Physics Textbook

Steven W. Stahler
Francesco Palla

WILEY-VCH

The Formation of Stars

Steven W. Stahler and Francesco Palla

The Formation of Stars

The Formation of Stars. Steven W. Stahler and Francesco Palla

ISBN: 3-527-40559-3

Steven W. Stahler and Francesco Palla

The Formation of Stars

WILEY-VCH

WILEY-VCH Verlag GmbH & Co. KGaA

Authors

Steven W. Stahler
University of California
Berkeley, USA
e-mail: stahler@astro.berkeley.edu

Francesco Palla
INAF-Osservatorio Astrofisico
di Arcetri
Florence, Italy
e-mail: palla@arcetri.astro.it

Cover Picture
T. A. Rector, B. Wolpa, M. Hanna;
AURA/NOAO/NSF:
The Rosette Nebula in Monoceros. Massive young
stars in the central cluster have cleared a hole
in the cloud. Ionizing radiation from these stars
causes surrounding gas to glow brightly.

Library of Congress Card No.: Applied for

British Library Cataloging-in-Publication Data:
A catalogue record for this book is available from the British Library

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available on the internet at <http://dnb.ddb.de>.

Printed in the Federal Republic of Germany
Printed on acid-free paper

Composition Uwe Krieg, Berlin
Printing Strauss GmbH, Mörlenbach
Bookbinding Litges & Dopf Buchbinderei GmbH, Heppenheim

ISBN 3-527-40559-3

Contents

III From Clouds to Stars

IV Environmental Impact of Young Stars

V Pre-Main-Sequence Stars

VI A Universe of Stars

Appendices

Preface

While there has long been theoretical speculation concerning the early life of stars, the subject first became an empirical one in the middle of the last century. Starting in the 1940s, the T Tauri class of objects, residing within dark clouds, was recognized and subsequently examined in considerable detail. This interest stemmed from the gradual realization that these peculiar variables represent a primitive phase of solar-type stars. It also became apparent that the observed objects must have condensed out of the dark clouds in which they are presently still found. By the mid 1950s, theorists began constructing numerical models for the pre-main-sequence evolutionary phase. The following decade saw advances in understanding the basic physics of cloud collapse.

The pace of discovery accelerated rapidly in the 1970s, largely as a result of new instrumentation. The advent of infrared astronomy allowed observers to peer behind the thick veil of obscuring dust and view even younger objects. Millimeter dishes, X-ray telescopes, and sensitive array detectors in the optical and near-infrared all had major impacts. Meanwhile, theoretical research kept apace, with studies of everything from chemical reaction networks in cloud environments to the interiors of the youngest stars. By the 1980s, star formation became one of the most vigorous areas of astronomical research.

There has been no sign of abatement in this activity. Indeed, parallel developments in other areas have underscored the central importance of the problem of star formation. Since the mid 1990s, observers have detected large numbers of giant planets encircling nearby stars. The properties of these systems range widely and call out for a more general understanding of the planet formation process. These bodies arise from dusty, circumstellar disks, which themselves appear during cloud collapse. Thus, a full account of planetary origins cannot ignore the early evolution of the central, stellar object.

Equally relevant are contemporary advances in cosmology. Numerical simulations of the last few decades follow the condensation of gas inside clumps of dark matter spread throughout the expanding Universe. This gas ultimately converts itself into stars. We can even track observationally the record of this transformation in both nearby and ancient, far-off galaxies. The pattern of star formation in a galaxy is a fundamental characteristic, one that correlates with, and in some measure determines, its observed structure. What had traditionally been viewed as a local phenomenon is now appreciated as a truly global one.

In the midst of such ongoing interest and progress, there is evidently a need to summarize the state of our knowledge and the pressing, unsolved questions. A number of excellent textbooks already cover the physics and chemistry of the interstellar medium, out of which stars are born. Equally well represented is the theory of stellar structure and evolution. Planetary science and galaxy formation have long been part of the standard curriculum. Still lacking, however, is a comprehensive treatment of the area touching all of these, the formation of stars themselves.

Any work attempting to encapsulate a rapidly evolving field must face the issue of timeliness. Won't any "facts" we present become outdated very quickly? The answer very much depends on the type of information conveyed. In the decade since we began this collaborative effort, it is certainly true that there have been many exciting discoveries. There is a core of

The Formation of Stars. Steven W. Stahler and Francesco Palla

ISBN: 3-527-40559-3

understanding, however, that has remained substantially intact. To put the matter succintly and somewhat glibly: Pictures and numbers indeed change rapidly; ideas do not.

This, then, is primarily a book about the ideas of star formation. Developing and illustrating these ideas has often required detailed exposition. This necessity, along with the sheer breadth of the field, has resulted in a much thicker volume than we originally anticipated. We ask the reader to be patient, trusting that the journey is worth some time and effort.

Two major hurdles confronted us at the very outset of the project. First was the daunting task of collecting a vast amount of research results on many diverse topics. As one collects, one also evaluates. Facts we had assumed were "well known" within the community often turned out, upon scrutiny, to be either wrong or to require significant qualification. Note also that we were not able to amass information through the time-honored method of constructing or borrowing from lecture notes in a star formation course. We stress that it was the very lack of such a course at our own and other institutions that provided an initial motivation for writing this work.

A second, and related, difficulty we faced was organizing all of this material in a logically compelling manner. Our solution here has been to group similar chapters into Parts. The ordering of these six Parts then dictates the overall flow of the narrative. We thus start from a general description of stars and their birth environments (Part I), before proceeding to a more detailed look at physical processes occurring within interstellar clouds (Part II). Part III spans the critical transition from clouds to stars. We describe the possible equilibrium configurations of clouds, and how these structures are disrupted through collapse. We also examine the primitive stars built up within the smallest collapsing entities. The profound thermal and mechanical effect of young stars on their surroundings is the subject of Part IV. In Part V, we see how stars, once divested of their cloud gas, evolve to maturity. Finally, the two chapters of Part VI return to a larger-scale view. Chapter 19 describes star formation activity in local galaxies and beyond, while Chapter 20 summarizes briefly our progress toward understanding key issues in the field.

We have aimed this book nominally at the level of graduate students in physics and astronomy. The complete text contains far more material than can be digested in a single semester. As a guide to the instructor, we list here the chapters and sections we consider most essential. We also suggest the time that could be devoted to each Part, assuming a 15-week course with 3 lecture hours per week.

Part	Weeks	Chapter/Sections
I	2	1; 2; 3; 4.1–4.3, 4.5
II	2	5.1–5.3; 6.1–6.2; 7; 8.1–8.2
III	4	9.1, 9.4–9.5; 10.1–10.2, 10.4; 11.1–11.3, 11.5; 12.1–12.3
IV	2	13.1–13.2; 14.1–14.2; 15.1–15.2,15.5
V	4	16.1–16.4; 17.1–17.3,17.5; 18.1–18.2, 18.4–18.5
VI	1	19.1–19.3; 20

Our book should also serve as a reference for professional researchers. This latter group will observe that we have *not* simply compiled all of the currently popular models in each topic. We have instead selected interpretations that fit best into the broad story of star formation, as we understand it. Some aspects of this story will undoubtedly change in the coming years. We nevertheless feel that anyone trying to master the subject is best served by a unified, coherent presentation. Whoever wishes to follow more closely the underlying debates or to acquire historical perspective on any topic will want to consult key research and review articles, many of which are cited as Suggested Reading at the end of each chapter.

As preparation, the prospective reader should have a solid background in physics at the undergraduate level. We do not assume similar training in astronomy. Many basic astronomy results, as well as much of its terminology, are presented within the text. Indeed, the study of star formation itself provides a good introduction to astrophysical concepts, simply because of the great variety of topics embraced by the field. In this context, the student should be comfortable with both theoretical arguments and observational results. Note that we have not shied away from describing observations for which no adequate model exists. To our minds, these results are among the most interesting, as they represent areas where fundamental progress can be made.

Over the years, we have received generous help and advice from many individuals. It is a pleasure to acknowledge them. We are indebted to Amanda McCoy for her tireless effort in producing hundreds of figures. Kevin Bundy read the entire manuscript and offered invaluable commentary on both the scientific content and the manner of its presentation. Eric Feigelson used a preliminary version of the book in a one-semester course; we are grateful to Eric and his students at Penn State for their comments and corrections. Others who reviewed selected chapters include Gibor Basri, Peter Bodenheimer, Jan Brand, Charles Curry, Daniele Galli, Dave Hollenbach, Richard Larson, Gary Melnick, Karl Menten, Mario Pérez, Steve Shore, and Hans Zinnecker. We have also benefited from discussions with Philippe André, Leo Blitz, Paola Caselli, Riccardo Cesaroni, Tom Dame, George Herbig, Ray Jayawardhana, Chung-Pei Ma, Thierry Montmerle, Antonella Natta, Sean Matt, Maria Sofia Randich, Leonardo Testi, Ed Thommes, Malcolm Walmsley, and Andrew Youdin.

Part I

Star Formation in Our Galaxy

The Formation of Stars
Steven W. Stahler and Francesco Palla
© 2004 WILEY-VCH Verlag GmbH & Co.

1 Overview

The complex of processes known as star formation must have occurred innumerable times in the remote past. The Big Bang, after all, did not produce a Universe full of stars but of diffuse gas. How gas turns into stars is the subject of this book. Anyone wishing to study the problem is aided immeasurably by the fact that star formation is also occurring now, and in regions close enough that the transformation can be examined in some detail. Indeed, most research activity in this field consists of our bold, if frequently misguided, attempts to interpret what is in fact happening all around us.

We begin, therefore, with the data. The four chapters of Part I describe various aspects of star formation activity in our own Galaxy. We discuss the properties of the gas in interstellar space, the structure of clouds that produce stars, and the morphology of young stellar groups. The treatment here is quite broad, since all these topics will be revisited later. Our very first task is to introduce the reader to the primary objects of interest, young stars themselves, and to the environments in which they are born. We start with a descriptive tour of two relatively nearby regions, before proceeding to a more quantitative and physical description of stars and their evolution.

1.1 Stellar Nurseries: Orion

The figure of Orion the Hunter is a familiar sight in the winter sky of the Northern hemisphere. It is one of the most easily recognized constellations and includes one tenth of the 70 brightest stars. Less familiar, perhaps, is the fact that this area is an extraordinarily active site of stellar formation. Over the years, no similar region has received such intense astronomical scrutiny, nor been studied with such a variety of observational tools. We refer the reader to the sky map of Figure 1.1. Here, some of the more conspicuous of the constellation's members are indicated, including the red supergiant Betelgeuse at the Hunter's right shoulder, and brilliant blue Rigel at his left foot. South of the three stars that comprise the belt of the Hunter is a bright, fuzzy patch. This is the Orion Nebula, a cloud of gas being heated by the intense radiation of the Trapezium stars embedded within it.

1.1.1 Giant Molecular Cloud

Young stars like those of the Trapezium are forming out of a huge body of gas known as the Orion Molecular Cloud. The extent of this object is indicated by the shading in Figure 1.1. In its longest dimension, the cloud covers 15° in the sky, or 120 pc at the distance of 450 pc.[1] Orion is but one of thousands of *giant molecular clouds*, or *cloud complexes*, found throughout the Milky Way. The gas here is predominantly molecular hydrogen, H_2. With their total masses of

---

[1] The reader unfamiliar with units commonly employed in astronomy, such as the parsec (pc), should consult Appendix A.

The Formation of Stars. Steven W. Stahler and Francesco Palla

ISBN: 3-527-40559-3

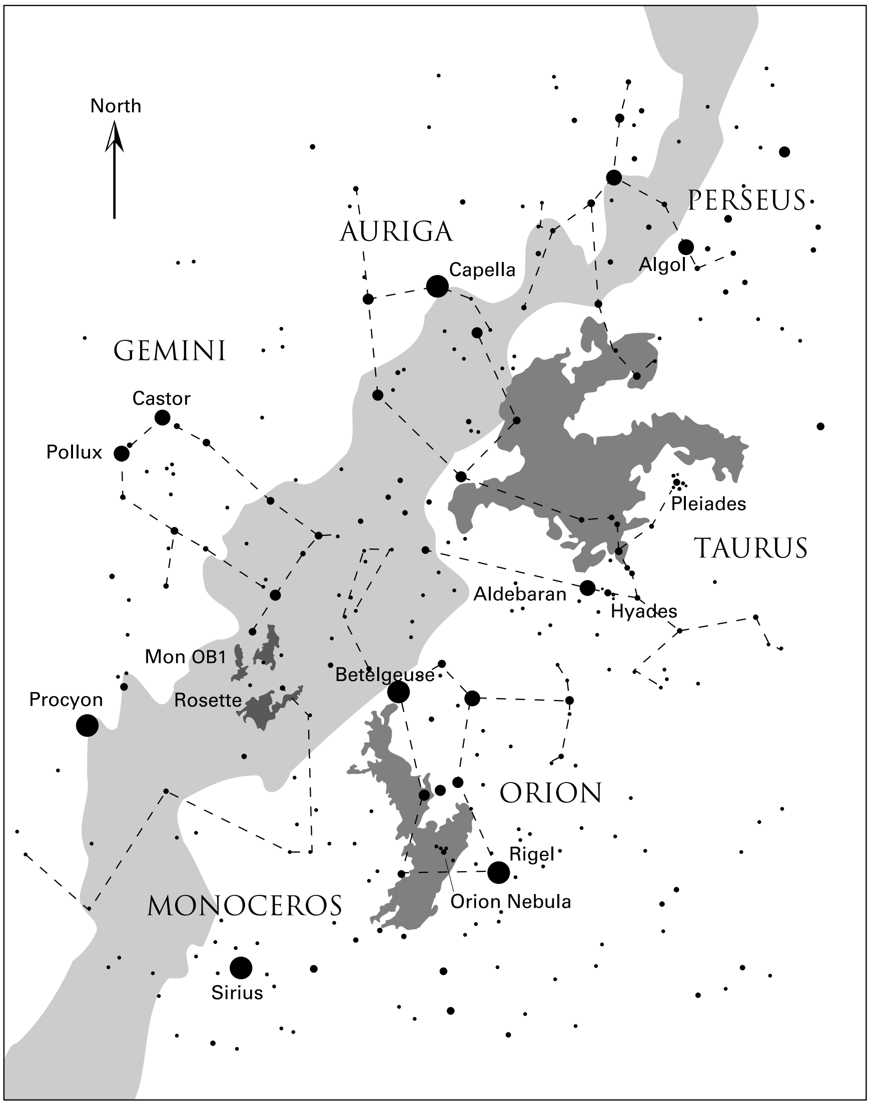


Figure 1.1 A portion of the Northern sky. The Milky Way is depicted as light grey, while the darker patches indicate giant molecular clouds. Also shown, according to their relative brightness, are the more prominent stars, along with principle constellations.

order $10^5\ M_\odot$, these structures are the largest cohesive entities in the Galaxy and are almost all producing new stars.

The fact that we know of molecular clouds at all is a triumph of radio astronomy. Gas in these regions is much too cold to radiate at visible wavelengths, but may be detected through its radio emission in trace molecules such as CO. Here observers most often rely on single dishes that can map extended areas of the sky. For more detailed studies of individual regions, one may utilize the fact that several dishes linked together effectively increase the detector area and hence the angular resolution. Such *interferometers* have been powerful research tools, especially for studying the distribution of matter around newly formed stars.

The left panel of Figure 1.2 is a high-resolution CO map of the entire Orion molecular cloud. In this case, the observations were made with a relatively large single-dish telescope. The spectral line being detected is the commonly used 2.6 mm transition of the main isotope, ^{12}C^{16}O. We have distinguished in the figure two major subunits, labeled Orion A and B. Both the elongated shape of the whole complex and its high degree of clumpiness are generic features of such structures.

Along with their gas, molecular clouds contain an admixture of small solid particles, the *interstellar dust grains*. These particles efficiently absorb light with wavelengths smaller than their diameters (about 0.1 μm) and reradiate this energy into the infrared. Regions where the dust effectively blocks the light from background stars are traditionally known as *dark clouds*. Generally, these represent higher-density subunits within a flocculent cloud complex, although they can also be found in isolation. Note that the extinction due to dust depends on its *column density*, *i. e.*, the volume density integrated along the line of sight. Figure 1.3 depicts the major dark clouds in Orion, determined by tracing the regions of strong obscuration in optical photographs. A number of the most prominent structures, such as L1630 and L1641, are labeled by their designations in the Lynds cloud catalogue. The shaded areas, including those with NGC numbers, are chiefly *reflection nebulae*, *i. e.*, dusty clouds that are scattering optical light from nearby stars into our direction.[2]

The mid- and far-infrared emission from warm dust particles provides yet another means to study the Orion region. The first instrument devoted exclusively to infrared mapping of the sky was IRAS (for *Infrared Astronomical Satellite*), launched in 1983. Figure 1.4, which spans the same angular scale as the previous two figures, shows the Orion Molecular Cloud as a composite of three monochromatic IRAS images taken at 12, 60, and 100 μm. Radiation in this spectral regime comes mainly from dust heated to roughly 100 K. The fact that even this modest temperature is maintained over such an extended region demands the presence of many stars of high intrinsic luminosity.

Returning to the ^{12}C^{16}O map of Figure 1.2, we see that several areas associated with reflection nebulae have closed, nested contours, indicating a local buildup in radio intensity. The received intensity in ^{12}C^{16}O is correlated with the hydrogen column density, so that such buildups mark the presence of embedded clumps. The 2.6 mm transition is most readily excited by gas with number density near 10^3 cm^{-3} and is relatively weak at higher values. However, other tracers are available to explore denser regions. The inset in Figure 1.2 is a map of Orion B in the 3.1 mm line of CS, a transition excited near 10^4 cm^{-3}. Here, most of the individual fragments have sizes of about 0.1 pc and inferred masses near 20 $M_\odot$, while the few largest ones have masses ten times as great. Such localized peaks within the broad sea of molecular cloud gas are known as *dense cores* and are the actual sites of star formation.

The stars being born within dense cores shine copiously in optical light, but none of this short-wavelength radiation can penetrate the high column densities in dust. As before, however, the same dust can be heated to emit at wavelengths that *can* escape. The shaded portions of the Figure 1.2 inset show regions detected at 2.2 μm. What is being seen, in fact, are several

---

[2] The *NGC* designation is historical and refers to the “New General Catalogue” of nebulous objects, dating from the 19th century.

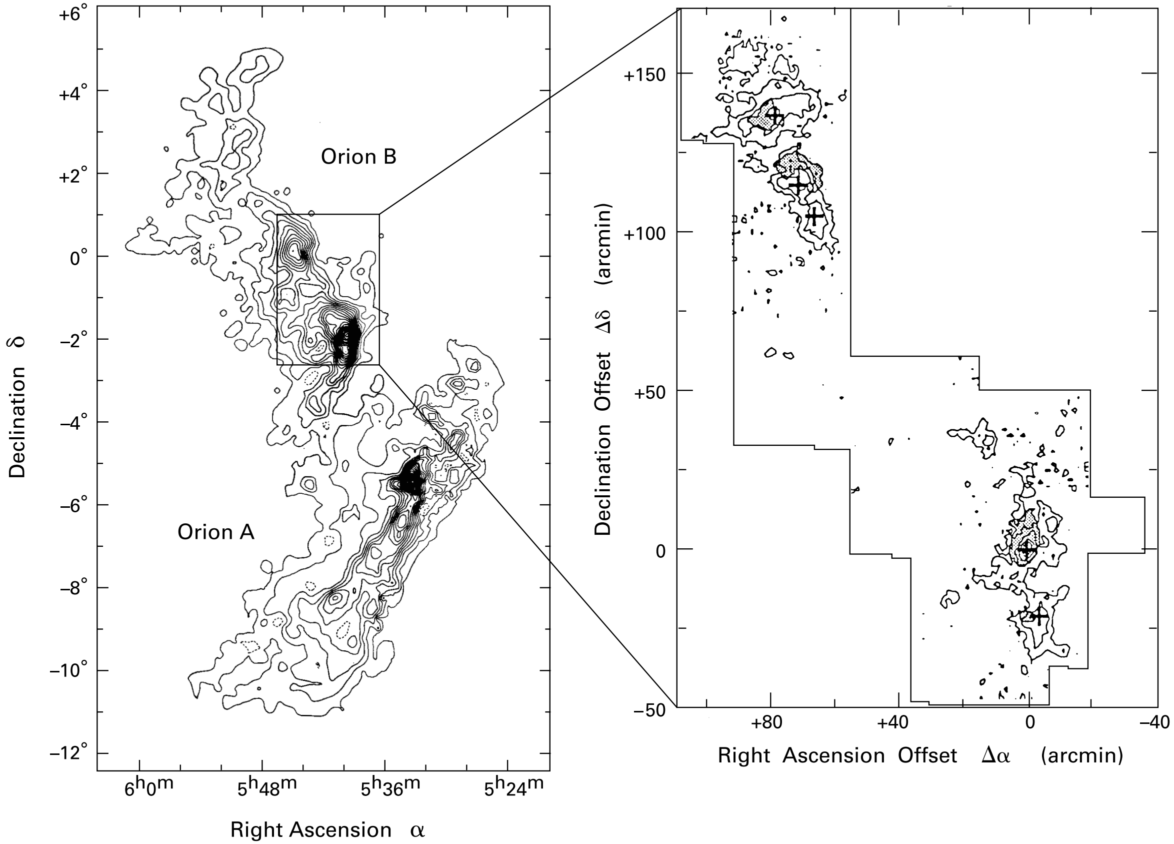


Figure 1.2 Map of the Orion Molecular Cloud in the 2.6 mm line of $^{12}C^{16}O$. See Appendix A for an explanation of the coordinates used here and elsewhere in the book. The insert shows a detail of Orion B in the 3.1 mm line of CS. Shading within the insert marks those regions which emit strongly at 2.2 μm. The (0,0) point coincides with the position of the reflection nebula NGC 2024. This nebula and others are indicated by crosses.

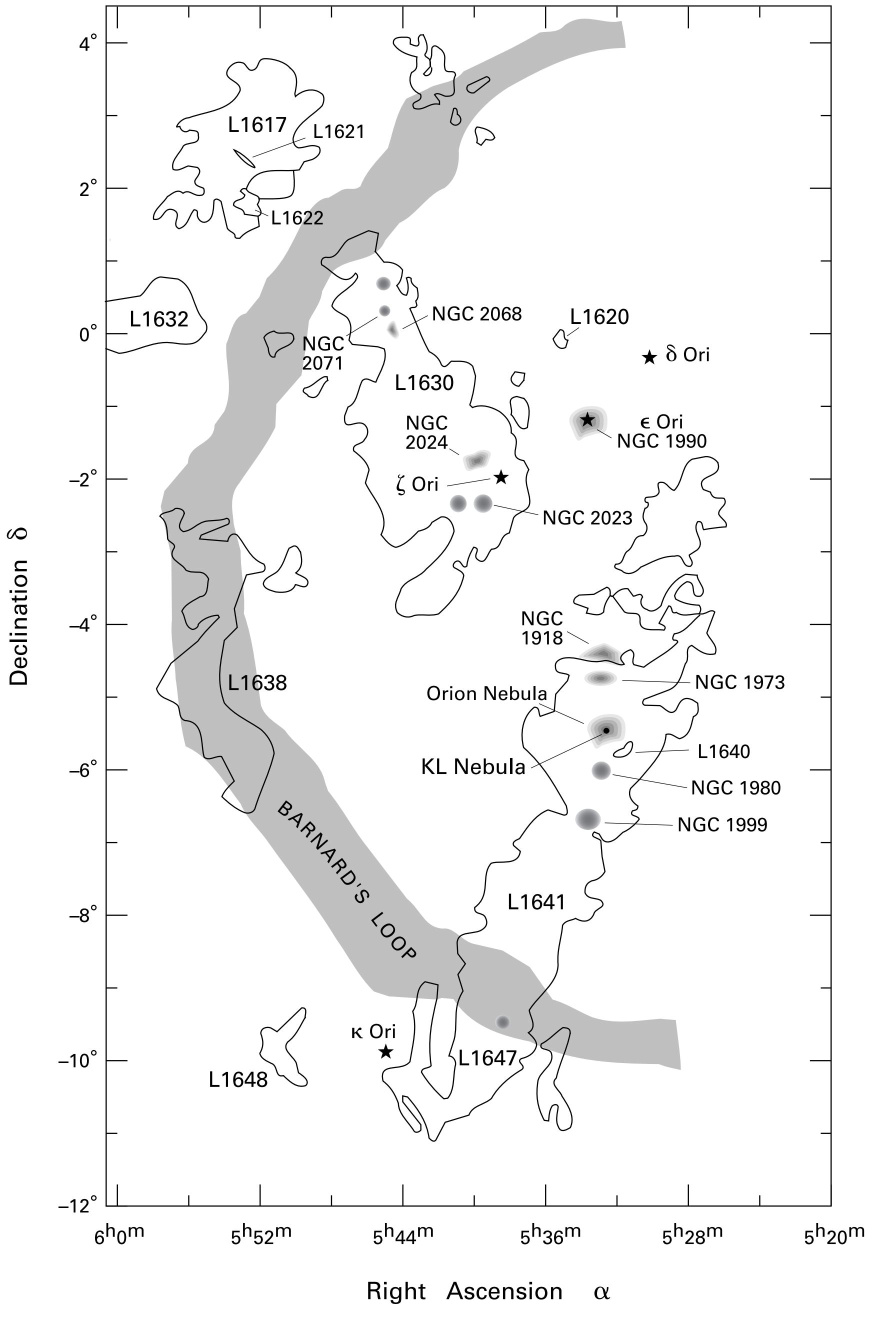


Figure 1.3 Dark clouds in Orion. The large swath is Barnard's loop, a diffuse region of enhanced optical emission. The labeled dark patches are reflection nebulae. Major clouds are also labeled by their Lynds catalogue numbers.

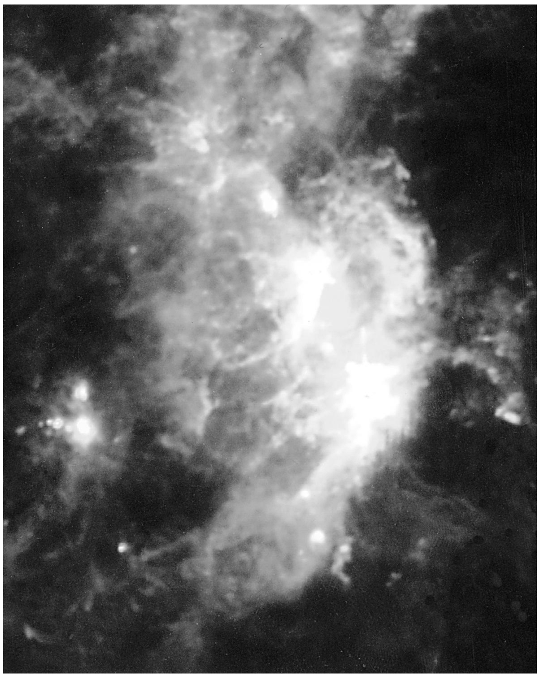


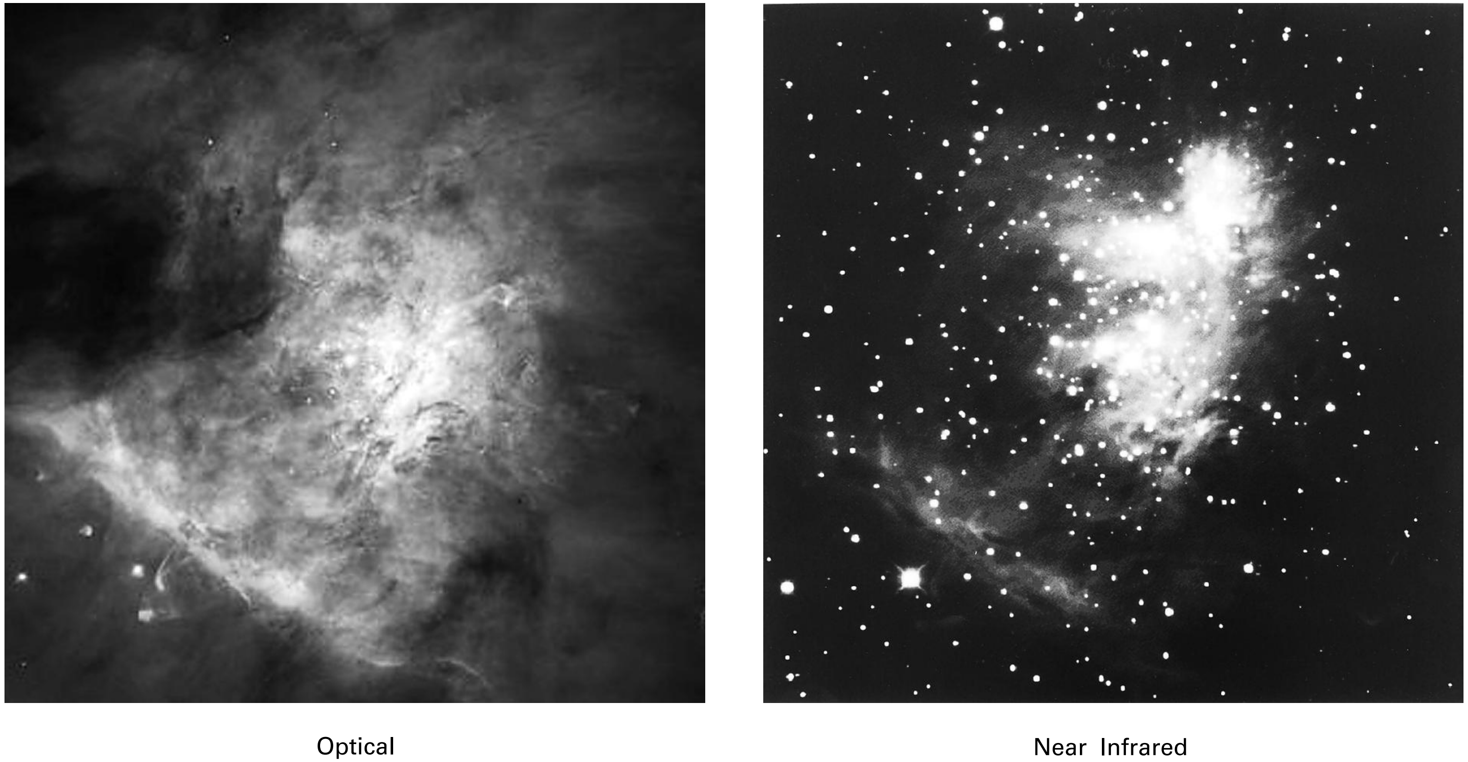
Figure 1.4 Infrared view of the Orion Molecular Cloud. This is a composite of three monochromatic images at 12, 60 and 100 μm.

embedded stellar *clusters*, compact groups containing tens to hundreds of members. Each cluster is associated with one of the more massive bodies of molecular gas, and virtually all cluster members are still nested within their parent dense cores. Thus, we see that stars form in giant molecular clouds at localized, massive peaks in the gas, and predominantly in a cluster mode, rather than in isolation.

1.1.2 Orion Nebula and BN-KL Region

To the south of Orion B, the Orion A cloud consists of a similar clumpy distribution of molecular gas. Within one elongated, high-density region is the famous Orion Nebula, also designated NGC 1976 or M42, the latter nomenclature referring to the 18th century Messier catalogue. As shown in the optical photograph of Figure 1.5 (*left panel*), the nebula is a turbulent expanse of gas, lit up by an embedded stellar cluster. The conspicuous ridge at the bottom of the photograph is the Orion Bar, whose emission also stems from gas heated and ionized by the cluster stars. This *ionization front* is seen edge on and is especially well defined because of the cool, dusty region just beyond it. The cluster responsible for such energetic activity is the Ori Id OB association, one of several small groups of massive stars in the giant complex. Near the center of the figure are the four stars of the Trapezium, whose most prominent member, the O star θ^1 Ori C, has a luminosity of $4 \times 10^5\ L_\odot$ and a surface temperature of 4×10^4 K.

Stars this hot emit most of their energy in the ultraviolet and are thus capable of ionizing hydrogen gas out to considerable distances. The Orion Nebula, which is about 0.5 pc in diameter, is the best studied example of such an *HII region*. Within the ionized plasma, the gas temperature is comparable to that at the surface of the exciting star. As electrons and nuclei recombine, the atoms emit a plethora of spectral lines, including optically visible radiation from



Optical

Near Infrared

Figure 1.5 *left panel*: Optical photograph of the Orion Nebula, with the Trapezium stars at the center. *right panel*: Infrared (2.2 μm) image of the same area. The BN-KL region lies at the upper right.

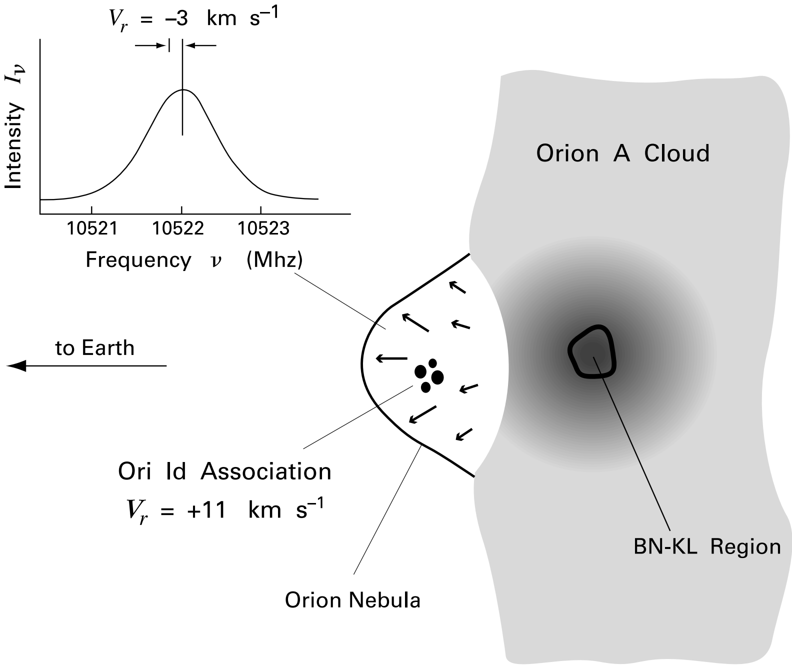


Figure 1.6 Expansion of the Orion Nebula (*schematic*). The insert depicts the blueshifted spectrum of the $n = 86 \to 85$ hydrogen recombination line, which has a rest-frame frequency of 10522.04 MHz.

both hydrogen and heavier elements. Here, these lines can be detected because the nebula sits near the edge of the Orion A cloud, as depicted in Figure 1.6. While the ionization created by the OB association slowly eats its way into the cloud, hot matter on the other side streams out into the more rarefied gas adjacent to the cloud. As sketched in the figure, this streaming motion in the direction of the Earth is evident from the Doppler shift toward the blue seen in hydrogen recombination lines. The ionized gas is approaching the Earth at a velocity of 3 km s^{-1}, while the Trapezium stars themselves are *retreating* at 11 km s^{-1}.

O and B stars, although intrinsically luminous, are produced only rarely within molecular clouds. Much more frequently, gas condenses into low-mass stars, *i. e.*, those of about $1\ M_\odot$ or less. Thus, the Ori Id association is merely the tip of the iceberg in terms of stellar production, as sensitive optical and infrared surveys have now made clear. The O and B stars lie in the center of the populous Trapezium Cluster (also called the Orion Nebula Cluster). The right panel of Figure 1.5 shows over 500 stars in this region, as imaged at 2.2 µm. The photograph covers about $4' \times 4'$ in angular extent, or 0.6×0.6 pc. With its central number density in stars exceeding 10^4 pc^{-3}, the Trapezium Cluster is among the most crowded star formation regions in the Galaxy.

The near-infrared image contains a bright region in the upper right that is invisible optically. This stellar cluster is actually located about 0.2 pc *behind* the Orion Nebula, in a portion of the cloud complex designated OMC-1 (see Figure 1.6). Here reside two mysterious and powerful sources of infrared radiation, the Becklin-Neugebauer (BN) Object and IRc2, each emitting

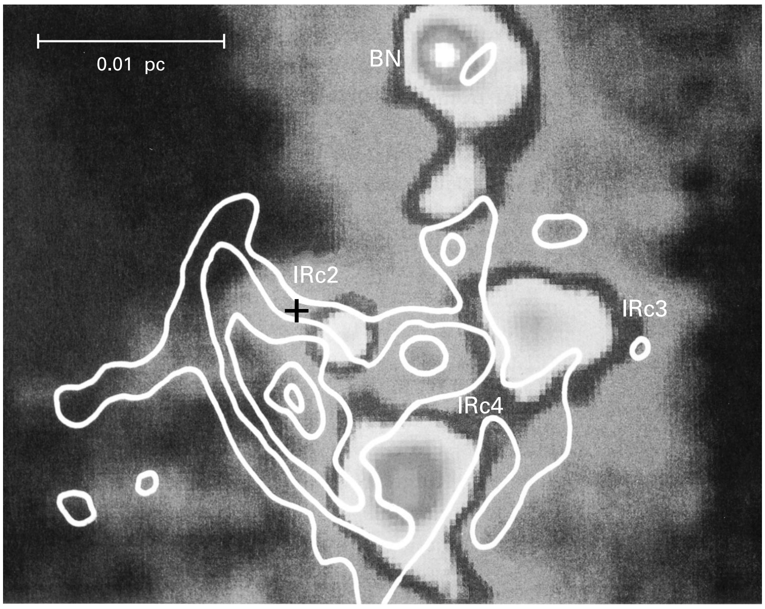


Figure 1.7 Mid-infrared (19 μm) continuum radiation from the KL Nebula (*greyscale*). Prominent infrared objects are labeled. The contours show emission from the 1.3 cm line of NH_3.

luminosities from 10^3 to 10^5 $L_\odot$. Both are part of the larger and more diffuse Kleinman-Low (KL) Nebula, which is about 0.1 pc in diameter. The greyscale image of Figure 1.7, representing emission at 19 μm, shows the area in more detail. To supplement this picture of the dust content, the superposed contours trace 1.3 cm radiation from heated NH_3 molecules. The NH_3 radio line is strongest where the dust emission is weak or absent. Since the molecule's emission is sensitive to the ambient temperature, this pattern suggests that the gas has been heated to the point where dust is thermally destroyed.

The BN-KL region is producing a massive *molecular outflow*, *i. e.*, high-velocity cloud gas streaming away from the infrared stars. The outflow phenomenon, ubiquitous in star formation regions, was first discovered here through detections of the Doppler shift in the $^{12}C^{16}O$ line. Observations of numerous other molecular species reveal that the impact of this wind on nearby gas has resulted in shock heating, which alters the pattern of chemical abundances in a characteristic manner. Also seen in the vicinity is copious near-infrared emission from H_2, resulting from collisional excitation of the molecule in shock fronts. Finally, the BN-KL region contains numerous *interstellar masers*, small regions of strongly beamed radiation from molecules such as H_2O and SiO. The measured intensity of these spots is vastly greater than that normally emitted at the ambient temperature of roughly 100 K. The maser phenomenon is yet another manifestation of wind-induced shocks in this extraordinarily active region.

1.2 Stellar Nurseries: Taurus-Auriga

Returning to Figure 1.1, we now proceed northwest from Orion, *i. e.*, along the direction of the Hunter's belt. We soon encounter the constellation Taurus (the Bull). Taurus is notable for the

Figure 1.8 Dark clouds in Taurus-Auriga. The large patches in the lower left and middle right correspond to TMC-1 and L1495 in Figure 1.9.

bright orange star Aldebaran, as well as for the Hyades, the V-shaped group of stars marking the Bull's face, and the Pleiades group riding his shoulder. Both the Hyades and Pleiades are nearby, young stellar clusters that continue to furnish valuable information for evolutionary studies. We are more interested, however, in an even younger region to the north, extending into the neighboring Auriga Constellation. Here, as in Orion, molecular clouds are actively producing a multitude of new stars.

1.2.1 Dark Clouds

The clouds of Taurus-Auriga, indicated as the shaded area of Figure 1.1, have long been noted even in optical images. Figure 1.8 is a photograph from early last century by E. E. Barnard, covering a region of about 50 square degrees. Here one sees prominent dark lanes in the otherwise rich stellar field. Referring to this photograph in his 1927 atlas of the Milky Way, Barnard wrote

> Very few regions of the sky are so remarkable as the Taurus region. Indeed, the photograph is one of the most important of the collection, and bears the strongest proof of the existence of obscuring matter in interstellar space.

Even more convincing – indeed, definitive – proof of interstellar dust was to come several years later, with J. Trumpler's demonstration of the progressive reddening of distant clusters.

The molecular gas accompanying the obscuring dust can be seen most readily in $^{12}C^{16}O$, as shown in Figure 1.9. The Taurus-Auriga region covers a greater angular area than the Orion

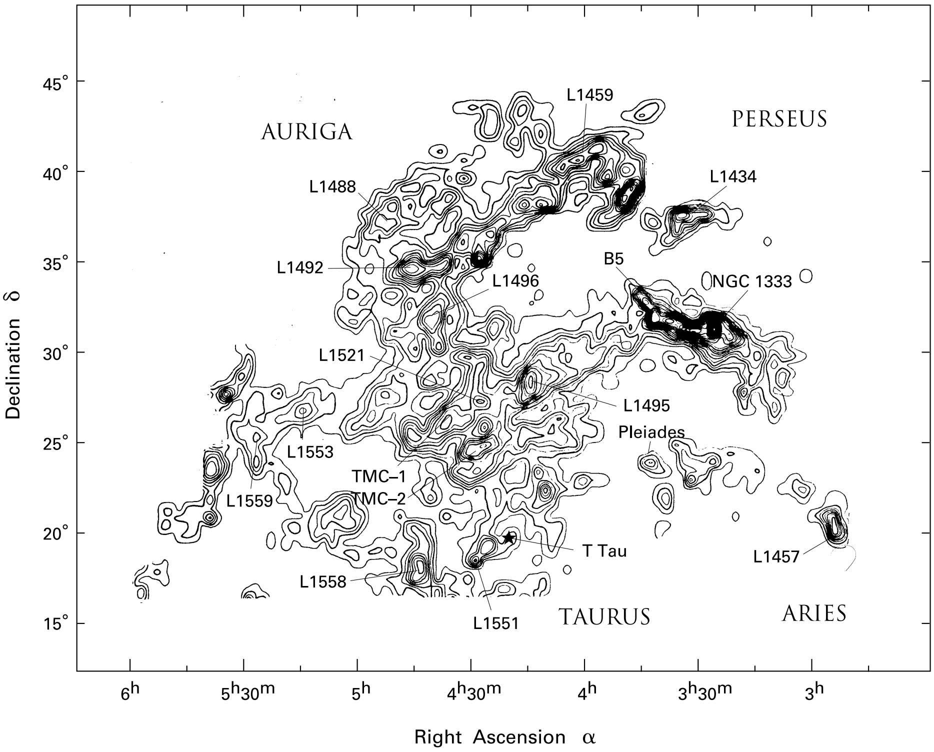


Figure 1.9 Map of Taurus-Auriga in the 2.6 mm line of $^{12}C^{16}O$. Prominent dark clouds are labeled.

Molecular Cloud (see Fig. 1.1), but only because it is closer. At a distance of 140 pc, the entire region measures about 30 pc in linear extent. As before, we have labeled some of the main dark clouds, which show up as local peaks in the CO emission; Barnard's optical image is centered on L1521 and L1495. The dense core TMC-1 has been closely studied for its wealth of interstellar molecules. Shown also in Figure 1.9 is the famous star T Tauri, to which we shall return shortly. Note finally that the elongated structure to the north, including L1459 and L1434, and the high-intensity area to the west containing Barnard 5 (B5) and NGC 1333, are not physically associated with the Taurus-Auriga system, but are more distant, massive clouds seen in projection.

In passing from Orion to Taurus-Auriga, we have not only diminished the physical scale of the star forming region but also the mode of stellar production. The gas in Taurus-Auriga is *not* part of a giant complex; the total mass in molecular hydrogen is of order $10^4 \, M_\odot$. Nor is the region forming high-mass stars, with their attendent reflection nebulae and HII regions. Finally, the existing low-mass stars are nowhere clustered as densely as in the Trapezium or even the less populous groups of Orion B.

Let us now focus on the dark clouds of the region, using two different tracers. The upper panel of Figure 1.10 is a map of the central portion of Taurus-Auriga, covering TMC-1, TMC-2 and the dark cloud L1495 in Figure 1.9. The image shows emission from $^{13}C^{16}O$. This isotope

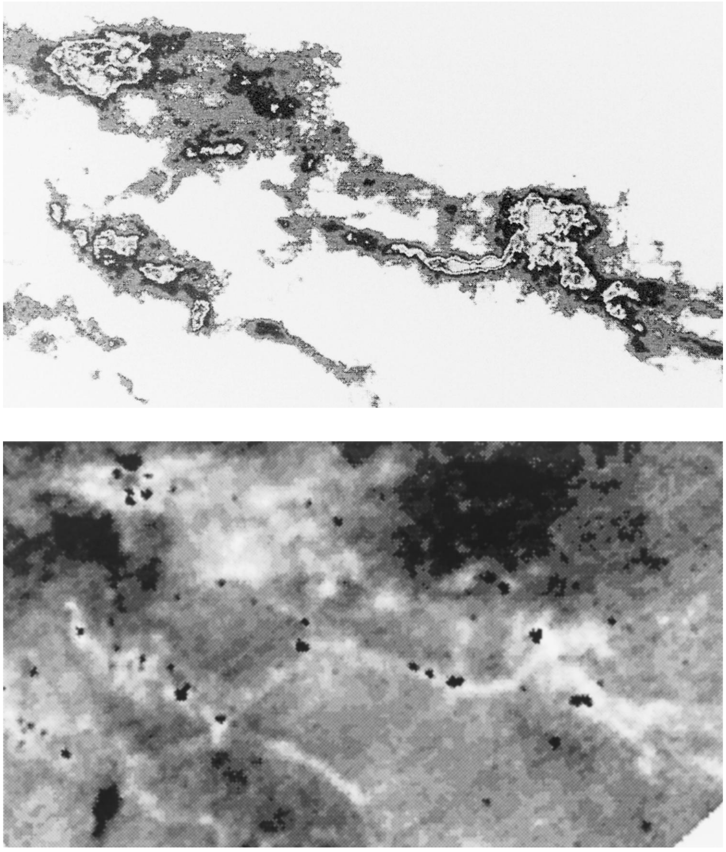


Figure 1.10 *upper panel*: Map of Taurus-Auriga in the 2.7 mm line of $^{13}C^{16}O$. *lower panel*: The same region as seen by IRAS. Here the brighter regions mark concentrations of relatively cold dust.

is excited at the same volume density as the more common $^{12}C^{16}O$. However, it effectively highlights regions of greater *column* density, which tend to trap the radiation from the more abundant species. Notice that the dark clouds have an elongated appearance, not unlike those in Orion (Figure 1.3).

As before, the infrared emission from warm dust grains is another useful tracer of molecular gas. The lower panel of Figure 1.10 is an IRAS map of the same region as in the upper panel. What is shown here is the ratio of the fluxes at 100 μm and 60 μm, a measure which emphasizes the contribution from the colder (near 30 K) grains. There is evidently an excellent match between the distributions of this dust component and the gaseous structures traced in $^{13}C^{16}O$.

1.2.2 T Association

The dark clouds are also the birthplaces for young stars, as seen in Figure 1.11, which is a composite of radio ($^{12}C^{16}O$), infrared, and optical data. For the stars, we have distinguished

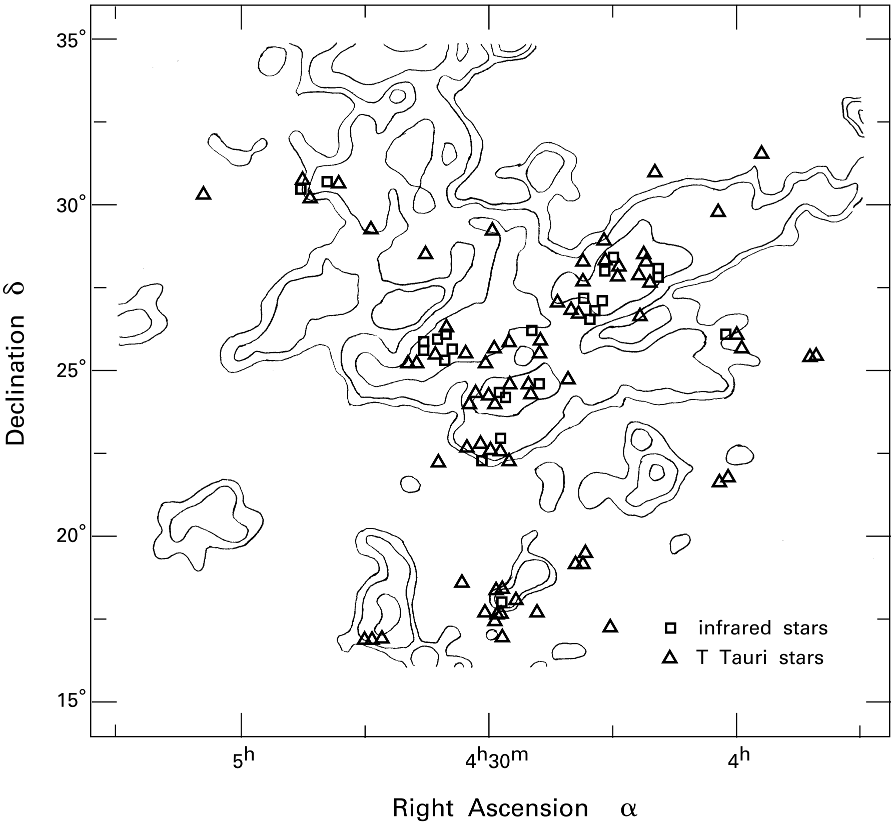


Figure 1.11 Central region of Taurus-Auriga in $^{12}C^{16}O$, along with the locations of infrared stars (*squares*) and optically visible T Tauri stars (*triangles*).

deeply embedded sources seen here in the infrared from optically visible, low-mass objects. The latter belong to a class known as *T Tauri stars*, named for the prototypical object indicated in Figure 1.9. Stars of the T Tauri class are but one type of *pre-main-sequence* star. In regions like Orion, one also finds more massive pre-main-sequence objects known as *Herbig Ae/Be stars*. Finally, we have seen that Orion (but not Taurus-Auriga) contains a number of O and B stars, which are *main-sequence* objects. Note that a given source is classified as a T Tauri or Herbig Ae/Be star based on specific observational criteria, while the pre-main-sequence and main-sequence designations refer to the presumed evolutionary state of the star.

It is evident from Figure 1.11 that both the infrared and optically visible young stars of Taurus-Auriga are confined to the denser molecular gas. The same is true for Orion, but the stellar distributions are quite different. The Taurus Molecular Cloud harbors a *T association*, in which stars are spread out more uniformly. Although some degree of clumping is evident, we do not see the extreme crowding of Orion. Recall, for example, that the entire Trapezium Cluster shown in Figure 1.5 has a diameter of about 0.4 pc, while each of the sparser groupings in Figure 1.10 extends over about ten times that length.

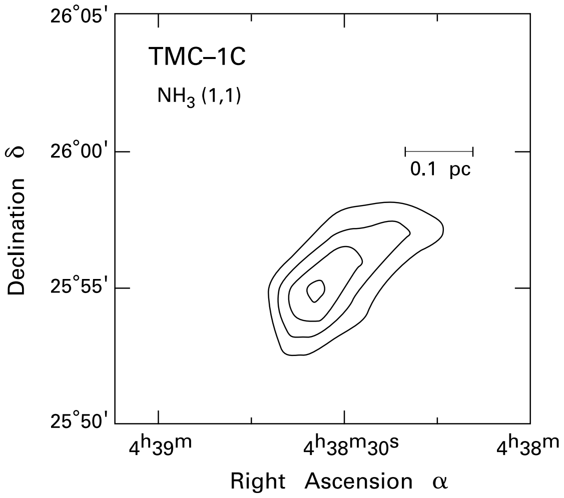


Figure 1.12 The dense core TMC-1C, as traced by the 1.3 cm line of NH_3.

The relative isolation of the young stars in Taurus-Auriga, together with the proximity of the region, have allowed more detailed study of individual stellar formation than in Orion. In particular, molecular transitions sensitive to higher densities than CO have been used effectively to examine the region's dense cores. Figure 1.12 is a radio map of TMC-1C, one of several subcondensations within TMC-1. The observations here were made in the 1.3 cm line of NH_3, which traces gas number densities near 10^4 cm^{-3}. In central density, linear extent, and total mass, the Taurus-Auriga dense cores are similar to most of those traced by CS in Orion, but no very massive fragments are detected.

The dense core shown in Figure 1.12 does not contain a young star, but over half of those observed in NH_3 have infrared point sources in their central regions. These embedded objects represent an earlier evolutionary stage than the visible T Tauri stars. Interestingly, most of these infrared stars are associated with molecular outflows. Figure 1.13 is a CO map of the dark cloud L1551, which the reader may locate in Figure 1.9. The striking bipolar lobes represent cloud gas being dragged to great distances from the central infrared source, which in this case is an embedded binary pair known collectively as IRS 5. These stars, with a total luminosity of less than 30 $L_\odot$, have a combined mass under 2 $M_\odot$. In contrast, a high-mass star or group of stars is creating the outflow in the Orion BN-KL region. Discovered in 1980, the L1551/IRS 5 system was the first detected low-mass outflow, of which hundreds are now known. Indeed, the high frequency of outflow observations probably indicates that *every* star turns on a powerful wind before it is optically visible.

1.3 Stars and Their Evolution

The foregoing examples demonstrate amply that there is no difficulty producing a variety of stars in the present-day Galaxy, given the right conditions. Of course, part of our task is to elucidate what those "right conditions" are. Another goal must be to discern the sequence of events by which a rarified cloud actually condenses to the stellar state. Clearly, both efforts should be informed by a working knowledge of stars as physical objects and of the conceptual tools that have been developed for their study.

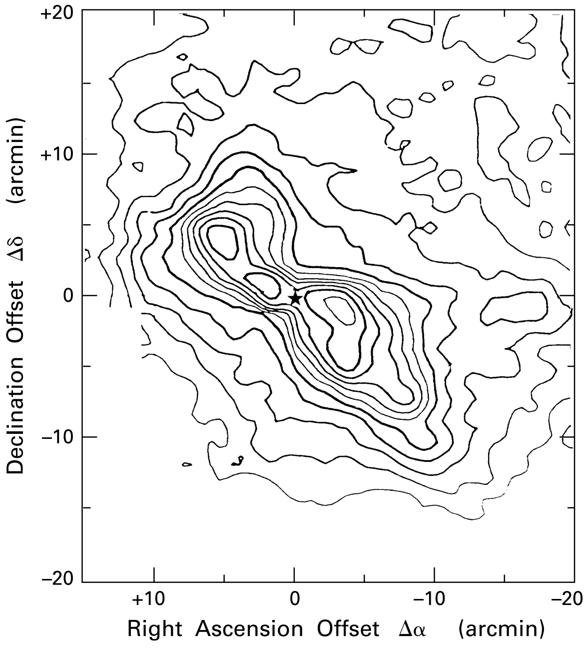


Figure 1.13 Map in ^{12}C^{16}O of the L1551 cloud in Taurus-Auriga. The star marks the position of the infrared source IRS 5.

1.3.1 Basic Properties

Virtually all the information we have about a star comes from the electromagnetic radiation it emits. Photons from the deeper layers break free and stream outward into space at the *photosphere*. The two most basic properties of the star are its luminosity, *i. e.*, the total energy emitted per unit time, and the temperature at the photosphere. When our focus is on the physics of stars, we will denote the luminosity as L_*, while the designation L_{bol} will be reserved for discussions of the emitted radiation. The latter notation emphasizes that we are interested in the *bolometric*, or total, luminosity, rather than that within a specific wavelength range.[3] For the second variable, it is conventional to use the *effective temperature* T_{eff}, *i. e.*, the temperature of an equivalent blackbody of the same radius (see Chapter 2).

If the distance to the star is somehow known, L_{bol} can be obtained in principle by flux measurements over a sufficiently broad wavelength range. The temperature T_{eff}, on the other hand, must always be deduced through theoretical modeling. Accordingly, astronomers often prefer to characterize stars through two related quantities that are more easily measured. In place of L_{bol}, one may use the logarithmic quantity M_V, the *absolute magnitude* of the star in the V (for *visual*) band, a relatively narrow wavelength interval centered on 5500 Å. Suppose F_V is the flux received in this band, *i. e.*, the energy in a unit wavelength interval per unit area per time. This flux is weaker for objects that are farther away. To measure the *intrinsic* brightness, we imagine the star of interest to be located at some fixed distance, conventionally taken to be

---

[3] We shall also use the symbol L_* when referring to the observationally estimated luminosity from the star alone, excluding any additional contribution from circumstellar matter. The quantity L_* is numerically identical to L_{bol} for mature stars that lack such material, but may differ in younger, more embedded objects.

10 pc. Then M_V is defined as

$$M_V \equiv -2.5 \log F_V(10\,\mathrm{pc}) + m_{V\circ} \ , \tag{1.1}$$

where $m_{V\circ}$ is a constant. Note that, by definition, fainter stars have numerically greater magnitudes. The V band used here is one of the standard Johnson-Morgan sequence of filters, which also include ultraviolet (U at 3650 Å) and blue (B at 4400 Å) wavebands. As detailed in Appendix A, another filter set, designated $R, I, J, ...Q$, extends into the infrared. Also used is the Strömgren four-color system. Here the filter sequence u, v, b, and y covers 3500 Å to 5500 Å.

Written in terms of an arbitrary wavelength λ, equation (1.1) becomes

$$M_\lambda \equiv -2.5 \log F_\lambda(10\,\mathrm{pc}) + m_{\lambda\circ} \ , \tag{1.2}$$

This equation can further be generalized to cover the case where the star is located, not at 10 pc, but at an arbitrary distance r. The new equation then defines a distance-dependent quantity known as the *apparent magnitude*:

$$m_\lambda \equiv -2.5 \log F_\lambda(r) + m_{\lambda\circ} \ . \tag{1.3}$$

Since the flux declines as r^{-2}, the apparent and absolute magnitudes are related by

$$m_\lambda = M_\lambda + 5 \log \left(\frac{r}{10\ \mathrm{pc}} \right) \ , \tag{1.4}$$

where the term added to M_λ is the *distance modulus*.

To find a surrogate quantity for T_{eff}, we take advantage of the fact that a star's surface temperature is correlated with its color. The latter can be quantified by forming the ratio of flux in the V band to some other, which is conventionally chosen as the B band. Equivalently, we define the $B - V$ color index as $M_B - M_V$. Notice, from equation (1.4), that this quantity is also equal to $m_B - m_V$. We denote by $(B - V)_\circ$ the *intrinsic* color index. Here, the fluxes are not those directly observed, but calculated under the assumption that the path to the star is free of dust. These particles redden starlight and alter the apparent color (see § 2.3). The intrinsic index at two arbitrary wavelengths λ_1 and λ_2 is written C°_{12}. By convention, λ_1 is less than λ_2.

The color index is a measure of surface temperature based on *photometric* observations, *i. e.*, those employing broadband filters. Another useful indicator is the *spectral type*. Astronomers evaluate this quantity by observing the relative strengths of sharp absorption lines in the spectrum. The sequence of spectral types in order of descending T_{eff} is designated O, B, A, F, G, K, and M, where the nomenclature has purely historical significance. Thus, the O stars that dominate the Orion Nebula are characterized by absorption lines from highly ionized heavy elements, such as C III (*i. e.*, C^{++}), O III, *etc.* On the other hand, the Taurus-Auriga region is rich in K and M stars, whose spectra show strong molecular bands in such species as CH and TiO. As a refinement of the system, each spectral type is further divided into ten subclasses, labeled 0 through 9, with higher numbers indicating cooler temperatures. The Sun, with its T_{eff} of 5800 K, is a G2 star. A star of type A0 serves to normalize the color indices. That is, the constants $m_{\lambda\circ}$ in equation (1.3) are selected so that the absolute magnitudes in all bands are identically zero for such an object.

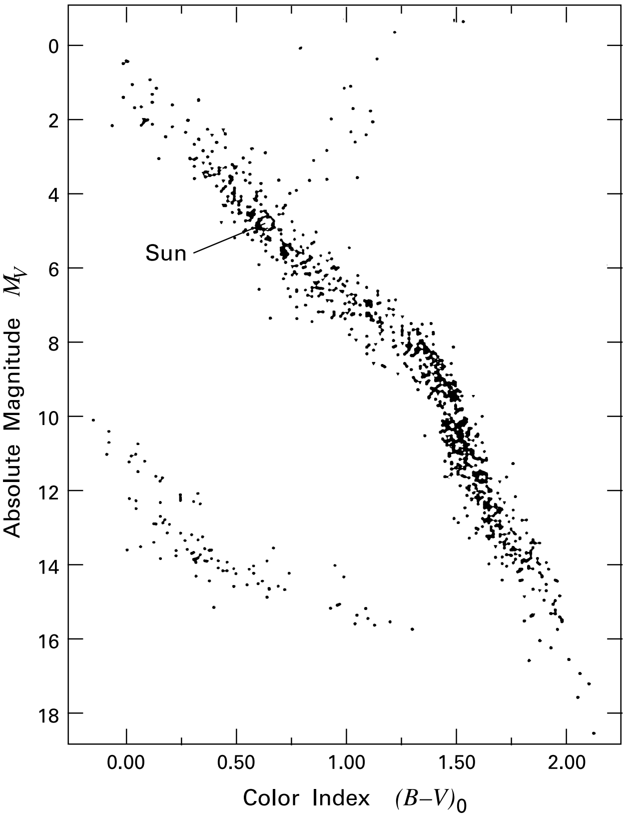


Figure 1.14 Color-magnitude diagram for 1094 stars in the solar neighborhood.

1.3.2 Main Sequence

The single most powerful conceptual tool in stellar astronomy is the *Hertzsprung-Russell (HR) diagram*. This is a plot of luminosity versus surface temperature (or their equivalents) for a single star or stellar group. A plot of M_V versus $(B-V)_\circ$ is also known as a *color-magnitude diagram*, while the L_*–$T_{\rm eff}$ plane is often called a *theoretical HR diagram*. Figure 1.14 is a color-magnitude diagram for relatively nearby stars. The vast majority, including the Sun itself, lie along a band known as the *main sequence*. The Sun's location is shown by the large open circle at $M_V = +4.82$ and $(B-V)_\circ = +0.65$. Astronomers frequently refer to main-sequence stars as *dwarfs*, to distinguish them from the sparser group to the upper right, the *giants*. Also apparent is a group to the lower left, the *white dwarfs*.

Figure 1.15 shows the main sequence in the theoretical HR diagram. The existence of this locus, in either set of coordinates, reflects the basic physics of stellar structure. A star is a self-gravitating ball of gas, supported against collapse by its internal thermal pressure. Throughout its life, the star continually radiates energy from its surface at the rate L_*. In a main-sequence object, this energy is resupplied by the fusion of hydrogen into helium at the center. The quantity L_* thus equals, in this case, $L_{\rm int}$, the luminosity crossing any interior

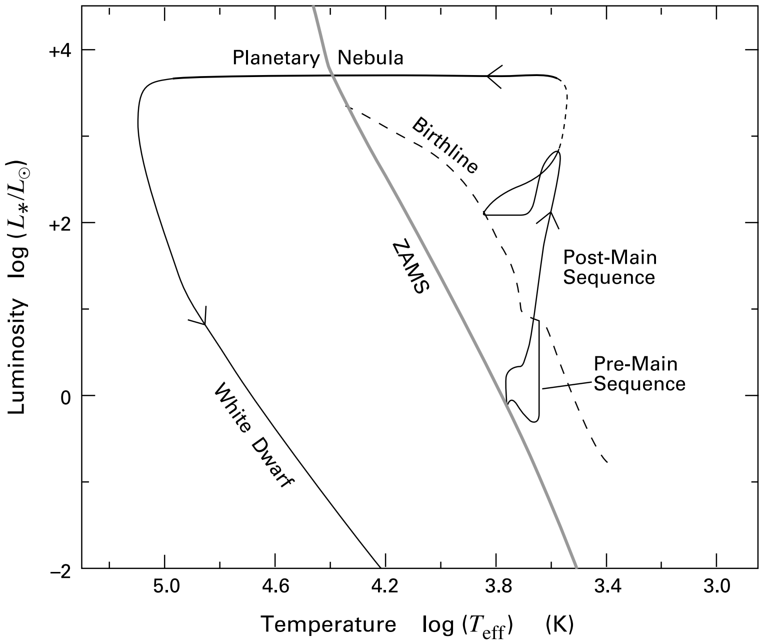


Figure 1.15 Evolutionary track of a 1 $M_\odot$ star in the theoretical HR diagram. The grey solid line represents the zero-age main sequence (ZAMS), while the dashed curve is the birthline.

spherical shell (see Figure 1.16). A main-sequence star is therefore in both *hydrostatic* and *thermal* equilibrium.

In a star of given mass M_* and radius R_*, the condition of hydrostatic equilibrium necessitates a certain interior run of temperature and density, where both quantities decrease outward. The radiative energy flux across any shell depends on the local temperature gradient, so that $L_{\rm int}$ is also specified for this object. If we now imagine squeezing the star to smaller R_*, its interior density rises. So must the temperature, in order to counteract the greater self-gravity. The luminosity crossing interior shells will also increase in response to the steeper temperature gradient. On the other hand, the star's nuclear reaction rate, which reflects both the frequency and energy of proton collisions, has its own functional dependence on the central density and temperature. Thus, for each M_*, we cannot really vary R_* at will; there is only one value for which the interior luminosity can be sustained by reactions at the center. But the stellar radius connects L_* and $T_{\rm eff}$ through the equation

$$L_* = 4\,\pi\,R_*^2\,\sigma_B\,T_{\rm eff}^4 \ . \tag{1.5}$$

where σ_B is the Stefan-Boltzmann constant. This blackbody relation, which we derive in Chapter 2, actually *defines* $T_{\rm eff}$. In summary, a main-sequence star of fixed mass has a unique L_* and $T_{\rm eff}$. The curve in the HR diagram is simply the functional relationship $L_*(T_{\rm eff})$ obtained by letting the stellar mass range freely.

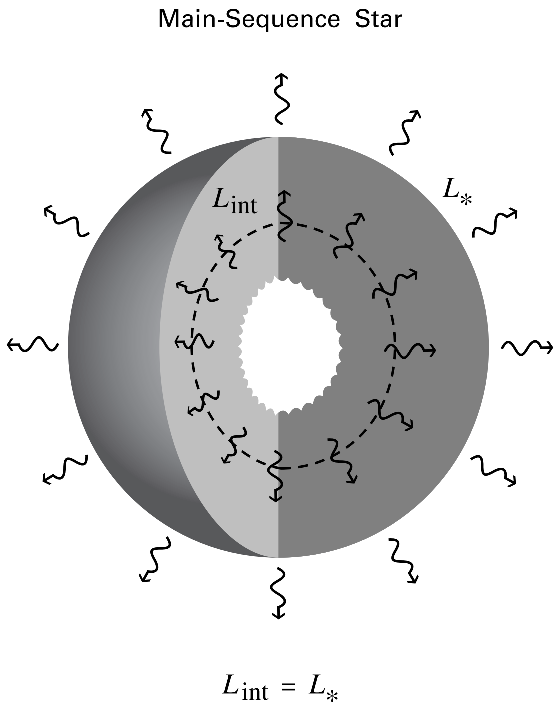


Figure 1.16 Energy transport in a main-sequence star. The interior luminosity $L_{\rm int}$ is generated in the nuclear-burning region near the center and is equal to the surface value L_*.

1.3.3 Early Phases

For an object that is *not* fusing hydrogen, both L_* and $T_{\rm eff}$ change with time. Correspondingly, its representative point moves in the HR diagram. The fact that most stars are observed to be on the main sequence reflects the longevity of the hydrogen-burning phase, during which L_* and $T_{\rm eff}$ change only slightly. Younger objects are more distended, with central temperatures that are too low for maintaining the fusion reaction. Nevertheless, these pre-main-sequence stars are relatively luminous. Since they are also detectable at visible wavelengths, their properties are well studied.

What supplies the star's luminosity during this early phase? The answer is the compressive work of gravity, which slowly squeezes the object to higher density. The local rate of energy loss from this process is zero at the center and increases monotonically outward. Thus, as illustrated in Figure 1.17, $L_{\rm int}$ across an arbitrary mass shell within a pre-main-sequence star is *less* than L_*. The surface radiation now causes a net drain of energy, leading to steady contraction and to continuous alteration of both L_* and $T_{\rm eff}$.

Determining pre-main-sequence tracks in the HR diagram for different stellar masses is an important aspect of star formation theory. Anticipating later results, Figure 1.15 shows the evolutionary track for a star of 1 $M_\odot$. The star first appears as an optically visible object on a curve called the *birthline*. As it then contracts, the star begins to descend a nearly vertical path. During this time, L_* is so high that energy is transported outward not by radiation but by *thermal convection*, *i. e.*, the mechanical motion of buoyant gas. By the time the star's path turns sharply upward and to the left, energy is partially transported by radiation, as well. After 3×10^7 yr, the star joins the main sequence.

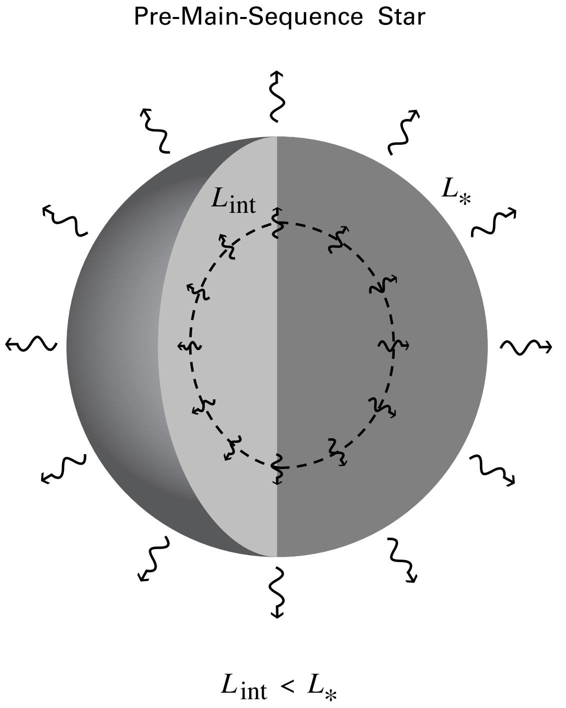


Figure 1.17 Energy transport in a pre-main-sequence star. In this case, there is no nuclear-burning region. The luminosity $L_{\rm int}$ monotonically increases from zero at the center to L_* at the surface.

Stars of other mass traverse analogous paths in the HR diagram but at very different rates. Less massive objects tend to have lower surface temperatures. According to equation (1.5), their values of L_* are also smaller for a given surface area, resulting in slower contraction. To quantify the rate, we first note that the sum of a star's thermal and gravitational energies is a negative quantity that is about GM_*^2/R_* in absolute value. The object radiates away an appreciable portion of this energy over the *Kelvin-Helmholtz time*:

$$\begin{aligned} t_{\rm KH} &\equiv \frac{G\,M_*^2}{R_*\,L_*} \\ &= 3\times 10^7\ {\rm yr}\left(\frac{M_*}{1\,M_\odot}\right)^2 \left(\frac{R_*}{1\,R_\odot}\right)^{-1} \left(\frac{L_*}{1\,L_\odot}\right)^{-1} \ . \end{aligned} \tag{1.6}$$

The significance of $t_{\rm KH}$ is that the star shrinks by about a factor of two over this interval, starting from any point during its pre-main-sequence phase. Notice that $t_{\rm KH}$ gets *longer* as the contraction proceeds. Thus, equation (1.6) also provides an approximate measure of the *total* time needed for a star to contract to its main-sequence values of M_*, R_*, and L_*.

Figure 1.18 displays pre-main-sequence tracks over a wide range of masses. All tracks descend from the birthline, which intersects the main sequence near 8 $M_\odot$. Higher-mass stars exhibit no optically visible pre-main-sequence phase, but first appear on the main sequence itself. If we imagine a group of stars all beginning contraction on the birthline at $t = 0$, their subsequent positions in the diagram at any fixed time would fall along a sequence of smooth curves. Figure 1.18 also shows a set of such pre-main-sequence *isochrones*. The reader should verify that the pattern of isochrones is consistent with the slower evolution of less massive objects and with the continual slowing of contraction at any fixed mass.

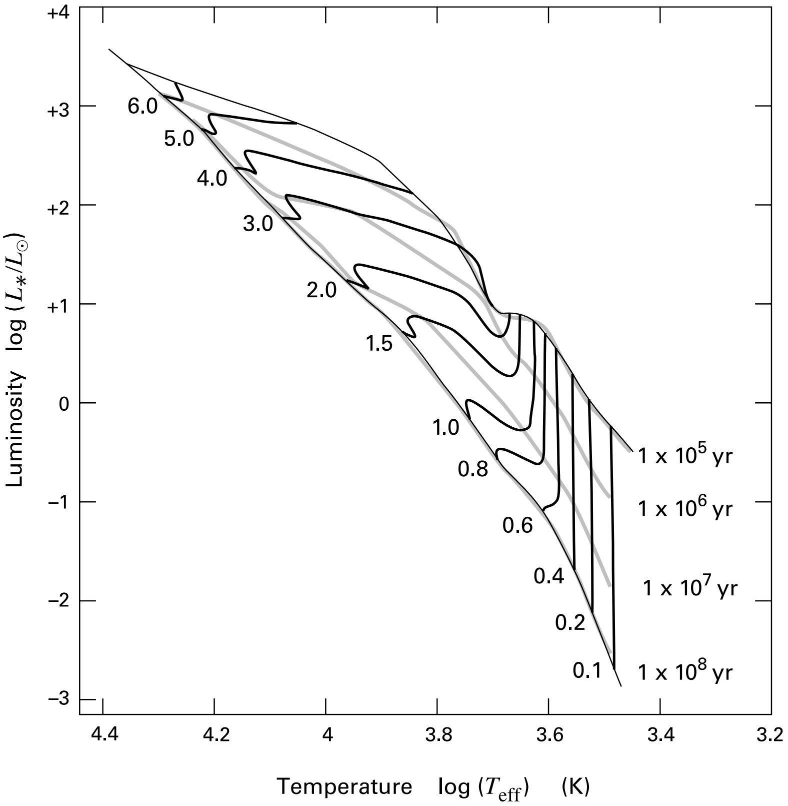


Figure 1.18 Pre-main-sequence evolutionary tracks. Each track is labeled by the stellar mass in units of $M_\odot$. The grey curves are isochrones, labeled in yr. The $t = 0$ isochrone coincides with the birthline, the lighter solid curve at the top. Note that the $t = 1 \times 10^8$ yr isochrone nearly matches the ZAMS, the lighter solid curve at the bottom.

The pre-main-sequence phase of evolution is not the first. At an earlier epoch, stars are still forming out of the gravitational collapse of their parent dense cores. Such *protostars* are even more luminous than pre-main-sequence stars, but are so obscured by interstellar dust that their emitted radiation lies entirely in the infrared and longer wavelengths. Under these circumstances, $T_{\rm eff}$ cannot be found by the traditional methods. Indeed, the character of the observed spectrum reflects more the properties of the dust surrounding the star than the stellar surface. Protostars therefore cannot be placed in a conventional HR diagram, and their identification within molecular clouds is still not secure.

Returning to pre-main-sequence evolution, we noted earlier that the contraction of any star causes its central temperature to rise. As long as the stellar mass exceeds 0.08 $M_\odot$, the temperature eventually reaches a value (near 10^7 K) where hydrogen fusion begins. Just at this point, the star is said to be on the *zero-age main sequence (ZAMS)*, and the corresponding relation between L_* (or $L_{\rm bol}$) and $T_{\rm eff}$ marks a rather precise locus in the HR diagram. The ZAMS is the curve actually shown in both Figures 1.15 and 1.18, and its properties are also listed in Table 1.1. The reader should bear in mind that the table represents a one-parameter family, where the basic physical variable is the stellar mass. Note also that astronomers sometimes use, in place of $L_{\rm bol}$, the *bolometric magnitude*:

$$M_{\rm bol} \equiv -2.5 \log \left(\frac{L_{\rm bol}}{L_\odot} \right) + m_\circ \ , \tag{1.7}$$

where the constant $m_\circ$ is +4.75. The difference $M_{\rm bol} - M_V$ for a main-sequence star of any spectral type is known as the *bolometric correction*, here evaluated in the V-waveband.

Table 1.1 Properties of the Main Sequence

Mass ($M_\odot$)	Sp. Type	M_V (mag)	log $L_{\rm bol}$ ($L_\odot$)	log $T_{\rm eff}$ (K)	$t_{\rm MS}$ (yr)
60	O5	−5.7	5.90	4.65	3.4×10^6
40	O6	−5.5	5.62	4.61	4.3×10^6
20	O9	−4.5	4.99	4.52	8.1×10^6
18	B0	−4.0	4.72	4.49	1.2×10^7
10	B2	−2.4	3.76	4.34	2.6×10^7
8	B3	−1.6	3.28	4.27	3.3×10^7
6	B5	−1.2	2.92	4.19	6.1×10^7
4	B8	−0.2	2.26	4.08	1.6×10^8
2	A5	1.9	1.15	3.91	1.1×10^9
1.5	F2	3.6	0.46	3.84	2.7×10^9
1	G2	4.7	0.04	3.77	1.0×10^{10}
0.8	K0	6.5	−0.55	3.66	2.5×10^{10}
0.6	K7	8.6	−1.10	3.59	
0.4	M2	10.5	−1.78	3.54	
0.2	M5	12.2	−2.05	3.52	
0.1	M7	14.6	−2.60	3.46	

1.3.4 Consumption of Nuclear Fuel

Stars remain on the main sequence for relatively long periods of time because of the vast supply of hydrogen available for fusion. To estimate the main-sequence lifetime $t_{\rm ms}$, we use the fact that the basic nuclear reaction is the fusion of four protons into ^{4}He. This process releases 26.7 MeV per helium nucleus, or $0.007\, m_p\, c^2$ for each proton of mass m_p. If f_H is the fraction of hydrogen consumed, then the star releases a total energy

$$E_{\rm tot} = 0.007\, f_H\, X\, M_*\, c^2 \tag{1.8}$$

during this period. Here, X is the star's mass fraction in hydrogen, which is typically 70 percent. The energy $E_{\rm tot}$ divided by $t_{\rm ms}$ must equal the (nearly constant) luminosity L_*. Detailed calculations show that $f_H \approx 0.1$ for most masses, so that the lifetime is approximately [4]

$$\begin{aligned} t_{\rm ms} &\approx 5 \times 10^{-4} \, \frac{M_* \, c^2}{L_*} \\ &= 1 \times 10^{10} \, \frac{(M_*/M_\odot)}{(L_*/L_\odot)} \ {\rm yr} \ . \end{aligned} \tag{1.9}$$

Here we have normalized the time to solar parameters. Because L_* increases very rapidly with M_*, $t_{\rm ms}$ falls correspondingly fast. Thus, the Sun, with a current age of 4.6×10^9 yr, is about midway through its main-sequence life. Table 1.1 gives more accurate values for $t_{\rm ms}$. We see that an O star of 40 $M_\odot$ lives for a relatively brief time, only 4.3×10^6 yr, while a K star of 0.6 $M_\odot$ burns hydrogen for 6.7×10^{10} yr, longer than the present age of the Galaxy.

As a star ends its life on the main sequence, its representative point in the HR diagram follows another well-defined path. Consider again the case of 1 $M_\odot$, whose post-main-sequence track is shown by the dashed curve in Figure 1.15. The exhaustion of hydrogen leaves a helium-rich central region that is thermally inert, *i. e.*, too cold to ignite its own nuclear reactions. Surrounding this region is a thick shell of fusing hydrogen. Continued shell burning adds more helium to the core, which eventually contracts from its self-gravity and from the weight of the overlying envelope of matter. The energy released from this contraction expands the envelope, and the star moves rapidly to the right in the HR diagram. This period ends once the envelope becomes convectively unstable. The greatly distended star then moves up the nearly vertical *red giant branch*, which resembles a time-reversed pre-main-sequence track.

While ascending the giant branch, the star is characterized by a contracting, inert core of helium, a hydrogen-burning shell, and an expanding envelope. The energy from the core and shell raises the luminosity by a factor of almost 10^3. The temperature in the core eventually reaches the point where helium can begin to fuse, forming ^{12}C. The star then moves to the left in the HR diagram, tracing the *horizontal branch*. After another 10^8 yr, helium burning exhausts itself in the core and shifts to an outer shell. The star, now containing two burning shells, ascends the *asymptotic giant branch*, so called because it approaches the original giant branch.

The ultimate fate of a star depends on its initial mass. In the 1 $M_\odot$ case, the asymptotic giant sheds a massive wind, exposing a small inert object at the center. The path in Figure 1.15 was computed under the plausible assumption that 0.4 $M_\odot$ was lost during this phase. The remnant central star is a *white dwarf*, an object so dense that it cannot be supported by ordinary gas pressure. Instead, the outward force stems from electron *degeneracy pressure*, *i. e.*, the mutual repulsion of the electrons' quantum mechanical wavefunctions. Such a star has no nuclear reserves and gradually fades from sight along the dashed curve shown in Figure 1.15.

More massive stars repeat the core-shell pattern of nuclear fusion numerous times while traversing the upper reaches of the HR diagram. These multiple ignition events create successively heavier elements. There is a limit, however, to the energy that can be extracted this way. In stars more massive than about 8 $M_\odot$, the central core eventually undergoes a violent collapse to become a *neutron star* or *black hole*. This collapse liberates so much energy that it completely disperses the outer layers of the star. Of the energy that is radiated, most goes into neutrinos.

---

[4] In this book, the symbol $\approx$ means "equals, to within a factor of two to three," while the symbol $\sim$ means "equals, to within an order of magnitude."

Nevertheless, the optical luminosity associated with this *supernova* briefly exceeds that of the entire Galaxy. The neutron star or black hole left behind gradually dims, until its presence is only known through its gravitating mass.

1.4 The Galactic Context

This, in broadest outline, is the evolution of a single star. However, our theory has a global aspect as well, since stellar birth and death are part of a vast evolutionary scheme being enacted in all galaxies like our own. Although we will continue, throughout this book, to emphasize the more local aspects of the problem, it is important to bear in mind this larger picture. Accordingly, let us take a brief look at our Galaxy and the role of star formation within it.

1.4.1 Structure of the Milky Way

The most conspicuous feature of the Galaxy is its highly flattened *disk* of stars. Our Sun orbits at a radius of 8.5 kpc, or about a third of the distance to the outer edge. The local rotation speed is 220 km s^{-1}, corresponding to a period of 2.4×10^8 yr within this differentially rotating structure. The local thickness of the disk *i. e.*, the average vertical excursion of the stars within it, varies systematically with spectral type. Thus, O and B stars have a characteristic half-thickness of 100 pc, while the figure increases to 350 pc for G stars like the Sun.[5] These observations refer to the *solar neighborhood*, *i. e.*, to objects closer than about 0.5 kpc.

The densest concentration of stars is in the *central bulge*. This nearly spherical configuration extends out of the disk plane and has a radius of 3 kpc. Even farther from the plane is an extended *stellar halo* (or *spheroid*) consisting of both *globular clusters* and a large population of field stars. Each cluster is a compact group with roughly 10^5 members. Halo stars, known collectively as Population II, are the oldest in the Galaxy, with only 1 percent or less of the heavy element content of those in the disk (Population I). The total mass of the stellar halo is at most comparable to that of the disk, about $6 \times 10^{10}\, M_\odot$. Finally, there is evidence for another nonplanar component, the unseen *dark halo*. The composition and spatial distribution of this (probably nonstellar) matter are still not known, but its total mass exceeds that of the disk and spheroid, perhaps by an order of magnitude.

What does the Galaxy look like? The solar system is embedded within the disk, so it is difficult to obtain a global view by direct observation, as opposed to theoretical reconstruction. While huge numbers of stars are visible at optical wavelengths, interstellar dust dims the light from the more distant ones. Hence, the effective viewing distance is limited to several kpc. However, red giant stars are relatively luminous and have such low surface temperatures that they emit copiously in the near infrared. Their radiation penetrates the dust and can thus be detected over much greater distances. The top panel of Figure 1.19 is a near-infrared portrait of the Galaxy obtained with the COBE satellite. Clearly evident in this remarkable image are the very thin disk and the central bulge. The bottom panel of the figure shows the more familiar, optical image.

---

[5] We define the *half-thickness* Δz of a planar distribution of matter as one half the ratio of total surface density to the volume density at the midplane. Alternatively, one may specify the *scale height* h, defined as the location where the volume density falls to e^{-1} of its midplane value. For a Gaussian distribution of density, $\Delta z = (\sqrt{\pi}/2)\, h = 0.89\, h$.

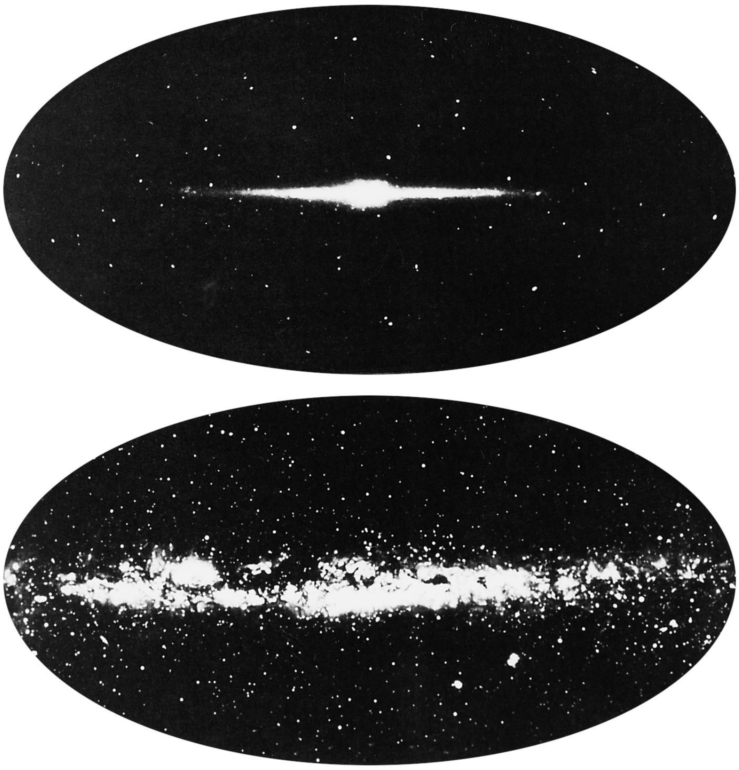


Figure 1.19 Two views of the Milky Way. The upper panel depicts the near-infrared emission, as seen by the COBE satellite. Beneath this is the optical image, shown at the same angular scale.

1.4.2 Spiral Arms

A true face-on view of our Galaxy is, of course, impossible to obtain, but would resemble the external galaxy M51 (NGC 5194), shown in the two panels of Figure 1.20. Here the most conspicuous feature is the presence of well-delineated *spiral arms*. Morphologically, M51 is a slightly *later-type* galaxy than the Milky Way, which has a somewhat larger nucleus and less extended arms. Other galaxies exhibit no spiral structure at all. Extreme *early-type* systems, the *ellipticals*, resemble three-dimensional spheroids rather than flattened disks.[6] In addition, there is a motley crew of *irregulars*, typified by the small companion to M51 seen in Figure 1.20.

According to density wave theory, the arms in spiral galaxies are not composed of a fixed group of stars, but represent a wave-like enhancement of density and luminosity rotating at a characteristic *pattern speed.* Both stars and gas in the underlying disk periodically overtake the arms and pass through them. Notice that, in Figure 1.20, the arms are more evident in the left image, taken in blue light, whereas the central bulge shows up strongly in the righthand,

---

[6] The designations "early" and "late," as applied to galaxies, refer to position along the Hubble morphological sequence. The same terms are used for stars, "early" ones being those with $T_{\rm eff}$ higher than solar. In neither case is there an implied *temporal* ordering.

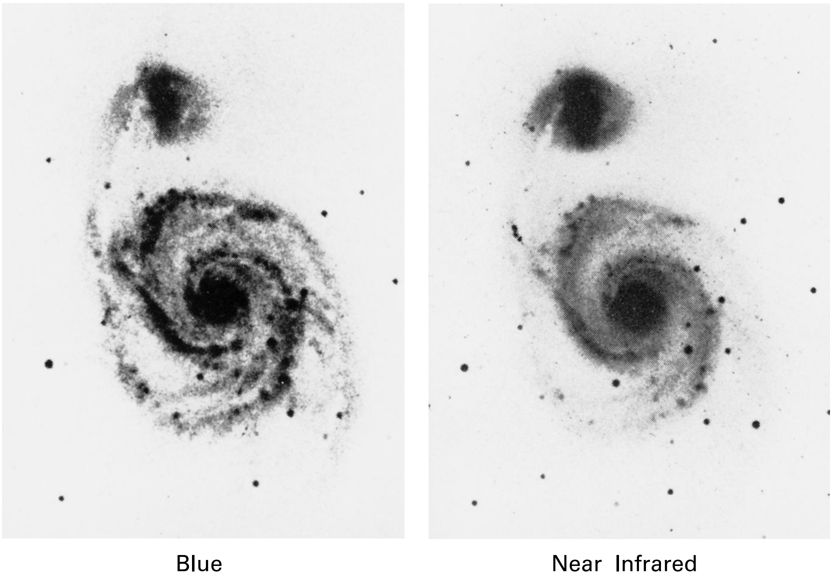


Figure 1.20 The galaxy M51 seen in blue light (*left panel*) and through a red filter (*right panel*). Note the prominence of the spiral arms in the blue image.

near-infrared photograph. The color of a galaxy is determined by the relative abundance and distribution in spectral type of its component stars. O and B stars, which have the highest surface temperature, dominate the blue image, while red giants, *i. e.*, dying stars of relatively low mass, are better seen through the infrared filter. Apparently, the most massive stars are concentrated heavily in spiral arms.

Recall that the main-sequence lifetime of a typical O star is of order 10^6 yr, only about 1 percent of the rotation period of a spiral arm. Thus spiral arms must produce, at each location in the disk, a temporary increase in gas density leading to a rise in the local star formation rate. Although stars of all mass are formed, it is the exceptionally bright O and B stars that are most conspicuous in optical photographs. Once the spiral wave has passed, the rate of forming new stars drops back down to its former low level.

This picture is reinforced by looking at the *gas* content of spiral galaxies. As we have seen, the molecular clouds that form stars can be detected most readily through their emission in CO. Figure 1.21 is yet another view of the galaxy M51. Here, an intensity map in the 2.6 mm CO line has been superposed on an optical image in the red Hα line, produced by atomic hydrogen heated to temperatures near 10^4 K. It is clear that the molecular gas is indeed concentrated along the spiral arms. Note that the radio contours seen here represent only 30 percent of the total CO emission. The remainder emanates from the interarm region, but is too smoothly distributed to be detected by the interferometer used for this observation. Closer inspection reveals that the Hα arms are displaced about 300 pc downstream from those seen in CO. Since Hα mainly stems from O and B stars, the implication is that cold gas entering the arms first condenses to form large cloud structures that later produce the massive stars. For a representative velocity of 100 km s^{-1} for material entering the arms, the corresponding time lag is 3×10^6 yr.

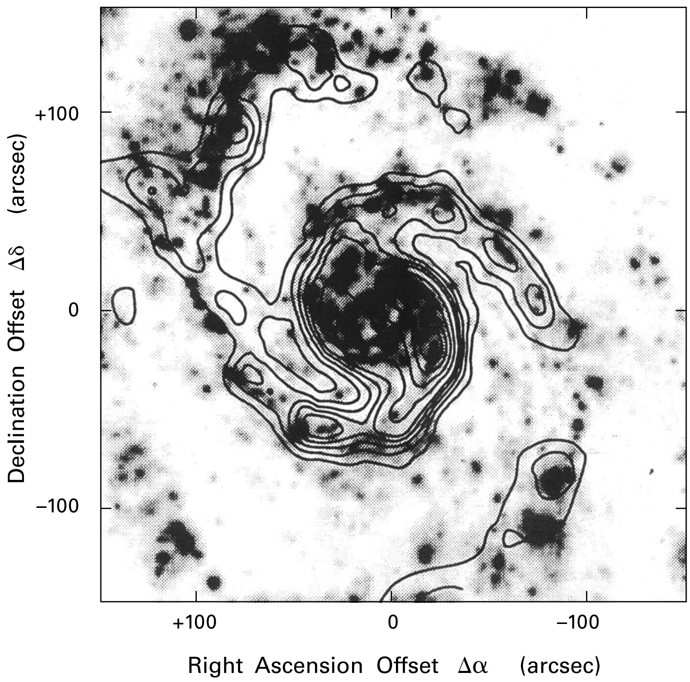


Figure 1.21 Map of M51 in the 2.6 mm line of $^{12}C^{16}O$ (*solid contours*) and the Hα line at 6563 Å (*dark patches*).

1.4.3 Recycling of Gas and Stars

The total mass of Galactic molecular gas, from which all new stars originate, is estimated to be $2 \times 10^9\, M_\odot$, or about 3 percent of the mass of the stellar disk. Much earlier in time, the very young Galaxy consisted entirely of diffuse gas. This material could have been produced cosmologically, during the nucleosynthesis occurring throughout the Universe in the first few minutes following the Big Bang. If so, the gas would have consisted solely of hydrogen, helium, and trace amounts of deuterium and lithium. On the other hand, both the present-day interstellar gas and the stars themselves, although composed predominantly of hydrogen and helium, contain a full complement of heavier elements, known in astronomical parlance as "metals." These could only have been produced through *stellar* nucleosynthesis. Thus, the primordial gas must first have condensed into stars, which later injected diffuse matter back into space. This gas itself had time to form a new generation of stars, which again injected gas following nuclear processing. The interstellar matter of today, therefore, has been recycled a number of times through stellar interiors.

Figure 1.22 illustrates the recycling process in a highly schematic fashion. Stars form continually through the condensation and collapse of interstellar gas clouds. These stars slowly consume their nuclear fuel, eventually ending their lives as white dwarfs, neutron stars, or black holes. The matter within such a *compact object* is irretrievably lost to the interstellar medium. But this mass is always less – often considerably less – than that of the original star, since stellar evolution itself inevitably produces mass loss.

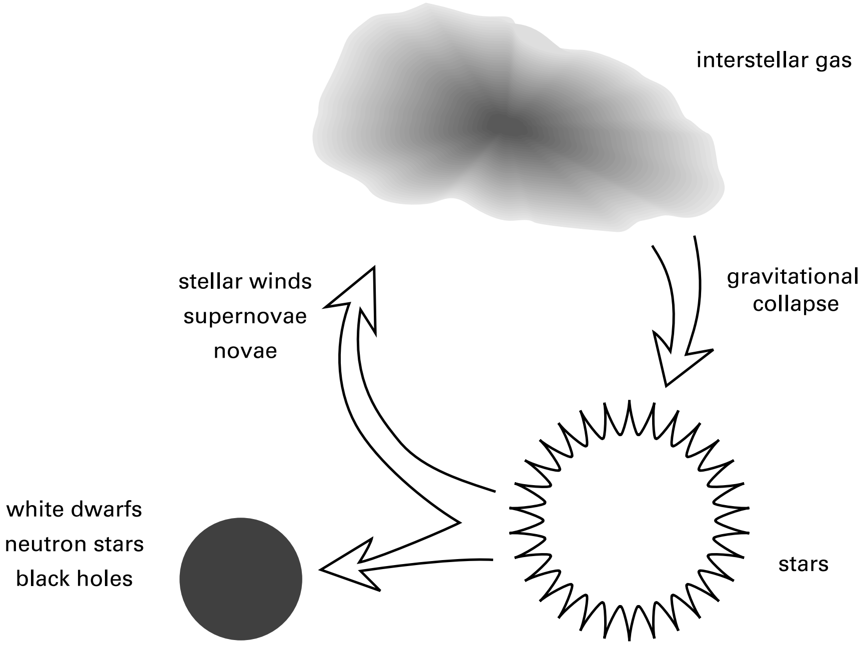


Figure 1.22 Recycling of stars and gas in the Galaxy.

The mass ejection process from stars is gentle and protracted in those objects destined to become white dwarfs. As we have seen, the outer envelopes of such stars inflate greatly after exhaustion of the neutral hydrogen. The loosely bound envelope then escapes as a stellar wind during the red giant and asymptotic giant phases. As this expanding gas cools, heavy elements within it nucleate into interstellar grains, which are then dispersed along with the gas. Although we have already mentioned grains for their obscuration of starlight, we will have much more to say about these submicron-size particles, which play an important role in the formation of stars.

Frequently, the white dwarf that survives mass ejection is part of a binary system. In the course of its own evolution, the companion may drive off matter that lands on the high-density surface of the compact object. Once it is sufficiently compressed, this hot material undergoes runaway nuclear fusion; the resulting *nova* ejects additional matter into space. Finally, stars too massive to become white dwarfs emit strong winds throughout their relatively brief lifetimes. The dispersal of their outer layers in a supernova is yet another source of Galactic gas.

Of the three mechanisms described, winds from stars with $M_* \lesssim 3\,M_\odot$ provide roughly 90 percent of the mass returned to the interstellar medium. Supernovae and the winds from massive stars account for the remainder. The latter two processes also yield the bulk of the heavy elements. All of the ejected gas mixes with that already present to form the raw material for new stars. The mixing process is partly nonlocal, since material from supernova explosions can be blown so far that it ultimately settles in a very different part of the Galaxy.

The formation of compact objects provides a continual drain on the Galactic gas content. Accordingly, the star formation rate has been falling with time for 1×10^{10} yr, the age of the oldest disk stars. This global formation rate, which we shall denote as $\dot{M}_*$, is currently about

$4\ M_\odot\ \mathrm{yr}^{-1}$. Here, the estimate relies on observations of O and B stars, which can be seen over the greatest distances. Using the known radiative output from a single star and the appropriate lifetime, the observed luminosity from a region yields the birthrate for stars of that type. The rate for all stars then follows by adopting a distribution of masses at birth. Another quantity of interest is $\dot{m}_*$, the *local* rate of star formation, as measured per unit area of the disk. Near our own Galactocentric radius $\varpi_\odot$, this quantity is about $3 \times 10^{-9}\ M_\odot\ \mathrm{yr}^{-1}\ \mathrm{pc}^{-2}$. The rate rises inward to a radius of about 3 kpc, inside of which it falls precipitously, before attaining a central peak. This behavior is correlated with the distribution of Galactic gas, and it is to this topic that we turn next.

Chapter Summary

The raw material for new stars within our Galaxy is a relatively small admixture of gas, concentrated especially near the spiral arms. Much of this diffuse matter is bound together into the extensive structures known as giant molecular clouds. At discrete sites, such as the Orion Nebula, both low-mass objects like the Sun and the much more luminous O and B stars form together within populous clusters. Other cloud complexes, such as that in Taurus-Auriga, are less dense and massive than giant molecular clouds. These sparser entities create loose associations of low- and intermediate-mass objects.

Each individual star originates in the collapse of a cloud fragment. After this optically invisible protostar phase, the subsequent evolution is conveniently pictured in the HR diagram, a plot of luminosity versus effective temperature. Starting at the locus known as the birthline, the star descends along a pre-main-sequence track, slowly contracting under its own gravity. The representative point in the diagram settles for a long time on the main sequence. Here the star derives energy from nuclear fusion. After exhausting this fuel, the object temporarily swells in radius, before eventually fading from sight. It also spews out gas, which joins the reservoir out of which additional stars will be made.

Suggested Reading

For an explanation of our abbreviations of journals and reviews, see the Sources list, toward the end of the book.

Section 1.1 The rich star-forming region of the Orion Nebula continues to yield fresh discoveries. Useful summaries, covering both the gas and stellar content, are

Genzel, R. & Stutzki, J. 1989, ARAA, 27, 41

O'Dell, C. R. 2003, *The Orion Nebula: Where Stars are Born* (Cambridge: Harvard U. Press).

Section 1.2 A brief but lucid summary of the Taurus-Auriga cloud complex and its stellar population is

Lada, E. A., Strom, K. M., & Myers, P. C. 1993, in *Protostars and Planets III*, ed. E. H. Levy and J. I. Lunine (Tucson: U. of Arizona Press), p. 245.

Section 1.3 Two texts on stellar structure and evolution are

> Clayton, D. D. 1983, *Principles of Stellar Structure and Nucleosynthesis* (Chicago: U. of Chicago)
>
> Hansen, C. J. & Kawaler, S. D. 1994, *Stellar Interiors: Physical Principles, Structure, and Evolution* (New York: Springer-Verlag).

As their titles imply, both books emphasize basic physics. Chapter 2 in the second text is a useful overview of the main phases of stellar evolution.

Section 1.4 The terminology used to describe the components of the Milky Way unfortunately varies among authors. Here we follow the convention established by

> King, I. R. 1990, in *The Milky Way as a Galaxy*, ed. G. Gilmore, I. R. King, and P. C. van der Kruit (Mill Valley: University Books), p. 1.

Star formation on the Galactic scale will be treated more fully in Chapter 19. For general texts, we recommend

> Binney, J. & Merrifield, M. 1998, *Galactic Astronomy* (Princeton: Princeton U. Press)
>
> Scheffler, H. & Elsasser, H. 1987, *Physics of the Galaxy and Interstellar Matter* (New York: Springer-Verlag).

The first emphasizes Galactic structure, while the second also includes a very broad array of topics relevant to the interstellar medium.

The Formation of Stars
Steven W. Stahler and Francesco Palla

2 The Interstellar Medium

A deeper understanding of early stellar evolution must begin by examining the rarefied medium out of which stars form. In this chapter, accordingly, we describe the overall physical properties of interstellar matter. We include both the gaseous constituent and the solid grains. For the former, we emphasize the physical state of the hydrogen and defer discussion of interstellar molecules until Chapter 5. Our treatment of the dust naturally brings in elements of radiation transport theory. The concepts introduced here will be extensively used throughout the remainder of this book.

2.1 Galactic Gas and Its Detection

We previously discussed the molecular hydrogen content of spiral galaxies, stressing the role of this gas in the production of stars. Equally important is the atomic component (HI), which has a total mass in our Galaxy exceeding H_2, and serves as the reservoir that ultimately produces molecular clouds. Finally, hydrogen can also be in ionized form (HII). While the total mass here is relatively small, this component is important as a tracer of massive stars.

2.1.1 Radio Emission from Atomic Hydrogen

The utility of CO for observing distant molecular clouds lies in the fact that the obscuring interstellar dust is transparent to millimeter photons (see § 2.4 below). To trace the HI gas, it is fortunate that there is another detectable, long-wavelength transition, the hydrogen line at 21.1 cm. Since its discovery in 1951, the 21 cm line has been of paramount importance for revealing the interstellar medium in our own and external galaxies.

We can understand the origin of this radiation by recalling the quantum mechanics of the hydrogen atom. In the nonrelativistic model based on Schrödinger's equation, the energy of the system, consisting of a single electron of mass m_e and charge e bound to a proton of mass m_p, is given by

$$\begin{aligned} E &= -\frac{\mu_{\rm ep}\, e^4}{2\,\hbar^2\, n^2} \\ &= 13.6\ \text{eV}\ n^{-2}\ , \end{aligned} \tag{2.1}$$

where n, the *principal quantum number*, can be 1, 2, 3, *etc.*, and where $\mu_{\rm ep}$ is the reduced mass, $m_e m_p/(m_e + m_p)$. The zero of energy, reached asymptotically for large n, represents the marginally bound state. The ultraviolet spectral lines generated by downward transitions to $n = 1$ form the *Lyman series*, where $n = 2 \rightarrow 1$ yields the Lyα line at 1216 Å, $n = 3 \rightarrow 1$ yields Lyβ at 1026 Å, *etc.* Similarly, the visible lines created by the jump to $n = 2$ constitute the *Balmer series*, with $n = 3 \rightarrow 2$ being the 6563 Å Hα line we have already encountered.

The Formation of Stars. Steven W. Stahler and Francesco Palla

ISBN: 3-527-40559-3

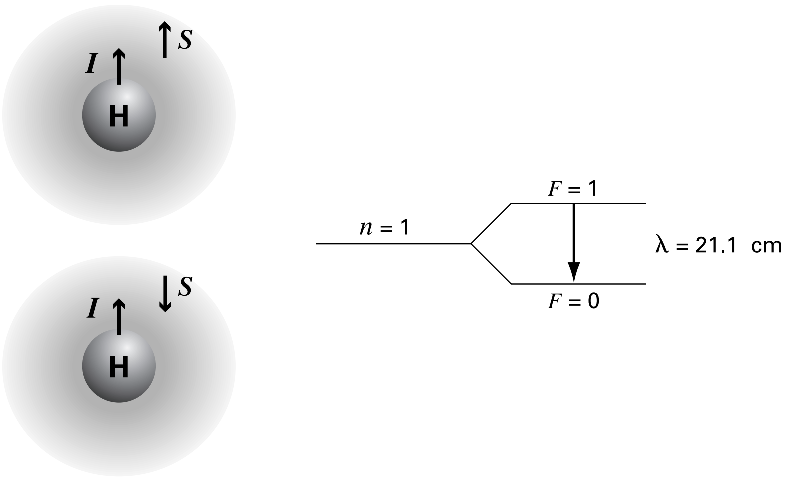


Figure 2.1 Origin of the 21 cm line. The hydrogen atom has greater energy when the spins of its proton and electron are parallel.

In the nonrelativistic treatment, a state with quantum number n actually consists of n^2 levels of identical energy. Each level is labeled not only by n, but also by a quantum number l, corresponding classically to the magnitude of the electron's orbital angular momentum $\boldsymbol{L}$, and by a third quantum number m_l, corresponding to the projection of $\boldsymbol{L}$ along any fixed axis. For a given n, l can take on any integer value from 0 to $n - 1$, while m_l can range from $-l$ to $+l$. Thus, in the $n = 1$ state normally found in interstellar gas, both l and m_l are zero.

The electron also has an intrinsic spin, which can be identified classically with another angular momentum vector $\boldsymbol{S}$ and an associated magnetic moment $\boldsymbol{\mu}$. In the electron's own reference frame, the motion of the charged proton creates a magnetic field which torques the spinning electron. The relativistic Dirac theory shows how this *spin-orbit interaction* causes the n^2 levels to have slightly different energies. Such *fine splitting* of the energy vanishes whenever $l = 0$. Even in this case, however, a smaller, *hyperfine splitting* is present due to the fact that the proton itself has an intrinsic spin $\boldsymbol{I}$ and therefore a magnetic moment. In a semi-classical picture, the energy of the atom depends on whether the vectors $\boldsymbol{S}$ and $\boldsymbol{I}$ are parallel or antiparallel (see Figure 2.1). Quantum mechanically, the two states are labeled by a quantum number F, which is 1 and 0 in the upper and lower states, respectively. The energy difference is so small, only 5.9×10^{-5} eV, that the wavelength of the emitted photon lies in the radio regime.

Within a region of HI gas, a hydrogen atom can be excited to the $F = 1$ state by collision with a neighboring atom. Usually, the same atom is later collisionally deexcited, but there is a small probability that the downward transition occurs through emission of a 21 cm photon. Despite the long interval between emission events for a given atom, 1×10^7 yr on average, an appreciable radio signal can be built up by a sufficiently large number of atoms. Indeed, the HI column density N_{HI}, *i. e.*, the number of atoms per unit area along any line of sight, is directly proportional to the received intensity of the 21 cm radiation.

Converting a column density into a volume density requires knowing the spatial location of the emitting gas. The actual received 21 cm signal is both shifted and spread out over a finite width in wavelength due to the Doppler effect. That is, any motion of a hydrogen atom toward or away from the observer changes the received wavelength from the intrinsically emitted value. The overall shift in peak intensity mainly stems from the differential rotation of the Galaxy. Since the pattern of rotation is well known, velocity shifts can be correlated with positions within the disk. Analysis of the emission profiles in many directions thus yields the number density n_{HI} throughout large regions.

Occasionally, one finds an extragalactic radio source, such as a quasar, behind a region of atomic hydrogen. If this source emits in the 21 cm regime, some of its photons will be absorbed by the intervening matter. Figure 2.2 shows a representative HI spectrum toward the quasar 3C 161.[1] The upper solid curve was obtained by averaging the signal from four lines of sight slightly offset from the point-like source. The resulting profile represents *emission* from HI gas, only some of which stems from the region of interest. When the telescope is pointed directly at the quasar, the additional signal is nearly constant in wavelength, except for the finite range removed by absorption. Hence, subtraction of the previous average from the on-source signal yields the *absorption* spectrum, also shown in the figure. In warmer gas, there is more collisionally induced emission of 21 cm radiation. The corresponding dip in the spectrum is therefore less pronounced. We can see, then, how 21 cm absorption profiles are useful probes of the gas temperature.

2.1.2 HI Distribution

Such radio studies, conducted over four decades, have been supplemented by numerous observations of optical and ultraviolet absorption lines seen against background stars. The net result is a rather detailed picture of the HI gas. This component of the interstellar medium pervades the Galactic disk, with a scale height of about 100 pc for Galactocentric radii between 4 and 8.5 kpc. In the outer Galaxy, the scale height appears to increase linearly with radius, while the HI layer at the furthest distances seen is warped by several kpc from a perfect plane. Throughout the Galaxy, most of the gas is in discrete clumps known as *HI clouds*. Although cloud properties vary broadly, typical number densities lie in the range from 10 to 100 cm^{-3}, and diameters range from 1 to 100 pc. The random motion of individual clouds, superposed on the general Galactic rotation, contributes significantly to the broadening of the observed 21 cm emission and creates the multiple emission peaks seen in Figure 2.2. A representative cloud temperature is 80 K, with a factor of two variation in either direction.

Figure 2.2 also indicates that a good deal of HI emission occurs outside the wavelength range in which the quasar radiation is being absorbed. The surplus emission, shown by the dashed profile in the figure, is generally smooth and stems from atomic gas *not* confined to individual clouds. This *warm neutral medium* is also distributed widely, with a scale height twice that of the cold component and a total mass that is also probably larger. The number density of this gas is about 0.5 cm^{-3} and its temperature, while difficult to ascertain because of the medium's transparency, is estimated at 8×10^3 K. Note that the term "neutral" is not strictly correct for either the cold or warm gas, since some metals in both are ionized by stellar photons. Ionized carbon, in particular, is an important coolant of the gas, as we shall discuss shortly.

The HI gas in the disk actually extends far beyond the visible stars. Indeed, observations of the circular velocity of this component in external spiral galaxies have provided the strongest

---

[1] We defer until Chapter 3 an explanation of the units used in this figure, which is essentially a plot of photon intensity versus wavelength.

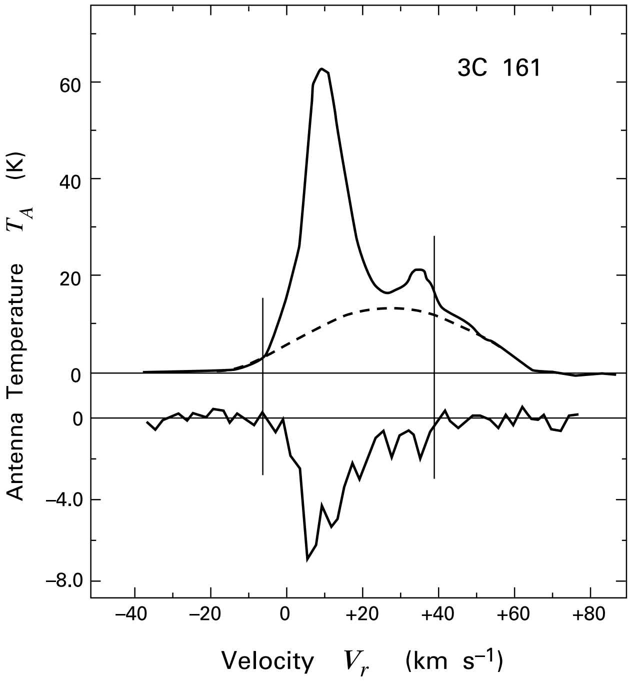


Figure 2.2 Radio spectrum of a quasar partially obscured by HI gas. The upper solid curve shows, as a function of velocity (or, equivalently, wavelength), the intensity of emission in the 21 cm line. The lower solid curve shows, with a different amplitude scale, the absorption. This absorption is mostly confined in velocity between the two vertical bars. Finally, the upper dashed curve indicates the contribution to the total emission from the warm neutral medium.

evidence for extended, dark halos. The distribution of HI surface density in our own Galaxy is shown in Figure 2.3. Within the central bulge, the density is very low, but rises to a fairly uniform level between radii of 4 and 14 kpc. At larger radii the surface density declines, but in a manner that is rather uncertain because of the poorly known Galactic rotation at these distances.

2.1.3 Molecular Gas

Returning to the molecular gas, the CO line emission commonly used to trace this component has a shorter wavelength than the 21 cm HI line and so provides higher spatial resolution. Numerous surveys have established that the molecular gas is also spread throughout the Galaxy and mostly contained in discrete clouds. A rotationally excited CO molecule decays much more quickly than an HI atom in the $F = 1$ state. Hence CO is intrinsically a stronger emitter. So much radiation is produced, however, that most molecular clouds are *optically thick* to the 2.6 mm line, *i. e.*, a photon is absorbed and reradiated many times before reaching the surface. Under these circumstances, the received intensity from a single cloud is *not* proportional to the column density. Nevertheless, the line is still useful for tracing the total H_2 contained in a gravitationally bound *ensemble* of clouds, as we will discuss in Chapter 6.

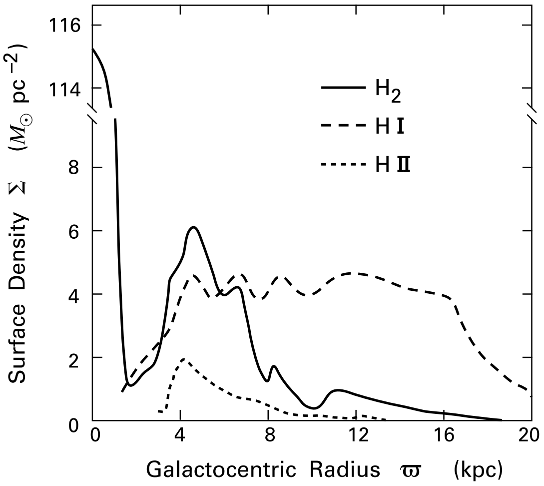


Figure 2.3 Galactic surface densities of H_2, HI, and HII gas. The densities are shown as a function of radial distance from the Galactic center. Note the high central peak of the molecular component. The ionized component shown is only that within HII regions.

The global distribution of H_2 is quite different from HI (see Figure 2.3). There is little molecular gas outside 10 kpc, but the surface density inside climbs quickly, reaching a broad maximum centered at 6 kpc. The origin of this massive *molecular ring* is uncertain, but is presumably related to Galactic spiral structure. The falloff in density inside the ring is broadly consistent with the decline in the measured star formation rate. Thus, the sharp rise in molecular density within 1 kpc of the Galactic center is especially intriguing. Whatever the origin of this gas, observations show that it is producing large numbers of massive stars. Note finally that the inward rise in star formation activity near the Sun's position is consistent with the fact that the observed metallicity in both stars and gas has a measured radial gradient in the same sense.

Molecular gas is more closely confined to the Galactic plane than the atomic component, with a scale height of 60 pc at the solar position, or about half the HI value. While the atomic gas fills much of the disk up to its nominal scale height, the molecular component occupies only about one percent of the available volume. The difference in scale heights reflects the lower random velocity of the molecular cloud ensemble. This velocity has a median value of 4 km s^{-1} and seems to vary little with either cloud mass or Galactocentric radius. The dominant gravitational force toward the plane is that from the stars, whose surface density falls steeply with radius. Hence the H_2 thickness should *increase* outward, a conclusion confirmed by the CO surveys.

2.1.4 HII Regions

In addition to its atomic and molecular forms, interstellar hydrogen also exists in the ionized state. Some of this gas is confined to HII regions surrounding individual O and B stars. Stellar *Lyman continuum* radiation, produced by hydrogen recombinations in which the photon energy exceeds 13.6 eV, ionizes a spherical volume several parsecs in radius. Recombining electrons and ions in this region emit a plethora of lines, including the optical Balmer series of

Figure 2.4 Emission in 1.3 cm continuum radiation (*white contours*) from the Orion Nebula. The optical photograph is a negative image in the Hα line.

hydrogen. Transitions between much higher levels of the atom produce radiation at centimeter wavelengths. This emission provides a radio beacon for HII regions over large distances.

Figure 2.4 shows the best studied HII region, that in Orion. Here, radio contours at 1.3 cm are overlaid on a negative Hα image. The contours are centered on the Trapezium stars, which also illuminate much of the region in Hα. Notice how much more symmetric the radio emission appears than the optical line, which is readily attenuated by dust. By summing the flux from such radio peaks over the Galactic plane, the total surface density can be obtained (Figure 2.3). Like the molecular component, the HII gas peaks at a radius near 5 kpc. Since HII regions are created by young, massive stars, this similarity further establishes the association of star formation activity with molecular gas. The total mass contained in these regions is about $1 \times 10^8\ M_\odot$, an order of magnitude less than the molecular or atomic hydrogen masses.[2]

The presence of emission lines from the heavier elements in HII regions allows reconstruction of the chemical composition of interstellar gas. Analysis of the absorption lines from stars lying behind HI clouds is another means to this end. The standard of comparison in such studies is *solar composition*, *i. e.*, the elemental distribution found in the solar photosphere and in primitive solar system bodies such as meteorites. Table 2.1 lists number abundances relative to hydrogen of the most prevalent elements in gas of solar composition. Also shown are observed abundances in the HII region M42. It is apparent that the metals in this region are systematically depleted relative to the solar standard. Such depletion is generally observed in interstellar gas and can be several orders of magnitude for iron and calcium. Since the total metallicity is presumed to be uniform throughout the solar neighborhood, this trend is evidence that an increasing portion of heavier atoms is locked up in solid matter, *i. e.*, interstellar grains.

---

[2] We stress that this mass estimate includes only discrete HII regions. A larger amount of gas is contained in the warm ionized medium, introduced below.

Table 2.1 Elemental Abundances in the Solar System and M42

Element	Solar System	M42
He/H	0.1	0.1
C/H	3.6×10^{-4}	3.4×10^{-4}
N/H	1.1×10^{-4}	6.8×10^{-5}
O/H	8.5×10^{-4}	3.8×10^{-4}
Si/H	3.6×10^{-5}	3.0×10^{-6}

All abundances are by number relative to hydrogen.

A useful concept for quantitative work is the *mean molecular weight* of interstellar gas. Symbolized μ, this quantity represents the average mass of a particle of gas relative to the hydrogen mass m_H. Thus, in a gas of mass density ρ, the total number density of particles is

$$n_{\rm tot} = \frac{\rho}{\mu\, m_H} \ . \tag{2.2}$$

Any one element contributes a number density of particles given by

$$n_i = \frac{X_i\, f_i\, \rho}{A_i\, m_H} \ . \tag{2.3}$$

Here, X_i is the *mass* fraction of the element, A_i its atomic weight relative to hydrogen, and f_i the number of free particles (including electrons) per atomic nucleus. The hydrogen mass fraction, conventionally designated X, is 0.70 in solar composition gas, while that for helium, denoted Y, is 0.28. For heavier elements, the mass fraction is the metallicity $Z = 0.02$, and the average atomic weight is $A_Z = 17$. By summing equation (2.3) over the three "elements" and comparing to equation (2.2), it follows that molecular gas of solar composition has $\mu = 2.4$, while HI gas has $\mu = 1.3$. In the case of a fully ionized gas, such as that found in stellar interiors, we may use the convenient fact that $f_Z \approx 2\, A_Z$ to derive $\mu = 0.61$.

2.2 Phases of the Interstellar Medium

Interstellar hydrogen exists in a variety of chemical forms– ionized, atomic, and molecular. In addition, atomic hydrogen itself is found in both discrete clouds and as warmer, more rarified gas. What is the origin of this diversity and how is it maintained? Our aim in this section is to see what insights theory can provide regarding these questions.

2.2.1 Pressure Equilibrium

We first need to draw a distinction between HII regions and the others. The ionized gas here arises from the ultraviolet radiation of nearby, massive stars. Each individual region is thus a transient phenomenon that disappears over the 10^6 yr lifetime of its parent object. Neutral hydrogen, however, persists for much longer periods and over vast regions devoid of stars.

Consider first the atomic component. With regard to its origin and stability, an essential clue is provided by the gas *pressure*. For a typical HI cloud with number density $n = 30$ cm^{-3} and temperature $T = 80$ K, the product nT is 2×10^3 cm^{-3} K. This figure is matched, within

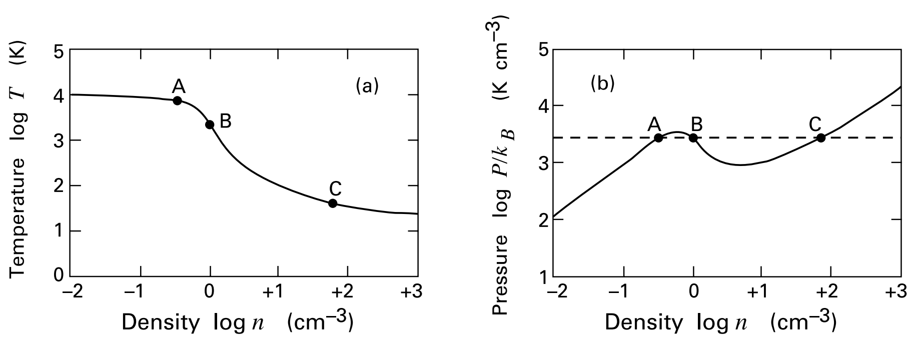


Figure 2.5 (a) Theoretical prediction for the equilibrium temperature of interstellar gas, displayed as a function of the number density n. (b) Equilibrium pressure nT as a function of number density. The horizontal dashed line indicates the empirical nT-value for the interstellar medium.

a factor of two, by the corresponding product for the warm neutral medium ($n = 0.5\ \mathrm{cm}^{-3}$; $T = 8 \times 10^3$ K). The numbers suggest that HI clouds and the warm neutral medium might profitably be viewed as two *phases* of the interstellar medium that coexist in pressure equilibrium.

We may pursue this idea quantitatively as follows. The thermal energy content of any gas results from the balance of heating and cooling processes. As we will detail in Chapter 7, rates for these processes can be determined by both theoretical and empirical methods. Assuming this information to be at hand, imagine slowly compressing a parcel of interstellar gas to successively higher densities. At each step, the gas will assume an equilibrium temperature and pressure, which can be determined by equating the heating and cooling rates. We can then examine those states that have pressures matching the known interstellar value.

The results of such a calculation, for gas of solar composition, are displayed in Figure 2.5. Over the range of densities shown, heating is provided mostly by starlight, through its ejection of electrons from the surfaces of interstellar grains. Figure 2.5a shows that this heating yields, at lower densities, temperatures approaching 10^4 K, where internal electron levels of hydrogen are excited. The gas in this regime cools primarily through emission of the Lyα line. At higher densities, hydrogen is in the ground electronic state and becomes an inefficient radiator. Atomic carbon, however, remains ionized by ambient ultraviolet photons, and the gas cools through the 158 μm transition of C II.

The solid curve in Figure 2.5b shows the predicted run of pressure (represented as $nT = P/k_B$, where k_B is Boltzmann's constant) as a function of density. A parcel of gas with pressure lying above this curve cools faster than it can be heated, while the reverse is true for points which lie below. The equilibrium curve crosses the mean empirical P/k_B-value of $3 \times 10^3\ \mathrm{K\ cm}^{-3}$ (*dashed horizontal line*) at three distinct points. Imagine first that the gas finds itself at point B. Suppose further that it is compressed slightly while maintaining pressure equilibrium with its surroundings. Since its representative point now crosses above the equilibrium curve, it must cool until it reaches point C. Conversely, any slight expansion of the same parcel causes it to heat up until it reaches point A. Point B, therefore, represents a *thermally unstable* state. Furthermore, the density and temperature at the stable point A ($n = 0.4\ \mathrm{cm}^{-3}$; $T = 7000$ K), are seen to match those in the warm neutral medium. Conditions at point C

($n = 60\ \mathrm{cm}^{-3}$; $T = 50$ K), where the gas is said to be in the *cold neutral medium*, are just those in typical HI clouds.

From this analysis, we have strong reason to believe that any large mass of atomic hydrogen gas naturally divides into two components with very different properties. Of course, we still do not know how the actual separation occurs in detail. Nor does the foregoing reasoning tell us the relative fractional volumes of the two phases, much less the sizes and masses of individual HI clouds. At present, our knowledge of the atomic component is simply too rudimentary to address these issues from a theoretical perspective.

What of the molecular gas? Can it, too, be regarded as a phase of the interstellar medium? In any molecular cloud, densities are sufficiently high that self-gravity plays a dominant role in the cloud's mechanical equilibrium. In other words, the product nT deep inside can be much greater than the background value, since the interior pressure must also resist the weight of overlying gas. On the other hand, the more rarified material in the cloud's outer layers must still match the external pressure if this region is not to expand or contract. In Chapter 8, we will use this requirement to establish plausible surface conditions for molecular clouds.

2.2.2 Vertical Distribution of HI

In the atomic gas, as well, thermal and mechanical conditions are intimately related. Consider, for example, the gas distribution above and below the Galactic disk midplane. Under the influence of the vertical gravitational force, the two components should separate spatially, with the colder and denser medium residing closer to the plane. Indeed, we may estimate the scale height of the cold neutral medium by balancing gravity against the internal pressure gradient. Let Φ_g be the Galactic gravitational potential, so that $\boldsymbol{\nabla}\Phi_g$ is the vector force per unit mass. The equation of *hydrostatic equilibrium* then reads

$$-\frac{1}{\rho_{\mathrm{HI}}}\,\boldsymbol{\nabla} P_{\mathrm{HI}} - \boldsymbol{\nabla}\Phi_g = 0 \ . \tag{2.4}$$

Let z represent the height above the midplane. Dotting the last equation with the unit vector $\hat{\boldsymbol{z}}$ and transposing yields

$$\frac{1}{\rho_{\mathrm{HI}}}\frac{\partial P_{\mathrm{HI}}}{\partial z} = -\frac{\partial \Phi_g}{\partial z} \ , \tag{2.5}$$

where P_{HI} and ρ_{HI} are, respectively, the pressure and mass density of the atomic gas. These two quantities are related by

$$P_{\mathrm{HI}} = \rho_{\mathrm{HI}}\, c_{\mathrm{HI}}^2 \ . \tag{2.6}$$

Here the quantity c_{HI}, which has the dimensions of velocity, represents the random internal motion of the medium. It is also the speed of sound, assuming the medium keeps a fixed temperature during passage of the waves. For an ideal gas, c_{HI} is given in terms of the temperature and mean molecular weight by $(\mathcal{R}T/\mu)^{1/2}$, with $\mathcal{R}$ being the gas constant. If we assume that c_{HI} does not vary with z, then substitution of equation (2.6) into (2.5) and a single integration yields the relation between ρ and Φ_g:

$$\frac{\rho_{\mathrm{HI}}(z)}{\rho_{\mathrm{HI}}(0)} = \exp\left[\frac{\Phi_g(0) - \Phi_g(z)}{c_{\mathrm{HI}}^2}\right] \ . \tag{2.7}$$

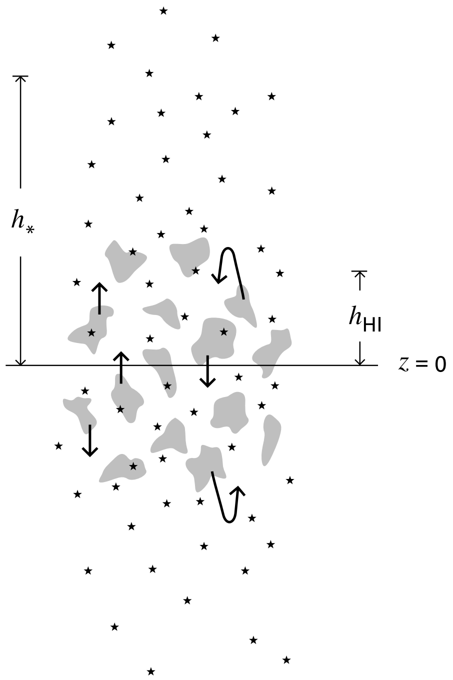


Figure 2.6 Motion of HI gas in the Galaxy. Discrete clouds move away from the Galactic midplane at $z = 0$, reaching an average distance $h_{\rm HI}$ above and below it. This distance is less than h_*, the scale height for low-mass stars.

To make further progress, we need to specify the gravitational potential. This quantity is related to ρ_*, the total mass density in the Galaxy, through Poisson's equation:

$$\nabla^2 \Phi_g = 4\pi G \rho_* \ . \tag{2.8}$$

In a thin disk, we may safely ignore horizontal gradients and write (2.8) as

$$\frac{\partial^2 \Phi_g}{\partial z^2} = 4\pi G \rho_*(z) \ . \tag{2.9}$$

Most of the Galactic matter consists of low-mass stars, which have a scale height h_* greater than that of the neutral gas (Figure 2.6). Hence, we may safely replace $\rho_*(z)$ in equation (2.9) by its midplane value $\rho_*(0)$. We then integrate this equation twice, noting that the gravitational force at the midplane, which is $-\partial\Phi_g/\partial z$, must vanish by symmetry. The resulting potential is given by

$$\Phi_g(z) = \Phi_g(0) + 2\pi G \rho_*(0)\, z^2 \ . \tag{2.10}$$

Equations (2.7) and (2.10) together imply that the cold neutral medium has a Gaussian distribution. The corresponding scale height is

$$h_{\rm HI} = \left[\frac{2\pi G \rho_*(0)}{c_{\rm HI}^2}\right]^{-1/2} \ . \tag{2.11}$$

How well does this theoretical prediction accord with observations? From measurements of the vertical distribution and velocity of stars of various masses, $\rho_*(0)$ in the solar neighborhood is known to be 0.18 $M_\odot$ pc^{-3}. If we then calculate $c_{\rm HI}$ using a temperature of 80 K, we find from equation (2.11) that $h_{\rm HI}$ is 10 pc, only 10 percent of the observed value! In hindsight, this failure is not too surprising. After all, 21 cm observations show that the cold gas actually consists of discrete clouds in seemingly random motion. A typical cloud velocity in the z-direction is 6 km s^{-1}. Substitution of *this* value for $c_{\rm HI}$ in (2.11) yields a scale height much closer to the observed one. The lesson here, as illustrated in Figure 2.6, is that the vertical distribution of cold HI gas stems from the *bulk motion of individual clouds* rather than internal pressure support. The salient question is now the energizing agent for this motion.

2.2.3 Warm and Hot Intercloud Gas

Widespread stirring of the neutral gas could be provided by supernovae. Although supernova explosions are rare, occurring only once every 50 years or so in a galaxy like our own, the expanding material from each event sweeps out a large volume and imparts considerable momentum to the surrounding gas. The extremely hot gas within the advancing shell of a supernova remnant cools so slowly that new remnants inevitably intersect older ones, resulting in a network of connected bubbles. Indeed, there is good evidence that the solar system is located within such a rarified region (see § 7.2). On a much greater scale, it is thought that the hot gas fills a large volume of the disk, surrounding the neutral component and effectively constituting a third phase of the interstellar medium. The temperature of this *hot intercloud medium* is predicted theoretically to be 10^6 K, and its density 0.003 $\rm cm^{-3}$.

Verification of the properties, including the spatial distribution, of the third phase has been slow, as the relevant data are difficult to obtain. Thus far, spectral lines from highly ionized heavy elements have provided the only solid clues. Since these lines lie in the ultraviolet, they must be observed from space. In the early 1970s, the Copernicus satellite observed pervasive emission from OVI. Other spaceborne instruments have subsequently detected additional species. It is interesting that optical lines from less highly ionized elements are also seen. These lines, together with widespread, diffuse emission in Hα, point to the existence of yet another gas component. This *warm ionized medium*, first systematically explored in the 1980s, shares the same density and temperature as the warm neutral medium, but its hydrogen is mostly ionized. Accounting for the energy input required to maintain this ionization poses another challenge to theory.

It should be clear from this brief discussion that many aspects of the interstellar gas remain unexplained. In particular, recent observations have emphasized the general principle that the medium is essentially a *dynamic* entity. This feature naturally renders the theorist's task much more difficult. Nevertheless, the insight that the gas consists of discrete phases in approximate pressure equilibrium will continue to play a key role. As a convenient reference for the reader, we summarize in Table 2.2 the basic properties of the known phases. Note that we have included the important molecular component, although this gas is not in pressure balance with the other phases.

2.3 Interstellar Dust: Extinction and Thermal Emission

Optical radiation from distant stars is attenuated by intervening dust. If this effect were not taken into account, an observed star would either be assigned too small a luminosity or else placed at an erroneously large distance. Indeed, the problem is even worse, since dust grains also

Table 2.2 Phases of the Interstellar Medium

Phase	$n_{\rm tot}$ (cm^{-3})	T (K)	M ($10^9\,M_\odot$)	f
molecular	> 300	10	2.0	0.01
cold neutral	50	80	3.0	0.04
warm neutral	0.5	8×10^3	4.0	0.30
warm ionized	0.3	8×10^3	1.0	0.15
hot ionized	3×10^{-3}	5×10^5	–	0.50

f is the volume filling factor.

redden the light passing through them. The dust content of interstellar matter must therefore be considered in determining the two most basic properties of any star – its luminosity and effective temperature. Extinction and reddening occur not only at visible wavelengths but, to varying degrees, throughout the electromagnetic spectrum. Since they enter astronomy so pervasively, we pause in our exploration of the interstellar medium to present the essential concepts and terminology related to these linked phenomena. In addition, we need to examine the emission produced by the dust as a result of heating by stellar light.

2.3.1 Extinction and Reddening

Let us first quantify the wavelength dependence of extinction. Recall from Chapter 1 that the brightness of a star near any wavelength λ is measured either by m_λ or M_λ, its apparent and absolute magnitudes, respectively. The presence of dust between the star and Earth alters the relation between these two quantities from equation (1.4) to

$$m_\lambda = M_\lambda \,+\, 5\log\left(\frac{r}{10\ {\rm pc}}\right) \,+\, A_\lambda \ , \tag{2.12}$$

where A_λ, a positive quantity measured in magnitudes, is known simply as the *extinction* at wavelength λ. Note that, even at the fiducial distance of 10 pc, the star now has $m_\lambda > M_\lambda$, *i. e.*, it is dimmer than its absolute magnitude.

The general tendency of dust to redden distant objects implies that A_λ must diminish with increasing λ, at least in the optical regime. Consider, then, writing equation (2.12) for the same star at two different wavelengths λ_1 and λ_2. Eliminating r between the equations yields

$$(m_{\lambda_1} - m_{\lambda_2}) = (M_{\lambda_1} - M_{\lambda_2}) + (A_{\lambda_1} - A_{\lambda_2}) \ . \tag{2.13}$$

The first righthand quantity is C_{12}°, the intrinsic color index measured at λ_1 and λ_2. As we noted in § 1.3, this quantity can also be written in terms of the UBV filter names as $(B-V)_\circ$, $(U-B)_\circ$, *etc.* The lefthand quantity in equation (2.13), denoted generically as C_{12}, or as $B-V$, $U-B$, *etc.*, is the *observed color index*. The difference between the intrinsic and observed color indices is a measure of reddening known as the *color excess*, E_{12}:

$$E_{12} \equiv C_{12} - C_{12}^\circ \tag{2.14a}$$

$$= A_{\lambda_1} - A_{\lambda_2} \ . \tag{2.14b}$$

The color excess is written in the UBV system as $E_{\rm B-V}$, $E_{\rm U-B}$, etc. At visual wavelengths, $A_{\lambda_1} > A_{\lambda_2}$ for $\lambda_1 < \lambda_2$, and E_{12} is a positive quantity.

Both the extinction and the color excess are proportional to the column density of dust grains along the line of sight.[3] Consider yet a third wavelength, λ_3. Then the ratios A_{λ_3}/E_{12} and E_{32}/E_{12} depend only on intrinsic grain properties. If we now fix λ_1 and λ_2, while letting the third wavelength have the arbitrary value λ, either ratio provides a measure of the wavelength-variation of dust extinction. Conventionally, λ_1 and λ_2 are chosen to correspond to the B and V filters, respectively. In this case, $E_{\lambda-\rm V}/E_{\rm B-V}$ is known as the normalized *selective extinction* at λ, while $A_\lambda/E_{\rm B-V}$ is the normalized *total extinction*. From equation (2.14b), these two quantities are simply related:

$$\frac{E_{\lambda-\rm V}}{E_{\rm B-V}} = \frac{A_\lambda}{E_{\rm B-V}} - \frac{A_V}{E_{\rm B-V}} \tag{2.15a}$$

$$= \frac{A_\lambda}{E_{\rm B-V}} - R\ . \tag{2.15b}$$

Here the fiducial ratio R is measured to be:

$$\begin{aligned} R &\equiv \frac{A_V}{E_{\rm B-V}} \\ &= 3.1\ . \end{aligned} \tag{2.16}$$

The quoted value applies to the diffuse interstellar medium, but R can be significantly higher in dense molecular clouds, including those where stars form.

A plot of either $E_{\lambda-\rm V}/E_{\rm B-V}$ or $A_\lambda/E_{\rm B-V}$ versus λ is known as the *interstellar extinction curve*. Figure 2.7 shows the standard curve, in which the independent variable is $1/\lambda$. This important result, summarizing the extinction and reddening properties of the interstellar medium, was obtained by applying equation (2.14b) empirically, *i. e.*, by comparing the observed color indices of many stars with unreddened objects of the same spectral type. The ratio R is then obtained from equation (2.15b) by measuring the limit of $E_{\lambda-\rm V}/E_{\rm B-V}$ for long wavelengths, under the assumption that A_λ tends to zero in this limit. In practice, observations out to the mid-infrared suffice for this purpose.

2.3.2 Transfer of Radiation

Thus far, we have considered extinction only as it alters the observations of stars at fixed distances. Let us now change perspective and look at the progressive effect of dust on the light propagating from the surface of a star. We begin with some definitions. Any radiation field may be described by I_ν, its *specific intensity*. This quantity is defined so that $I_\nu\,\Delta\nu\,\Delta\mathcal{A}\,\Delta\Omega$ is the energy per unit time, with frequency between ν and $\nu + \Delta\nu$, that propagates normal to the area $\Delta\mathcal{A}$ and within the solid angle $\Delta\Omega$ (see Figure 2.8). Note that we could just as easily have defined a quantity I_λ within a certain *wavelength* range; we will freely alternate between ν and λ as our independent variable. In general, I_ν is a function not only of ν and the spatial position relative to the radiation source, but also of the propagation direction, where the latter

---

[3] A useful fact is that A_V, when averaged over all lines of sight within the Galactic plane, increases by 1.9 mag kpc^{-1} away from the Sun. In Chapter 3, we will present a general relation between A_V and hydrogen column density.

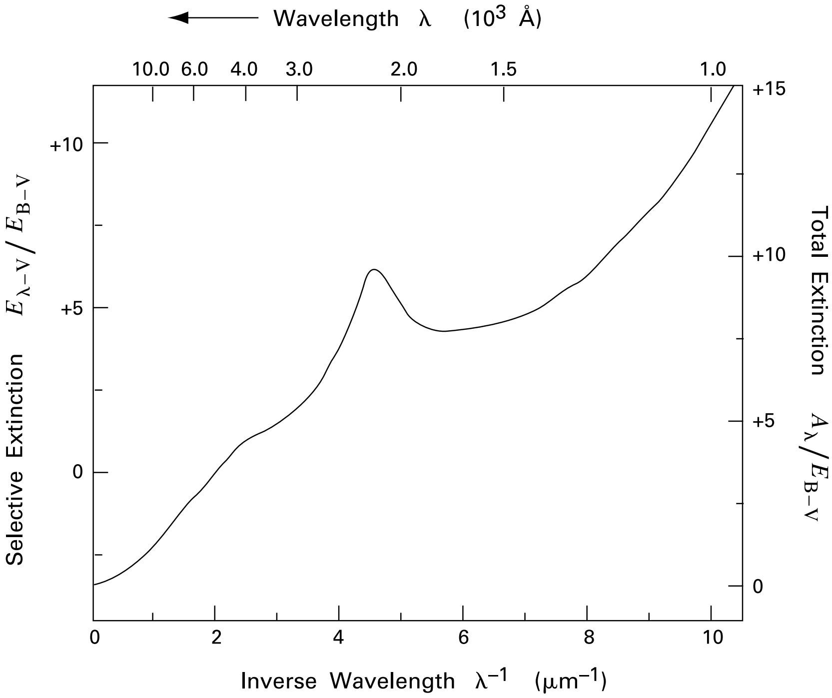


Figure 2.7 Interstellar extinction curve. Plotted as a function of inverse wavelength λ^{-1} are both the selective and total extinction. Note that the true wavelength λ is shown on the top horizontal scale.

is indicated by the unit vector $\hat{\boldsymbol{n}}$ in Figure 2.8. We next define the *specific flux* F_ν, also known as the *flux density*. This is the monochromatic energy per area per unit time passing through a surface of *fixed* orientation $\hat{\boldsymbol{z}}$, and is given by

$$F_\nu \equiv \int I_\nu \, \mu \, d\Omega \ . \tag{2.17}$$

Here $\mu \equiv \hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{z}}$ is the cosine of the angle between the propagation direction and the surface normal (Figure 2.8). Since light travels at the speed c, I_ν/c represents the energy density propagating in the direction $\hat{\boldsymbol{n}}$. The total energy density per unit frequency at a fixed location is therefore

$$u_\nu \equiv \frac{1}{c} \int I_\nu \, d\Omega \ . \tag{2.18}$$

This quantity is closely related to the *mean intensity* J_ν, *i. e.*, the average of I_ν over all directions:

$$J_\nu \equiv \frac{1}{4\pi} \int I_\nu \, d\Omega \ . \tag{2.19}$$

Consider now ΔI_ν, the change of I_ν along the direction of propagation (Figure 2.9). Over a small distance Δs, dust grains can remove radiation from the beam through several processes. One is *absorption*, in which the radiative energy is transformed to internal motion of the grain

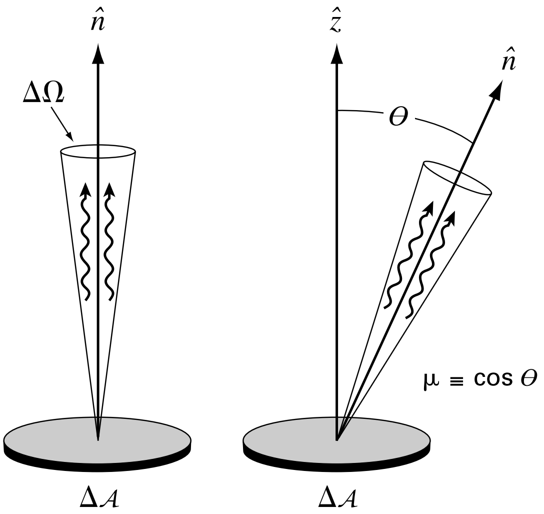


Figure 2.8 Definition of the specific intensity I_ν. We imagine light propagating within a cone of solid angle $\Delta\Omega$ centered on a direction $\hat{\boldsymbol{n}}$. The vector $\hat{n}$ can itself be tilted at some angle θ relative to $\hat{\boldsymbol{z}}$. The latter defines the normal direction to the surface area $\Delta\mathcal{A}$ of interest.

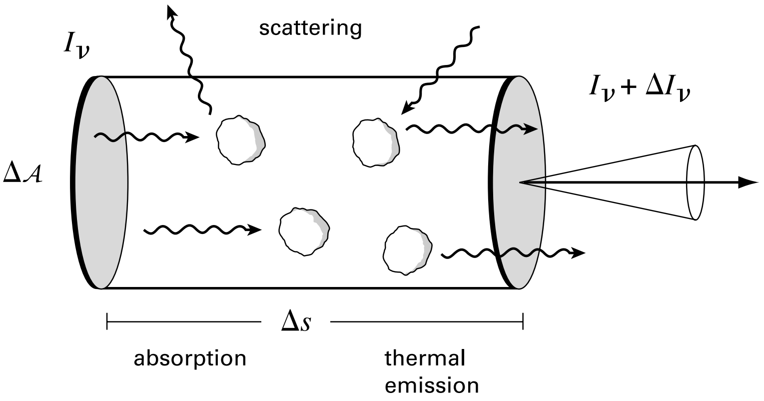


Figure 2.9 Passage of light through dust grains. Radiation with specific intensity I_ν enters an area $\Delta\mathcal{A}$ at the left. By the time it travels a distance Δs, the intensity changes to $I_\nu + \Delta I_\nu$ through absorption, scattering and thermal emission by the grains.

lattice. Photons are also lost from the beam by *scattering*. Here, the induced excitation quickly decays and reemits another photon in a different direction. This second photon has a frequency identical to the first in the rest frame of the grain. An external observer sees a slight Doppler shift associated with the finite grain velocity.

For present purposes, we may lump absorption and scattering together, and simply note that the total extinction must be proportional to Δs. In addition, the rate of photon removal should vary linearly with the incident flux, *i. e.*, with I_ν itself. Finally, for a gaseous medium with a given admixture of grains, the extinction must be proportional to the total density ρ. We therefore write this negative contribution to ΔI_ν as $-\rho\kappa_\nu I_\nu$. Here κ_ν is the *opacity*, a quantity measured in cm^2 g^{-1} that depends on the incident frequency ν, the relative number of grains, and their intrinsic physical properties.

There are also several ways that I_ν can be increased in Δs. The lattice vibrations excited by radiation also emit it, generally at infrared wavelengths. In addition to such *thermal emission*, photons are added to the beam by being scattered from beams propagating in other directions. As before, we lump these processes together and define an *emissivity* j_ν, such that $j_\nu\,\Delta\nu\,\Delta\Omega$ is the energy per volume per unit time emitted into the direction $\hat{\boldsymbol{n}}$. Writing the elementary volume as $\Delta\mathcal{A}\,\Delta s$ and recalling that I_ν is defined as an energy per unit *area*, we see that the augmentation to I_ν is simply $j_\nu\,\Delta s$. The full change to I_ν is thus

$$\Delta I_\nu = -\rho\,\kappa_\nu\,I_\nu\,\Delta s \,+\, j_\nu \Delta s \ ,$$

Dividing through by Δs, we obtain the *equation of transfer*:

$$\frac{dI_\nu}{ds} = -\rho\,\kappa_\nu\,I_\nu \,+\, j_\nu \ . \tag{2.20}$$

Although we have motivated this result for the specific case of radiation propagating through dust grains, both equation (2.20) and the associated terminology are applicable to any continuous medium which can remove or generate photons. For example, we will use the equation of transfer in Chapter 6 to discuss the interaction of radio waves with interstellar molecules.

The quantity $1/\rho\kappa_\nu$, which has the dimensions of a length, is known as the *photon mean free path*. The ratio of Δs to this length, *i. e.*, the product $\rho\kappa_\nu\Delta s$, is the *optical depth*, denoted $\Delta\tau_\nu$.[4] When a photon propagates through an *optically-thick* medium, *i. e.*, one for which $\Delta\tau_\nu \gg 1$, it has a high probability of extinction. Conversely, radiation can travel freely in an *optically-thin* environment ($\Delta\tau_\nu \ll 1$). Note that the same physical medium can be optically thin or thick, depending on the frequency of the radiation in question.

2.3.3 Extinction and Optical Depth

At this point, we have two nondimensional measures of extinction, the optical depth and the quantity A_λ introduced in equation (2.12). To see the relation between them, we first use the equation of transfer to obtain the specific intensity at a point P located a distance r from the center of a star (Figure 2.10). The plan is to use this result to obtain the star's apparent magnitude at P. Suppose we measure the specific intensity at a frequency where the intervening dust has negligible emission, so that $j_\nu = 0$ in equation (2.20). Reverting from frequency to wavelength, we integrate the equation along any ray from the stellar surface to P, obtaining

$$I_\lambda(r) = I_\lambda(R_*)\,\exp\left(-\Delta\tau_\lambda\right) \ . \tag{2.21}$$

Here $\Delta\tau_\lambda$ denotes the optical depth from the stellar radius to P. Referring to Figure 2.10, we see that this depth depends on the precise location of the emitting point within the cone converging on P. However, we suppose that $r \gg R_*$, so that this variation can safely be ignored. Similarly, $I_\lambda(R_*)$ is independent of propagation direction within the cone if the stellar surface is assumed to radiate like a blackbody (see below).

---

[4] Here we follow the imprecise, but accepted, procedure of employing the term "optical" depth even for frequencies outside the visual regime.

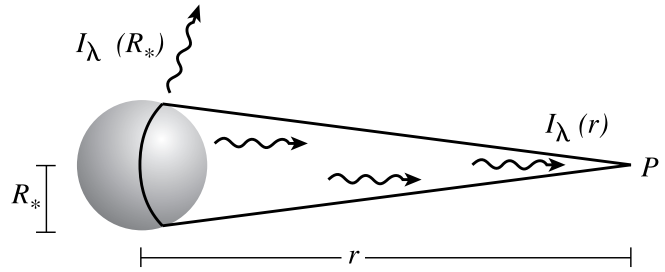


Figure 2.10 Attenuation of starlight by the interstellar medium. The point P, located a distance r from a star of radius R_*, receives a specific intensity $I_\lambda(r)$. This intensity is reduced from the value $I_\lambda(R_*)$ emitted isotropically from the stellar surface.

Our next step is to use equation (2.17) to find the specific flux received at P. In the present case, $\mu \approx 1$, and the small solid angle $\Delta\Omega$ subtended by the star is $\pi R_*^2/r^2$. Thus

$$F_\lambda(r) = \pi\, I_\lambda(R_*) \left(\frac{R_*}{r}\right)^2 \exp\left(-\Delta\tau_\lambda\right) \ . \tag{2.22}$$

Imagine now that the same star were located at some other distance $r_\circ$ from P, and that there were no intervening extinction. Writing the received flux in this case as $F_\lambda^*(r_\circ)$, we have simply

$$F_\lambda^*(r_\circ) = \pi\, I_\lambda(R_*) \left(\frac{R_*}{r_\circ}\right)^2 \ . \tag{2.23}$$

If we now divide equation (2.22) by equation (2.23), the result can be written in the form

$$-2.5 \log F_\lambda(r) = -2.5 \log F_\lambda^*(r_\circ) \,+\, 5 \log\!\left(\frac{r}{r_\circ}\right) \,+\, 2.5\ (\log e)\, \Delta\tau_\lambda \ . \tag{2.24}$$

Referring to Equations (1.2) and (1.3), we let $r_\circ$ equal 10 pc. We then add the constant $m_{\lambda\circ}$ to both sides of equation (2.24) to find

$$m_\lambda = M_\lambda \,+\, 5 \log\left(\frac{r}{10\ \mathrm{pc}}\right) \,+\, 2.5\ (\log e)\, \Delta\tau_\lambda \ . \tag{2.25}$$

Comparison of this equation with (2.12) yields the desired result:

$$\begin{aligned} A_\lambda &= 2.5\ (\log e)\, \Delta\tau_\lambda \\ &= 1.086\ \Delta\tau_\lambda \ . \end{aligned} \tag{2.26}$$

Thus, the two measures of extinction actually give very similar numerical values.

2.3.4 Blackbody Radiation

Let us now examine quantitatively the grains' thermal emission. Imagine surrounding a portion of interstellar gas, with its admixture of grains, by a container whose walls are maintained at temperature T. Suppose further that these walls absorb all the photons impinging on them, while the gas is transparent to this radiation. Then the heated walls will generate their own photons, and the interior of the container will be filled with radiation that is in *thermal equilibrium* with the walls. This means that the distribution of photons among the available quantum states is the most probable one consistent with the free exchange of energy between radiation and matter. The radiation energy density under these conditions is given by the Planck formula:

$$u_\nu = \frac{8\pi h\nu^3/c^3}{\exp(h\nu/k_BT) - 1} \ . \tag{2.27}$$

The so-called *blackbody radiation* we have just described is also isotropic, *i. e.*, the specific intensity I_ν is independent of direction. From equation (2.18), we have $I_\nu = cu_\nu/4\pi$. The specific intensity in this case is given the special symbol B_ν and is a function of temperature alone:

$$B_\nu(T) = \frac{2h\nu^3/c^2}{\exp(h\nu/k_BT) - 1} \ . \tag{2.28}$$

We may also define a $B_\lambda(T)$ by using

$$\begin{aligned} B_\lambda &= B_\nu \left(\frac{d\nu}{d\lambda}\right) \\ &= -\left(\frac{c}{\lambda^2}\right) B_\nu \end{aligned}$$

to obtain

$$B_\lambda(T) = \frac{2hc^2/\lambda^5}{\exp(hc/\lambda k_BT) - 1} \ . \tag{2.29}$$

Within our hypothetical container, the spatial uniformity of the radiation field implies, from equation (2.20), that the emissivity of the matter obeys

$$(j_\nu)_{\rm therm} = \rho\, \kappa_{\nu,{\rm abs}}\, B_\nu(T) \ , \tag{2.30}$$

where we have indicated that only the *absorption* component of κ_ν counts here. Now remove the container walls. The same matter must emit thermally at precisely the same rate. In applying equation (2.30), we must take care to use T_d, the *temperature of the dust grains*, which may be quite different from the temperature of the surrounding gas (see Chapter 7). [5]

Figure 2.11 plots the important function $B_\nu(T)$ for several values of T. We see that increasing T raises the intensity at all frequencies but maintains the shape of the curve. By making the substitution $x \equiv h\nu/k_BT$ in equation (2.28), the reader may verify that $B_\nu(T)$ reaches its maximum at $x_\circ = 2.82$, so that

$$\begin{aligned} \frac{\nu_{\rm max}}{T} &= \frac{x_\circ\, k_B}{h} \\ &= 5.88 \times 10^{10}\ {\rm Hz\ K^{-1}} \ . \end{aligned} \tag{2.31}$$

Similarly, from substitution of $y \equiv hc/\lambda k_BT$ into equation (2.29), it follows that $B_\lambda(T)$ peaks at the wavelength $\lambda_{\rm max}$, where

$$\lambda_{\rm max}\, T \ = \ 0.29\ {\rm cm\ K} \ . \tag{2.32}$$

---

[5] Equation (2.30) is a statement of Kirkhoff's law; see also Appendix E. Note that we adopt the convention that a subscript d denotes *dust*, while g denotes *gas*.

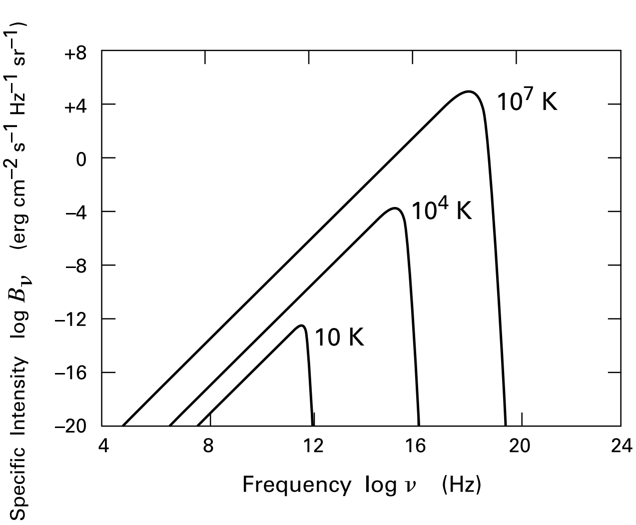


Figure 2.11 The specific intensity $B_\nu(T)$ of blackbody radiation. The quantity is shown as a function of frequency for three representative temperatures.

Equations (2.31) and (2.32) are alternate forms of *Wien's displacement law*. This relation, especially in the form of equation (2.32), has great practical utility.

For future reference, we note that the blackbody relations in Equations (2.27) through (2.29) also apply to the interiors of stars. Any star is so optically thick at all frequencies that its matter and radiation are very nearly in thermal equilibrium. Hence, the radiation energy density at an interior point with temperature T is given by equation (2.27). Moreover, the specific intensity of the light *leaving* the star is given by Equations (2.28) and (2.29), with T now being the effective temperature T_{eff} introduced in § 1.3. We therefore have

$$I_\lambda(R_*) \;=\; B_\lambda(T_{\text{eff}}) \; . \tag{2.33}$$

To find the specific *flux* leaving the surface, we return to equation (2.17), but recast in terms of wavelength. Writing $d\Omega \;=\; 2\pi d\mu$ and integrating from $\mu = 0$ to 1, we find:

$$F_\lambda(R_*) \;=\; \pi\, B_\lambda(T_{\text{eff}}) \; . \tag{2.34}$$

We may integrate this equation over all wavelengths to obtain the bolometric flux. Again letting $y \equiv hc/\lambda k_B T$, we find

$$F_{\text{bol}} \;=\; \frac{2\,\pi\, k_B^4\, T^4}{c^2\, h^3} \int_0^\infty \frac{y^3\, dy}{e^y \,-\, 1} \; .$$

The integral has the value $\pi^4/15$, so we may write

$$F_{\text{bol}} \;=\; \sigma_B\, T_{\text{eff}}^4 \; , \tag{2.35}$$

where the Stefan-Boltzmann constant is given by

$$\begin{aligned}\sigma_B &\equiv \frac{2\,\pi^5\, k_B^4}{15\, c^2\, h^3} \\ &= 5.67 \times 10^{-5}\ \text{erg}\,\text{cm}^{-2}\,\text{s}^{-1}\,\text{K}^{-4} \; .\end{aligned} \tag{2.36}$$

Multiplication of equation (2.35) by the stellar surface area $4\pi R_*^2$ leads to the relation between $L_{\rm bol}$ (or L_*) and $T_{\rm eff}$ in equation (1.5).

Our manipulations, incidentally, have allowed us to determine the *total* energy density in any blackbody radiation field. That is, we can now integrate equation (2.27) over all frequencies. It is conventional to write the result of this integration as

$$u_{\rm rad} \,=\, a\,T^4 \ . \tag{2.37}$$

Here, we have introduced the radiation density constant $a \,\equiv\, 4\sigma_B/c$, which has the numerical value $7.56 \times 10^{-15}\,{\rm erg\,cm^{-3}\,K^{-4}}$.

2.4 Interstellar Dust: Properties of the Grains

We next turn to the physical characteristics of the dust grains themselves, as deduced from both their effect on background starlight and their own emission. This information adds, of course, to our general understanding of the interstellar medium. But grain properties are also intimately linked, as we shall see, with those of star-forming clouds.

2.4.1 Efficiency of Extinction

The opacity κ_ν represents the total extinction cross section per mass of interstellar material. If we now wish to explore the contribution of each individual grain, a change of notation is in order. Since the gas itself contributes only a minor part of the extinction in the interstellar medium, we may write

$$\rho\,\kappa_\nu \,=\, n_d\,\sigma_d\,Q_\nu \ . \tag{2.38}$$

Here n_d is the number of dust grains per unit volume.The quantity σ_d, with the dimensions of area, is the geometrical cross section of a typical grain, taken for simplicity to be spherical. That is, $\sigma_d \,\equiv\, \pi a_d^2$, where a_d is the grain radius. The ratio of the actual extinction cross section to this projected cross section is Q_ν, the *extinction efficiency factor*. As usual, Q_ν can be written as the sum of absorption and scattering components, which we will denote $Q_{\nu,{\rm abs}}$ and $Q_{\nu,{\rm sca}}$, respectively.

The empirical form of Q_ν (or its equivalent Q_λ) can be obtained from the interstellar extinction curve, Figure 2.7, along with some theoretical input to set the scale. Equations (2.26) and (2.38) make it clear that the wavelength variation of the efficiency factor is just

$$\frac{Q_\lambda}{Q_{\lambda_\circ}} = \frac{A_\lambda/E_{\rm B-V}}{A_{\lambda_\circ}/E_{\rm B-V}} \ , \tag{2.39}$$

where $\lambda_\circ$ is any reference wavelength. In the limit of *zero* wavelength, classical electromagnetic theory gives the result that both $Q_{\nu,{\rm abs}}$ and $Q_{\nu,{\rm sca}}$ approach unity, yielding a total efficiency factor of 2. Figure 2.7 shows that $A_\lambda/E_{\rm B-V} \,=\, 14$ at the last available data point of $\lambda \,=\, 1000\,\text{Å}$. Theory indicates that the extinction rises only slightly at shorter wavelengths, so we may apply this number as our asymptotic value to conclude that

$$Q_\lambda \,=\, 0.14\,\frac{A_\lambda}{E_{\rm B-V}} \ . \tag{2.40}$$

Any theoretical model of grain composition must reproduce this basic connection to observations. For example, the behavior of the extinction curve tells us that Q_λ varies roughly as λ^{-1} in the optical regime. At longer wavelengths, there is a local maximum near 10 µm, not evident in the figure. There is also a more prominent peak in the ultraviolet, at 2200 Å. Finally, grain models must account for the scattering and polarization of the interstellar medium observed at visible and near-infrared wavelengths.

In the context of star formation, there has been much effort to determine Q_λ at far-infrared and millimeter wavelengths. The general interstellar medium is transparent in this regime, so one turns instead to the observed emission from heated dust clouds. Suppose the total optical depth $\Delta\tau_\lambda$ through a cloud subtending an angle $\Delta\Omega$ is less than unity. Suppose further that the cloud is sufficiently transparent in the optical that A_V can be determined through counts of background stars, a technique we will discuss in Chapter 6. Then Equations (2.20) and (2.30) imply that the specific intensity leaving the cloud is approximately

$$I_\lambda \,=\, B_\lambda(T_d)\,\Delta\tau_\lambda \ .$$

Here, we have ignored the absorption term in (2.20). Application of equation (2.17) shows that the received flux can be written as

$$F_\lambda \,=\, B_\lambda(T_d)\,\Delta\Omega\,\Delta\tau_\lambda \qquad\qquad \Delta\tau_\lambda \ll 1 \ , \tag{2.41}$$

where we have also assumed that $\Delta\Omega$ is small. If the dust temperature is known through other observations or theoretical considerations, equation (2.41) yields $\Delta\tau_\lambda$. Knowledge of A_V, together with Equations (2.16), (2.26), and (2.40) then gives Q_λ.

The determination of T_d and A_V is usually problematic, so that the wavelength dependence of Q_λ in this regime is also poorly determined. It is conventional to write this dependence, for 30 µm $\lesssim \lambda \lesssim 1$ mm, as $\lambda^{-\beta}$. Here, β is a positive number, which is thought to decline from about 1 to 2 over this wavelength range. There is also evidence that β is generally smaller in the densest clouds and circumstellar disks, but closer to 2 in more diffuse environments. This difference may reflect physical agglomeration of the grains in the denser regions.[6] The cumulative effect of these considerations is that Q_λ, and hence the opacity, is presently uncertain by an order of magnitude at $\lambda = 1\ \mathrm{mm}$. We will see later that this circumstance hinders the attempt to measure disk masses.

2.4.2 Size Distribution and Abundance

Returning to the issue of grain structure, most extinction observations can be accommodated if the particles consist of refractory *cores* surrounded by icy *mantles*. The cores are rich in silicates, such as the mineral olivine found in terrestrial rocks. Silicates can account for the 10 µm feature through the vibration of the SiO bond. Traditionally, the 2200 Å spike has been attributed to electronic excitation of graphite, which has therefore been added as another core material. However, the failure to detect graphite in comets and meteorites has cast some doubt on this interpretation. The mantles consist of a mixture of water ice and other molecules presumably

---

[6] As we have just seen, the opacity of a grain does not depend on the wavelength λ of absorbed radiation once the geometric size becomes larger than λ. From equation (2.38), the wavelength-dependence of the efficiency factor also vanishes in this limit. Hence, centimeter-size grains have a small exponent β in the millimeter.

adsorbed from the surrounding gas. Such mantles can persist within cold interstellar clouds, but sublimate once the grain temperature exceeds about 100 K.

How large are the grains? In most models, a radius of $a_d \sim 0.1\,\mu\text{m}$ is adopted, and this frequently serves as a rough estimate. However, it is clear that a continuous distribution of sizes is necessary to match the extinction data. The most commonly used distribution is that of Mathis, Rumpl, and Nordsiek. Here, the relative number of grains per interval in radius varies as $a_d^{-3.5}$, with upper and lower cutoffs at 0.25 µm and 0.005 µm, respectively. Equations such as (2.38) are therefore more correctly written as integrals over the size distribution, but we will not need this refinement. [7]

Within our simplified picture of uniform spherical grains, we imagine an HI cloud with hydrogen number density n_H. We will later find it useful to know Σ_d, the total geometric cross section of grains per hydrogen atom:

$$\Sigma_d \equiv \frac{n_d\,\sigma_d}{n_H} \ . \tag{2.42}$$

If our hypothetical cloud has a length L along the line of sight, then Σ_d can also be written as a ratio of column densities:

$$\Sigma_d = \frac{N_d\,\sigma_d}{N_H} \ , \tag{2.43}$$

where $N_d \equiv n_d L$ and $N_H \equiv n_H L$. We can evaluate the numerator in this relation by first noting that equation (2.38) can be rewritten, after multiplication by L, as

$$\Delta\tau_\lambda = N_d\,\sigma_d\,Q_\lambda \ . \tag{2.44}$$

The quantity $\Delta\tau_\lambda/Q_\lambda$ can be found using equations (2.26) and (2.40). Applying the result to equation (2.42), we find

$$\Sigma_d = 7.8\left(\frac{E_{\rm B-V}}{N_H}\right)\,\text{cm}^2 \ . \tag{2.45}$$

The ratio of color excess to hydrogen column density in equation (2.45) has been established empirically, through an important set of observations. To measure N_H in any region, we take advantage of the fact that O and B stars located behind clouds can excite electronic transitions in the intervening hydrogen. For diffuse clouds, the stars are still visible but have additional absorption lines superposed. These lines are in the ultraviolet and can only be observed above the Earth's atmosphere. In 1972, an ultraviolet spectrometer aboard the OAO-2 satellite first measured the Lyα transition in the spectra of 69 O and B stars. The depth of the absorption line translates into a hydrogen column density. In addition, the B and V magnitudes of these same stars were observed in order to ascertain their color excess. The two quantities are well correlated, with a linear relationship:

$$\frac{E_{\rm B-V}}{N_H} = 1.7\times 10^{-22}\,\text{mag}\,\text{cm}^2 \ . \tag{2.46}$$

---

[7] The concept of an average grain size is useful because a number of important effects increase with the particle radius. Such effects include the extinction of starlight and the catalysis of H_2 formation (Chapter 5). In these cases, the properly weighted average radius is not far below the upper cutoff of the distribution. An exception is photoelectric heating (Chapter 7), which is so efficient at small radii that a full integration is necessary.

Applying this result to equation (2.45) finally yields

$$\Sigma_d = 1.0 \times 10^{-21}\,\mathrm{cm}^2 \ . \tag{2.47}$$

This evaluation of Σ_d allows us to obtain a convenient expression for κ_λ in terms of the standard extinction curve. Referring to equation (2.38), we first write ρ as $m_H n_H / X$. We then solve this equation for the opacity to find

$$\kappa_\lambda = \frac{X\, n_d\, \sigma_d\, Q_\lambda}{m_H\, n_H} \ .$$

Using the definition of Σ_d and its numerical value, we have

$$\begin{aligned} \kappa_\lambda &= 420\ \mathrm{cm}^2\,\mathrm{g}^{-1}\, Q_\lambda \\ &= 59\ \mathrm{cm}^2\,\mathrm{g}^{-1}\, \frac{A_\lambda}{E_{\mathrm{B-V}}} \ , \end{aligned} \tag{2.48}$$

where we have also employed equation (2.40).

It is also instructive to estimate f_d, the mass fraction of the interstellar medium contained in grains. Within the picture of uniform spheres, this fraction is

$$f_d = \frac{4\pi a_d^3 \rho_d}{3\mu m_H} \left(\frac{n_d}{n_H} \right) \ , \tag{2.49}$$

where ρ_d, the internal grain density, is about 3 g cm^{-3}. The number fraction n_d/n_H is just $\Sigma_d/\pi a_d^2$, which is 3×10^{-12} for $a_d = 0.1\,\mu\mathrm{m}$. We thus find for the mass fraction

$$\begin{aligned} f_d &= \frac{4\rho_d a_d \Sigma_d}{3\mu m_H} \\ &= 0.02 \ . \end{aligned} \tag{2.50}$$

Since this number matches the metallicity of the gas, we confirm that a large proportion of heavy elements must be locked up in solid form.

2.4.3 Polarization of Starlight

Another important aspect of grains is their ability to polarize radiation. Consider the visible reflection nebulae surrounding bright, young stars. Here, stellar photons travel unimpeded through gaps in patchy cloud gas until they encounter a grain and are scattered in our direction. Prior to the scattering event, the incident electric field vector $\boldsymbol{E}$ oscillates randomly within the plane normal to the propagation direction $\hat{\boldsymbol{n}}$ (see Figure 2.12). Now focus on radiation that scatters into directions 90° from $\hat{\boldsymbol{n}}$, such as $\hat{\boldsymbol{s}}$ or $\hat{\boldsymbol{s}}'$ in the figure. These new directional vectors define their own normal planes. The scattered field $\boldsymbol{E}$ only oscillates along the line that is the projection of the new plane and the old. Hence, this radiation is linearly polarized. Scattering into other directions, such as $\hat{\boldsymbol{s}}''$, results in partial polarization. That is, $\boldsymbol{E}$ again oscillates along two orthogonal lines, but with unequal amplitude.

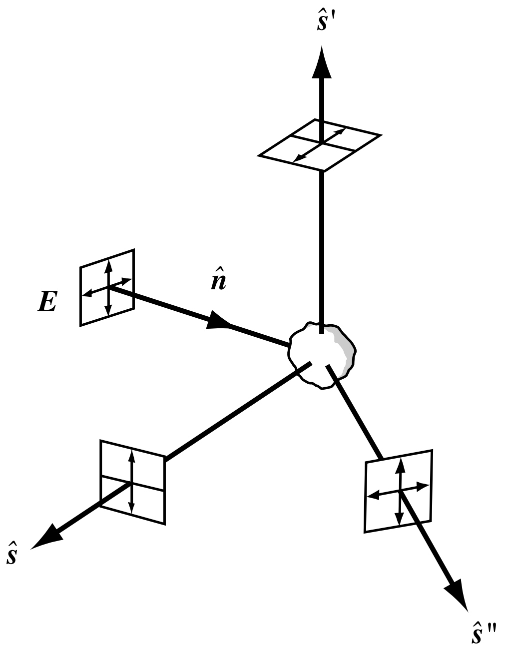


Figure 2.12 Polarization of light by grain scattering. The incident radiation propagates along the direction $\hat{n}$. That portion scattered into orthogonal directions $\hat{s}$ or $\hat{s}'$ is linearly polarized, while that in an arbitrary direction $\hat{s}''$ is only partially polarized.

If one rotates a polarizing filter in front of a source, the received intensity at any wavelength varies from some lowest value $I_{\rm min}$ up to $I_{\rm max}$. The orientation of the polarizer corresponding to $I_{\rm max}$ is the position angle of $\boldsymbol{E}$, while the degree of polarization is defined to be

$$P \equiv \frac{I_{\rm max} - I_{\rm min}}{I_{\rm max} + I_{\rm min}} \ . \tag{2.51}$$

This quantity is a function of wavelength, since $Q_{\lambda,\rm sca}$ generally decreases with λ. Consequently, reflection nebulae appear bluer than their central stars. Values of P in the optical are typically a few percent in the reflection nebulae surrounding young stars, but can be as high as 0.2. Note that the change in polarization direction (*i. e.*, the orientation of $\boldsymbol{E}$) across such a nebula depends on the position of the star relative to the line of sight. In many cases, the pattern of $\boldsymbol{E}$-vectors accurately locates an illuminating source that is too embedded for direct observation.

In the presence of a magnetic field, grains also polarize radiation through *dichroic extinction*. This phenomenon depends on the fact that the particles are *not* perfect spheres, but irregular structures that tend to rotate end over end, *i. e.*, about their shortest principal axes. Grain material also has a small electric charge and is paramagnetic. Both features cause it to acquire a magnetic moment $\boldsymbol{M}$ that points along the instantaneous axis of rotation. Interaction with the ambient magnetic field then creates a torque $\boldsymbol{M} \times \boldsymbol{B}$ that gradually forces the grain's short axis to align with the field. Figure 2.13 illustrates this situation, for the case of an idealized, cylindrical grain.

Consider now unpolarized radiation that impinges on this rotating grain. The electric field is most effective in driving charges down the body's *long* axis, This direction therefore becomes the one of maximum absorption for the electromagnetic wave. As seen in Figure 2.13, the electric vector of the transmitted radiation lies along the ambient $\boldsymbol{B}$. The reader may check that the effect is greatest when the propagation direction of the incident radiation is nearly orthogonal

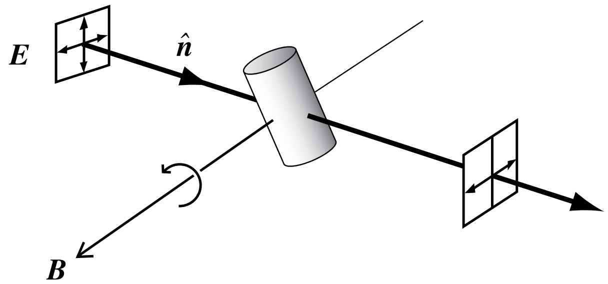


Figure 2.13 Polarization by dichroic extinction. Radiation propagating in direction $\hat{\boldsymbol{n}}$ impacts an elongated grain rotating about the magnetic field $\boldsymbol{B}$. The transmitted radiation is polarized in the direction of $\boldsymbol{B}$.

to the field. The observation of polarized light from background stars is a powerful method for mapping the magnetic field throughout the Galaxy.

2.4.4 PAHs

Returning to the reflection nebulae around massive stars, we often find that such regions emit copiously at far-infrared wavelengths, in addition to the optical and ultraviolet. This radiation represents the thermal emission from the grains themselves. For a star at distance r, the flux on the grain falls as r^{-2}. With $Q_{\lambda,\mathrm{abs}}$ varying as λ^{-2}, the thermal emission is proportional to T_d^6 (see Chapter 7). Hence, the grain temperature falls as $r^{-1/3}$. For a representative distance of 0.1 pc from a B star, T_d is about 100 K. The temperature is somewhat higher for smaller grains, for which the efficiency factor governing emission is lower. However, the high temperature-sensitivity of the emission implies that this effect is not strong.

A number of reflection nebulae display a quantitatively small, but significant, anomaly in their emission. While the lion's share of the luminosity is in the ultraviolet, visual and far infrared, about one percent is emitted at *near*-infrared wavelengths. Figure 2.14 gives one example. Here, the observed near-infrared flux is well matched by a blackbody at a temperature of about 1500 K, along with a narrower emission feature near 3 µm. The broadband emission could not possibly be reflected starlight, since the radiation of the illuminating B star is predominantly in the visible and ultraviolet. Furthermore, the derived temperature is an order of magnitude higher than the value for heated dust, even at the smallest sizes posited in the Mathis, Rumpl, and Nordsiek distribution. Most intriguingly, the temperature does *not* decline with distance from the star.

The solution to this puzzle is to consider grains so tiny that their total thermal energy content is much less than the energy of a *single stellar photon*. Since the energy of a photon, unlike the stellar flux F_ν, does not decrease with distance, the peak temperature T_d could also be distance-independent. Under such circumstances the very concept of an equilibrium dust temperature loses significance. The temperature of each grain would undergo stochastic jumps with every photon impact and then rapidly fall as the grain cools. Because of the steep temperature-dependence of the emissivity, the observed T_d would be heavily weighted toward the peak values.

How small does such a grain have to be? If ΔT_d is the temperature jump, then the grain lattice gains energy $3Nk_B\Delta T_d$, where N is the number of atoms. We equate this energy to the photon value $h\nu$ and solve for N:

$$N = \frac{h\nu}{3\,k_B\,\Delta T_d}\ , \tag{2.52}$$

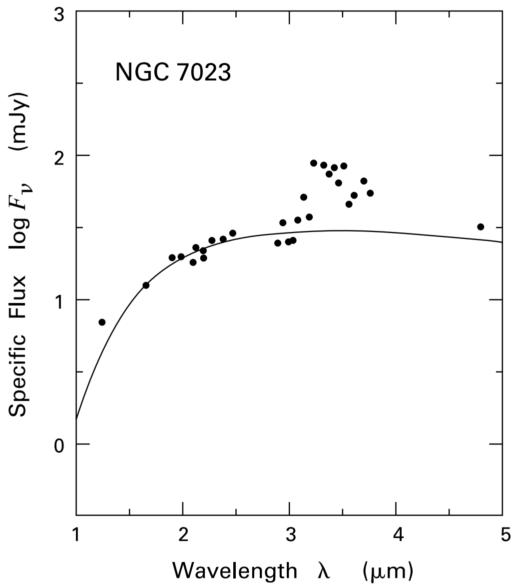


Figure 2.14 Near-infrared radiation from the reflection nebula NGC 7023. The specific flux F_ν is displayed as a function of wavelength. The solid curve shows the flux from a blackbody at 1500 K. Note the excess emission between 3 and 4 μm.

For a representative photon energy of 10 eV and temperature jump of 1000 K, we find that $N = 40$. If Δr is the lattice spacing between atoms, this number can be packed into a sphere with radius $R \approx N^{1/3}\Delta r = 10\,\text{Å}$, assuming a typical Δr of 3 Å.

The 3 μm emission spike seen in Figure 2.14 is actually one of several that were unexplained in the standard grain models. Since the 1980s, much effort has been devoted to the laboratory study of tiny grains (or macromolecules) that can reproduce these features. The most promising are known chemically as *polycyclic aromatic hydrocarbons*, or PAHs. These consist of linked carbon rings lying in a single plane. Figure 2.15 shows a representative PAH together with its emission spectrum in the near infrared. Here, the peaks arise from vibration of the C-H and C-C bonds.

Despite their small cross section, PAHs within clouds are important as heating agents, facilitating the transfer of the energy in starlight to the interstellar gas. In this role, their planar structure helps, allowing electrons to be readily liberated by ultraviolet photons. Theoretical studies which include PAHs in order to match observed cloud temperatures find that their population relative to ordinary grains can still be found by extrapolation of the standard size distribution. This fact lends support to the view that the distribution itself has a universal character, reflecting essential aspects of the origin and dissemination of interstellar grains.

Chapter Summary

Much of the matter between stars consists of gas at about 100 K, where hydrogen is in atomic form (HI). Discrete HI clouds permeate the Galaxy and have large velocities in and out of the plane. Warmer and more rarefied gas, with a temperature closer to 10^4 K, also exists, but its structure is poorly known. So too is the morphology of even hotter gas at 10^6 K. All of these components appear to coexist in rough pressure equilibrium.

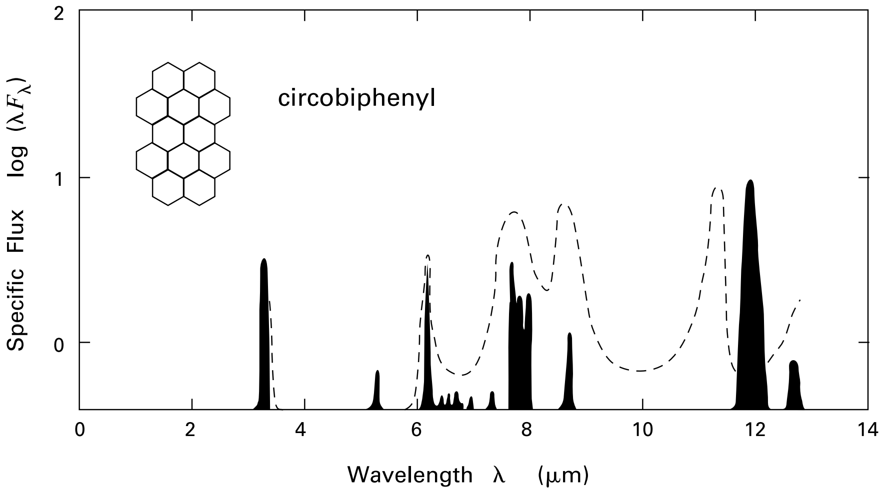


Figure 2.15 Emission from a candidate PAH (arbitrary units). The filled peaks represent emission calculated from the laboratory-measured absorption spectrum of the organic compound shown. The dashed curve displays, for comparison, the observed emission from the reflection nebula NGC 2023. Note again the spike near 3 μm.

Yet another form of hydrogen, and the most interesting from the perspective of star formation, is the molecular (H_2). While the total mass here is comparable to that in HI, the spatial distribution is very different. The molecular gas surface density in the Galactic disk attains a local peak at a radius of 6 kpc and then climbs steeply toward the center. Molecular cloud interiors are under a higher pressure than the other phases of the interstellar medium because of the influence of self-gravity.

Solid dust grains mixed in with the gas absorb background radiation and redden its color. Both effects are quantified as a function of wavelength by the interstellar extinction curve. Each grain consists of a relatively dense core surrounded by an icy mantle. Radiation from the dust around hot stars shows that the larger grains attain equilibrium between stellar heating and thermal emission. However, the tiniest structures, known as PAHs, reveal themselves through their anomalously high temperatures.

Suggested Reading

Section 2.1 The astronomical significance of the 21 cm line from atomic hydrogen was pointed out by van de Hulst in 1945. For an English translation of the paper, see

van de Hulst, H. C. 1979, in *A Source Book of Astronomy and Astrophysics: 1900-1975*, ed. K. R. Lang and O. Gingerich (Cambridge: Harvard U. Press), p. 627.

The line was actually first detected by

Ewen, H. I. & Purcell, E. M. 1951, Nature, 168, 356

Muller, C. A. & Oort, J. 1951, Nature, 168, 357.

A useful review of the HI gas in the Galaxy is

> Burton, W. B. 1992, in *The Galactic Interstellar Medium*, ed. D. Pfenninger and P. Bartholdi (Berlin: Springer-Verlag), p. 1.

The distribution of molecular hydrogen, including the implications for spiral structure and cloud properties, is covered in

> Combes, F. 1991, ARAA, 29, 195,

while the HII component is summarized by

> Gordon, M. A. 1988, in *Galactic and Extragalactic Radio Astronomy*, ed. G. L. Verschuur and K. I. Kellerman (Berlin: Springer-Verlag).

Section 2.2 The modern concept of phases in the interstellar medium was introduced by

> Field, G. B., Goldsmith, D. W., & Habing, H. J. 1969, ApJ, 155, L149.

Two other important, early contributions were

> Cox, D. P. & Smith, B. W. 1974, ApJ, 189, L105
>
> McKee, C. F. & Ostriker, J. P. 1977, ApJ, 218, 148.

The first paper discusses effects of supernova remnants, while the second introduced the three-phase medium. A current review of the field is

> Dopita, M. A. & Sutherland, R. S. 2002, in *Astrophysics of the Diffuse Universe* (Berlin: Springer-Verlag), Chapter 14.

Sections 2.3 and 2.4 An introduction to the physical properties of interstellar dust grains and their astronomical significance is

> Whittet, D. C. B. 1992, *Dust in the Galactic Environment* (Bristol: Institute of Physics).

For a more detailed treatment of various aspects of grain physics, see

> Krügel, E. 2003, *The Physics of Interstellar Dust* (Bristol: Institute of Physics).

The grain size distribution was derived empirically by

> Mathis, J. S., Rumpl, W., & Nordsieck, K. H. 1977, ApJ, 217, 425,

while the relation between $E_{\rm B-V}$ and N_H is due to

> Bohlin, R. C., Savage, B. D., & Drake, J. F. 1978, ApJ, 224, 132.

A detailed theoretical study of dust opacity is

> Draine, B. T. & Lee, H. M. 1984, ApJ, 285, 89.

The ongoing effort to quantify the extinction at far-infrared and millimeter wavelengths is reviewed in

> Henning, Th., Michel, B., & Stognienko, R. 1995, Planet. Sp. Sci., 43, 1333.

The Formation of Stars
Steven W. Stahler and Francesco Palla

3 Molecular Clouds

Having surveyed the gas content of the Galaxy, we now take a closer look at the molecular component. Table 3.1 summarizes the physical properties of Galactic molecular clouds. For later reference, we have distinguished a number of cloud types, but any such classification scheme necessarily has a degree of arbitrariness. The diameter L, for example, is really a characteristic value within a range that blends into the adjacent type. Note that our listing is in order of increasing A_V, the typical visual extinction along a line of sight through the cloud interior. At the low end are *diffuse clouds*. These are relatively isolated entities, with comparable amounts of atomic and molecular hydrogen. The fact that A_V is near unity means that much of the light from background stars can actually traverse these objects. Absorption lines seen in such radiation, particularly in the ultraviolet, have proved to be of considerable value for studies of molecular abundances and chemical reaction networks. However, diffuse molecular clouds represent a minor fraction of interstellar gas and are never found to produce stars, so we will not be examining them in any detail.

We pass instead to the next category in Table 3.1, the giant molecular clouds. Here, the reader has already encountered one important example, the complex in Orion. We first discuss more systematically the properties of such structures. To aid in the analysis, we introduce and then utilize the virial theorem, a powerful tool for understanding the mechanical equilibrium of self-gravitating bodies. We then turn our attention to dense cores and Bok globules, the much tinier entities associated with the birth of individual stars.

3.1 Giant Molecular Clouds

Both HI and diffuse molecular clouds can persist for long periods by means of pressure balance. That is, the internal thermal motion of the gas is prevented from dispersing the cloud by the confining presence of a surrounding, more rarefied and warmer medium. In giant molecular clouds, we encounter an entirely different dynamical situation. Here the main cohesive force is the cloud's own gravity, while internal thermal pressure plays only a minor role in the overall

Table 3.1 Physical Properties of Molecular Clouds

Cloud Type	A_V (mag)	$n_{\rm tot}$ (cm^{-3})	L (pc)	T (K)	M (M$_\odot$)	Examples
Diffuse	1	500	3	50	50	ζ Ophiuchi
Giant Molecular Clouds	2	100	50	15	10^5	Orion
Dark Clouds						
Complexes	5	500	10	10	10^4	Taurus-Auriga
Individual	10	10^3	2	10	30	B1
Dense Cores/Bok Globules	10	10^4	0.1	10	10	TMC–1/B335

The Formation of Stars. Steven W. Stahler and Francesco Palla

ISBN: 3-527-40559-3

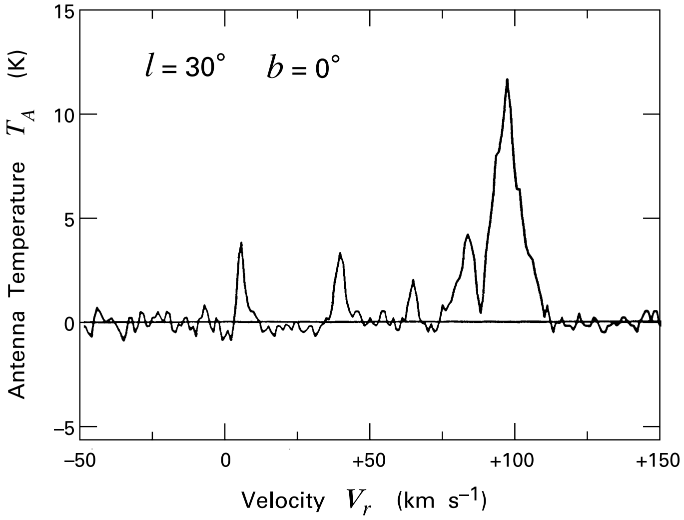


Figure 3.1 Typical 2.6 mm profile of $^{12}C^{16}O$ in the Galactic plane.

force balance. Such an environment is clearly more favorable to star formation, which proceeds through gravitational condensation.

3.1.1 Galactic Distribution

Within the Milky Way, over 80 percent of the molecular hydrogen resides in giant cloud complexes. The latter also account for most of the new star production. A typical giant cloud, we will show later, survives for 3×10^7 yr before it is destroyed by the intense winds from embedded O and B stars. On average, the cloud converts about 3 percent of its mass into stars during this time. Given the total H_2 mass in the Galactic disk of $2 \times 10^9\ M_\odot$, it follows that the star formation rate from giant clouds is about $2\ M_\odot$ yr^{-1}. This rate is a bit less than, but consistent with, $\dot{M}_*$ quoted in Chapter 1, which was based on the observed luminosities of massive stars and their theoretical lifetimes. Bearing in mind the significant uncertainty associated with the underlying data, such consistency is gratifying.

In the case of Orion, we have seen how the relatively rare O and B stars found throughout the complex are signposts for the local production of numerous low-mass objects. There are undoubtedly other pockets of star formation still hidden from view by the large column densities in dust. Generalizing from Orion, it is significant that *every* Galactic OB association thus far observed is closely associated with a giant molecular cloud. This fact was established by extensive observations in the 2.6 mm line of $^{12}C^{16}O$. Indeed, cloud complexes were initially discovered in the 1970s through CO studies of the sky near known HII regions, infrared sources, and areas of high visual extinction. Because it can be detected over large distances, the 2.6 mm line has remained the tool of choice for largescale surveys, both in our own and external galaxies.

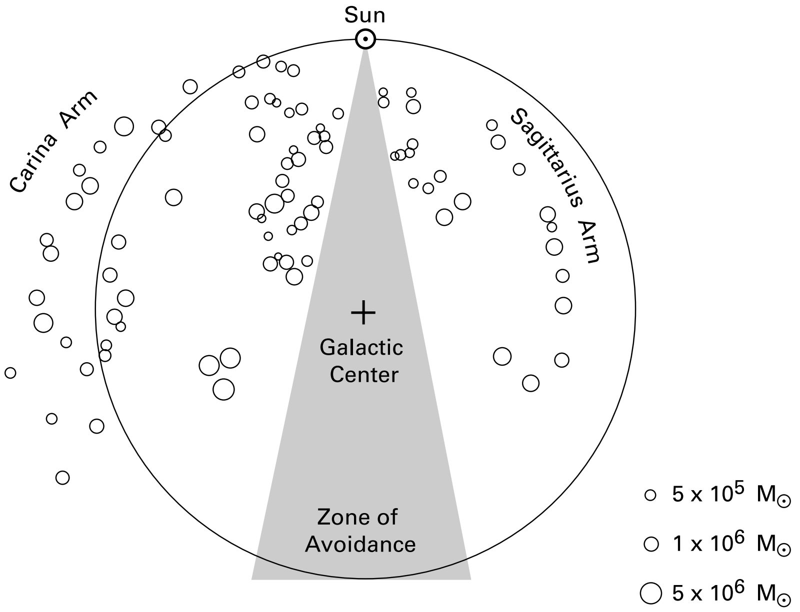


Figure 3.2 Galactic distribution of giant molecular clouds interior to the solar position. Note the "zone of avoidance," where radial velocities are too small for accurate position determination.

Figure 3.1 is a representative $^{12}C^{16}O$ spectrum. The figure follows radio astronomical convention by plotting, instead of the specific intensity I_ν, a proportional quantity known as the *antenna temperature*, T_A; this quantity is defined precisely in Appendix C. Note also that the independent variable in Figure 3.1 is not the frequency ν itself, but rather the "radial" gas *velocity* V_r along the line of sight. This is the velocity that would, through the Doppler effect, shift the line-center frequency $\nu_\circ$ into ν. Thus, V_r is given by $c\,(\nu_\circ - \nu)/\nu_\circ$. A positive V_r-value thus corresponds to a redshifted line ($\nu < \nu_\circ$).

The spectrum in Figure 3.1 shows a number of discrete peaks. Each of these represents an individual giant molecular cloud along the line of sight, which is here given in Galactic coordinates by $l = 30°$ and $b = 0°$ (*i. e.*, in the disk plane). The radial velocity of each cloud reflects its particular Galactocentric orbital speed. This speed varies in a known way with distance from the Galactic center. By taking a large number of spectra, therefore, one can map the distribution of clouds over a wide area of the disk. Moreover, the integrated value of T_A under each peak is a measure of the total cloud mass (see Chapter 6). Figure 3.2 displays the results of a cloud survey utilizing these tools. Here the observations are limited to Galactocentric radii comparable to or less than solar, and to very massive cloud complexes in excess of $10^5\,M_\odot$. The "zone of avoidance" toward the Galactic center is the region where the line-of-sight component of the clouds' circular velocity is too small to yield a reliable distance. Notice how the alignment of clouds in two regions delineates fragments of spiral arms. [1]

---

[1] In some renderings of Galactic structure, the two local features shown in Figure 3.2 are pieces of a contiguous "Sagittarius-Cygnus arm." A second global structure, the Perseus arm, passes outside the solar position.

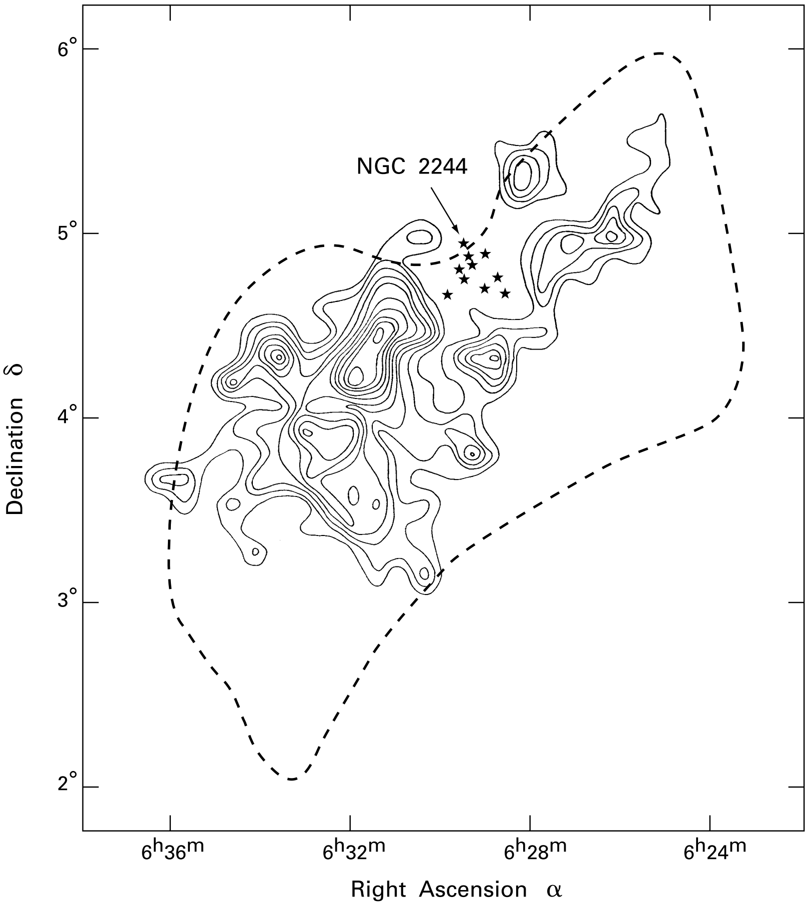


Figure 3.3 The Rosette Molecular Cloud, as seen in the 2.6 mm line of $^{12}C^{16}O$. The dashed contour indicates the boundary of the HI envelope.

Figure 3.3 is a map in $^{12}C^{16}O$ of the Rosette Molecular Cloud. This well-studied giant complex is located in the Monoceros region at a distance of 1.5 kpc. It appears as the small shaded region in Figure 1.1, just to the left of the Orion Molecular Cloud and inside the Milky Way band. The cloud contains the Rosette Nebula, an HII region generated by a compact group of five O stars. As was the case for the Orion Nebula, the massive stars are themselves embedded within a more extended cluster of low-mass members. The location of this cluster, designated NGC 2244, is also indicated in Figure 3.3.

3.1.2 Internal Clumps

The Rosette Molecular Cloud has an elongated, clumpy appearance reminiscent of the Orion cloud (Figure 1.2). We must bear in mind that the 2.6 mm line of $^{12}C^{16}O$ is always optically thick in giant molecular clouds and so emanates only from their surface layers. To probe internal structure, one can utilize the analogous line from a rarer isotope, such as $^{13}C^{16}O$. In this case, a photon from the deep interior has a smaller probability of absorption in the lower column density of ^{13}C^{16}O molecules. On the other hand, the received intensity is also less than for the main isotope, necessitating longer integration times at the telescope. Figure 3.4 is a ^{13}C^{16}O map of an interior portion of the Rosette cloud. Again we see that the emission presents a highly clumpy appearance. However, we can be certain that the peaks represent true internal density enhancements.

Any map is a two-dimensional projection onto the plane of the sky and therefore blends together structures that are physically distinct. As before, the radial velocity of the emitting gas can be effectively used as the third coordinate. The distribution of velocities is contained in the line spectrum at each sampled position, but it is cumbersome to display a two-dimensional array of all the spectra. Alternatively, one can take a linear slice through the full map and give the values of I_ν (or rather $T_A(V_r)$) only at points along that slice. Figure 3.5 shows the resulting *position-velocity diagram* for the cut at $b = -1.^\circ 75$; this cut is indicated in Figure 3.4. The diagram shows, for example, that the clump seen near $l = 207.^\circ 75$ actually consists of two overlapping structures separated in velocity by about 5 km s^{-1}.

There is an important difference between the velocity variation encountered in this high-resolution scan and that displayed by the complexes as a whole, illustrated in Figure 3.1. In the latter case, the velocities of the giant molecular clouds correlate systematically with Galactocentric radius. In contrast, the clump velocities within a complex appear to be *randomly dispersed* about a mean value. For the Rosette Molecular Cloud, this mean is +13 km s^{-1}, while the one-dimensional dispersion, *i. e.*, the root-mean-square deviation, of the radial velocity is 2.3 km s^{-1}. The simplest interpretation of this local dispersion is that the clumps represent a swarm of relatively high-density parcels that maintain their integrity as they move within the interior of the cloud complex.

By integrating the ^{13}C^{16}O intensity within the borders of individual clumps, one can reliably determine the clump mass and the volume-averaged density. We will be detailing this technique in Chapter 6. In the Rosette cloud, where the typical clump radius is 1.5 pc, the average mass thus obtained is 250 $M_\odot$, corresponding to a hydrogen density $n_H = 550$ cm^{-3}. Since the mean density for the entire complex is only 60 cm^{-3}, the clumps cannot occupy a large fraction of the volume.

Figure 3.6 shows the actual distribution of the clump mass M. Above a certain minimum, the number of clumps per unit mass, $\mathcal{N}$, falls off as a power law:

$$\mathcal{N} = \mathcal{N}_\circ \left(\frac{M}{M_{\rm min}}\right)^{-1.5} \qquad (M \geq M_{\rm min}) \ , \tag{3.1}$$

where $\mathcal{N}_\circ$ is a constant, and where $M_{\rm min} \approx 30\, M_\odot$. The pure power-law dependence is indicated in the figure by a dashed line. Other giant molecular clouds yield a similar result. Interestingly, cloud surveys such as that displayed in Figure 3.2 also find this power law for the masses of complexes as a whole. Such universality suggests that giant clouds are built up by the agglomeration of many clumps which were already distributed in mass according to equation (3.1).

Returning to our Rosette example, the structure of the complex is further clarified by examining the visual extinction through the cloud gas. By multiplying Equations (2.16) and (2.46), we first obtain the generally useful result

$$A_V/N_H = 5.3 \times 10^{-22}\,{\rm mag\,cm^2} \ . \tag{3.2}$$

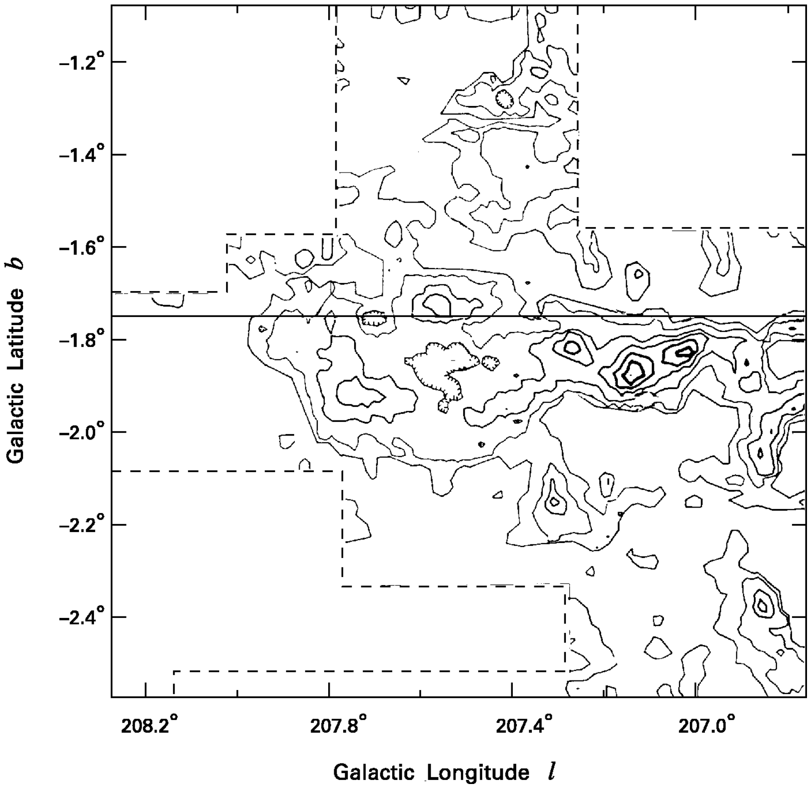


Figure 3.4 Map of the Rosette Molecular Cloud in $^{13}C^{16}O$. The horizontal line is the cut used in Figure 3.5.

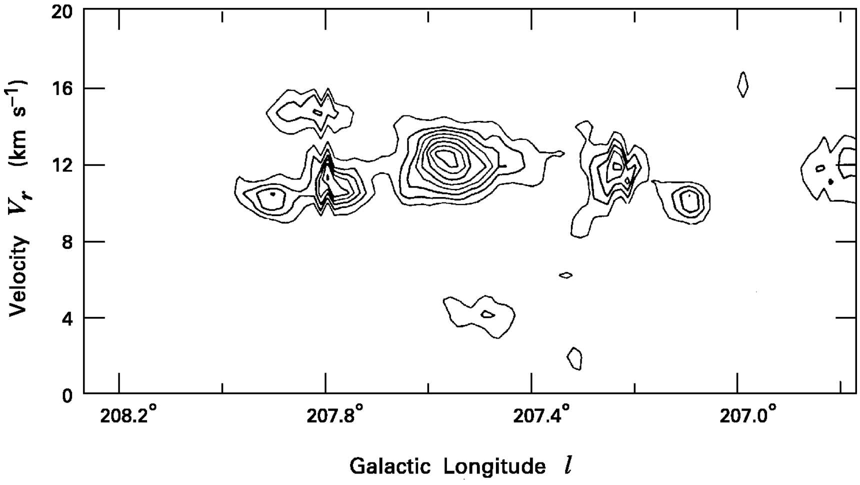


Figure 3.5 Position-velocity diagram for the cut indicated in Figure 3.4.

A typical line of sight through a spherical cloud of density n_H and radius R will penetrate a column density N_H of $n_H R$. We thus find that the average Rosette clump has an A_V of 1.0. For the complex as a whole, the mean A_V-value is 1.9, as determined by both $^{12}C^{16}O$ and $^{13}C^{16}O$ mapping. Since the interclump gas appears too tenuous to contribute significantly to A_V, these results imply that our generic line of sight through the complex intersects two clumps. In summary, the clumps, which comprise as much as 90 percent of the total mass of $1 \times 10^5\ M_\odot$, fill the projected *area* of the complex but not its interior *volume*.

3.1.3 Atomic Constituent

The space between clumps is occupied by a lower-density gas, whose properties are not yet known in detail. A fraction of this material is certainly molecular, since it can be detected at low intensity in $^{13}C^{16}O$. The remainder is probably atomic gas at a temperature 2 to 4 times the 10 K characterizing the clumps themselves. Evidence for this component comes from 21 cm observations of molecular clouds, both giant complexes and more isolated dark clouds. Such studies often find a central absorption dip superposed on the ubiquitous emission from HI. As an illustration, Figure 3.7 shows two 21 cm profiles observed near the star-forming dark cloud ρ Ophiuchi. The dip seen directly toward the cloud (Figure 3.7a) stems from relatively cold HI gas partially absorbing the emission from warmer, background material. This dip is absent when the line of sight is slightly offset (Figure 3.7b).

The interclump gas represents a minor fraction of the complex by mass. However, both the Rosette and other systems also have extended, massive *envelopes* of atomic hydrogen. These envelopes span linear sizes several times that of the enclosed complexes and have compara-

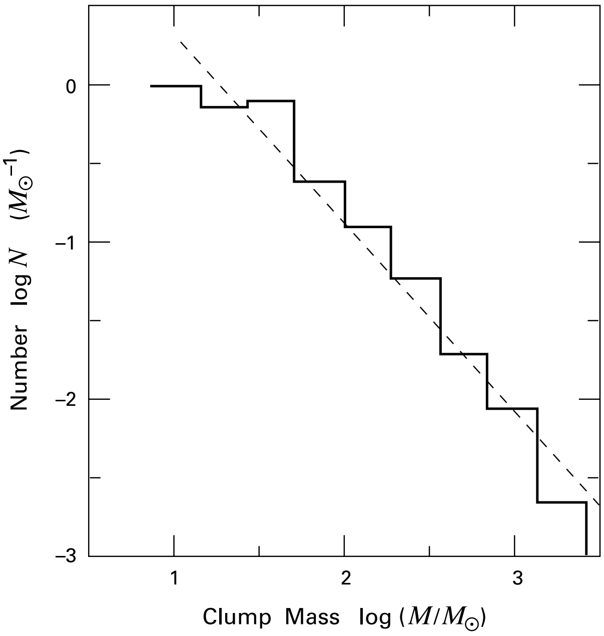


Figure 3.6 Distribution of clump masses in the Rosette Molecular Cloud.

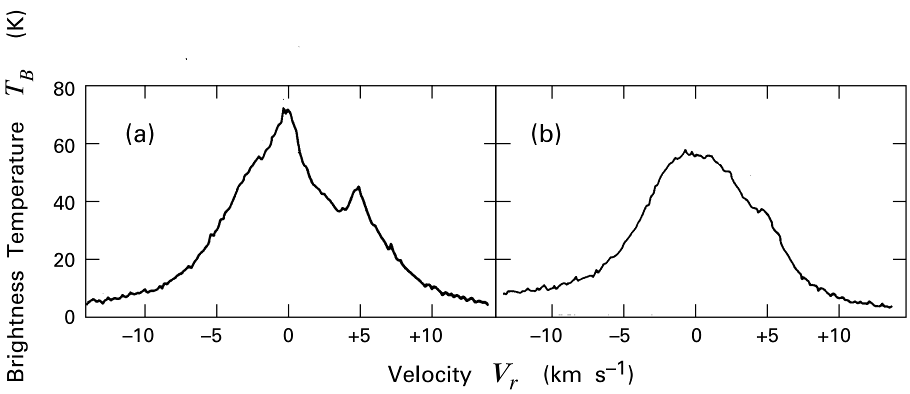


Figure 3.7 Profiles of HI emission in the dark cloud complex ρ Ophiuchi (a) directly toward the complex, and (b) adjacent to the molecular gas.

ble total masses. They appear as an excess of 21 cm emission associated in both position and velocity with giant molecular clouds. The temperature of the envelope gas lies in the range of 50 to 150 K, *i. e.*, similar to other HI clouds found throughout the Galaxy. Because of the much poorer spatial resolution available with centimeter radiation, these structures have not been mapped with nearly the detail of their molecular interiors, but there is evidence for filaments, arcs, and other inhomogeneities. The dashed contour in Figure 3.3 traces, for the Rosette system, the locus of the HI intensity at half the peak value. Similar envelopes, but on a reduced scale, are also detected around isolated dark clouds.

3.1.4 Origin and Demise

We have already seen one indication that giant molecular clouds are formed through the accumulation of many individual clumps. The clustering of the complexes along Galactic spiral arms suggests that this buildup occurs as gas flows into the potential wells associated with the arms (recall Figure 1.19). Here, the original gas is presumably atomic. Molecular clumps could form inside the condensing medium through their self-shielding from ultraviolet radiation (see Chapter 8), leaving behind the present HI envelopes. Note further that the observed drop in the H_2 surface density between the arms implies that a typical giant cloud cannot survive as long as the interarm travel time of the Galactic gas, *i. e.*, about 10^8 yr at the solar position.

What destroys the complexes? There is strong empirical evidence that the powerful winds and radiative heating associated with massive, embedded stars are the primary agents. In the Rosette cloud, the O stars within the Nebula have already dispersed much of the adjacent atomic and molecular gas, and are driving a thick HI shell into the remainder. The shell radius of 18 pc, in combination with the expansion velocity of 5 km s^{-1} obtained from the 21 cm line, implies that the expansion has proceeded for 4×10^6 yr. This time matches the age of the partially embedded NGC 2244 cluster, as determined by the main-sequence turnoff method (see Chapter 4).

For other systems, it is also possible to witness the progressive alteration in cloud properties as a function of the age of an associated stellar cluster. Older complexes generally contain

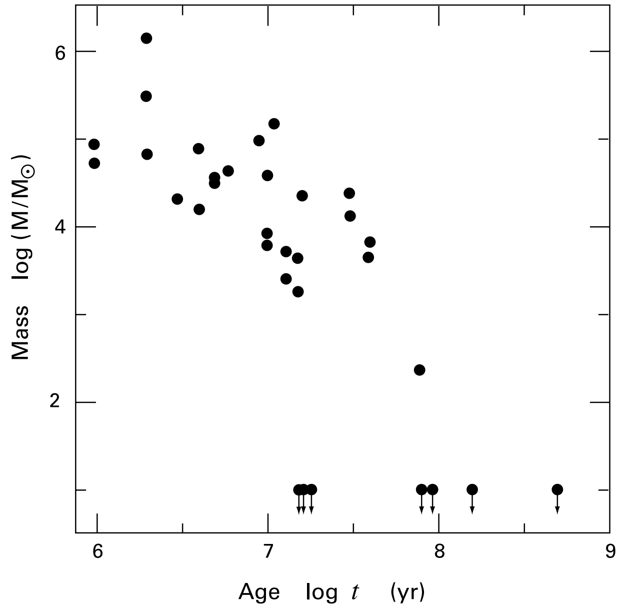


Figure 3.8 Molecular mass near open stellar clusters as a function of the cluster age. The vertical arrows signify upper bounds.

more clumps of smaller diameter and show evidence for streams of ionized gas. In addition, the cloud fragments are receding from the stars, with a typical speed of 10 km s^{-1}. Figure 3.8 displays the total molecular mass, as gauged by CO emission, lying within 25 pc of known stellar clusters. This mass is plotted as a function of the cluster age. (We will see how the second quantity is obtained in Chapter 4.) Note the marked decline in the mass by an age of 5×10^6 yr ($\log t = 6.7$), with essentially complete disappearance after 5×10^7 yr. The first O stars must appear relatively soon after the formation of a complex, since the majority of giant clouds seen today contain a massive association. Thus, the disappearance time also provides an estimate of the maximum duration of complexes.

3.2 Virial Theorem Analysis

The random nature of the clump velocities within complexes is an important clue to cloud morphology. Let us now take a more quantitative approach and ask if the magnitude of the velocity dispersion is consistent with the internal gravitational field. To do so, we pause in our survey of molecular clouds to introduce the virial theorem, which not only allows us to address this question for giant complexes and other cloud types, but also to assess generally the balance of forces within *any* structure in hydrostatic equilibrium.

3.2.1 Statement of the Theorem

Our starting point is the equation of motion for an inviscid fluid. We generalize the hydrostatic equation (2.4) to include acceleration as well as the effect of an ambient magnetic field:

$$\rho \frac{D\boldsymbol{u}}{Dt} = -\boldsymbol{\nabla} P - \rho \boldsymbol{\nabla} \Phi_g + \frac{1}{c} \boldsymbol{j} \times \boldsymbol{B} \ . \tag{3.3}$$

The quantity $D\boldsymbol{u}/Dt$ represents the full (or *convective*) time derivative of the fluid velocity $\boldsymbol{u}$, including both its rate of change at fixed spatial position $\boldsymbol{x}$ and the change induced by transporting the element to a new location with differing velocity:

$$\frac{D\boldsymbol{u}}{Dt} \equiv \left(\frac{\partial \boldsymbol{u}}{\partial t} \right)_{\boldsymbol{x}} + (\boldsymbol{u} \cdot \boldsymbol{\nabla})\, \boldsymbol{u} \ . \tag{3.4}$$

The last term in equation (3.3) is the magnetic force per unit volume acting on a current density $\boldsymbol{j}$. The two quantities $\boldsymbol{j}$ and $\boldsymbol{B}$ are related through Ampère's law:

$$\boldsymbol{\nabla} \times \boldsymbol{B} = \frac{4\pi}{c} \boldsymbol{j} \ . \tag{3.5}$$

The full Maxwell equation corresponding to (3.5) also contains on the right side the displacement current $1/c\,(\partial \boldsymbol{E}/\partial t)_{\boldsymbol{x}}$, where $\boldsymbol{E}$ is the electric field. This term may be safely ignored for the relatively slow changes of interest. Using (3.5), we recast equation (3.3) as

$$\rho \frac{D\boldsymbol{u}}{Dt} = -\boldsymbol{\nabla} P - \rho \boldsymbol{\nabla} \Phi_g + \frac{1}{4\pi} (\boldsymbol{B} \cdot \boldsymbol{\nabla}) \boldsymbol{B} - \frac{1}{8\pi} \boldsymbol{\nabla} |\boldsymbol{B}|^2 \ . \tag{3.6}$$

Here we have also employed the vector identity for triple cross products. The third righthand term in (3.6) represents the tension associated with curved magnetic field lines, while the last term is the gradient of a scalar *magnetic pressure* of magnitude $|\boldsymbol{B}|^2/8\pi$, analogous to the thermal pressure but not generally isotropic. Note that the existence of tension requires bending of the field lines, while pressure arises from the crowding of the lines, whether they be curved or straight.

Equation (3.6) governs the *local* behavior of the fluid. To derive a relation between *global* properties of the gaseous body, we form the scalar product of (3.6) with the position vector $\boldsymbol{r}$ and integrate over volume. We also avail ourselves of the equation of mass continuity:

$$-\left(\frac{\partial \rho}{\partial t} \right)_{\boldsymbol{x}} = \boldsymbol{\nabla} \cdot (\rho \boldsymbol{u}) \ , \tag{3.7}$$

as well as Poisson's equation, where the right side of (2.8) now contains the gas density ρ. We show in Appendix D how interchange of the order of differentiation and repeated integration by parts leads to the virial theorem:

$$\frac{1}{2} \frac{\partial^2 I}{\partial t^2} = 2\mathcal{T} + 2U + \mathcal{W} + \mathcal{M} \ , \tag{3.8}$$

where we have dropped the subscript on the partial derivative. Here I is a quantity resembling the moment of inertia:

$$I \equiv \int \rho\, |\boldsymbol{r}|^2\, d^3\boldsymbol{x} \ . \tag{3.9}$$

Of the righthand terms in (3.8), $\mathcal{T}$ is the total kinetic energy in bulk motion:

$$\mathcal{T} \equiv \frac{1}{2}\int \rho\,|\boldsymbol{u}|^2\, d^3\boldsymbol{x} \ , \tag{3.10}$$

while U is the energy contained in random, thermal motion:

$$\begin{aligned} U &\equiv \frac{3}{2}\int n\, k_B\, T\, d^3\boldsymbol{x} \\ &= \frac{3}{2}\int P\, d^3\boldsymbol{x} \ . \end{aligned} \tag{3.11}$$

The quantity $\mathcal{W}$ is the gravitational potential energy:

$$\mathcal{W} \equiv \frac{1}{2}\int \rho\,\Phi_g\, d^3\boldsymbol{x} \ , \tag{3.12}$$

and $\mathcal{M}$ is the energy associated with the magnetic field:

$$\mathcal{M} \equiv \frac{1}{8\pi}\int |\boldsymbol{B}|^2\, d^3\boldsymbol{x} \ . \tag{3.13}$$

In writing equation (3.8), we have ignored a number of surface integrals, including one representing the effect of any external pressure. (See equation (D.12) in Appendix D.) Such an approximation would *not* be justified for HI or diffuse clouds, but is valid for the strongly self-gravitating giant complexes. The issue here is which of the terms in (3.8) can balance the gravitational binding energy $\mathcal{W}$. Note that the integrals U, $\mathcal{M}$, and $\mathcal{T}$ are all positive, while $\mathcal{W}$ is negative. If *none* of these three terms can match $\mathcal{W}$ in magnitude, then a typical giant molecular cloud would be in a state of gravitational collapse.

3.2.2 Free-Fall Time

Let us briefly explore this latter possibility. We denote by $R \equiv L/2$ the characteristic "radius" for a cloud of mass M. Then, to within a factor of order unity, $\mathcal{W}$ is $-GM^2/R$. Equation (3.8) reduces in this case to

$$\frac{1}{2}\,\frac{\partial^2 I}{\partial\, t^2} \approx -\frac{G\,M^2}{R} \ .$$

If we further approximate I as MR^2, then dimensional analysis of this relation tells us that R in the collapsing cloud shrinks by roughly a factor of two over a characteristic *free-fall time* $t_{\rm ff}$. This is approximately

$$\begin{aligned} t_{\rm ff} &\approx \left(\frac{R^3}{G\,M}\right)^{1/2} \\ &= 7\times 10^6 \ {\rm yr}\left(\frac{M}{10^5\ M_\odot}\right)^{-1/2}\left(\frac{R}{25\ {\rm pc}}\right)^{3/2} \ , \end{aligned} \tag{3.14}$$

where we have inserted representative numbers from Table 3.1. Note further that, since $M/R^3 \approx \rho$, the time scale can also be written as $(G\,\rho)^{-1/2}$. It is conventional to define $t_{\rm ff}$ precisely as

$$t_{\rm ff} \equiv \left(\frac{3\,\pi}{32\,G\,\rho}\right)^{-1/2} \ . \tag{3.15}$$

This expression actually gives the time for a homogeneous sphere with zero internal pressure to collapse to a point (Chapter 12).

Equation (3.14) indicates that giant molecular clouds have free-fall times comparable to their observed lifetimes. Does this mean that the clouds are indeed collapsing? One issue here is our questionable use of a global, volume-averaged density in equation (3.14). The higher density appropriate for an individual clump would have yielded a $t_{\rm ff}$-value closer to 10^6 yr. In any case, there is no convincing empirical evidence for large-scale shrinking or flattening over such an interval. The complexes' internal velocities also appear to be random, not systematically directed toward a collapsing center. Apparently, the entities survive until they are destroyed from within, by the massive stars they spawn.

3.2.3 Support of Giant Complexes

If the complexes are in approximate force balance over their lifetimes, we may actually ignore the left side in equation (3.8) and obtain the form of the virial theorem appropriate for longterm stability:

$$2\mathcal{T} + 2U + \mathcal{W} + \mathcal{M} = 0 \, . \tag{3.16}$$

For a cloud in such "virial equilibrium," the question is again how to balance $\mathcal{W}$. In order to gauge the effectiveness of the internal pressure in supporting the cloud, we first note that U is given, in the spirit of our approximations, by $M\mathcal{R}T/\mu$, where T is a representative gas temperature. We then form the ratio

$$\begin{aligned} \frac{U}{|\mathcal{W}|} &\approx \frac{M\mathcal{R}\,T}{\mu}\left(\frac{G\,M^2}{R}\right)^{-1} \\ &= 3\times 10^{-3}\left(\frac{M}{10^5\,M_\odot}\right)^{-1}\left(\frac{R}{25\ \rm pc}\right)\left(\frac{T}{15\ \rm K}\right) \, . \end{aligned} \tag{3.17}$$

Even after considering a reasonable spread in the values of M, R, and T, this result shows unambiguously that giant complexes are *not* sustained by thermal pressure.

We turn to the final, magnetic term in equation (3.16). Studies of the polarization of background starlight reveal the existence of a large-scale B-field throughout the plane of the Galaxy. This field penetrates giant molecular clouds and could exert a major dynamical influence, helping to prevent collapse. The strength of the magnetic force relative to self-gravity is measured by

$$\begin{aligned} \frac{\mathcal{M}}{|\mathcal{W}|} &\equiv \frac{|\boldsymbol{B}|^2\,R^3}{6\,\pi}\left(\frac{G\,M^2}{R}\right)^{-1} \\ &= 0.3\left(\frac{B}{20\ \mu\rm G}\right)^2\left(\frac{R}{25\ \rm pc}\right)^4\left(\frac{M}{10^5\ M_\odot}\right)^{-2} \, , \end{aligned} \tag{3.18}$$

where we have crudely represented the cloud as a sphere of radius R in estimating $\mathcal{M}$. The magnetic field strength in molecular clouds is currently obtained from the Zeeman splitting of either the 21 cm line of HI or of a cluster of lines near 18 cm from OH (see Chapter 6). Unfortunately, this technique has not yet provided results for either the clump or interclump gas in giant clouds. Our representative B-value in equation (3.18) comes from measurements in nearby dark clouds and is consistent with the somewhat lower values detected in the warm HI envelopes of giant complexes.

The numerical estimate of $\mathcal{M}/|\mathcal{W}|$ indicates that the magnetic field *is* important, but in precisely what sense? According to equation (3.3), the associated force acts on a fluid element in a direction orthogonal to $\boldsymbol{B}$. Thus any self-gravitating cloud supported mainly by a well-ordered field can slide freely along field lines until it settles into a nearly planar configuration. Such flattening is *not* evident in the giant complexes, so we are forced to reject the hypothesis of a perfectly smooth internal field. Now the direction of $\boldsymbol{B}$ in the plane of the sky, but not its magnitude, can be ascertained from the optical polarization mentioned earlier. We saw in Chapter 2 how elongated grains lying perpendicular to the field polarize starlight through dichroic extinction. However, the observed $\boldsymbol{E}$ vectors in any local region are never perfectly aligned, but display significant scatter. This scatter indicates the presence of a *random* component to $\boldsymbol{B}$, coexisting with the smooth background. The field distortion arises, at least in part, from *magnetohydrodynamic (MHD) waves*, also called *hydromagnetic* waves. As we will discuss in Chapter 9, these waves may provide the isotropic support preventing the cloud from flattening.

The final term to consider in equation (3.16) is the kinetic energy $\mathcal{T}$. The bulk velocity within giant clouds stems mostly from the random motion of their clumps. Denoting by ΔV the mean value of this speed, we find

$$\begin{aligned}\frac{\mathcal{T}}{|\mathcal{W}|} &\approx \frac{1}{2} M\,\Delta V^2 \left(\frac{GM^2}{R}\right)^{-1} \\ &= 0.5 \left(\frac{\Delta V}{4\ \mathrm{km\ s^{-1}}}\right)^2 \left(\frac{M}{10^5\ M_\odot}\right)^{-1} \left(\frac{R}{25\ \mathrm{pc}}\right) .\end{aligned} \tag{3.19}$$

To obtain a representative ΔV, which we take to be the three-dimensional velocity dispersion, we have increased the Rosette line-of-sight dispersion of 2.3 km s^{-1} by a factor of $\sqrt{3}$, as would be appropriate for a random, three-dimensional velocity field. Our numerical result for $\mathcal{T}/|\mathcal{W}|$ implies that the typical internal ΔV is close to the *virial velocity* V_{vir}, which we define as

$$V_{\mathrm{vir}} \equiv \left(\frac{G\,M}{R}\right)^{1/2} . \tag{3.20}$$

Comparison with equation (3.14) shows that V_{vir} is the velocity of a parcel of gas that traverses the cloud over the free-fall time t_{ff}. In other words, it is the typical speed attained by matter under the influence of the cloud's internal gravitational field.

The actual velocity dispersion in any cloud can readily be determined by the broadening of some spectral line, generally one of CO. Figure 3.9 shows that, despite considerable scatter, ΔV roughly matches V_{vir} (or, equivalently, that $\mathcal{T}$ matches $|\mathcal{W}|$) not only in our typical giant molecular cloud, but over a much wider range of sizes. In giant complexes, this approximate equality is consistent with the picture of a swarm of relatively small clumps, each one moving in the gravitational field created by the whole ensemble. The kinetic energy in clump motion is matched by that of the internal magnetic field, which also has a significant random component. Since energy can be exchanged between the matter and field through hydromagnetic waves, such equipartition is not too surprising. Nevertheless, there are as yet no quantitative, theoretical models of giant clouds, incorporating both the uniform and fluctuating field, that account naturally for this result.

Viewed as separate entities, the clumps within giant molecular clouds fall under the "individual dark cloud" category in Table 3.1. The largest clumps have masses of order $10^3\ M_\odot$, but

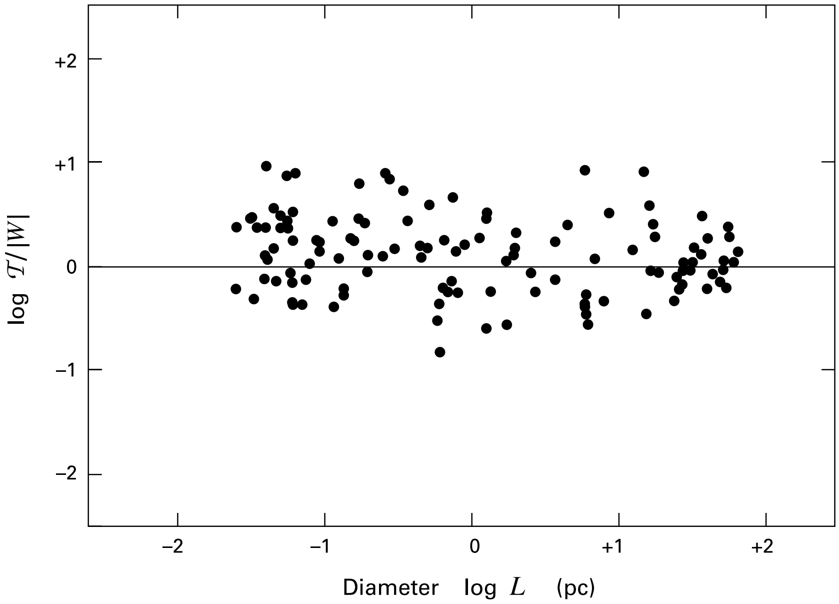


Figure 3.9 The ratio of the bulk kinetic energy $\mathcal{T}$ to the gravitational potential energy $|\mathcal{W}|$ for molecular clouds, plotted as a function of the cloud diameter L.

even more massive dark clouds exist. The Taurus-Auriga system described in Chapter 1 and the well-studied ρ Ophiuchi region are two prime examples. Yet another at roughly the same distance is Corona Australis in the Southern hemisphere. Though differing morphologically from each other, all have mean densities similar to the clumps, but total masses closer to $10^4\,M_\odot$. We shall term such objects *dark cloud complexes*. While accounting for a significant fraction of Galactic star formation, it is noteworthy that they do *not* produce the OB associations that are a hallmark of the more massive systems. On the other hand, complexes such as ρ Ophiuchi do contain regions with peak A_V-values of order 100. Such locales always harbor a multitude of embedded young stars.

3.3 Dense Cores and Bok Globules

We now focus on the smallest cloud entities. These dense cores and Bok globules are seen, in many cases, to harbor infrared point sources of emission. We know, therefore, that they are sites of star formation. How this production occurs will occupy much of our attention. We begin, however, with an overview of the cloud properties.

3.3.1 Quiescent Gas

The hierarchy of molecular clouds displays several intriguing patterns. One we have seen is the similarity of $\mathcal{T}$ and $|\mathcal{W}|$. The much sparser observations of magnetic field strengths are also consistent with a rough equality of $\mathcal{M}$ and $|\mathcal{W}|$, holding over an equally broad range of cloud parameters. Finally, there is a third trend not evidently tied to energy considerations. Suppose

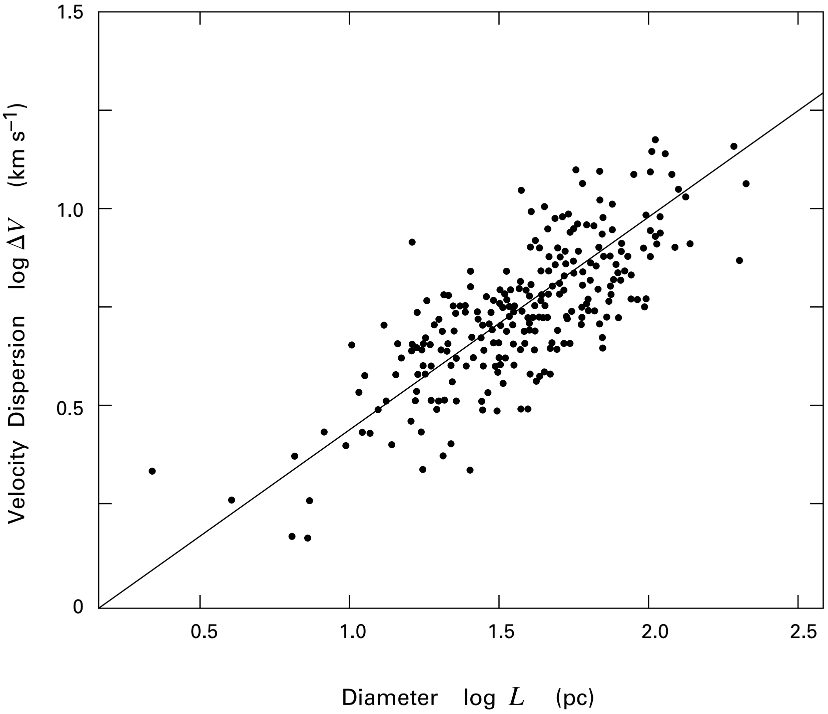


Figure 3.10 Three-dimensional velocity dispersion of molecular clouds as a function of their diameter.

one observes the broadening of a spectral line toward a cloud without spatially resolving that object. This broadening, as we have seen, can be translated directly into a velocity dispersion. Figure 3.10 demonstrates that the measured dispersion varies systematically with cloud size:

$$\Delta V = \Delta V_\circ \left(\frac{L}{L_\circ} \right)^n , \tag{3.21}$$

where $n \approx 0.5$, and where $\Delta V_\circ \approx 1\ \mathrm{km\,s^{-1}}$ for $L_\circ = 1$ pc. This empirical relation, whose basis is still not fully understood, is often called *Larson's law*. Taken literally, equation (3.21), in conjunction with (3.20), implies that the *column density* ML^{-2} is unchanged from cloud to cloud. In fact, there is a slow increase of the proportional quantity A_V toward denser and smaller clouds, as indicated in Table 3.1. Nevertheless, the near constancy of A_V is also of interest and could well be a clue toward deciphering the puzzle of cloud origin and structure.

The broadening of a spectral line from any molecular cloud stems from both thermal motion and bulk velocity of the emitting gas. The quantity ΔV in equation (3.21) refers only to the nonthermal component, which is presumably induced by hydromagnetic waves.[2] As we consider successively smaller clouds, ΔV eventually reaches the ambient thermal speed, whose root-mean-square value is $(3\mathcal{R}T/\mu)^{1/2}$. Equation (3.21) implies that this transition is reached

---

[2] In contrast, the kinetic energy $\mathcal{T}$ in Figure 3.9 was calculated using both the thermal and nonthermal velocity dispersions.

at a size scale $L_{\rm therm}$ found from

$$\begin{aligned} L_{\rm therm} &= \frac{3\,\mathcal{R}\,T\,L_\circ}{\mu\,\Delta V_\circ^2} \\ &= 0.1\ {\rm pc}\left(\frac{T}{10\ {\rm K}}\right)\ . \end{aligned}$$

An object of size $L_{\rm therm}$, which also has $U \approx |\mathcal{W}|$, is a more quiescent environment than larger dark clouds with their energetic internal waves. In fact, both the clumps within giant complexes and isolated dark clouds do contain distinguishable substructures of this dimension. These entities are the *dense cores* responsible for individual star formation.

Our surveys of Orion and Taurus-Auriga showed how these cores are found throughout the interiors of dark clouds (recall Figure 1.2). With densities exceeding $10^4\ {\rm cm}^{-3}$, such structures cannot be probed by the usual $^{12}C^{16}O$ or even $^{13}C^{16}O$ lines, which are both optically thick. Other species, such as NH_3, $^{12}C^{18}O$, and CS, emit radio lines that are optically thin and can be used both for discerning global properties and for spatial mapping. These observations, carried out since the early 1980s, show that the typical dense core comprises several solar masses of gas, and has a temperature near 10 K. Although a few individual cores within a dark cloud may be far more massive, the aggregate of all cores generally comprises less than 10 percent of the total gas supply.

The most extensive studies have been conducted using the 1.3 cm line of NH_3. Figure 3.11 is a typical NH_3 spectrum toward the center of the dense core L260, at a distance of 160 pc. The dashed curve is the profile expected from thermal motion alone, as calculated for the internal temperature of 9 K. (We will later see how such a figure is obtained.) Here, it is assumed that the probability of any line-of-sight velocity V_r is given by the Maxwell-Boltzmann distribution, so that the observed intensity varies as

$$T_A(V_r) \propto \exp\left(-\frac{m_{\rm NH_3}\,V_r^2}{2\,k_B\,T}\right) \tag{3.22}$$

where $m_{\rm NH_3}$ is the mass of the ammonia molecule. (See Appendix E for a general discussion of line broadening.) Note that the root-mean-square value of V_r is $(k_B T/m_{\rm NH_3})^{1/2}$; this is the one-dimensional dispersion for thermal motion.

The profile in equation (3.22) has a full-width half-maximum extent of

$$\Delta V_{\rm FWHM}\ (\textit{therm}) = \left(\frac{8\ \ln 2\,k_B T}{m_{\rm NH_3}}\right)^{1/2}\ , \tag{3.23}$$

which has the value 0.15 km s^{-1} for L260. Figure 3.11 shows that the observed profile is actually broader than this. The additional broadening can be ascribed to a random field of turbulent velocities; the solid line is the calculated model result. The turbulent velocities in the model obey a Gaussian probability distribution like that in equation (3.22), but with an associated width denoted $\Delta V_{\rm FWHM}\,(\textit{turb})$. Since the two distributions are uncorrelated, the *total* profile width is

$$\Delta V_{\rm FWHM}^2\,(\textit{tot}) = \Delta V_{\rm FWHM}^2\,(\textit{therm}) + \Delta V_{\rm FWHM}^2\,(\textit{turb})\ , \tag{3.24}$$

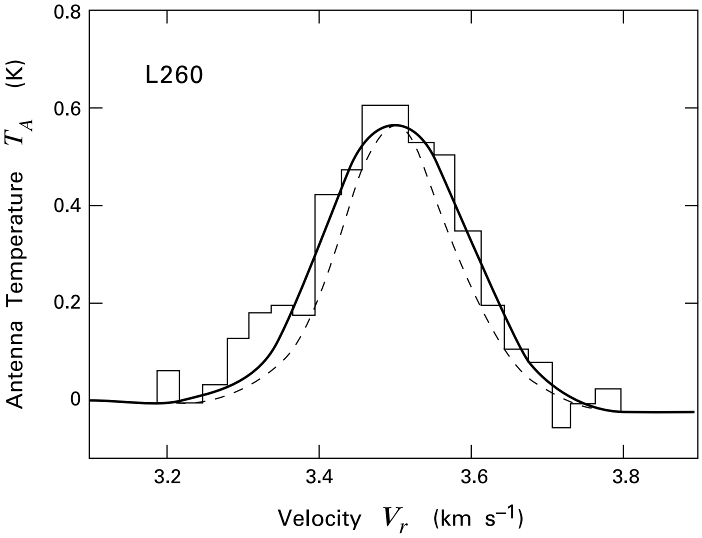


Figure 3.11 Profile in the 1.3 mm line of NH_3 of the dense core L260. The histogram represents the observed profile, while the dashed curve is a theoretical result assuming purely thermal motion. The heavy solid curve is the theoretical profile including a Gaussian distribution of turbulent speeds.

as we show in Appendix E. In our example, we find that $\Delta V_{\rm FWHM}\,(turb)$ must be 0.11 km s^{-1} in order to reproduce the observed total width of 0.19 km s^{-1}.

It is instructive to compare the spatial maps of a given core in a number of different lines. Figure 3.12 is a composite map of L1489, a dense core containing an infrared point source, as indicated. Shown are the half-maximum intensity contours in NH_3 (the 1.3 cm line), CS (at 3.0 mm), and $^{12}C^{18}O$ (at 2.7 mm). The observed lines are excited into emission by collisions with ambient hydrogen molecules. The fact that different molecular transitions require different threshold hydrogen densities for excitation accounts in part for the systematic widening of the maps. Thus, the 1.3 cm line from NH_3 has a higher threshold than the line from $^{12}C^{18}O$ and therefore samples a compact, interior region. Smaller regions are less turbulent, according to equation (3.21), so that we can also understand the relatively low value of the nonthermal NH_3 line width.

However, there are complications to this picture. The CS line has an even larger critical hydrogen density than NH_3. Yet Figure 3.12 shows its map extending farther out. A plausible explanation is that the CS molecules stick to dust grains at the highest ambient densities. Then the observed central intensity would be lowered, and the radial falloff in flux made more shallow. The half-maximum contour thus moves outward, as seen. Another puzzle is the offset of *all* the maps in Figure 3.12 from the embedded star. The latter would naturally be expected to form at the density peak. This offset is seen in other cases but not well understood.

The presence of pointlike infrared sources in cores such as L1489 is, of course, the most direct evidence that these structures indeed form stars. The IRAS satellite found such embedded stars in about half the dense cores in Taurus-Auriga and ρ Ophiuchi that had been previously identified through molecular lines. While some of these sources have optical counterparts, an

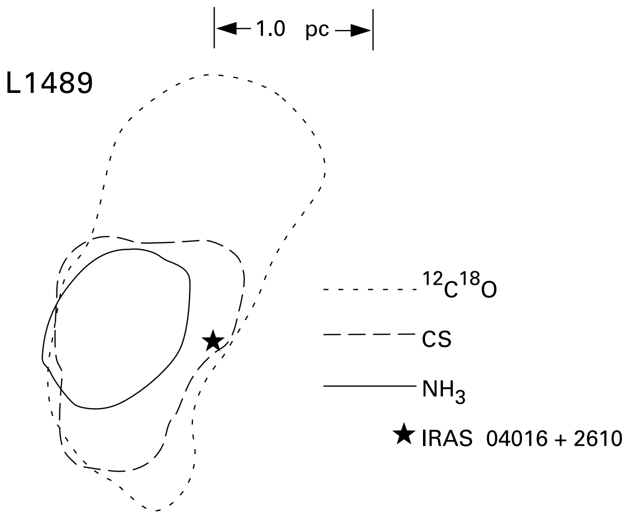


Figure 3.12 Composite map of the dense core L1489 in the lines of NH_3, CS, and $^{12}C^{18}O$. Each contour corresponds to the intensity at half its peak value. Note the offset position of the embedded infrared source.

equal number do not. Finally, most of the optically invisible stars within cores are associated with outflows, as detected in CO. These important findings tell us, first, that dense cores exist for a substantial time prior to forming stars. Second, since optically invisible stars are presumably the youngest, outflow generation occurs extremely early in stellar evolution. Third, a dense core does not suddenly vanish after forming a star, but gradually dissipates as the object inside ages.

From an observational perspective, there are no outstanding differences between the dense cores lacking infrared sources and those containing young stars. Both types, for example, have visual extinctions that range from about 5 to 15 mag.[3] On the other hand, the sample having deeply embedded, optically invisible stars includes some cores with significantly higher NH_3 line widths, as illustrated in Figure 3.13. Since the difference in the measured gas temperature between cores with and without stars is negligible, the larger widths must be due to turbulent motion. It is tempting to associate the higher level of turbulence with the molecular outflows created by the youngest stars. As we will see in Chapter 13, turbulence is indeed an integral feature of outflows, but a more convincing link to the observed line widths needs to be made.

3.3.2 Intrinsic Shapes

Let us now turn to the more detailed properties of dense cores, beginning with their shapes. Here the type of composite map shown in Figure 3.12 represents the best data available. For the commonly used NH_3 line at 1.3 cm, the beam diameter with a 40-meter telescope is $80''$, or 0.05 pc at the distance of Taurus-Auriga. The corresponding figure is 0.01 pc for the 3.0 mm line of CS. Thus, the current resolution of single-dish (as opposed to interferometric) observations is modest for NH_3, but adequate at shorter wavelengths for discerning gross spatial features.

---

[3] Extinction measurements in dense cores presently rely on an empirical correlation between A_V and the column density in an optically thin tracer such as $^{12}C^{18}O$. We discuss such relationships in Chapter 6.

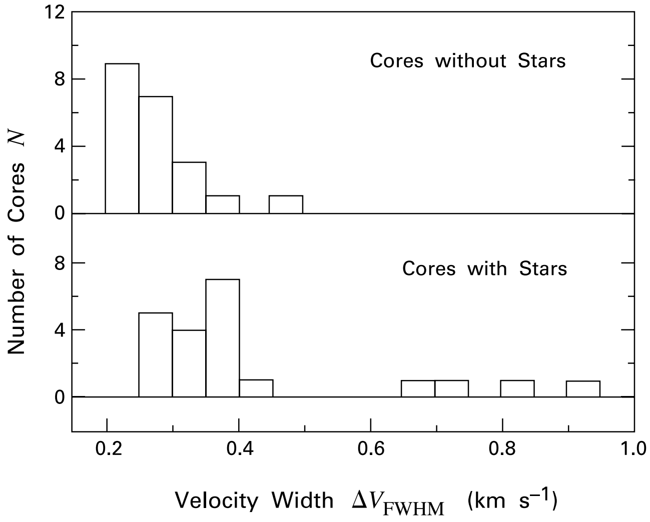


Figure 3.13 Distribution of dense cores as a function of their NH3 velocity width ΔV_{FWHM}. The top histogram shows those cores whose nearest star lies outside their boundaries, again as seen in NH3. The bottom histogram is for cores with embedded stars.

In particular, examination of several dozen dense cores reveals mean axial ratios of about 0.6, similar to the L1489 case shown here. This ratio, as well as the orientation of the long axis, does not vary greatly from line to line for a given core. The question now is which three-dimensional structures can yield such shapes when projected onto the sky.

As a first approximation, we assume that all cores have a single intrinsic shape, which we take to be spheroidal. Figure 3.14 shows how both oblate (flattened) and prolate (elongated) spheroids can appear to have identical axial ratios when projected on the sky. However, if we further assume that the spheroids are randomly oriented, then the oblate configurations are less likely.

The argument here is a bit technical but worth some effort. Refer to Figure 3.14, and define true and apparent (*i. e.*, projected) axial ratios for the oblate spheroid as $R_{\mathrm{true}} \equiv b/a$ and $R_{\mathrm{app}} \equiv b'/a$, respectively. Note that $R_{\mathrm{true}} < R_{\mathrm{app}} < 1$ for any angle i between the line of sight and the spheroid's axis of rotation. Figure 3.15 shows more explicitly the relation between b' and the true axes. It also defines a length $y_\circ$, which is related to b' by $b' = y_\circ \sin i$. As an exercise in geometry, the reader may verify that

$$y_\circ^2 = a^2 \cot^2 i + b^2 , \tag{3.25}$$

from which we deduce

$$(b')^2 = a^2 \cos^2 i + b^2 \sin^2 i . \tag{3.26}$$

We thus find that

$$\sin^2 i = \frac{1 - R_{\mathrm{app}}^2}{1 - R_{\mathrm{true}}^2} . \qquad \textit{oblate} \tag{3.27}$$

Consider now the *average* observed ratio for a number of randomly oriented spheroids, all with identical R_{true}. Denoting by $\langle\ \rangle$ the average over solid angle and recalling that

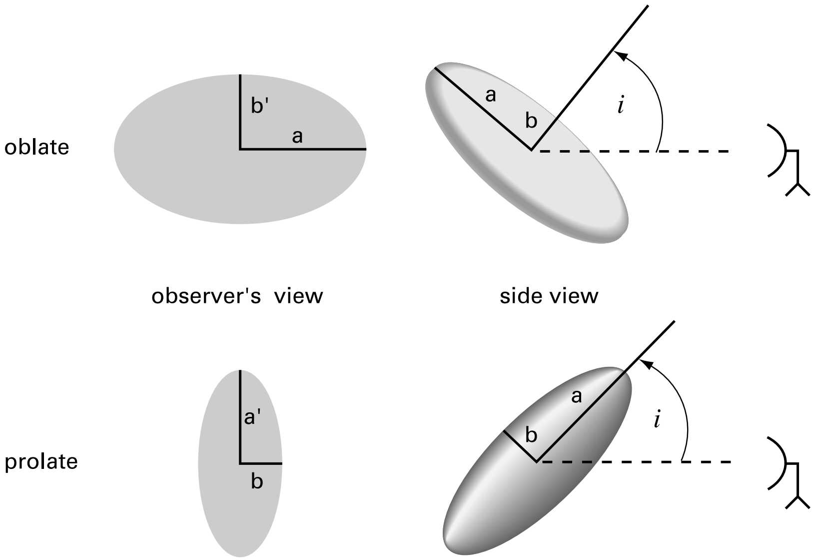


Figure 3.14 Projected views of oblate and prolate spheroids. Both objects generally appear as ellipses in the plane of the sky.

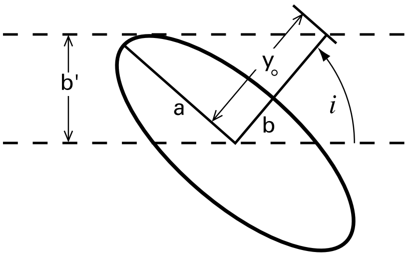


Figure 3.15 Geometric relations within the cross section of an oblate spheroid.

$\langle \sin^2 i \rangle = 2/3$, we see that *no* $R_{\rm true}$, however small, will reproduce a given $\langle R^2_{\rm app} \rangle$ unless $\langle R^2_{\rm app} \rangle > 1/3$. The observations to date marginally satisfy this latter criterion and demand that $R_{\rm true}$ be about 0.2, *i. e.*, that the putative oblate spheroids be highly flattened.

Turning to the prolate case, the analogous relation to equation (3.26) is

$$(a')^2 \, = \, a^2 \, \sin^2 i \, + \, b^2 \, \cos^2 i \ , \tag{3.28}$$

where the axis labels are defined in Figure 3.14. Since $R_{\rm true}$ is still b/a but $R_{\rm app}$ is now b/a', we find

$$\cos^2 i \, = \, \frac{R^{-2}_{\rm true} - R^{-2}_{\rm app}}{R^{-2}_{\rm true} - 1} \ . \qquad \textit{prolate} \tag{3.29}$$

Recognizing that $\langle \cos^2 i \rangle = 1/3$, the averaged form of equation (3.29) may be written

$$R^{-2}_{\rm true} \, = \, \langle R^{-2}_{\rm app} \rangle \, + \, \frac{1}{2} \left(\langle R^{-2}_{\rm app} \rangle - 1 \right) \ . \tag{3.30}$$

The difference with the oblate case is now apparent. The last equation yields a value of $R_{\rm true}$ for *any* $\langle R_{\rm app}^{-2} \rangle$. Moreover, the first quantity only slightly exceeds the second for modest ratios. Reproducing the observations requires only that $R_{\rm true}$ lie between 0.4 and 0.5. That is, the structure can be rounder than in the oblate case.

We may recast the essential argument in non-mathematical terms. Even a razor-thin disk usually presents a rather high aspect ratio, when projected randomly onto the sky. A perfectly thin needle, on the other hand, looks just the same in projection, except when it is exactly pole-on. Thus, a prolate object needs more *intrinsic* thickness for its projected image to have a sizable *apparent* thickness.

Which case is more reasonable: a highly flattened disk or a thick cigar? While not demanding the rounder, prolate shapes, the observations certainly favor them. The measured line widths in dense cores indicate that they are largely, though not totally, supported by the thermal pressure gradient, which acts isotropically. In addition, the dark cloud regions containing the cores often have a striated appearance (recall Figure 1.9). In a number of cases, there is clear alignment of the cores' long axes with these "fingers" of visual obscuration. Our geometric argument indicates that the larger structures, with their extreme axial ratios, are unlikely to be flattened. The same should then be true for their embedded cores. We will see in Chapter 12 that prolate clouds are also attractive theoretically as progenitors for binary stars.

Molecular lines are not the only tools for probing the cores' structure. The dust grains coexisting with the gas also emit continuum radiation as a result of their finite temperatures. Detection of this emission at millimeter wavelengths has yielded maps of high spatial resolution. Figure 3.16 shows the starless dense core L63, situated in a relatively isolated dark cloud about 160 pc away. The left panel is a map in the 1.3 cm line of $\mathrm{NH_3}$, while the right is the same object viewed in 1.3 mm continuum radiation. Although the second image covers a smaller spatial extent, it displays a similar elongation as the $\mathrm{NH_3}$ map. The continuum version also reveals more small-scale structure. Here, the angular resolution of the telescope is $12''$, corresponding to 0.01 pc at the estimated distance.

Both the continuum radiation and the molecular line are optically thin in this case, and thus yield conditions in the core's deep interior. As we have indicated, line emission can only be produced when the ambient density exceeds the critical value $n_{\rm crit}$ associated with that transition. The emission also tends to fall off significantly for $n > n_{\rm crit}$, so that any given line samples a relatively narrow density range. Dust emission does not suffer from this limitation and requires only that the density and temperature along the line of sight be high enough for detection. The continuum maps are thus potentially useful for reconstructing the internal density profile, although detailed results still require accurate knowledge of the internal temperature distribution. The studies have indicated that starless cores like L63 have densities that rise steeply from the outside, but then reach a shallower plateau somewhat above 10^5 $\mathrm{cm^{-3}}$.

3.3.3 Magnetic Fields

Since magnetic forces are expected to be significant in dense cores, it would be extremely valuable if field strengths could be ascertained directly by observation in these regions. Unfortunately, this is not yet the case. To date, measurements by Zeeman splitting are limited to a handful of dark clouds of larger size and lower density. Optical and near-infrared polarization maps using background field stars, which yield the direction but not the magnitude of $\boldsymbol{B}$, are also confined to these sparser regions.

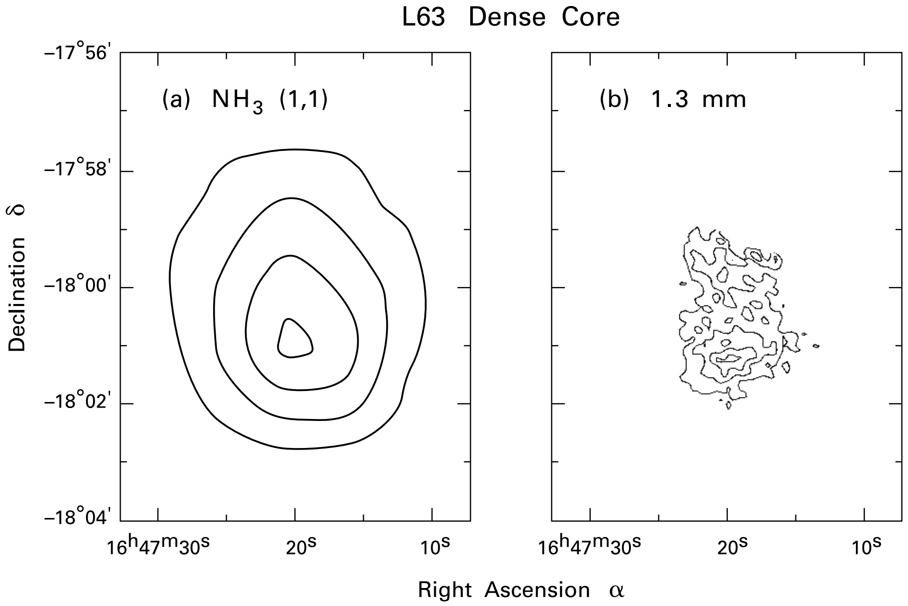


Figure 3.16 (a) The dense core L63, mapped in the 1.3 cm line of NH_3. (b) The core as mapped in 1.3 mm continuum radiation.

Figure 3.17 is an optical polarization map of the ρ Ophiuchi dark cloud complex. Here, the more filamentary density contours display the elongated substructure mentioned previously. The short line segments indicate the direction of the electric vector of the stellar radiation. Assuming the polarization is due to magnetically aligned grains, this direction is also that of the ambient magnetic field. Note that most of the mass of this complex, as well as the intense star formation activity, is contained in the L1688 cloud toward the lower right. In this region, there is no strong correspondence between the field direction and the cloud morphology. However, there *is* striking alignment in the lower-mass fragments L1709 and L1755 stretching to the northeast. Further south, this same field direction is preserved, creating a systematic offset from the orientation of the clouds L1729 and L1712.

A particularly well-studied object, with magnetic field measurements spanning a range of densities, is B1, a fragment some 3 pc in diameter within the Perseus dark cloud complex. Measurements in OH emission lines yield field strengths ranging from 10 μG in the more rarified outer portion of the cloud to 54 μG in a compact central region of diameter 0.2 pc. Since this latter region, which includes an embedded star, contains some 10 $M_\odot$ of cloud gas, the magnetic virial term $\mathcal{M}$ is about $1/3$ the gravitational potential energy $\mathcal{W}$. Such a fraction is consistent with the analysis of NH_3 line profiles in typical dense cores, provided that the amplitude of the fluctuating field is comparable to that of the more uniform component.

One promising development in this field is the observation of polarized, submillimeter emission from heated dust grains. This technique allows us, at least in principle, to trace directly the field geometry within dense cores themselves. Curiously, the maps obtained thus far show a steep falloff in the degree of polarization toward the core center. It is unclear whether this trend is due to the magnetic field topology, or to altered properties of the grains themselves. In

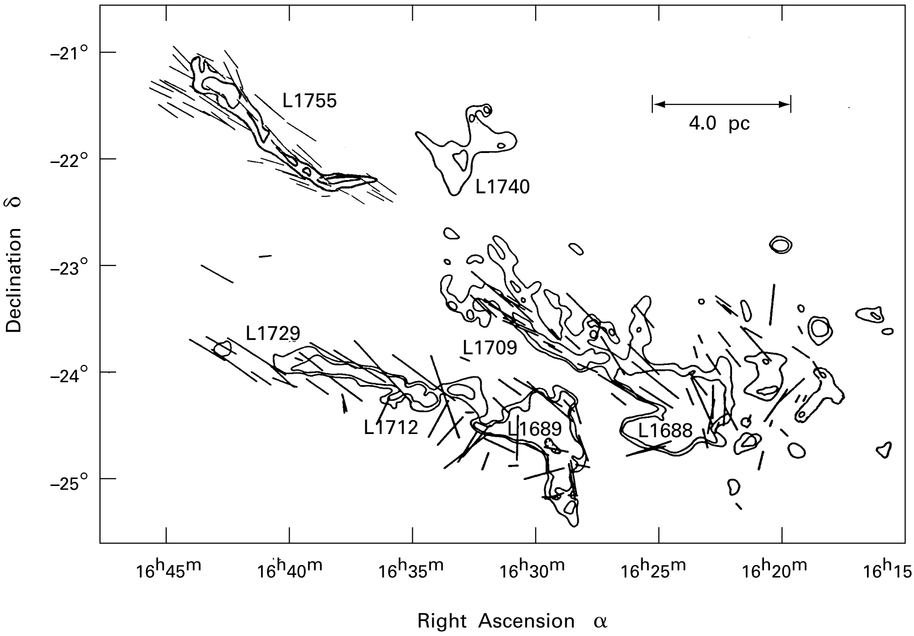


Figure 3.17 Dark clouds in the ρ Ophiuchi complex. The contours represent $^{13}C^{16}O$ emission. The short line segments indicate the direction and degree of polarization of the electric field vector.

the case of B1, the polarization dips, but is nonzero, at each of several internal density clumps, including the one with the embedded star. The magnetic field vectors of the various clumps are not aligned.

3.3.4 Rotation

As cores collapse to form stars, field lines are pulled in, following the gas motion. The resulting buildup in magnetic pressure acts as an impediment to further collapse and alters the direction of the dynamical evolution. In a similar manner, any initial rotational motion in the core must increase through angular momentum conservation, ultimately creating a centrifugal barrier to collapse. It is therefore also important to search observationally for core rotation. Here the idea is to look for variation in the radial velocity V_r across the cloud face. As usual, we gauge V_r by the Doppler-induced shift in some spectral line. The majority of dense cores analyzed thus far indeed display the expected variation.

As one illustration, Figure 3.18 is an NH_3 map of a dense core within L1251A, an elongated dark cloud at a distance of 200 pc. The filled squares superposed on the contour map of the 1.3 cm line have sizes proportional to the measured V_r at each position. A velocity gradient from left to right is clearly present. Its magnitude of 1.3 km s^{-1} pc^{-1}, is typical of those observed, which range from the detection limit near 0.3 km s^{-1} pc^{-1} to a factor of ten higher. Tracer molecules other than NH_3 yield similar figures. If every dense core were rotating as a solid body with its rotation axis perpendicular to the line of sight, the measured gradient would

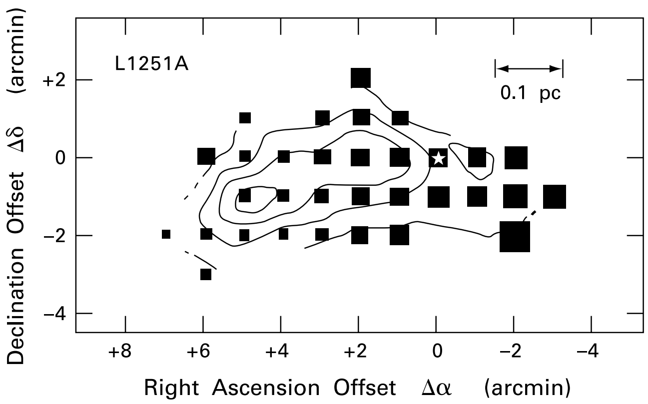


Figure 3.18 Rotation of the dense core L1251A. The size of each filled square is proportional to the radial velocity, as gauged from the 1.3 cm line of NH_3. The star symbol marks the position of the embedded source IRAS 22290+7458.

correspond to Ω, the cloud's angular velocity. For tipped axes that form an angle i with the line of sight, the gradient would be $\Omega/\sin\, i$.

These results must be viewed with some caution. For example, the projected motion of a molecular outflow can also create gradients in the radial velocity. However, this effect cannot be dominant, as similar velocity gradients are seen in cores with and without young stars. Assuming, then, that the observations are indicative of true rotation, we may gauge its dynamical significance. A dense core of mass M and diameter L has a rotational kinetic energy given by $\mathcal{T}_{\rm rot} = (1/20)\, M L^2 \Omega^2$, if the object is idealized as a uniform-density sphere. The result for a prolate configuration of average diameter L and modest axial ratio is the same within a factor of two. Since the potential energy in the spherical case is $\mathcal{W} = (6/5)\, G M^2/L$, we can estimate the ratio of $\mathcal{T}_{\rm rot}$ to $|\mathcal{W}|$:

$$\begin{aligned} \frac{\mathcal{T}_{\rm rot}}{|\mathcal{W}|} &\approx \frac{\Omega^2\, L^3}{24\, G\, M} \\ &= 1 \times 10^{-3} \left(\frac{\Omega}{1\ {\rm km\ s^{-1}\ pc^{-1}}}\right)^2 \left(\frac{L}{0.1\ {\rm pc}}\right)^3 \left(\frac{M}{10\ M_\odot}\right)^{-1} . \end{aligned} \tag{3.31}$$

Equation (3.31) implies that a dense core rotates so slowly (with a representative period of $2\pi/\Omega = 6 \times 10^6$ yr) that the associated centrifugal force is negligible compared to the self-gravity and pressure gradients that truly determine its equilibrium structure. It is hardly surprising, then, that the observed orientations of the rotation axes bear no apparent relation to the cores' spatial elongations. The true significance of the rotation lies not in cloud structure, but in the effect on collapse itself.

3.3.5 Globule Structure

We finally discuss compact, dense regions *not* embedded within larger complexes. These are the *Bok globules*, named in honor of the astronomer who first recognized, in the 1940s, their potential role in stellar birth. Bok's insight has since been amply confirmed through observations of embedded infrared sources and energetic molecular outflows associated with a large fraction of these objects. Apart from their relative isolation, globules resemble the more common dense cores in most respects. About 200 of them lie within 500 pc of the Sun, where they can be picked out as small patches of visual extinction. The optical photograph in Figure 3.19 shows the extraordinarily sharp boundaries of the globule B68 in Ophiuchus. Radio observations in CO have determined that such structures are actually surrounded by envelopes of more diffuse gas, extending over a few parsecs.

The relatively simple appearance of Bok globules, together with the sparsity of surrounding matter, render them attractive candidates for high-resolution mapping. The most well-studied object is B335, at a distance of 250 pc. This globule consists of a visually opaque core of 11 $M_\odot$ and an elongated envelope of twice that mass (Figure 3.20). Near the peak density of the core is a far-infrared star with luminosity 3 $L_\odot$. This star is driving an extended molecular outflow. Observations in CO, which form the basis of Figure 3.20, have been supplemented by others using a variety of tracers, in order to establish the density profile as a function of the distance r from the star. The data are consistent with a profile $n_H(r)$ that rises as r^{-2} for r decreasing from 0.3 to 0.03 pc, and with a flatter profile interior to that.

We will describe in Chapter 6 how molecular line observations can provide temperatures in molecular cloud interiors by utilizing multiple transitions from the same species. In B335, this technique has established an inner temperature near 10 K. At larger distances, the temperature estimate is less secure and relies partially on the chemical enhancement of CO isotopes (see Chapter 6). Figure 3.21 displays the empirical temperature distribution in B335 as a function of n_H. Notice how the measured density values surpass $10^4\ \mathrm{cm}^{-3}$, just as in dense cores. The fact that the temperature *increases* outward (*i. e.*, toward lower n_H) is noteworthy and somewhat surprising in an object containing an embedded star. However, the stellar heating is confined to much smaller radii than these observations probe. The outward rise in the envelope actually stems from the cosmic rays and ultraviolet radiation field in which globules are immersed.

The infrared stars that have so far been detected within Bok globules are of relatively low luminosity. The same holds for the stars within dense cores. What, then, are the cloud fragments that give rise to high-mass stars? The answer to this basic question is not known. It is true that there are infrared point sources of very large luminosity, but these are generally so far away that the surrounding cloud structures are not amenable to high-resolution studies. Moreover, such observations are unfavorable in a statistical sense. Protostars and embedded main-sequence stars of high mass must disperse their parent clouds in a time short compared with the typical stellar lifetime of a few million years. Hence it is difficult to witness the actual birth of an O or B star, but relatively easy to observe the destructive aftermath of that birth in the surrounding, fragmentary molecular gas.

The damage wrought by massive stars may help explain the otherwise puzzling existence of Bok globules themselves. That is, some globules could represent the remnants of larger clouds dispersed by stellar winds and radiation pressure. Finally. the influence of massive stars on their surroundings can also be discerned by comparing dense cores in Taurus-Auriga with those in the Orion Molecular Cloud. Studies of several dozen examples in both environments, carried out with comparable spatial resolution, show the Orion cores to be systematically warmer, by about 5 K. These objects also have NH_3 line widths two to three times broader, indicating a higher degree of turbulent support. Moreover, we will see in Chapter 5 that the elevated gas temperatures close to Orion's massive stars promote a qualitatively distinct chemical reaction network. Note finally that the typical Orion dense core has three times the mass and twice the

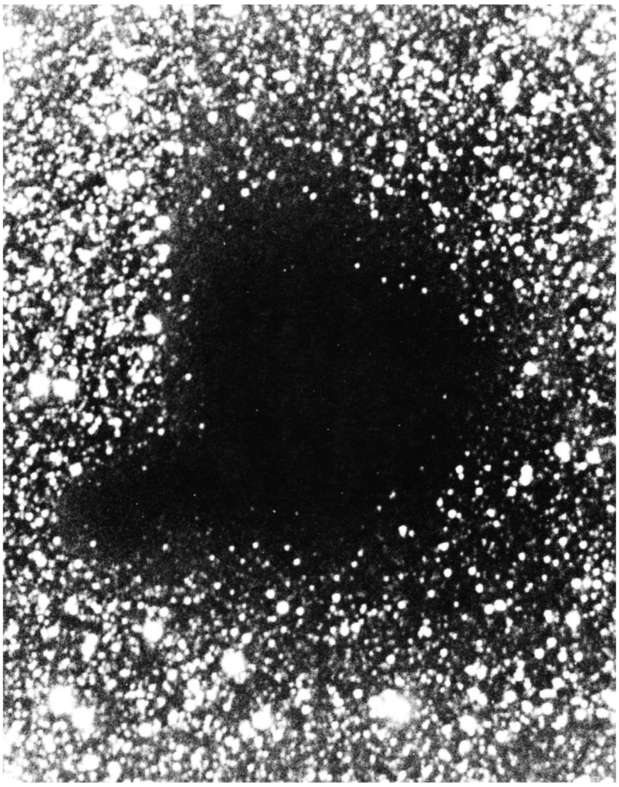


Figure 3.19 Optical photograph of the Bok globule B68 in Ophiuchus.

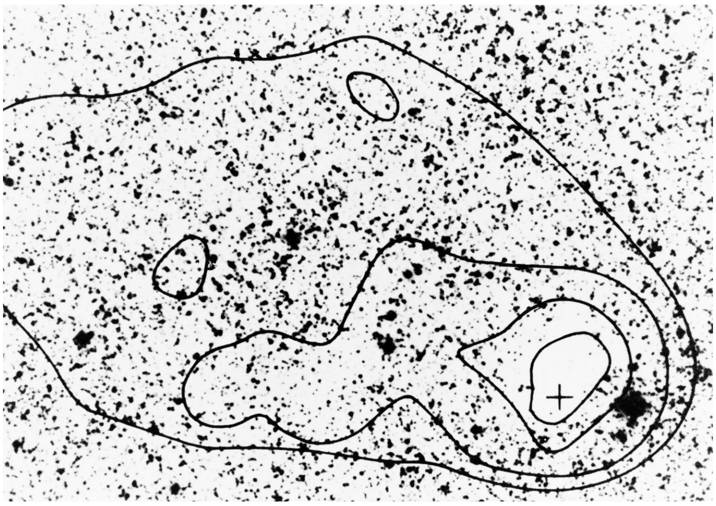


Figure 3.20 Core and envelope of the globule B335. Contours represent the column density in molecular hydrogen, as estimated from CO measurements. The outermost contour spans a diameter of 2.5 pc. The interior cross marks the peak value of the H_2 column density.

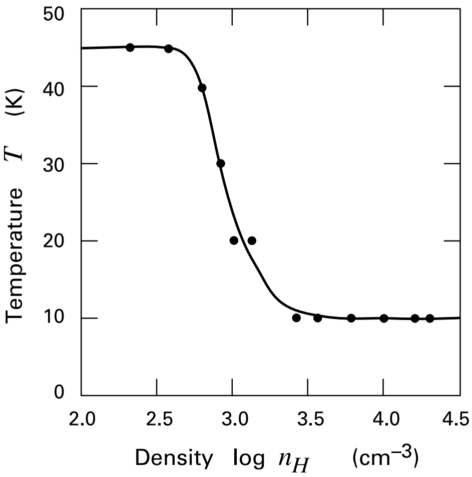


Figure 3.21 Gas temperature as a function of n_H in the globule B335.

diameter of its counterpart in Taurus-Auriga. This difference should be related to the proliferation of high-mass stars in giant molecular clouds and their scarcity in dark cloud complexes. Elucidating the connection is another task for future research.

Chapter Summary

Most Galactic star formation occurs in giant molecular clouds. These structures are clumpy and possess extended envelopes of HI gas. Both the random motion of clumps and the pressure associated with an internal magnetic field prevent immediate gravitational collapse. Thermal pressure is unimportant in this regard. After a period not greatly exceeding 10^7 yr, the entire cloud is dispersed by winds and radiative heating from massive stars it previously created.

A convenient tool for analyzing molecular clouds, or any other self-gravitating structure, is the virial theorem. Here we compare the magnitudes of global, integrated quantities such as the gravitational potential and bulk kinetic energies. In this way, one avoids constructing a detailed interior model, while still gaining understanding of the dominant forces at play.

Giant molecular clouds lie at one end of a hierarchy of morphological types, differing in both linear size and mass density. The internal motion of all clouds decreases systematically with size. Near the other end of the hierarchy are the dense cores and Bok globules that actually produce individual stars and binaries. Cores and globules are intrinsically elongated and rotate slowly in space. Prior to the start of protostellar collapse, they are supported against self-gravity by a combination of thermal and magnetic pressure.

Suggested Reading

Section 3.1 The properties of giant molecular clouds are summarized in

Blitz, L. 1993, in *Protostars and Planets III*, ed. E. H. Levy and J. I. Lunine (Tucson: U. of Arizona Press), p. 125.

Much of the data on these clouds comes from large-scale CO surveys, such as that of

Solomon, P. M. & Rivolo, A. R. 1989, ApJ, 339, 919.

Section 3.2 The role of the magnetic field in the virial theorem is covered by

Shu, F.H. 1991, *The Physics of Astrophysics, Vol. II: Gas Dynamics*, (Mill Valley: University Science Books), Chapter 24.

Section 3.3 A cogent summary of dense cores and Bok globules, written from an observational perspective, is

Myers, P. C. 1995, in *Molecular Clouds and Star Formation*, ed. C. Yuan and J. You (Singapore: World Scientific), p. 47.

Historically, the significance of the globules in star formation was first recognized in the prescient work of

Bok, B. J. & Reilly, E. F. 1947, ApJ, 105, 255.

Our discussion of B335 is based on the comprehensive study by

Frerking, M. A., Langer, W. D., & Wilson, R. W. 1987, ApJ, 313, 320.

The relation between cloud sizes and velocity dispersions was first uncovered by

Larson, R. B. 1981, MNRAS, 194, 809.

For the three-dimensional shapes of dense cores, see

Ryden, B. S. 1996, ApJ, 471, 822,

which also contains earlier references. The use of dust emission to probe internal structure is exemplified by

André, P., Ward-Thompson, D., & Motte, F. 1996, AA, 314, 625.

The Formation of Stars
Steven W. Stahler and Francesco Palla

4 Young Stellar Systems

The fraction of visible stars in the sky that are located in well-defined groups like the Pleiades is relatively small. However, such stars are always found to be young, either through their photospheric properties or by the presence of nearby gas. In addition, the probability that any randomly chosen star belongs to some aggregate rises sharply with stellar mass. Since more massive stars also tend to be younger on average, these facts suggest that clustering could be an important feature of early stellar evolution. We now believe, in fact, that *all* stars are born within groups scattered throughout the Galactic disk. These are either true *clusters*– entities with a high space density that are usually gravitationally bound – or the much looser *associations*, which can extend over 100 pc or more.

We noted in Chapter 1 how the theoretical lifetimes of massive stars, together with estimates for the total mass in OB associations, yield the current stellar production rate in such groups throughout the Galaxy. Similarly, the total masses of other types of associations and clusters give their present birthrates. Comparison of the results shows that star formation in all optically visible clusters accounts for only about 10 percent of the Galactic total. The remainder occurs primarily in OB associations, although their precise contribution is still uncertain. The difficulty is that the typical OB association is so far away that its lower-luminosity members cannot be seen directly, and the total mass must be inferred by theoretical extrapolation. Fortunately, this situation is rapidly improving through the use of sensitive detectors at infrared and X-ray wavelengths.

This chapter is a descriptive survey of the various groups into which stars are born. We begin with optically invisible aggregates still embedded within the gas and dust of molecular clouds. Partially revealed associations are the next topic. We cover in turn T, R, and OB associations, each distinguished by the masses of its prominent members. This traditional nomenclature, while motivated observationally, is somewhat misleading, as each region includes a substantial mass range. Section 4.3 is a brief overview of the fully exposed open clusters, for which observations are most complete. Finally, we discuss the distribution of stellar masses, both in nascent groups and within the general field population.

4.1 Embedded Clusters

Clusters of optically revealed stars are the scant remains of much more populous systems created within the dense interiors of molecular cloud complexes. Although no complete census exists, observations are consistent with the hypothesis that most stars form in such environments. We shall use the term "embedded cluster" generically, to signify any group of physically related stars so obscured by ambient molecular gas that most can be detected only at infrared and longer wavelengths. In applying this terminology, we recognize that the issue of whether any particular group will remain gravitationally bound following dispersal of its gas is rarely, if ever, known with confidence.

The Formation of Stars. Steven W. Stahler and Francesco Palla

ISBN: 3-527-40559-3

4.1.1 Near-Infrared Surveys

The discovery of embedded stellar aggregates resulted from a key technological advance in infrared astronomy. Prior to the early 1980s, observational surveys at these longer wavelengths could not attain the fine detail available with optical instruments. The situation dramatically changed with the advent of *near-infrared array detectors*. These solid state devices provide in a relatively short exposure time detailed views of embedded systems, even those of large angular size. A filter in one of the standard wavelengths precedes the detector, so that a monochromatic image results. Combining several such images also allows one to produce a composite, false-color rendition. We will later show examples of both types.

The utility of near-infrared radiation for penetrating large columns of molecular cloud gas is evident from the interstellar extinction curve (Figure 2.7). It can be seen that a photon in the K band, centered at 2.22 µm, has an extinction 0.1 times that of a V-band photon at 0.555 µm. Consider now a representative T Tauri star of spectral type K7, with $M_V = +6.5$ and $M_K = +2.2$. If such a star were at a distance of 200 pc, equation (2.12) tells us that it would have an apparent V-magnitude above a reasonable detection threshold, *e. g.*, $m_V = +25$, only if the associated cloud extinction A_V were under 12 mag. On the other hand, the same star could be inside a cloud with $A_V = 100$ and still be detectable at K, where the limiting magnitude at large, groundbased telescopes is currently about +20. Of course, observations in the mid- and far-infrared regimes would be even more effective in this regard. Such radiation, however, is so strongly absorbed by the Earth's atmosphere that its detection, at the longest wavelengths, requires spaceborne instruments. The 1983 launch of IRAS first allowed reconstruction of the nearly complete spectral distribution of emitted energy from numerous embedded stars.

Clusters inside molecular clouds have thus far mostly been discovered through surveys in a single near-infrared waveband. A cluster is usually first identified as a region with a significant overdensity of sources compared with nearby fields. However, this initial reconnaissance work is never sufficient to establish the true membership. Many, if not most, of the stars in the group will be intrinsically fainter than the ones first seen. Besides accounting for completeness, it is also necessary to separate out background objects that are reddened by the same cloud. One may estimate the total cluster population statistically by using off-cloud observations to subtract the expected number of background and foreground stars in the region. Individual cluster members can be selected in principle by their proper motion (movement in the sky relative to the background), but one needs at least two observations widely separated in time. Other identification techniques include spectroscopy and multicolor photometry.

Let us consider further the photometric method, which often combines K-band observations with those at J (1.25 µm) and H (1.65 µm). The principal tool in such studies is a near-infrared *color-color diagram*. As illustrated in Figure 4.1, one plots the $J - H$ color on the vertical axis and $H - K$ horizontally. The magnitude conventions imply that the numerical values of both $J - H$ and $H - K$ increase for redder, cooler stars. In any individual case, the observed colors depend both on the photospheric properties and on the extinction provided by the cloud. Background stars, however, exhibit a well-defined relationship between the two colors.

To obtain this relationship, we assume that all sources outside the cluster are either main-sequence stars or the rarer but more luminous red giants. If stellar surfaces radiated as perfect blackbodies, equation (2.29) indicates that the ratio of emergent fluxes at any two wavelengths would be a unique function of the temperature $T_{\rm eff}$. The dotted line in Figure 4.1 displays the blackbody values of $J-H$ and $H-K$ for the indicated range of $T_{\rm eff}$. Stellar photospheres depart

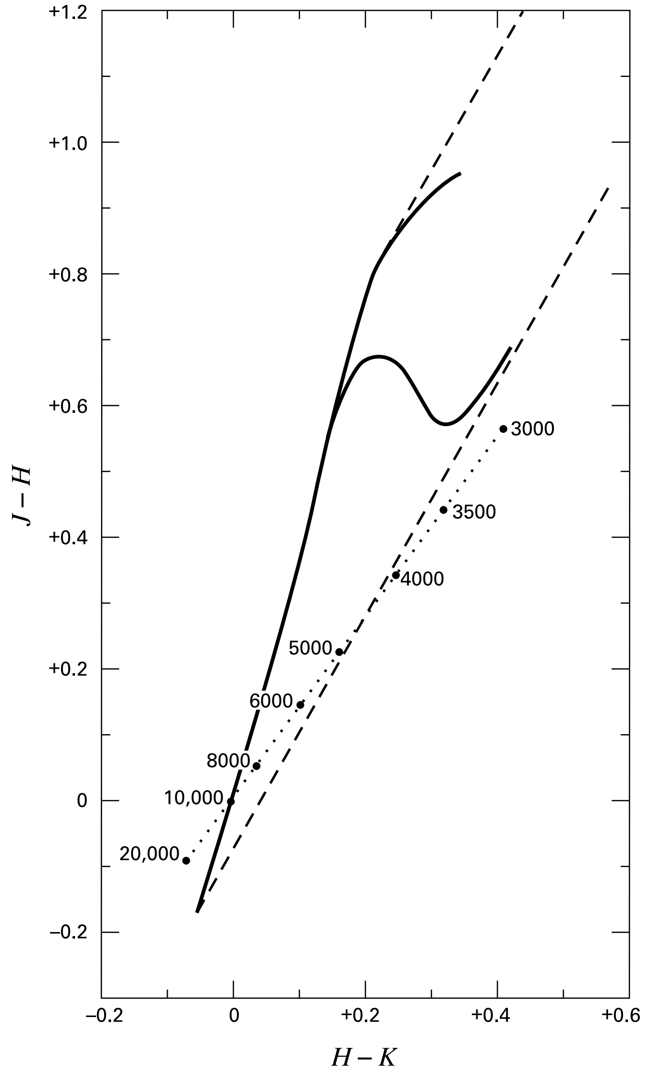


Figure 4.1 Near-infrared color-color diagram. The $J - H$ and $H - K$ color indices are displayed as the vertical and horizontal axes, respectively. The solid curve shows the relation between these indices for main-sequence stars (*lower branch*) and giants (*upper branch*). The dotted curve shows the colors of blackbody spectra at the indicated temperatures. Straight dashed lines indicate the relative color changes due to interstellar reddening.

from a blackbody because they are not equally opaque at all wavelengths. A dominant source of opacity in the outer layers of low-mass stars (including the Sun) is the H^- ion. The opacity from H^- has a broad minimum near 1.6 µm, close to the H-band. Consequently, cool stars that emit much of their energy in this regime have near-infrared colors that differ significantly from blackbodies. This deviation is evident in the solid curve in the figure, which represents a sequence of photospheres for both main-sequence stars and giants. Moving to the right along this curve toward lower $T_{\rm eff}$, the surface opacity is increasingly dominated by molecular lines. A prime source in the near-infrared is CO, whose rich spectrum we shall encounter again in Chapter 5. The depth of the numerous CO absorption lines, and consequently the broadband color, is sensitive to the stellar surface gravity, which is lower in the giants. Hence, the solid curve eventually bifurcates, at a spectral type of early K. The upper branch represents the giants, while the lower is the main sequence.

Thus far we have considered only *unreddened* stars, such as those found in front of the cluster of interest. As we saw in Chapter 2, colors are also modified through dust extinction. The actual values of the color excesses $E_{\rm J-H}$ and $E_{\rm H-K}$ depend on the column density, but a general relation between the two may be deduced from the extinction curve. Referring once more to Figure 2.7, we find

$$\begin{aligned}\frac{E_{\rm J-H}}{E_{\rm H-K}} &= \left(\frac{E_{\rm J-V}}{E_{\rm B-V}} - \frac{E_{\rm H-V}}{E_{\rm B-V}}\right)\left(\frac{E_{\rm H-V}}{E_{\rm B-V}} - \frac{E_{\rm K-V}}{E_{\rm B-V}}\right)^{-1} \\ &= \frac{2.58\,-\,2.25}{2.77\,-\,2.58} = 1.74\ .\end{aligned} \tag{4.1}$$

Thus, for any background star with intrinsic colors $(J-H)_\circ$ and $(H-K)_\circ$, the observed colors lie along the *reddening vector* given by

$$(J-H)\,-\,(J-H)_\circ\,=\,1.74\,[(H-K)\,-\,(H-K)_\circ]\ . \tag{4.2}$$

Returning to Figure 4.1, the measured $J-H$ and $H-K$ values for an ensemble of stars should fall within the band enclosed by dashed lines.

In Figure 4.2 we show a color-color diagram for IC 348, a compact cluster located in the Perseus Molecular Cloud, at a distance of 320 pc. This system does contain numerous optically visible stars, but many more are still embedded. Of the 342 sources plotted, comparison with adjacent fields indicates that about 60 percent should be cluster members. The majority of stars in the diagram have near-infrared colors that are close to, but above, the main-sequence curve, with a displacement along the reddening vector that corresponds to $A_V \lesssim 5$ mag. A smaller number of sources have significantly larger A_V, while some 20 percent lie well outside the dashed boundaries altogether. These latter stars, which are redder in $H-K$ than would be expected from their $J-H$ colors, are said to exhibit an *infrared excess.*

4.1.2 Classification of Member Stars

The infrared excess of a young star arises not from the reddening due to distant grains, but from emission relatively close to the stellar surface. That is, the phenomenon is *circumstellar*, rather than *interstellar*, in origin. As our present example illustrates, all stars within embedded clusters are subject to reddening by foreground dust, whether or not they have an infrared excess. Once the full spectral energy distribution of a source is established from observations, one may use the estimated A_V through the cloud, together with the extinction curve, to calculate A_λ, the extinction at all other wavelengths. In this manner, the "dereddened" energy distribution stemming from the star and its circumstellar matter can be reconstructed. Stars with heavy infrared excesses and those with none lie at the extremes of a rather well-defined morphological sequence of broadband spectra.

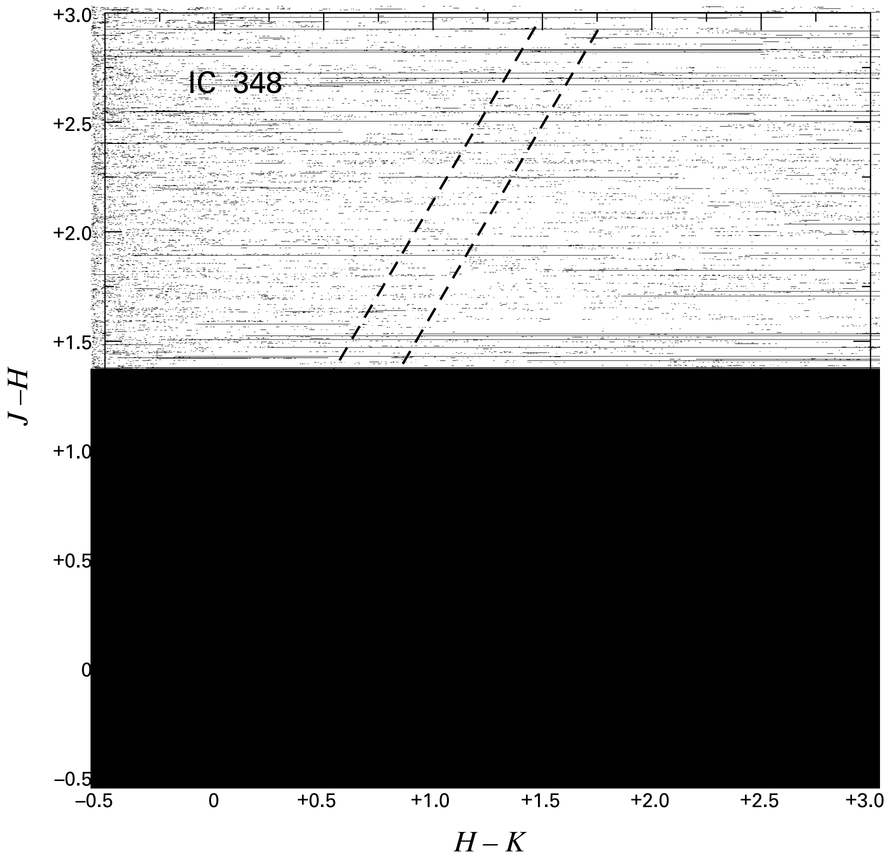


Figure 4.2 Color-color diagram of the embedded cluster IC 348. The open circles represent member stars. The solid curve at the lower left, partially hidden from view, is the main-sequence relation from Figure 4.1, shown with the associated reddening band (*dashed lines*). The reddening vector corresponding to $A_V = 5$ mag is at the lower right.

To illustrate the trend, we turn to another well-studied region, the dense cluster at the heart of the ρ Ophiuchi dark cloud complex. Near-infrared surveys have detected hundreds of embedded stars in the L1688 cloud, at the western end of the complex (recall Figure 3.17). Toward the center of this cloud, the visual extinction exceeds 50 mag. The dereddened spectral energy distributions of three representative sources are shown in Figure 4.3. Plotted as a function of wavelength is λF_λ, the flux measured per logarithmic wavelength interval. The spectrum of the star WL 12 shows a pronounced infrared excess, in that λF_λ, which would normally peak near 1 µm in a late-type star, is still increasing out to 60 µm. The flux eventually does fall at submillimeter and longer wavelengths, which are not included here. For the star SR 24, the flux has a shallow negative slope in the mid- and far-infrared regime, while for SR 20 it falls steeply. In this last example, the spectrum is beginning to resemble a blackbody curve, also shown in the figure.

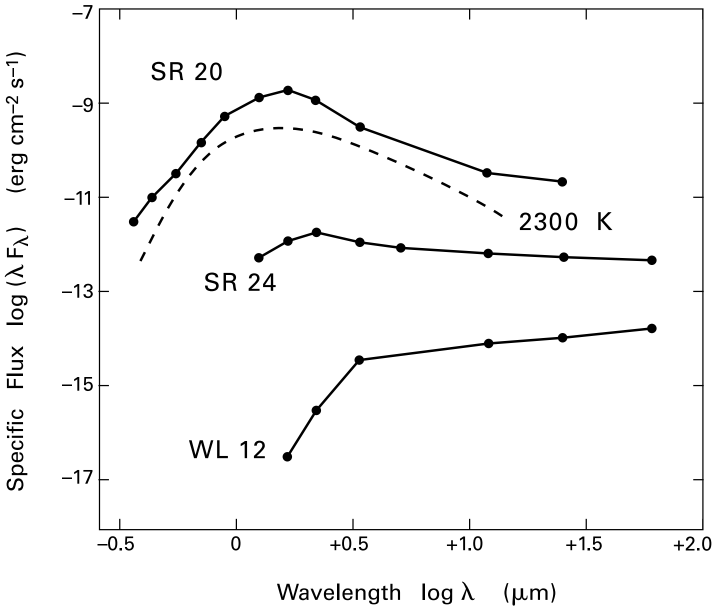


Figure 4.3 Spectral energy distribution of three stars in the ρ Ophiuchi dark cloud complex. The dashed curve corresponds to a blackbody at 2300 K. From bottom to top, these broadband spectra exemplify Class I, II, and III sources, respectively.

The copious infrared emission from the first two sources stems from heated dust grains. Note from equation (2.32) that thermal radiation which peaks at 10 μm has an associated temperature near 300 K. The typical infrared excess does *not* display a pure blackbody spectrum, indicating a significant range of temperatures. Moreover, these temperatures are high enough that the dust in question must be relatively close to the star. It is natural to suppose that these grains are part of cloud material that is either participating in protostellar collapse or else was left behind after collapse ended. Such remnant matter gradually disappears with time.

Pursuing this line of reasoning, we may use the infrared excess as an empirical measure of stellar youth. We quantify matters by considering the infrared spectral index $\alpha_{\rm IR}$:

$$\alpha_{\rm IR} \equiv \frac{d\log\,(\lambda F_\lambda)}{d\log\lambda} \ . \tag{4.3}$$

It is conventional to evaluate the derivative by numerically differencing the flux between 2.2 and 10 μm. Infrared sources such as WL 12, with $\alpha_{\rm IR} > 0$, are said to be in Class I. Such objects are generally associated with dense cores, as seen by NH_3 emission. The less embedded star SR 24 is an example of a Class II source, for which $-1.5 < \alpha_{\rm IR} < 0$. Class III stars like SR 20 have $\alpha_{\rm IR} < -1.5$. Finally, a "Class 0" has been added to incorporate sources so deeply buried that they can only be detected at far-infrared and millimeter wavelengths. One example, shown in Figure 4.4, is the object known as L1448/mm. This star of 10 $L_\odot$ lies inside a Bok globule in the Perseus region, some 300 pc distant. Notice how its spectral energy distribution is shifted to much longer wavelengths than those in Figure 4.3. Like all Class 0 objects, L1448/mm is driving a powerful molecular outflow.

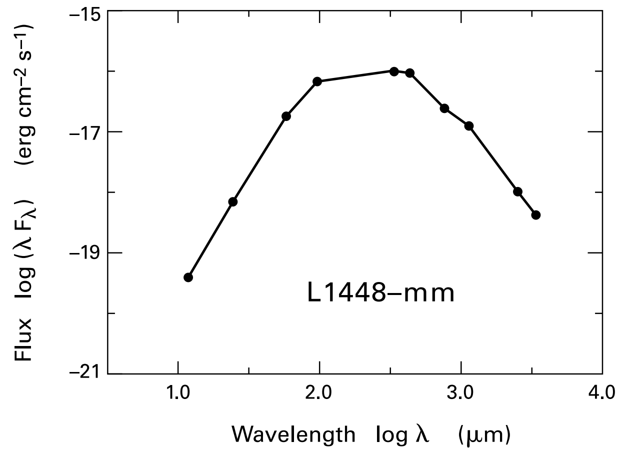


Figure 4.4 Spectral energy distribution of the Class 0 source L1448/mm in Perseus.

4.1.3 Cluster Luminosity Functions

The spectral classification scheme is a convenient means to gauge the evolutionary status of cluster stars that are inaccessible to more direct observation. A complementary tool of equal importance is the cluster *luminosity function.* Here one counts up the number of stars ΔN that have luminosities in the range L_* to $L_* + \Delta L_*$, where L_* itself spans the range of observed values. We recall from Chapter 1 that pre-main-sequence stars have luminosities that evolve during contraction, at rates that are highly mass-dependent. Since any sufficiently young cluster contains a large fraction of such objects, the form of $\Delta N(L_*)$ changes with time until most members have settled onto the main sequence. The potential value of the luminosity function as an evolutionary probe is thus apparent, but observers have not yet taken full advantage of this technique.

The main impediment has been the practical difficulty of obtaining bolometric luminosities for large numbers of embedded stars. The satellite observations used at long wavelengths have so far lacked the high spatial resolution required to sample the crowded fields in dense clusters. Consequently, the most complete luminosity functions at present are those in a single near-infrared wavelength, usually the K band. Figure 4.5 shows the K luminosity function for 90 stars in the central region (L1688 cloud) of the ρ Ophiuchi complex. Note that the falloff in the population for $m_K \gtrsim 10$ simply reflects the incompleteness of observations at this magnitude. Subsequent observations have found the population to increase down to at least $m_K = 14$. At this faint level and beyond, the membership of the numerous sources in the vicinity is less secure.

Measurements at other wavelengths have allowed the determination of bolometric luminosities for about 50 of the ρ Ophiuchi stars in L1688. Figure 4.6 displays the currently known bolometric luminosity function. Also shown here is the distribution of sources among the spectral classes. Once again, the decline in the population at the highest luminosities is real, but that at low $L_{\rm bol}$ is not, and is being pushed back by more sensitive surveys. We will return to the physical interpretation of luminosity functions later in this chapter and again in Chapter 12, when we revisit cluster evolution from a more theoretical perspective.

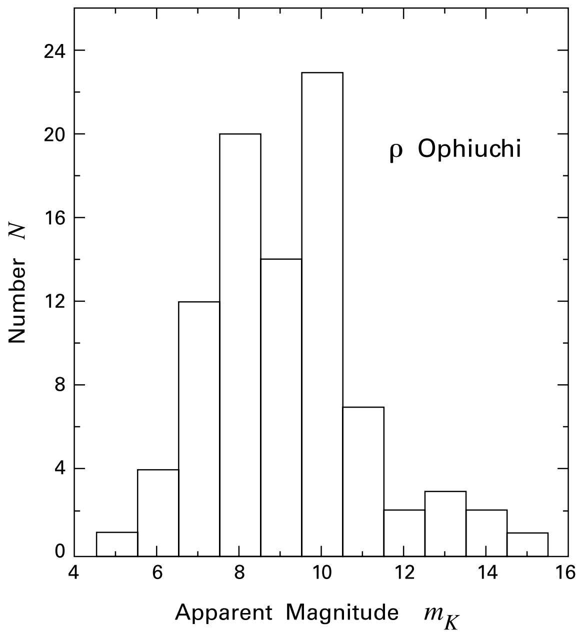


Figure 4.5 Luminosity function in the K band for 90 stars in the L1688 region of ρ Ophiuchi.

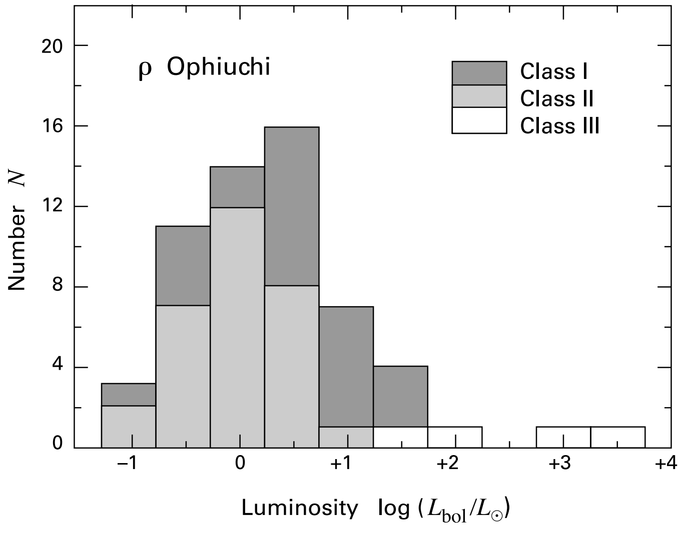


Figure 4.6 Bolometric luminosity function for 55 ρ Ophiuchi stars, again in L1688. The distribution among three spectral classes is indicated.

4.1.4 Morphology of the Groups

Let us now turn to the morphology of actual clusters. A more complete picture is obtained by combining infrared studies of stars and their attendant dust with radio observations of the gas. One important example is the ensemble of clusters within the L1630 region of Orion B. Here, star formation is almost exclusively confined to four discrete regions: NGC 2071, 2068, 2023, and 2024 (recall Figures 1.2 and 1.3). Each of these clusters, roughly 1 pc in diameter, is associated with a previously known HII region or reflection nebula, signaling the presence of at least one O or B star. Stellar densities are of order 100 pc^{-3}, similar to the nucleus of the L1688 cloud in ρ Ophiuchi. In both locations, molecular gas with $n_H > 10^4$ cm^{-3} comprises from 50 to 90 percent of the total mass of several hundred $M_\odot$ within the cluster borders.

The massive, bright stars spawned within many clusters allow these systems to be seen across the Galaxy. Subsequent mapping with near-infrared arrays has then revealed the lower-mass population. Plates 1 through 8 (see end of chapter) are a sequence of images showing groups in formation, at the distance of Orion and beyond. First is NGC 2024, richest of the L1630 clusters. On the left in Plate 1 is the optical view, which is notable for the broad, vertical dust lane obscuring most of the interior stars. Several hundred of these stars are revealed in the infrared image shown as the righthand part of Plate 1. Here it is apparent that the very brightest stars tend to lie in the most crowded portion of the cluster, a region that is invisible optically.

Most embedded systems like this one are *not* destined to form open clusters, but will become unbound after their gas is dispersed. The argument is a statistical one. Assuming that star formation in the Orion Molecular Cloud is representative of other complexes, then roughly 50 clusters like NGC 2024 should be forming now within 2 kpc of the Sun. We noted in Chapter 3 that much of the molecular gas associated with a cluster disappears by 5×10^6 yr, which we may take as a representative formation time. Thus, after 10^8 yr of steady cluster production, about $50 \times 20 = 10^3$ systems should be found in the same Galactic area. In fact, the total number of open clusters this age or younger is less than 100, so that their formation must be quite inefficient.

Plate 2 is a near-infrared view of S106, the nearest (at 600 pc) and best-studied example of a bipolar nebula illuminated by a massive star. Long known as an optical HII region, S106 again has a prominent dark lane in its central region. Here radio studies have uncovered a rich concentration of molecular gas. Situated within the obscuring slab is the infrared source IRS 4, a late-O or early-B star of $10^4\ L_\odot$ that is driving high-velocity, ionized flows into each lobe. These latter two structures span a total length of 0.7 pc. Over 200 stars with $m_K < +14$ are located within a 0.3 pc radius of IRS 4. The stellar density, which again peaks near the massive star, exceeds 10^3 pc^{-3}, approaching that in the Trapezium cluster.

Not all massive stars coincide with low-luminosity clusters, at least according to present observations. Plate 3 is a red photograph showing three HII regions in the Gem OB1 association, 2.5 kpc distant. The left and center regions are denoted S255 and S257, respectively, while the more diffuse region to the right is S254. The bottom panel (Plate 4) is a near-infrared image covering a smaller scale. The two brightest objects on the extreme left and right are the isolated B0 stars exciting S255 and S257, respectively. These stars have no detected low-mass companions. On the other hand, the prominent cluster between them, revealed only in the infrared, contains about 70 members within a 0.5 pc radius. Near the cluster center is an embedded star, designated S255/IR, that radiates close to $10^5\ L_\odot$, largely in the far-infrared. The cluster itself is sandwiched between two peaks of radio emission from a compact molecular cloud.

Intermediate-mass stars are also frequently located within clusters of less massive objects, but this retinue is less dense. Consider the Herbig Be star BD+40°4124, located 1 kpc away, in the direction of the Cygnus spiral arm. Near-infrared imaging (Plate 5) shows the star attended by several dozen embedded sources; these outnumber the nearby visible T Tauri stars by a factor of three. Observations in $^{12}C^{18}O$ and CS reveal two such visible stars, V1318 Cygni and V1686 Cygni, to be situated in a ridge of very dense gas, comprising several hundred $M_\odot$. The molecular outflow and maser activity also detected in the region stem from a bright infrared binary companion to V1318 Cygni, rather than the optical Be star. This companion has the largest infrared excess of any cluster member. Whether other visible Herbig stars actually drive molecular outflows remains an open question.

Occasionally, star forming regions of differing ages are found in proximity, as illustrated in Plate 6. This image shows the environment of NGC 7538, a previously known HII region in the Cas OB2 association. Diffuse radiation from the HII region itself, consisting of both reflected starlight and thermal gas emission, appears as the crescent-shaped nebulosity surrounding the hot, white OB stars. The two prominent red patches are even younger regions containing molecular gas and compact clusters of embedded stars.

One of the most spectacular HII regions is NGC 3603, shown in the central portion of Plate 7. This massive cluster, seen here in the near-infrared, is located in the Carina spiral arm, at a distance of 6 to 7 kpc. The O and B stars alone comprise some 2000 $M_\odot$ and have 100 times the ionizing luminosity of the Trapezium Cluster. Indeed, the center of NGC 3603 is one of the densest concentrations of high-mass objects in the Galaxy. Many of the O and B stars are visible optically, despite the ambient dust. However, the far more numerous low-mass stars are only discernible in the infrared. Their placement in a color-magnitude diagram yields pre-main-sequence contraction ages of 3×10^5 to 1×10^6 yr.

The production of many thousands of stars in such a brief period makes NGC 3603 an impressive HII region but still not on the scale of a true *starburst*. To find these, we need to go outside the Milky Way. The neighboring Large Magellanic Cloud, for example, contains 30 Doradus (Plate 8). This giant HII region is morphologically similar to NGC 3603, but has a total luminosity ten times higher, *i. e.*, of order $10^8\ L_\odot$. The central cluster is bright enough to be seen optically, even at a distance of 50 kpc. Note the great tendrils of gas around the compact, stellar group. These structures give the region its other name, The Tarantula Nebula. Even brighter starbursts lie within dwarf irregular and spiral galaxies, all located at greater distances.

4.2 T and R Associations

The fate of an embedded cluster depends partially on how its gas is dispersed. In many cases, one or more high-mass stars drive off the interstellar matter relatively quickly. The result is the expanding group of stars known as an OB association. Other systems were born in dark cloud complexes that never contained massive stars. The extended distribution of low-mass stars in Taurus-Auriga (Figure 1.11) is not the product of rapid dispersal, but largely reflects the initial extent of the parent cloud. Not only are many of the visible stars too young to have spread far, but a significant number of even younger, embedded sources are commingled spatially with the others. Since the vast majority of the members in this system are T Tauri stars, the Taurus-Auriga stellar complex is designated generically as a *T association*. The term was introduced by V. Ambartsumian in 1949, four years after the identification of the T Tauri class by A. H. Joy.

4.2.1 T Tauri Star Birthsites

While the existence of an embedded cluster can be initially established by simply examining near-infrared images, finding a T association requires also the identification of T Tauri stars *per se*. These come in two varieties. The *classical* T Tauri stars are conspicuous spectroscopically for their strong optical emission lines in Hα, as well as the H and K-lines of Ca II at 3968 Å and 3934 Å, respectively. A practical and efficient search strategy, then, is to equip wide-field telescopes with objective prisms that can simultaneously record many stellar spectra over a few square degrees of the sky. Such surveys, however, capture only part of the population of interest. At least as numerous as the classical members are the *weak-lined* T Tauri stars. As their name implies, these stars lack strong emission lines, although the two types overlap substantially in age. The weak-lined population was actually discovered through its enhanced X-ray emission relative to main-sequence field stars. X-ray detection was made initially with the Einstein satellite, launched in 1979 and operational through 1981. The launch of the more sensitive ROSAT satellite in 1991 provided more weak-lined candidates.

Figure 4.7 shows the distribution of nearby T associations in the Galactic plane. The groups range from TW Hydrae, a small aggregate of T Tauri stars only 50 pc away, to the populous and highly active NGC 1333 in the Perseus cloud complex. Named for a bright reflection nebula, the latter region contains dozens of visible young stars and over a hundred embedded ones, including many that drive molecular outflows. Within the Serpens region, also depicted here, most visible stars are associated with the L572 cloud, which has a dense complement of infrared sources. Finally, the large molecular complex in Cygnus has both revealed T Tauri stars and embedded sources in the L984 and L988 clouds.

Our figure also includes major embedded clusters, such as ρ Ophiuchi and IC 348 (also part of the Perseus cloud). Indeed, *all* the regions shown in this map contain both obscured and visible objects. Within ρ Ophiuchi, for example, is a large group of revealed low-mass objects on the outskirts of the compact L1688 core. The numerous visible stars of IC 348 are more centrally concentrated. Surveys in Hα have discerned a scattered population of visible stars lying outside the four obscured clusters of Orion B. In summary, embedded clusters and T associations should be viewed as extremes along a continuum of morphological types.

The properties of T associations are well illustrated by the nearest prominent example, Taurus-Auriga, which has been scrutinized thoroughly in the infrared, optical, and X-ray regimes. Here, it has been possible to establish the membership of most stars kinematically. One first obtains a radial velocity, V_r, for each star by examining a convenient portion of its optical spectrum. We compare absorption lines with those of a standard star of the same spectral type, yielding the Doppler shift and hence V_r. In Taurus-Auriga, the results for both classical and weak-lined members agree well with the velocities of the local cloud gas, as obtained from molecular lines. Note that both cloud and stellar V_r-values change over the length of the complex, but the entity as a whole appears to be gravitationally bound.

Obtaining the orthogonal component of velocity, *i. e.*, the proper motion, requires comparison of at least two wide-field images well separated in time. Such comparison shows that the proper motions of Taurus-Auriga stars cluster tightly about a single vector, as expected. The one-dimensional dispersion is from 2 to 3 km s^{-1}. No velocity information is available for the more deeply embedded (Class I) members, which constitute about 10 percent of the population. The total number of kinematically confirmed members within the boundary of the complex shown in Figure 1.11 stands at about 100, with 60 being classical and the rest weak-lined T Tauri stars. This tally is probably complete down to a V-magnitude of +15.5.

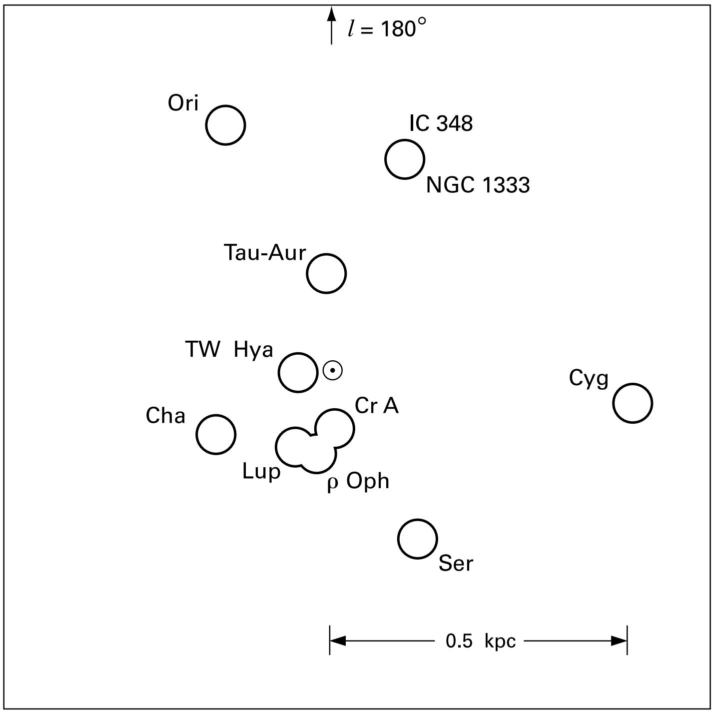


Figure 4.7 Distribution within the Galactic plane of nearby T associations and embedded clusters. Note the longitude convention: the Galactic center is at $l = 0^\circ$, while Cygnus lies near $l = 90^\circ$. Note also the symbol representing the Sun's location.

Although associations do not exhibit the high degree of central concentration seen in embedded clusters, all large molecular clouds, including dark cloud complexes, are intrinsically clumpy. Hence, some degree of clustering is to be expected in the stars born from such structures. Figure 4.8 confirms this hypothesis in the case of Taurus-Auriga. The heavy contours are lines of constant stellar surface density in the plane of the sky, while the lighter curve is the CO boundary from Figure 1.11. Each of the 6 groups, containing 5 to 20 stars each, has a projected radius under 1.0 pc and an internal, radial velocity dispersion of 0.5 to 1.0 km s^{-1}. Thus, most of the total measured dispersion across the complex actually stems from the relative motion between these subunits.

4.2.2 HR Diagrams: Main-Sequence Turnon

A particularly effective means of visualizing the composition and evolutionary status of a T association is to place its optically visible members in the theoretical HR diagram. Figure 4.9 displays the diagrams for four associations. In most cases, the quantity L_* has been extrapolated from the luminosity in the J-band. To this monochromatic value was added a bolometric correction. Strictly speaking, the latter is appropriate only for a *main-sequence* star. Detailed comparisons find, however, that the resulting L_*-values are accurate to within an error of about 20 percent.

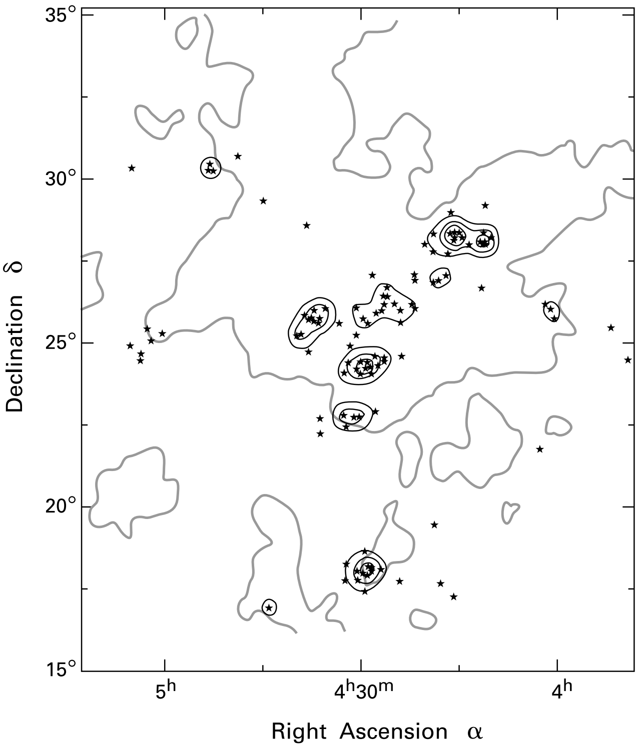


Figure 4.8 Clumping of stars in Taurus-Auriga. The heavy contours represent stellar surface density, while the solid grey contour is the border of CO emission from Figure 1.11.

Beginning with Taurus-Auriga (Figure 4.9a), we see that many stars are crowded near the birthline, also shown in each panel. These members have only recently dispersed their obscuring envelopes of dust and gas. An even younger population is represented by the embedded, infrared sources, which, in the absence of a measurable effective temperature, cannot be placed in a conventional HR diagram. The youngest optical members are mostly classical T Tauri stars, symbolized by the filled circles, but contain an admixture of weak-lined members (open circles). Below the birthline, the number of low-mass stars drops off before the main sequence is reached, at isochrones corresponding to several million years. Thus, the complex as a whole began forming stars at that epoch and continues to produce them today.

It is also apparent that the proportion of non-emission stars increases markedly as a function of age. This trend is consistent with the fact that the strong emission lines present in the classical T Tauri population are absent in ZAMS stars of the same mass. In addition, most of the weak-lined stars, including the ones shown here, lack the infrared excess of their classical counterparts. [1] Thus, both the hot gas ($T \sim 10^4$ K) creating optical emission lines and the cooler

---

[1] An infrared excess always signifies the presence of circumstellar dust. Hence those weak-lined stars with main-sequence spectral energy distributions in the infrared were formerly designated “naked” T Tauri stars. The open circles in Figure 4.9 also include some older, “post-T Tauri” stars, to be defined shortly.

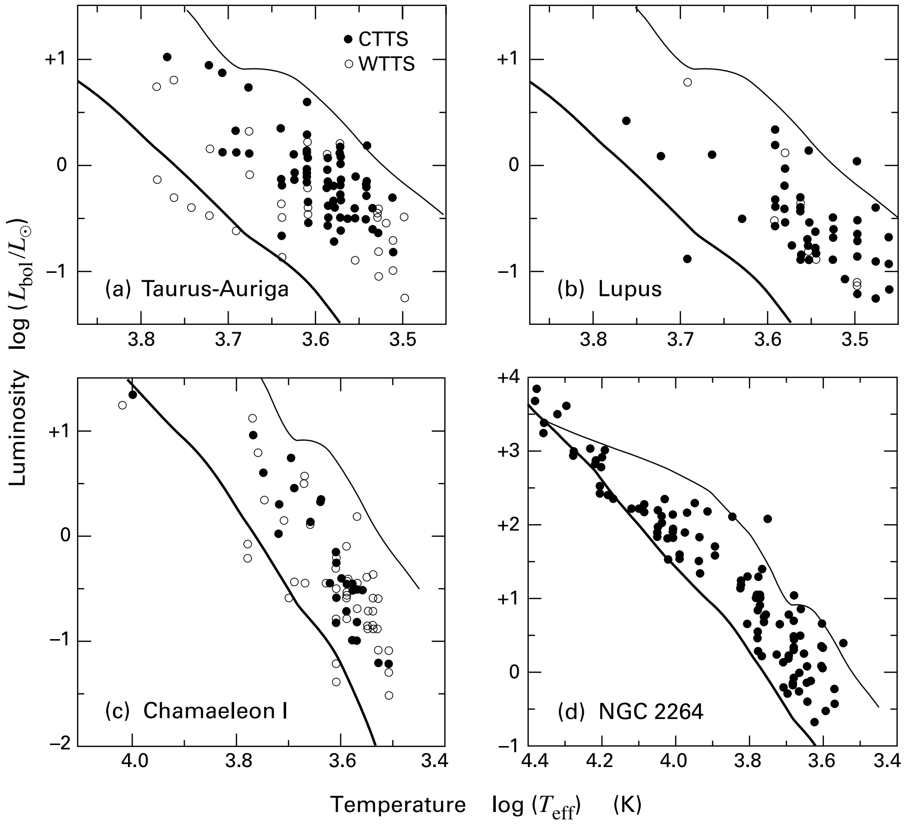


Figure 4.9 HR diagrams for four stellar associations. In panels (a)–(c), closed circles represent classical T Tauri stars, while open circles are weak-lined and post-T Tauri stars. For NGC 2264, we show both classical T Tauri and Herbig Ae/Be stars, as well as main-sequence objects. The upper and lower solid curves in each panel are the birthline and ZAMS, respectively.

dust ($T \sim 10^2 - 10^3$ K) emitting in the near- and mid-infrared regimes gradually disappear in a low-mass pre-main-sequence object.

The highest stellar density in a nearby T association occurs in the Lupus constellation in the southern sky. Because of its location, about 10° south of ρ Ophiuchi, this large and active association is less well-studied than Taurus-Auriga. Star formation is mainly confined to four subgroups embedded in an extended dark cloud complex (see Figure 4.10). The total mass in molecular gas is $3 \times 10^4\ M_\odot$, close to that for Taurus-Auriga, with a substantial fraction located in the isolated B228 cloud. The greatest concentrations of CO emission correspond to filamentary dust lanes apparent in optical photographs.

Originally, the young stars catalogued in Lupus were classical T Tauri's discovered in objective prism surveys. Later X-ray studies found other, weak-lined objects. In addition, there is a Herbig Ae star of 71 $L_\odot$. Fully half of the association members are in the Lupus 3 subgroup; these stars are only indicated schematically in Figure 4.10. The Ae star (HR 5999) is part of a binary pair at the center of this highly compact stellar birthplace. The HR diagram

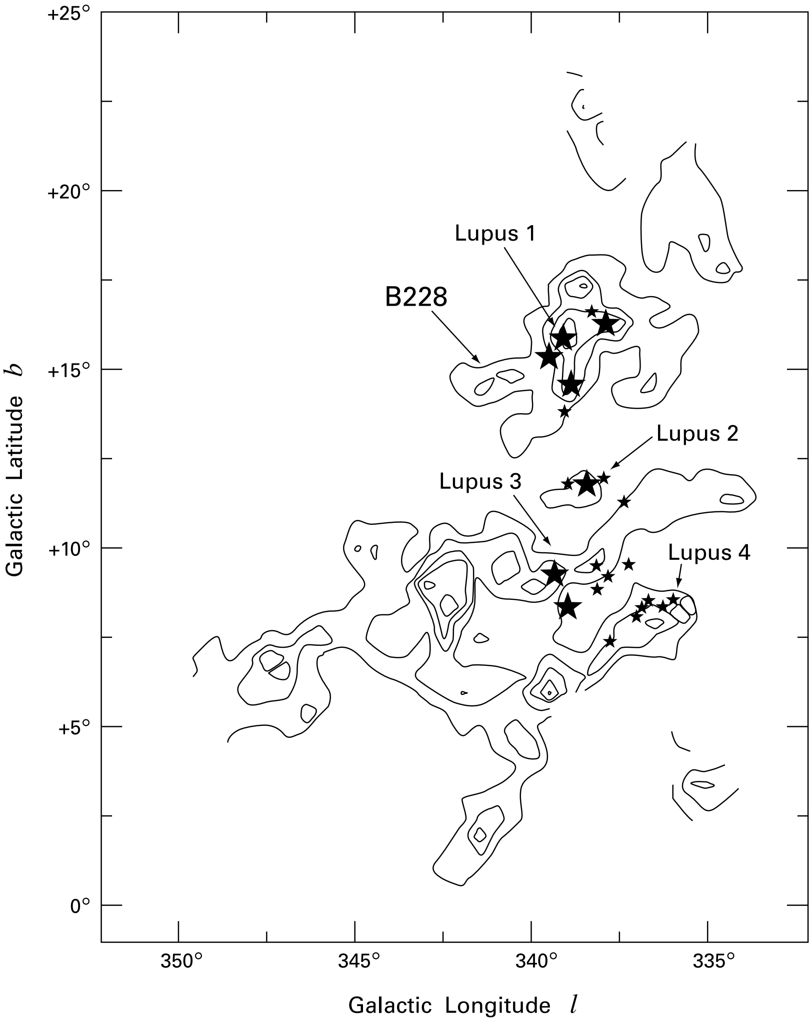


Figure 4.10 Stars and molecular gas in the Lupus association. The contours trace $^{12}C^{16}O$ intensity. Large star symbols represent tight clusters.

for all the subgroups together (Figure 4.9b) again shows a population near the birthline, indicating current star formation activity. The picture will be more complete once infrared and weak-lined sources are studied more systematically. Finally, we note that a polarization map of background starlight reveals a well-ordered magnetic field oriented roughly perpendicular to the most prominent filaments.

In the case of Taurus-Auriga, the distance of 140 pc can be determined reliably by comparing the absolute and dereddened apparent magnitudes of main-sequence stars that have reflection nebulae, and are therefore physically associated with the dark clouds. For the more sparsely sampled Lupus region, a more uncertain figure of 150 pc follows from its proximity to the Scorpius-Centaurus OB association. Yet a third T association at a similar distance is that of

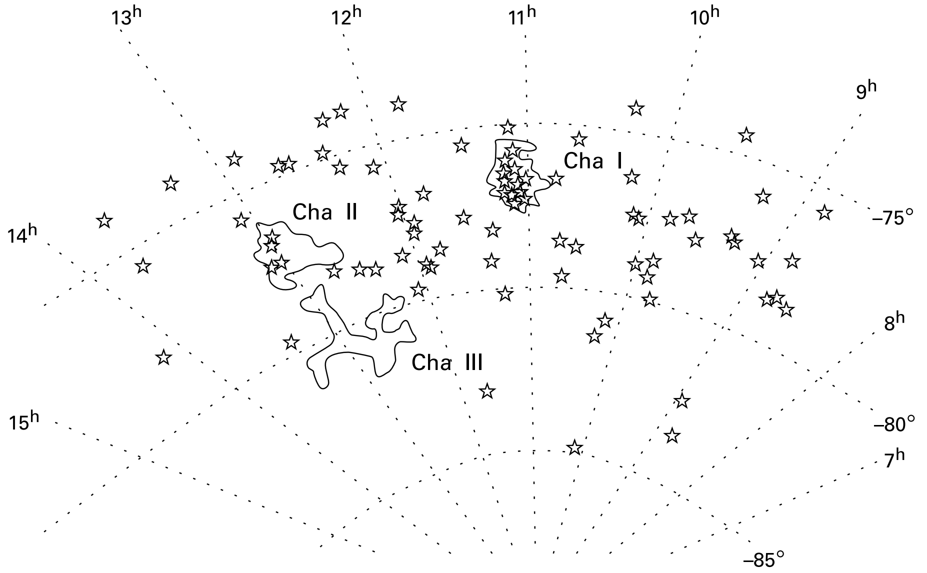


Figure 4.11 Halo of X-ray emitting stars in the Chamaeleon region. The solid contours trace 100 µm continuum emission, delineating the dust and gas. Dotted lines and curves indicate right ascension and declination.

Chamaeleon. In this region, close to the South celestial pole, the molecular gas is confined to three well-defined, coherent structures, as labeled in Figure 4.11. Of these, Chamaeleon I has both the highest visual extinction and the largest population of young stars, while the more irregular Chamaeleon III appears to lack any star formation activity. Since the Chamaeleon constellation is 15° from the Galactic plane, all three clouds are relatively free of background stars.

The HR diagram for Chamaeleon I includes some 80 association members. About half were first identified optically, either from Hα emission surveys or else by virtue of their photometric variability, another characteristic property of classical T Tauri stars. The remainder are weak-lined stars discovered initially through their X-ray emission by the ROSAT satellite. It is clear from the diagram in Figure 4.9c that the classical and weak-lined stars are thoroughly intermingled in both mass and age, with many close to the birthline. Further evidence of ongoing star formation is the presence of infrared sources with no optical counterparts; some of these are exciting Herbig-Haro objects and molecular outflows.

X-ray observations have revealed an additional population of stars in the Chamaeleon region. Some lie inside the molecular cloud boundaries, but most do not, and are distributed over tens of parsecs (Figure 4.11). Similar halos surround the Taurus-Auriga, Orion, and Lupus regions. A relatively young age is indicated for some of these objects by the presence of surface lithium, which is gradually destroyed in the course of stellar evolution (Chapter 16). This subset consists largely of post-T Tauri stars, *i. e.*, contracting objects intermediate in their properties between classical and weak-lined stars on the one hand, and those already settled onto the main sequence (Chapter 17). The rest are even older, with ages perhaps as great as 10^8 yr. In any case,

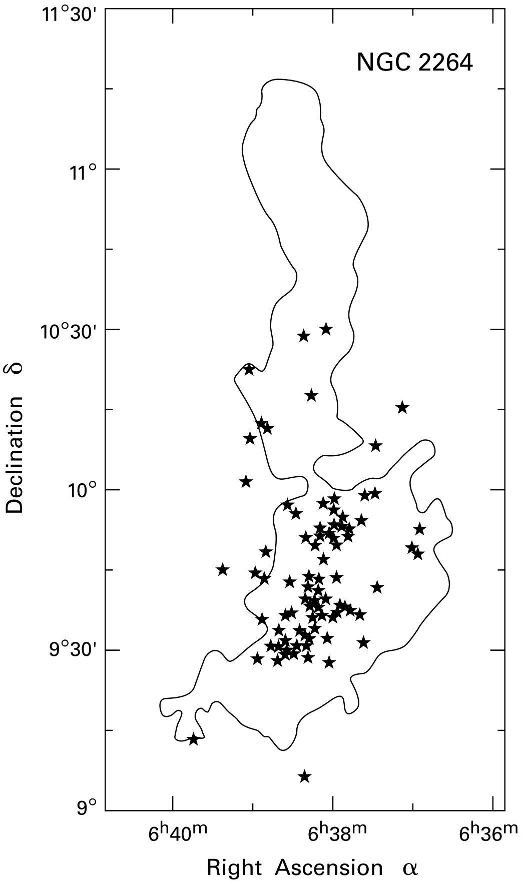


Figure 4.12 Stars and gas in NGC 2264. The solid contour represents the boundary of emission in $^{12}C^{16}O$.

this halo population either migrated from the present-day Chamaeleon region or else formed out of molecular gas that has long since vanished.

The fourth HR diagram of Figure 4.9 shows NGC 2264, a populous grouping of stars in Monoceros. In both evolution and morphology, this region represents a transition from a fully embedded to an open cluster. Several hundred optically visible members range in mass from the O7 star S Mon to very late-type T Tauri stars. The system lies at a distance of 800 pc, too far for complete X-ray identification of weak-lined stars. As seen in Figure 4.12, most visible members are crowded into the southern portion of a large molecular cloud. This cloud in turn is part of the Mon OB1 complex, whose boundary we have also included in Figure 1.1. Judging from the slight color excesses observed, the visible cluster sits just in front of the cloud, while a host of embedded infrared stars – some driving vigorous molecular outflows – extend further behind. The cloud itself, some 25 pc long and with a mass of $3 \times 10^4\ M_\odot$, conveniently blocks background starlight at optical wavelengths, facilitating study of the cluster members.

The HR diagram of NGC 2264 again shows numerous stars close to the birthline, including now the most massive pre-main-sequence objects near 10 $M_\odot$. In addition, there is a clearly defined main sequence, but only down to 3 $M_\odot$, corresponding to a spectral type of A0. Less massive stars, with their slower contraction rates, have not yet had time to reach the ZAMS. His-

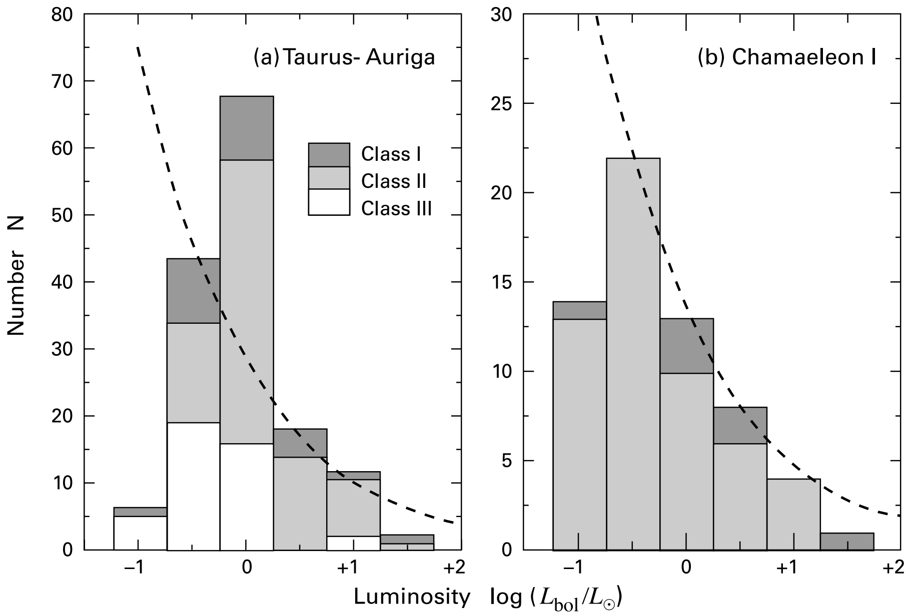


Figure 4.13 Bolometric luminosity functions for (a) Taurus-Auriga and (b) Chamaeleon I. The dashed curve is the initial luminosity function, defined here as the relative number of field stars per logarithmic unit of $L_{\rm bol}$ (see § 4.5.1.). Shading indicates different infrared classes, as explained in panel (a).

torically, it was the discovery of this *main-sequence turnon* in the equivalent color-magnitude diagram that enabled M. Walker in 1956 to demonstrate unambiguously the existence of the pre-main-sequence phase. Through spectroscopic analysis, Walker then confirmed that the youngest low-mass members were the recently identified T Tauri class. Note that the contraction time from the birthline to the ZAMS for a 3 $M_\odot$ star is 2×10^6 yr. This figure represents the time in the past when visible stars first began to emerge from the front face of the Mon OB1 cloud. Clearly, this process is continuing today.

4.2.3 Association Luminosity Functions

The bolometric luminosity functions of T associations, like those for embedded clusters, are another potentially valuable diagnostic and less plagued in this case by incompleteness of the sample. Figure 4.13a shows the luminosity function for 130 stars in Taurus-Auriga. Here, as in Figure 4.6, we have subdivided the members by their near-infrared class. The peak near $L_{\rm bol} \sim 1\,L_\odot$ is real, since the limiting luminosity of the survey is closer to $0.1\,L_\odot$.

Figure 4.13b displays the same data for 62 stars in Chamaeleon I. The existence of a maximum in this case is more problematic. In addition, the absence of Class III objects simply reflects the selection criteria for membership, which included the detection of a near-infrared excess. For both associations, the luminosity functions have more stars than the respective HR diagrams, which require spectroscopic analysis to establish $T_{\rm eff}$.

How are we to gauge the significance of Figure 4.13? It is instructive to compare these results with the theoretical luminosity distribution of field stars as they first appear on the ZAMS. This so-called *initial luminosity function*, denoted $\Psi\,(L_{\rm bol})$, is shown by the dashed curves in Figure 4.13. As described in § 4.5 below, we obtain the function by combining stellar luminosities with main-sequence lifetimes. In both Taurus-Auriga and Chamaeleon I, $\Psi\,(L_{\rm bol})$ matches rather well the observed falloff at high luminosity. For Taurus, however, the data display a steeper decline near 1 $L_\odot$. Figure 4.6 shows that the same is true for the embedded ρ Ophiuchi cluster. Another difference from the smooth field-star curve is the maximum in the Taurus-Auriga luminosity function. Both these characteristics are in accord with theoretical expectations (Chapter 12). The "initial" luminosity function is in fact only reached gradually in a cluster or association, as pre-main-sequence members contract and cool.

4.2.4 Intermediate-Mass Objects

We have seen how the discovery of a T association generally begins with an objective prism search for emission-line objects. Included in this category are the rarer Herbig Ae and Be stars, which are often picked out through the same technique. Most intermediate-mass stars, however, are even closer to the ZAMS and lack prominent emission lines. On the other hand, their contraction times are so short that such stars are frequently still illuminating nearby molecular gas. Groups of young, intermediate-mass stars are therefore conspicuous on optical photographs by the *reflection nebulae* accompanying them. These appear as fuzzy patches that are bluer in color than their host stars, since the reflection of visible light from dust grains is more efficient at shorter wavelengths. We call such stellar groups *R associations*.

Because intermediate-mass stars lack both the brilliance of more massive objects and the large numbers of T Tauri stars, R associations have not received a great deal of scrutiny. The

Table 4.1 The Nearest R Associations

Name	Distance (pc)	B stars
Taurus R1	110	4
Taurus R2	140	2
Scorpius R1	150	9
Perseus R1	330	4
Taurus-Orion R1	360	5
Cepheus R2	400	5
Vela R1	460	3
Cassiopeia R1	530	5
Orion R1/R2	470	6
Cepheus R1	660	3
Canis Major R1	690	8
Monoceros R1	800	4
Monoceros R2	830	7
Vela R2	870	6
Scorpius R5	870	4

typical system consists of a dozen or so A- and B-type stars spread out over perhaps 10 pc. Intermingled with these objects is a larger population of low-mass stars, many of which are T Tauri's. The majority of intermediate-mass stars in the association are on the main sequence, but some have the emission lines indicative of pre-main-sequence contraction. From Figure 1.18, these latter members have nearly identical luminosities as ZAMS stars of the same mass but lower effective temperatures. One well-known R association is Mon R1, located near the NGC 2264 cluster. Other examples include Ori R1 and Ori R2. The first is within the L1630 region of the Orion B molecular complex (Figure 1.3), while the second is a more widely dispersed collection of reflection nebulae centered on the Trapezium in Orion A. Table 4.1 lists all the R associations within 1 kpc. Here we give both the distance and the number of identified B stars.

Photometric techniques are employed to study the properties of the dust within R associations. Recall that stars already on the main sequence have well-known absolute magnitudes and intrinsic colors as a function of their spectral type. Given the latter, the apparent $B - V$ color of a main-sequence star illuminating a reflection nebula immediately yields its color excess $E_{\rm B-V}$. Suppose now that the distance to the association is already established. Then the apparent V-magnitude of the same member star gives its associated A_V through equation (2.12). One common result of such investigations is that A_V varies substantially from one star to another within the R association, indicating a clumpy distribution of dust. More intriguing is the fact that the ratio $A_V/E_{\rm B-V}$ is often higher than the fiducial interstellar value given by equation (2.16). Such "greyer" extinction is a sign that the typical grain is abnormally large, the result presumably of continuing mantle growth within the denser portions of the enveloping clouds.

4.3 OB Associations

As we consider progressively more massive young stars, any cloud material in proximity not only reflects starlight, but starts to generate its own optical radiation. This emission stems from ionization created by the ultraviolet component of the stellar spectrum. The O and early-B stars capable of such ionization are themselves often found in loose collections of a few dozen members. Although their boundaries are often difficult to locate with any precision, these OB associations extend over regions that can be as small as ordinary open clusters, or as large as several hundred parsecs in diameter.

Historically, the tendency for O and B-type stars to cluster was recognized as soon as precise spectral classification became available, at the beginning of the 20th century. Spectroscopic and proper motion studies gave a physical basis to the observed grouping by establishing common spatial velocities for bright stars in Orion, Perseus, and the Scorpius-Centaurus region. It gradually became clear that their large *internal* velocities, typically about 4 km $\rm s^{-1}$, doom these systems to expansion and eventual dispersal. Observational verification of the expansion came in 1952, when A. Blaauw measured proper motions in what is now called Per OB2 (signifying the second OB association in the Perseus region). With this discovery came understanding of the great size range of these systems. The largest are the oldest, with maximum inferred ages of about 3×10^7 yr.

4.3.1 Location within the Galaxy

None of the massive stars in OB associations have optically visible pre-main-sequence contraction phases. In any given system, therefore, most of the luminous members lie on the main sequence, while a smaller fraction are supergiants caught in the act of leaving it. This fact allows determination of the distance to the association, through the technique known as *spectroscopic parallax.* The first step is to obtain, through analysis of absorption lines, the spectral types of as many member stars as possible. Photometry in the V-band then allows placement of the association in a diagram plotting m_V against spectral type. At this point, we again utilize the fact that the absolute magnitude M_V is a known function of spectral type along the main sequence (Table 1.1). The difference between M_V and m_V for each star then yields the distance via equation (2.12). Of course, this equation cannot be used without knowledge of the interstellar extinction A_V. Since the extinction is itself proportional to distance, one must obtain a self-consistent solution through trial and error.

For individual stars located outside associations, a variant of the method is still feasible. First, one uses spectroscopic analysis to verify that the object is on the main sequence. Photometry in two wavebands then establishes the apparent color. This may be compared with the intrinsic color index to yield the reddening and extinction, and hence the distance. Note that a basic assumption underlying spectroscopic parallax is that the standard relations hold between reddening and extinction, and between extinction and distance. The method is therefore unsuited for investigating clumpy clouds or those with anomalously large grains, although such local effects presumably fade in significance with greater distance. Indeed, many visible OB associations are too far away for accurate spectroscopy. In these cases, photometry in *three* wavebands suffices to place the member stars in a color-color diagram analogous to Figure 4.2. If the stars lie off the appropriate main-sequence curve, one may draw the reddening vector to read off the extinction and thus the distance. When applied to the $(U-B, B-V)$ diagram, this procedure is traditionally known as the *Q method*.[2]

Such techniques have facilitated the location of hundreds of O and B stars throughout the Galactic disk. Early efforts yielded the first convincing delineation of the local spiral arms, a discovery soon corroborated by researchers employing the 21 cm line of HI. OB associations trace the spiral structure as reliably as the interstellar gas because their constituent stars are too young to have moved far from their birthsites. Out of 200 O stars within 3 kpc of the Sun, some 75 percent are within associations. The figure falls to 50 percent for B0 – B2 stars, whose population is not as completely sampled. Modern study of OB associations has been greatly facilitated by the remarkable sensitivity of *charge-coupled detectors (CCDs).* In addition, near-infrared arrays and X-ray detectors have enabled us to probe the lower-mass component of these systems.

---

[2] Consider, for any observed star, the quantity $E_{\rm U-B}/E_{\rm B-V}$. The relation analogous to equation (4.1) together with the interstellar extinction curve (Figure 2.7) imply that this ratio has the numerical value 0.71, independently of the actual degree of reddening to the star or its spectral type. From the definition of the color excess (equation (2.14)), it follows that

$$(U-B) - 0.71\,(B-V) = (U-B)_\circ - 0.71\,(B-V)_\circ \equiv Q\,.$$

The quantity Q is thus also reddening-independent, but varies with spectral type. Note that $(U-B)_\circ$ and $(B-V)_\circ$ are functionally related along the main sequence. Knowledge of Q from the *apparent* magnitudes therefore also yields the star's *intrinsic* colors and hence the reddening.

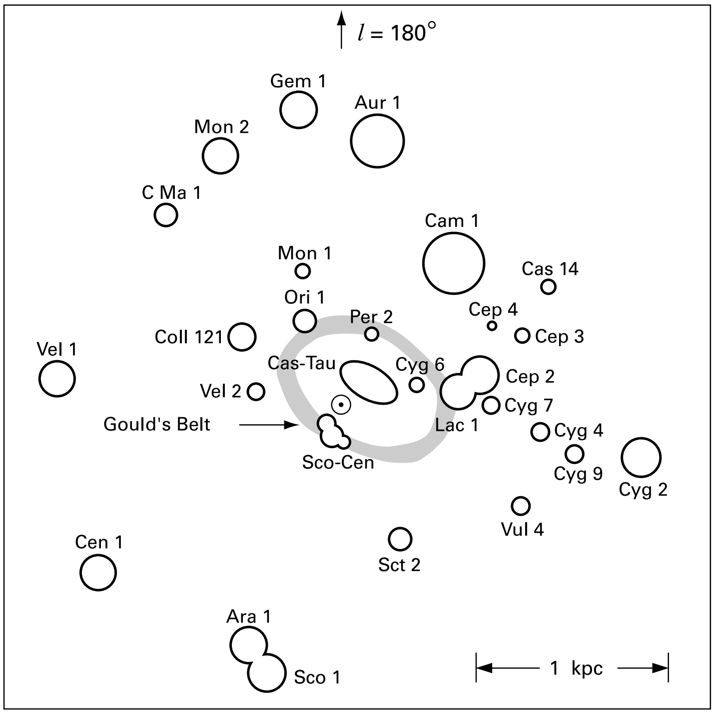


Figure 4.14 Galactic distribution of the closest OB associations. The Sun is indicated near the center of the frame. Sizes of the circles represent the physical dimension of each system and, in a few cases, the gross morphology. Note the ring of stellar and gaseous emission in Gould's Belt (*shaded region*), surrounding the solar symbol.

Before delving into the morphology of individual associations, let us first consider their spatial distribution. Figure 4.14 depicts all the systems within 1.5 kpc, as projected onto the Galactic plane. The reader may profitably compare Figure 3.2 for giant molecular clouds, which focuses on objects *inside* the Sun's galactocentric radius. Note that the nearest segment of the Perseus spiral arm, containing such associations as Cep OB1 and Per OB1, lies beyond the upper right border of Figure 4.14, between 2 and 3 kpc away. Note, too, we have already encountered many of these systems in other contexts. Thus, Mon OB1 is the relatively small association containing the cluster NGC 2264, and Mon OB2 encompasses NGC 2244 in the Rosette molecular cloud. Ori OB1 is the large association that includes the Trapezium, while Per OB2 surrounds the embedded cluster IC 348. These examples illustrate how expanding stellar systems may harbor smaller, interior clusters. They further remind us of the intimate relationship between OB associations and giant molecular clouds. Indeed, it is the rare association that is *not* in close proximity to some cloud complex.

Table 4.2 summarizes essential properties of the associations closer than about 600 pc, as measured by the Hipparcos satellite. Listed here are both a distance and an estimated physical diameter. The latter, indicated approximately in Figure 4.14, was obtained using the average angular extent of each system together with its distance. The table also gives separately the

Table 4.2 The Nearest OB Associations

Name	Distance (pc)	Diameter (pc)	O stars	B Stars
Lower Centaurus-Crux	120	50	0	42
Upper Centaurus-Lupus	140	75	0	66
Upper Scorpius	150	30	0	49
α Persei (Perseus 3)	180	10	0	30
Cassiopeia-Taurus	210	200	0	83
Cepheus 6	270	40	0	6
Perseus 2	320	50	0	17
Trumpler 10	360	45	0	22
Lacerta 1	370	60	1	35
Vela 2	410	75	1	81
Orion1	470	75	9	327
Collinder 121	590	120	2	85
Cepheus 2	610	110	1	53

numbers of O and B-star members, as established through both spectroscopy and proper motion. Notice that over half the systems have no O stars, which are rare objects indeed. Note finally that the relatively small association designated Trumpler 10 is not depicted in the spatial map, as it lies in front of the larger Vela OB2.

Returning to Figure 4.14, the shaded, open structure surrounding the solar position is *Gould's Belt*. This huge ring of bright stars and gas, up to 700 pc in diameter, links a number of the closest associations. Near its center lies the Cassiopeia-Taurus (Cas-Tau) system. Containing no stars brighter than $M_V = -5$ and extending over some 200 pc, this diffuse grouping stands out from the background field only by virtue of the similar and parallel proper motions of its members. The entire Taurus-Auriga cloud complex lies within its borders, as does the small and possibly bound system α Persei listed in Table 4.2. The Cas-Tau association appears to represent the largely dispersed remnant from an earlier epoch of massive star formation. Other systems in such an advanced state of disintegration must exist throughout the Galaxy, but are currently impossible to find outside the solar neighborhood.

4.3.2 Expansion

The more compact associations strung out along Gould's Belt have velocities indicating a general expansion from the Cas-Tau region. The best-studied such system is that of Scorpius-Centaurus (Sco-Cen). With a maximum size that rivals Cas-Tau, this association consists of a sequence of three, spatially discrete subgroups (Figure 4.15). At one end lie the embedded stars of the ρ Ophiuchi molecular clouds. The Lupus T association and its molecular complex are just inside the border of the middle (Upper Centaurus-Lupus) subgroup, as shown in the figure. Neither cloud region is forming O stars, but the ρ Ophiuchi complex in particular has clearly been disturbed by such activity nearby. In Figure 3.17, the change in both the cloud morphology and the pattern of polarization vectors near L1688 suggest compression from the Upper Scorpius subgroup to its right. This impression is reinforced by 21 cm data revealing a

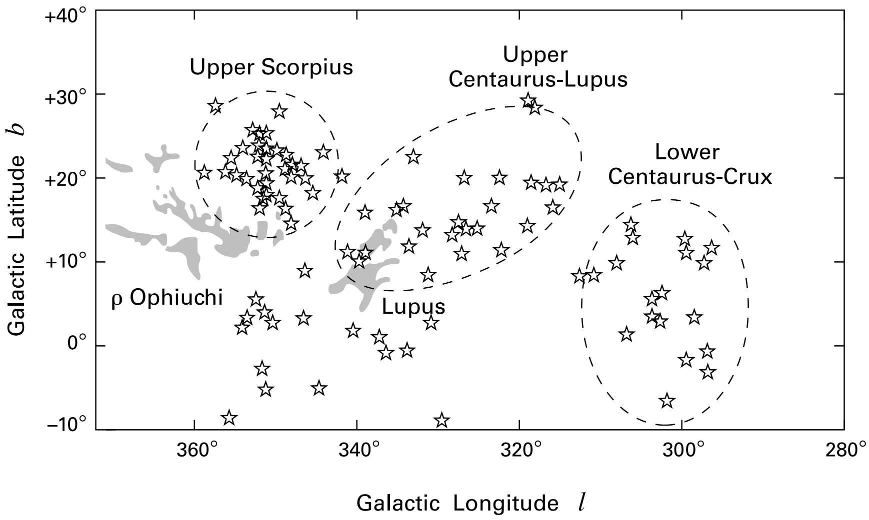


Figure 4.15 Subgroups within the Sco-Cen OB association. Shown are the prominent stars, as well as the molecular clouds in ρ Ophiuchi and Lupus.

shell of atomic hydrogen centered on the massive stars in Upper Scorpius and impinging on the ρ Ophiuchi clouds.

From the dashed boundaries shown in Figure 4.15, it is apparent that the three optically visible subgroups within Sco-Cen have differing sizes and hence ages. That is, these groups are separated both spatially and temporally. It is natural to suppose that all three originated in a giant molecular complex, of which the ρ Ophiuchi and Lupus clouds are the sole remains. The pattern of subgroup ages then corresponds to the order in which various high-density regions of the parent complex underwent gravitational collapse. Thus, the first subgroup to form massive stars was Upper Centaurus-Lupus, followed by Lower Centaurus-Crux, and then Upper Scorpius.

To quantify matters and obtain actual ages, one may utilize the expansion velocities of individual stars. Tracing their motion backwards in time leads to a unique configuration for which the stellar density is highest. The corresponding time then gives the age of the subgroup in question. In practice, both radial velocities and proper motions of the expected magnitude (a few km s^{-1}) are difficult to obtain. Within Sco-Cen, accurate proper motions are available in Upper Scorpius, while radial velocities here are too small for detection. Figure 4.16 depicts the proper motion vectors, as well as the inferred initial configuration, which has an associated age of 4×10^6 yr. The longest dimension of this configuration is about 45 pc, in good agreement with present-day giant cloud complexes. Note that all the velocities shown are those relative to the mean proper motion of the subgroup; the latter reflects the global expansion of Gould's Belt mentioned previously.

4.3.3 Main-Sequence Turnoff

An independent check on this kinematic method comes from another clock– the HR diagram. As was the case for the low-mass T associations, the distribution of stars in the diagram constitutes a record of star formation history, but now supplies complementary information. Fig-

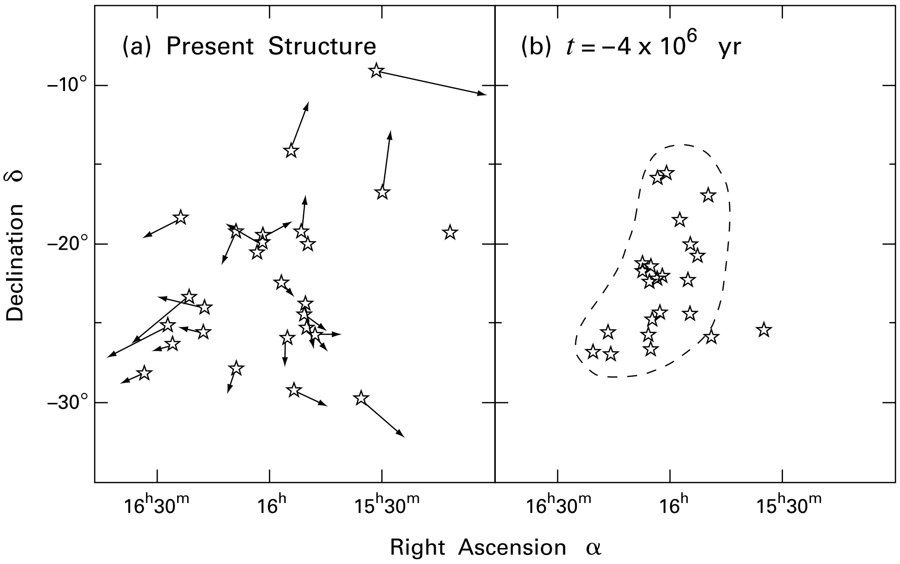


Figure 4.16 Reconstructing the initial configuration of the Upper Scorpius subgroup. (a) Proper motion vectors of prominent stars. (b) Most compact structure leading to the present configuration.

ure 4.17 displays the HR diagram for the Upper Scorpius subgroup. While the intermediate-mass stars fall along the main sequence, higher-mass members begin to deviate from it, and the most massive stars are absent altogether. This *main-sequence turnoff* reflects the age of the system. The deviation occurs at a mass of about 30 $M_\odot$, which translates into a main-sequence lifetime of 5×10^6 yr. The Upper Scorpius region must have begun producing stars *at least* that far back in the past. Stars of significantly greater mass formed then or earlier would have finished burning hydrogen by now and migrated out of the diagram, thus accounting for the present truncation of the main sequence. As illustrated in Figure 4.17, we may conveniently read off the system age (in the foregoing sense) by matching the empirical turnoff with the set of theoretical post-main-sequence isochrones, *i. e.*, the loci of constant evolutionary time for stars of various mass. Here, $t = 0$ corresponds to the initiation of hydrogen fusion on the ZAMS.

These considerations will naturally remind the reader of our previous discussion of the main-sequence turnon in T associations. Both features of the HR diagram are age indicators, but their conceptual difference is noteworthy. The turnon point singles out the *oldest* pre-main-sequence stars in the association, *i. e.*, it indicates when the formation of relatively low-mass objects began. Conversely, the turnoff identifies the *youngest* post-main-sequence members and thus tells us when the last high-mass stars were born. In a region currently devoid of molecular gas, this latter time marks the end of the star formation process. Quite generally, the turnon "age" should always exceed that given by the turnoff, with their difference being a measure of the total duration of star formation activity.

Because of the statistics of stellar masses, not all forming groups exhibit both turnon and turnoff points. Pure T associations like Lupus or Taurus-Auriga simply lack the high-mass component that would include a turnoff. With regard to OB associations, however, numerous surveys support the view that regions harboring massive stars invariably contain many more

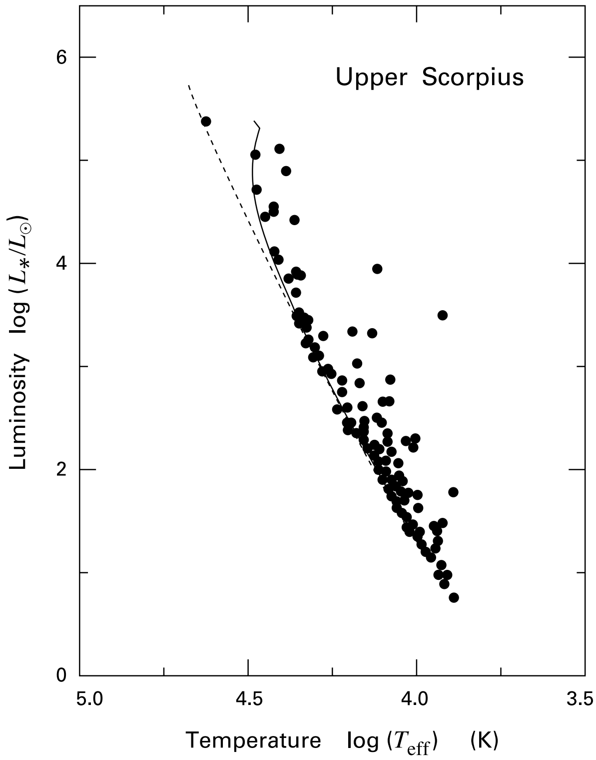


Figure 4.17 Main-sequence turnoff in Upper Scorpius. The dashed curve represents the ZAMS, while the solid curve is the 5×10^6 yr isochrone.

of lower mass. Within Upper Scorpius, X-ray observations by the Einstein satellite turned up dozens of previously unknown members. Subsequent photometry and spectroscopy showed that most of these are weak-lined T Tauri stars. Deeper surveys by the ROSAT and Chandra X-ray satellites have uncovered even more sources, which by now outnumber the massive stars shown in Figure 4.17.

4.3.4 The Orion Association

Let us apply these ideas to the best known of all OB associations, that in Orion. Figure 4.18 shows the familiar outline of CO emission from the giant molecular cloud, along with the approximate boundaries of the four identified subgroups. That labeled 1c largely coincides with the Ori R1 association, while the small 1d subgroup is the region of radius 2.5 pc that includes the even more compact Trapezium cluster.

The spatial pattern of all the subgroups again suggests vividly the progression of massive star formation and further demonstrates how this process serves to clear out molecular gas. Thus, the oldest and largest 1a subgroup lies in an area currently free of CO emission. With a little imagination, one can picture how the Orion A cloud once extended northward into this region. The somewhat younger 1b system still partially encompasses dense gas, while the smallest 1c and 1d groups are wholly embedded within Orion A. This temporal ordering is confirmed by the stellar distribution. For example, the most luminous star in 1b is the supergiant ζ Ori, with a

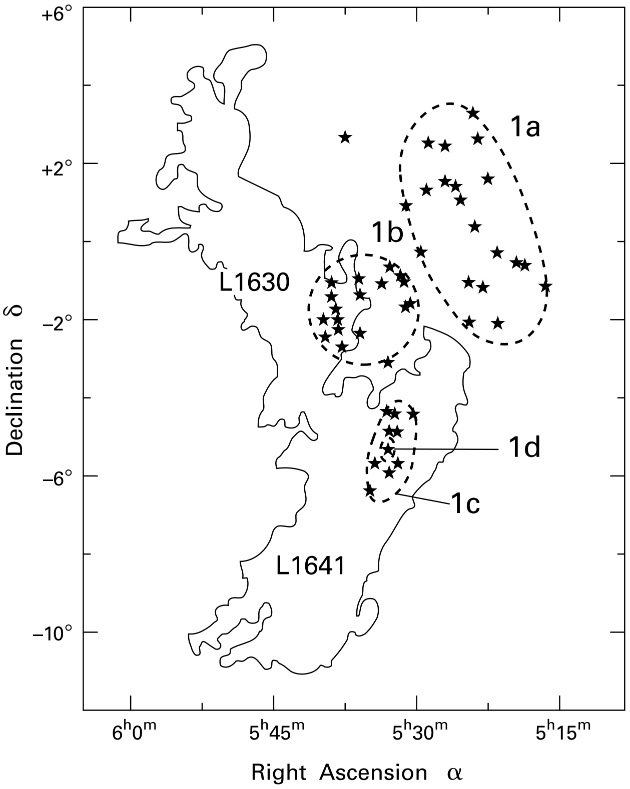


Figure 4.18 Subgroups in the Orion OB association. The outline of CO emission is also shown.

mass of 49 $M_\odot$ and a main-sequence lifetime of 4×10^6 yr. The corresponding member of 1a is η Ori, a B1 main-sequence star of 16 $M_\odot$ with a lifetime of 1.4×10^7 yr.

We saw in Chapter 1 how near-infrared surveys have revealed many young, low-mass stars in the Orion B cloud. Similarly, in the 1d subgroup, a large, low-mass population is manifest both in the optical and infrared. A similar distribution surely holds in Orion 1a, where a turnon should exist somewhat below 1.5 $M_\odot$, the mass with a pre-main-sequence contraction time of 1.4×10^7 yr. Picking out F and G stars from the plethora of background sources over such a wide area is a challenging task.

Returning, then, to Figure 4.18, we can now appreciate how the OB association represents but one aspect (albeit the most conspicuous one) of much more extensive formation activity. Failure to recognize this fact has misled some into positing a *causal* link in subgroup formation. The idea is that the creation of O and B stars in one locale somehow induces collapse in a neighboring region, leading to a kind of OB chain reaction. However, while there is ample evidence that massive stars can terminate formation activity over a substantial volume, there is little to suggest that they also initiate it (except on a restricted spatial scale; see Chapter 15.) The relative ages and locations of OB subgroups are certainly of interest, but the true global pattern of stellar birth can only be discerned through multiple observations. A case in point is the L1641 region of Orion A, also depicted in Figure 4.18. Here, X-ray and infrared studies have uncovered a distributed population consisting of hundreds of low-mass stars. The oldest are weak-lined T Tauri's, with ages that rival high-mass members of the 1a subgroup, while the youngest are embedded infrared sources or classical T Tauri stars near the birthline.

The physical picture emerging is that the Orion Molecular Cloud has been gravitationally settling over some period exceeding 10^7 yr. This contraction has proceeded locally at different rates and with diverse outcomes. It apparently began within the present 1a subgroup, at an epoch which can best be measured once a turnon in low-mass stars is observed. Eventually, enough massive stars formed here to disperse the surrounding gas. Sometime later, the L1641 region also condensed to the point of star formation, but never attained the compactness necessary for massive stars. The 1c and 1d regions followed suit, and intense formation activity continues today in both Orion A and B. The details in this highly incomplete picture will undoubtedly change as future studies focus increasingly on the low-mass stellar component. What seems secure, both in Orion and elsewhere, is that star formation in any particular region can occur without an external trigger, purely through the gravitational contraction of a large cloud region. We shall elaborate this key idea in subsequent chapters.

4.3.5 Embedded and Runaway Stars

We have been focusing on massive stars that are optically visible. These have either moved away from, or else destroyed, the gas and dust in their immediate vicinity within the last few million years. The HII regions they excite are extended structures, with typical diameters of 10^{18} cm (*i. e.*, 0.3 pc). Even younger O and early-B stars exist, for which the enshrouding matter completely absorbs all ultraviolet and visible photons. Constituting some 10 percent of the massive star population in the solar neighborhood, these objects are detectable through their reemission of stellar photons at radio and infrared wavelengths. Such *ultracompact HII regions*, roughly 10^{17} cm in size, are among the most luminous Galactic objects in the far infrared. The powerful radio source W49 in Aquila contains at least seven of these regions crowded into an area only 0.8 pc in diameter. Dense systems like this one could represent the ancestors of Trapezium-like clusters within visible OB associations and probably contain numerous low-mass members that are currently beyond detection.

Figure 4.19 shows the distribution of ultracompact HII regions in the Galactic plane, as revealed through IRAS observations in the far infrared. There is nearly a perfect match with the corresponding distribution of giant molecular clouds, delineated by the CO contours in the same figure. This agreement underscores the extreme youth of the deeply embedded stars.

At the other extreme are massive objects with little or no associated interstellar matter. As we mentioned, about 25 percent of O stars do not appear to be members of clusters or associations. These field objects tend to be farther from the Galactic midplane than their counterparts within groups. Their radial velocities also exhibit more dispersion about the local value expected from Galactic rotation alone. Statistical analysis of the velocities reveals that most objects are leaving the plane, rather than entering it from above and below. Thus, the stars were likely born in ordinary associations, but with speeds that were higher than average.

A large fraction of the field objects are *runaway OB stars*. These have exceptionally high spatial velocities, typically from 50 to 150 km s^{-1}, and are sometimes located high above the Galactic plane. In one sense, the origin of runaways is not a mystery, since their proper motions can often be traced back to a known OB association. The real problem is their velocities, which indicate that the objects were once subject to strong forces. Thus, each runaway might originally

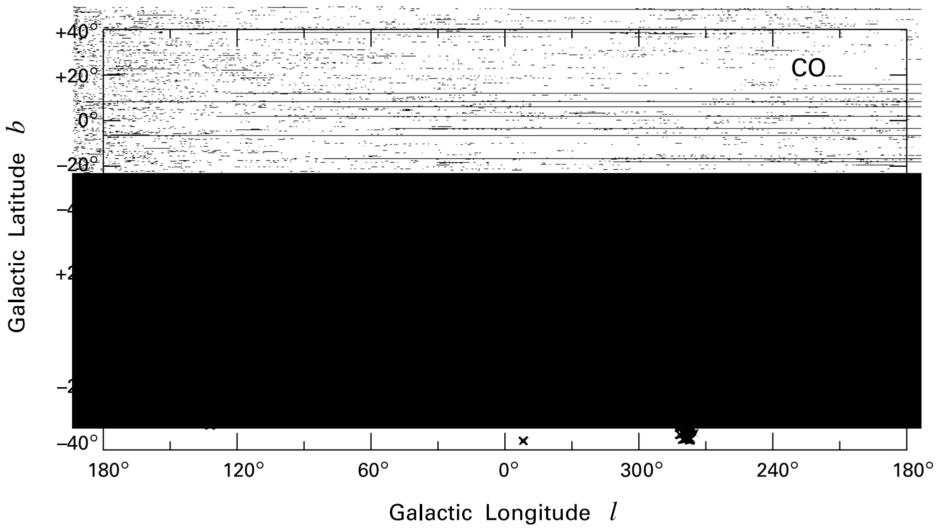


Figure 4.19 *upper panel*: Emission in ^{12}C^{16}O within the Galactic plane. *lower panel*: Distribution of far-infrared point sources, each of which corresponds to an ultracompact HII region.

have been a member of a close binary pair consisting of two massive stars. If the companion exploded as a supernova, the star of interest could escape with a speed equal to the orbital value. Alternatively, the star may have been launched after a close encounter with other stars inside a dense cluster, like those often found at the centers of OB associations.

Neither hypothesis is free of difficulty. Calculations of binary evolution show that the initially more massive star (the *primary*) transfers a great deal of mass to the secondary before the primary becomes a supernova. The explosion consequently ejects too little mass to unbind the system. The binary thereafter consists of the now-massive secondary along with a neutron star or black hole. Such, at least, is the prediction of theory. In fact, no such pair has ever been seen. Almost all observed runaway OB stars are single, in contrast to the general stellar population (Chapter 12).

A simple modification of this picture may resolve the difficulty. Suppose the supernova explosion were *anisotropic*. Then the strong recoil force acting on the compact object would easily free it from its companion star, despite the relatively small ejected mass. Observations show, in fact, that the youngest neutron stars (which happen to be radio pulsars) have very high velocities, often 500 km s^{-1} or even more. Speeds of this magnitude could plausibly result from an asymmetric supernova. Unfortunately, the cause of the putative anisotropy in the explosion is not yet understood.

As to the cluster hypothesis, numerical simulations have indeed produced single, high-speed stars through encounters between *two* binaries. The usual outcome is for the *least* massive of the four stars to be ejected. Low-mass runaways, however, are apparently very rare. For this hypothesis to survive, the cluster must have an unusual mass distribution by normal Galactic standards, one that is very strongly skewed toward the upper end. Alternatively, the binaries might consist of four intermediate-mass objects, two of which physically merge during the encounter. The numerical studies indicate that such collisions indeed occur, although the escaping, merged star naturally has a lower speed than would a single, runaway object. The reader should keep these possibilities in mind as we periodically return to explore the special problems connected with massive star formation.

4.4 Open Clusters

Our final category of young systems includes those most easily recognized by the unaided eye. These are called *open clusters* because their individual members can often be clearly distinguished. In contrast, only the largest telescopes can resolve the central regions of the much more massive and dense globular clusters, very old systems residing far outside the plane of the Milky Way. Open (or "galactic") clusters are the most evolved aggregates we have discussed. Half of them reach ages of 1×10^8 yr, while about 10 percent survive as long as 1×10^9 yr. Even at 10^8 yr, however, all stars of less than 0.5 $M_\odot$ are still in the pre-main-sequence phase, so most of these systems are young enough for our purposes.

4.4.1 Basic Properties

The largest open clusters are those near the centers of the more compact OB associations. Examples include NGC 2244, the group of diameter 11 pc within the Mon OB2 association, and IC 1396, a 12 pc cluster located inside Cep OB2 (see Figure 4.14). Despite appearances, systems of this type are rarely gravitationally bound, as few remain within older and larger associations. Most other open clusters are indeed bound and have diameters ranging from 2 to 10 pc, with a median value close to 4 pc. The total number of members in any one is always difficult to assess, but the population seen in photographs can be anywhere from 10 to 10^3. Within the Galaxy, open clusters concentrate strongly toward the plane; the scale height of the distribution in the solar neighborhood is 65 pc. There are currently over 1200 systems known, almost all of them less than 6 kpc from the Sun. The sample is essentially complete within 2 kpc, while dust extinction increasingly hampers visual observations at greater distances.

Table 4.3 lists all known open clusters within 300 pc. We shall discuss shortly how the actual distances are established. The diameters were obtained by eye from photographic plates, after applying the distances. While clearly subject to some uncertainty, these figures should represent the volume within which the majority of component stars reside. The tabulated ages follow from analyzing the main-sequence turnoff in the HR diagram. Finally, we note that the membership figures given in the last column represent true counts (or estimates) of the total number of objects for which there is optical photometry in the U-, B-, and V-bands.

Open clusters contain little molecular gas and are not currently forming protostars. They therefore represent groups whose members, all located at nearly the same distance, share a common age and chemical composition. These characteristics have made them invaluable research tools for over a century. In 1930, Trumpler first demonstrated the phenomenon of interstellar absorption by showing that clusters of smaller apparent diameter, which are more distant on average, are also systematically dimmer, above and beyond the usual inverse square falloff in flux. Over the next 30 years, it was the intercomparison of open clusters of different ages that provided the empirical foundation for the developing theories of pre- and post-main-sequence evolution. More recently, studies of stellar surface activity (such as X-ray emission), rotation, and the depletion of light elements have also relied heavily on these systems.

Table 4.3 The Nearest Open Clusters

Name	Distance (pc)	Diameter (pc)	Age (10^6 yr)	Members
Ursa Major	25	0.9	300	25
Hyades	41	4.3	630	550
Coma Berenices	96	2.7	450	45
Melotte 227	120	2.4	370	25
Pleiades	130	5.2	130	800
IC 2602	160	4.7	32	120
IC 2391	175	3.1	46	80
α Persei	185	16	72	380
Praesepe	190	3.8	730	500
Collinder 359	250	17	32	13
Blanco 1	270	5.5	63	190
NGC 6475	300	7.0	250	120
NGC 2451	300	4.4	45	180

We have already mentioned the indirect means that must be employed to obtain distances to OB associations. In contrast, the familiar open cluster of the Hyades is one of the few stellar groups close enough that its distance follows from velocity measurements alone. The proper motion vectors of the stars all converge to a single point, indicating that the cluster is receding. If θ is the system's present angular diameter, the proper motions yield $\dot{\theta}$, the rate at which this diameter is shrinking. For small θ, the fractional rate of shrinking, $\dot{\theta}/\theta$, is also equal to V_r/d. Here, V_r is the common radial velocity of the stars at distance d. Once the velocity is determined from Doppler shifts of the spectral lines, this *moving-cluster method* directly gives d, which is 41 pc in the present case.

The actual value of the Hyades distance is not as important as the role of the moving-cluster method in astronomical calibration. Within the Hyades, knowing the cluster distance allows one to assign luminosities to all members with measured apparent magnitudes. Since many of these stars are on the main sequence, this assignment, together with spectroscopic temperature determinations, establishes the ZAMS empirically over a finite range of L_* and T_{eff}. Turning to other open clusters, one can now apply the technique of spectroscopic parallax described in § 4.3. That is, we vertically slide each diagram of m_V versus spectral type until it matches the Hyades, thereby both establishing the cluster distance and completing the calibration of the main sequence itself. Such *main-sequence fitting* is the basis of the ZAMS represented by the theoretical curves in Figures 1.15 and 1.18. The distances provided by the moving-cluster method and spectroscopic parallax form the lowest rungs of the *cosmic distance ladder*. To go beyond our Galaxy, we must utilize other techniques – including observations of pulsating stars and supernovae – to bootstrap our way outward. At each step, however, the most reliable measurements are relative ones, so that even the greatest cosmological distances rest ultimately on those few established kinematically for nearby open clusters.

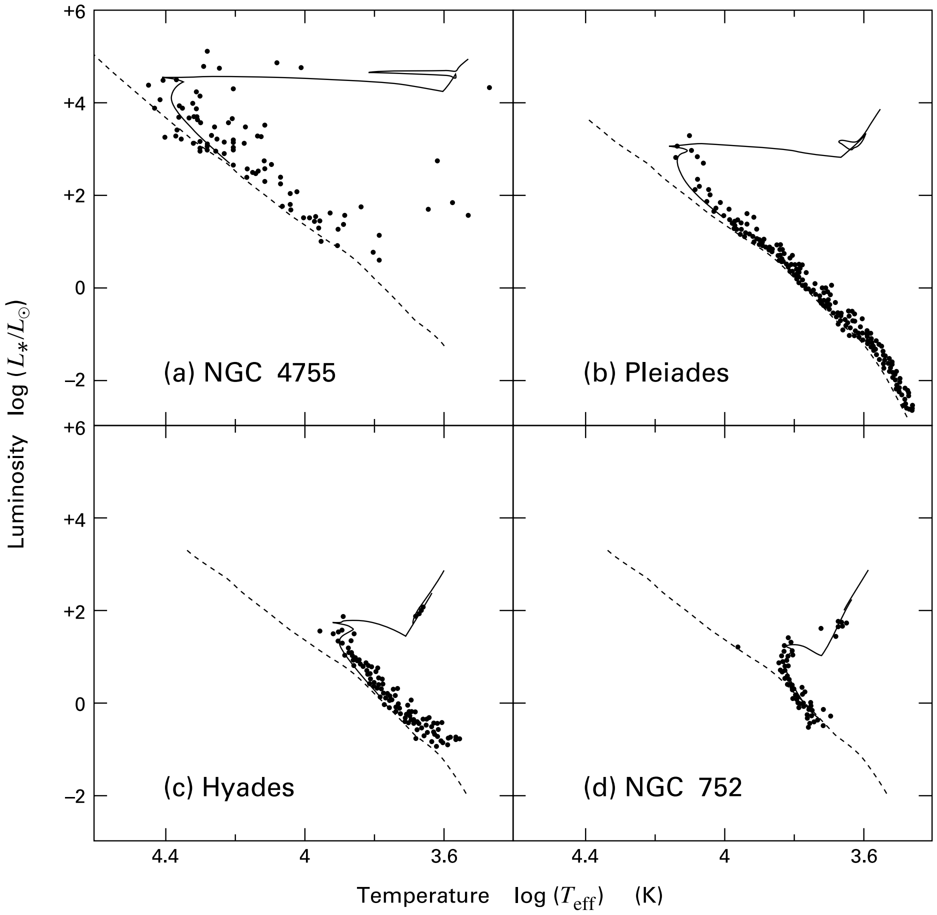


Figure 4.20 HR diagrams of four open clusters, arranged by age. For each system, both the ZAMS (*dashed curve*) and the best-fit isochrone (*solid curve*) are also displayed.

4.4.2 Evolution in the HR Diagram

As before, the HR diagram is a powerful tool for gauging the evolutionary status of any observed system. The vexing incompleteness problem that plagues more embedded T and OB associations is here greatly diminished. In addition, the age span of open clusters is such that both main-sequence turnoffs and turnons may be observed, sometimes within a single cluster. Figure 4.20 is a composite of four diagrams, in order of increasing age. The values of L_* and $T_{\rm eff}$ were derived in all cases from photometric observations at visual and near-infrared wavelengths, after applying a global extinction correction for each cluster. In addition to the ZAMS, the figure also includes the theoretical post-main-sequence isochrones that best fit the high-mass turnoff in each case.

Our youngest example (Figure 4.20a) is NGC 4755, or "Herschel's Jewel Box," a rich system of several hundred members. Located in the Southern Crux constellation, its distance of 2.1 kpc is too great for adequate study of the fainter objects, which undoubtedly contain an admixture of interlopers from the field. Even within the brighter population, the HR diagram shows considerable scatter, most of which stems from patchy extinction contaminating the luminosity estimates. Nevertheless, the stellar distribution displays, in addition to the turnoff, a clear departure from the ZAMS at low masses, below about $\log T_{\text{eff}} = 3.9$. From Table 1.1, this temperature corresponds to a mass of $2\,M_\odot$. Such a star has a pre-main-sequence contraction time of 8×10^6 yr. The post-main-sequence isochrone in the figure has the similar associated age of 1×10^7 yr.

Systems this young containing massive stars are not uncommon and some may actually be OB associations rather than open clusters. Most often, classification is a matter of historical accident. The difference, however, is a true physical one, since it involves the eventual fate of the system. Will it quickly disperse, or will it remain gravitationally bound for an extended period? In principle, the answer may be obtained empirically, through accurate measurement of the stars' spatial velocities.

The advantages of proximity are evident for the Pleiades (Figure 4.20b), which lies only 130 pc away. Here, the scatter in the HR diagram is much less than for NGC 4755, and the low-mass portion of the 800 or so known members is better sampled. The very brightest of these, familiar in the Northern sky as the Seven Sisters, are part of a central core of stars within an extended halo, some 4° (*i. e.*, 10 pc) in radius. The haze seen in optical photographs attests to the presence of interstellar matter, but the extinction is modest, with $A_V = 0.12$ mag. In the HR diagram, the turnoff from the main sequence is clear. The displayed isochrone corresponds to an age of 1×10^8 yr. The main-sequence turnon is less apparent, but a careful examination confirms its presence near $L_* = 0.1\,L_\odot$. As in our previous example, there is rough agreement between the turnon and turnoff ages, but the measurements are still too inexact to warrant further assessment of their difference.

One difficulty in obtaining more precise turnon ages is that even the *empirical* ZAMS is not known to great accuracy at the lowest luminosities. As we have seen, a crucial building block in this enterprise is the Hyades, whose diagram we display as Figure 4.20c. Spectroscopic analysis reveals that the metallicity in the Hyades is higher than that of other nearby clusters by a factor of about 1.5. This difference is enough to shift the Hyades main sequence slightly toward lower temperatures, and a proper compensation is necessary in constructing the fiducial ZAMS. As for the evolutionary status of the cluster itself, its nuclear age of 6×10^8 yr is relatively secure, and implies that the main sequence is populated only up to $2\,M_\odot$. Correspondingly, the turnon is now lowered to about $0.1\,M_\odot$, or a luminosity of $1 \times 10^{-3}\,L_\odot$. This point lies well below the observational cutoff present in the Figure.

Figure 4.20d depicts NGC 752, one of a handful of open clusters significantly older than the Hyades. At a distance of 400 pc, this sparsely populated system has fewer than a hundred observed members. Its advanced evolutionary state is apparent from the absence of all high- and intermediate-mass stars. At the turnoff age, estimated to be 2×10^9 yr, stars of $1.5\,M_\odot$ are just completing main-sequence hydrogen fusion. There are undoubtedly additional cluster members below the $0.8\,M_\odot$ minimum mass shown here. Only those of less than $0.09\,M_\odot$, however, are still in their pre-main-sequence phase.

An interesting feature in the HR diagram of NGC 752 is the clump of stars above and to the right of the main sequence. An analogous, but smaller, group is also visible in the Hyades. Evolutionary calculations show these stars to be red giants undergoing core helium burning. Finally, we see that the diagram again displays considerable scatter about the ZAMS, despite the fact that the cluster is at a high Galactic latitude and suffers little extinction. One plausible source for this scatter is the presence of unresolved binaries, which can raise the apparent luminosity if their mass ratios are close to unity.

4.4.3 Mass Segregation

Returning to the Pleiades, the central concentration of its brightest and most massive members is a phenomenon we have encountered before. We recall the deeply embedded clusters of L1630 in Orion, with their luminous cores of O and B stars (Figure 1.2), or the buildup of stars surrounding the massive object in S106 (Plate 2). Within the more exposed NGC 2264 cluster, careful mapping of the stellar density reveals *two* concentrations – one surrounding S Mon and another associated with a star that is again the most massive in its local region. In principle, more refined observations of mass segregation should be possible for open clusters, but less than two dozen systems have so far been examined in sufficient detail. For most of these, the average stellar mass drops steadily from the center outwards.

Since mass segregation is present to some extent in the very youngest systems, it is evidently part of the star formation process itself. Thus, the salient question, which we shall explore later, is *not* how the most massive objects find their way to the densest regions, but why they form there in the first place. Having said this, it is also true that open clusters are old enough that the process of *dynamical relaxation* can further promote the settling of massive stars toward their centers.

To understand dynamical relaxation, consider a hypothetical cluster of 1000 stars, with a total mass of 500 $M_\odot$ and diameter of 5 pc. The typical velocity of a cluster member is given by the virial value in equation (3.20), and is about 1 km s^{-1}, if we assume that no gas remains in the system. The *crossing time* over which the star can traverse most of the cluster is therefore 5×10^6 yr. During each such passage, the stellar orbit is determined mainly by the smoothly varying gravitational force arising from the system as a whole. However, each interaction with an individual field star produces an additional tug, and many such tugs change the orbit completely. The system gradually relaxes toward a state independent of initial conditions, one in which the total available energy is apportioned roughly equally among the members.

Under the new conditions, the least massive stars have the highest velocities and therefore fill out the largest volume. Conversely, the high-mass members tend to crowd toward the middle. For our sample cluster, theory predicts that such a state prevails within about 15 crossing times, where the precise figure depends on the stellar mass spectrum. Thus, the *relaxation time* is roughly 7×10^7 yr, too long for embedded systems but within the range for open clusters. One might expect, from this argument, that older clusters would exhibit a steeper outward falloff in the average mass of their members, but no such effect is evident in the data at hand.

4.4.4 Destruction by Giant Clouds

The centrally peaked appearance of open clusters is strong, though not conclusive, evidence that they are gravitationally bound. This is not to say they they remain intact for all time. Dynamical relaxation gradually inflates a halo of lighter stars, some of which actually escape. Such "evaporation," however, typically requires 100 crossing times to deplete an isolated system. Why, then, do so few observed clusters survive to even 10^9 yr? Clearly, some external process is at work that destroys them more efficiently.

There is little direct observational evidence bearing on this question, but theory suggests that the main culprit is encounters with giant molecular clouds. The rate of such encounters is low– about one for each rotation of the cluster about the Galaxy– but the cloud mass is so huge that the effect can be devastating. Both molecular clouds and clusters have similar random motions within the Galactic disk. Their typical relative velocity exceeds a cluster's *internal* velocity dispersion by about an order of magnitude. During an encounter, the cloud effectively imparts a brief impulse to each star, in a manner somewhat akin to dynamical relaxation. In this case, however, there is a net energy gain by the cluster as a whole.

The additional energy arises from the tidal component of the gravitational interaction. Stars that are closest to the passing cloud respond most strongly, causing the stellar system to stretch along the line joining the centers of mass. Incidentally, the same effect, but arising from the general Galactic field, strips stars from the cluster halos and truncates their radii to about 10 pc in the solar neighborhood. Often only a single encounter with a giant molecular cloud is sufficient to disrupt a cluster entirely. If not, the cumulative tidal stretching from several such encounters does the job. It is ironic, then, that the very structures giving rise to all young clusters appear responsible for their ultimate demise.

4.5 The Initial Mass Function

Any attempt to understand the origin of stellar groups must address the issue of their internal mass distribution. It is not obvious, of course, that any single function will adequately describe all existing systems. In principle, the natural variation in such environmental factors as the ambient magnetic field or the molecular cloud temperature prior to cluster formation could yield a wide variety of distributions. However, we have already seen from numerous examples that massive stars are intrinsically rarer than their low-mass counterparts. We now seek to quantify this notion. As a practical matter, the masses of embedded stars are difficult to obtain empirically, so we first look to *field* stars in the solar neighborhood. We will then show that the mass distribution found here also appears to hold, at least approximately, for discrete clusters and associations. This important finding bolsters the view that all stars are born within such groups.

4.5.1 Luminosities Past and Present

Even for an unobscured field star at a known distance, it is the luminosity within a certain wavelength range, rather than the mass, that is directly observable. A fundamental statistical property of field stars is thus the *general luminosity function*, $\Phi\,(M_V)$. This function is defined so that $\Phi\,(M_V)\,\Delta M_V$ is the number of stars per cubic parsec in the solar neighborhood with absolute visual magnitude between $M_V - \Delta M_V/2$ and $M_V + \Delta M_V/2$. Obtaining the general luminosity function is no trivial matter. One must derive distances to large numbers of stars and make proper extrapolations for the even larger numbers whose distances are unavailable directly. Other, more subtle complications abound. For example, the Galactic scale heights of stars vary inversely with mass, the brightest stars hovering close to the midplane during their relatively short lifetimes. These same stars can be seen out to distances much greater than their scale heights. Thus, they appear to occupy flattened disks, whose volumes must be accurately assessed when obtaining densities. Beginning with the pioneering efforts of P. J. van Rhijn in

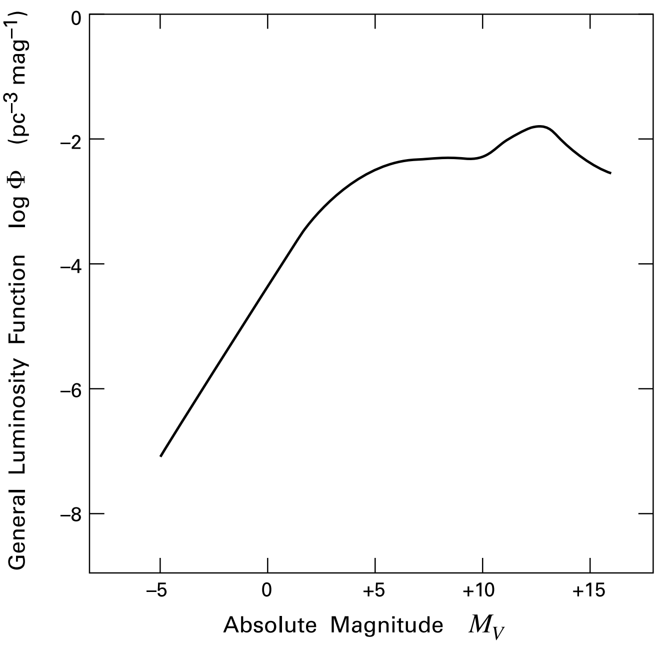


Figure 4.21 General luminosity function for stars in the solar neighborhood.

the 1920s, such difficulties have gradually been overcome, and a modern result is displayed as the solid curve in Figure 4.21.

Over most of the magnitude range shown, the general luminosity function rises with increasing M_V, *i. e.*, there are more dim stars than bright ones in any fixed magnitude interval. This trend continues up to $M_V \approx +12$, after which there is a steady decline. The details of this slow falloff remain somewhat uncertain, since many of the observed "stars" in this regime are actually binaries that have not been spatially resolved. On the other hand, there is no doubt that the slope of $\Phi(M_V)$ is considerably steeper for the most luminous stars. This initial sharp rise levels off rather abruptly at about $M_V^* = +5$. Reference to Table 1.1 reveals that a star at this transitional magnitude has a main-sequence lifetime $t_{\rm MS}$ slightly greater than 10^{10} yr. This time is close to the age of the Galactic disk, currently estimated at $t_{\rm gal} = 1 \times 10^{10}$ yr. If we recall that any star's post-main-sequence lifetime is brief compared to $t_{\rm MS}$, the origin of the change in slope becomes clear. Relatively dim, low-mass stars with $M_V \gtrsim M_V^*$ have been accumulating steadily over the lifetime of the Galaxy, while only a fraction of the brighter, short-lived stars with $M_V \lesssim M_V^*$ have survived.

It is apparent, then, that $\Phi(M_V)$ itself does not accurately reflect the relative production rate of stars with various M_V-values. However, the foregoing argument can readily be quantified to make the necessary modification. Following the notation of Chapter 1, let $\dot{m}_*(t)$ be the total Galactic star formation rate per square parsec near the solar position. Notice that we use the rate integrated over the disk thickness, in order to account for the diffusion of stars from the midplane during their evolution. To a fair approximation, in fact, the stellar *volume* density falls off exponentially away from the plane, with a scale height H that is a function of M_V. We further define the *initial luminosity function* $\Psi(M_V)$ to be the relative frequency with which stars of a given M_V first appear. This function is normalized to unity: $\int \Psi(M_V)\, dM_V = 1$. Resetting M_V^* to that magnitude for which $t_{\rm MS} = t_{\rm gal}$ precisely, we can write the general luminosity function as an integral over time, where the integration limits depend on M_V:

$$\Phi(M_V) = \begin{cases} \int_{t_{\rm gal}-t_{\rm MS}}^{t_{\rm gal}} dt\, \dot{m}_*(t)\, \Psi(M_V)\, [2\, H(M_V)]^{-1} & \text{if } M_V < M_V^* \\ \int_0^{t_{\rm gal}} dt\, \dot{m}_*(t)\, \Psi(M_V)\, [2\, H(M_V)]^{-1} & \text{if } M_V \geq M_V^* \ . \end{cases} \tag{4.4}$$

In writing equation (4.4), we have ignored any possible time-dependence in $\Psi(M_V)$ or $H(M_V)$. Furthermore, we are actually interested in the appearance of *main-sequence* stars with various magnitudes, while $\Phi(M_V)$ encompasses bright field stars that are giants and supergiants. Accordingly, the luminosity function on the left side of equation (4.4) must be diminished at the lowest values of M_V, corresponding to the brightest stars.

After making this correction, knowledge of the main-sequence lifetimes $t_{\rm MS}(M_V)$ allows us to invert equation (4.4) and obtain $\Psi(M_V)$, *provided* we know the rate $\dot{m}_*(t)$. Neither theory nor observation is of much help in this regard, beyond the general statement that $\dot{m}_*(t)$ should diminish with time. Fortunately, the final result is rather insensitive to the prescription adopted here, so we follow the standard expedient of ignoring the time dependence and adopting a fixed rate. With $\Psi(M_V)$ in hand, it is a straightforward matter to apply bolometric corrections and obtain the relative birthrates of stars as a function of $L_{\rm bol}$ rather than M_V. This is $\Psi(L_{\rm bol})$, the form of the initial luminosity function already shown in Figure 4.13. As anticipated, the curve here is smoother than $\Phi(M_V)$, since it lacks the age-dependent falloff for the brightest stars.

4.5.2 Character of the Mass Distribution

Our true goal, however, is the distribution at birth of various stellar *masses*. We accordingly define $\xi\,(M_*)$ to be the *initial mass function (IMF)*, the relative number of stars produced per unit mass interval. Again normalizing this function to unity, we have simply

$$\xi\,(M_*) = \Psi(M_V) \frac{dM_V}{dM_*} \ . \tag{4.5}$$

The derivative on the righthand side refers to variations along the main sequence and can be obtained numerically from Table 1.1 or its equivalent.[3]

Historically, E. E. Salpeter proceeded in the manner we have outlined to find that $\xi\,(M_*)$ varies as M_*^γ, with $\gamma = -2.35$. This simple power law is still frequently employed to obtain approximate results, but has long since been supplanted by other investigations using more extensive data. Figure 4.22 displays the results of a later study. For convenience, we may approximate the mass function as a sequence of power laws:

$$\xi\,(M_*) = \begin{cases} C\ (M_*/M_\odot)^{-1.2} & 0.1 < M_*/M_\odot < 1.0 \\ C\ (M_*/M_\odot)^{-2.7} & 1.0 < M_*/M_\odot < 10 \\ 0.40\, C\ (M_*/M_\odot)^{-2.3} & 10 < M_*/M_\odot \ , \end{cases} \tag{4.6}$$

---

[3] Many authors define the IMF as the relative number of stars per *logarithmic* mass interval, *i. e.*, as $M_*\,\xi(M_*)$. The reader should check carefully in each case.

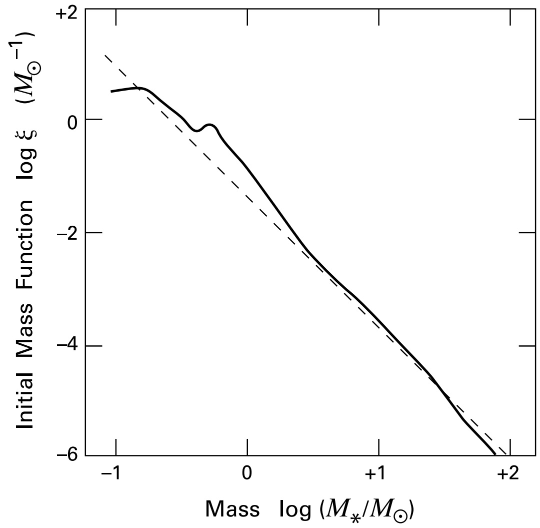


Figure 4.22 Initial mass function for solar neighborhood stars. The dashed line has the same slope as Salpeter's power law.

where C is a normalization constant. It is apparent that $\xi\,(M_*)$ is considerably flatter than Salpeter's function (*dashed line*) below $1.0\,M_\odot$ and approaches it for $M_* \gtrsim 10\ M_\odot$. Both of these general features have been amply confirmed. Our simple power law does not capture the broad maximum near $0.1\ M_\odot$ seen in the figure, but the true behavior at the lowest masses remains unclear. Their small luminosities make objects in this regime difficult to detect. After extensive searches, a growing population of *brown dwarfs*, *i. e.*, objects less massive than the hydrogen-burning limit of $0.08\ M_\odot$, is presently being uncovered. These and other data indicate that $\xi\,(M_*)$ is relatively flat near the brown-dwarf limit. Establishing its precise form will require additional effort.

This lingering uncertainty should not obscure the essential message of Figure 4.22. Within any volume undergoing star formation, the number of new stars per unit mass falls rapidly above roughly $0.1\ M_\odot$. If we make the simplifying assumption that $\xi\,(M_*)$ is constant below this value, then equation (4.6) implies that half of all stars are produced with $M_* \geq 0.2\ M_\odot$. Only 12 percent have masses exceeding $1\ M_\odot$, while the fraction drops to 0.3 percent for stars above $10\ M_\odot$. Conversely, 70 percent of stars have $M_* \geq 0.1\ M_\odot$. We conclude that *the star formation process yields objects with a characteristic mass of a few tenths of* $M_\odot$. Things might have been otherwise. One could imagine stars forming with a pure power-law mass spectrum down to some very low level at the planetary scale. That this is *not* the case is surely significant. Unfortunately, current theory cannot explain, in any convincing manner, the form of $\xi\,(M_*)$. Even the origin of the basic mass scale itself remains uncertain. Needless to say, we shall revisit this central issue in later chapters.

4.5.3 Mass Function in Stellar Groups

It is worth reemphasizing that $\xi\,(M_*)$ represents an average over thousands of field stars, *i. e.*, objects outside of known clusters or associations. What is the corresponding function within such groups? In any one system of a few hundred stars or less, statistical fluctuations become significant when addressing this issue. With this caveat in mind, the best candidates for investigation are the relatively unobscured open clusters. Here, the members are no longer accreting molecular gas, while the stellar masses themselves can be read with confidence from the HR diagram, given knowledge of both L_* and $T_{\rm eff}$. On the other hand, the main-sequence turnoffs limit the highest observed masses to modest values, about 15 $M_\odot$ for a cluster age of 10^7 yr. One must turn to OB associations to probe the upper end of the distribution. Unfortunately, both the greater distances and the presence of main-sequence turnons make the lowest-mass members difficult to access. Since no one system is ideal, one is forced to sample the mass spectrum within groups in a piecemeal fashion.

Figure 4.23 illustrates the complementary roles of open clusters and OB associations. The left panel shows the number of stars per unit mass, denoted $\xi_{\rm cluster}(M_*)$, for most known members of the Pleiades, along with the curve from equation (4.6). The flattening below $M_* \approx 0.15\,M_\odot$ is similar to that seen in Figure 4.22 for field objects. Note that the luminosities and effective temperatures of many cluster members were obtained here from their R- and I-band magnitudes. The stellar masses then followed by comparison with theoretical pre-main-sequence tracks. In contrast, single measurements at V sufficed for the more massive stars, which are certain to be on the main sequence. It is apparent that the mass distribution in this accessible and populous system also matches the field (*i. e.*, equation (4.6)) between 0.15 and 5 $M_\odot$. Other open clusters yield mass functions that are similar but exhibit significant variation. Such deviations from the field-star result do not appear to be correlated with cluster morphology or age, and largely disappear if one adds together the populations of at least a dozen systems.

Turning to OB associations, the most reliable procedure is to focus on rich, spatially compact subgroups. One such subgoup is NGC 6611, a cluster within the Serpens OB1 association, 2.2 kpc distant. The stars here illuminate the Eagle Nebula (M16), long known as a visible HII region traversed by broad lanes of obscuration. Most of the stars in NGC 6611 are still embedded within the local molecular cloud. However, the 150 or so members above 5 $M_\odot$ are bright enough that they can be placed in the HR diagram through optical measurements. Figure 4.23b shows the masses at this upper end of the distribution. A power-law falloff is evident. Indeed, the best-fit line has $\Gamma \equiv {\rm d}\log\xi/{\rm d}\log(M_*/M_\odot) = -2.1$, close to the Salpeter value of -2.35. Other, less populous clusters within associations have Γ-values that range from -1.7 to -3.0 and a mean consistent with the slope of the field-star initial mass function.

We conclude, then, that all groups of sufficient membership display a similar decline in population with mass, at least above the brown-dwarf regime. The steepness of this falloff for stars exceeding several $M_\odot$ makes it difficult to obtain the complete mass or population of any group by extrapolation from its very brightest components. We stress that the actual form of the initial mass function must currently be viewed as a purely empirical result, one whose proper explanation awaits better understanding of both cluster formation and the termination of protostellar collapse. In particular, the absence of any obvious breaks in the observed distribution does *not* necessarily imply that a single mechanism is at work in the origin of all stellar masses. On the

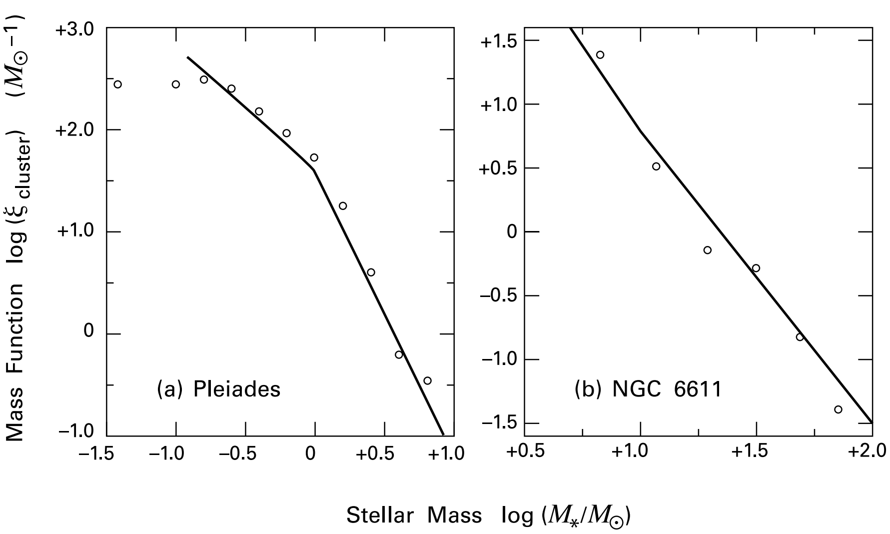


Figure 4.23 Empirical mass functions for (a) the Pleiades, and (b) the cluster NGC 6611. The solid curves represent the initial mass function from equation (4.6).

contrary, their distinctive central concentration in bound clusters suggests that the most massive stars form very differently from their more numerous, low-mass counterparts.

Chapter Summary

Stars are not formed separately, but in groups within molecular clouds. All the observed members of such primitive, embedded clusters suffer heavy extinction and reddening because of interstellar dust. Additionally, they exhibit varying degrees of excess infrared emission created by circumstellar matter, *i. e.*, dust immediately adjacent to the stars themselves. The ensemble of observed spectral energy distributions fall into four classes that appear to define an evolutionary sequence.

Somewhat older, visible objects are generally located in either T or OB associations. The former contain classical and weak-lined T Tauri stars along with residual cloud gas, and remain intact for periods up to 10^7 yr. OB associations, so called because they also include a few massive objects, are already dispersing into the field population. Here, the molecular gas was driven off violently by the winds and radiation pressure of the highest-mass members. Finally, a relatively small fraction of stars survives dissipation of their parent cloud as gravitationally bound clusters.

The observed luminosities of field stars, together with their main-sequence lifetimes, allow one to deduce the statistical distribution of masses at birth. This initial mass function peaks between 0.1 and 1.0 $M_\odot$, a fundamental and unexplained fact. The distribution within specific groups, obtained by placing members in the HR diagram, agrees broadly with the field-star result.

Suggested Reading

Section 4.1 For the relative birthrates of stars in different cluster environments, see

Miller, G. E. & Scalo, J. M. 1978, PASP, 90, 506.

Two reviews of embedded clusters, both stressing near-infrared observations, are

Zinnecker, H., McCaughrean, M. J., & Wilking, B. A. 1993, in *Protostars and Planets III*, ed. E. H. Levy and J. I. Lunine (Tucson: U. of Arizona Press), p. 429

Lada, C. J. & Lada, E. A. 2003, ARAA, 41, 57.

The 2MASS (Two Micron All Sky Survey) project has provided a view of the Milky Way in three near-infrared bands:

Skrutskie, M. F. et al. 1997, in *The Impact of Large Scale Near-Infrared Surveys*, ed. F. Garzon et al. (Dordrecht: Reidel), p. 25.

The 2MASS data were released in 2003. For the original classification of embedded stars by their infrared spectral index, see

Lada, C. J. 1987, in *Star Forming Regions*, ed. M. Peimbert and J. Jugaku (Dordrecht: Reidel), p. 1.

Section 4.2 The objects now called classical T Tauri stars were first recognized as a distinct group by

Joy, A. H. 1945, ApJ, 102, 168,

while their weak-lined counterparts were classified by

Walter, F. W. 1986, ApJ, 306, 573.

Note that Walter actually identified the “naked” group, *i. e.*, those weak-lined T Tauri stars lacking near-infrared excess. For the morphology of nearby T associations, see the discussion of Chamaeleon by

Schwarz, R. D. 1991, in *Low-Mass Star Formation in Southern Molecular Clouds*, ed. B. Reipurth (ESO Publication), p. 93,

and that of Taurus-Auriga in

Palla, F. & Stahler, S. W. 2002, ApJ, 581, 1194.

The identification of R associations is due to

Van den Bergh, S. 1966, AJ, 71, 990.

Section 4.3 Two contributions of historical interest on the dynamical expansion of OB associations are

Blaauw, A. 1952, BAN, 11, 405

Ambartsumian, V. A. 1955, Observatory, 75, 72.

A modern review of these associations is

Garmany, C. D. 1994, PASP, 106, 25.

The relationship of these groups to molecular clouds is analyzed in

Williams, J. P. & McKee, C. F. 1997, ApJ, 476, 166.

Section 4.4 The reader wishing to see how spectroscopic parallax is used in practice to derive the distances to young clusters may consult

Perez, M. R., Thé, P. S., & Westerlund, B. E. 1987, PASP, 99, 1050.

For properties of the Galactic system of open clusters, see

Janes, K. A., Tilley, C., & Lyngå, G. 1988, AJ, 95, 771.

A theoretical work that lucidly discusses the main issues in their evolution is

Terlevich, E. 1987, MNRAS, 224, 193.

Section 4.5 The concept of the initial mass function and its first determination are due to

Salpeter, E. E. 1955, ApJ, 121, 161.

A comprehensive review that delves into the many subtleties in this continuing endeavor is

Scalo, J. M. 1986, Fund. Cosm. Phys., 11, 1.

For a more recent discussion, see

Kroupa, P. 2002, Science, 295, 82.

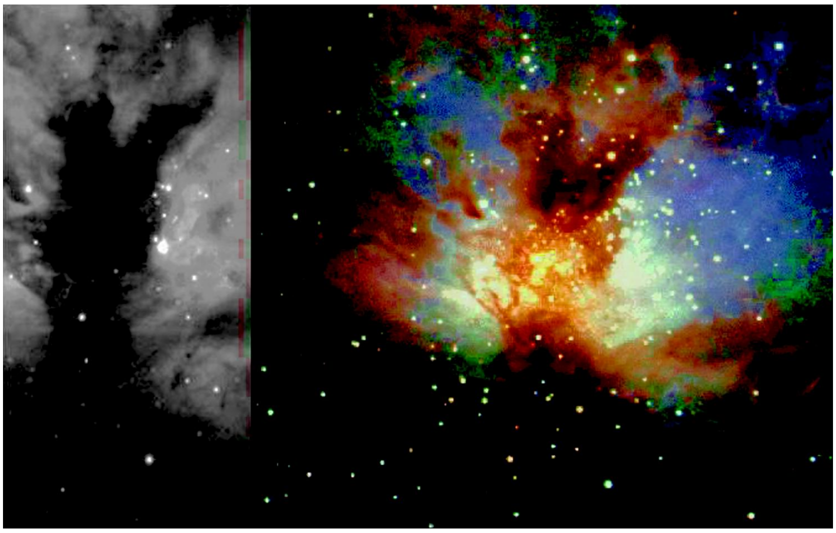


Plate 1 *left:* Optical photograph of the NGC 2024 cluster in Orion B. The image covers an angular size of $4' \times 10'$, or $0.4 \times 1\ \mathrm{pc}$. *right:* Near-infrared image of the same cluster. The vertical scale matches that of the optical image. This is a composite picture combining three separate mosaics in the J, H, and K wavebands. The color coding is blue, green, and red, respectively.

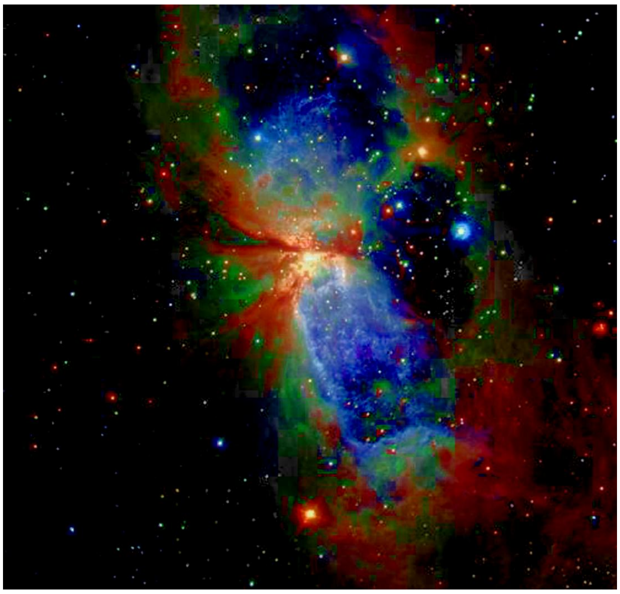


Plate 2 Near-infrared image of the S106 bipolar nebula. This is a composite of the J, H, and K wavebands. The color coding in this and all other composite, near-infrared images is the same as in Plate 1.

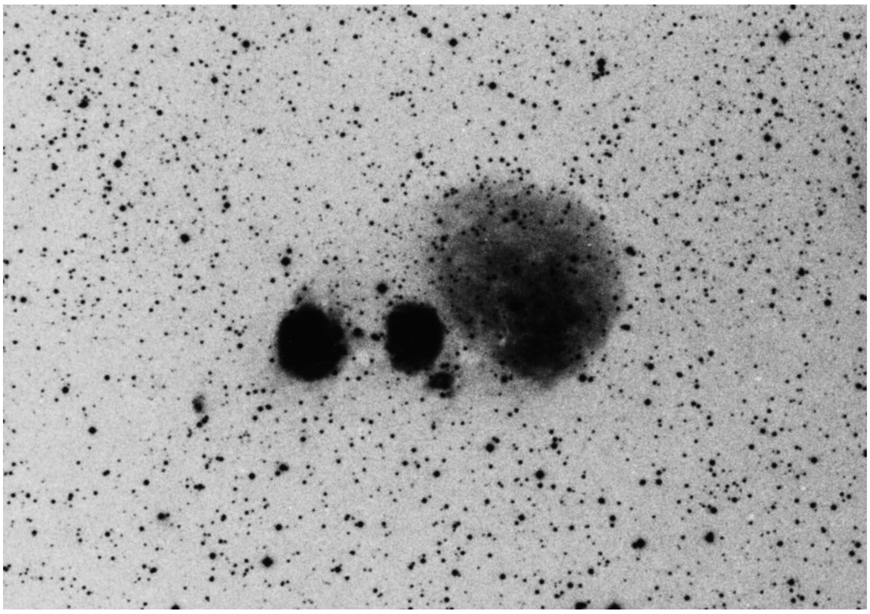


Plate 3 Negative optical photograph of three HII regions in the Gem OB1 association. The regions span a total distance of 9 pc.

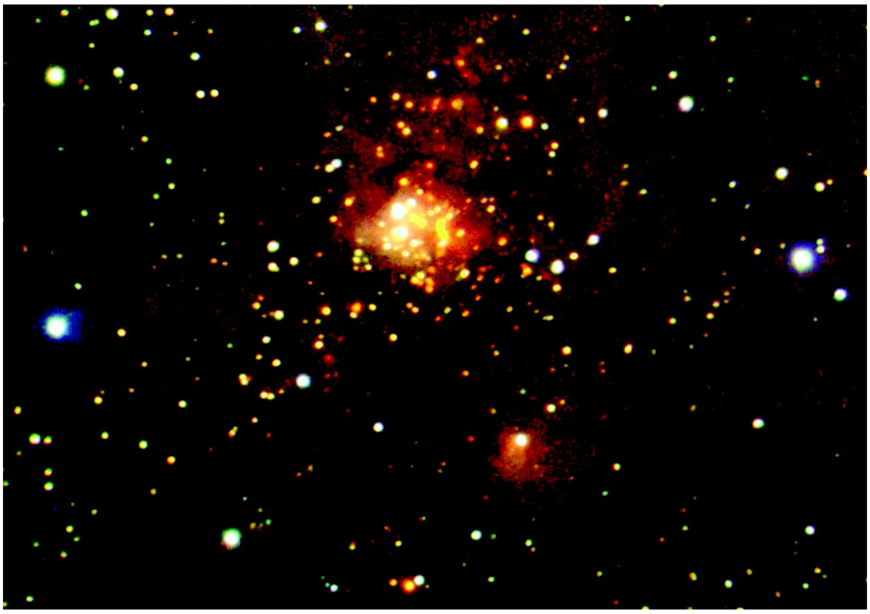


Plate 4 Expanded near-infrared (J, H, and K) image of Gem OB1. The bright nebula in the center lies between the left and center HII regions of Plate 3, and is invisible optically.

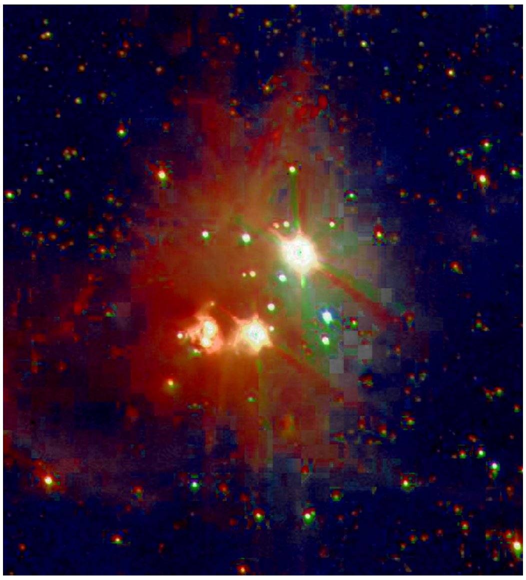


Plate 5 Composite, near-infrared image (J, H, and K) of the region surrounding the Herbig Be star BD$+40^{\circ}4124$. This object is the brightest central spot. The most prominent companion is an emission-line star, V1686 Cyg, that is also optically visible, but most of the other, nearby objects can only be seen in the near-infrared.

Plate 6 Near-infrared image (J, H, and K) of NGC 7538. The entire image covers $12' \times 12'$, or 9.5×9.5 pc at the 2.7 kpc distance of the region. The red patches are embedded clusters that appear to be younger than the prominent HII region.

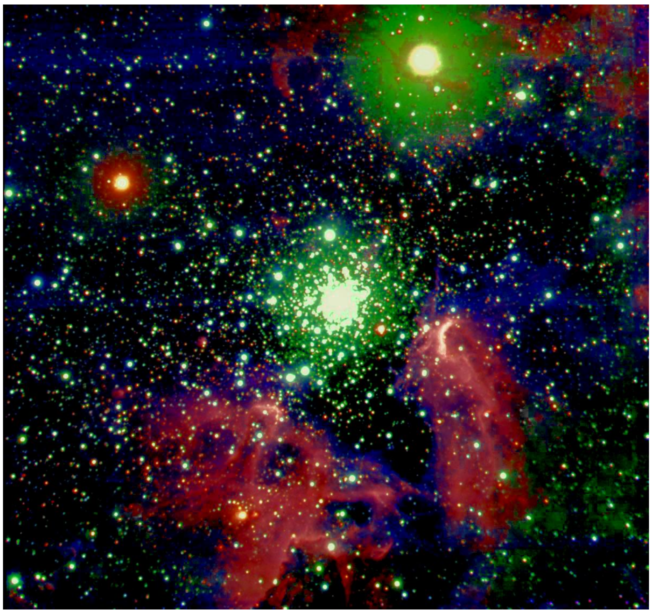


Plate 7 Composite J, H, and K-band near-infrared image of the dense cluster NGC 3603. The field of view is $3\rlap{.}'5 \times 3\rlap{.}'5$, or 6×6 pc at a distance of 6 kpc.

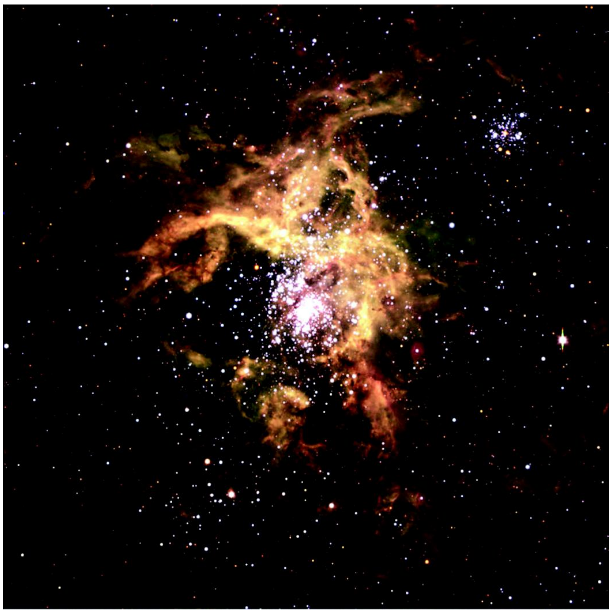


Plate 8 Composite image in B, V, and R-bands of 30 Doradus in the Large Magellanic Cloud. The gas filaments surrounding the central cluster span about 50 pc.

Part II

Physical Processes in Molecular Clouds

The Formation of Stars
Steven W. Stahler and Francesco Palla

5 Molecular Transitions: Basic Physics

With this chapter, we begin a broad discussion of molecular clouds from a physical perspective. Our treatment throughout Part II will be more theoretical and quantitative than previously, as the underlying goal is to provide a basis for understanding cloud structure and protostellar collapse, the subjects of Part III. We will not attempt an exhaustive coverage of molecular cloud physics, but concentrate instead on those areas that seem at present to be most relevant for star formation.

To explore conditions within the clouds that produce stars, astronomers rely mainly on observations of spectral lines emitted by various molecules. Hence, our initial goals in this chapter are to describe how such species form throughout interstellar space and how their abundances reflect local conditions. We next turn to the simplest and most common molecules that have been employed as tracers of cloud properties. The aim is to present, succinctly but accurately, the physical principles underlying the most readily observed transitions. Chapter 6 will then demonstrate how these transitions are used in practice to determine cloud properties.

5.1 Interstellar Molecules

From a chemical viewpoint, an important feature of dark clouds is their relatively high column density in dust, which effectively blocks ambient radiation at both optical and ultraviolet wavelengths. Hence interstellar molecules, which would have short lifetimes against ultraviolet photodissociation in unshielded regions of space, are able to survive and proliferate. To date, over 100 molecules have been identified, ranging from the simplest diatomic species to long chains like the cyanopolyyne $HC_{11}N$. Dense cores contain many of the more complex species found so far, but the shock-heated regions of Orion and the Galactic center cloud Sagittarius B2 have also been rich sources. Equally important are the distended envelopes of evolved, giant stars, which we have already noted as the birth sites of interstellar grains.

5.1.1 Reaction Energetics

Molecular astrophysics began in the late 1930s, with the discovery of CH, CH^+, and CN in diffuse clouds. These simple molecules were detected by their absorption of optical light from background stars. The question of how such species form immediately posed a theoretical challenge, one which deepened with the discovery in the 1960s of OH, NH_3, and H_2O. The problem is one of energetics. Consider first the collision of two atoms. The particles approach each other with positive total energy. Unless energy can somehow be given to a third body, the atoms will simply rebound after their encounter. The simultaneous collision of a third atom can occur with appreciable frequency at terrestrial densities, but not in the vastly more rarefied interior of a molecular cloud. It is also possible for the energy sink to be a photon, *i. e.*, for

The Formation of Stars. Steven W. Stahler and Francesco Palla

ISBN: 3-527-40559-3

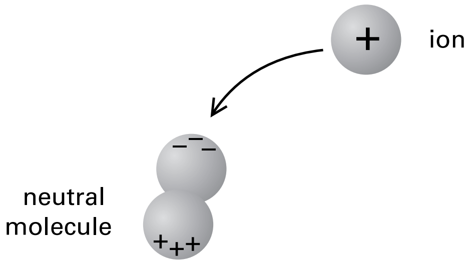


Figure 5.1 Mechanics of ion-molecule reactions. The charged ion induces a dipole moment in the neutral molecule, creating an electrostatic attraction between the two species.

the two atoms to form an excited molecule which radiatively decays to the ground state before it can dissociate. Again, the probability of such *radiative association* is generally too low in molecular clouds to be of interest.

In the laboratory, molecules also form through *neutral-neutral* reactions. Here, the colliding species, whether atoms or molecules, combine temporarily into a configuration known as an *activated complex*. This then separates into two or more product species that share the total energy. The process involves the making and breaking of chemical bonds, and usually requires a net expenditure of energy. The associated *activation barrier* has a typical energy, expressed in temperature units, of $\Delta E/k_B \sim 100$ K. A barrier of this magnitude is not insurmountable in shock-heated clouds, but completely suppresses neutral-neutral reactions in the very cold interiors of quiescent clouds.

By the early 1970s, it had become clear that *ion-molecule* reactions can alleviate the energy difficulty. When a charged ion approaches a neutral molecule (or atom), it induces a dipole moment in the latter, creating an electrostatic attraction between the two. (See Figure 5.1.) The long-range nature of this attraction means that the effective cross section is greatly increased above the geometric value for direct collision. Even at temperatures near 10 K, such reactions can proceed fast enough to account for a large fraction of observed interstellar molecules. Since, however, the fraction of ions available at any time is relatively small, the huge abundance of H_2 itself cannot be explained in this manner. In this important special case, to which we shall return later, two neutral atoms *can* react, but only through the catalytic action of an interstellar grain surface.

In symbolic form, the generic ion-molecule reaction may be written

$$\mathrm{A}^+ + \mathrm{B} \rightarrow \mathrm{C}^+ + \mathrm{D} \ , \tag{5.1}$$

where all species can be atoms or molecules. If $\mathrm{C} = \mathrm{B}$ and $\mathrm{D} = \mathrm{A}$, the reaction is simple *charge exchange*. Reactions involving negative ions and neutrals are also possible, in which case one of the products is a free electron. If n_{A^+} and n_{B} denote the number densities of the reactants, then we let $k_{\mathrm{im}}\,(n_{\mathrm{A}^+})\,(n_{\mathrm{B}})$ be the reaction rate per unit volume per unit time. For either positive or negative ions, the rate coefficient k_{im} is of order 10^{-9} cm^3 s^{-1} and is only weakly dependent on temperature.

Positively charged molecules within clouds are also destroyed by ambient free electrons. The electron recombines to create an energetic, unstable neutral molecule. Most of the time, this molecule simply "autoionizes," spitting back the electron. However, if its constituent atoms separate before autoionization occurs, the molecule falls apart into neutral species:

$$\mathrm{A}^+ + \mathrm{e}^- \rightarrow \mathrm{B} + \mathrm{C}\ . \tag{5.2}$$

The rate coefficient for such *dissociative recombination* is $k_{\rm dr} \sim 10^{-7}\ {\rm cm}^3\ {\rm s}^{-1}$ for a temperature T near 100 K and *increases* slowly as the temperature falls. Note that, if A represents an atom instead of a molecule, reaction with an electron yields the neutral form of the atom plus a photon. The typical rate for this *radiative recombination*, $k_{\rm rr} \sim 10^{-11}\ {\rm cm}^3\ {\rm s}^{-1}$, is again very low for most circumstances.

5.1.2 Abundance Patterns

Suppose we wish to observe some molecule through its emission in a particular spectral line. The detection of this line, which results from a transition between discrete energy levels, requires an ambient temperature high enough to excite the upper level of interest. Within quiescent clouds, this requirement generally singles out low-lying rotational transitions, with associated photon wavelengths in the millimeter regime. Even when it has such a transition, however, a highly complex molecule is intrinsically difficult to detect. In this case, a great many levels exist in any appreciable energy range. Many of these become populated in a sufficiently warm environment, so that the power in any one transition is relatively small.

Most of the molecules observed to date contain one or more carbon atoms. While no inorganic species found in space contain more atoms than NH_3, organic molecules exist in complex rings and chains. Hence the carbon bond, which plays such a dominant role in terrestrial chemistry, is also important in the interstellar environment. Furthermore, since the cosmic abundance of oxygen exceeds that of carbon, it is no surprise that the relatively tightly bound CO is the most abundant species after H_2 itself. The observations of CO, pursued since 1970, have yielded more information on star-forming regions than any other molecule.

Theorists employ time-dependent computer models to understand the pattern of chemical abundances in any region. These programs simulate large reaction networks, usually operating at fixed ambient density and temperature. With time, the reactions that create and destroy various species equilibrate, and the abundances approach steady-state values. Dense cores with no internal stars provide a particularly simple environment to test such schemes. Starting with reasonable initial conditions, the models have little difficulty matching the observed abundances of simple species like CO, CS, or HCO^+. Such agreement represents a gratifying confirmation of the ion-molecule chemistry at the root of these networks. On the other hand, the models are not without problems. One finds that complex organics always build up in time initially and then disappear as carbon becomes locked up in CO. At the density and temperature of a typical dense core such as TMC-1, there would be nearly complete conversion of atomic carbon to CO by 1×10^6 yr.

Such a time is difficult to reconcile with our understanding of cloud history. Although no accurate ages for dense cores are available, the process of gravitational settling that creates them operates over a period of order 10^7 yr. The observed presence of organics therefore remains puzzling. In addition, the current chemical models have difficulty explaining the significant spatial variation in molecular abundances seen across TMC-1 and other starless cores. The two simplest possibilities– gradients in age or elemental composition– both seem rather contrived as

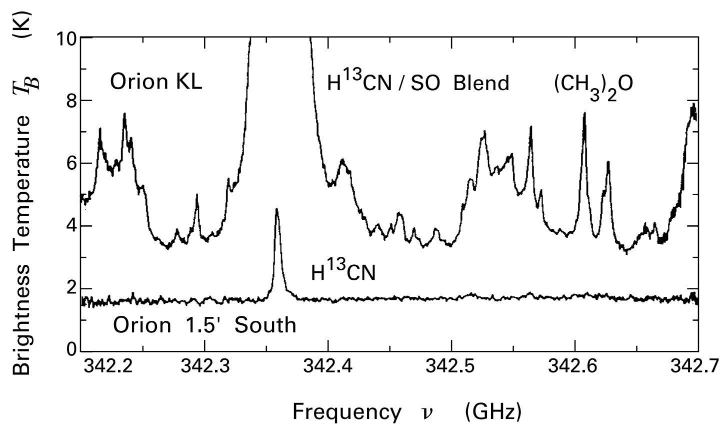


Figure 5.2 Submillimeter spectrum of two regions in Orion. The KL Nebula (*upper spectrum*) shows many more molecular lines than the dense core in Orion 1.5′ South (*lower spectrum*).

a general explanation. The observed gradients could also reflect the presence of still undetected young stars, whose luminosities would thermally alter the local chemistry.

Chemical gradients are also present in sites of high-mass star formation, such as the Orion BN-KL region. The fact that the luminous infrared sources IRc2 and BN have significantly affected their surroundings is evident from Figure 5.2. Here a portion of the spectrum near the 868 μm ($\nu = 345$ GHz) transition of H^{13}CN is shown, both for the KL Nebula and for a starless dense core known as Orion 1.5′ South, located 0.2 pc from IRc2. (Recall Figure 1.7.) Although the two cores have similar total column densities through their centers, the one in the KL Nebula has a far richer spectrum of molecular lines. At the same time, this region has a *lower* abundance of complex organics. In fact, Orion 1.5′ South has a similar chemical makeup as TMC-1, although it has 5 times the temperature and 10 times the density of the latter. Clearly, the critical factor here is proximity to a massive star.

Several effects are at work to alter and enrich the molecular population near IRc2. The core material surrounding the star has a temperature of some 200 K, sufficient to vaporize grain mantles and thereby reinject molecules into the gas phase. Even closer to the star, shock waves driven by the massive outflow can heat the gas to the point of destroying the grain cores themselves. Such thermal processing explains why sulfur and silicon, both major grain constituents, are relatively abundant in this and other outflow regions. At the same time, such environments are unfavorable to long chains, which apparently need heavily shielded dense cores to form and survive.

Of the total list of known interstellar molecules, about half were first discovered in Sagittarius B2, a giant molecular cloud located within 200 pc of the Galactic center. With a visual extinction of some 30 mag, even near-infrared observations of the region are unable to detect any but the brightest sources. Figure 5.3 is a high-resolution map in 800 μm continuum emission from heated dust. The cloud complex Sgr B2 lies along the dust ridge that also contains Sgr A*, a compact radio source only a few parsecs from the Galaxy's true center. As in the case of Orion, the abundance of molecules is correlated with an elevated gas temperature created by the presence of many embedded stars. Most molecules are seen in Sgr B2/North, a compact clump with gas temperature 200 K and a density in H_2 as high as 10^7 cm^{-3}. Here, it is impossible to see the stellar cluster directly, but its presence is inferred from the $5 \times 10^6\ L_\odot$ in radiation emitted by heated dust. The pattern in molecular abundances broadly resembles Orion KL. It is intriguing that an equally luminous clump located just 2 parsecs away from Sgr B2/North is relatively sparse in molecular lines. The luminosity in this latter region, known as Sgr B2/Middle, stems from identifiable HII regions associated with several dozen O and B stars. The lesson here is that the ultraviolet radiation inflating an HII region also destroys the molecules that previously surrounded the star when it was more deeply embedded.

5.1.3 Adherence to Grains

Any account of the chemistry within dense clouds must consider the propensity of molecules to stick to grain surfaces. Consider a volume V of cloud gas with a number density n_d of spherical grains, each with radius a_d. In the reference frame of a molecule, the grains are all moving with $V_{\rm therm}$, the molecule's thermal speed:

$$V_{\rm therm} = \left(\frac{3\,k_B\,T}{2\,A\,m_H}\right)^{1/2} . \tag{5.3}$$

Here A is the molecule's mass relative to hydrogen. In a time interval Δt, each grain sweeps out the cylindrical volume $\pi a_d^2\, V_{\rm therm}\,\Delta t$, and all grains sweep out a volume larger by $n_d V$. Hence, the probability per unit time that the molecule is struck by some grain is the ratio of this total volume to V, or $n_d\,\pi a_d^2\, V_{\rm therm}$. Inverting this probability gives the average time for a collision to occur:

$$\begin{aligned} t_{\rm coll} &= \frac{1}{n_d\,\pi a_d^2\, V_{\rm therm}} \\ &= \frac{1}{n_H\,\Sigma_d\, V_{\rm therm}} . \end{aligned} \tag{5.4}$$

Here we have used equation (2.42) for the total geometric cross section of the grains per hydrogen atom.

The quantity $t_{\rm coll}$ measures the time to deplete significantly a given molecule, provided there is a high probability of sticking upon collision. Such is the case for all molecules except H_2, which does not readily adhere to grain mantles. Consider CS, for which $V_{\rm therm} = 5.3 \times 10^3$ cm s^{-1} at $T = 10$ K. From equation (5.4), we find that $t_{\rm coll}$ is only 6×10^5 yr at the center of a dense core with $n_H = 10^4$ cm^{-3}. Once again, we face the dilemma that the disappearance time is brief compared to the expected cloud age. To put the matter another way, chemical models *without* grain depletion of molecules give a reasonable match to the observed CS abundance.

There evidently must exist some mechanism for reinjecting molecules from grain surfaces back into the gas phase. Ultraviolet photons would serve the purpose, but too few of them penetrate dark cloud interiors. In sufficiently small grains, the heat from surface chemical reactions could raise the grain temperature enough to sublimate many species. However, for standard grains within dark clouds, the problem of rapid depletion of the molecules remains unsolved.

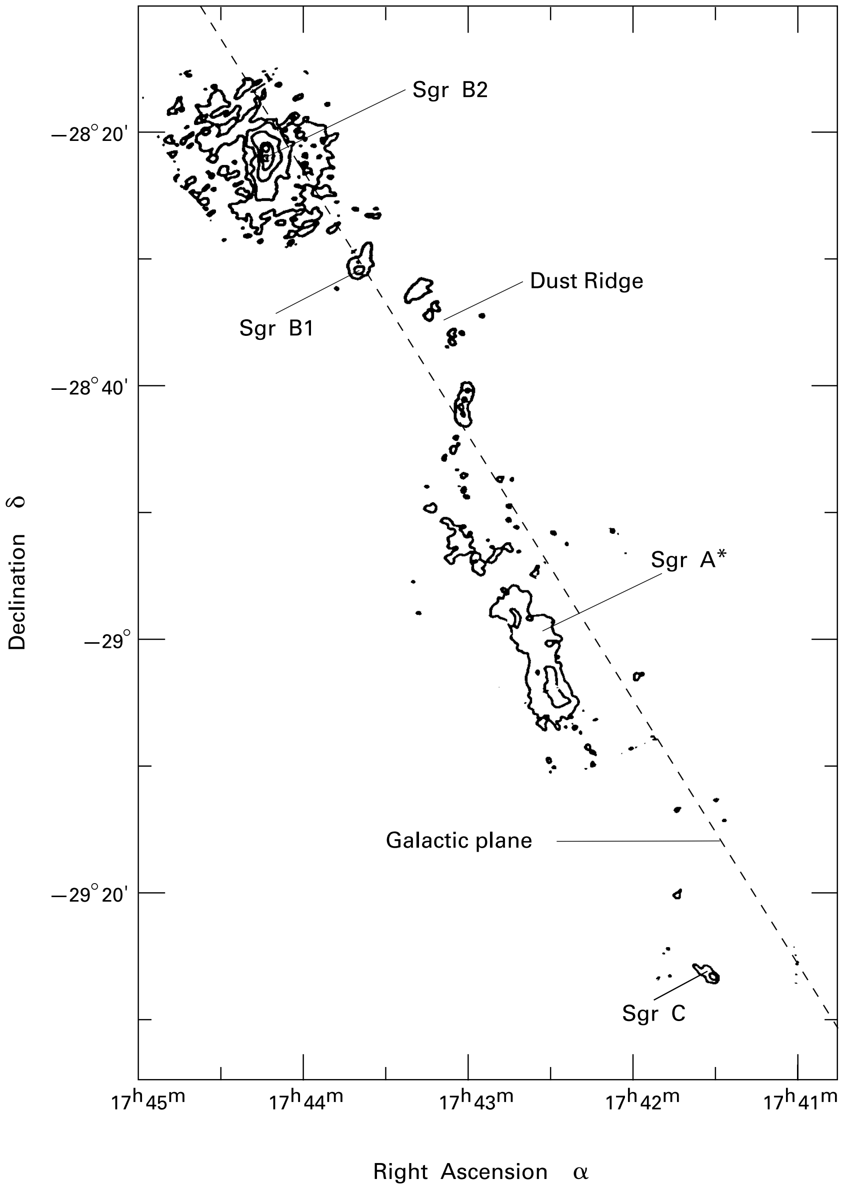


Figure 5.3 Continuum emission at 800 μm from heated dust near the Galactic center. Many molecules have been discovered in the giant molecular cloud Sgr B2, located 200 pc from the strong radio source Sgr A*. The dashed line demarcates the Galactic plane.

5.2 Hydrogen (H_2)

We now consider a few of the species that have been especially fruitful in the study of molecular clouds. Table 5.1 lists abundances and important transitions for all the molecules discussed in this chapter, as well as some others of interest. The sixth column gives the energy difference between the upper and lower states. This energy is expressed as an equivalent temperature by using the Boltzmann constant k_B. Also given is the Einstein coefficient $A_{\rm ul}$, *i. e.*, the probability per time of spontaneous decay from the upper to the lower state. Appendix B introduces the Einstein coefficients more systematically. Finally, Table 5.1 lists the *critical density*, $n_{\rm crit} \equiv A_{\rm ul}/\gamma_{\rm ul}$, where $\gamma_{\rm ul}$ is the rate of collisionally induced downward transitions in the molecule, as measured per unit colliding partner. This quantity, as we have noted, is an estimate of the minimum ambient density at which collisions depopulate the upper state before it can decay through radiation. The utility of $n_{\rm crit}$ will become evident as we proceed through specific examples.

5.2.1 Allowed Transitions

It is an unfortunate fact that the chief constituent of cold interstellar clouds, the hydrogen molecule, is also among the most difficult to detect. Even the lowest excited energy levels, those corresponding to molecular rotation, are too far above the ground state to be easily populated. In addition, H_2 consists of two identical hydrogen atoms and therefore lacks a permanent electric dipole moment. A rotationally excited molecule must radiate through a relatively slow quadrupole transition. To find H_2, it is best to look in hotter environments, such as clouds irradiated by a luminous star or shocked by a stellar wind. Here, photons or particle collisions can excite vibrational and electronic states that do decay in a relatively brief time. The molecule was actually first detected in 1970, through rocket observations that found several ultraviolet absorption lines in the direction of the O star ξ Persei. These lines arise from the photo-excitation of electronic states in H_2 within an intervening diffuse cloud.

Since the hydrogen molecule plays a dominant role in many aspects of star formation, it is worthwhile to examine its transitions in some detail. We begin with the rotational levels. In classical mechanics, the kinetic energy of a dumbbell rotating about an axis through the center of mass and perpendicular to the plane of rotation is given by

$$E_{\rm rot} = \frac{J^2}{2\,I} \ , \tag{5.5}$$

where I is the moment of inertia and J the angular momentum. The quantum-mechanical analog of (5.5) is

$$\begin{aligned} E_{\rm rot} &= \frac{\hbar^2}{2\,I}\,J\,(J+1) \\ &\equiv B\,h\,J\,(J+1) \ . \end{aligned} \tag{5.6}$$

Here J is now the dimensionless *rotational quantum number*, which can be 0, 1, 2, *etc.* The quantity B in the second form of equation (5.6) is known as the *rotational constant* and has the units of frequency.

Table 5.1 Some Useful Molecules

molecule	abundance[a]	transition	type	λ	$T_\circ^b$ (K)	A_{ul} (s^{-1})	$n_{\rm crit}^c$ (cm^{-3})	comments
H_2	1	1→0 S(1)	vibrational	2.1 μm	6600	8.5×10^{-7}	7.8×10^{7}	shock tracer
CO	8×10^{-5}	$J=1\rightarrow0$	rotational	2.6 mm	5.5	7.5×10^{-8}	3.0×10^{3}	low density probe
OH	3×10^{-7}	$^2\Pi_{3/2}$;J=3/2	Λ-doubling	18 cm	0.08	7.2×10^{-11}	1.4×10^{0}	magnetic field probe
NH_3	2×10^{-8}	(J,K)=(1,1)	inversion	1.3 cm	1.1	1.7×10^{-7}	1.9×10^{4}	temperature probe
H_2CO	2×10^{-8}	$2_{12}\rightarrow1_{11}$	rotational	2.1 mm	6.9	5.3×10^{-5}	1.3×10^{6}	high density probe
CS	1×10^{-8}	$J=2\rightarrow1$	rotational	3.1 mm	4.6	1.7×10^{-5}	4.2×10^{5}	high density probe
HCO^+	8×10^{-9}	$J=1\rightarrow0$	rotational	3.4 mm	4.3	5.5×10^{-5}	1.5×10^{5}	tracer of ionization
H_2O		$6_{16}\rightarrow5_{23}$	rotational	1.3 cm	1.1	1.9×10^{-9}	1.4×10^{3}	maser
″	$<7\times10^{-8}$	$1_{10}\rightarrow1_{11}$	rotational	527 μm	27.3	3.5×10^{-3}	1.7×10^{7}	warm gas probe

[a] number density of main isotope relative to hydrogen, as measured in the dense core TMC–1

[b] equivalent temperature of the transition energy; $T_\circ \equiv \Delta E_{\rm ul}/k_B$

[c] evaluated at T=10 K, except for H_2 (T=2000 K) and H_2O at 527 μm (T=20 K)

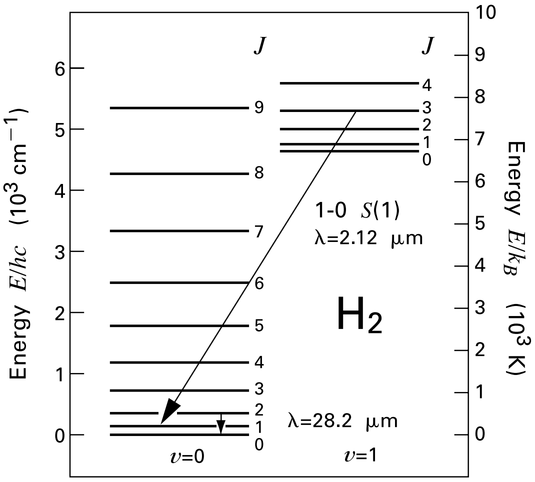


Figure 5.4 Rotational levels of H_2 for the first two vibrational states. Within the $v = 0$ state, the $J = 2 \rightarrow 0$ transition at 28.2 μm is displayed. Also shown is the transition giving the 1-0 S(1) rovibrational line at 2.12 μm. Note that two different energy scales are used.

The moment of inertia of H_2 is the smallest of any diatomic molecule, so equation (5.6) shows why its energy levels are widely spaced. Figure 5.4 displays the ladder of rotational levels. Here the energy E is given both as an equivalent temperature, E/k_B, and in wave numbers, E/hc. Note that the two scales are similar numerically, since $k_B/hc = 0.70\ \mathrm{deg}^{-1}\ \mathrm{cm}^{-1}$. The second measure is convenient in that the energy difference between two states yields directly the inverse of the wavelength of the emitted photon.[1] As already noted, the rotational levels of H_2 decay principally through electric quadrupole transitions, in which J decreases by 2. The lowest possible transition, $J = 2 \rightarrow 0$, has an associated energy change of 510 K and the relatively low Einstein A-value of $3.0 \times 10^{-11}\ \mathrm{s}^{-1}$. Each decay produces a photon of wavelength 28.2 μm. This far-infrared line has been detected through spaceborne observations.

The total energy of the hydrogen molecule is the sum of its rotational, vibrational, and electronic contributions:

$$E_{\mathrm{tot}} = E_{\mathrm{rot}} + E_{\mathrm{vib}} + E_{\mathrm{elect}}\ . \tag{5.7}$$

In quantum mechanics, the energy of a simple harmonic oscillator of natural frequency $\nu_\circ$ is

$$E_{\mathrm{vib}} = \hbar\varpi_\circ\,(v + 1/2)\ , \tag{5.8}$$

where $\varpi_\circ \equiv 2\pi\nu_\circ$, and where the *vibrational quantum number* v can be 0, 1, 2, *etc.* The $J = 0$ rotational level of the $v = 1$ state has an energy above ground equivalent to 6.6×10^3 K and an A-value of $8.5 \times 10^{-7}\ \mathrm{s}^{-1}$. In the hot environments where vibrational states are excited, the molecule relaxes through *rovibrational* transitions, in which both J and v change. (See Figure 5.4.) Here, the change in v is unrestricted, while ΔJ can be 0 or ± 2. Suppose v' and v''

---

[1] In this chapter only, we follow the spectroscopic definition of wave number, $k \equiv 1/\lambda$.

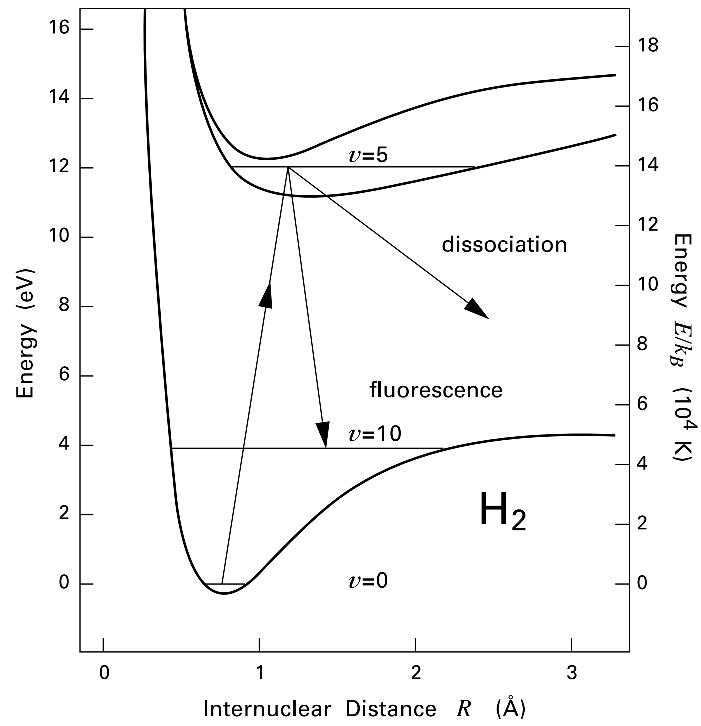


Figure 5.5 Potential energy of H_2 as a function of internuclear separation. Note again the use of two alternate energy scales. The three solid curves correspond to the ground and first two excited electronic states. The horizontal lines represent vibrational levels. Arrows depict photo-excitation into the Lyman band, followed by either fluorescent decay or dissociation.

denote the initial and final vibrational states, respectively. Then we designate the rovibrational transitions as $v' - v''\, O(J'')$, $Q(J'')$, or $S(J'')$ for $J'' - J' = 2,\ 0$, or -2. Thus, the commonly observed $1 - 0\, S(1)$ line, with a wavelength of $2.12\,\mu$m, represents the transition from the $(v' = 1,\ J' = 3)$ to the $(v'' = 0,\ J'' = 1)$ level.

The electronic states are even more widely spaced, with an energy separation of order 10^5 K. Figure 5.5 is a potential energy diagram for the ground and first two excited states, which are formed by promoting one electron to a higher orbital. The figure displays the Coulomb energy of the protons plus the binding energy of the electrons, calculated as a function of internuclear separation. This total energy acts as the potential well in which the protons oscillate. For each electronic state, the equilibrium separation of the nuclei is located at the minimum of its potential.

The ground electronic state contains 14 vibrational levels, plus a continuum of levels with $E > \Delta E_{\text{diss}} = 4.48$ eV, the molecule's binding energy. In regions of high density, such as shocks or collapsing clouds, H_2 may be destroyed through energetic collisions. Otherwise, dissociation occurs by the absorption of an ultraviolet photon. For photon energies exceeding 14.7 eV, the process is direct and leaves one of the hydrogen atoms in an excited electronic state. However, indirect radiative dissociation is more common. Here, a photon with $E > 11.2$ eV excites H_2 to a higher electronic state. About 85 percent of the time, the excited molecule drops back to ground through electronic and rovibrational transitions. The cluster of transitions connecting levels of varying v and J within the first excited electronic state to levels in the ground electronic state is known as the *Lyman band*, while the transitions between the second excited electronic and ground states fall within the *Werner band*. During the entire decay process, known as *fluorescence*, lines are emitted with wavelengths ranging from the ultraviolet to the infrared. Observers have detected many of these in hot cloud regions. In order to dissociate, the molecule must decay from an excited electronic state to a vibrational continuum level lying above the $v = 14$ level in the ground state. (See Figure 5.5.) After dissociation, the molecule's energy surplus goes into both radiation and kinetic energy of the constituent atoms.

5.2.2 Formation Rates

Let us now return to the question of H_2 formation. The allowed radiative transitions within the ground electronic state are so slow that simple association of two free atoms is rarely productive. What is needed is a third body to absorb the released energy. Interstellar dust grains play this role. Two hydrogen atoms that have just landed on a grain wander along its surface until they encounter one another. The grain's heat capacity is so large that it can easily absorb the recombination energy without an appreciable rise in temperature.

To quantify the H_2 formation rate, consider a gas containing $n_{\rm HI}$ hydrogen atoms per unit volume. Suppose these have a thermal velocity $V_{\rm therm}$. Then a single grain with geometric cross section σ_d is struck by an atom, on average, within the collision time $t_{\rm coll}$:

$$t_{\rm coll} = \left(n_{\rm HI}\,\sigma_d\,V_{\rm therm}\right)^{-1} . \tag{5.9}$$

The incident atom is attracted to the grain surface not by a chemical bond, but by the weaker Van der Waals force, with a typical binding energy of 0.04 eV. The atom rapidly explores the surface, through quantum mechanical tunneling, until it comes to rest at a lattice defect. Here the unpaired electron forms a somewhat stronger bond to the lattice, with an energy of order 0.1 eV. Within another interval $t_{\rm coll}$, a second atom lands on the grain and quickly finds a binding site adjacent to the first. Only then do the two atoms combine.

The resulting H_2 molecule has no unpaired electrons and so binds only weakly to the defect site where it formed. Thus, it soon returns to the gas phase. We may write the total H_2 formation rate per unit volume as

$$\begin{aligned} \mathcal{R}_{\rm H_2} &= \frac{1}{2}\,\gamma_H\,n_d\,t_{\rm coll}^{-1} \\ &= \frac{1}{2}\,\gamma_H\,n_d\,\sigma_d\,n_{\rm HI}\,V_{\rm therm} . \end{aligned} \tag{5.10}$$

Here n_d is the grain number density, and γ_H is the sticking probability, *i. e.*, the fraction of atoms striking a grain that eventually recombine. This probability is about 0.3 at the gas and grain temperatures of quiescent molecular clouds.

Even in gas that is completely devoid of dust grains, H_2 can still form through purely gas-phase processes, as long as the temperature and density are high enough. Assuming that the gas is at least lightly ionized, there will be a supply of free electrons. Two coupled reactions then occur:

$$\begin{aligned} &\mathrm{H} + \mathrm{e}^- \rightarrow \mathrm{H}^- + h\nu \\ &\mathrm{H}^- + \mathrm{H} \rightarrow \mathrm{H}_2 + \mathrm{e}^- \ , \end{aligned} \tag{5.11}$$

where we use $h\nu$ to denote a photon. Additionally, the ambient protons provide H_2 through

$$\begin{aligned} &\mathrm{H} + \mathrm{H}^+ \rightarrow \mathrm{H}_2^+ + h\nu \\ &\mathrm{H}_2^+ + \mathrm{H} \rightarrow \mathrm{H}_2 + \mathrm{H}^+ \ . \end{aligned} \tag{5.12}$$

Molecular hydrogen may have formed by these reactions in the early Universe, before dust grains condensed from the heavy-element debris of previous stellar generations. Because of the limited supply of free electrons and protons, only a small fraction of the atomic hydrogen could have turned molecular by this route. However, if the primordial hydrogen gas reached sufficiently high densities, then the three-body processes

$$\mathrm{H} + \mathrm{H} + \mathrm{H} \rightarrow \mathrm{H}_2 + \mathrm{H} \tag{5.13a}$$

and

$$\mathrm{H} + \mathrm{H} + \mathrm{H}_2 \rightarrow 2\,\mathrm{H}_2 \tag{5.13b}$$

may have produced the first molecular clouds.

We have mentioned that the main destruction mechanism for H_2 at lower densities is photodissociation by ultraviolet photons of energy 11.2 eV or higher. Such radiation, largely produced by O and B stars, permeates interstellar space with a flux sufficient to dissociate each molecule within a time period of about 400 years. A cloud of pure H_2, therefore, cannot exist, since all the molecules near the surface are effectively destroyed. However, this very process, along with efficient absorption of the radiation by dust grains, cuts down the ultraviolet flux until molecules further inside *can* survive. The interstellar H_2 is therefore said to be *self-shielding*. We will return to this phenomenon in Chapter 8, when we discuss the chemical makeup of star-forming clouds.

5.3 Carbon Monoxide (CO)

The simple and abundant molecule CO forms only through gas-phase reactions. Its strong binding energy of 11.1 eV then helps preserve the molecule against further destructive reactions. Like H_2, therefore, CO is self-shielding in the ambient field of ultraviolet radiation. In the outer regions of a molecular cloud, the two species build up in a similar manner, although CO remains dissociated to a greater depth. (See Chapter 8.)

It is fortunate for astrophysics that the CO molecule *does* have a permanent electric dipole moment and emits strongly at radio frequencies. Since its 1970 discovery in the Orion molecular cloud, CO has served as the primary tracer of molecular gas, both in our own and in external galaxies. The most abundant isotope, $^{12}C^{16}O$, is naturally the easiest to detect, but $^{13}C^{16}O$, $^{12}C^{18}O$, and occasionally $^{12}C^{17}O$ and $^{13}C^{18}O$, have also proved useful.

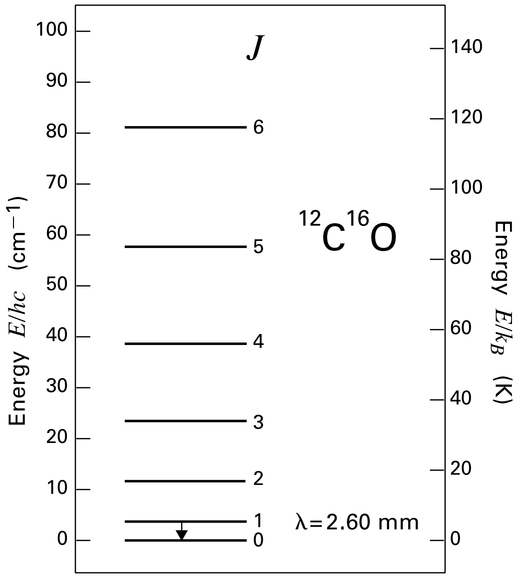


Figure 5.6 Rotational levels of $^{12}C^{16}O$ within the ground ($v = 0$) vibrational state. The astrophysically important $J = 1 \rightarrow 0$ transition at 2.60 mm is shown.

5.3.1 Populating the Rotational Ladder

The rotational levels again have energies given by equation (5.6). These levels are more closely spaced than for H_2 because the moment of inertia is greater. (See Figure 5.6.) More importantly, the faster electric dipole transitions can now occur. Here, J changes by ± 1. The $J = 1$ state is elevated above the ground state in $^{12}C^{16}O$ by $\Delta E_{10} = 4.8 \times 10^{-4}$ eV, or by an equivalent temperature of only 5.5 K. It is therefore easy to excite this level inside a quiescent cloud and even to populate $J = 2$, which lies 16 K above the ground state. When the $J = 1 \rightarrow 0$ transition is made radiatively, the emitted photon has a wavelength of 2.60 mm.

Within a molecular cloud, excitation of CO to the $J = 1$ level occurs primarily through collisions with the ambient H_2. In a cloud of relatively low total number density $n_{\rm tot}$, each upward transition is followed promptly by emission of a photon. Conversely, when $n_{\rm tot}$ is high, the excited CO usually transfers its excess energy to a colliding H_2 molecule, with no emission of a photon. The critical density separating the two regimes is given by A_{10}/γ_{10}. Using $A_{10} = 7.5 \times 10^{-8}\ {\rm s}^{-1}$ and $\gamma_{10} = 2.4 \times 10^{-11}\ {\rm cm}^3\ {\rm s}^{-1}$ (the appropriate value for collisions with H_2 at a temperature of 10 K), we find that $n_{\rm crit}$ is $3 \times 10^3\ {\rm cm}^{-3}$.

In general, the rate of spontaneous photon emission per unit volume from the $J = 1 \rightarrow 0$ transition is $n_1 A_{10}$. Here n_J denotes the number density of CO molecules in the level with quantum number J. We may determine these populations by balancing the rates of collisional and radiative excitation and deexcitation. Appendix B presents the simplified but instructive example of a two-level system. The ratio of the densities n_1 and n_0 is usually expressed through the *excitation temperature*, $T_{\rm ex}$. We define this quantity through a generalization of Boltzmann's

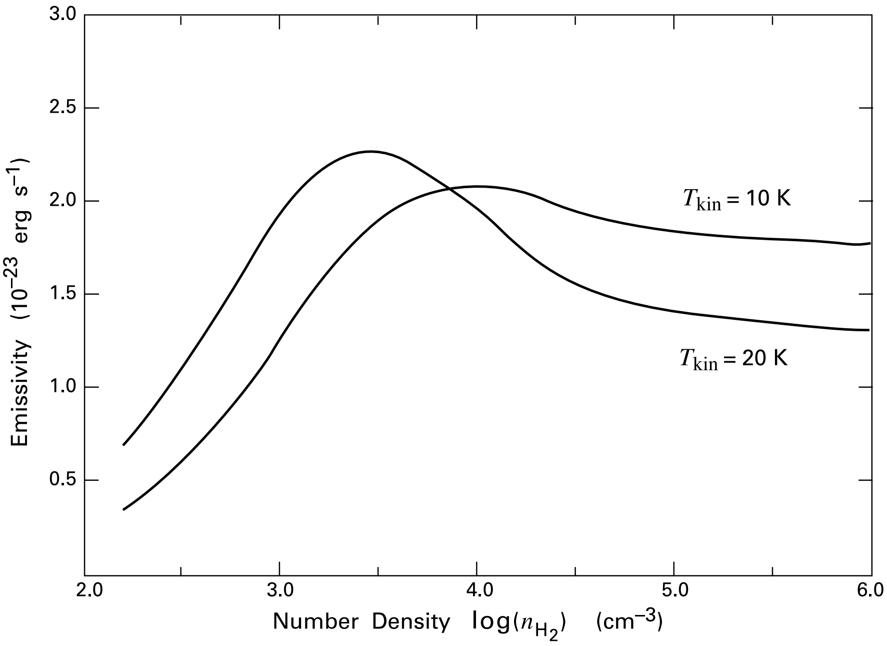


Figure 5.7 Emission in the $J = 1 \rightarrow 0$ line of ^{12}C^{16}O. The CO is immersed in a gas of pure H_2. Emission is measured per CO molecule and is displayed as a function of H_2 number density.

law:

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} \exp\left(-\frac{\Delta E_{10}}{k_B T_{ex}}\right) , \tag{5.14}$$

where g_1 and g_0 are the degeneracies of the two levels. In the case of CO rotational states, $g_J = 2J+1$. For $n_{tot} \ll n_{crit}$, n_1/n_0 is small and proportional to n_{tot}. (See Appendix B.) The excitation temperature in this case is less than T_{kin}, the *kinetic temperature* that characterizes the velocity distribution of the colliding molecules. For $n_{tot} \gg n_{crit}$, however, the CO molecule comes into *local thermodynamic equilibrium* (LTE) with its environment. The $J = 1$ and $J = 0$ level populations are again related through equation (5.14), but with T_{ex} now equal to T_{kin}.

Increasing the density in a cloud can therefore enhance the $J = 1 \rightarrow 0$ emission, but only for subcritical values of n_{tot}. We show the full behavior in Figure 5.7, which was obtained by numerical calculation of the level populations. At a fixed value of T_{kin}, the $1 \rightarrow 0$ emission rate peaks as n_{tot} increases to n_{crit} and beyond. The high-density decline is caused by the increasing excitation of molecules to the $J > 1$ states. This effect slowly drains the population of the $J = 1$ level, eventually forcing it down to the LTE value. The calculation here ignores the fact that many of the emitted $1 \rightarrow 0$ photons can themselves excite CO molecules instead of leaving the cloud. Inclusion of this *radiative trapping* would cause the peak in emission to be achieved at densities somewhat less than n_{crit}.

Figure 5.7 concerns a parcel of gas with uniform density n_{tot}. Along any line of sight through a molecular cloud, the true density will vary, with most material being at some minimum “background” value. If this ambient gas has an $n_{\rm tot}$ well below $n_{\rm crit}$ for the transition of interest, it will not contribute appreciably to the emission. On the other hand, we have seen that the emissivity declines for $n_{\rm tot}$ far above $n_{\rm crit}$, where there is often little material in any case. The conclusion, valid beyond the specific case of CO, is that *observations in a given transition are most sensitive to gas with densities near the corresponding* $n_{\rm crit}$. This fact should be borne in mind when interpreting molecular line studies.

5.3.2 Vibrational Band Emission

In gas that is being heated by nearby young stars, the upper vibrational levels of CO become significantly populated. Equation (5.8) still gives the level energies to fair accuracy, but understanding the observed complex spectra requires that we consider the next higher correction. Specifically, the picture of CO oscillating in a parabolic potential well must break down at large amplitude, where the molecule is eventually torn apart. Thus, $E_{\rm vib}$ does not increase precisely as $v + 1/2$, but contains a relatively small negative term proportional to $(v + 1/2)^2$. As a consequence, the frequencies of photons emitted by the transitions with $v = 1 \rightarrow 0,\ 2 \rightarrow 1,\ etc.$, decrease slowly with the starting v-value. Such *fundamental* vibrational transitions have the largest A-values. Other *overtone* transitions with $\Delta v = -2, -3,\ etc.$ also occur. Figure 5.8 shows the system of first overtones observed toward the BN object in Orion. Here, the vibrational levels are collisionally excited in a gas with kinetic temperature near 4000 K.

The emission spikes in Figure 5.8 are actually *bands* consisting of many closely spaced lines. Each band corresponds to a pair of vibrational quantum numbers, say v' and v'', while the lines within a band are individual rovibrotional transitions. The spacing between these lines gradually decreases toward higher frequency. Eventually, the lines merge in a *band head*. Let us see how this behavior, evident in the higher-resolution spectrum of Figure 5.9, can be understood in physical terms.

For downward dipole transitions within the same vibrational state ($J \rightarrow J - 1$), equation (5.6) predicts that $\Delta E_{\rm rot}$ is proportional to J, so that the lines are equally spaced. However, a CO molecule in a higher v-state has a slightly larger average separation between atoms. Its moment of inertia, I_v, is therefore greater and its rotational constant, B_v, is lower. If v' and v'' again denote the upper and lower vibrational states, respectively, then the $J \rightarrow J - 1$ transition now yields an energy of

$$\begin{aligned} \Delta E\,(v', J \rightarrow v'', J-1) &= \Delta E_{v'v''} + (B_{v'} + B_{v''})\,h\,J + (B_{v'} - B_{v''})\,h\,J^2 \\ &\approx \Delta E_{v'v''} + 2B_{v'}\,h\,J + (B_{v'} - B_{v''})\,h\,J^2 \ . \end{aligned} \tag{5.15}$$

Here $J = 1, 2, 3,\ etc.$ and $\Delta E_{v'v''}$ is the energy difference between two $J = 0$ states. Note that $J = 0 \rightarrow 0$ dipole transitions do not exist, so that the band contains a gap at the corresponding frequency. Since $B_{v''}$ is slightly greater than $B_{v'}$, equation (5.15) shows that the frequencies of the $J \rightarrow J - 1$ lines, known collectively as the *R-branch* of the band, increase, reach a maximum, and then begin to decline with higher J. For the fundamental $v = 1 \rightarrow 0$ band, this maximum is reached at $\lambda = 4.30\,\mu$m, while it occurs at 2.29 μm in the first overtone $v = 2 \rightarrow 0$ band. Figure 5.9 shows the $v = 2 \rightarrow 0$ band head in SSV 13, an embedded infrared source driving a molecular outflow.

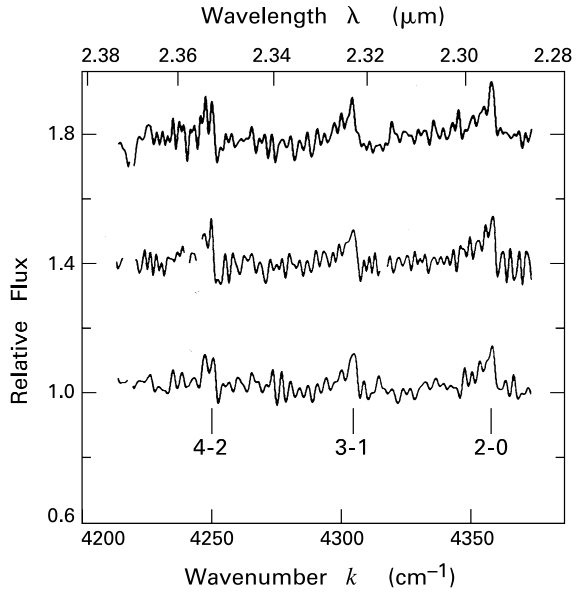


Figure 5.8 Near-infrared spectrum of the BN object in Orion, shown at three different observing times. The relative flux is plotted against the wave number k, defined here as $1/\lambda$.

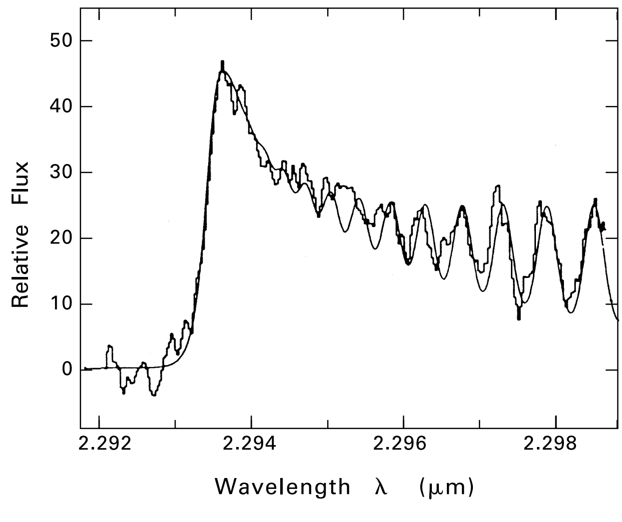


Figure 5.9 High-resolution near-infrared spectrum of the embedded stellar source SSV 13. The structure of the $v = 2 \rightarrow 0$ band head in $^{12}C^{16}O$ is evident. The smooth curve is from a theoretical model that employs an isothermal slab at 3500 K. Note that the spectrum here represents only a portion of the R-branch.

For the CO molecule, J can change by *either* -1 or $+1$ between different rovibrational states. The lines from $J-1 \rightarrow J$ fall on the other side of the central gap, which lies to the right of the spectrum shown in Figure 5.9. The energies in this *P-branch* are

$$\begin{aligned}\Delta E\,(v', J-1 \rightarrow v'', J) &= \Delta E_{v'v''} - (B_{v'} + B_{v''})\,h\,J + (B_{v'} - B_{v''})\,h\,J^2 \\ &\approx \Delta E_{v'v''} - 2B_{v'}\,h\,J + (B_{v'} - B_{v''})\,h\,J^2\ ,\end{aligned} \tag{5.16}$$

where, again, $J = 1, 2, 3,$ *etc.* Because of the sign change from equation (5.15), the frequency falls below the gap for any J. The spacing between successive lines widens, and there is no convergence to a band head.

Populating the upper electronic levels of CO requires even more energetic environments. The first excited electronic level lies at an equivalent temperature of 9.3×10^4 K above ground. Downward rovibrational transitions give rise to bands in the ultraviolet regime. Fewer lines now separate the band head from the gap, since the difference $B_{v'} - B_{v''}$ in equation (5.15) is significantly greater. If enough energy is available, CO dissociates. As in H_2, collisional dissociation is direct, while photodissociation occurs through a two-step process. That is, the molecule must first be excited by line absorption to a higher electronic level, from which it can either relax to the ground state or else fall apart into separate carbon and oxygen atoms.

5.4 Ammonia (NH_3)

We have seen how the lowest rotational levels of CO saturate in population as the cloud density climbs above $n_{\rm crit}$ for that species. The molecule then ceases to be a useful observational gauge of $n_{\rm tot}$. Additionally, we have noted that photons from a given CO molecule are absorbed and reemitted many times by other, identical molecules before leaving the cloud. In other words, the cloud becomes optically thick to this radiation, a condition that sets in first for the lowest transitions of the main isotope $^{12}C^{16}O$. The detected radiation is then only sampling conditions in the sparse, outer regions of the cloud.

Since its discovery in 1969, interstellar ammonia has been one of the most widely used probes for higher-density molecular regions. Formed through a network of gas-phase reactions, the polyatomic NH_3 has a much more complex set of transitions than CO and is therefore a more sensitive diagnostic of cloud conditions. Many useful transitions fall within a narrow frequency range, thus greatly reducing relative errors in instrumental calibration.

5.4.1 The Symmetric Top

Let us consider first the rotational transitions of this molecule, a pyramid in which the hydrogen atoms form an equilateral triangle (Figure 5.10). In classical mechanics, the kinetic energy of a three-dimensional rotator is found from a generalization of equation (5.5):

$$E_{\rm rot} = \frac{J_A^2}{2\,I_A} + \frac{J_B^2}{2\,I_B} + \frac{J_C^2}{2\,I_C}\ . \tag{5.17}$$

Here, I_A, I_B, and I_C are the moments of inertia about the principal axes of rotation, while J_A, J_B, and J_C are the corresponding projections of the total angular momentum vector $\boldsymbol{J}$. The

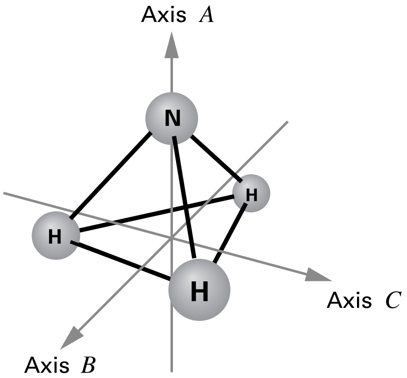


Figure 5.10 Molecular structure of NH_3, showing the three principal axes of rotation.

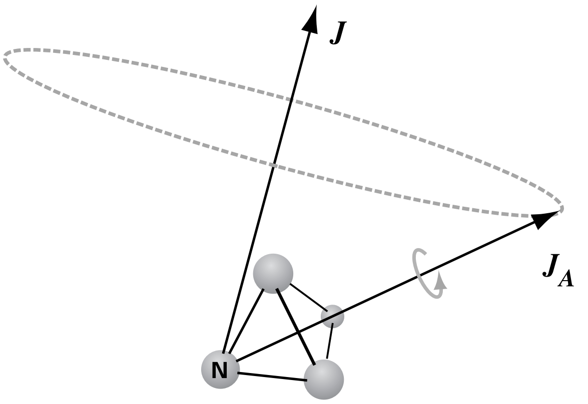


Figure 5.11 Torque-free motion of NH_3. The molecule rotates about the axes perpendicular to the plane of hydrogen atoms, with associated angular momentum $\boldsymbol{J}_A$. The latter vector in turn precesses about the axis lying along the total angular momentum $\boldsymbol{J}$.

principal axes for NH_3 are also shown in Figure 5.10. The molecule is a *symmetric top*, where two of the axes, here labeled B and C, have identical moments of inertia. Note that $I_A < I_B$, *i. e.*, the molecule is a prolate (as opposed to oblate) rotator. In the absence of external torques, a symmetric top rotates about its symmetry (A) axis, which in turn precesses about the fixed vector $\boldsymbol{J}$ (Figure 5.11). The scalars J and J_A are therefore both constants of the motion.

To find the quantum mechanical energy levels, we first use the equality of I_B and I_C to rewrite equation (5.17) as

$$E_{\rm rot} = \frac{J^2}{2\,I_B} + J_A^2\left(\frac{1}{2\,I_A} - \frac{1}{2\,I_B}\right) \ . \tag{5.18}$$

Since J and J_A are conserved classically, the generalization of equation (5.6) is

$$E_{\rm rot} = B\,h\,J\,(J+1) + (A-B)\,h\,K^2 \ . \tag{5.19}$$

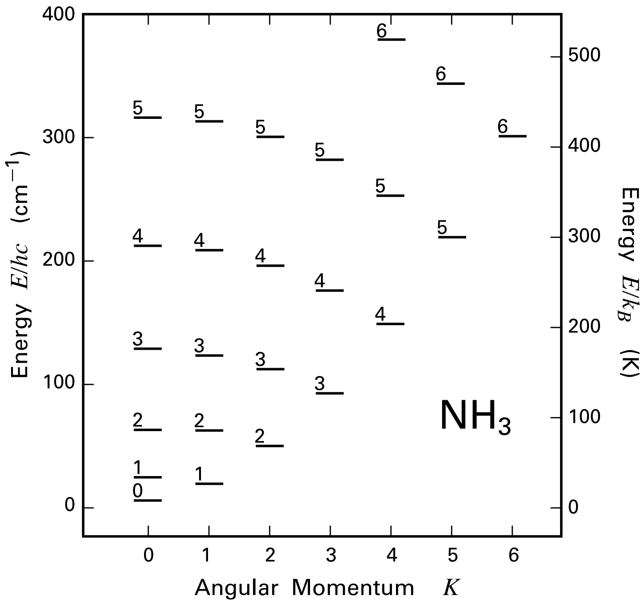


Figure 5.12 Rotational levels of NH$_3$. Those levels lying on the lower border constitute the rotational backbone.

Here, A and B are the two rotational constants, while J and K are quantum numbers that measure, respectively, the magnitude of the total angular momentum and its component along the symmetry axis. For a given value of J, the possible K-values range from $-J$ to $+J$. Since, according to (5.19), states with $\pm K$ have the same energy, it is conventional, when labeling states, to restrict K to values greater than or equal to 0.

The total set of rotational levels is conveniently arranged into columns of fixed K-value (Figure 5.12). Within a given column, the state of lowest energy has $J = K$. Downward transitions from (J, K) to $(J - 1, K)$ occur very rapidly, with typical A-values from 10^{-2} to 10^{-1} s^{-1}. By symmetry, the molecule's electric dipole moment vector $\boldsymbol{\mu}$ lies along the central axis. Classically, rotation about this axis can therefore produce no dipole radiation. Correspondingly, quantum-mechanical dipole transitions with nonzero ΔK are forbidden. States along the lower border in the diagram consitute the *rotational backbone*. Downward quadrupole transitions along this border, $(J, K) \rightarrow (J-1, K-1)$, do occur, but with A-values of order 10^{-9} s^{-1}. Thus, the states along the backbone are *metastable*.

5.4.2 Inversion Lines

The most useful spectral lines from NH$_3$, as we shall demonstrate in Chapter 6, arise from *inversion*, the oscillation of the nitrogen atom through the hydrogen plane. In most molecules, vibrational transitions yield infrared photons of much higher frequency than those from rotational modes. The inversion transition of NH$_3$, however, produces microwave photons, in contrast to the far-infrared rotational lines. The reason is that, from the classical viewpoint, the nitrogen atom does not have sufficient energy to cross the central plane, *i.e.*, there exists a barrier in the potential well (Figure 5.13). In quantum mechanics, the atom's wavefunction can *tunnel* through this barrier, given sufficient time. Oscillation thus occurs, but at a far lower rate than in a simple parabolic potential well. To produce the low-frequency emission, each rotational level (J, K) with $K > 0$ is split into two sublevels with an energy separation of order 10^{-4} eV (Figure 5.14). The transition from the upper to the lower sublevel yields the *main line* of the NH_3 microwave spectrum. For the (1,1) state, this line has a wavelength of 1.27 cm.

Additional effects further split the two inversion states. The nitrogen nucleus has a non-spherical charge distribution and an electric quadrupole moment. It can therefore be torqued in the presence of an electric field gradient. The system energy depends on the relative orientation of the nuclear spin and the total angular momentum vector of the electrons, which in turn varies with the rotational state of the molecule. Consequently, each inversion state splits into three sublevels, as illustrated in Figure 5.14. When the appropriate selection rules are enforced, the allowed transitions between the upper and lower levels give rise to a total of five lines – the original main line and two pairs of *satellite lines*, separated from the main line by about 1 MHz.

Finally, even weaker, magnetic interactions between the spins of the various nuclei split the lines again, with typical separations of 40 KHz. The net result is that the observationally important (1,1) and (2,2) rotational states each produce a total of 18 lines. The original main line, now split into 8 closely spaced components, has about half the total intensity of the inversion transition, while each cluster of satellite lines carries roughly equal intensity.

5.5 Water (H_2O)

The water molecule has a relatively large dipole moment, almost 20 times that of CO. Here, the associated vector $\boldsymbol{\mu}$ is directed along the symmetry axis through the oxygen atom (Figure 5.15). Within the ground vibrational state are a large number of allowed rotational transitions at far-infrared and millimeter wavelengths. Excitation of these levels followed by prompt radiative decay provides an important cooling mechanism in shock-heated clouds, where accelerated chemical reactions produce a relatively high abundance of the molecule. Absorption by atmospheric water is an impediment that has prevented H_2O from becoming a primary diagnostic of cloud conditions. Spaceborne observations, first by the SWAS satellite, have yielded an upper bound to the H_2O abundance in quiescent clouds (Table 5.1). A number of higher transitions had previously been detected, starting with the important 22.2 GHz (1.35 cm) line discovered in 1969. This line, along with others subsequently found, are actually *maser* transitions, in which an enhanced population of the upper state creates extraordinarily strong emission. We will discuss water masers and their application in Chapter 14.

5.5.1 The Asymmetric Top

The rotational emission spectrum of H_2O is more complex than that of NH_3 because the molecule is an *asymmetric top*, with three unequal moments of inertia along its principal axes (Figure 5.15). Classically, the quantities conserved during rotation are the total vector angu-

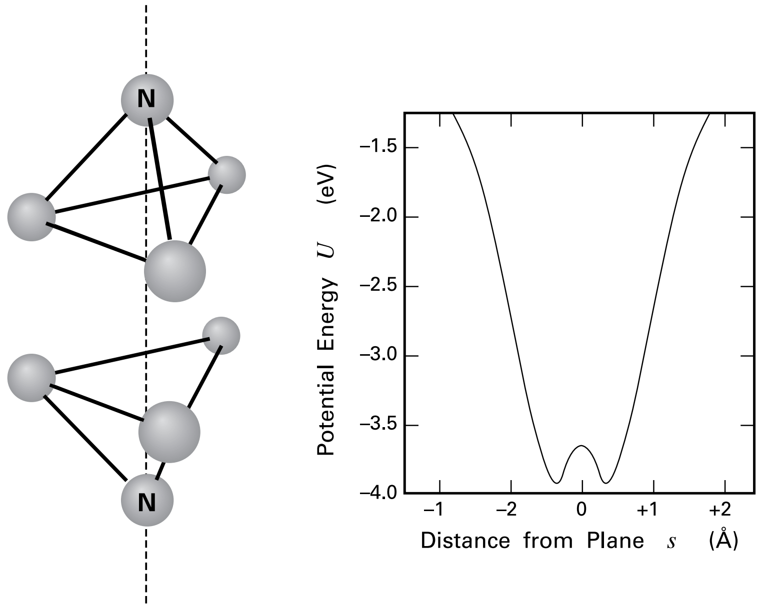


Figure 5.13 *left panel*: The inversion of NH_3, as the nitrogen atom tunnels through the plane of hydrogens. *right panel*: The potential energy of the molecule, shown as a function of the nitrogen atom's distance from the hydrogen plane. Note the central energy barrier.

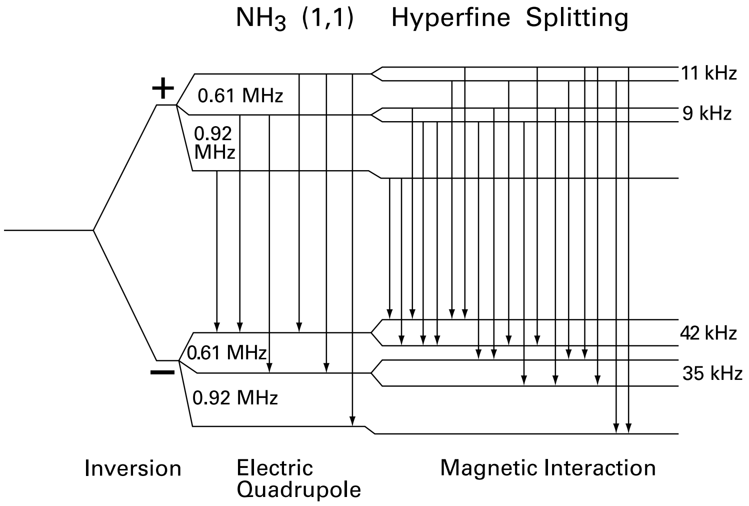


Figure 5.14 Splitting of the inversion line in the NH_3 (1,1) state. The various frequency differences are indicated, along with allowed transitions.

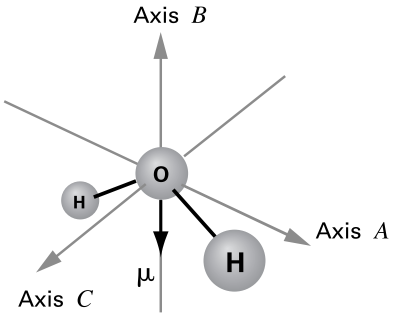


Figure 5.15 Molecular structure of H_2O, showing the three principal axes and the electric dipole moment $\boldsymbol{\mu}$.

lar momentum $\boldsymbol{J}$ and its projection along an axis fixed in inertial space, but *not* the projection along any axis tied to the molecule itself. Thus there is no second quantum number beyond J to parametrize the rotational energy. This energy may be expressed as a complicated function of J and the rotational constants A, B, and C, which correspond to the three principal axes in Figure 5.15. It is conventional to order the constants so that $A > B > C$. Since B is numerically closer in value to C than to A, the molecule is more prolate than oblate.

Generalizing from the J_K notation for symmetric tops, the rotational states are labeled with three numbers in the form J_{K_{-1},K_1}. The first subscript is the K-value of that *prolate* symmetric top state obtained if the rotational constant B were changed to C. Similarly, K_1 is the subscript for the *oblate* configuration created by letting B tend toward A.[2] The dipole selection rules allow J to change by 0 or ± 1, while K_{-1} or K_1 can each change by ± 1 or ± 3. The so-called eo states, *i. e.*, those with even K_{-1} and odd K_1, can only change into oe states and vice versa, while ee and oo states are similarly linked. Physically, these two separate classes are distinguished by the sign change of the molecular wave function under a 180° rotation about the symmetry axis.

Any given rotational state has a number of equal-energy sublevels corresponding to different orientations of the spins of the two hydrogen nuclei. This degeneracy is three times greater for the eo and oe ("ortho") class of states, which are correspondingly more populated than the ee and oo ("para") class. The energy-level diagram of Figure 5.16 displays separately the lower ortho- and para-rotational states. These are arranged so that states with a common J-value occupy the same column.

5.5.2 Observed Rotational Lines

The water molecule's large dipole moment implies that many downward transitions have relatively high A-values, particularly those between levels with the same J and neighboring K-values. As a consequence, it is rather difficult to excite the higher levels collisionally. For example, the $1_{10} \rightarrow 1_{01}$ transition has $A = 3.5 \times 10^{-3}$ s^{-1} and a collisional deexcitation rate of 2.0×10^{-10} cm^{-3} s^{-1} at $T_{\rm kin} = 20$ K. The corresponding value of $n_{\rm crit}$ is 2×10^7 cm^{-3}. This is far greater than densities in quiescent clouds, but attainable in shocked regions near massive

---

[2] The rather cumbersome notation for the subscripts reflects the fact that a standard parameter measuring molecular asymmetry tends toward -1 for prolate configurations and toward $+1$ for oblate structures.

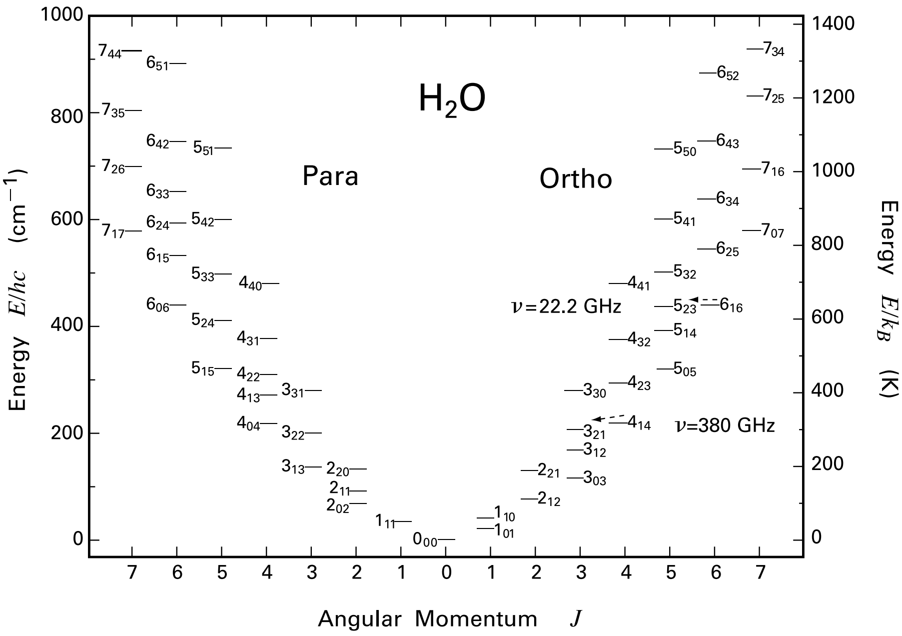


Figure 5.16 Rotational levels of H_2O. Two maser transitions, $6_{16} \rightarrow 5_{23}$ and $4_{14} \rightarrow 3_{21}$, are indicated on the right.

stars. At lower densities, rotational levels can still be excited by infrared continuum photons from heated dust grains.

Whatever their source of energy, the excited molecules tend to cascade downward to the rotational backbone, consisting of the states 1_{01}, 2_{12}, 3_{03}, *etc.* Inspection of Figure 5.16 reveals only two allowed transitions among the ortho-states, $4_{14} \rightarrow 3_{21}$ and $6_{16} \rightarrow 5_{23}$, by which molecules can leave the backbone without dropping directly to the next lower backbone level. The levels 3_{21} and 5_{23} thus accumulate molecules, but remain underpopulated relative to their adjacent backbone states. In fact, both transitions in question are observed as masers, the second being the 22.2 GHz line.

Like CO and NH_3, H_2O forms through gas-phase reactions. Within shocked cloud regions, the major sequence is

$$\begin{aligned} H_2 + O &\rightarrow OH + H \\ OH + H_2 &\rightarrow H_2O + H \; . \end{aligned} \tag{5.20}$$

As we noted in §5.1, such neutral-neutral reactions are inoperative within cold, dark clouds.

Here, ion-molecule reactions can still proceed. The formation of H_2O involves the chain

$$\begin{aligned} \mathrm{H_3^+} + \mathrm{O} &\rightarrow \mathrm{H_2} + \mathrm{OH^+} \\ \mathrm{OH^+} + \mathrm{H_2} &\rightarrow \mathrm{OH_2^+} + \mathrm{H} \\ \mathrm{OH_2}^+ + \mathrm{H_2} &\rightarrow \mathrm{H_3O^+} + \mathrm{H} \\ \mathrm{H_3O^+} + \mathrm{e^-} &\rightarrow \mathrm{H_2O} + \mathrm{H} \ . \end{aligned} \tag{5.21}$$

The last reaction is an example of the dissociative recombination process introduced in equation (5.2). Such reactions proceed very quickly given an adequate supply of free electrons.

5.6 Hydroxyl (OH)

We saw in Chapter 3 how molecular clouds are threaded by the magnetic field lines that permeate interstellar space. Compression of this field creates an effective pressure that partially supports clouds against gravitational collapse. To measure $\boldsymbol{B}$ accurately, a molecular probe should have a relatively large magnetic moment. Especially important in this regard are molecules with one unpaired electron and therefore a nonzero electronic angular momentum. Classified chemically as *free radicals*, such compounds are violently reactive in the laboratory, but can survive for long periods in the rarefied environments of molecular clouds. The most widely used species of this kind is OH.

5.6.1 Nature of Rotational Motion

The rotational levels of OH again have energies given by equation (5.6). However, it is the more finely spaced hyperfine transitions that have found application in magnetic field measurements. A given rotational state, again labeled by the quantum number J, is split into two sublevels of nearly equal energy by a phenomenon known as Λ-*doubling*. Here, $\Lambda\hbar$ is the projection of the orbital angular momentum of the unpaired electron along the molecule's internuclear axis. The molecule is rotationally symmetric about this axis, so there are no torques to alter the corresponding component of the angular momentum. It follows that Λ is a valid quantum number. Since a state with $-\Lambda$, *i. e.*, with the orbital motion reversed, has almost the same energy as the $+\Lambda$ state, the label Λ is conventionally restricted to nonnegative integer values. The projection of the electronic spin angular momentum along the internuclear axis is another good quantum number, denoted Σ. Since OH has only one unpaired electron, Σ is restricted in value to $\pm 1/2$. The projection of the electron's total angular momentum, a quantity denoted Ω, is given by $|\Lambda + \Sigma|$.

We emphasize that it is only the *axial projections* of the electronic angular momenta that are constants of the motion. In a semi-classical description, the spin and orbital angular momentum vectors, denoted $\boldsymbol{S}$ and $\boldsymbol{L}$ respectively, are *not* fixed in space, but undergo a complex motion (Figure 5.17). First, the unpaired electron is attracted toward the internuclear axis by the powerful electrostatic force of the chemical bond. The resulting torque causes $\boldsymbol{L}$ to precess rapidly about this axis. Second, the electron sees, in its own reference frame, a magnetic field

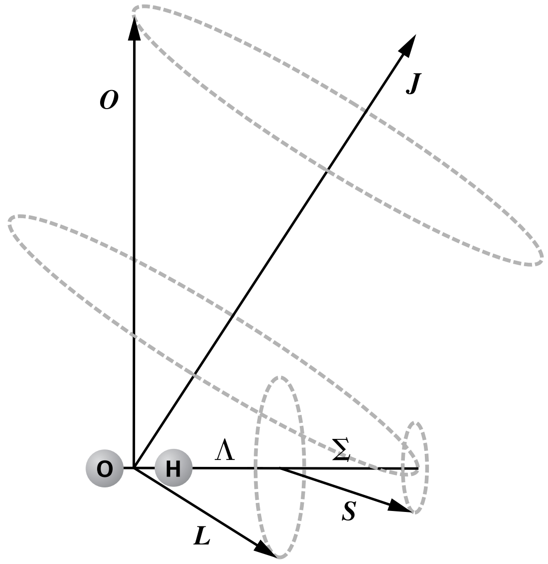


Figure 5.17 Torque-free motion of OH. The vectors $\boldsymbol{L}$ and $\boldsymbol{S}$, representing respectively the unpaired electron's orbital and spin angular momentum, have projections Λ and Σ along the internuclear axis. Both $\boldsymbol{L}$ and $\boldsymbol{S}$ precess rapidly about this axis. Meanwhile, the internuclear axis itself, and the associated nuclear angular momentum $\boldsymbol{O}$, precess slowly about the total angular momentum $\boldsymbol{J}$.

from the motion of the nuclei and remaining electrons. The field creates a torque on the magnetic moment associated with $\boldsymbol{S}$. (Recall the discussion of the hydrogen atom in § 2.1.) As a result of this spin-orbit coupling, $\boldsymbol{S}$ also precesses quickly about the internuclear axis. Finally, the axis itself tumbles slowly end over end, through rotation of the O and H nuclei. Note that the angular momentum associated with this rotation, denoted $\boldsymbol{O}$, lies perpendicular to the axis, since the atomic nuclei have negligible moments of inertia about the line joining them.

Regardless of these internal torques, the angular momentum $\boldsymbol{J}$, formed by adding vectorially $\boldsymbol{S}$, $\boldsymbol{L}$, and $\boldsymbol{O}$, is very nearly constant in magnitude and direction. We may picture the vector $\boldsymbol{O}$ and the projection of $\boldsymbol{S} + \boldsymbol{L}$ along the internuclear axis both precessing about the fixed $\boldsymbol{J}$ (Figure 5.17). The motion is analogous to that of a symmetric top. (Recall Figure 5.11.)[3] Thus, J is another good quantum number, whose possible values are given by Ω, $\Omega + 1$, $\Omega + 2$, *etc.* This sequence is a generalization of $J = 0, 1, 2,$ *etc.* for molecules without electronic angular momentum.

The rotational states of the molecule are thus labeled, not only by J, but also by the quantum numbers Λ and Ω. In addition, one must specify *S*, the magnitude of the electronic spin. This number, fixed for all rotational states, is here equal to 1/2. In spectroscopic notation, the ground state for OH is symbolized ${}^2\Pi_{3/2}$ and has $J = 3/2$. The Π denotes $\Lambda = 1$, while the subscript is Ω. Since $\Sigma = +1/2$ for this state, $\Omega = |1+1/2| = 3/2$. Figure 5.18 shows the rotational ladder of ${}^2\Pi_{3/2}$ levels with $J = 3/2, 5/2, 7/2,$ *etc.* The superscript in the spectroscopic symbol is the *multiplicity*, equal to $2S + 1$. In this case, the multiplicity of 2 indicates that there exists another state with the same Λ, but with the unpaired electron's spin oriented oppositely, *i. e.*, with $\Sigma = -1/2$ and therefore $\Omega = 1/2$. Any state with these values of Λ and Ω is denoted

---

[3] The precession of angular momentum vectors described here applies only to the lower rotational states of OH. In states of higher J, the spin-orbit coupling can no longer effectively lock $\boldsymbol{S}$ to $\boldsymbol{L}$ directly.

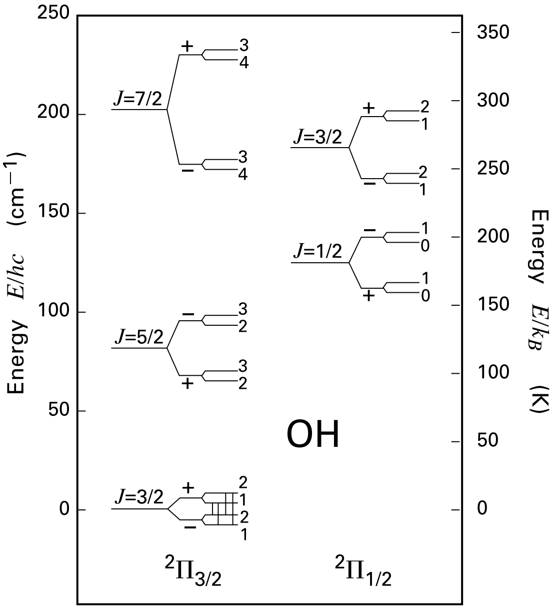


Figure 5.18 Rotational states of OH. The two ladders correspond to opposite orientations of the unpaired electron's spin. The splitting of the levels due to both Λ-doubling and the magnetic hyperfine interaction is shown schematically. Also indicated are the allowed transitions within the ground rotational state.

$^2\Pi_{1/2}$; the lowest one has $J = 1/2$. Figure 5.18 includes the separate ladder of $^2\Pi_{1/2}$ states with $J = 1/2$, $3/2$, $5/2$, *etc.*

5.6.2 Λ-Doubling

Focusing again on an individual rotational state, we saw how the motion of the internuclear axis causes the axial projection of $\boldsymbol{S} + \boldsymbol{L}$ to precess about $\boldsymbol{J}$. Such precession by itself would not affect the molecule's energy. However, the nuclear rotation also distorts slightly the electron's orbital motion. Figure 5.19 depicts two orthogonal probability distributions for the electron in the plane perpendicular to the internuclear axis. Because of the molecule's symmetry about the axis, the two cases (a) and (b) might be expected to have identical energy. However, the molecule as a whole rotates about the perpendicular axis indicated in the figure. This rotation induces a centrifugal force, so that the distribution in (a) has higher energy than in (b), where the electron is, on average, farther from the rotation axis.

This splitting of the previously degenerate $\pm\Lambda$ levels constitutes Λ-doubling. Note that the actual wavefunctions for the two eigenstates (known as the $+$ and $-$ states) are linear combinations of those corresponding to orthogonal directions of electronic rotation. Since the nuclear rotation of the molecule is very slow compared to the electron's orbital speed, the resulting perturbations to the energy are slight. Figure 5.18 shows schematically that the energy split increases going up the rotational ladder to states of higher J, *i. e.*, to faster molecular rotation. For the $^2\Pi_{3/2}$ $(J = 3/2)$ state, the temperature equivalent of the energy difference is 8.0×10^{-2} K. A photon emitted during the transition between sublevels has a frequency of about 1700 MHz and a wavelength of 18 cm.

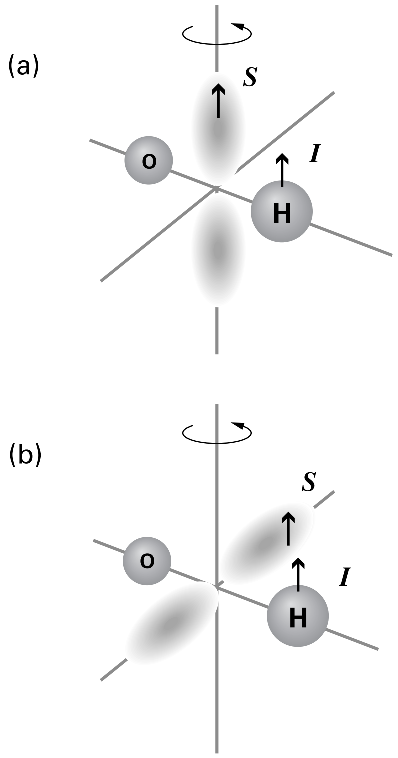


Figure 5.19 Physical origin of Λ-doubling in OH. The molecule's energy depends on whether the symmetry axis of the unpaired electron's orbital motion (a) coincides with the internuclear axis, or (b) lies orthogonal to that axis.

5.6.3 Magnetic Hyperfine Splitting

Each of the sublevels of the ${}^2\Pi$ states is further split by an interaction between the spins of the unpaired electron and the hydrogen nucleus. With reference to Figure 5.17, the result is that even $\boldsymbol{J}$ is not strictly constant, but precesses with $\boldsymbol{I}$, the proton angular momentum, about their sum, the grand total angular momentum $\boldsymbol{F}$. The new interaction arises because the magnetic field from the spinning electron at the position of the proton depends on the relative orientation of the two spin axes. Thus, in Figure 5.19, the electronic angular momentum $\boldsymbol{S}$ appears parallel to $\boldsymbol{I}$ in both panels. However, the magnetic field from the electron is actually parallel to $\boldsymbol{I}$ in (a) and antiparallel in (b). Reversing the direction of $\boldsymbol{I}$ would result in two other distinct states.

The net effect of this *magnetic hyperfine* splitting is quantitatively small; each sublevel of the ground ${}^2\Pi_{3/2}$ state is split in frequency by only about 60 MHz. The F-values of the final states are given in Figure 5.18. Transitions that connect states of the same F are said to produce *main* lines, while the others emit *satellite* lines. The reader may recall from Chapter 2 that the same magnetic interaction is responsible for the famous 21 cm line of atomic hydrogen. The split in energy, corresponding in that case to a frequency of 1420 MHz, is larger than in OH because the average separation of the electron and proton is less.

The short vertical line segments in Figure 5.18 indicate the allowed radiative transitions within the ${}^2\Pi_{3/2}$ ($J = 3/2$) ground state of OH. Historically, the 1963 detection of the four lines at 1612, 1665, 1667, and 1720 MHz constituted the first radio identification of an interstellar molecule. In this case, the lines appeared in absorption against the radio source Sgr A*. Their relative intensities were close to the theoretically predicted values in an optically-thin medium, with the strongest lines being the main ones at 1665 and 1667 MHz. Soon after, observers found three of the four lines in emission, but the one at 1665 MHz was far brighter than the others. This line, initially dubbed "mysterium," was also much narrower than expected and strongly variable. What was seen, in fact, was the first instance of an interstellar maser.

Chapter Summary

Even in the cold environments of dark clouds, one finds a great variety of molecules. Most are created when an ion of one species polarizes a nearby, electrically neutral atom, increasing their mutual attraction. A significant exception is H_2 itself, which forms on the surfaces of dust grains. It is puzzling that these grains do not promptly sweep up other molecules within a cloud and remove them from the gas phase. Apparently, there are processes that liberate molecules from grains, but they are poorly understood at present.

Rotational, vibrational, and electronic transitions of molecules require increasing energy for excitation. A given transition radiates strongly when the ambient density is critical, *i. e.*, just high enough to offset radiative decay through collisonal pumping. In most molecular clouds, H_2 is undetectable, so observers principally utilize the rotational lines of CO. Denser cloud material may be traced through a complex of microwave NH_3 lines that arise from quantum mechanical tunneling of the nitrogen atom. The high temperatures and densities created by shocks can overpopulate rotational levels of H_2O, leading to maser emission. Finally, rotation of OH subtly affects the motion of this molecule's unpaired electron. The resulting quartet of lines near 18 cm is a valuable diagnostic of interstellar magnetic fields.

Suggested Reading

Section 5.1 Two sources of general interest are

Hartquist, T. W. & Williams, D. A. 1995, *The Chemically Controlled Cosmos* (Cambridge: Cambridge U. Press)

Van Dishoek, E. F. & Blake, G. A. 1998, ARAA, 36, 317.

The first is a broad survey of interstellar chemistry, while the second reviews the molecular composition of star-forming regions. For thorough, physically oriented discussions of the more important transitions, the reader should consult

Townes, C. H. & Schawlow, A. L. 1975, *Molecular Spectroscopy* (New York: Dover)

Gordy, W. & Cook, R. L. 1984, *Microwave Molecular Spectra, Techniques of Chemistry, Vol. XVIII* (New York: Wiley).

Section 5.2 The first detection of H_2 in space was by

Carruthers, G. R. 1970, ApJ, 161, L81.

The theory of its formation on grain surfaces is due to

Hollenbach, D. & Salpeter, E. E. 1971, ApJ, 163, 155.

The molecule has been observed in a variety of astrophysical settings. A useful review is

Shull, J. M. & Beckwith, S. 1982, ARAA, 20, 163.

This article does not include the later observations of H_2 in stellar jets; see Chapter 13.

Section 5.3 The discovery of interstellar CO was announced by

Wilson, R. W., Jefferts, K. B., & Penzias, A. A. 1970, ApJ, 161, L43,

in the same issue of The Astrophysical Journal as that of H_2. Many astrophysically interesting properties of the molecule are covered in

van Dishoek, E. F. & Black, J. H. 1987, in *Physical Processes in Interstellar Clouds*, ed. G. E. Morfill and M. Scholer (Dordrecht: Reidel), p. 241.

Section 5.4 Interstellar NH_3 was first found in the direction of Sagittarius A:

Cheung, A. C., Rank, D. M., Townes, C. H., Thornton, D. D., & Welch, W. J. 1968, Phys. Rev. Lett., 21, 1701.

Subsequent observations of the molecule are reviewed by

Ho, P. T. P. & Townes, C. H. 1983, ARAA, 21, 239.

Section 5.5 The initial detection of H_2O was through its 22.2 GHz maser line:

Cheung, A. C., Rank, D. M., Townes, C. H., Thornton, D. D., & Welch, W. J. 1969, Nature, 221, 626.

For satellite observations of the ground-state rotational transition, see

Melnick, G. J. et al. 2000, ApJ, 539, L87.

Section 5.6 The the first recorded spectrum of OH was in absorption toward the Galactic center:

Weinreb, S., Barrett, A. H., Meeks, M. L., & Henry, J. C. 1963, Nature, 200, 829,

while the maser phenomenon emerged from the subsequent emission spectrum recorded by

Weaver, H., Williams, D. R. W., Dieter, N. H., & Lum, W. T. 1965, Nature, 208, 209.

The Formation of Stars
Steven W. Stahler and Francesco Palla
© 2004 WILEY-VCH Verlag GmbH & Co.

6 Molecular Transitions: Applications

Having acquired some physical understanding of the most important molecular transitions, we next examine how these lines are employed in practice. The infrared emission from H_2 requires more energetic conditions than are present in quiescent clouds. Hence, these lines are not useful as general diagnostics. Note, however, that the 2.12 μm line has proved effective in tracing shocks and stellar jets. The potentially important rotational transitions of H_2O are obscured by atmospheric absorption; some of these lines have now been detected by satellite. Of the molecules we have considered previously, the most heavily used are CO, NH_3, and OH. As we consider each species in turn, we focus on just a few of its common applications. We defer treatment of the maser emission from H_2O until Chapter 14.

6.1 Carbon Monoxide

Because its critical density is relatively low, CO has most often been used for studying massive clouds, rather than the dense cores within them. In these larger regions, the $J = 1 \rightarrow 0$ transition of $^{12}C^{16}O$ is almost always optically thick, while the same line from rarer isotopes is frequently not. The issue of optical thickness plays a key role when interpreting observations from this molecule.

6.1.1 Observed Profiles

Figure 6.1 shows three representative $J = 1 \rightarrow 0$ line profiles, all from the same region in Taurus-Auriga. As usual, we plot the antenna temperature T_A instead of I_ν itself, and the line-of-sight velocity V_r in place of the frequency ν. The $^{12}C^{16}O$ profile in the figure has a flat-topped or *saturated* appearance, the characteristic sign of optical thickness. Any photon emitted near the line center $\nu_\circ$ is very quickly absorbed by nearby $^{12}C^{16}O$ molecules. Because the absorbers have a finite relative velocity, the reemitted photons are slightly Doppler-shifted in frequency. As this process is repeated a number of times, photons diffuse in frequency into the line wings, *i. e.*, the original emission profile is broadened.

Most molecules in the cloud have a small relative speed. Thus, $^{12}C^{16}O$ photons are still optically thick over some frequency range centered on $\nu_\circ$. Within this range, the cloud radiates from its surface like a blackbody. The observed intensity I_ν is proportional to the Planckian function B_ν evaluated at the excitation temperature near the surface (see Appendix C). This function varies little over the relatively narrow frequency range in which the radiation is optically thick. Hence, the profile appears flat. Sufficiently far from line center, there are few enough absorbing molecules that the photons can escape, and the intensity drops.

For more optically-thin transitions, such as the $^{13}C^{16}O$ and $^{12}C^{18}O$ lines also shown in Figure 6.1, the profile is reduced in amplitude and more sharply peaked. In these cases, every

The Formation of Stars. Steven W. Stahler and Francesco Palla
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-40559-3

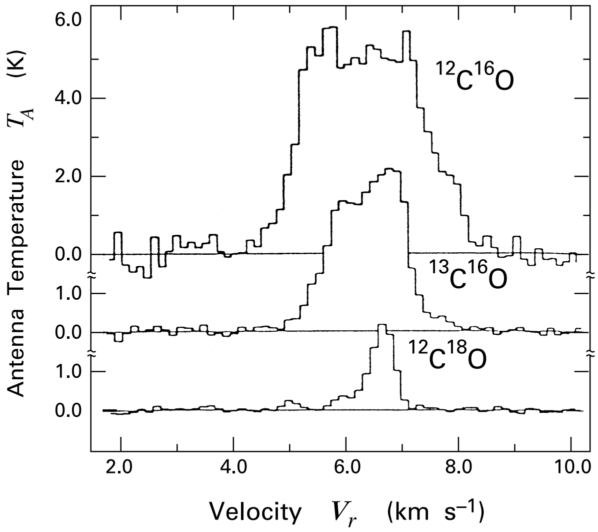


Figure 6.1 Intensity profiles of the $J = 1 \rightarrow 0$ line in three CO isotopes, observed toward Taurus-Auriga.

molecule along the line of sight contributes to the emission, so that the intensity integrated over all frequencies (or, equivalently, T_A integrated over V_r) is proportional to the total column density of the isotope in question. It is important to understand that this proportionality *cannot* hold for $^{12}C^{16}O$, whose radiation emanates only from the cloud's surface layers.

To see the matter more quantitatively, consider the *ratio* of the column densities of the two species $^{12}C^{16}O$ and $^{13}C^{16}O$. Since the oxygen nuclei are identical, this number in nearby clouds should approximately equal the terrestrial value for the ratio of carbon isotopes, which we denote as $[^{12}C/^{13}C]^*$. This ratio is measured to be 89. We do not expect the equality to be exact, because a number of chemical reactions in clouds slightly favor the rarer isotope. Since $^{13}C^{16}O$ is more tightly bound than $^{12}C^{16}O$ by 3.0×10^{-3} eV, *i. e.*, by an equivalent temperature of 35 K, such *chemical fractionation* is significant at molecular cloud temperatures. In warmer clouds, the inferred values of $[^{12}C/^{13}C]$ *do* fall reasonably close to the terrestrial figure, although there is evidence for a systematic decline toward the Galactic center. For the example shown in Figure 6.1, however, the measured ratio of $\int T_A dV_r$ for the two lines is only 2.2, much too small to be caused by chemical fractionation alone. The true explanation is that, for the optically-thick $^{12}C^{16}O$ emission, $\int T_A dV_r$ does not trace the full column density, which must be inferred by other means.

6.1.2 Temperature and Optical Depth

The general problem we are addressing is how to use the received intensity of any spectral line to deduce the physical conditions within a cloud. Let us focus first on temperature and density as the quantities of interest. A key relation, derived in Appendix C, is the *detection equation*:

$$T_{B_\circ} = T_\circ \left[f(T_{\rm ex}) - f(T_{\rm bg})\right] \left[1 - \exp(-\Delta\tau_\circ)\right] \ . \tag{6.1}$$

Here, $\Delta\tau_\circ$ and $T_{B_\circ}$ are, respectively, the cloud's optical thickness at line center and the received *brightness temperature* at the same frequency. As discussed in Appendix C, $T_B(\nu)$ is related to the directly observed $T_A(\nu)$ through the beam efficiency and the beam dilution factor. The quantity $T_{\rm bg}$ in (6.1) is the blackbody temperature associated with any background radiation field, assumed here to be approximately Planckian in its energy distribution. Finally, the function $f(T)$ is defined by

$$f(T) \equiv \left[\exp(T_\circ/T) - 1\right]^{-1} \ . \tag{6.2}$$

Here $T_\circ$ is the equivalent temperature of the transition, *i. e.*, $T_\circ \equiv h\nu_\circ/k_B$.

Let us apply the detection equation to the profiles in Figure 6.1. For this particular observation, both the beam efficiency and dilution factor happen to be close to unity, so that $T_B \approx T_A$. Consider first the ^{13}C^{16}O line. While $T^{13}_{B_\circ}$ is known directly from the observed profile, equation (6.1) still contains, beside $T_{\rm bg}$, the two unknowns $T^{13}_{\rm ex}$ and $\Delta\tau^{13}_\circ$; here we use the superscript to specify the carbon isotope. Interpretation of the ^{13}C^{16}O profile clearly requires more information. For the optically-thick ^{12}C^{16}O line, however, it is generally true that $\Delta\tau^{12}_\circ \gg 1$, so that the rightmost factor in (6.1) reduces to unity. We then have, to a good approximation,

$$T^{12}_{B_\circ} = T^{12}_\circ \left[f(T^{12}_{\rm ex}) - f(T_{\rm bg})\right] \ . \tag{6.3}$$

Assuming the radiation source behind the cloud to be the cosmic microwave background, we set $T_{\rm bg}$ equal to 2.7 K. We also know that $T^{12}_\circ$ is 5.5 K. We may now use equation (6.3) to solve for $T^{12}_{\rm ex}$ in terms of the observed quantity $T^{12}_{B_\circ}$. For our profile, $T^{12}_{B_\circ} = 5.8$ K, so we find that $T^{12}_{\rm ex} = 9.1$ K. The ^{12}C^{16}O line is usually so optically thick that the $J = 0$ and $J = 1$ level populations can be taken to be in LTE, even if the ambient density is less than $n_{\rm crit}$ (see § 6.2 below). We therefore have a measure of the interior kinetic temperature:

$$T_{\rm kin} = T^{12}_{\rm ex} \ . \tag{6.4}$$

It is also of interest to evaluate the optical thickness of the cloud to the observed spectral lines. In the case of ^{13}C^{16}O, its collisional and radiative transition rates per molecule, and therefore also its value of $n_{\rm crit}$, are very close to those of ^{12}C^{16}O. At a fixed ambient density $n_{\rm tot}$, one must obtain $T_{\rm ex}$ by considering not only $n_{\rm crit}$ but also the local radiation intensity (see Appendix B). In the absence of a detailed model, it is difficult to assess the amount of radiative trapping, but it is generally safe to take the lower levels of ^{13}C^{16}O to be in LTE if ^{12}C^{16}O is very optically thick. The relative populations in such levels for the two isotopes are then the same, so that we have

$$T^{13}_{\rm ex} = T^{12}_{\rm ex} \ . \tag{6.5}$$

This equality finally allows us to apply the detection equation to ^{13}C^{16}O. With $T^{13}_{B_\circ}$ known from the observation to be 4.1 K, and $T^{13}_\circ = 5.3$ K, we find $\Delta\tau^{13}_\circ = 1.2$. The fact that the ^{13}C^{16}O line is only marginally optically thin is consistent with the profile's appearance in Figure 6.1, where we can see the beginning of saturation broadening.

To estimate the optical thickness of the ^{12}C^{16}O line, we first note that both $\Delta\tau_\circ^{12}$ and $\Delta\tau_\circ^{13}$ must be proportional to the total column densities of their respective isotopes:

$$\frac{\Delta\tau_\circ^{12}}{\Delta\tau_\circ^{13}} = \frac{N_{\rm CO}^{12}}{N_{\rm CO}^{13}} \ . \tag{6.6}$$

We have already seen, however, that

$$\frac{N_{\rm CO}^{12}}{N_{\rm CO}^{13}} = \gamma \left[\frac{^{12}C}{^{13}C}\right]^* \ . \tag{6.7}$$

Here γ is a number less than unity, representing both chemical fractionation and the lesser effect of photodissociation by ambient ultraviolet radiation. At a cloud temperature near 10 K, γ ranges from 0.1 to 0.3, with higher values occurring at greater optical depths. Combining (6.6) and (6.7), we solve for $\Delta\tau_\circ^{12}$:

$$\begin{aligned}\Delta\tau_\circ^{12} &= \gamma \left[\frac{^{12}C}{^{13}C}\right]^* \Delta\tau_\circ^{13} \\ &= 89\,\gamma\,\Delta\tau_\circ^{13} \ .\end{aligned} \tag{6.8}$$

For our sample observation, we use $\gamma = 0.3$ to derive $\Delta\tau_\circ^{12} = 27$.

There are many variations on the basic method we have outlined. For example, an observer wishing to obtain higher angular resolution may prefer to use the $J = 2 \rightarrow 1$ line of ^{12}C^{16}O at 1.3 mm instead of $J = 1 \rightarrow 0$. In this case, the optical depth will generally not be high enough for equation (6.3) to hold. Suppose, however, that the analogous line in ^{13}C^{16}O is still optically thin. Then one may observe another, even higher transition in ^{12}C^{16}O. One writes equation (6.1) for all three lines and solves for the unknowns $T_{\rm ex}$, $\Delta\tau_\circ^{13}$, and $\Delta\tau_\circ^{12}$. These optical depths refer to the $J = 2 \rightarrow 1$ transition only, since the higher one can be obtained through the assumption of LTE. We will encounter a practical application of this technique in Chapter 13, when discussing the temperatures within molecular outflows.

6.1.3 Column Density

Let us now consider the determination of the cloud volume density, $n_{\rm tot}$. Here the idea is first to evaluate *column* densities along each line of sight. An optically-thin transition is clearly the optimal choice for this purpose. Although ^{13}C^{16}O often only marginally fulfills the requirement, it is much more easily detected than rarer isotopes, such as ^{12}C^{18}O, that are unequivocally optically thin. Accordingly, many CO-based estimates of $n_{\rm tot}$ start with the column density $N_{\rm CO}^{13}$.

We have already used the fact that the column density of any species is proportional to the optical thickness of an emitted spectral line. The actual proportionality constant between $N_{\rm CO}^{13}$ and $\Delta\tau_\circ^{13}$ follows in a straightforward manner from the transfer equation for spectral lines. From Equations (C.15) and (C.16) in Appendix C, we find

$$N_{\rm CO}^{13} = \frac{8\pi\nu_\circ^2\Delta\nu^{13}Q^{13}\Delta\tau_\circ^{13}}{c^2 A_{10}} \left(\frac{g_0}{g_1}\right) \left[1 - \exp\left(-\frac{T_\circ^{13}}{T_{\rm ex}^{13}}\right)\right]^{-1} \ . \tag{6.9}$$

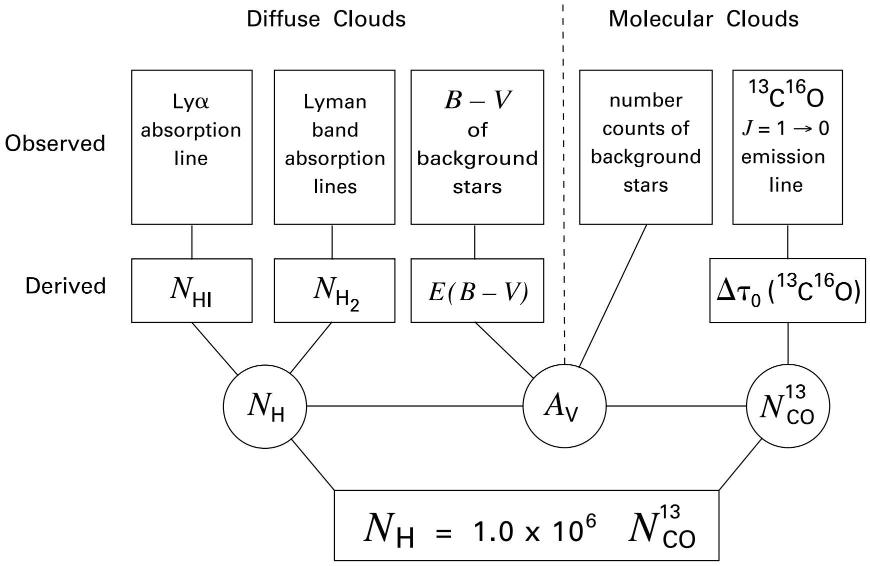


Figure 6.2 Steps in deriving the relationship between the $^{13}C^{16}O$ and the total hydrogen column densities.

Here $\Delta\nu^{13}$ is the observed full-width half-maximum of the $J = 1 \rightarrow 0$ line. The derivation of equation (6.9) assumes this width is intrinsic and neglects any saturation broadening. The quantity Q^{13} is the partition function for the rotational levels. Equation (C.18) implies that $Q^{13} = 2\,T^{13}_{\rm ex}/T^{13}_\circ$. Returning once more to our example, the observed velocity width, $\Delta V^{13}_r \equiv (\Delta\nu^{13}/\nu_\circ)\,c$, is 1.5 km s^{-1}. Having already determined $\Delta\tau^{13}_\circ$, we deduce that $N^{13}_{\rm CO} = 8.8 \times 10^{15}$ cm^{-2}.

For a given abundance of CO relative to hydrogen, $N^{13}_{\rm CO}$ should be proportional to the total hydrogen column density $N_H \equiv N_{\rm HI} + 2\,N_{\rm H_2}$. Knowing N_H will not give us, of course, the local volume density at interior points. However, if the cloud is well mapped, we may be able to estimate the physical depth of the cloud along our line of sight at a number of points. Dividing the column density by this depth then gives the average hydrogen number density n_H in the appropriate column.

6.1.4 Relation to Hydrogen Content

How, then, do we obtain the hydrogen column density from the CO data? The usual procedure is to invoke an empirical $N_H - N^{13}_{\rm CO}$ relationship. As this relationship, or analogous ones involving other CO isotopes, underlies the mass and density estimates for many molecular clouds and cloud complexes, we should understand its derivation. The essential steps, which rely on observations from ultraviolet to millimeter wavelengths, are shown schematically in Figure 6.2.

We recognize at the outset that the desired relationship cannot be established simply from measurements of quiescent molecular clouds, where it is impossible to observe N_H directly. We saw in Chapter 2, however, that O and B stars located behind *diffuse* clouds can yield estimates of $N_{\rm HI}$ through photo-excitation of the Lyα transition. A similar procedure, but employing Lyman band excitation, may be used to obtain H_2 column densities in front of early-type stars. In this manner, we may estimate the total N_H for clouds that have both atomic and molecular hydrogen. Unfortunately, the clouds for which the procedure works are not detectable in $^{13}C^{16}O$. What we must do is relate the column density to some other parameter that can be observed in *both* diffuse and molecular clouds. Such a parameter is A_V. This is obtained from the color excesses of the same background O and B stars used to determine hydrogen column densities. In fact, stellar absorption dips and color excesses were used to establish the linear $N_H - E_{\rm B-V}$ relation of equation (2.46). We then used equation (2.16) to establish the proportionality of N_H and A_V in equation (3.2). Although derived from observations of diffuse clouds, this latter equation can safely be extended to molecular clouds, as long as the composition of embedded dust grains is not significantly different.

The final step is to establish a connection between $N^{13}_{\rm CO}$ and A_V. Returning to molecular clouds, we have already seen how the first quantity can be obtained from the measured intensity of the $J = 1 \rightarrow 0$ emission line. If a cloud is not too opaque, we may estimate A_V by considering the obscuration of background stars. Consider for simplicity a uniform distribution of identical field stars, with spatial density n_* and a single absolute visual magnitude M_V. Suppose further that a molecular cloud, subtending a solid angle Ω_c, exists between distances r_1 and r_2, and provides the only source of visual extinction (see Figure 6.3). Then the apparent magnitude of any star located at $r > r_2$ is larger by A_V than it would have been without the cloud. Conversely, the radial distances of these background stars, considered as a function of m_V, are uniformly *lower*. According to equation (2.12), these distances increase with m_V as

$$\Delta \log r = 0.2\, \Delta m_V \ . \tag{6.10}$$

This linear relation holds only in front of the cloud or beyond it. In between, there must be a kink.

There is no practical value to writing the full equation relating $\log r$ to m_V, since the first quantity is not observed directly. Within any radial interval Δr, however, the number of stars included in Ω_c is

$$\begin{aligned} \Delta N_* &= n_*\, \Omega_c\, r^2\, \Delta r \\ &= 2.3\, n_*\, \Omega_c\, r^3\, \Delta \log r \ . \end{aligned} \tag{6.11}$$

Thus, if $\mathcal{N}$ represents dN_*/dm_V, the observed number of stars per interval of apparent magnitude, Equations (6.10) and (6.11) imply that $\mathcal{N}$ is proportional to r^3 for $r < r_1$ or $r > r_2$. Thus, $\log \mathcal{N} = 3 \log r$ to within an additive constant. A plot of the two observable quantities $\log \mathcal{N}$ and m_V should therefore show the same features as the $\log r - m_V$ relation– a smooth initial rise, a temporary break in slope between two magnitudes m_1 and m_2 corresponding to r_1 and r_2, respectively, and then a resumption of the initial slope. Figure 6.3 indicates how A_V may be read directly from such a plot.

In practice, one does not have the luxury of observing identical stars with a uniform spatial distribution. However, a generalized method based on the same principal does establish the

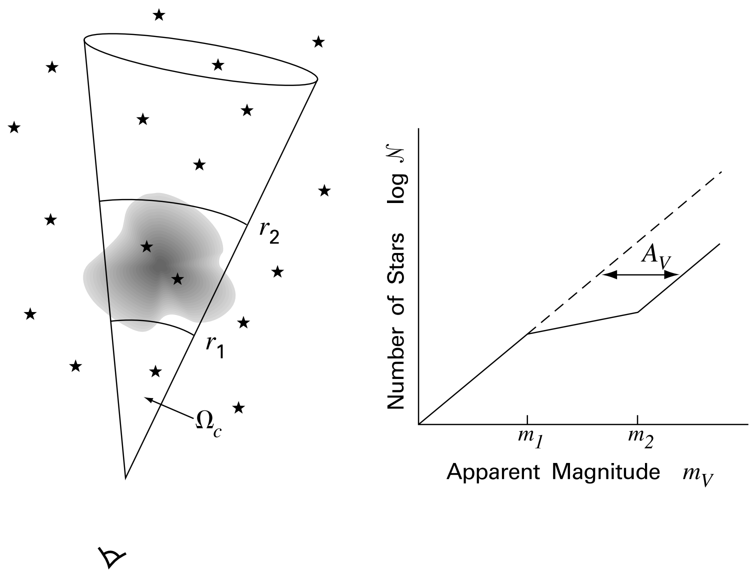


Figure 6.3 Procedure for obtaining A_V toward a molecular cloud through star counts.

$N_{\rm CO}^{13} - A_V$ relation. This is found to be

$$\frac{N_{\rm CO}^{13}}{A_V} = 2.5 \times 10^{15}\,{\rm cm}^{-2}\,{\rm mag}^{-1}\ . \tag{6.12}$$

Equations (3.2) and (6.12) finally yield the connection between the hydrogen and CO column densities:

$$N_H = 7.5 \times 10^5\, N_{\rm CO}^{13}\ . \tag{6.13}$$

Here, there is an estimated uncertainty of 50 percent in the proportionality constant. Returning one last time to our Taurus example, we derive $N_H = 6.6 \times 10^{21}$ cm^{-2}.

One should bear in mind that the correlations underlying equation (6.13) represent an average within the solar neighborhood. Under a variety of circumstances, such as the exposure of a cloud to intense ultraviolet emission from nearby stars, the concentration of CO relative to hydrogen changes. Moreover, a given cloud might well have substantial optical thickness even to ^{13}C^{16}O. The derivation of $\Delta\tau_\circ^{13}$ through equation (6.1) is then no longer feasible. To apply equation (6.13), one must first observe the cloud in a rarer CO isotope and then relate the new column density to that of ^{13}C^{16}O through the natural abundance ratios of the respective isotopes.

6.1.5 The X-Factor

As one considers more distant clouds within the Milky Way, or those in other galaxies, the emission from ^{13}C^{16}O becomes too feeble for practical use, even if equation (6.13) still applies in principle. Observational surveys on the largest scales revert to ^{12}C^{16}O and utilize only the velocity-integrated profile. If we again assume beam dilution and efficiency factors of unity, the hydrogen column density follows from the simple empirical relation:

$$N_H \,=\, X \int T_A^{12}\, dV_r \ . \tag{6.14}$$

The proportionality constant X is currently estimated at 2×10^{20} cm^{-2} K^{-1} km^{-1} s, again with a 50 percent uncertainty in either direction. Satellite observations of γ-rays have played a major role both in establishing equation (6.14) and in setting the value of X. We will see in Chapter 7 that cosmic-ray protons penetrating molecular clouds produce these high-energy photons through collisions with hydrogen nuclei. With sufficient knowledge of the cosmic-ray flux, the γ-radiation provides a direct measure of N_H, to be compared with ^{12}C^{16}O observations of the same region.

What is the theoretical basis for equation (6.14)? Earlier, we emphasized that the optical thickness of ^{12}C^{16}O implies that its integrated intensity should *not* be proportional to the column density along any line of sight. However, this statement refers to a well-resolved cloud region. The emission at great distances actually stems from an *ensemble* of many clouds. For equation (6.14) to hold, this diffuse collection must radiate, in some sense, as if it were optically thin, even if each individual cloud is not.

This viewpoint presupposes a certain degree of uniformity among all the radiating entities, at least in a statistical sense. To elucidate the underlying assumptions, let us crudely estimate the integral in equation (6.14) as $T_{A_\circ}^{12} \Delta V_r$, where $T_{A_\circ}^{12}$ is the antenna temperature at line center from a single, unresolved cloud, and ΔV_r is the cloud's observed line width. Equations (6.3) and (6.4) together imply that $T_{A_\circ}^{12}$ depends only on the local gas kinetic temperature. From the discussion in § 3.3, we may relate ΔV_r to the virial velocity in equation (3.20). This velocity, in turn, is proportional to $n_H^{1/2}\, L$. Here, n_H and L represent mean values within the smallest radiating unit, which, for large-scale surveys, could be an entire cloud complex. We conclude that the "constant" X in equation (6.14) is actually proportional to $n_H^{1/2}/T_{A_\circ}^{12}$. It is hardly surprising, then, that careful studies have indicated variations in X across our own and external galaxies, with a significant decline close to the galactic centers. Nevertheless, the figure quoted above serves as a reasonable first approximation in many circumstances.[1]

6.2 Ammonia

In using ammonia to determine cloud densities and kinetic temperatures, one takes advantage of the many observable lines associated with a single inversion transition. The basic idea is to alter the guessed density and temperature until the observed spectrum is reproduced. In this section, we shall explore this method in some detail, as the molecule has played a key role in determining the properties of dense cores.

---

[1] When observing an external galaxy, the telescope beam usually contains many unresolved clouds. The line width, set by the intercloud velocity, is much greater than ΔV_r for a single object. We assume, however, that these entities are so sparsely distributed that they do not shadow one another. Then the integrated antenna temperature is what it would be for a collection of static clouds, with the same total column density. Thus, our scaling argument for X still holds.

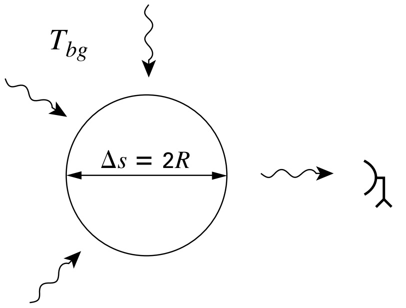


Figure 6.4 Cloud geometry assumed for deriving the NH_3 inversion spectrum.

6.2.1 Modeling the Inversion Lines

The fact that the states along the rotational backbone are metastable means that we may effectively treat each one as an isolated two-level system. That is, we lump together all the levels on either side of the inversion transition into two fictitious "super-levels." Using such a model, one successful strategy has been to determine first the excitation temperature and optical thickness for both the (1,1) and (2,2) backbone states by matching theoretical and observed emission spectra. Next, one solves the rate equations governing the super-level populations within each backbone state, in order to relate the previously derived quantities to the cloud density and kinetic temperature.

Let us follow this procedure, covering the essential steps. For simplicity, we take the cloud to be a uniform-density sphere of radius R, and we consider a line of sight that passes through its center (Figure 6.4). We focus on the transfer of radiation through molecules that are in a specific backbone state, *e.g.*, (1,1). Referring to the figure, the path length for radiation emanating from the cloud's far side is $\Delta s = 2\,R$. The corresponding optical thickness is therefore $\Delta\tau_\nu = 2\,\rho\,\kappa_\nu R$, where ρ and κ_ν are the cloud's mass density and opacity, respectively. We now apply the detection equation (6.1), generalized to a frequency off line center:

$$T_B(\nu) = (T_{\rm ex} - T_{\rm bg})\;[1 - \exp\left(-\Delta\tau_\nu\right)] \ . \tag{6.15}$$

In writing (6.15), we have used the Rayleigh-Jeans approximation to the function f, *i. e.*, we have assumed $T_\circ \ll T_{\rm ex}$ and $T_\circ \ll T_{\rm bg}$. This approximation is, in fact, only marginally applicable because $T_\circ$ for the inversion transition is about 1 K, while $T_{\rm ex}$ and $T_{\rm bg}$ are between 3 and 20 K.

To generate theoretical spectra from equation (6.15), we must somehow account for the complex line structure. Within our two-level model, we let the opacity κ_ν peak at discrete frequency intervals. Specifically, we represent $\Delta\tau_\nu$ as a sum:

$$\Delta\tau_\nu = \Delta\tau_\circ^{\rm tot}\sum_{i=0}^{N}\alpha_i \exp\left[-\left(\frac{\nu-\nu_\circ-\Delta\nu_i}{\Delta\nu}\right)^2\right] \ , \tag{6.16}$$

where the index i runs from 0 for the most intense central line to $N = 17$. Each line, assumed to be a Gaussian of identical width $\Delta\nu$, is displaced in frequency by $\Delta\nu_i$ from the $i = 0$ line centered at $\nu_\circ$. The coefficients α_i are the relative absorption probabilities for each hyperfine transition. These probabilities are known from theory. Finally, $\Delta\tau_\circ^{\rm tot}$ is the sum of the optical

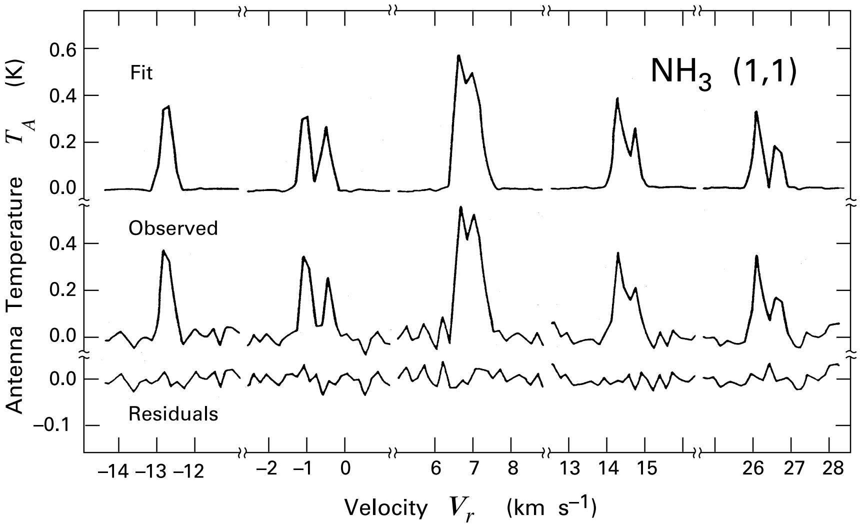


Figure 6.5 Theoretically derived emission spectrum of the NH_3 (1,1) transition in the L1489 dense core. The observed spectrum is shown below, along with the difference between the two.

thicknesses at all the line centers. Equation (6.16) implies that the central optical thickness of the $i = 0$ line is

$$\Delta\tau_\circ = \alpha_\circ \, \Delta\tau_\circ^{\rm tot} \ , \tag{6.17}$$

where the probability $\alpha_\circ$ is 0.23.

Assuming a background temperature $T_{\rm bg}$ of 2.7 K, Equations (6.15) and (6.16) give a spectrum, $T_B(\nu)$, for each choice of the three parameters $T_{\rm ex}$, $\Delta\tau_\circ^{\rm tot}$, and $\Delta\nu$. The idea is to tune these parameters for the (1,1) and (2,2) states until their observed emission spectra are matched. Figure 6.5 shows a pair of observed and calculated (1,1) spectra of the dense core L1489. Notice again that the independent variable is V_r rather than ν. Thus, for obtaining the theoretical spectrum, the quotient $(\nu - \nu_\circ - \Delta\nu_i)$ in (6.16) is replaced by $(V_r - \Delta V_i)/\sigma$. Here ΔV_i is the shift of each satellite from the $i = 0$ line and σ is the common velocity width of each Gaussian profile. Note also that the figure displays the antenna temperature T_A, which may be converted to T_B through the beam efficiency and the filling factor. These numbers are here 0.25 and 0.60, respectively. The best match to the spectrum then corresponds to $T_{\rm ex} = 6.6$ K, $\Delta\tau_\circ^{\rm tot} = 14$, and $\sigma = 0.23$ km s^{-1}.

6.2.2 Radiative Trapping

The next step is to relate $T_{\rm ex}$ and $\Delta\tau_\circ^{\rm tot}$ for the two backbone states to the values of $n_{\rm tot}$ and $T_{\rm kin}$. Physically, we need to find that unique density and temperature that produces, through both radiation and collisions, the observed excitation of the two inversion transitions. We therefore solve, within the framework of our highly simplified model, the coupled equations of radiative transfer and statistical equilibrium.

According to equation (B.3) in Appendix B, the populations n_u and n_l of the upper and lower super-levels within a single backbone state remain constant in time if collisional and radiative effects balance:

$$\gamma_{\rm lu}\, n_{\rm tot}\, n_l \;+\; B_{\rm lu}\, \bar{J}\, n_l = \gamma_{\rm ul}\, n_{\rm tot}\, n_u \;+\; B_{\rm ul}\, \bar{J}\, n_u \;+\; A_{\rm ul}\, n_u\ . \tag{6.18}$$

The quantity $\bar{J}$ is related to the mean intensity J_ν of the ambient field through equation (B.1):

$$\bar{J} \equiv \int_0^\infty J_\nu\, \phi(\nu)\, d\nu\ . \tag{6.19}$$

We take the absorption profile $\phi(\nu)$ to have the same frequency dependence as $\Delta\tau_\nu$ in equation (6.16):

$$\phi(\nu) = \phi_\circ \sum_{i=0}^{N} (\alpha_i/\alpha_\circ) \exp\left[-\left(\frac{\nu - \nu_\circ + \Delta\nu_i}{\Delta\nu}\right)^2\right]\ . \tag{6.20}$$

In this relation, $\phi_\circ \equiv \phi(\nu{=}\nu_\circ)$ is determined by the requirement that the profile be normalized to unity:

$$\int_0^\infty \phi(\nu)\, d\nu \;=\; 1\ .$$

Our solution to equation (6.18) will assume, for simplicity, that all quantities are spatially constant. We obtain $\bar{J}$ by starting with the exact solution of the radiative transfer equation for I_ν (Appendix C). We then evaluate the specific intensity at the center of the sphere in Figure 6.4. Equations (C.3) and (C.8) together imply

$$I_\nu = \frac{2\,\nu_\circ^2\, k_B\, T_{\rm bg}}{c^2} \exp\left(-\frac{\Delta\tau_\nu}{2}\right) + \frac{2\,\nu_\circ^2\, k_B\, T_{\rm ex}}{c^2} \left[1 - \exp\left(-\frac{\Delta\tau_\nu}{2}\right)\right]\ . \tag{6.21}$$

Here we have set the optical thickness $\alpha_\nu \Delta s$ in equation (B.3) equal to $\Delta\tau_\nu/2$ and have again applied the Rayleigh-Jeans approximation. Since I_ν at the center is isotropic, $J_\nu = I_\nu$. To obtain $\bar{J}$, we use Equations (6.19) and (6.20). Writing $\Delta\tau_\nu$ in (6.21) as $\Delta\tau_\circ \phi(\nu)/\phi_\circ$, we find

$$\bar{J} = \frac{2\,\nu_\circ^2\, k_B}{c^2} \left[\beta\, T_{\rm bg} + (1-\beta)\, T_{\rm ex}\right]\ , \tag{6.22}$$

where

$$\beta \equiv \int_0^\infty d\nu\, \phi(\nu) \exp\left[-\frac{\Delta\tau_\circ\, \phi(\nu)}{2\, \phi_\circ}\right]\ . \tag{6.23}$$

Note that the parameter β is small for large optical thickness $\Delta\tau_\circ$ and approaches unity as $\Delta\tau_\circ$ goes to zero. This quantity is therefore a measure of the *escape probability* for a photon with frequency near $\nu_\circ$.

Returning to equation (6.18), Appendix B derives the level populations for arbitrary $\bar{J}$. Equation (B.10) casts this solution in terms of $T_{\rm ex}$ and the radiation temperature $T_{\rm rad}$ associated with $\bar{J}$. Additionally, equation (B.7) implies that the parameter $f_{\rm coll}$ of Equations (B.10) and (B.11) reduces, in the Rayleigh-Jeans limit, to

$$f_{\rm coll} \approx \left(1 + \frac{n_{\rm crit}\, T_{\rm rad}}{n_{\rm tot}\, T_\circ}\right)^{-1}\ .$$

According to equation (6.22), $T_{\rm rad} = \beta T_{\rm bg} + (1-\beta) T_{\rm ex}$ in the same limit. After substituting these results into equation (B.10) and expanding the exponentials, we find

$$\frac{T_{\rm bg}}{T_{\rm ex}} + \frac{n_{\rm tot}}{\beta\, n_{\rm crit}} \left(\frac{T_\circ}{T_{\rm ex}} - \frac{T_\circ}{T_{\rm kin}} \right) = 1 \ . \tag{6.24}$$

We recall that $n_{\rm crit} \equiv A_{\rm ul}/\gamma_{\rm ul}$, so that the dimensionless quantity $n_{\rm tot}/\beta\, n_{\rm crit}$ in (6.25) measures the relative rates of collisional and radiative deexcitation. The factor β modifies the net emission rate to account for the sphere's finite optical thickness.

It is instructive to rewrite equation (6.24) in the form

$$T_{\rm ex} = \alpha\, T_{\rm bg} + (1-\alpha)\, T_{\rm kin} \ , \tag{6.25}$$

where

$$\alpha \equiv \left(1 + \frac{n_{\rm tot}\, T_\circ}{\beta\, n_{\rm crit}\, T_{\rm kin}} \right)^{-1} \ . \tag{6.26}$$

We see that $T_{\rm ex}$ is a weighted average of $T_{\rm bg}$ and $T_{\rm kin}$. Under conditions such that

$$\frac{n_{\rm tot}}{\beta\, n_{\rm crit}} \gg \frac{T_{\rm kin}}{T_\circ} \ ,$$

we have $\alpha \ll 1$ and LTE applies: $T_{\rm ex} \approx T_{\rm kin}$. In an optically-thin environment, *i.e.*, when $\beta \lesssim 1$, this condition is fulfilled once the density of colliders $n_{\rm tot}$ exceeds $n_{\rm crit}$ by a sufficient amount. Conversely, when collisions are relatively rare, $\alpha \approx 1$ and the two-level system comes into thermal equilibrium with the background radiation field: $T_{\rm ex} \approx T_{\rm bg}$. Notice, finally, that α can be much less than unity at *any* density $n_{\rm tot}$, provided β is sufficiently small. In other words, *large optical thickness drives the level populations into LTE, even if the ambient density is subcritical.* This effect of radiative trapping, illustrated here for the specific case of a two-level system, is of general importance in molecular clouds.

Suppose we know both $T_{\rm ex}$ and $\Delta\tau_\circ^{\rm tot}$ for the (1,1) system through matching to the observed hyperfine spectrum. With $\phi(\nu)$ given by equation (6.20), we may calculate the escape probability β from (6.23), after finding $\Delta\tau_\circ$ from (6.17). We then substitute the values of β and $T_{\rm ex}$ into equation (6.24) to yield a relation between $n_{\rm tot}$ and $T_{\rm kin}$. If we now repeat the procedure for the (2,2) system, we obtain a second such relationship, from which the density and kinetic temperature follow. For the representative dense core in Figure 6.5, this method yields $n_{\rm tot} = 2.5 \times 10^4$ cm^{-3} and $T_{\rm kin} = 10$ K.

6.3 Hydroxyl

In quiescent clouds far from any luminous stars, it is difficult to excite the rotational states of OH. Referring again to Figure 5.18, the energy gap between the ground and first excited level of the $^2\Pi_{3/2}$ ladder has an equivalent temperature of 120 K, while the jump to the ground state of the $^2\Pi_{1/2}$ ladder is 180 K. We may therefore restrict our attention to the four ground-state hyperfine transitions depicted in the figure. The Einstein A-coefficient for the strongest line at 1667 MHz ($F = 2^+ \rightarrow 2^-$) is quite low at 7.2×10^{-11} s^{-1}. Hence, the critical density for

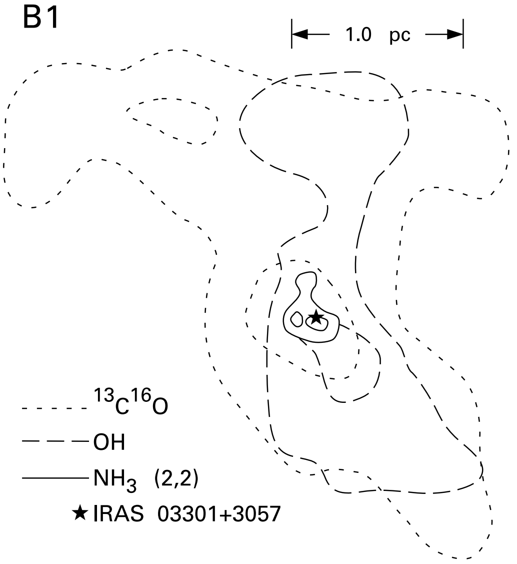


Figure 6.6 Molecular line observations of the B1 cloud in Perseus, along with the central embedded star. Two contours are shown for each line, an inner one near the peak intensity and an outer one at half that value.

collisional deexcitation of the line is also low, and OH serves as a tracer only of rather diffuse material. Figure 6.6 provides one example, the well-studied B1 cloud in Perseus. Here, the OH map covers a region almost as extended as that seen in the more abundant $^{13}C^{16}O$. Both species saturate in intensity outside the densest central region. The (2,2) transition of NH_3, on the other hand, is seen *only* near the embedded star.

6.3.1 Zeeman Splitting

The main utility of OH is not as a tracer of cloud gas but as a probe of the local magnetic field strength. The physical basis for these measurements is the *Zeeman effect.*[2] The energy of an OH molecule situated in a magnetic field $\boldsymbol{B}$ depends on the relative orientation of the field and the molecule's magnetic moment $\boldsymbol{\mu}$:

$$E_{\rm mag} = -\boldsymbol{\mu}\cdot\boldsymbol{B} \ . \tag{6.27}$$

Since the magnetic moment of OH arises principally from the presence of an unpaired electron, proper evaluation of $E_{\rm mag}$ requires a perturbative analysis of the electronic wave function. Such analysis shows that $\boldsymbol{\mu}$ has contributions from both the electron's orbital and spin angular momenta:

$$\boldsymbol{\mu} = \boldsymbol{\mu}_l + \boldsymbol{\mu}_s \ . \tag{6.28}$$

The first term is

$$\boldsymbol{\mu}_l = -\frac{e}{2m_e c}\boldsymbol{L} \ , \tag{6.29}$$

where $-e$ and m_e are, respectively, the electronic charge and mass. The reader may verify that equation (6.29) would also result classically if the electron's magnetic moment arose from the

---

[2] Magnetic fields in the atomic envelopes of molecular clouds are observed through the Zeeman effect in HI.

current associated with the particle's orbital motion. Since the orbital angular momentum is $\hbar$ in order of magnitude, it is convenient to rewrite (6.29) as

$$\boldsymbol{\mu}_l = -g_l\,\mu_B\,(\boldsymbol{L}/\hbar)\ . \tag{6.30}$$

Here g_l is a dimensionless number called the *Landé g-factor*, equal to unity in this case. The quantity μ_B is the *Bohr magneton*, given by

$$\begin{aligned}\mu_B &\equiv \frac{e\hbar}{2m_e c}\\ &= 5.8\times 10^{-3}\,\mathrm{eV}\,\mu\mathrm{G}^{-1}\ .\end{aligned} \tag{6.31}$$

The spin contribution $\boldsymbol{\mu}_s$ obeys an equation analogous to (6.30), but with a value for g_s of 2.0.

Given $\boldsymbol{\mu}_l$ and $\boldsymbol{\mu}_s$, we may use Equations (6.27) and (6.28) to evaluate E_{mag}. However, in the usual representation of the molecular wavefunction, neither the component of $\boldsymbol{L}$ nor of $\boldsymbol{S}$ in the direction of $\boldsymbol{B}$ has an associated quantum number. As we saw in § 5.5, both $\boldsymbol{L}$ and $\boldsymbol{S}$ precess rapidly about the internuclear axis because of the torque exerted on the electron by the magnetic field of the nuclei and remaining, paired electrons. Within each Λ-doubled rotational state shown in Figure 5.15, the precise energy depends on the relative orientation of $\boldsymbol{B}$ and the grand total angular momentum $\boldsymbol{F}$. Let us denote by ΔE_{mag} the perturbation of the state energy from its value at $\boldsymbol{B} = 0$. Then ΔE_{mag} depends on all the quantum numbers characterizing the state (Λ, Ω, S, J, I, and F), and on the additional quantum number M_F. The latter represents the projection of $\boldsymbol{F}$ along $\boldsymbol{B}$ and can take on $2F+1$ integer values, ranging from $M_F = -F$ to $+F$. The final expression for the energy perturbation is then

$$\Delta E_{\mathrm{mag}} = -g\,\mu_B\,M_F\,B\ . \tag{6.32}$$

Here the full Landé g-factor is again of order unity and is a function of all the quantum numbers except M_F.

Dividing equation (6.32) by h, we obtain the splitting in terms of frequency. We may write the result as

$$\Delta\nu_{\mathrm{mag}} = \left(\frac{b}{2}\right)B\ . \tag{6.33}$$

The factor 1/2 is convenient because the frequency splitting is symmetric with respect to the unperturbed line. For the dominant 1665 and 1667 MHz transitions, the constant b has the values 3.27 and 1.96 Hz $\mu\mathrm{G}^{-1}$, respectively. The actual $\Delta\nu_{\mathrm{mag}}$ for typical cloud field strengths is therefore relatively small ($\Delta\nu_{\mathrm{mag}}/\nu_\circ \sim 10^{-8}$). Nevertheless, it is *much* greater than in molecules lacking an unpaired electron. Here, Λ-doubling is absent, and the magnetic interaction that splits each rotational state involves the spins of two nucleons, rather than an electron and a nucleon. From equation (6.31), the *nuclear magneton* μ_N is smaller than μ_B by the electron-nucleon mass ratio of 1/1836. The b-values of NH_3 and H_2O, for example, are only 7.2×10^{-4} and 2.3×10^{-3} Hz $\mu\mathrm{G}^{-1}$, respectively.

6.3.2 Polarized Lines

Figure 6.7 shows schematically the full splitting of the ground rotational state in the presence of an external $\boldsymbol{B}$-field. The various possible lines can be classified according to the corresponding

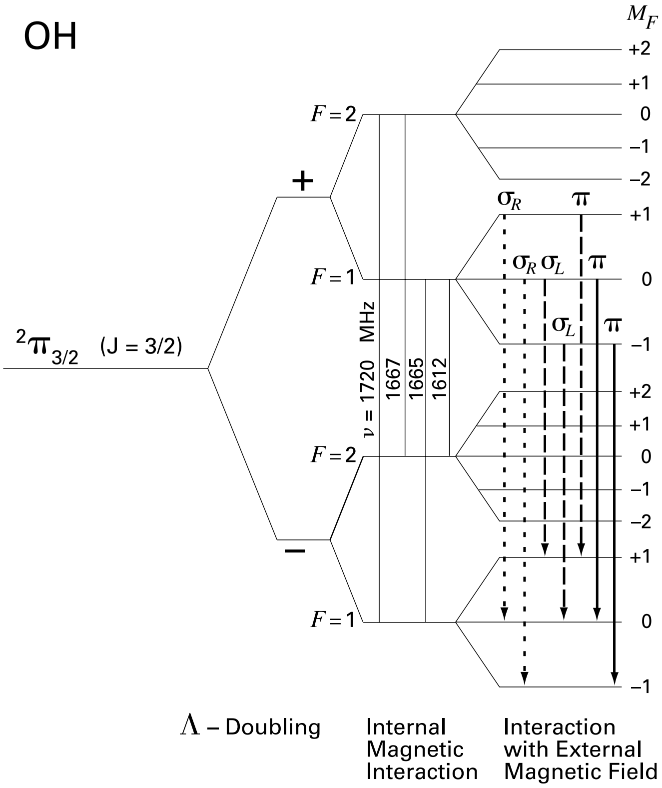


Figure 6.7 Zeeman splitting of the ground rotational state of OH. The multiplicity of levels arises from Λ-doubling, internal spin-orbit coupling, and the effect of an external magnetic field. The 7 possible transitions near the 1665 MHz line are shown.

values of ΔM_F. Selection rules limit the possibilities to $\Delta M_F = 0, +1$, and -1, which are designated π-, σ_R-, and σ_L-transitions, respectively. The symbols pertain to the state of polarization of the radiation, *i. e.*, the manner in which the electric field vector oscillates.

Let us focus on the cluster of lines near 1665 MHz; the seven allowed level transitions (comprising only three distinct lines) are shown in Figure 6.7. The observed pattern of line splitting and polarization depends on the angle θ between the magnetic field and the line of sight (see Figure 6.8). When $\theta = \pi/2$, three separate lines appear: a π-transition at the original frequency $\nu_\circ$, and two weaker σ-lines at $\nu_\circ \pm \Delta\nu_{\rm mag}$. Here the frequency shift is that in equation (6.33). As shown in the figure, all three lines are linearly polarized, with the electric field vector oscillating in a direction parallel to $\boldsymbol{B}$ for the unshifted line, and perpendicular for the two σ-components.

When $\boldsymbol{B}$ is pointing directly toward the observer (*i. e.*, $\theta = 0$), Figure 6.8 indicates that only two lines appear. The σ_R line, at a frequency of $\nu_\circ + \Delta\nu_{\rm mag}$, has right-handed circular polarization, *i. e.*, the electric vector rotates counter-clockwise from the observer's point of view.[3]

---

[3] We follow the IEEE definition, in which right circular polarization is clockwise rotation of the electric field vector as seen by the transmitter of the radiation, *i. e.*, counter-clockwise to the observer.

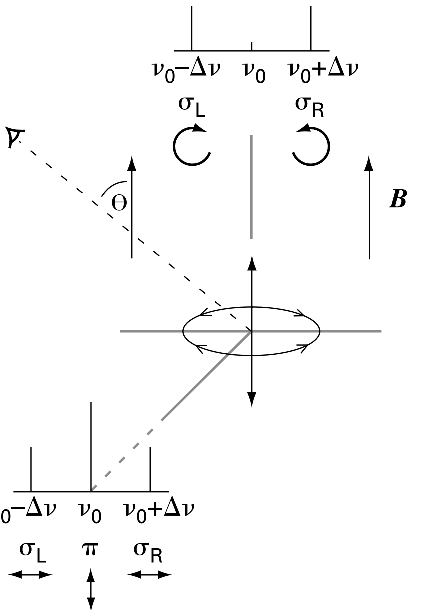


Figure 6.8 Intensity and polarization state of the three OH lines near 1665 MHz. The pattern is identical to that predicted by classical electromagnetic theory for an electron both oscillating in the direction of $\boldsymbol{B}$ and revolving in the perpendicular plane.

Conversely, the σ_L line at $\nu_\circ - \Delta\nu_{\rm mag}$, has left-handed circular polarization. For an arbitrary inclination angle θ, each line at $\nu_\circ \pm \Delta\nu_{\rm mag}$ is an admixture of linearly and circularly polarized components, *i. e.*, it is elliptically polarized. Moreover, the intensity of the circularly polarized component at each frequency varies as $\cos\theta$.

It is interesting that this splitting and polarization of the 1665 MHz line is in precise agreement with a naïve classical model of electromagnetic radiation arising from an oscillating electron. Here, the unshifted line is produced by the electron's motion along the direction of $\boldsymbol{B}$, while the two σ-components arise from circular motion in the plane perpendicular to the magnetic field (see Figure 6.8). When $\boldsymbol{B}$ lies along the line of sight, the central component disappears, since an oscillating charge does not radiate in its direction of motion. However, this agreement between the classical and quantum accounts is fortuitous. The 1612 MHz ($F = 1^+ \rightarrow 2^-$) line, for example, splits into *more* than three lines for any $\boldsymbol{B}$-field orientation. The reader may verify, incidentally, that the frequency shifts in the classical picture correspond to a Landé g-factor of exactly unity.

6.3.3 Measurement of B-Fields

Since, for any line, the pattern of splitting depends on the angle θ, it would seem straightforward to read off both $B_{||}$, the field component along the line of sight, and the orthogonal component $B_\perp$. However, our discussion has thus far assumed an idealized spectral line of zero width. In fact, outside of maser environments, each OH line is at least broadened thermally, *i. e.*, by the Doppler shifts resulting from the molecules' random motion. The magnitude of such broadening is, from Appendix E,

$$\frac{\Delta\nu_{\rm therm}}{\nu_\circ} \approx \frac{v_s}{c} \ , \tag{6.34}$$

where v_s is the sound speed in the cloud. The estimate here is good to within a factor of order unity. The quantity v_s is $(\mathcal{R}T_{\rm kin}/\mu)^{1/2}$, where $\mathcal{R}$ is the ideal gas constant and μ, the mean molecular weight relative to hydrogen, is about 2. Comparing this line broadening to the magnetic line splitting in equation (6.33), we find

$$\begin{aligned}\frac{\Delta\nu_{\rm mag}}{\Delta\nu_{\rm therm}} &\approx \frac{b\,B}{\nu_\circ}\left(\frac{c}{v_s}\right) \\ &\approx 10^{-3}\left(\frac{B}{\mu\,G}\right)\left(\frac{T_{\rm kin}}{10\,{\rm K}}\right)^{-1/2} \ .\end{aligned} \tag{6.35}$$

Apart from masers, the actual B-values of interest range in magnitude from roughly $10\,\mu$G in cold, quiescent molecular clouds to $100\,\mu$G near HII regions. Hence, the well-separated lines in Figure 6.8 are generally *not* seen, and a more indirect approach is necessary.

The mechanism of thermal broadening operates identically on all polarization components of the radiation. Imagine viewing the 1665 MHz line, for example, through two circularly polarizing filters of opposite helicity. The two profiles, $I_R(\nu - \nu_\circ)$ and $I_L(\nu - \nu_\circ)$, would be identical in shape, but shifted in frequency by $\Delta\nu_{\rm mag}$ (see Figure 6.9). This fact suggests that we consider the *difference* of the two profiles. In general, the electric field vector of any monochromatic radiation can be decomposed into right- and left-handed circularly polarized waves. The intensity difference of these components is known as the *Stokes V-parameter*, one of three independent scalar quantities that fully characterize the polarization state of the radiation. In the present case, the radiation is *not* monochromatic, but $I_R(\nu - \nu_\circ) - I_L(\nu - \nu_\circ)$ is still termed the *V-spectrum* of the source.[4]

We noted earlier that both $I_R(\nu - \nu_\circ)$ and $I_L(\nu - \nu_\circ)$ are reduced in magnitude from the total intensity $I(\nu - \nu_\circ)$ by $\cos\,\theta$. Hence we may write their difference as

$$\begin{aligned}I_R(\nu - \nu_\circ) - I_L(\nu - \nu_\circ) &= \cos\,\theta\,[I(\nu - \nu_\circ - \Delta\nu_{\rm mag}) - I(\nu - \nu_\circ + \Delta\nu_{\rm mag})] \\ &\approx -2\,\cos\,\theta\,\Delta\nu_{\rm mag}\,\frac{dI}{d\nu}(\nu - \nu_\circ) \ .\end{aligned} \tag{6.36}$$

Using equation (6.33) for $\Delta\nu$, we have the approximate equality

$$\begin{aligned}I_R(\nu - \nu_\circ) - I_L(\nu - \nu_\circ) &= -b\,B\,\cos\,\theta\,\frac{dI}{d\nu}(\nu - \nu_\circ) \\ &= -b\,B_\parallel\,\frac{dI}{d\nu}\,(\nu - \nu_\circ) \ .\end{aligned} \tag{6.37}$$

Equation (6.37) represents a practical means of determining the field component $B_\parallel$. One first measures directly the V-spectrum on the left side of the equation. It is best to evaluate the

---

[4] In practice, one measures all three Stokes parameters by employing polarizing filters. One may then define the fractional polarization of the beam, whether circular or linear, through algebraic combinations of these parameters; see also equation (2.51).

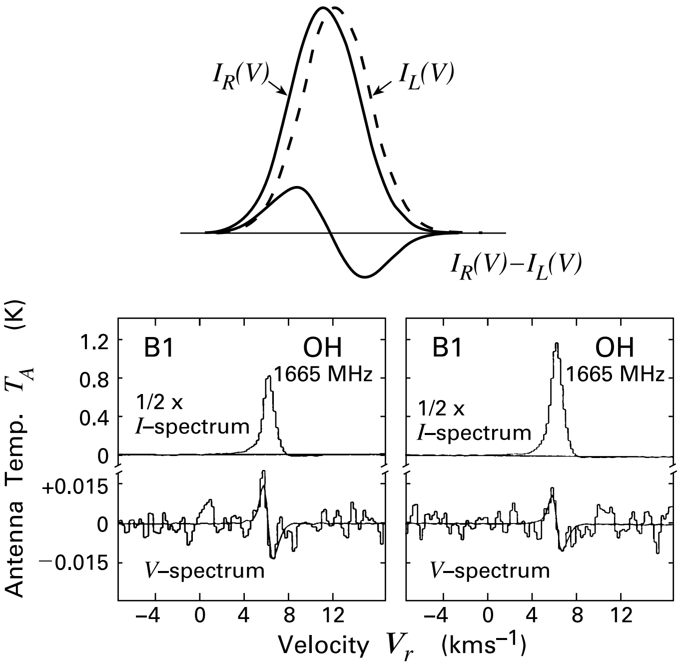


Figure 6.9 *upper panel*: Subtraction of the right- and left-circularly polarized components of an OH line to yield the V-spectrum. *lower panel*: Observed I- and V-spectra of the 1665 MHz and 1667 MHz lines in B1.

derivative on the right side using the *average* of $I_R(\nu - \nu_\circ)$ and $I_L(\nu - \nu_\circ)$. Such an average is equal to one half of the *I-spectrum*, where the Stokes I-parameter is the sum of the two circularly polarized components. Finally, one determines $B_{||}$ through a fitting procedure. Note that a negative value of $B_{||}$ indicates a field pointing *toward* the observer.

The first successful application of this procedure in molecular clouds was toward NGC 2024 in the Orion A complex (Plate 1). In this case, the 1665 and 1667 MHz lines appear in absorption against the continuum radiation from a background HII region, and the derived $B_{||}$ is 38 μG. Thus far, it has proved difficult to extend the technique to dense cores. The OH molecule is created through ion-molecule reactions, and the relative abundance of the species diminishes in very optically thick regions. Nevertheless, there are a handful of relevant observations. Figure 6.9 shows the Zeeman splitting in OH emission from the central region of the B1 cloud. The best-fit value for $B_{||}$ of -27 μG indicates that the field is providing substantial support against the cloud's self-gravity. We will later study in depth this important aspect of cloud structure.

Chapter Summary

The rotational transitions of CO effectively measure both cloud temperature and column density along a given line of sight. Here one may profitably use a combination of optically thick lines, emanating from the cloud surface, and optically thin lines that probe the interior. For more distant clouds in our own and other galaxies, only the strongest, optically thick lines from ^{12}C^{16}O are detectable. In all cases, one must rely on an empirical correlation to find the hydrogen column density.

Observation of the NH$_3$ inversion lines give access to both temperature and local *volume* density. The procedure is to model the collisional excitation and radiative decay of numerous levels, while including the effect of partially trapped photons. The levels are populated as if the molecules had a temperature intermediate between that characterizing the background radiation and that governing their translational motion.

The unpaired electron in OH endows the molecule with a net magnetic moment. Consequently, an ambient magnetic field reveals itself by shifting the frequencies of circularly polarized lines. This Zeeman splitting has proved most useful in clouds of relatively modest density. In these structures, the magnetic field is strong enough to play a key role in the clouds' mechanical equilibrium.

Suggested Reading

Section 6.1 For a discussion of the isotopic ratios of CO, see

Langer, W. D. 1997, in *CO: Twenty-Five Years of Millimeter-Wave Spectroscopy*, ed. W. B. Latter, S. J. E. Radford, P. R. Jewell, J. G. Magnum, and J. Bally (Dordrecht: Kluwer), p. 98.

Application of the detection equation to observations of both CO and NH$_3$ is covered in

Martin, R. N. & Barrett, A. H. 1978, ApJSS, 36, 1.

The N_H-A_V correlation was first obtained by

Bohlin, R. C., Savage, B. D., & Drake, J. F. 1978, ApJ, 224, 132.

while the $N_{\rm CO}^{13}$-A_V and $N_{\rm CO}^{13}$-N_H relations are due to

Dickman, R. L. 1978, ApJSS, 37, 407.

The proportionality constant X between N_H and the integrated ^{12}C^{16}O intensity is a central topic in

Combes, F. 1991, ARAA, 29, 195.

Section 6.2 The analysis of NH$_3$ hyperfine transitions is presented in

Barrett, A. H., Ho, P. T. P., & Myers, P. C. 1977, ApJ, 211, L39,

and applied extensively to dense cores by

Jijina, J., Myers, P. C., & Adams, F. C. 1999, ApJSS, 125, 161.

For a critical discussion of the standard interpretation of NH$_3$ spectra, see

Stutzki, J. & Winnewisser, G. 1985, AA, 148, 254.

Section 6.3 The physics of the Zeeman effect is lucidly explained in

> Powell, J. L. & Crasemann, B. 1961, *Quantum Mechanics* (Reading: Addison-Wesley), Chapter 10.

The first detection of the effect in molecular clouds was due to

> Crutcher, R. M. & Kazés, I. 1983, AA, 125, L23.

Reviews of subsequent observations, using OH and other tracers, are in

> Crutcher, R. M. 1999, ApJ, 520, 706
>
> Bourke, T. L., Myers, P. C., Robinson, G., & Hyland, A. R. 2001, ApJ, 554, 916.

7 Heating and Cooling

A proper understanding of molecular clouds must include an account of how they absorb energy and reradiate it into space. Our purpose in this chapter is to elucidate the most important mechanisms. We first cover the heating from interstellar cosmic rays and photons. Included in the latter are both the diffuse background of radiation and the X-rays from pre-main-sequence stars. We next describe how clouds cool themselves through emission from their constituent atoms, molecules, and dust grains.

In each case, our main goal is to provide a physical account of the relevant process, but we also give the reader practical formulas for heating and cooling rates.[1] We limit ourselves to the case of quiescent clouds, *i. e.*, we ignore any heating from nearby massive stars or internal, bulk motion of the gas. Both of these topics will be covered in later chapters. For convenient reference, Table 7.1 lists all the thermal processes discussed here.

7.1 Cosmic Rays

One significant source of cloud heating is a ubiquitous flux of particles whose existence has been known for almost a century, but whose origin is only now becoming clear. By 1900, physicists were routinely using ionization chambers to measure minute amounts of emission from radioactive elements. Even when *no* substance was introduced, the chambers recorded the presence of some ionizing agent. Using such a device on board a balloon, V. Hess found, in 1912, that the flux increased with altitude and had no diurnal variation, and therefore was of extraterrestrial and extrasolar origin. It took several more decades to unravel the detailed nature of this bombarding stream.

7.1.1 Composition and Energetics

Cosmic rays, we now understand, mostly consist of relativistic protons, with an admixture of heavy elements and electrons. Their composition is, in fact, roughly solar, but with significant differences. An over-abundance of the light elements lithium, beryllium, and boron reflects the creation of secondary particles through nuclear reactions between the protons and the interstellar medium. These secondaries result from *spallation*, *i. e.*, ejection from more massive target nuclei. More frequently, cosmic ray protons collide with the slow-moving protons of interstellar atomic hydrogen. Collision excites the target proton, causing it to emit a π° meson. The meson, in turn, decays into two γ-rays. This process accounts for much of the diffuse background of

---

[1] The formulas we present are simplified, approximate versions of those found in the research literature. Quantitative rates are trustworthy to about a factor of two, under the conditions specified.

The Formation of Stars. Steven W. Stahler and Francesco Palla

ISBN: 3-527-40559-3

Table 7.1

Cloud Heating

Process	Reaction	Equation
cosmic rays on HI	$p^+ + H \rightarrow H^+ + e^- + p^+$	(7.12)
cosmic rays on H_2	$p^+ + H_2 \rightarrow H_2^+ + e^- + p^+$	(7.14)
carbon ionization	$C + h\nu \rightarrow C^+ + e^-$	(7.16)
photoelectric ejection		(7.18)
dust irradiation		(7.20)
stellar X-rays	$H + h\nu \rightarrow H^+ + e^-$	(7.23)

Cloud Cooling

Process	Reaction	Equation
O collisional excitation	$O + H \rightarrow O + H + h\nu$	(7.26)
C^+ fine structure excitation	$C^+ + H \rightarrow C^+ + H + h\nu$	(7.27)
CO rotational excitation	$CO + H_2 \rightarrow CO + H_2 + h\nu$	(7.35)
dust thermal emission		(7.39)
gas-grain collisions		(7.40)

interstellar γ-radiation found by satellites in the 1960s and 1970s. Protons impinging on molecular clouds produce more localized sources of γ-rays. Observations by the COS-B satellite and, subsequently, the Compton Gamma Ray Observatory, were useful as an independent means of tracing the distribution of H_2 in giant cloud complexes.

The observed energies of cosmic rays span an enormous range, from 10 to 10^{14} MeV. Note that a single proton of 10^{14} MeV has the kinetic energy of a well-hit tennis ball. From about 10^3 to 10^9 MeV, the received particle flux, $\Phi_{\rm CR}(E)$, follows a power law: $\Phi_{\rm CR}(E) \sim E^{-2.7}$. Here, the flux is measured per unit energy E (see Figure 7.1). At higher energies than those shown in the figure, the flux declines as E^{-3}. Conversely, $\Phi_{\rm CR}(E)$ flattens and then turns over as E falls below about 10^3 MeV. This turnover is an effect of the solar wind, that both sweeps out incoming particles and creates a periodic modulation of the flux in step with the solar activity cycle.

Where do cosmic rays come from and how do they attain such high energies? With regard to the latter issue, all the particles are electrically charged and thus subject to magnetic deflection.

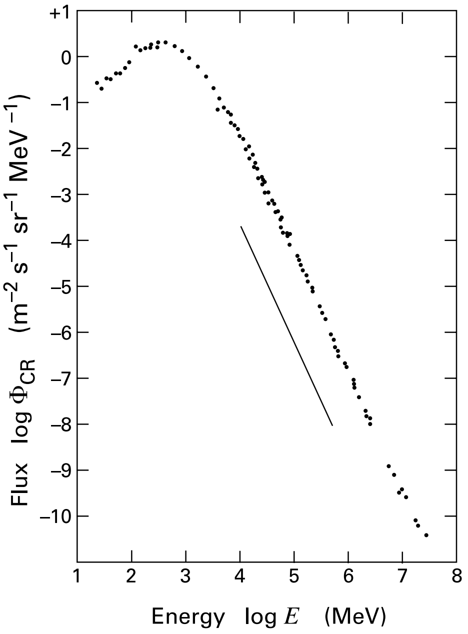


Figure 7.1 The observed particle flux in cosmic rays, as measured per unit solid angle and per unit energy. The flux is plotted as a function of E, the energy per particle. The straight line is a power law with slope -2.7.

Indeed, it is the magnetic field within the solar wind that prevents low-energy particles from reaching the Earth. A time-invariant field cannot alter a proton's energy, since the Lorentz force is perpendicular to the velocity vector. A *changing* magnetic field, however, generates an electric field that *can* do work. Thus, as E. Fermi pointed out in 1949, cosmic rays could be accelerated by the fields frozen into a turbulent interstellar plasma. Fifteen years earlier, W. Baade and F. Zwicky had proposed that supernovae constitute the basic source of cosmic rays. In recent decades, these two ideas have merged in a fruitful way. It is now believed that cosmic rays with energies up to 10^9 MeV are produced by particle acceleration within the magnetized shocks created by supernova remnants as they plow through the interstellar medium. More energetic particles are probably extragalactic in origin.

The energy density of cosmic rays within our Galaxy is about 0.8 eV cm^{-3}, close to that associated with a typical interstellar magnetic field of $3\,\mu$G. This similarity suggests that the Galactic field is strong enough to contain, at least temporarily, the cosmic rays produced from within. Any charged particle entering a magnetized region executes circular motion. The size of the gyroradius, r_B, for a particle of mass m, charge q, and velocity v in a field of magnitude B is

$$r_B = \frac{\gamma\, m\, c\, v}{q\, B} \, , \tag{7.1}$$

where γ is the Lorentz factor $(1 - v^2/c^2)^{-1/2}$. Even for a 10^6 MeV proton in a 3 μG field, r_B is only of order 10^{15} cm, far less than Galactic dimensions. Thus, the orbit is a tight helix, in which the velocity is closely confined in the plane perpendicular to the field, but is unconstrained in the parallel direction (Figure 7.2). Irregularities in the field change the orientation of these helical orbits and lead to a gradual drift of the particles out of the Galaxy. From measurements

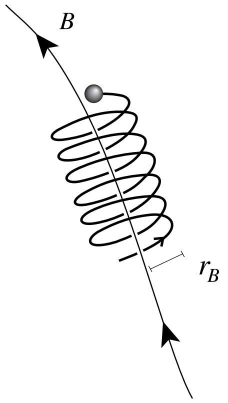


Figure 7.2 Helical motion of a cosmic-ray proton in a magnetic field. The gyroradius r_B is indicated.

of the isotopic abundances of heavier nuclei, we know that the mean confinement time is of order 10^7 yr. Diffusion across field lines also accounts for the high degree of isotropy seen in the cosmic rays entering our solar system.

7.1.2 Interaction with Clouds

Upon entering a molecular cloud, a gyrating cosmic-ray proton interacts with ambient nuclei and electrons through both the Coulomb and nuclear forces. The nuclear excitations principally decay through emission of γ-rays, which promptly escape. Elastic scattering due to the Coulomb repulsion between the proton and H_2 is negligible at cosmic ray energies. Instead, the proton scatters inelastically, exciting H_2 to a higher electronic state that leads to dissociation. On average, each scattering transfers 1.6 eV to the gas as kinetic energy. The most common result of proton impact, however, is ionization:

$$\mathrm{p}^+ + \mathrm{H}_2 \rightarrow \mathrm{H}_2^+ + \mathrm{e}^- + \mathrm{p}^+ \ . \tag{7.2}$$

Besides heating the cloud, ionization supplies the electric current that couples the gas to any internal magnetic field. This coupling occurs even though the net fractional ionization is quite low. The presence of charged species also serves to initiate the ion-molecule reactions described in Chapter 5.

It is the secondary electron in (7.2) that actually provides heat through its subsequent interactions with ambient hydrogen molecules. Let us denote the rate of heat deposition in the cloud per unit volume as $\Gamma_{\mathrm{CR}}(\mathrm{H}_2)$. We may write this quantity as

$$\Gamma_{\mathrm{CR}}(\mathrm{H}_2) = \zeta(\mathrm{H}_2)\, n_{\mathrm{H}_2}\, \Delta E(\mathrm{H}_2) \ . \tag{7.3}$$

Here, $\zeta(\mathrm{H}_2)$ is the ionization rate (probability per unit time) for a single hydrogen molecule, n_{H_2} is the volume density of these molecules, and $\Delta E(\mathrm{H}_2)$ is the thermal energy added to the gas as a result of each ionization event.

To obtain $\Delta E(\mathrm{H}_2)$, we must consider the energy of the secondary electron, which naturally depends on that of the cosmic-ray proton. Protons with energy $E \gtrsim 1$ GeV have no heating effect at all. Instead of ionizing H_2, these produce nuclear excitation and γ-rays. One might have expected that the electron production and heating from lower-energy protons would peak and then fall off significantly with diminishing E. However, it is a peculiarity of the long-range Coulomb interaction involved in the p-H_2 collision that the energy of the secondary electron is very insensitive to that of the incident proton. Thus, we can focus on a "typical" electron of 30 eV, which is produced by a 10 MeV proton.

There are a number of routes this electron can take. One possibility is the further ionization of ambient molecules:

$$\mathrm{e}^- + \mathrm{H}_2 \rightarrow \mathrm{H}_2^+ + \mathrm{e}^- + \mathrm{e}^- \ . \tag{7.4}$$

This reaction does not itself supply thermal energy, but produces an additional electron that can heat the cloud. Although the incident electron is much less energetic than the original cosmic ray proton, the cross section for ionization of H_2 by electrons peaks at an energy several times the threshold value of 15.4 eV. Thus, the process described by equation (7.4) is significant. For our 10 MeV proton, this reaction enhances the total ionization rate $\zeta(\mathrm{H}_2)$ by a factor of 1.6 over that from the protons alone.

However it is produced, the electron may elastically scatter off a hydrogen molecule. Such a collision imparts relatively little kinetic energy to the gas because of the tiny electron-to-proton mass ratio. The electron may also scatter inelastically, exciting internal energy levels in H_2. If the excited level is a rotational one within the ground electronic and vibrational state, there is time to transfer energy to neighboring molecules through collisions. More commonly, it is the higher electronic and vibrational states that are populated. These have such large A-values that they always decay radiatively rather than collisionally. However, most of these energetic photons are absorbed by dust. If the emitting H_2 molecule is left in a high vibrational level of the ground electronic state, it continues to decay until it ends up in an excited rotational level within the ground state. Once again, these higher J-states heat the gas collisionally.

The most important means for the electron to provide heating is dissociation:

$$\mathrm{e}^- + \mathrm{H}_2 \rightarrow \mathrm{H} + \mathrm{H} + \mathrm{e}^- \ . \tag{7.5}$$

The energy of the incoming electron beyond that required to dissociate H_2 goes into motion of the two hydrogen atoms. Collisions quickly disperse this energy throughout the gas. After summing all possible processes with the appropriate branching ratios, we find that $\Delta E(\mathrm{H}_2)$ is 7.0 eV. For comparison, the corresponding figures for 1 and 100 MeV cosmic ray proton are 6.3 and 7.6 eV, respectively.

Within diffuse clouds, the heating of *atomic* hydrogen by cosmic rays is also important. As in the molecular case, the most common result of proton impact is ionization:

$$\mathrm{p}^+ + \mathrm{H} \rightarrow \mathrm{H}^+ + \mathrm{e}^- + \mathrm{p}^+ \ . \tag{7.6}$$

It is again the electron, now ejected at a typical energy of 35 eV, that actually heats the gas, at a rate

$$\Gamma_{\mathrm{CR}}(\mathrm{HI}) = \zeta(\mathrm{HI})\, n_{\mathrm{HI}}\, \Delta E(\mathrm{HI}) \ . \tag{7.7}$$

In a weakly ionized HI gas, the secondary electron initially slows down by ionization and excitation of additional hydrogen atoms. The effect of secondary ionizations is to increase the total ionization rate by a factor of 1.7. The excited and ionized atoms quickly radiate away their excess energy. Thus, true heating cannot begin until the energy of the electron falls below the 10.2 eV needed to excite hydrogen to its first excited ($n = 2$) electronic state. From then on, the electron gradually loses its remaining energy through many elastic collisions. Numerical calculation reveals that $\Delta E(\mathrm{HI})$, the average thermal energy transfer per ionization event, is 6.0 eV.

7.1.3 Hydrogen Ionization Rate

Determining either $\Gamma_{\mathrm{CR}}(\mathrm{H_2})$ or $\Gamma_{\mathrm{CR}}(\mathrm{HI})$ still requires measurement of the respective ionization rates, $\zeta(\mathrm{H_2})$ or $\zeta(\mathrm{HI})$. Here, the main impediment is our ignorance of the interstellar flux of low-energy cosmic rays. We have noted, however, that the ionizations are the first steps in a network of ion-molecule reactions that lead to the formation of more complex species. We may therefore use the observed abundances of selected molecules to infer the ionization rates indirectly.

One such molecule is OH, which, along with hydrogen itself, can sometimes be detected in diffuse clouds through its ultraviolet absorption lines. In this manner, observers have measured the typical number density of OH relative to HI as 2×10^{-7}. (Compare Table 5.1.) The sequence of reactions forming OH from atomic oxygen and hydrogen begins when the $\mathrm{H^+}$ created by cosmic rays encounters an oxygen atom from the ambient gas:

$$\mathrm{H^+ + O \rightarrow O^+ + H} \; . \tag{7.8}$$

Note that the $\mathrm{H^+}$ may also recombine with free electrons before such charge exchange can occur. We will simplify the analysis by ignoring this possibility. As illustrated in Figure 7.3, the production of $\mathrm{O^+}$ initiates a chain of reactions with $\mathrm{H_2}$:

$$\begin{aligned} \mathrm{O^+ + H_2} &\rightarrow \mathrm{OH^+ + H} \\ \mathrm{OH^+ + H_2} &\rightarrow \mathrm{H_2O^+ + H} \\ \mathrm{H_2O^+ + H_2} &\rightarrow \mathrm{H_3O^+ + H} \; . \end{aligned} \tag{7.9}$$

The $\mathrm{H_3O^+}$ formed in the last reaction cannot gain any more hydrogen atoms, and instead undergoes dissociative recombination with two possible outcomes:

$$\mathrm{H_3O^+ + e^-} \rightarrow \begin{cases} \mathrm{OH + H_2} \\ \mathrm{H_2O + H} \; . \end{cases} \tag{7.10}$$

The first reaction in (7.10) is known from laboratory measurements to occur with a relative probability $p_1 = 0.75$.

Since each ionization of a hydrogen atom by a cosmic-ray proton leads to an OH with probability p_1, the volumetric production rate of OH is $p_1\,\zeta(\mathrm{HI})\,n_{\mathrm{HI}}$. In steady state, this rate is balanced by the destruction of OH. The molecule can react with ambient ions such as $\mathrm{C^+}$, but more frequently dissociates from ultraviolet radiation penetrating the cloud interior. Writing the characteristic photodissociation time as τ_{photo}, we balance creation and destruction to find the desired expression for the rate of cosmic-ray ionization:

$$\begin{aligned} \zeta(\mathrm{HI}) &= \frac{n_{\mathrm{OH}}}{n_{\mathrm{HI}}}\,(p_1\,\tau_{\mathrm{photo}})^{-1} \\ &= 2\times10^{-17}\ \mathrm{s^{-1}} \; . \end{aligned} \tag{7.11}$$

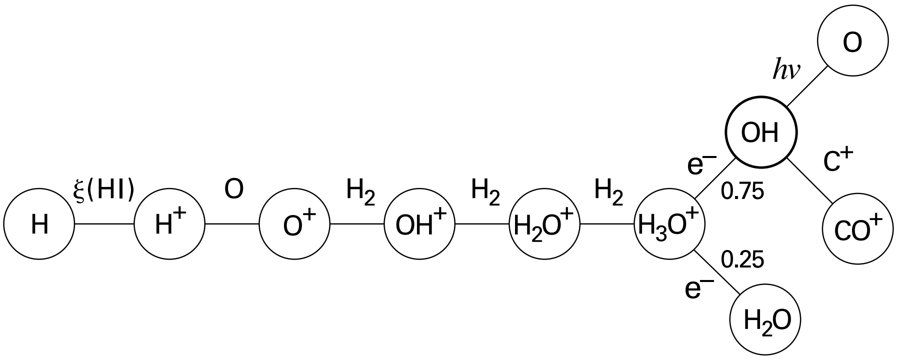


Figure 7.3 Reaction sequence forming OH within diffuse clouds. Note the branching ratios for the two possible outcomes of H_3O^+ dissociative recombination. The OH itself is either photodissociated or destroyed by ions such as C^+.

In the numerical evaluation, we have inserted the theoretical result $\tau_{\rm photo} = 2 \times 10^{10}$ s, relevant for the diffuse clouds of interest. Note that our empirical value of $\zeta(\rm HI)$ includes both direct ionizations by protons and secondary ionizations by ejected electrons. Substituting this value into equation (7.7), we find

$$\Gamma_{\rm CR}({\rm HI}) = 1 \times 10^{-13} \left(\frac{n_{\rm HI}}{10^3\,{\rm cm}^{-3}}\right)\ {\rm eV\,cm^{-3}\,s^{-1}} \tag{7.12}$$

It remains to determine the corresponding rates for molecular hydrogen. Theoretical calculations show that the probability of a hydrogen molecule being ionized by a cosmic-ray proton is 1.6 times the atomic value. Utilizing the enhancement factors from secondary ionizations, we have

$$\begin{aligned}\zeta({\rm H_2}) &= \frac{1.6 \times 1.6 \times \zeta({\rm HI})}{1.7} \\ &= 3 \times 10^{-17}\,{\rm s}^{-1}\ ,\end{aligned} \tag{7.13}$$

so that

$$\Gamma_{\rm CR}({\rm H_2}) = 2 \times 10^{-13} \left(\frac{n_{\rm H_2}}{10^3\,{\rm cm}^{-3}}\right)\ {\rm eV\,cm^{-3}\,s^{-1}}\ . \tag{7.14}$$

Comparison with equation (7.12) gives the simple result that, whether the gas is in atomic or molecular form, the cosmic ray heating rate is the same as measured per hydrogen atom.

7.2 Interstellar Radiation

A second important heating agent for molecular clouds is the diffuse radiation field that permeates interstellar space. We need to understand in detail how these photons impinging on the gas create thermal energy. We should also look to stars embedded within the clouds as additional heating sources.

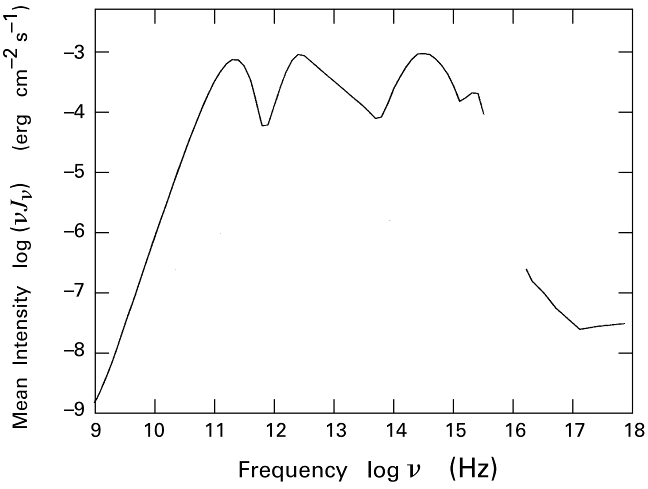


Figure 7.4 Mean intensity of the interstellar radiation field, expressed as a function of frequency.

7.2.1 Major Constituents

Figure 7.4 shows the intensity of the interstellar radiation field as a function of frequency. The form of this distribution is a consequence of the space density and mass spectrum of both stars in the solar neighborhood and of the gas and dust obscuring their light. Thus, we may reliably use Figure 7.4 when considering relatively nearby interstellar clouds, but not to describe conditions out of the plane of the Galaxy or closer to its center. The figure actually plots νJ_ν, where J_ν is the specific intensity averaged over all solid angles in the sky. From the discussion in Chapter 2, the latter quantity is related to the monochromatic energy density u_ν by

$$J_\nu = \frac{c\,u_\nu}{4\,\pi}$$

Integration of the empirical J_ν over all frequencies yields, after using this equation, a total radiation energy density of 1.1 eV cm^{-3}. This figure is intriguingly close to that for the cosmic rays.

The energetically dominant components of the radiation field are those at millimeter, far-infrared, and optical wavelengths, peaking at $\log \nu_{\rm max} = 11.3$, 12.4, and 14.5, respectively, Here each frequency $\nu_{\rm max}$ is measured in Hz. The first component stems from the cosmic background radiation, a blackbody distribution with an associated temperature of 2.74 K. Background photons heat clouds primarily by exciting the lowest rotational transitions in such abundant molecules as CO. The far-infrared component in Figure 7.4 arises from interstellar dust warmed by starlight. Molecular clouds are transparent to this radiation, which is therefore not a heating agent.

The optical component consists of light from field stars. Suppose we crudely model the energy distribution as arising from a diluted blackbody of temperature $\bar{T}$. To estimate this temperature, we identify the peak frequency $\nu_{\rm max}$ as that for which the blackbody specific intensity

$B_\nu(T)$ has a maximum. We thus obtain $\bar{T} = h\nu_{\rm max}/x_\circ k_B = 5400$ K, where we have used the result from § 2.3 that $x_\circ = 2.82$. This temperature is the surface value for a main-sequence G3 star. We may regard this spectral type as an average over the A-dwarfs and K- and M-type red giants that actually dominate the luminosity in the solar neighborhood. A true blackbody radiation field of temperature 5400 K has a peak value for νB_ν of 9.3×10^9 erg cm^{-2} s^{-1} sr^{-1}. To match the observed peak of $\nu J_\nu = 9.1 \times 10^{-4}$ erg cm^{-2} s^{-1} sr^{-1}, the blackbody intensity at every frequency must be multiplied by a *dilution factor* of $W = 1 \times 10^{-13}$. This factor is essentially the fractional solid angle of the sky occupied by stellar surfaces.

Figure 7.4 shows a smaller local maximum in the ultraviolet (log $\nu_{\rm max} = 15.3$). Matching $\nu_{\rm max}$ to blackbody values yields an equivalent temperature of 3.4×10^4 K, with a dilution factor of 1×10^{-17}. To gauge the effect of massive stars on molecular clouds, theorists often consider an ultraviolet flux that is far above the local background. The convention is to assume an isotropic radiation field in this regime, but with an intensity distribution that is scaled upward by some factor, traditionally denoted $G_\circ$.

At the highest frequencies, Figure 7.4 displays a contribution in the soft X-ray regime ($16.5 < \log\,\nu < 16.8$) and a conspicuous lack of data in the extreme ultraviolet ($15.5 < \log\,\nu < 16.5$). Radiation of both types was first observed through rocket experiments in the 1960s and 1970s. The ultraviolet spectrum was later explored by the EUVE (Extreme Ultraviolet Explorer) satellite. A diffuse gas with temperature of order 10^6 K can emit photons in these regimes. Within the context of the three-phase model of the interstellar medium, such a hot plasma gains its energy from supernova remnants and fast winds from massive stars. Note that the stars themselves produce extreme ultraviolet photons. However, this contribution is mostly absorbed through its ionization of hydrogen and helium in the stars' own HII regions or in intervening clouds. To avoid such absorption, the inferred 10^6 K gas must be relatively nearby, probably within a distance of order 100 pc.

7.2.2 Carbon Ionization

The ultraviolet component of the radiation field is too weak at energies above 13.6 eV to significantly ionize the hydrogen or helium in molecular clouds. However, a number of heavier elements have lower ionization potentials. Of these, atomic carbon (C I) is the most abundant, with a number density relative to hydrogen of $n_C/n_H = 3 \times 10^{-4}$ (see Table 2.1). Any photon more energetic than 11.2 eV will eject an electron from C I. Since the kinetic energy of this electron quickly disperses to surrounding atoms through collisions, carbon ionization is an effective heating mechanism. The volumetric heating rate, which we denote as $\Gamma_{\rm CI}$, is then

$$\Gamma_{\rm CI} \,=\, \zeta({\rm C\,I})\, n_C\, \Delta E({\rm C\,I}) \ . \tag{7.15}$$

Here, $\zeta({\rm C\,I})$ is the ionization rate of a single carbon atom and $\Delta E({\rm C\,I})$ the average energy of the ejected electron.

The evaluation of both $\zeta({\rm C\,I})$ and $\Delta E({\rm C\,I})$ entails an integration over frequency of the radiation intensity J_ν, weighted by the ionization cross section of C I. The result is that $\zeta({\rm C\,I}) = 1 \times 10^{-10}$ s^{-1} and $\Delta E({\rm C\,I}) = 1$ eV. Substituting these values into equation (7.15) along with the carbon abundance, yields

$$\Gamma_{\rm CI} = 4 \times 10^{-11} \left(\frac{n_H}{10^3\,{\rm cm}^{-3}}\right)\ {\rm eV\,cm^{-3}\,s^{-1}}\ . \tag{7.16}$$

Note that our result implicitly assumes that carbon is still mostly in neutral form. In practice, the element can be largely ionized in cloud regions where the heating is significant, and one must reduce $\Gamma_{\rm CI}$ appropriately.

7.2.3 Photoelectric Heating

Ultraviolet photons also eject electrons from interstellar dust grains; these electrons heat surrounding gas. The energy required to separate a single electron from the neutral grain surface, *i. e.*, the analog of the ionization potential, is known as the *work function* and is about 6 eV for standard grain composition. The photons actually liberate liberate electrons about 100 Å *inside* the grain (see Figure 7.5). Only about 10 percent of these make their way to the surface. Those that do must still overcome the work function and end up leaving the grain with energies of only about 1 eV. Comparing with the typical photon energy of 10 eV, we see that the net energy efficiency $\epsilon_{\rm PE}$ is only about 0.01. The photoelectric process is nevertheless a major heating source for the gas because of the large grain cross section. We may write the associated rate $\Gamma_{\rm PE}$ in terms of the grain number density and geometric cross section:

$$\Gamma_{\rm PE} = 4\pi\, n_d\, \sigma_d\, \epsilon_{\rm PE} \int_{\rm FUV} J_\nu\, d\nu\ . \tag{7.17}$$

The frequency integration extends over the far-ultraviolet regime above 6 eV, while the factor 4π comes from geometric considerations. For an isotropic specific intensity, $J_\nu = I_\nu$, the flux impinging on a *planar* surface element would be $\int J_\nu \mu d\Omega = 2\pi \int_0^1 J_\nu \mu d\mu = \pi J_\nu$. However, we picture standard grains as being (roughly) spherical in shape. The additional factor of 4 in equation (7.17) is simply the ratio of the surface area of a sphere to the cross-sectional area which that sphere presents to a plane-parallel flux. The quantity σ_d is the cross section in the latter sense.

We saw in Chapter 2 how $\Sigma_d \equiv n_d \sigma_d / n_H$ is determined, through a combination of empirical and theoretical steps, to be 1.5×10^{-21} cm^2. The quantity $4\pi \int_{\rm FUV} J_\nu\, d\nu$ is estimated from observations to be 1.6×10^{-3} erg cm^{-2} s^{-1}, a figure traditionally known as the "Habing flux." Using a representative $\epsilon_{\rm PE}$ of 0.01 in equation (7.17), we find that $\Gamma_{\rm PE}$ has the value $2 \times 10^{-11}\,(n_H/10^3\,{\rm cm}^{-3})\ {\rm eV\,cm^{-3}\,s^{-1}}$. This estimate, however, ignores the influence of very small grains, including PAHs. For these, the ejected electrons can more easily reach the surface, *i. e.*, the efficiency factor is higher. In addition, the number density of such grains is relatively high. Recall that the abundance of 10 Å grains exceeds that of 0.1 µm particles by $(10^{-7}/10^{-5})^{-3.5} = 10^7$. Despite the small cross section of each PAH, their cumulative effect is thus appreciable. By integrating over a realistic grain size distribution, we arrive at a final heating rate of

$$\Gamma_{\rm PE} = 3 \times 10^{-11} \left(\frac{n_H}{10^3\,{\rm cm}^{-3}}\right)\ {\rm eV\,cm^{-3}\,s^{-1}}\ . \tag{7.18}$$

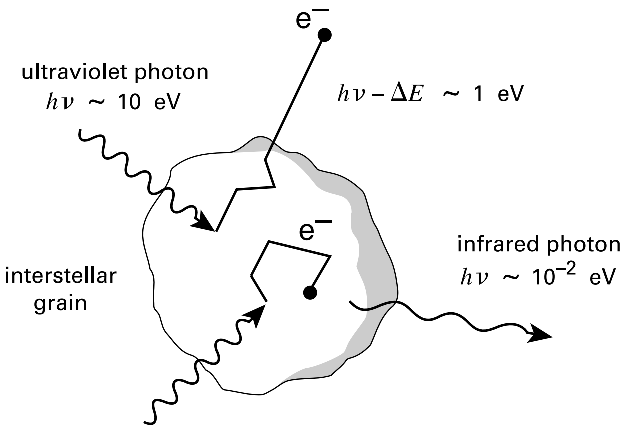


Figure 7.5 Interaction of ultraviolet photons with interstellar grains. The electron that is ejected sometimes escapes from the grain surface, but more often deposits its energy to the lattice, which in turn radiates infrared photons.

This figure assumes that each grain is electrically neutral, a condition that is violated in sufficiently strong ultraviolet radiation fields. We will consider the resulting alteration to $\Gamma_{\rm PE}$ when we discuss the photodissociation regions near massive stars in Chapter 8.

7.2.4 Irradiation of Grains

The liberated electrons that do *not* leave a grain impart their energy through collisions to the lattice (Figure 7.5). Most of the ultraviolet radiation, therefore, serves to raise the grain temperature T_d. Somewhat more dust heating is provided by the stronger flux of optical photons, which also excite internal electrons. Counting only this visible component, the total dust heating rate per unit volume is

$$\Gamma_d = 4\pi\, n_d\, \sigma_d \int_{\rm VIS} Q_{\nu,\rm abs} J_\nu d\nu \ . \tag{7.19}$$

Here the grain absorption efficiency $Q_{\nu,\rm abs}$ is proportional to ν in the visual regime. As before, we model J_ν as arising from a diluted blackbody at the characteristic temperature $\bar{T}$ and peak frequency $\nu_{\rm max}$. Equation (7.19) then becomes

$$\Gamma_d = 4\pi\, W\, n_d\, \sigma_d\, Q_{\nu_{\rm max}} \int_0^\infty \left(\frac{\nu}{\nu_{\rm max}}\right) \frac{2h\nu^3/c^2}{\exp(h\nu/k_B\bar{T}) - 1}\, d\nu \ ,$$

where W is the dilution factor. We define the nondimensional variable $x \equiv h\nu/k_B\bar{T}$ and find

$$\Gamma_d = \frac{8\pi\, W\, Q_{\nu_{\rm max}} n_d\, \sigma_d\, k_B \bar{T} \nu_{\rm max}^3}{x_\circ^4\, c^2} \int_0^\infty \frac{x^4}{e^x - 1}\, dx \ .$$

The nondimensional integral has the numerical value 24.9. We next use the relation between $n_d\sigma_d$ and n_H, and set $Q_{\nu_{\rm max}}$ equal to 0.1, the value appropriate for the optical peak frequency of $\nu_{\rm max} = 3 \times 10^{14}\ {\rm s}^{-1}$. (Recall equation (2.40) and Figure 2.10.) The expression for Γ_d then becomes

$$\Gamma_d = 2 \times 10^{-9} \left(\frac{n_H}{10^3\,{\rm cm}^{-3}}\right)\ {\rm eV\,cm^{-3}\,s^{-1}}\ . \tag{7.20}$$

We emphasize that this relation governs the heating of the *dust grains*, not the gas. The grains can transfer energy to the gas only at cloud densities high enough for good collisional coupling between the two components.

We have not yet mentioned the effect of interstellar radiation on the dominant constituent of star-forming clouds, molecular hydrogen itself. As we saw in Chapter 5, the absorption of a photon with energy $h\nu$ greater than 11.2 eV promotes H_2 to an excited electronic state. Most often, the excited molecule drops to the electronic ground state and then cascades down the rovibrational levels within that state. In quiescent molecular clouds, this decay occurs through the emission of ultraviolet and infrared photons, but the energy can be given to other colliding species at the high densities and temperatures behind shocks. If the excited H_2 instead dissociates, it emits a photon of energy $h\nu - \Delta E_{\rm diss} - \epsilon$ and imparts kinetic energy ϵ to the separate atoms. On average, ϵ is about 2 eV. This energy quickly spreads into the gas through collisions.

In a quiescent cloud of molecular hydrogen exposed to the full interstellar radiation field, this secondary effect of H_2 dissociation would completely dominate the heating. As we have already noted, however, such a cloud cannot exist. Molecular hydrogen is only found at a depth below the cloud surface where the ultraviolet flux has already been severely attenuated by excitation and dissociation of the outer H_2 molecules and by dust absorption. Consequently, heating by dissociation is relegated to a minor role.

7.2.5 Stellar X-Rays

Molecular clouds are also heated by the radiation from new stars they create. Inside a large complex, the densest regions are dispersed by winds and outflows generated from within, but lower-density gas can remain. High-mass stars emit copiously in the ultraviolet. This radiation heats nearby dust by direct absorption and gas through the grain photoelectric effect. The $G_\circ$-factor measuring the enhancement of the ultraviolet flux above interstellar may be as high as 10^6 for a cloud neighboring an O or B star. In contrast, low-mass pre-main-sequence objects, *i. e.*, T Tauri stars, emit most of their luminosity at optical and near-infrared wavelengths. This radiation is readily absorbed by dust grains, at a rate given by equation (7.19), but with an appropriately enhanced J_ν. At the typical densities of molecular cloud interiors, the grains radiate away this energy before they can transfer it to the gas through collisions.

About 10^{-4} of the luminosity from low-mass stars is in X-rays, which *do* heat the gas directly. X-rays interact with molecular cloud gas by ionizing its constituent atoms. Photons with energy below about 0.5 keV mostly ionize hydrogen and helium, while more energetic ones eject the innermost (K-shell) electrons from heavy elements. The total cross section for photoionization, σ_ν, falls approximately as ν^{-3}, apart from upward jumps whenever the photon energy matches some ionization threshold. Recall from Chapter 2 that this falloff with frequency contrasts directly with the behavior of infrared and optical photons encountering interstellar dust.

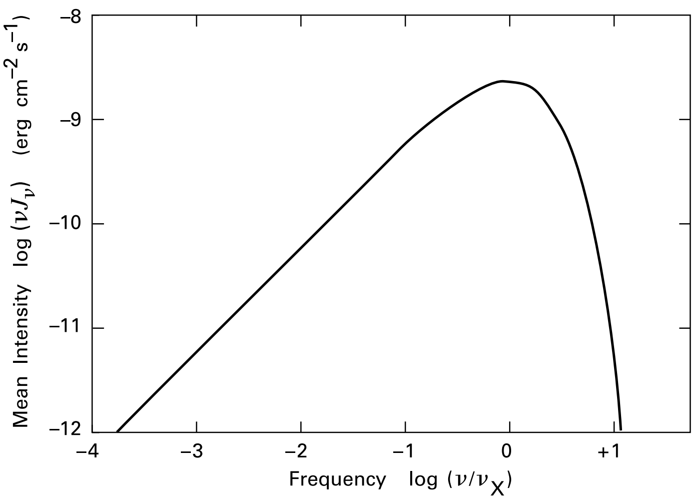


Figure 7.6 Mean intensity of X-rays at a distance of 0.1 pc from a pre-main-sequence star. The star has a total X-ray luminosity of $L_X = 1 \times 10^{30}$ erg s^{-1}. The radiation is assumed to be thermal bremsstrahlung from a plasma with temperature $T_X = 1 \times 10^7$ K.

The X-rays in T Tauri stars originate from a hot plasma of temperature $T_X \approx 1 \times 10^7$ K, or a value of $k_B T_X$ near 1 keV. It is the acceleration of the randomly moving electrons in this plasma that actually provides the radiation; this emission mechanism is called *thermal bremsstrahlung*. Very few of these electrons have kinetic energies far above $k_B T_X$. Consequently, the mean intensity J_ν produced by the plasma, which is nearly independent of frequency for $\nu \lesssim \nu_X \equiv k_B T_X/h$, falls off rapidly at higher frequencies. Figure 7.6 shows νJ_ν from thermal bremsstrahlung, as would be detected close to a typical low-mass pre-main-sequence star. This intensity plot ignores any interstellar absorption. In fact, stellar X-rays emitted into a cloud of density n_H deposit most of their energy within a characteristic distance r_X given by

$$\begin{aligned} r_X &\equiv (n_H\, \sigma_X)^{-1} \\ &= 2 \left(\frac{n_H}{10^3\,\mathrm{cm}^{-3}}\right)^{-1} \mathrm{pc}\ . \end{aligned} \tag{7.21}$$

Here, σ_X, is the absorption cross section of each hydrogen atom at ν_X, or 2×10^{-22} cm^2 for $k_B T_X = 1$ keV.

Consider now the volumetric heating rate within a sphere of radius r_X surrounding a star with total X-ray luminosity L_X (Figure 7.7). Absorption causes the specific intensity at any frequency to fall exponentially with distance. Taking the intrinsic J_ν to be flat for $\nu \leq \nu_X$, the flux of photons $F_\nu \Delta\nu$ at radius r with frequency between ν and $\nu + \Delta\nu$ must be

$$F_\nu \Delta\nu = \left(\frac{\Delta\nu}{\nu_X}\right) \frac{L_X}{4\pi r^2} \exp\left(-\tau_\nu\right)\ .$$

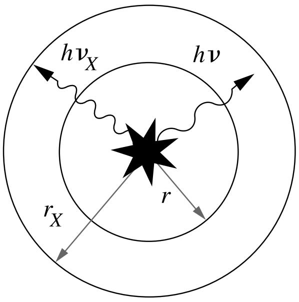


Figure 7.7 Cloud heating by stellar X-rays. Photons with the maximum frequency ν_X can penetrate out to a maximum radial distance r_X. Photons with frequency $\nu < \nu_X$ see a larger cross section. At any interior radius r, some of these are absorbed and the rest stream outward.

Here the optical depth τ_ν is given by $n_H \sigma_\nu r$, where the cross section σ_ν is $\sigma_X\, (\nu_X/\nu)^3$. Let us assume that the energy of each ionizing photon is converted entirely to heat. Then to obtain the volumetric heating rate at distance r, we first multiply the flux by $n_H\, \sigma_\nu$. Integrating over frequency, we find

$$\Gamma_X = \frac{n_H\, L_X}{4\pi\, r^2\, \nu_X} \int_0^{\nu_X} d\nu\, \sigma_\nu \exp(-\tau_\nu) \ . \tag{7.22}$$

To evaluate this last expression, we first let τ_X represent the optical depth to r at $\nu = \nu_X$, *i. e.*, $\tau_X \equiv n_H \sigma_X r$. From this definition, we also have $\tau_\nu = \tau_X\, (\nu_X/\nu)^3$. We next define a new variable $y \equiv \nu/\nu_X$, so that equation (7.22) becomes

$$\Gamma_X = \frac{n_H^3\, \sigma_X^3\, L_X}{4\pi\, \tau_X^2} \int_0^1 dy\, y^{-3} \exp(-\tau_X y^{-3}) \ .$$

The exponential that appears on the righthand side, and therefore Γ_X itself, vanishes for distances well beyond r_X, *i. e.*, for $\tau_X \gg 1$. To evaluate Γ_X interior to this radius, we first note that the integrand goes to zero for both small and large y, and peaks at $y_\circ = \tau_X^{1/3}$. We may crudely approximate the function as a Gaussian centered at $y = y_\circ$, with the correct height and curvature at that point. It is then convenient to extend the integration limits from $y = -\infty$ to $y = +\infty$. The heating rate then reduces to

$$\begin{aligned}\Gamma_X &= \frac{1}{6e\sqrt{2\pi}} \frac{L_X}{\tau_X^{8/3}\, r_X^3} \\ &= 2 \times 10^{-13} \left(\frac{n_H}{10^3\,\mathrm{cm^{-3}}}\right)^{1/3} \left(\frac{L_X}{10^{30}\,\mathrm{erg\,s^{-1}}}\right) \left(\frac{r}{0.1\,\mathrm{pc}}\right)^{-8/3} \mathrm{eV\,cm^{-3}\,s^{-1}} \ ,\end{aligned} \tag{7.23}$$

We stress that this formula only applies for $r < r_X$. The reader may verify that a volume integration of Γ_X out to this radius yields a total heat input rate of $\sqrt{2\pi}/e\, L_X$, close to the correct value of L_X.

7.3 Cooling by Atoms

We now turn our attention to the mechanisms by which molecular clouds lose energy to interstellar space. Within a quiescent cloud, neither hydrogen nor helium can radiate away an appreciable fraction of the total thermal energy content. Hence, any cooling of the gas is necessarily indirect and relies on the presence of minor constituents that *do* act as effective radiators. Hydrogen and helium collide inelastically with ambient atoms, molecules, and dust grains, exciting internal degrees of freedom. These excitations decay through the emission of photons.

7.3.1 Density Dependence

Consider first a simplified atom with only an upper and lower energy level, separated by $\Delta E_{\rm ul}$. If this species is excited by collisions with hydrogen at the rate $n_H \gamma_{\rm lu}$ per atom, what is $\Lambda_{\rm ul}$, the volumetric rate of cooling? We know that, for $n_H \ll n_{\rm crit} \equiv A_{\rm ul}/\gamma_{\rm ul}$, each upward collisional excitation is followed promptly by a downward radiative transition. In this low-density regime, therefore, the cooling rate by hydrogen impact is

$$\begin{aligned}\Lambda_{\rm ul}\,(n_H \ll n_{\rm crit}) &= n_l\, n_H\, \gamma_{\rm lu}\, \Delta E_{\rm ul} \\ &= \frac{g_u}{g_l} n_l\, n_H\, \gamma_{\rm ul}\, \Delta E_{\rm ul}\, \exp(-T_\circ/T_g)\ ,\end{aligned} \tag{7.24}$$

where n_l is the population of the lower level. Here, we have used equation (B.4) to relate $\gamma_{\rm lu}$ to the less temperature-sensitive rate $\gamma_{\rm ul}$ through the gas temperature $T_g \equiv T_{\rm kin}$. Since only a small portion of the atoms or molecules is excited at any time, the product $n_l n_H$ appearing in (7.24) is quite accurately given by the product of n_H^2 and the fractional number abundance of the coolant.

At higher densities, $n_H \gg n_{\rm crit}$, each upward collisional excitation is usually followed by collisional deexcitation rather than by radiative decay. Under these conditions, the two level populations are in LTE, and no longer depend on n_H. The cooling rate is the product of n_u and $A_{\rm ul}\Delta E_{\rm ul}$, the energy loss rate from the upper level:

$$\begin{aligned}\Lambda_{\rm ul}\,(n_H \gg n_{\rm crit}) &= n_u\, A_{\rm ul}\, \Delta E_{\rm ul} \\ &= \frac{g_u}{g_l} n_l\, A_{\rm ul}\, \Delta E_{\rm ul}\, \exp(-T_\circ/T_g)\ .\end{aligned} \tag{7.25}$$

We see that $\Lambda_{\rm ul}$ in this supercritical regime may formally be obtained from the subcritical result by replacing n_H with $n_{\rm crit}$. It is therefore often said that collisions "quench" atomic cooling. This terminology is somewhat misleading, as $\Lambda_{\rm ul}$ still rises with increasing density, albeit at a slower rate.

7.3.2 Fine-Structure Splitting

These general considerations find practical application in the fact that many atoms have low-lying *fine-structure levels* that are prone to collisional excitation. The existence of such levels stems from the spin-orbit interaction. Following our discussion of the hydrogen atom in § 2.1, the relative motion of any orbiting electron and a charged nucleus creates a torque on the magnetic moment associated with the electron's intrinsic spin. The electrons of hydrogen and helium are described by single-particle wavefunctions that have no associated orbital angular momentum ($l = 0$). Hence, the electronic ground states of these atoms lack the internal torque and exhibit no fine-structure splitting. We must turn to oxygen, the next most abundant element.

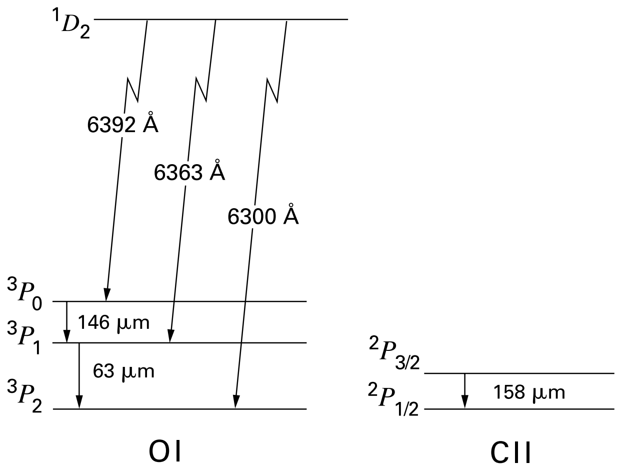


Figure 7.8 Fine-structure splitting of the electronic ground states of O I and C II. The far-infrared lines associated with transitions between these levels are indicated. Also shown in the case of O I are the optical transitions from the first excited electronic state. No optical lines arise from the analogous transition in C II.

Oxygen exists in neutral, atomic form (O I) throughout large regions of molecular clouds. In its electronic ground state, O I has four p-electrons ($l = 1$). Their individual orbital angular momenta combine vectorially to yield a total orbital quantum number $L = 1$, while the electron spins add together to $S = 1$. Although other values for L and S are possible, all result in greater Coulomb repulsion between the electrons and hence higher energy. The electronic ground state of O I is symbolized spectroscopically as ^3P, where the symbol P denotes $L = 1$ and the superscript is equal to $2S + 1$. As sketched in Figure 7.8, the first excited state, corresponding to $L = 2$ and $S = 0$, is denoted ^1D.

The figure also shows that the ^3P state is actually a *multiplet* of three different levels of slightly different energy. Each level is distinguished by its degree of spin-orbit interaction. One may picture the orbital and spin angular momentum vectors, $\boldsymbol{L}$ and $\boldsymbol{S}$, precessing about their fixed sum, $\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}$. (Compare the discussion of the OH molecule in § 5.5.) The spin-orbit energy is proportional to $\boldsymbol{L}{\cdot}\boldsymbol{S}$, which remains constant during precession. By convention, each level is labeled with the subscript J, the magnitude of the total angular momentum. This number can take the values 0, 1, or 2 for the ^3P state, but is restricted to 2 for the first excited (^1D) state. The $^3P_2$ level of O I is the true ground state, while the $^3P_0$ level is highest in energy. This ordering of energies within the multiplet is called *inverted*, since, in the semi-classical model, $\boldsymbol{L}$ and $\boldsymbol{S}$ prefer to be antiparallel, implying that J is normally minimized in the ground state.[2]

---

[2] In the electron's reference frame, the nucleus creates a magnetic field $\boldsymbol{B}$ parallel to $\boldsymbol{L}$. For an electron with spin magnetic moment $\boldsymbol{\mu}_s$, the lowest energy occurs when $\boldsymbol{\mu}_s$ is parallel to $\boldsymbol{B}$, and hence to $\boldsymbol{L}$. Because of the electron's negative charge, $\boldsymbol{\mu}_s$ is antiparallel to $\boldsymbol{S}$. Therefore, $\boldsymbol{L}$ and $\boldsymbol{S}$ are expected to be antiparallel.

7.3.3 Emission from O I and C II

The energy difference between the $^3P_1$ and $^3P_2$ levels of O I is 2.0×10^{-2} eV, corresponding to a temperature $T_\circ$ of 230 K. The upper level can be collisionally excited in the warmer regions of a molecular cloud and decays with an associated A-value of 9.0×10^{-5} s^{-1}. Note that all the downward fine-structure transitions are "forbidden" by the electric-dipole selection rules of quantum mechanics and therefore occur through slower *magnetic* dipole interactions. For the $^3P_1 \rightarrow {}^3P_2$ transition of O I, the emitted spectral line is symbolized [O I] 63 µm, where the braces denote a forbidden transition. Except in the very densest cloud regions, this far-infrared radiation readily escapes and provides an important source of cooling. The deexcitation rate $\gamma_{\rm ul}$ from hydrogen impact has the value 4×10^{-11} cm^3 s^{-1} at the temperatures of interest ($T_g \lesssim 40$ K), so that $n_{\rm crit} = 2 \times 10^6$ cm^{-3}. We may therefore evaluate $\Lambda_{\rm OI}$, the volumetric cooling rate, from the subcritical expression, equation (7.24). Using a number density of oxygen relative to hydrogen of 4×10^{-4} (Table 2.1) and noting that the degeneracy of each J-state is $2J + 1$, we find

$$\Lambda_{\rm OI} = 2 \times 10^{-10} \left(\frac{n_H}{10^3\,{\rm cm}^{-3}}\right)^2 \exp\left(-\frac{230\,{\rm K}}{T_g}\right) \ {\rm eV\,cm^{-3}\,s^{-1}} \ . \tag{7.26}$$

This expression would need to be modified in denser regions, where much of the oxygen is in CO or grain mantles. In such locales, however, other cooling mechanisms dominate (Chapter 8).

Fine-structure cooling is also available from carbon, which has an elemental abundance comparable to oxygen. As we have seen, atomic carbon can be readily ionized by the ultraviolet component of the interstellar radiation field, so that it is actually C II that provides most of the cooling. The ion has only one p-electron, and the spin-orbit interaction splits its ground state into $^2P_{3/2}$ and $^2P_{1/2}$ levels. Figure 7.8 shows that, in this case, the energy ordering is normal, *i. e.*, the $^2P_{1/2}$ level is actually the ground state. The energy difference between the levels is 7.93×10^{-3} eV, corresponding to a $T_\circ$ of 92 K and a photon wavelength of 158 µm.

The appropriate value of $\gamma_{\rm ul}$ is now 6×10^{-10} cm^3 s^{-1}, and $A_{\rm ul}$ is 2.4×10^{-6} s^{-1}. Thus, the critical density is only 3×10^3 cm^{-3}, a value often reached in molecular clouds. However, most of the carbon is locked into CO at higher density, so that the subcritical cooling rate is still appropriate. We find

$$\Lambda_{\rm CII} = 3 \times 10^{-9} \left(\frac{n_H}{10^3\,{\rm cm}^{-3}}\right)^2 \exp\left(-\frac{92\,{\rm K}}{T_g}\right) \ {\rm eV\,cm^{-3}\,s^{-1}} \ . \tag{7.27}$$

Both Equations (7.26) and (7.27) have assumed that promotion to upper fine-structure levels occurs only through impact with ambient hydrogen atoms. In fact, the levels are much more easily excited by free *electrons*. It is only because the fractional ionization throughout quiescent molecular clouds is very low that we may safely ignore the electronic contribution.

7.4 Cooling by Molecules and Dust

Turning from atoms to molecules, it is the closely spaced rotational levels that are readily populated through collisions. At subcritical densities, these excitations decay quickly through radiation, but the emitted photons may suffer absorption within the cloud. We need to account carefully for this radiative trapping to assess the net energy loss. Another important cloud constituent is the dust, which can either heat or cool the gas.

7.4.1 Trapping in CO Lines

Although a number of molecules, such as H_2O and O_2, act as cooling agents, the dominant one is CO. Here, the $J = 1 \rightarrow 0$ rotational transition is always optically thick in molecular clouds. The relevant cross sections are so large, in fact, that a number of higher transitions are also optically thick. Internal trapping reduces the emission in any given line, while enhancing the populations of the higher levels above those attainable through collisions alone.

Consider now an idealized cloud of spatially uniform density and temperature. What is the optical depth at line center for the $J = 1 \rightarrow 0$ transition? By using equation (6.9), now applied to the main isotope $^{12}C^{16}O$, we may write τ_{10} in terms of the column density $N_{\rm CO}$:

$$\tau_{10} = \left(\frac{g_1}{g_0}\right) \frac{A_{10}\, N_{\rm CO}\, c^3}{8\pi\, \nu_{10}^3\, Q\, \Delta V} \left[1 - \exp\left(-\frac{h\nu_{10}}{k_B T_g}\right)\right] \; . \tag{7.28}$$

Here we have denoted the line-center frequency by ν_{10} and have set $T_{\rm ex}$ equal to T_g. As we remarked in § 6.2, this LTE assumption is justified once the line radiation becomes optically thick. In writing (7.28), we have also assumed that the broadening of the spectral line is mainly due to bulk motion in the cloud interior. Thus, we have set the line width $\Delta\nu_{10}$ equal to $\nu_{10}\, \Delta V/c$, where ΔV is the internal velocity dispersion. For the partition function Q, we have used equation (C.18) to write $Q = 2k_B T_g/h\nu_{10}$. After expanding the exponential, whose argument is less than unity, we may evaluate τ_{10} numerically as

$$\tau_{10} = 8 \times 10^2 \left(\frac{N_H}{1 \times 10^{22}\ {\rm cm}^{-2}}\right) \left(\frac{T_g}{10\ {\rm K}}\right)^{-2} \left(\frac{\Delta V}{1\ {\rm km\ s}^{-1}}\right)^{-1} \; . \tag{7.29}$$

This expression assumes that nearly all the carbon, with its number abundance of 3×10^{-4} relative to hydrogen, is in the form of CO.

Despite the very large value of τ_{10} under typical cloud conditions, the optical depth for higher transitions, $\tau_{J+1,J}$, falls quickly with increasing J. The reason is the decline in N_J, the column density in the J-level. To quantify the trend, we generalize equation (7.29), utilizing the result for quantum mechanical rotators that $A_{J+1,J}$ scales as $(J+1)/(2J+3)\, \nu_{J+1,J}^3$. Applying the LTE assumption to N_J and using equation (5.6) to obtain the J-dependence of $\nu_{J+1,J}$, we find

$$\tau_{J+1,J} = \tau_{10} \left(\frac{J+1}{2J+1}\right) \frac{1 - \exp[-(J+1)/\theta]}{1 - \exp(-1/\theta)} \exp\left[-\frac{J(J+1)}{2\theta}\right] \; . \tag{7.30}$$

Here, θ is a dimensionless temperature:

$$\theta \equiv \frac{k_B\, T_g}{h\nu_{10}} = \frac{T_g}{5.5\,{\rm K}} \; . \tag{7.31}$$

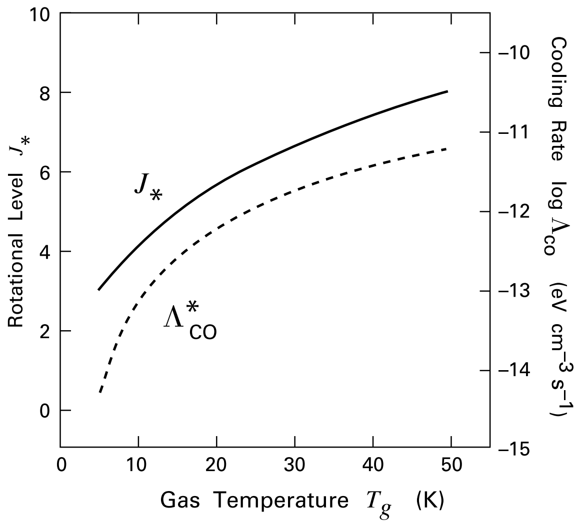


Figure 7.9 Plot of J_*, the highest CO rotational level creating an optically thick line, shown as a function of gas temperature. Also plotted is the volumetric cooling rate $\Lambda^*_{\rm CO}$. Here we have displayed J_* as a smoothly varying real number, to avoid artificial jumps in the associated cooling rate.

At fixed θ, the presence of the Boltzmann factor $\exp[-J(J+1)/2\theta]$ ensures that $\tau_{J+1,J}$ falls to unity at a modest J-value, which we denote as J_*. We may find this critical level by solving

$$\begin{aligned}\exp\left[\frac{J_*(J_*+1)}{2\theta}\right] &= \tau_{10}\left(\frac{J_*+1}{2J_*+1}\right)\frac{1-\exp[-(J_*+1)/\theta]}{1-\exp(-1/\theta)}\\ &= \frac{3\,A_{10}\,N_{\rm CO}\,c^3}{16\,\pi\,\nu_{10}^3\,\Delta V}\left(\frac{J_*+1}{2J_*+1}\right)\frac{1-\exp[-(J_*+1)/\theta]}{\theta}\ .\end{aligned} \tag{7.32}$$

In the second form of this equation, we have inserted the expression for τ_{10} from (7.28). Figure 7.9 shows J_* as a function of temperature, computed for a cloud with $\Delta V = 1\ {\rm km\ s^{-1}}$, $n_{\rm H_2} = 1\times10^3\ {\rm cm^{-3}}$, and a diameter of 1 pc, *i. e.*, with an associated hydrogen column density of $N_H = 6\times10^{21}\ {\rm cm^{-2}}$. In this example, $N_{\rm CO} = 2\times10^{18}\ {\rm cm^{-2}}$. The rise in J_* with temperature demonstrates the increasing importance of radiative trapping.

7.4.2 CO Cooling

All those transitions with $J \lesssim J_*$ are both optically thick and have level populations in LTE. Deep within the cloud, the photons at these frequencies transport energy by diffusion from one region to another, but only escape at the cloud surface. Here, the flux is $F_{\rm CO}(J+1,J) = \pi B_\nu(T_g)\Delta\nu_{J+1,J}$, where the Planckian specific intensity $B_\nu(T_g)$ is evaluated at line center. That is,

$$F_{\rm CO}(J+1,J) = \frac{2\,\pi\,h\nu_{J+1,J}^4}{c^3}\,\frac{1}{\exp\left(h\nu_{J+1,J}/k_B T_g\right)-1}\,\Delta V\ . \tag{7.33}$$

After rewriting $\nu_{J+1,J}$ in terms of ν_{10}, we find

$$\begin{aligned} F_{\rm CO}(J+1,J) &= \frac{2\pi h\nu_{10}^4}{c^3}\,\frac{(J+1)^4}{\exp[(J+1)/\theta]-1}\,\Delta V \\ &= 4\times 10^{-20}\,\frac{(J+1)^4}{\exp[(J+1)/\theta]-1}\left(\frac{\Delta V}{1\ {\rm km\,s^{-1}}}\right)\ {\rm eV\,cm^{-2}\,s^{-1}}\ . \end{aligned} \tag{7.34}$$

At fixed T_g and ΔV, equation (7.34) shows that $F_{\rm CO}(J+1,J)$ is a function only of the variable $\alpha \equiv (J+1)/\theta$. The cooling flux initially rises as α^3, peaks at $\alpha \approx 4$, and thereafter declines. It would seem, then, that the peak flux might emanate from that transition with $J \approx 4\theta - 1$. In practice, however, this line is always optically thin, with associated level populations so far below LTE that it contributes little to the net cooling. The true maximum in emission occurs at the lower value J_* given implicitly by equation (7.32) and depicted in Figure 7.9. Because of the steep rise in emission with α (or J_*), much of the total CO cooling arises from this single, critical line.

To obtain the luminosity from the $J_*+1 \rightarrow J_*$ transition, we multiply the flux in equation (7.34) by the cloud surface area. Alternatively, since the line is only marginally optically thick, we may calculate a volumetric cooling rate $\Lambda_{\rm CO}^*$. This form is more convenient when comparing the rate to that from other sources. For the volumetric form, we may employ our previous results for two-level atoms. Assuming that radiative trapping can maintain the J_*+1 level in LTE, we use equation (7.25) to find

$$\begin{aligned} \Lambda_{\rm CO}^* &= \frac{f(J_*)\,\Delta E_{10}\,A_{10}\,n_{\rm CO}}{2\theta}\,\exp\left[-\frac{(J_*+1)(J_*+2)}{2\theta}\right] \\ &= 5\times 10^{-12}\,\frac{f(J_*)}{\theta}\,\exp\left[-\frac{(J_*+1)(J_*+2)}{2\theta}\right]\left(\frac{n_H}{10^3\ {\rm cm^{-3}}}\right)\ {\rm eV\,cm^{-3}\,s^{-1}}\ , \end{aligned} \tag{7.35}$$

where

$$f(J_*) \equiv \frac{(J_*+1)^5}{2J_*+1}\ .$$

The reader may verify that $\Lambda_{\rm CO}^*$, when multiplied by the cloud volume, yields a luminosity which matches that implied by equation (7.34) to within a factor of order unity.

In summary, the volumetric cooling rate from a cloud with given N_H, T_g, and ΔV may be found by first estimating J_* from equation (7.32). We then obtain the radiative emission from the dominant $J_*+1 \rightarrow J_*$ line by using equation (7.35). For a more accurate assessment of $\Lambda_{\rm CO}$, the *total* CO cooling, one should add the emission from several optically thick lines, *i. e.*, those with $J < J_*$. Here one needs to specify the cloud surface area. Figure 7.9 plots, in addition to J_*, $\Lambda_{\rm CO}^*$ as a function of gas temperature for the same cloud parameters. Notice how the cooling rate increases exponentially with T_g, as radiative trapping excites more molecules into the critical J_*-level.

7.4.3 Thermal Effect of Dust

Interstellar dust grains also serve as important coolants. In this case, collisions with gas atoms and molecules lead to lattice vibrations, which decay through the emission of infrared photons. As we have seen, the same grains are also heated through the absorption of optical and ultraviolet photons. Thus, the grain temperature, T_d, is generally different from the gas temperature T_g.

Let us first consider Λ_d, the volumetric cooling rate due to emission from the grains themselves. Note that a volumetric rate is appropriate here since the cloud is usually optically thin to these infrared photons. We may determine Λ_d from the general formula for thermal emission in equation (2.30), after replacing $\rho\kappa_{\nu,\mathrm{abs}}$ by the equivalent $n_d\sigma_d Q_{\nu,\mathrm{abs}}$. Integrating $(j_\nu)_{\mathrm{therm}}$ over all solid angles and noting that the radiation is isotropic, we find

$$\Lambda_d = 4\pi\, n_d\, \sigma_d \int_0^\infty Q_{\nu,\mathrm{abs}}\, B_\nu(T_d) d\nu \ . \tag{7.36}$$

For a typical grain temperature of 30 K, $B_\nu(T_d)$ peaks at a wavelength of about 100 μm, much greater than grain dimensions. As we noted in Chapter 2, the absorption efficiency $Q_{\nu,\mathrm{abs}}$ at long wavelengths tends to the quadratic form

$$Q_{\nu,\mathrm{abs}} = Q_{\nu_{\mathrm{max}}} (\nu/\nu_{\mathrm{max}})^2 \ . \tag{7.37}$$

Using this last relation in equation (7.36), along with $n_d\, \sigma_d = \Sigma_d\, n_H$, we find

$$\begin{aligned}\Lambda_d &= \frac{4\pi\, Q_{\nu_{\mathrm{max}}}\, \Sigma_d\, n_H}{\nu_{\mathrm{max}}^2} \int_0^\infty \nu^2\, B_\nu(T_d) d\nu \\ &= \frac{8\pi\, Q_{\nu_{\mathrm{max}}}\, \Sigma_d\, n_H\, k_B^6\, T_d^6}{h^5\, c^2\, \nu_{\mathrm{max}}^2} \int_0^\infty \frac{x^5}{e^x - 1} dx \ .\end{aligned} \tag{7.38}$$

To evaluate Λ_d numerically, we also recall from Chapter 2 that $Q_{\nu_{\mathrm{max}}}$ may be written in terms of the corresponding opacity as $\mu m_H\, \kappa_{\nu_{\mathrm{max}}}/\Sigma_d$. Here, μ is the mean mass per particle relative to the atomic hydrogen value m_H (see Equations (2.2) and (2.38)). For the opacity, we employ the theoretical result that $\kappa = 0.34\,\mathrm{cm}^2\,\mathrm{g}^{-1}$ at a wavelength of 100 μm ($\nu_{\mathrm{max}} = 3.0 \times 10^{12}\ \mathrm{s}^{-1}$). Using $\mu = 1.3$ for an HI gas and noting that the nondimensional integral has a value of 122, we find

$$\Lambda_d = 1 \times 10^{-10} \left(\frac{n_H}{10^3\,\mathrm{cm}^{-3}}\right) \left(\frac{T_d}{10\,\mathrm{K}}\right)^6 \ \mathrm{eV\,cm^{-3}\,s^{-1}} \ . \tag{7.39}$$

In writing the last equation, we have implicitly assumed that grains of all size have the same temperature. This cannot generally be true. A grain bathed by ultraviolet light, for example, has an absorption Q that is independent of the grain radius a, as long as the grain diameter exceeds the incident wavelength. On the other hand, this same grain emits in the infrared, where the associated Q varies as a. Smaller grains therefore need to be warmer in order to compensate for their lower intrinsic emission rates. In practice, however, the range in temperature is modest enough that an average T_d is still a useful concept.

Finally, we discuss gas cooling by collisions with grains. We stress that this process represents the transfer of energy between two components *within* a cloud, rather than a direct loss to interstellar space. Knowledge of this rate, which we denote $\Lambda_{g\rightarrow d}$, is often essential to establishing the temperature of both gas and grains. A single grain is struck by a hydrogen molecule once in the time $t_{\rm coll}$ given in equation (5.9), where we now replace each HI subscript by H_2. The impacting molecule brings with it translational kinetic energy $(3/2)k_B T_g$, which it imparts to the grain lattice.[3] Assuming there is time for the molecule to reach thermal equilibrium with the lattice before departing, it leaves with energy $(3/2)k_B T_d$. The net cooling rate for the gas is therefore

$$\begin{aligned}\Lambda_{g\rightarrow d} &= \frac{3}{2}\,k_B\,(T_g - T_d)\,\frac{n_d}{t_{\rm coll}} \\ &= \frac{3}{4}\,k_B\,\Sigma_d\,n_H^2\,v_{\rm H_2}\,(T_g - T_d) \\ &= 2\times 10^{-14}\left(\frac{n_H}{10^3\,{\rm cm^{-3}}}\right)^2\left(\frac{T_g}{10\,{\rm K}}\right)^{1/2}\left(\frac{T_g - T_d}{10\,{\rm K}}\right)\ {\rm eV\,cm^{-3}\,s^{-1}}\ .\end{aligned} \tag{7.40}$$

Chapter Summary

The most important external source of gas heating for opaque clouds is cosmic rays. Relativistic protons ionize H_2, liberating electrons which then dissociate other, still intact molecules. The creation of ions within a cloud facilitates reactions that produce most molecules. One of these is OH, measurement of whose abundance allows one to determine empirically the cosmic-ray ionization rate.

Ultraviolet radiation is another source of thermal energy. The flux from field stars readily ionizes carbon in HI clouds. Again, it is the ejected electron that actually provides heating. The same radiation can also liberate electrons directly from the surfaces of dust grains. Those photons absorbed internally by the dust serve to raise its temperature to a value generally different from that of the gas. The ultraviolet radiation from a massive star can warm gas out to a great distance. Low-mass stars, on the other hand, provide heating through their X-rays. Here, the effect is more localized.

It is the minor constituents within a cloud, rather than the hydrogen itself, that emit energy into space. Hydrogen collides with O I and C II, exciting fine-structure levels that decay through far-infrared lines. The most important coolant in molecular clouds is CO. Photons from the lowest-lying rotational transitions become trapped inside the cloud, while those from higher levels escape from its surface. Finally, gas transfers energy by collisions to dust grains. These, in turn, radiate into the infrared continuum.

---

[3] The average kinetic energy a molecule imparts to the grain surface is actually $2\,k_B\,T_g$, since faster molecules, while rarer, hit more often. We ignore this correction, as well as the finite probability that molecules bounce instead of sticking to the grain.

Suggested Reading

Section 7.1 The phenomenology of cosmic rays, together with an historical account of early research, may be found in

Friedlander, M. W. 1989, *Cosmic Rays* (Cambridge: Harvard U. Press).

A broader and somewhat more technical book is

Schlickeiser, R. 2000, *Cosmic Ray Astrophysics* (Berlin: Springer-Verlag).

Our discussion of the heat deposition in molecular clouds has utilized the theoretical studies of

Cravens, T. E. & Dalgarno, A. 1978, ApJ, 219, 750

Van Dishoeck, E. F. & Dalgarno, A. 1984, ApJ, 277, 576,

where the latter obtains the lifetime of OH against photodissociation.

Section 7.2 Cloud heating by the diffuse radiation field is covered in the useful review of

Black, J. H. 1987, in *Interstellar Processes*, ed. D. J. Hollenbach & H. A. Thronson (Dordrecht: Reidel), p. 731.

The physics of the grain photoelectric effect has been summarized by

Hollenbach, D. J. 1990, in *The Evolution of the Interstellar Medium*, ed. L. Blitz (ASP Conf. Ser. Vol. 12), p. 167.

The contribution to this process from PAHs is in

Bakes, E. L. O. & Tielens, A. G. G. M. 1994, ApJ, 427, 822.

For cloud heating by stellar X-rays, see

Krolik, J. H. & Kallman, T. R. 1983, ApJ, 267, 810,

although our treatment is quite different.

Section 7.3 The L-S coupling underlying atomic fine-structure transitions is explained in

Messiah, A. 1975, *Quantum Mechanics, Vol. II* (Amsterdam: North Holland), Chapter 16.

Observations of these lines in molecular clouds have been reviewed by

Melnick, G. J. 1990, in *Molecular Astrophysics*, ed. T. W. Hartquist (Cambridge: Cambridge U. Press), p. 273.

Section 7.4 Our presentation of CO cooling is largely based on the detailed analysis by

Goldreich, P. & Kwan, J. 1974, ApJ, 189, 441,

and gives similar numerical results to the later work of

Neufeld, D. A., Lepp, S., & Melnick, G. J. 1995, ApJSS, 100, 132.

This paper also includes the cooling rates for other abundant molecules.

The Formation of Stars
Steven W. Stahler and Francesco Palla

8 Cloud Thermal Structure

The physical processes we have just detailed enable us to reexamine, in a more quantitative manner, the structure of molecular clouds. In this chapter, we focus on the variation of gas and dust temperatures, and on the dissociation and ionization of the constituent molecules and atoms. We consider both quiescent clouds and those subjected to the strong ultraviolet radiation fields and shocks generated by nearby, massive stars. In the quiescent case, we analyze in some detail the process of self-shielding, which governs the transition from atomic to molecular hydrogen. We also investigate the residual ionization within molecular gas, as this feature controls the dynamical influence of magnetic fields. These fields, we show, strongly modify the shocks that arise from stellar winds. Note that our treatment of clouds themselves is based largely on the application of thermal and dissociative equilibrium. Thus, we can say little at this point concerning *mechanical* issues, such as the internal distribution of density. To make progress on that front, we must also invoke force balance, as we shall do in Chapter 9.

8.1 The Buildup of Molecules

Figure 8.1 depicts, in a schematic fashion, the generic molecular cloud that will be our object of study. With reference to Table 3.1, the entity shown could represent an individual dark cloud or a massive clump within a giant complex. The outermost region is the *atomic envelope*, in which ultraviolet photons from the interstellar radiation field or nearby massive stars promptly dissociate any molecular hydrogen that forms. As we discussed in Chapter 3, 21 cm observations of HI have established such a tenuous component around at least some giant complexes (recall Figure 3.3), but have yet to map the analogous structure on smaller scales. Some clumps that are especially embedded may lack this outer layer. In any case, most of the cloud mass is taken up by the *molecular interior*. This is the region principally observed through various CO isotopes. Inside it we find dense cores, a few of which are sketched here. Some of these cores, in turn, contain young stars. Our present goal is *not* to explain this hierarchical structuring, but to explore the interior run of temperature and chemical composition as one passes through regions of successively higher density.

8.1.1 The Atomic Envelope

Let us begin our analysis with the outermost region. In the absence of X-rays from nearby young stars, the gaseous component here is heated mainly by cosmic rays, at a volumetric rate given by equation (7.12), and by the grain photoelectric effect (equation (7.18)). The additional heating mechanism of carbon ionization (equation (7.16)) is unavailable, since that element is already fully ionized throughout the envelope. The gas cools through emission of infrared photons from

The Formation of Stars. Steven W. Stahler and Francesco Palla

ISBN: 3-527-40559-3

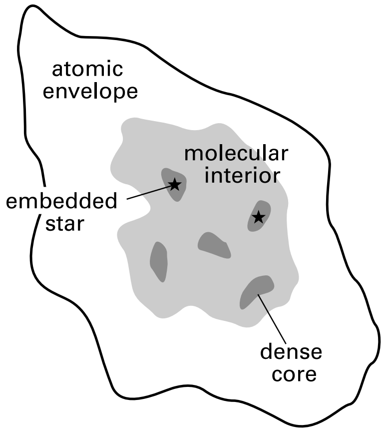


Figure 8.1 Sketch of a molecular cloud, including dense cores with and without interior stars.

the fine-structure transitions of atoms and ions. The most effective coolants are the 63 µm line of O I (equation (7.26)) and the 158 µm line of C II (equation (7.27)), both of which are optically thin at typical envelope densities of $n_H \lesssim 10^3\ \mathrm{cm}^{-3}$.

Comparing the magnitudes of the various rates, we find that Λ_{CII} exceeds Λ_{OI} at subcritical densities, so that the requirement of thermal balance is effectively

$$\Gamma_{\mathrm{PE}} = \Lambda_{\mathrm{CII}}\ . \tag{8.1}$$

The photoelectric heating rate varies linearly with the density n_H, while the fine-structure cooling rate is proportional to n_H^2. Thus, equation (8.1) yields immediately the desired temperature-density relation:

$$T_g = \frac{40\,\mathrm{K}}{2.0 + \log\left(n_H/10^3\,\mathrm{cm}^{-3}\right)}\ . \tag{8.2}$$

Figure 8.2 plots this relation as the upper solid curve. The steady decline in T_g at higher density stems from the increasing efficiency of fine-structure cooling.

From the last equation, we may derive a representative density and temperature at the cloud's outer boundary. We discussed in Chapter 2 how the various components of the interstellar medium, including molecular clouds, all appear to be roughly in pressure balance. In a self-gravitating molecular cloud, this ambient pressure applies at the edge, but climbs higher within the deep interior. To determine conditions at the boundary, we use equation (8.2) as one relation between n_H and T_g. For a second relation, we equate the product $n_H T_g$ with the empirical figure of 3000 cm^{-3} K. In this way, we obtain a density of 56 cm^{-3} and a gas temperature of 54 K. The open circle in Figure 8.2 indicates this pair of boundary values.

The gas in the envelope is too rarefied to be thermally coupled to the dust, so we must independently determine the temperature of the latter. The visual extinction through the envelope is relatively small. Thus, optical photons can heat the dust at the rate Γ_d given in equation (7.20). Dust cooling occurs through the emission of far-infrared radiation at the rate Λ_d (equation (7.39)). Since both Γ_d and Λ_d vary linearly with n_H, equating the two yields a unique temperature, here equal to 16 K. The dashed curve in Figure 8.2 shows this uniform distribution of T_d.

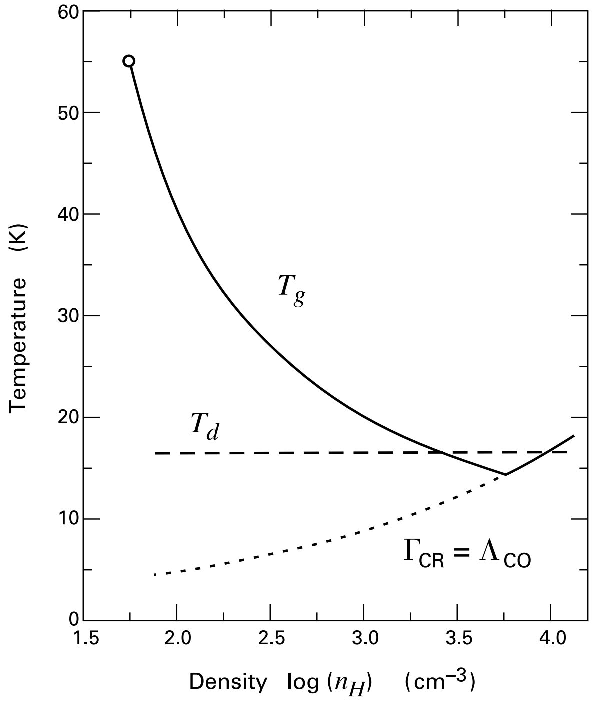


Figure 8.2 Temperature profiles in the lower-density region of a molecular cloud. The solid curve represents the gas temperature, while the dashed line refers to dust grains. The lower, dotted curve results from balancing cosmic-ray heating and CO cooling in the gas.

8.1.2 Destruction and Formation of H_2

The transition from the atomic envelope to the molecular interior occurs through the process of self-shielding, *i.e.*, the progressive loss through H_2–absorption of those interstellar photons capable of dissociating the molecule (see Figure 8.3). To gain a more quantitative understanding of the buildup in H_2, let us consider the key elements in the theory. The basic idea is to equate, at each location, the rates of dissociation and recombination of H_2, thereby obtaining the spatial variation of the molecular fraction $f_{\rm H_2} \equiv 2\, n_{\rm H_2}/n_H$.

We shall denote by $\sigma_i(\nu)$ the cross section for a photon of frequency ν to promote H_2 into one of the levels – each labeled by the index i – within the first or second excited electronic states (recall Figure 5.5). Note that our generic index i really stands for some combination (v, J) of vibrational and rotational quantum numbers. We further denote by β_i the probability that the subsequent decay from such an excited level actually leads to dissociation of the molecule. Summing over all available levels, we find the volumetric dissociation rate to be

$$\mathcal{D}_{\rm H_2} = 4\pi\, n_{\rm H_2} \sum_i \beta_i \int \frac{J_\nu}{h\nu} \sigma_i(\nu)\, d\nu \ . \tag{8.3}$$

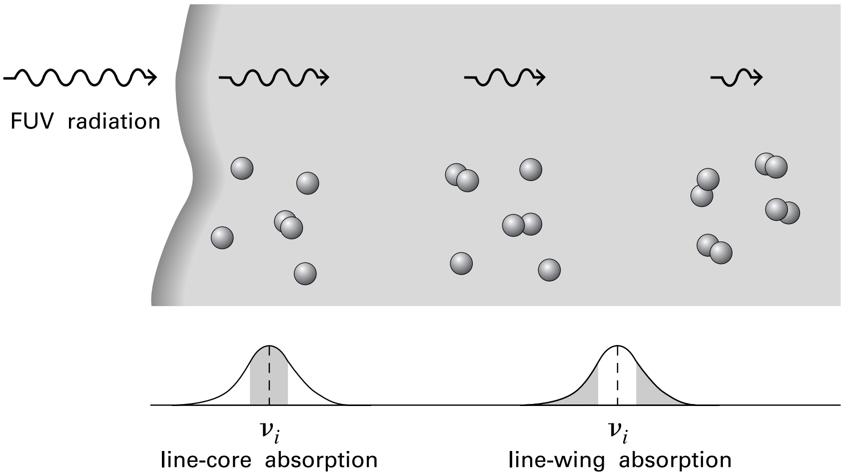


Figure 8.3 Photodissociation of H_2 at the surface of a cloud. The initial absorption of far-ultraviolet photons is at frequencies close to line center for each rovibrational transition. Absorption in the line wings occurs further inside, where the radiation intensity is lower and most atoms are recombined.

Here, J_ν is the average specific intensity of the local radiation field. If we first use the *interstellar* field, then equation (8.3) gives us $t_{\rm diss}$, the characteristic H_2 dissociation time in unshielded regions. This time may be found from

$$t_{\rm diss}^{-1} = 4\pi \sum_i \frac{\beta_i \, J_{\nu_i}}{h\nu_i} \int \sigma_i(\nu) \, d\nu \ , \tag{8.4}$$

and has a numerical value of 400 yr. Here, ν_i denotes the line-center frequency of each electronic transition. In going from equation (8.3) to Equation (8.4), we have used the fact that the most rapid frequency variation is in each $\sigma_i(\nu)$. The integral of $\sigma_i(\nu)$ is proportional, through atomic constants, to the *oscillator strength* of the transition, a quantity that has been calculated quantum mechanically for all the relevant lines. Note that one may determine the oscillator strengths, and hence $t_{\rm diss}$, without knowledge of the line-broadening mechanism, *i. e.*, the precise functional form of $\sigma_i(\nu)$. For the transitions of interest, ν_i is typically 3×10^{-15} s^{-1} and the integrated cross section is about 1×10^{-4} cm^2 s^{-1}.

We next equate $\mathcal{D}_{\rm H_2}$ with $\mathcal{R}_{\rm H_2}$, the recombination rate given in equation (5.10). Applying this procedure first at the envelope boundary, we set $V_{\rm therm}$ in that equation equal to $(3\mathcal{R}T_g/2\mu)^{1/2}$, where $T_g = 54$ K and $\mu = 1.3$. The latter value is appropriate for an HI region. After further equating $n_d \sigma_d$ to $n_H \Sigma_d$, we find the boundary value of the molecular hydrogen fraction:

$$\begin{aligned} f_{\rm H_2} &= \gamma_H \, \Sigma_d \, n_{\rm HI} \, V_{\rm therm} \, t_{\rm diss} \\ &= 3 \times 10^{-5} \ . \end{aligned} \tag{8.5}$$

Here we have set the HI boundary density to 56 cm^{-3}. Thus, the actual molecular hydrogen density, n_{H_2}, has the very low value of 9×10^{-4} cm^{-4}.

8.1.3 Photon Penetration

Proceeding inward, we must modify $\mathcal{D}_{\mathrm{H}_2}$ in equation (8.3) to account for the absorption of ultraviolet photons. Dust provides some of this absorption, as does the electronic excitation of H_2. Thus, a hydrogen column density N_H, with a molecular component N_{H_2}, corresponds to an optical depth of $N_H \Sigma_d + N_{\mathrm{H}_2} \sigma_i(\nu)$. Here, we are neglecting dust scattering and using the fact that a grain's cross section for absorbing the far-ultraviolet photons that dissociate H_2 is essentially its geometrical area. (The latter statement fails for the very smallest grains and PAHs.) A proper determination of the new average radiation intensity at each radius would entail a numerical calculation in spherical geometry. We adopt instead the simpler prescription, valid in a plane-parallel slab, of reducing the boundary intensity at each frequency by $\exp[-N_H \Sigma_d - N_{\mathrm{H}_2} \sigma_i(\nu)]$, where both column densities are measured radially inward from the cloud boundary. The destruction rate now becomes

$$\mathcal{D}_{\mathrm{H}_2} = 4\pi\, n_{\mathrm{H}_2} \exp\left[-N_H \Sigma_d\right] \sum_i \frac{\beta_i\, J_{\nu_i}}{h\nu_i} \int \sigma_i(\nu) \exp\left[-N_{\mathrm{H}_2} \sigma_i(\nu)\right]\, d\nu\ . \tag{8.6}$$

Evaluation of the integral in equation (8.6) requires that we specify the form of $\sigma_i(\nu)$. We do so through use of the normalized line profile function $\phi(\nu - \nu_i)$, introduced in Appendix B. The cross section $\sigma_i(\nu)$ is proportional to the profile function. Here the constant of proportionality is the oscillator strength, as we see by applying the normalization condition of equation (B.2):

$$\sigma_i(\nu) = \phi(\nu - \nu_i) \int \sigma_i(\nu')\, d\nu'\ . \tag{8.7}$$

We may now rewrite equation (8.6) in the form

$$\mathcal{D}_{\mathrm{H}_2} = 4\pi\, n_{\mathrm{H}_2} \exp\left[-N_H \Sigma_d\right] \sum_i \frac{\beta_i\, J_{\nu_i}\, \delta_i}{h\nu_i} \int \sigma_i(\nu)\, d\nu\ . \tag{8.8}$$

Here, δ_i is the *penetration probability* for a line photon to reach the column density N_{H_2} in the presence of self-shielding:

$$\delta_i \equiv \int d\nu\, \phi(\nu - \nu_i) \exp\left[-N_{\mathrm{H}_2} \phi(\nu - \nu_i) \int \sigma_i(\nu') d\nu'\right]\ . \tag{8.9}$$

Note that δ_i for a given line, being an integral over all frequencies, depends only on N_{H_2} and decreases as the column density rises. Note also that, apart from a change in notation, equation (8.9) is identical to equation (6.23) for the photon *escape* probability β in a spherical medium.

Suppose that the lines are broadened by Gaussian turbulent motion (Appendix E). Then all photons with frequencies within a Doppler width $\Delta\nu_D$ of ν_i are quickly absorbed through electronic excitation. At the shallow depths where this absorption occurs, the molecule fraction is still small. The large cross section, however, compensates to give a rapid falloff in the incident flux. Referring to equation (8.9), the peaking of the profile function in the *line core* greatly diminishes the exponential factor, even for modest N_{H_2}. For a velocity dispersion ΔV of 1 km s^{-1}, $\Delta\nu_D = \nu_i\,\Delta V/c = 9 \times 10^9\ \mathrm{s}^{-1}$. Now the fraction of photons at frequency ν absorbed in column depth N_{H_2} is $N_{\mathrm{H}_2}\sigma_i(\nu)$, or $N_{\mathrm{H}_2}\int\sigma_i(\nu)d\nu/\Delta\nu_D$ averaged over the line core. Conversely, the initial, line-core absorption is essentially complete within a column density $\Delta N_{\mathrm{H}_2} = \Delta\nu_D/\int\sigma_i(\nu)d\nu = 9\times 10^{13}\ \mathrm{cm}^{-2}$. Even with our low boundary value for n_{H_2} of $9\times 10^{-4}\ \mathrm{cm}^{-3}$, this column is reached in the relatively short distance of 1×10^{17} cm.

At greater depths, the only penetrating photons are those in the *line wings*, with a frequency differing from ν_i by more than $\Delta\nu_D$. Referring again to equation (8.9), the relatively small value of the profile function in the exponential means that the penetration probability now diminishes slowly with column density. To follow this decline, we use the result from Appendix E that the profile in the line wings is

$$\begin{aligned}\phi(\nu-\nu_i) &= \frac{\gamma_i/4\pi^2}{(\nu-\nu_i)^2+(\gamma_i/4\pi^2)^2}\\ &\approx \frac{\gamma_i}{4\pi^2(\nu-\nu_i)^2}\ .\end{aligned} \tag{8.10}$$

Here the damping constant γ_i equals the Einstein A-value for the transition. The second, approximate form of (8.10) assumes $|\nu-\nu_i| \gg \gamma_i$. Substituting this profile into equation (8.9) and performing the resulting integration, we find

$$\delta_i = \left(\frac{N_i}{N_{\mathrm{H}_2}}\right)^{1/2}\ . \tag{8.11}$$

The column density N_i sets the scale for δ_i, and is given by

$$N_i \equiv \frac{\gamma_i}{4\pi\int\sigma_i(\nu)\,d\nu}\ . \tag{8.12}$$

For a typical γ_i of $2\times 10^9\ \mathrm{s}^{-1}$, we find that N_i is $1\times 10^{12}\ \mathrm{cm}^{-2}$. We now have a more complete picture of the aborption process. Once the photons have traversed the initial column density ΔN_{H_2}, equation (8.11) indeed shows that their subsequent penetration probability decreases gradually with depth, as $N_{\mathrm{H}_2}^{-1/2}$. This behavior, in turn, stems from the gentle decline with frequency of the absorption cross section in the line wings.

8.1.4 Appearance of H_2 and CO

Most of the buildup in H_2 actually occurs within this interior region, *i. e.*, at a column density above ΔN_{H_2}. Accordingly, we may follow the transition explicitly by employing equation (8.11) for δ_i in equation (8.8). After equating $\mathcal{D}_{\mathrm{H}_2}$ and $\mathcal{R}_{\mathrm{H}_2}$, equation (8.4) allows us to write

$$\begin{aligned}\frac{n_{\mathrm{H}_2}}{t_{\mathrm{diss}}}\frac{\langle N_i\rangle^{1/2}}{N_{\mathrm{H}_2}^{1/2}}\exp\left[-N_H\Sigma_d\right] &= \frac{1}{2}\,\gamma_H\,\Sigma_d\,n_H\,n_{\mathrm{HI}}\,V_{\mathrm{therm}} \\ &= \frac{1}{2}\,\gamma_H\,\Sigma_d\,n_H\,(n_H-2n_{\mathrm{H}_2})\,V_{\mathrm{therm}}\;.\end{aligned} \tag{8.13}$$

The angular brackets on the lefthand side denote an average over the various lines.

Our purpose, again, is to obtain the distribution of the molecular fraction f_{H_2}. Both N_H and N_{H_2} are integrals of volume density over the depth $\Delta r \equiv R - r$. Here, R and r are the radial positions of the cloud surface and the point of interest, respectively. It is clear that we need the functional form of $n_H(\Delta r)$ before we can solve equation (8.13) for f_{H_2}. For simplicity, let us consider the case of a spatially uniform n_H. Then we may recast equation (8.13) into

$$N_{\mathrm{H}_2} \;=\; \int_0^{\Delta r} n_{\mathrm{H}_2}\,d(\Delta r) \;=\; \frac{4\,\langle N_i\rangle\exp(-2n_H\Sigma_d\Delta r)}{\left[\gamma_H\,t_{\mathrm{diss}}\,\Sigma_d\,V_{\mathrm{therm}}\,(n_H/n_{\mathrm{H}_2}-2)\right]^2}\;. \tag{8.14}$$

We introduce a nondimensional parameter $\eta_\circ$, defined as

$$\eta_\circ \;\equiv\; \frac{4\,\langle N_i\rangle}{\gamma_H^2\,t_{\mathrm{diss}}^2\,\Sigma_d\,V_{\mathrm{therm}}^2\,n_H^2}\;, \tag{8.15}$$

and a dust "optical" depth at far-ultraviolet frequencies: $\tau \equiv n_H\Delta r\Sigma_d = N_H\Sigma_d$. If we further change the dependent variable to $h \equiv (1-f_{\mathrm{H}_2})^{-1}$, then equation (8.14) becomes

$$\int_0^\tau \left(\frac{h\,-\,1}{h}\right)d\tau \;=\; \frac{\eta_\circ}{2}\,(h\,-\,1)^2\exp(-2\tau)\;.$$

Differentiation of this equation with respect to τ and rearrangement finally yields

$$\frac{dh}{d\tau} \;=\; \frac{1}{\eta_\circ h}\exp(2\tau)\,+\,h\,-\,1\;. \tag{8.16}$$

After solving (8.16) numerically, we may recover the fraction f_{H_2}. Figure 8.4 shows the spatial variation of this fraction. Here we have set n_H to 100 cm^{-3} and the temperature (needed to compute V_{therm}) to 30 K. We see how dust attenuation of ultraviolet photons ensures that f_{H_2} reaches unity by a depth of $\Delta r \gtrsim 2\,(n_H\Sigma_d)^{-1}$, *i. e.*, by a visual extinction A_V from the surface of about 2. This result continues to hold in a more detailed treatment of the problem.

The change from atomic to molecular hydrogen occurs in step with another important transition, the buildup of CO. Like H_2, CO is dissociated by ultraviolet photons through electronic excitation. This molecule, however, does *not* form out of neutral atomic components, via grain-surface catalysis. Instead, it builds up through ion-molecule reactions in the gas phase. The essential ingredient in these reactions is the ionized species C II, which quickly reacts with either atomic or molecular hydrogen. For example, one major pathway to CO production in the surface region is

$$\begin{aligned}\mathrm{C}^+ + \mathrm{H}_2 &\rightarrow \mathrm{CH}_2^+ \\ \mathrm{CH}_2^+ + \mathrm{e}^- &\rightarrow \mathrm{CH} + \mathrm{H} \\ \mathrm{CH} + \mathrm{O} &\rightarrow \mathrm{CO} + \mathrm{H}\;.\end{aligned} \tag{8.17}$$

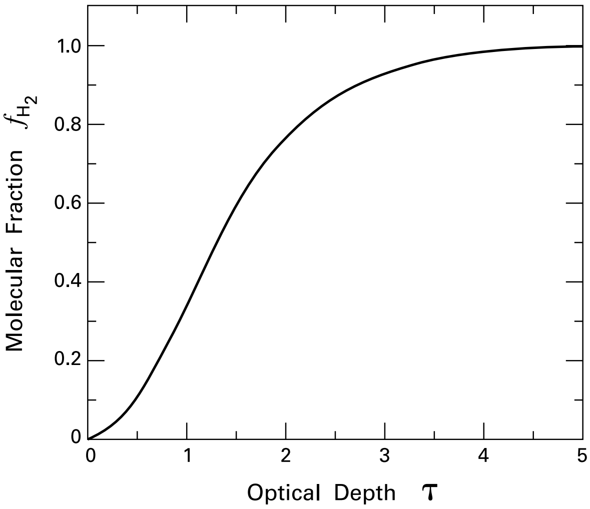


Figure 8.4 Buildup of H_2 inside the cloud surface. Plotted is f_{H_2}, the fraction of hydrogen atoms bound into molecules, as a function of the dust optical depth at far-ultraviolet frequencies.

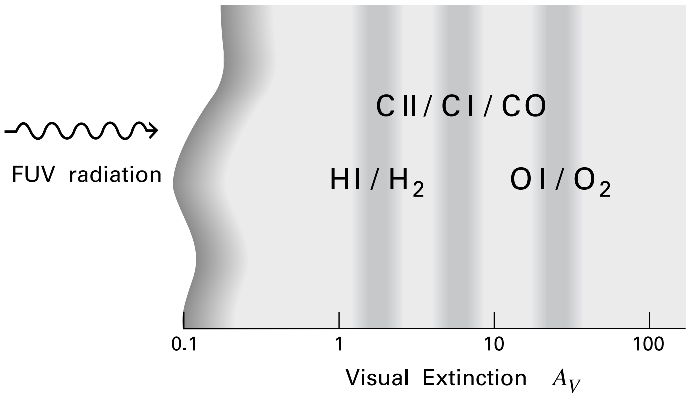


Figure 8.5 Chemical transformations in molecular gas exposed to interstellar far-ultraviolet radiation. The changes are indicated as a function of the visual extinction A_V, measured inward from the surface. Extinction values are appropriate for Orion-like conditions, *i. e.*, for an ultraviolet enhancement factor $G_\circ \sim 10^5$.

It happens that carbon's first ionization potential of 11.2 eV closely matches the excitation energy of the Lyman band of H_2. Throughout the envelope, therefore, photons of higher energy are depleted by both elements, and H_2 and C I concurrently grow in abundance. It is important to remember, however, that H_2 is dissociated through absorption in discrete spectral *lines*, while any *continuum* photon with frequency above 11.2 eV is capable of ionizing C I. Thus, even when hydrogen has nearly all recombined and the intensity at each line center ν_i is essentially zero, there are still enough energetic photons at intermediate frequencies to maintain an appreciable abundance of C II. The ion-molecule reactions continually use this C II to form CO, sometimes by first producing C I. Once formed, CO then shields itself in a manner similar to H_2. Essentially all of the gas-phase carbon is thus transformed, with increasing depth, from C II to CO, with an intermediate layer of C I generally present, as well. The remaining free oxygen persists in neutral, atomic form (O I) even deeper inside. In this case, the relatively small O_2 dissociation energy of 5.1 eV renders the molecule susceptible to even a small remnant of the incident ultraviolet flux. Figure 8.5 sketches the major chemical transformations in the cloud, here one bathed in a relatively strong radiation field. The transitions are shown as functions of A_V, measured from the surface.

8.2 The Molecular Interior

As we move inside the atomic envelope, we enter a region that is at least partially shielded from the external radiation field. The ambient temperature accordingly falls, as we noted observationally in the case of B335 (Figure 3.21). Despite this drop, and the concurrent rise in density, there is still some nonzero level of ionization. We need to gauge this level carefully, as it determines the influence of any internal magnetic field.

8.2.1 Temperature Profile

Consider first the run of temperature in the molecular interior. Since the grain photoelectric effect relies on ultraviolet photons, it can no longer heat the gas within the molecular interior. We are left with cosmic rays as the principal heating agent. Almost all the carbon is in CO, which now becomes the major coolant. The gas temperature in this region therefore follows from

$$\Gamma_{\rm CR}({\rm H}_2) = \Lambda_{\rm CO} \ , \tag{8.18}$$

where equation (7.14) gives the cosmic-ray contribution. The rising curve in Figure 8.2 is the temperature-density relation implied by equation (8.18). This behavior of $T_g(n_H)$ stems from the fact that $\Lambda_{\rm CO}$ is rather insensitive to density in this regime. Thus, the rise in $\Gamma_{\rm CR}$ with density can only be offest by an increase in temperature.[1] Considering the decline of T_g in the envelope, it is clear that the gas temperature must reach a minimum at some point. This location conveniently demarcates the boundary with the molecular interior. For simplicity, we have shown the minimum lying at the intersection of the two curves, but its true position, as well as the local temperature profile, depend on the detailed falloff of the ultraviolet flux with increasing depth.

The visual extinction A_V as we enter the molecular interior is of order unity, so that the grains are still heated mostly by optical interstellar photons. Since the dust cools by infrared emission, its temperature retains the envelope value of 16 K in our representative model. According to the rough profile in Figure 8.2, the gas temperature dips temporarily below this value

---

[1] Here we have used, for the volumetric CO cooling rate, $\Lambda^*_{\rm CO}$ from equation (7.35). Additional flux from optically thick CO lines can raise the total rate by as much as a factor of 5 for typical cloud densities and sizes. The rising temperature profile shown in Figure 8.2 would consequently be lowered by about 30 percent.

at the envelope-interior boundary before resuming its climb. For values of n_H in excess of about 10^4 cm^{-3}, we enter the regime of dense cores. Here, gas-dust collisions become frequent enough that significant heat transfer occurs between the two components. If the dust grains are relatively cold, this exchange acts to cool the gas, at the volumetric rate $\Lambda_{g\rightarrow d}$ given by equation (7.40).

The condition of thermal equilibrium of the gas now reads

$$\Gamma_{\mathrm{CR}}(\mathrm{H}_2) \,=\, \Lambda_{\mathrm{CO}} \,+\, \Lambda_{g\rightarrow d} \ . \tag{8.19}$$

Because of the coupling term, we cannot solve this equation without simultaneously determining T_d. At these high densities, the typical value of A_V is so large that optical photons no longer penetrate to heat the dust. There is also a rapid decline in the flux of interstellar mid-infrared photons, which could potentially heat the grains because of the spike in opacity at 10 µm. (Recall the discussion in § 2.3.) Photons of even longer wavelength are provided by the outer, heated dust. If we neglect this component of the radiation field, then the interior dust is heated principally by its collisional coupling to the gas. We denote the associated heating rate by the new symbol $\Gamma_{d\rightarrow g}$, although its magnitude is just given by the negative of equation (7.40). As long as the column density N_{H} is not so high that it prevents the escape of even far-infrared radiation, the cooling rate is still Λ_d in equation (7.39). The dust temperature then follows from

$$\Gamma_{d\rightarrow g} \,=\, \Lambda_d \ . \tag{8.20}$$

For a given value of n_H, we may solve Equations (8.19) and (8.20) simultaneously for T_g and T_d. As seen in Figure 8.6, the previous rise in T_g has now stopped, and the gas cools because of increasing thermal contact with the very cold dust. Thus, $\Lambda_{g\rightarrow d}$, which is already twice Λ_{CO} at $n_H \,=\, 10^4$ cm^{-3}, rapidly increases in importance because of its quadratic dependence on density. The dust itself is also considerably colder than in the envelope because of the cutoff of interstellar radiation. Proper accounting of this extinction and the reradiation of incident photons at longer wavelengths would result in a smoother decline in T_d and turnover in T_g from the outer region.

Despite the simplifications we have made, the generic temperature profiles in Figures 8.2 and 8.6 are at least broadly consistent with those derived empirically from continuum and molecular-line studies of individual clouds. Our discussion in Chapter 6 makes it clear that mapping the density and temperature with good spatial resolution is still highly problematic. Hence, while more careful theoretical calculations exist, systematic comparisons of these with observations are still lacking. Note finally that we have extended Figure 8.6 to densities higher than those in typical dense cores to emphasize the point that T_g and T_d must eventually approach a common value. Physically, thermal contact between the two components becomes so strong that they can be considered a single species, whose temperature is determined by

$$\Gamma_{\mathrm{CR}}(\mathrm{H}_2) \,=\, \Lambda_d \ . \tag{8.21}$$

Here we have used the fact that Λ_{CO} becomes insignificant in this limit. Equation (8.21) can be solved to yield the unique temperature of 4 K.

Thus far, our discussion has centered on a relatively isolated cloud. When embedded within other cold molecular gas, such as the interior of a giant complex where there are no nearby,

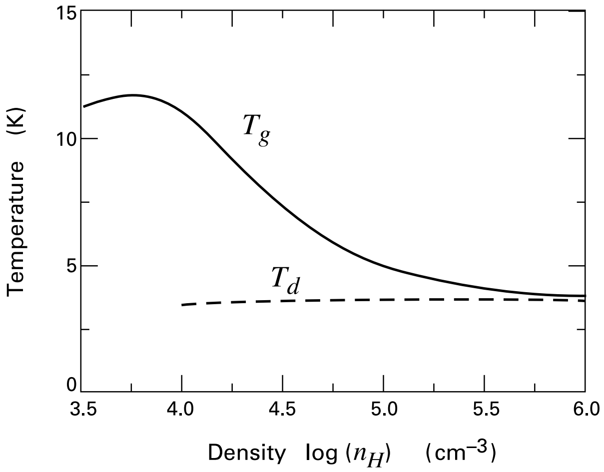


Figure 8.6 Temperature profiles in the higher-density region of a molecular cloud. As in Figure 8.2, the gas (*solid curve*) and dust (*dashed curve*) are shown separately.

massive stars, a cloud receives less ultraviolet flux than the interstellar field assumed here. Thus, both Γ_{PE} and the envelope values of T_g diminish. Moreover, the buildup of H_2 and CO occurs at lower column density. The net result is that the cloud's envelope shrinks relative to its interior. The thermal properties of the dense core, which is shielded from the interstellar radiation, are unchanged.

8.2.2 Measuring the Ionization Level

As we will discuss in Chapter 10, dense cores evolve by the slippage of gas through the ambient magnetic field. This process, in turn, depends critically on the *ionization fraction*, since it is only the charged species that sense the field directly. Let us now examine how this fraction is actually determined empirically. We will then look at the theory that allows us to follow the ionization up to densities much higher than those presently observed.

We focus on the relative density of free electrons: $[\mathrm{e}^-] \equiv n_{\mathrm{e}^-}/n_H$. Since electrons are the dominant negative species, charge neutrality dictates that their density nearly equals the total for positive ions. The idea is to relate $[\mathrm{e}^-]$, which cannot be observed directly, to the concentration of another species that can. One practical choice is HCO^+, a relatively abundant molecule which, along with CO, is detected through its rotational transitions (see Table 5.1).

Figure 8.7 depicts the principal creation and destruction pathways for HCO^+ in dark clouds or dense cores. The process begins with the generation of H_2^+ by cosmic-ray impact of H_2. The ionized molecule then reacts with neutral H_2:

$$\mathrm{H}_2^+ + \mathrm{H}_2 \rightarrow \mathrm{H}_3^+ + \mathrm{H} \ . \tag{8.22}$$

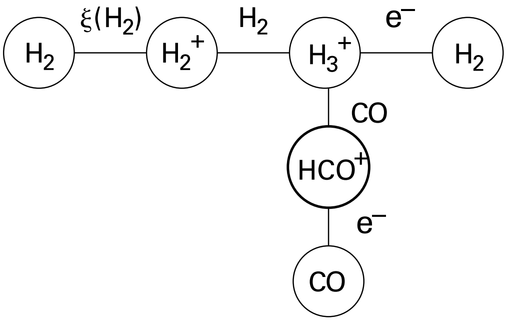


Figure 8.7 Principal reactions creating and destroying HCO$^+$ in dark clouds or dense cores.

We designate the associated rate constant as k_1, with units of cm^3 s^{-1}. The H_3^+ produced usually undergoes dissociative recombination:

$$\mathrm{H}_3^+ + \mathrm{e}^- \rightarrow \mathrm{H}_2 + \mathrm{H} \; . \tag{8.23}$$

Less frequently, H_3^+ reacts with CO to form HCO$^+$:

$$\mathrm{H}_3^+ + \mathrm{CO} \rightarrow \mathrm{HCO}^+ + \mathrm{H}_2 \; , \tag{8.24}$$

where we denote as k_2 the reaction rate.[2]

The HCO$^+$ formed through this reaction sequence is itself destroyed by dissociative recombination:

$$\mathrm{HCO}^+ + \mathrm{e}^- \rightarrow \mathrm{CO} + \mathrm{H} \; . \tag{8.25}$$

Let $k_{\mathrm{dr}}(\mathrm{HCO}^+)$ be the recombination rate. In equilibrium, the number density of HCO$^+$ is

$$[\mathrm{HCO}^+] = \frac{k_2\,[\mathrm{H}_3^+]\,[\mathrm{CO}]}{k_{\mathrm{dr}}(\mathrm{HCO}^+)\,[\mathrm{e}^-]} \; , \tag{8.26}$$

where the square brackets again signify the density relative to n_H. To determine $[\mathrm{H}_3^+]$, we similarly use Equations (8.22) and (8.23), now ignoring equation (8.24) as a quantitatively insignificant depletion channel. We find

$$[\mathrm{H}_3^+] = \frac{k_1\,[\mathrm{H}_2^+]}{k_{\mathrm{dr}}(\mathrm{H}_3^+)\,[\mathrm{e}^-]} \; . \tag{8.27}$$

Here, $k_{\mathrm{dr}}(\mathrm{H}_3^+)$ is the reaction rate corresponding to equation (8.23). Finally, we obtain $[\mathrm{H}_2^+]$ by balancing cosmic-ray ionization, at the volumetric rate $\zeta(\mathrm{H}_2)\,n_{\mathrm{H}_2}$, against destruction by neutral hydrogen molecules:

$$[\mathrm{H}_2^+] = \frac{\zeta(\mathrm{H}_2)}{k_1\,n_H} \; . \tag{8.28}$$

---

[2] The dissociative recombination rate of H_3^+ in equation (8.23) is still uncertain. So too, therefore, is the relative importance of destruction by CO.

We now combine Equations (8.26), (8.27), and (8.28) to obtain the desired expression for $[\mathrm{e}^-]$ in terms of known and observable quantities:

$$[\mathrm{e}^-]^2 = \frac{k_2\,\zeta(\mathrm{H}_2)}{k_{\rm dr}(\mathrm{H}_3^+)\,k_{\rm dr}(\mathrm{HCO}^+)}\,\frac{[\mathrm{CO}]}{[\mathrm{HCO}^+]}\,\frac{1}{n_H}\ . \tag{8.29}$$

For numerical evaluation, we use $k_2 = 2 \times 10^{-9}$ cm^3 s^{-1}, within the range of current experimental values. For a temperature of 10 K, we also have $k_{\rm dr}(\mathrm{H}_3^+) = 4 \times 10^{-6}$ cm^3 s^{-1} and $k_{\rm dr}(\mathrm{HCO}^+) = 3 \times 10^{-6}$ cm^3 s^{-1}. We see, from Table 5.1, that the number density of HCO$^+$ is typically 10^{-4} that of CO. Thus, at a representative dense core n_H of 10^4 cm^{-3}, equation (8.29) indicates that $[\mathrm{e}^-]$ is of order 10^{-7}.

8.2.3 Theoretical Derivation

How are we to understand this number from a theoretical perspective? Species such as HCO$^+$ are important practically, but constitute only a fraction of the cloud's ions. The majority are singly charged atoms, chiefly Na$^+$, Mg$^+$, Ca$^+$, and Fe$^+$. Most heavy elements within a dense core are actually locked up in solid grains, and only a few percent by mass exist in the gas phase. Nevertheless, the density of the gaseous component is sufficiently high that the condition of overall charge neutrality is, to good accuracy,

$$n_{\mathrm{e}^-} = n_{\mathrm{M}^+}\ . \tag{8.30}$$

Here, M$^+$ signifies a metal ion. The right side of this relation is actually a sum over all such species.

Metallic ions are created when neutral atoms undergo charge exchange with *molecular* ions, such as H_3^+ and HCO$^+$. Denoting the molecules generically as m, we may symbolize this process as

$$\mathrm{m}^+ + \mathrm{M} \rightarrow \mathrm{m} + \mathrm{M}^+\ , \tag{8.31}$$

and write the associated reaction rate as $k_{\rm ce}$. A metal ion can be destroyed when it encounters a free electron. However, the rate for such radiative recombination is negligible for the electron fractions of interest (recall § 5.1). Of greater significance are collisions with dust grains. The latter have a small negative charge (equivalent to one or two free electrons), so the sticking probability in these collisions is high. Recalling that the dust grain number density n_d is proportional to n_H, we may write the volumetric collision rate as $k_{\rm dM}\,n_H\,n_{\mathrm{M}^+}$. The coefficient $k_{\rm dM}$, properly averaged, has the value 5×10^{-17} cm^3 s^{-1} near a temperature of 10 K. The metallic ion abundance follows by equating the creation and destruction rates:

$$k_{\rm ce}\,n_{\rm M}\,n_{\mathrm{m}^+} = k_{\rm dM}\,n_H\,n_{\mathrm{M}^+}\ . \tag{8.32}$$

Turning to molecular ions, these form when cosmic rays bombard H$_2$, which then undergoes charge exchange with other species. Their depletion partially stems from charge exchange with neutral metallic atoms. We have also noted that they undergo dissociative recombination with free electrons, as in equation (8.25). Their abundance in steady state therefore obeys

$$\zeta(\mathrm{H}_2)\,n_{\mathrm{H}_2} = k_{\rm ce}\,n_{\mathrm{m}^+}\,n_{\rm M} + k_{\rm dr}\,n_{\mathrm{m}^+}\,n_{\mathrm{e}^-}\ \ , \tag{8.33}$$

Of the two means for depleting molecular ions, dissociative recombination is faster provided that the electron abundance is not too low. We expect $[\mathrm{e}^-]$ to fall with rising n_H, since higher density promotes recombination. The practical condition for ignoring the first righthand term in equation (8.33) is $n_H \lesssim 10^8\,\mathrm{cm}^{-3}$. Within this regime, we may eliminate n_{m^+} between (8.32) and the simplified equation (8.33). After utilizing equation (8.30), we find

$$n_{\mathrm{e}^-} = \left[\frac{\zeta(\mathrm{H}_2)\,k_{\mathrm{ce}}\,n_{\mathrm{M}}}{2\,k_{\mathrm{dM}}\,k_{\mathrm{dr}}}\right]^{1/2} . \qquad (8.34)$$

Here, we have also set n_H equal to $2\,n_{\mathrm{H}_2}$. Since the number density of metals is proportional to that of hydrogen, equation (8.34) predicts that $[\mathrm{e}^-]$ varies as $n_H^{-1/2}$. A more careful treatment, summing over the various metallic and molecular species, gives the same result and supplies the numerical coefficient:

$$[\mathrm{e}^-] = 1\times 10^{-5}\,n_H^{-1/2} . \qquad n_H \lesssim 10^8\,\mathrm{cm}^{-3} \qquad (8.35)$$

This relation is consistent with the values of $[\mathrm{e}^-]$ obtained empirically through the abundance of HCO^+ and other tracers. Equation (8.35) is frequently recast as

$$\rho_i = C\,\rho^{1/2} , \qquad (8.36)$$

where the constant C is 3×10^{-16} $\mathrm{cm}^{-3/2}$ $\mathrm{g}^{1/2}$ at a temperature of 10 K.

The last two equations do not accurately track the ionization level within a collapsing cloud, where the density climbs to very large values. For $n_H \gtrsim 10^8$ cm^{-3}, charge exchange with metals becomes the dominant mode of destroying molecular ions. We again eliminate n_{m^+} between Equations (8.32) and (8.33), and use (8.30) to find

$$n_{\mathrm{e}^-} = \frac{\zeta(\mathrm{H}_2)}{2\,k_{\mathrm{dM}}} . \qquad n_H \gtrsim 10^8\,\mathrm{cm}^{-3} \qquad (8.37)$$

After using equation (7.14) for $\zeta(\mathrm{H}_2)$, we find that n_{e^-} has the constant value of 0.3 cm^{-3}. In this regime, therefore, $[\mathrm{e}^-]$ is proportional to n_H^{-1}.

As n_H reaches even higher values, the number density of electrons and ions eventually falls below that of the charged grains. From equation (2.47), the grain number fraction is

$$\begin{aligned}\frac{n_d}{n_H} &= \frac{\Sigma_d}{\pi\,a_d^2} \\ &= 3\times 10^{-12} ,\end{aligned} \qquad (8.38)$$

where the numerical estimate uses a grain radius a_d of 1×10^{-5} cm. Thus, from equation (8.37), $[\mathrm{e}^-]$ falls below n_d/n_H for n_H larger than 1×10^{11} cm^{-3}. In fact, deviations from equation (8.37) appear a bit earlier, for $n_H \gtrsim 10^{10}$ cm^{-3}. Once the grains themselves become the dominant carriers of both positive and negative charge, their collisions with neutral gas govern the slippage of the magnetic field. As the density continues to rise, cosmic rays become attenuated, and the dominant source of ionization shifts to radioactive elements (chiefly ^{40}K) or to X-rays from the central star. Finally, once temperatures exceed about 10^3 K, energetic collisions among the gas particles supply most of the free electrons. Metallic atoms are again the first species to be ionized. The electron fraction in the streaming gas starts to increase dramatically, and coupling to the local magnetic field becomes much stronger.

8.3 Photodissociation Regions

The most significant modification of cloud thermal structure is from nearby stars. Those born in the center of a dense core radiatively heat a large volume of the surrounding dust. As we will discuss in future chapters, it is this warm dust that has actually provided the best means for identifying embedded young stars. Molecular gas is also heated by the shocks associated with stellar winds.

We have already seen how O and B stars, although rare in number, wreak havoc with vast quantities of molecular gas, to the point of dispersing the giant complexes into which they are born. The effects here are both *mechanical*, through the impact of powerful stellar winds, and *thermal*, through radiative and shock heating of cloud material. In this section, we consider in detail the radiative effect. The theory has wide applicability. For example, infrared emission from hot dust has been the principal tool for locating star formation regions in distant galaxies.

8.3.1 Grain Heating and Emission

Massive stars emit the bulk of their energy in the ultraviolet region of the spectrum. A molecular cloud near such a star receives an ultraviolet flux orders of magnitude greater than from interstellar radiation. Consider, for example, a main-sequence star of spectral type B0 embedded within a giant complex. This star has an effective temperature of 3×10^4 K, equal to the value of $\bar{T}$ characterizing the ultraviolet component of the interstellar radiation field, and a bolometric luminosity of $L_* = 5 \times 10^4 \, L_\odot$. The star emits photons at all energies up to 13.6 eV; those of higher energy are absorbed in the surrounding HII region before reaching cloud gas. If located a distance D from a clump within the complex, the star bathes the cloud face with a flux F_* equal to $L_*/4\pi D^2$. On the other hand, the interstellar radiation contributes a flux $F_{\rm int} = \pi \int_{\rm UV} J_\nu d\nu$. For F_* to equal $F_{\rm int}$, the star must be at a distance

$$\begin{aligned} D &= \left(\frac{L_*}{4\pi \, F_{\rm int}}\right)^{1/2} \\ &= 50 \ {\rm pc} \ . \end{aligned} \tag{8.39}$$

Here, we have estimated the integral of J_ν over the ultraviolet range to be twice that for the far-ultraviolet range of 6 eV and above. Since the latter is known empirically from the Habing flux, we have

$$\pi \int_{\rm UV} J_\nu \, d\nu \ \approx \ 2\pi \int_{\rm FUV} J_\nu \, d\nu \ = \ 8 \times 10^{-4} \, {\rm erg \, cm^{-2} \, s^{-1}} \ .$$

Some clumps within the complex will indeed be at a distance of order 50 pc. On the other hand, we will see in Chapter 15 that molecular gas can persist as close as 0.1 pc from the star in question. This neighboring gas, located just outside the star's HII region, receives a flux enhanced over the interstellar value by the factor $G_\circ$, here equal to $(50/0.1)^2 \ = \ 2 \times 10^5$.

Such an intense radiation field disturbs the physical and chemical equilibrium of the cloud to great depths. Since all the molecules in the outer layers are promptly dissociated, the affected area is known as a *photodissociation region*, a term coined by A. Tielens and D. Hollenbach in

1985. Photodissociation regions are seen not only near massive stars, as in the Orion (M42) and Omega (M17) Nebulae, but also in planetary nebulae and the nuclei of external galaxies.

What is the fate of the ultraviolet radiation? The answer is clear from the observations, which show that photodissociation regions all emit copiously in the infrared. The nearest and best studied example, the Orion Nebula, has a far-infrared luminosity estimated at $3 \times 10^5\, L_\odot$. This prodigious energy output can only stem from dust heated by stars in the region, both the visible members of the Trapezium and others hidden from view by this same dust. The cooling rate of dust grains, given by Λ_d in equation (7.39), rises steeply with the dust temperature T_d. Conversely, T_d responds sluggishly to large increases in the incident radiative flux. For a quantitative assessment, we must first evaluate $\Gamma_d(\mathrm{UV})$, the dust heating rate in the ultraviolet. Employing the enhancement factor $G_\circ$, this rate is given by the appropriate modification of equation (7.19):

$$\begin{aligned}\Gamma_d(\mathrm{UV}) &= 4\pi\, G_\circ\, n_d\, \sigma_d \int_{\mathrm{UV}} Q_\nu\, J_\nu\, d\nu \\ &= 4\pi\, G_\circ\, \Sigma_d\, n_H\, Q_{\nu_{\max}} \int_{\mathrm{UV}} J_\nu\, d\nu \\ &= 2 \times 10^{-9}\, G_\circ \left(\frac{n_H}{10^3\,\mathrm{cm}^{-3}}\right)\ \mathrm{eV\,cm^{-3}\,s^{-1}}\ .\end{aligned} \tag{8.40}$$

In the second form of this equation, we have pulled Q_ν out of the integrand because it varies slowly in the ultraviolet. We used a value of 0.7 for $Q_{\nu_{\max}}$, corresponding to an opacity κ of 500 g cm^{-2} at the peak frequency $\nu_{\max} = 2 \times 10^{15}\ \mathrm{s}^{-1}$ of the interstellar field. Equating $\Gamma_d(\mathrm{UV})$ to Λ_d, we find

$$T_d \;=\; 16\, G_\circ^{1/6}\ \mathrm{K}\ . \tag{8.41}$$

In our example of the B0 star, T_d would be raised to 120 K. A blackbody at this temperature emits at a peak wavelength of 30 μm.

8.3.2 Fine-Structure Cooling

The cloud matter near O and B stars is also observed to radiate in atomic fine-structure lines, principally those of O I and C II. Figure 8.8 shows the 158 μm C II emission from the famous Horsehead Nebula in Orion. The Nebula lies at the edge of a heavily obscured region, about 1° south of the Belt star ζ Ori (see Figure 1.3). The O9 star, and the equally massive σ Ori to the west, illuminate the cloud edge with ultraviolet light, providing an equivalent $G_\circ$ of about 100. In Figure 8.8, the optical image of the Nebula is shown in the negative, while the contours represent the 158 μm intensity distribution. We have grown accustomed to thinking of clouds being delineated by their CO emission, but the luminosity in C II is here far greater.

The strong atomic line emission from photodissociation regions, which itself totals only about 1 percent of the continuum infrared luminosity, stems from hot gas located alongside the dust. The fraction of 1 percent can be neatly explained if we assume that, as in quiescent clouds, the gas is heated primarily through the grain photoelectric effect. As we have seen, ϵ_{PE}, the efficiency with which an ultraviolet photon converts its energy to heating the gas, as opposed to the dust, is indeed about 0.01. The fine-structure luminosity from regions of massive

Figure 8.8 Negative optical photograph of the Horsehead Nebula in Orion. The white contours show emission in the 158 µm fine-structure line of C II.

star formation is so large that it is here, rather than in quiescent clouds, that the transitions were first detected in space. Thus, the 63 µm emission of O I was found in M42 and M17, with luminosities in the line of 600 and 3000 $L_\odot$, respectively.

Previously we argued, using Equations (7.26) and (7.27), that the cooling from C II should dominate that from O I in cloud envelopes. When these lines are seen toward dense photodissociation regions, O I is generally stronger. Let us see how this reversal comes about. Equations (7.26) and (7.27) were derived assuming n_H, the density of colliding hydrogen atoms, to be below the critical values for both transitions. This assumption is safe enough in the envelopes of quiescent molecular clouds, but not in the gas surrounding massive stars, where shock compression can lead to much higher densities. Consider, then, the fine-structure emission from cloud material with arbitrary n_H, bathed in an increasing ultraviolet flux. The two fiducial densities to bear in mind are 3×10^3 cm^{-3}, which is $n_{\rm crit}$ for the C II line, and 5×10^5 cm^{-3}, the corresponding value for the 63 µm line of O I at a typical gas temperature of 300 K, where $\gamma_{\rm ul} = 2 \times 10^{-10}\,{\rm cm^{-3}\,s^{-1}}$.

For a cloud density below 3×10^3 cm^{-3}, both $\Lambda_{\rm OI}$ and $\Lambda_{\rm CII}$ are still given by equation (7.24), and their ratio is

$$\frac{\Lambda_{\rm OI}}{\Lambda_{\rm CII}} = 0.07 \exp\left(-\frac{138}{T_g}\right) , \tag{8.42}$$

which is indeed less than unity. The emission in the dominant C II line relative to that from the dust is

$$\frac{\Lambda_{\rm CII}}{\Lambda_d} = \frac{\Gamma_{\rm PE}}{\Gamma_d({\rm UV})} = 0.01 \ , \tag{8.43}$$

where we have scaled up $\Gamma_{\rm PE}$ in equation (7.18) by the factor $G_\circ$ and have used equation (8.40) for $\Gamma_d({\rm UV})$. The gas temperature follows by equating $\Gamma_{\rm PE}$ to $\Lambda_{\rm CII}$. The appropriate modification to equation (8.2) for quiescent envelopes is then

$$T_g = \frac{40\ {\rm K}}{2.0 + \log\left(n_H/10^3\ {\rm cm^{-3}}\right) - \log G_\circ} \ . \tag{8.44}$$

Notice that this last equation, if evaluated at an n_H of 10^3 cm^{-3}, formally yields an *infinite* T_g for $G_\circ$ greater than the modest value $10^{2.0} = 100$. The true situation is that the photoelectric heating rate is self-limiting. As $G_\circ$ climbs above unity, T_g at the cloud edge does rise at first and can reach several hundred K, as predicted by equation (8.44). However, if electrons are ejected from grain surfaces too rapidly, the resulting buildup of positive charge creates a strong electrostatic force. This attractive force inhibits further ejection, effectively lowering $\epsilon_{\rm PE}$ until thermal balance can be once more achieved. From this point on, further increase of $G_\circ$ actually *lowers* T_g at the cloud edge.

For densities between 3×10^3 cm^{-3} and 5×10^5 cm^{-3}, $\Lambda_{\rm OI}$ remains unchanged, but $\Lambda_{\rm CII}$ must now be found from equation (7.25), which effectively reduces the rate by $(3 \times 10^3\ {\rm cm^{-3}}/n_H)$. The ratio of $\Lambda_{\rm OI}$ to $\Lambda_{\rm CII}$ is now

$$\frac{\Lambda_{\rm OI}}{\Lambda_{\rm CII}} = 2 \times 10^{-2} \left(\frac{n_H}{10^3\ {\rm cm^{-3}}}\right) \exp\left(-\frac{138}{T_g}\right) \ , \tag{8.45}$$

so that $\Lambda_{\rm CII}$ still dominates at lower densities. However, the two cooling rates become comparable at the highest n_H because of the quadratic increase of $\Lambda_{\rm OI}$ with density. It should be remembered here that we are only discussing *local* rates. The emission from the entire cloud favors O I even more, since neutral oxygen persists to greater depths. Regardless of which line dominates, our derivation of equation (8.43) makes it clear that the ratio of fine-structure to dust cooling remains about 0.01, as long as we can ignore the lowering of $\epsilon_{\rm PE}$.

Finally, for cloud densities above 5×10^5 cm^{-3}, both O I and C II are in the supercritical regime and we find

$$\frac{\Lambda_{\rm OI}}{\Lambda_{\rm CII}} = 11 \exp\left(-\frac{138}{T_g}\right) \ . \tag{8.46}$$

Since all the level populations have now reached LTE, it is the product $A_{\rm ul}\Delta E_{\rm ul}$, the emission rate per atom from the upper level, which now governs both cooling rates. This product, weighted by the relative chemical abundances and degeneracy factors, is higher for the O I transition. Although clumps with $n_H \gtrsim 5 \times 10^5$ cm^{-3} may indeed exist around massive stars, the radiation in both lines can be optically thick in such environments. The relevant cooling rate for both species is then a surface flux, given by the appropriate modification of equation (7.33). The ratio of the two fluxes is

$$\frac{F_{\rm OI}}{F_{\rm CII}} = \left(\frac{230}{92}\right)^4 \frac{\exp\left(92/T_g\right) - 1}{\exp\left(230/T_g\right) - 1} \ , \tag{8.47}$$

so that the O I line dominates for any T_g greater than 40 K.

The basic chemical transformations within photodissociation regions are just those found in quiescent cloud envelopes. Indeed, the various recombinations also occur at similar A_V-values, since the greater $G_\circ$ tends to compensate for the higher ambient density. In the surface region, the ultraviolet flux not only dissociates H_2, but also ionizes those atoms, such as carbon, with ionization potentials below 13.6 eV. Carbon still becomes CO within the molecular interior, where it again acts as a principal coolant of the gas. The other important thermal processes also remain the same, except in clouds of the very highest densities and $G_\circ$-values, in which infrared radiation from hot dust can excite the fine-structure lines and thus warm the gas. The temperature structure of both gas and dust is thus qualitatively similar to the quiescent case (Figure 8.2), with both T_g and T_d scaling upward as $G_\circ$ increases, and with T_g always exceeding T_d by a wide margin at the cloud edge. One significant difference is that T_g always climbs initially before turning over and undergoing its characteristic slow decline within the envelope. This temporary increase stems from the rise in $\epsilon_{\rm PE}$ accompanying the attenuation of the ultraviolet flux. For $G_\circ$-values near 10^5, the peak gas temperature can exceed 10^3 K. Deeper in the cloud, increasing thermal contact between the gas and dust again drives them toward a common temperature that is relatively insensitive to the external ultraviolet flux.

8.3.3 Heated H_2

Let us finally consider observations of molecular hydrogen. We earlier noted that photodissociation through excitation of the Lyman and Werner bands is an inefficient process, as the excited molecule usually relaxes intact to its ground state. The fluorescent emission accompanying such relaxation is another important signature of photodissociation regions. Radiation from the lower rovibrational transitions occurs at near-infrared wavelengths accessible to ground-based telescopes. If the ambient density n_H is subcritical with respect to these transitions ($n_{\rm crit} \sim 10^6$ cm^{-3}), the branching ratios during fluorescent cascade depend only on the transition rates for radiative decay, *i. e.*, on internal molecular constants. Thus, the *relative* intensities of the various lines are also determined, although the *absolute* intensities still vary with $G_\circ$ and n_H. Referring back to Figure 5.4, consider the $1-0\ S(1)$ and $2-1\ S(1)$ lines, which are transitions between the same rotational sublevels for $v = 1 \rightarrow 0$ and $v = 2 \rightarrow 1$, respectively. Since the A-values of the two transitions are nearly identical, it is not surprising that the theoretical fluorescent intensities are also close, with the $1-0\ S(1)$ line predicted to be 1.8 times as strong.

In fact, the observed relative line intensities often differ considerably from the predicted fluorescent values. To quantify the difference, we recall that the volumetric emission rate associated with any transition from an upper to a lower state is given by

$$\Lambda_{\rm ul} = n_u\, A_{\rm ul}\, \Delta E_{\rm ul}\ . \tag{8.48}$$

If we express the level population n_u using the generalized Boltzmann relation of equation (5.14), then equation (8.48) becomes

$$\ln\left(\Lambda_{\rm ul}/g_{\rm ul}\, A_{\rm ul}\, \Delta E_{\rm ul}\right) = C_\circ - \Delta E_u/k_B\, T_{\rm ex}\ . \tag{8.49}$$

Here $C_\circ$ is a nondimensional number depending on $n_{\rm H_2}$, $T_{\rm ex}$, and molecular constants, and ΔE_u is the energy of the upper level above ground. In an optically-thin environment, the observed intensity $I_{\rm ul}$ is proportional to $\Lambda_{\rm ul}$. Now *if* the region in question were in LTE, $T_{\rm ex}$ would equal the gas temperature T_g. Then a plot of $\ln(I_{\rm ul}/g_u A_{\rm ul} \Delta E_{\rm ul})$ versus ΔE_u for all the observed H_2 lines would yield a straight line with slope $-1/k_B T_g$. Of course, the molecular hydrogen in a photodissociation region of subcritical density is *not* in LTE, so that such an *excitation diagram* should not exhibit a unique slope. In some cases of observed H_2 line emission, this is true. In others, however, the diagrams clearly do indicate a single temperature.

Figure 8.9 shows an example of each type. In the first panel is an excitation diagram for NGC 2023, a reflection nebula illuminated by a B star in the molecular cloud L1630. The open and closed circles symbolize, respectively, the “para” and “ortho” forms of H_2. (In classical language, the two proton spins in parahydrogen point in opposite directions.) From the observed intensities, no single slope is evident, although there is a pattern of slopes, and therefore excitation temperatures, among subgroups of lines. In particular, the $1-0$ $S(1)$ and $2-1$ $S(1)$ lines, distinguished by additional outer circles in the figure, have an intensity ratio of 3.7. Since this is larger than 1.8, the region exhibits collisional pumping of the levels, in addition to fluorescent decay. The observed ratio yields, after applying equation (8.49) to each line and subtracting, a $T_{\rm ex}$ of 3600 K. Such a “vibrational temperature” is useful information, but it does not necessarily correspond to any actual T_g.

The second panel shows results from the area of peak emission in the Orion BN-KL region; this area coincides with the infrared source IRc2. Here, a single slope evidently does fit the data. The $1-0$ $S(1)$ to $2-1$ $S(1)$ intensity ratio is now much higher, about 10, and the best-fit $T_{\rm ex}$ using all the points is 2000 K. This figure probably does represent a gas kinetic temperature. In a photodissociation region, however, such high gas temperatures are only attained in cloud envelopes exposed to the largest $G_\circ$-values, where very little H_2 could survive. What then is the excitation mechanism? It is generally agreed that all such regions, of which Orion BN-KL was the first example discovered, represent gas that previously passed through a *shock*, became violently heated as a result, and is now cooling back down to normal cloud temperatures. For the H_2 levels to be maintained in LTE through collisions, the ambient density must be very high, at least 10^6 cm^{-3}. In the present example, the shock creating the emission is generated where the wind from a massive star strikes nearby cloud gas. The generation of shocks, however, is associated with stars of *all* mass and enters so many aspects of stellar formation that we should explore the phenomenon from a broader perspective.

8.4 J-Shocks

Shocks are sharp transitions generated in a fluid when it is subjected to a large pressure gradient. O and B stars, for example, create shocks in two different ways. The wind from such a star encounters a stationary shock front, *i.e.*, a sudden, adverse pressure gradient, when it impacts a molecular cloud. At greater distances, the star creates as HII region, an extended volume heated and ionized by ultraviolet radiation. The pressure of this hot gas on the surrounding material generates supersonic motion and a moving shock wave, which compresses and heats the cloud gas ahead of it as it propagates away from the star. Of course, shock “waves” can always be viewed as stationary “fronts” by an appropriate change of reference frame; we shall henceforth employ that special frame for our discussion.

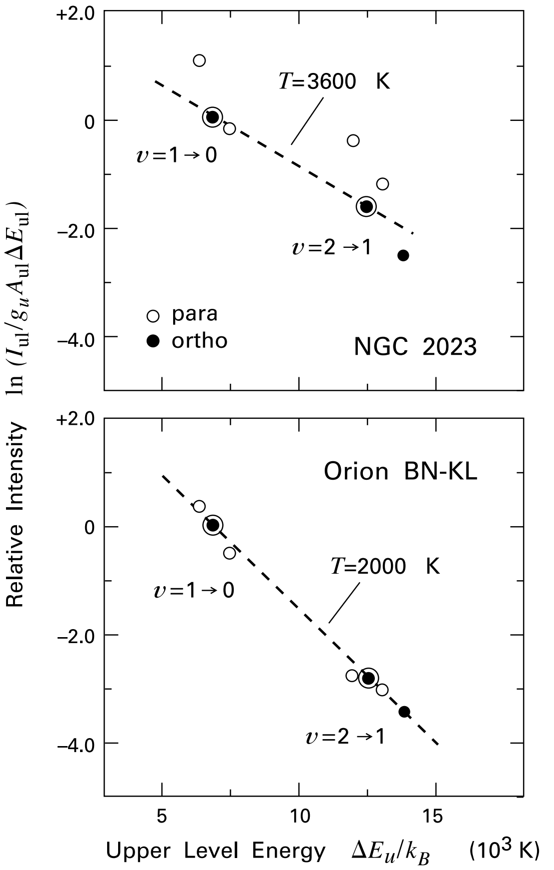


Figure 8.9 Excitation diagrams for H_2 emission in (a) the reflection nebula NGC 2023, and (b) the Orion BN-KL region. The observed intensity is plotted as a function of the energy in each line above the H_2 ground state. Groups of lines corresponding to the $v = 1 \rightarrow 0$ and $v = 2 \rightarrow 1$ transitions are indicated separately.

The properties of the shock transition depend on the preshock density and speed. The latter is known as the *shock velocity* and will be designated $V_{\rm shock}$. A shock arises only when $V_{\rm shock}$ exceeds the local sound speed. A moving fluid element then has no time to be "warned" by sound waves of the approaching high-pressure region. It therefore undergoes a sudden change created by direct contact with the hotter and denser postshock gas. We will later examine relatively slow shocks where this change can occur more gradually. For now, we focus on the faster *J-shocks*, in which all fluid variables *jump* to their postshock values.

8.4.1 Temperature and Density Changes

From a kinetic viewpoint, shocks convert much of the ordered, bulk motion of the preshock gas into random, thermal motion. For sufficiently high shock speeds, the hot postshock gas radiates, and this radiation further heats more distant gas both upstream and downstream from the front itself. Figure 8.10 shows the temperature and density profiles associated with a modestly strong shock in a molecular cloud. The horizontal axis represents distance on either side of the front, as measured by the column density N_H. In Figure 8.10a, preshock gas with number

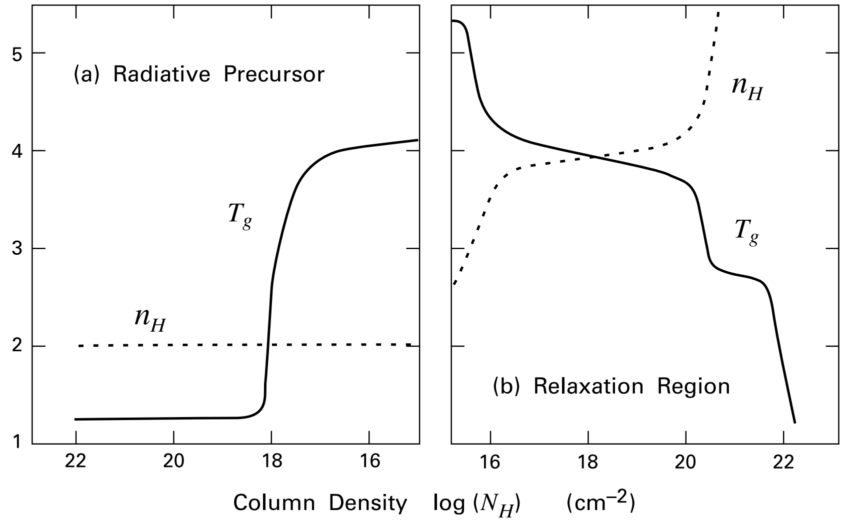


Figure 8.10 Gas temperature and hydrogen number density in a shocked molecular cloud, where $V_{\text{shock}} = 80$ km s^{-1}. The graph actually displays $\log(T_g)$ and $\log(n_H) - 3$. Both quantities are shown as functions of hydrogen column density (a) upstream from the shock front, and (b) behind it. The reference frame is that for which the front is stationary.

density n_H streams to the right at velocity V_{shock}, here equal to 80 km s^{-1}. Meanwhile, the gas temperature T_g is increased by the radiation emitted just downstream from the front. The warmed preshock region is known as the *radiative precursor*. The front itself is the transition layer in which the actual thermalization of motion occurs through collisions between the pre- and postshock atoms and molecules.[3] The thickness of this layer is roughly one particle mean-free-path in the postshock gas. For the example shown in the figure, the relevant mean-free-path pertains to collisions between ions and electrons, and is 4×10^{10} cm. This distance is so small compared to the length scale for variations outside the front that the fluid essentially undergoes a discontinuous change in its properties. The temperature, density, and pressure jump upward, while the velocity is reduced to a value that is subsonic with respect to the local sound speed. Downstream from the front, the gas temperature falls, first quickly and then more slowly, in an extended *relaxation region* (Figure 8.10b). It is the radiation generated in this cooling zone that has provided our knowledge of shocks in star-forming environments.

The changes in the gas properties across the shock are independent of the thermalization mechanism within the front and are readily determined through conservation of mass, momentum, and energy. We derive the corresponding set of *Rankine-Hugoniot jump conditions* in Appendix F. We also derive the ratios of various post– to preshock quantities in terms of the upstream Mach number $M_1 \equiv V_{\text{shock}}/a_1$. Here, we let the subscripts 1, 2, and 3 represent far upstream, immediate postshock, and final, downstream quantities, respectively. For

---

[3] Such momentum exchange through interparticle collisions underlies ordinary fluid viscosity. For this reason, J-shocks are also known as *viscous* shocks.

the case of a strong $(M_1 \gg 1)$ shock in a perfect gas with $\gamma = 5/3$, equation (F.16) tells us that $T_2/T_1 = (5/16)\, M_1^2$. The adiabatic sound speed a_1 is equal to $(5k_B T_1/3\mu m_H)^{1/2}$, so the postshock temperature is

$$\begin{aligned} T_2 &= \frac{3\,\mu\, m_H\, V_{\rm shock}^2}{16\, k_B} \\ &= 2.9 \times 10^5\,{\rm K} \left(\frac{V_{\rm shock}}{100\ {\rm km\ s^{-1}}}\right)^2 . \end{aligned} \tag{8.50}$$

In the second form of this equation, we have assumed $\mu = 1.3$ (neutral preshock gas) and have used, as a fiducial $V_{\rm shock}$, the typical wind speed for a low-mass young star. Note that the massive stars we have been discussing can have wind speeds exceeding 1000 km $\rm s^{-1}$. In any case, we see that wind-generated shocks in molecular clouds create temperatures far in excess of any we have encountered thus far.

The hot postshock gas both decelerates and cools. Once the velocity is well below the local sound speed, the pressure within the relaxation region no longer changes with depth. Hence the density, which is proportional to P/T_g in a perfect gas, can increase considerably by the time the temperature has relaxed to its undisturbed value T_3 (see Figure 8.10b). This increase is, in practice, limited by the rise of magnetic pressure within the gas. Nevertheless, the compressive effect of shocks plausibly underlies the very high densities seen in such regions as Orion BN-KL.

We may also employ the conservation relations to express the total rate of energy loss from the relaxation region. Applying equation (F.17) from Appendix F to a perfect gas, we have

$$\left[\frac{1}{2}\, v^2 + \frac{\gamma}{\gamma - 1}\,\frac{P}{\rho}\right]_1^3 = -\frac{2\, F_{\rm rad}}{\rho_1\, V_{\rm shock}} , \tag{8.51}$$

where v is the fluid velocity relative to the front, and $F_{\rm rad}$ is the flux emitted in either direction. For a strong shock, the term $(1/2)V_{\rm shock}^2$ dominates all others on the left side, and we have simply

$$F_{\rm rad} \approx \frac{1}{4}\,\rho_1\, V_{\rm shock}^3 . \tag{8.52}$$

Thus, $F_{\rm rad}$ increases sharply with the shock speed.

8.4.2 Hydrogen Ionization

The intense radiation from a strong shock can easily destroy any molecules in the preshock flow. In addition, the photons radiated by the postshock gas can have energies exceeding the 13.6 eV limit from HII regions. If their energies also exceed 15.4 eV, the photodissociation of $\rm H_2$ proceeds not by the usual excitation of the Lyman and Werner bands, but rather through direct ionization followed by radiative recombination:

$$\begin{aligned} &{\rm H_2} + h\nu \rightarrow {\rm H_2^+} + {\rm e^-} \\ &{\rm H_2^+} + {\rm e^-} \rightarrow {\rm H} + {\rm H} . \end{aligned} \tag{8.53}$$

Consider now the spatial variation of the ionizing flux through the precursor, *i. e.*, for $x < x_\circ$, where $x_\circ$ denotes the shock front location. Let $\mathcal{F}_{\rm rad}(x)$ denote this flux, as measured in photons cm^{-2} s^{-1}. If we assume that all photons are eventually absorbed, then $\mathcal{F}_{\rm rad}(x)$ starts at zero and increases to some finite value at the front itself. For preshock gas whose hydrogen is initially all molecular, each two H atoms produced in this manner absorb one ionizing photon from the radiation field generated at the front. An additional two photons are required to ionize both atoms. Thus, over a small distance interval Δx within the precursor, the increase in flux is related to the changes in atomic and ionized hydrogen number density by

$$\Delta\mathcal{F}_{\rm rad} \,=\, V_{\rm shock}\left(\frac{1}{2}\,\Delta n_{\rm HI} \,+\, \frac{3}{2}\,\Delta n_{\rm HII}\right) \;.$$

At the shock front, we therefore have

$$\frac{\mathcal{F}_{\rm rad}(x_\circ)}{(n_H)_1\,V_{\rm shock}} \,=\, \frac{1}{2}\,f_{\rm HI}(x_\circ) \,+\, \frac{3}{2}\,f_{\rm HII}(x_\circ) \;. \tag{8.54}$$

Here, $(n_H)_1$ is the incoming number density, while $f_{\rm HI}$ and $f_{\rm HII}$ are, respectively, the number fractions of atomic and ionized hydrogen. In order to completely ionize the precursor flow $(f_{\rm HI} = 0, f_{\rm HII} = 1)$, the outgoing photon flux must be a factor 3/2 larger than the incoming particle flux. Determination of the actual emitted spectrum requires detailed numerical calculations. These find that a shock speed of 120 km s^{-1} is necessary to meet this critical ionization condition. At even higher shock velocities, HI is converted to HII at an *ionization front* which stands off from the shock at some fixed location $x_i < x_\circ$.

Implicit in our discussion is the assumption that the electrons ejected from both H$_2$ and HI do not have time to recombine before the preshock gas is swept into the front. The focusing effect of the Coulomb force causes recombination cross sections to vary with electron velocity as v^{-2}. At the high temperatures of concern, these cross sections are sufficiently small that the assumption is justified. By the same token, the hydrogen just inside the front also cannot recombine promptly, since its density is only increased by at most a factor of 4 (see Appendix F). The initial ionization state of the postshock gas is therefore inherited from the precursor and is *not* set by the local density and temperature, as it would be in LTE. This statement also applies if the incoming gas is lightly ionized. Consider, for example, a shock with $(n_H)_1 = 10^5$ cm^{-3} and $V_{\rm shock} = 80$ km s^{-1}. Calculations show that the hydrogen crossing the front is only 1 percent ionized in this case. On the other hand, the postshock temperature, from equation (8.50), is 2×10^5 K, which would result in essentially complete ionization in LTE.

8.4.3 Nonequilibrium Cooling

The manner in which the postshock gas radiates away its internal energy depends sensitively on its ionization state. Since the latter can differ significantly from LTE, the gas is said to undergo *nonequilibrium cooling*. Much of the radiation behind the front follows the excitation, by electron impact, of electronic levels within relatively abundant heavy elements, such as C, N, and O. Their initial ionization states depend on the character of the shock-generated radiation field and therefore $V_{\rm shock}$. Once excited, the higher electronic levels decay very rapidly, through allowed electric-dipole transitions. The resulting ultraviolet photons are those which actually ionize the precursor.

Whatever hydrogen survives in atomic form can also be excited collisionally. Here, the emitted photons are quickly absorbed by neighboring hydrogen atoms, which most often reemit a photon of the same energy. This process of *resonant scattering* continues until an excited atom drops first to an intermediate level before reaching the ground state. The lower-energy photons produced scatter until most of the original radiation is converted into the Lyα line emitted during the $n = 2 \rightarrow 1$ transition. This line, together with additional ultraviolet photons from heavy ions, constitutes most of the shock's total radiative output.

The hydrogen further downstream at first becomes increasingly ionized through both collisions and the ion-produced radiation field. Meanwhile, the gas temperature falls until, at about 10^4 K, recombination begins to offset photoionization. The cooling of the gas is now largely due to electron collisions with HI, followed eventually by Lyα emission. This cooling is highly sensitive to temperature, since lower electron thermal velocities cannot excite hydrogen's electronic levels. Hence, the temperature maintains a plateau near 10^4 K, which extends until all the shock ultraviolet photons more energetic than 13.6 eV are absorbed and recombination is complete (see Figure 8.10b). Note that nonequilibrium cooling dominates also in the plateau, but that the hydrogen ionization fraction is now *greater* than LTE because of the relative slowness of recombination.

Since HI by itself is a poor cooling agent, the end of hydrogen recombination allows the emission to be dominated once more by the metals. These are neutral or at most singly ionized. The ambient temperature is now too low for collisions to populate those levels within metal atoms corresponding to permitted transitions. However, any levels lying roughly $k_B T_g$ above ground, *i. e.*, about 1 eV for $T_g \approx 10^4$ K, can still be excited. Such *metastable states* actually exist in abundance. An important example is the ${}^1D_2$ state of O I. This is the level, characterized by total electronic quantum numbers $L = 2$ and $S = 0$, that we encountered when discussing fine-structure splitting (recall § 7.3 and Figure 7.8). Electric-dipole transitions to the ground state ($L = 1, S = 1$) are forbidden, but the decay can still occur through a "semi-forbidden," magnetic-dipole transition. Here, the associated A-value is 6.3×10^{-3} s^{-1}. The emitted [OI] 6300 Å line is an important tracer of wind-generated shocks from young stars.

8.4.4 Molecule Formation

The excitation of metastable states allows the postshock temperature to fall steeply once more, until the familiar fine-structure transitions dominate the cooling. Of particular importance is the [OI] 63 µm line, which here overwhelms [CII] 158 µm emission because of the high temperature and density. Once the temperature has dropped to several 10^3 K, molecules start to reform and thereafter control both heating and cooling. First to appear is H_2, initially produced by the H^- formed out of the residual electron supply (recall equation (5.11)). After the electrons are exhausted, H_2 formation continues through grain-surface catalysis. The new molecules are vibrationally excited when first injected back into the gas phase. If n_H at this point is at least 10^5 cm^{-3}, decay of these levels occurs collisionally. Under these conditions, the formation of H_2 becomes the major heating source for the gas, stabilizing T_g at a second plateau of about 500 K until all the hydrogen becomes molecular. This feature is also evident in Figure 8.10b.

Other molecules also form, through purely gas-phase processes. The low ionization level and prolonged temperature elevation of the plateau are conducive to neutral-neutral reactions. A network of such reactions is activated which yields a variety of species, including the important coolants CO, OH, and H_2O. The latter two form mainly through

$$\begin{aligned} H_2 + O &\rightarrow OH + H \\ OH + H_2 &\rightarrow H_2O + H \,. \end{aligned} \tag{8.55}$$

The far-infrared and millimeter rotational emission from these molecules, together with increasing thermal contact with the relatively cold dust, allow the cloud gas finally to settle back down to its original, undisturbed temperature. Note that the total postshock cooling time is only a few years for shocks impacting molecular clouds. In the example shown in Figure 8.10, the distance covered by the relaxation region is of order 10^{13} cm.

8.4.5 Dust Heating and Destruction

For deeply embedded shocks, very little of the optical and near-infrared emission produced behind the front can escape the cloud. As in photodissociation regions, most of this radiation, excluding the ionizing component, is absorbed by dust and reemitted at far-infrared wavelengths. The absorption heats the dust grains, particularly those which actually enter the shock front. Such a grain receives an energy per unit time of $\sigma_d F_{\rm rad}$. Here, $F_{\rm rad}$ is given by equation (8.52), and σ_d is the geometrical cross section, appropriate for ultraviolet radiation. To estimate T_d, we may equate the volumetric heating rate $\sigma_d n_d F_{\rm rad} = \Sigma_d n_H F_{\rm rad}$, to Λ_d in equation (7.39). For $n_H = 10^3$ cm^{-3} and $V_{\rm shock} = 100$ km s^{-1}, we thus find $T_d = 190$ K.

The thermal effect on dust entering high-velocity shocks is minor compared to the *mechanical* effect associated with collisions in the postshock region. Grain-grain collisions drive shocks *inside* the solid material that vaporize it once the deposited energy exceeds several times the binding energy of the lattice. The latter is typically 5 eV per atom. Such collisions can also shatter the grain directly. Most importantly, fast-moving ions from the gas phase chip away the grains' surface layers. This phenomenon of *sputtering* effectively destroys most incoming grains for shock speeds in excess of 200 km s^{-1}. Sputtering by the shocks associated with supernova remnants is the principal destruction mechanism for dust throughout the interstellar medium.

We mentioned previously that the emissivity of the postshock gas is sensitive to its state of ionization. For increasing shock velocity, first hydrogen and then the heavy elements become completely ionized and therefore ineffective as coolants. Suppose that the depth of the relaxation region is limited by some constraint, such as the geometrical thickness of the shocked cloud. Then there exists a critical shock speed above which the postshock gas cannot cool in the available flow time. The energy injected by the preshock gas thus remains trapped for an extended period. Such *nonradiative shocks* are created, for example, by the impact of O- and B-star winds on surrounding cloud matter, a circumstance we shall consider in Chapter 15.

8.5 C-Shocks

We finally consider shocks in which the fluid variables do not undergo a discontinuous jump, but change smoothly and continuously. This possibility was first conjectured theoretically. Later, observations of heated molecular hydrogen seemed to call for just such a transition. Subsequent studies have confirmed the idea and demonstrated its applicability in star-forming regions.

8.5.1 Maximal Compression

The rovibrational transitions of H_2 supply most of the copious emission from the Orion BN-KL region. Figure 8.9 demonstrates that, despite the presence of highly luminous sources, the molecule is not fluorescing, but is in LTE at a temperature near 2000 K. As we stated earlier, the ultimate source of energy is likely to be a wind-driven shock. This hypothesis is bolstered by the presence of the [OI] 63 µm line, which should indeed be a major postshock coolant. On the other hand, it is not clear how the excitation of H_2 actually occurs. Even a moderately fast shock dissociates the molecule, which only later reforms on grain surfaces farther downstream. At this point, the ambient temperature is about 500 K, well below that observed. One could invoke a shock so weak that H_2 passes through the front without being dissociated. However, the resulting infrared emission would then be too low. Increasing the preshock atomic density n_H does raise the collisional excitation rate of H_2, but also leads to the rapid formation of other molecules like H_2O, which then dominate the cooling.

How is it possible, then, for molecular hydrogen to pass through the shock front intact and yet be heated enough to radiate away much of the incoming energy flux? The answer is that the preshock gas in this case contains a relatively strong magnetic field. As we shall detail in Chapter 9, magnetized interstellar gas in motion effectively carries along its internal field. Passage through the front compresses *both* the gas and the magnetic field. Field compression absorbs some of the incoming momentum. The postshock gas pressure, as well as the kinetic temperature, is *lower* than for a nonmagnetized fluid entering with identical velocity. The reduction in temperature prevents molecular dissociation. On the other hand, if the shock speed is high enough, even the diminished postshock temperature is sufficient for H_2 and other species to emit strongly.

The cushioning effect of the magnetic field also limits the density increase attainable by matter as it traverses the shock. Suppose that the pressure associated with the field, $B^2/8\pi$, is negligible in the *preshock* state, as compared to the initial ram pressure, $\rho_1 v_1^2$. Crossing the shock both compresses the gas and raises the magnetic pressure, since the field is frozen into the matter. Indeed, the postshock magnetic field strength, B_2, is related to the preshock value by

$$\frac{B_2}{B_1} = \frac{\rho_2}{\rho_1} \,. \tag{8.56}$$

This relation applies only when the field is perpendicular to the direction of the flow, as in Figure 8.11.

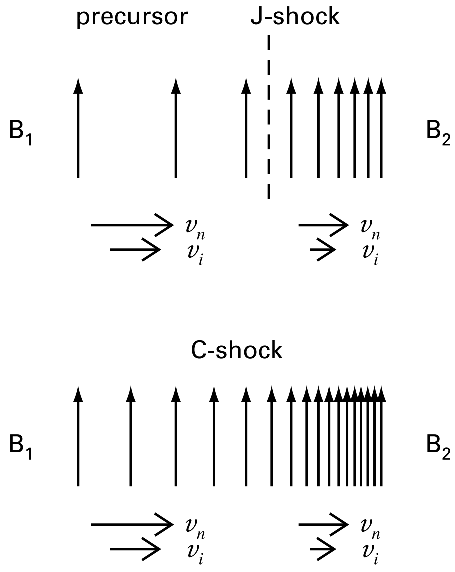


Figure 8.11 Shocks in a magnetized fluid. Depicted are the magnetic field, along with the neutral and ion velocities. Here the field is perpendicular to the flow, and the reference frame is again that in which the shock front is stationary. *Upper panel:* At low field strengths, the neutrals undergo a J-shock after crossing the magnetic precursor. *Lower panel:* For stronger fields, the neutral velocity declines smoothly. In all cases, both the field strength and ion speed change smoothly. The net change in the field is always governed by equation (8.56) in the text.

Maximal compression occurs when the *postshock* value of $B^2/8\pi$ greatly exceeds both the thermal and ram pressures. In this limiting case, momentum conservation implies that

$$\frac{B_2^2}{8\,\pi} = \rho_1\,v_1^2 \; . \tag{8.57}$$

Combining Equations (8.56) and (8.57), we find

$$\begin{aligned}\left(\frac{\rho_2}{\rho_1}\right)_{\rm max} &= \frac{\sqrt{8\pi\rho_1}\,v_1}{B_1} \\ &= \sqrt{2}\left(\frac{v}{V_A}\right)_1 \; .\end{aligned} \tag{8.58}$$

Here, V_A is the *Alfvén velocity*, defined as

$$V_A \equiv \frac{B}{\sqrt{4\pi\rho}} \; . \tag{8.59}$$

We shall later demonstrate that V_A represents the speed at which disturbances propagate along the magnetic field. The ratio of fluid to Alfvén velocities is known as the *Alfvénic Mach number*. Equation (8.58) tells us that the largest possible compression is $\sqrt{2}$ times this ratio, evaluated in the preshock state.

It is instructive to compare our result with that for an unmagnetized gas. In this case, the compression is the *square* of the ordinary Mach number, *i. e.*, the ratio of the fluid velocity to the local sound speed (see Appendix F). Here it is assumed that the postshock temperature has fallen to its preshock value, the circumstance allowing the highest shock compression. The Alfvén velocity within interstellar clouds generally exceeds the sound speed by an order of magnitude. Hence, for any given fluid speed, the ordinary Mach number exceeds the Alfvénic one. We see, then, that the presence of the field severely limits the degree of shock compression.

8.5.2 Magnetic Precursor

The field has another important effect on shock structure. It is only the ions and electrons within the fluid that are subject to the Lorentz force. These charged species then transit the force to the neutral fluid via ion-neutral and, less significantly, electron-neutral collisions. At the very low ionization levels of molecular clouds, the ion-electron fluid and the neutrals may have substantially different speeds. In particular, the magnetic field may decelerate the charged species before the neutrals can significantly alter their speed. The neutral velocity eventually undergoes a sharp jump at an ordinary J-shock, *i. e.*, one meditated by collisions among neutral atoms and molecules. The region of gradually increasing field strength and declining ion speed ahead of the J-shock is known as the *magnetic precursor*.

No fluid can shock unless it is moving faster than the speed of internal pressure disturbances. In the neutral component of the gas, this signal speed is the sound velocity. Now the collisions between neutrals and ions in the magnetic precursor not only transmit momentum, but also raise the temperature within the neutral gas, and hence the sound speed. Since the incoming velocity is less supersonic, the density enhancement across the J-shock is reduced. For a sufficiently strong field, the J-shock disappears altogether, and *neither* fluid undergoes a sharp discontinuity. We call the transition in this case a *C-shock*, where the prefix denotes "continuous." The smooth deceleration of the neutrals stems from collisional drag with the relatively slow ions. [4]

In summary, there are two possible types of shocks within a magnetized gas (see Figure 8.11). If the field is relatively weak, the neutral matter still undergoes a J-shock, characterized by a sharp increase in temperature and density, and a corresponding drop in velocity from supersonic to subsonic values. Upstream is the extended magnetic precursor, in which the field strength builds and the velocity of electrons and ions declines. If the ambient magnetic field is stronger, there is no viscous shock at all. The fluid temperature and density increase smoothly, with the rise in density being limited by equation (8.58). The neutral velocity still falls, but the decline is again gradual, and the velocity itself remains supersonic through the transition region.

Figure 8.12 displays numerical results for the structure of a C-shock in a molecular cloud. As in the previous figure, gas enters the stationary front from the left. In this particular example, the preshock cloud density is $n_H = 10^4\,\mathrm{cm}^{-3}$, while V_{shock} is 25 km s^{-1}. The preshock magnetic field strength, B_1, is 100 μG. If the shock were propagating at the same speed into a nonmagnetic cloud, equation (8.50) would give a postshock temperature of 3.4×10^4 K. Detailed calculations confirm that, at a preshock density of 10^4 cm^{-3}, all the molecular hydrogen would be dissociated. Figure 8.12 shows, however, that the actual peak temperature is only 1200 K, and the dissociation is negligible. The figure also demonstrates that the ions are the first to decelerate, while the neutral velocity is only later reduced. Eventually, however, all species are comoving again. The whole C-shock in this case spans a distance of 3×10^{16} cm.

---

[4] The incoming neutrals are moving faster than either the sound speed or V_A, as given in equation (8.59). However, the signal speed in the ion-electron fluid is a modified Alfvén velocity, obtained by replacing ρ in equation (8.49) by the total density of ions and electrons, $\rho_i + \rho_e \approx \rho_i \ll \rho$. Since the actual ion velocity is always well under this new signal speed, the charged species decelerate smoothly, creating a drag on the neutrals.

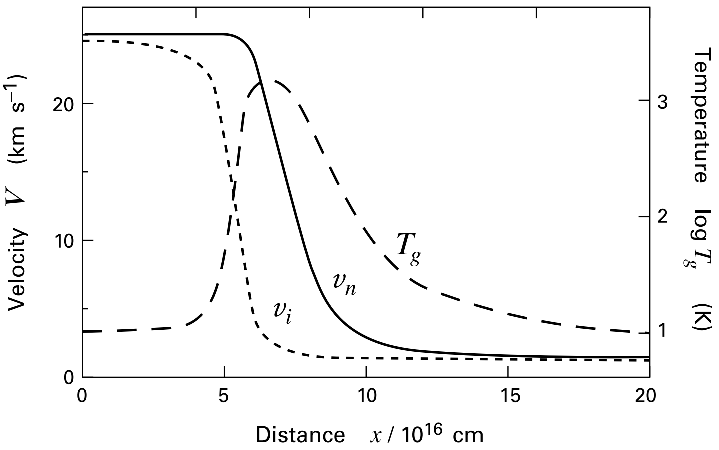


Figure 8.12 Structure of a C-shock in a molecular cloud. Here, $V_{\rm shock}$ = 25 km s^{-1}, $n_H = 10^4$ cm^{-3}, and B = 100 µG, where n_H and B refer to the preshocked state. Displayed are the gas temperature T_g, the neutral velocity v_n, and the ion velocity v_i. The shock is stationary in the adopted reference frame.

8.5.3 Heating and Cooling Mechanisms

Within a J-shock, the immediate postshock temperature follows from the Rankine-Hugoniot jump conditions, while the subsequent decline in the relaxation region is due to the various cooling mechanisms already discussed. In a C-shock, both the temperature rise and decline are smooth and, as we have seen, cover a larger total distance. The temperature profile depends sensitively on both the internal heating and cooling. Nevertheless, the jump conditions may still be used to relate upstream and downstream material outside the interaction region, *i. e.*, in which both charged and neutral species have a common speed and kinetic temperature. Heating within the shock mainly stems from the differential velocity between the ions and neutrals. This slip generates random motion in the neutral atoms and molecules, as a result of repeated scattering.

Interstellar grains similarly impart thermal energy through collisions, although their temperature remains well below that of the gas. Additionally, the dust effectively transfers momentum between the neutrals and ions. The reason is that individual grains carry an electrical charge, even within quiescent clouds. (Recall § 8.3 for the situation near massive stars.) The charged bodies, like ions and electrons, gyrate about the local magnetic field. Their collisions with the neutrals then exert a drag force on this fluid.

Much of the cooling is from molecular lines, all in the infrared regime. We have already mentioned the rovibrational transitions of H_2. Rotational lines from CO, OH, and H_2O are also significant, and sometimes paramount in the energy balance. Note that the CO lines include much higher levels (*e. g.*, $J = 15 \rightarrow 14$) than in quiescent clouds. Cooling by H_2O does not dominate until the preshock density is about 10^6 cm^{-3}. Atomic fine-structure lines also contribute, principally [OI] 63 µm. Notably absent are the ultraviolet and optical transitions that characterize J-shocks. In the example of Figure 8.12, the initial temperature rise reflects the dominance of ion-neutral heating over O I cooling. The peaking and decline of the temperature occur once H_2 cooling begins to take over. In addition, the neutrals eventually decelerate to the ion speed, so the velocity slip is no longer a heat source.

Thus far, we have described C-shocks as a phenomenon occurring when the ambient magnetic field exceeds a certain strength. We may alternatively consider a *fixed* ambient field and ask what happens when we change the preshock flow velocity. It now becomes apparent that efficient cooling is fundamental for the existence of C-shocks. As long as the ambient field strength is above some threshold value, modestly supersonic velocities give rise to these continuous transitions. For too large a speed, the increased heating from ion-neutral slippage overwhelms the available cooling, and the molecules begin to dissociate. However, the molecules themselves supply much of the cooling. Hence the effect is catastrophic. Even a small amount of dissociation lowers the cooling and raises the temperature, leading to further dissociation, and eventually ionization. The temperature approaches T_2, as given in equation (8.50), and we have a J-shock, albeit with a magnetic precursor. Where this transition occurs depends not only on the field strength, but also on the preshock degree of ionization. Numerical studies involving magnetic fields with reasonable field values and ionizations for molecular cloud environments find that the critical shock velocity lies between 40 and 50 km s^{-1}.

8.5.4 The Wardle Instability

Returning to the Orion BN-KL region, the H_2 emission has features that remain problematic, at least for the simplest C-shock models. The individual lines tend to be quite broad, with velocity widths often exceeding 100 km s^{-1}. A planar C-shock normal to the flow can yield emission over a range of velocities, depending on the actual speed of the preshocked gas. The total range of velocities, however, cannot exceed the upper limit to V_{shock} of about 50 km s^{-1}. On the other hand, suppose that gas enters the planar shock obliquely. Then the velocity component parallel to the front is preserved. This component could easily exceed 100 km s^{-1}, depending on the wind speed and the orientation of the shock relative to this flow. Broadened lines could then be produced by the superposition of many such oblique shocks, each with a different orientation; *i. e.*, from a *curved*, rather than planar, shock front. Over part of the curved surface, the impact speed could be high enough to result in a J-shock, while the rest could have a C-type character.

We shall detail in Chapter 13 how curved shock fronts naturally arise whenever collimated, jet-like winds impact molecular clouds, and how the fronts indeed yield broad emission lines. In the Orion region itself, infrared observations of high spatial resolution have found numerous small arcs glowing in H_2. These appear to be individual bowshocks, produced when a wide-angle spray of material from IRc2 impacts surrounding cloud gas. Whether all of the hydrogen emission from the region can be modeled in this way remains to be seen.

Even the simplest, planar C-shock must be, on purely theoretical grounds, an idealization. Consider the forces acting on a typical ion. In the lower panel of Figure 8.11, the drag force arising from collisions with neutrals acts to the right and is proportional to the velocity difference $\boldsymbol{v}_n - \boldsymbol{v}_i$. The ions are also subject to the Lorentz force, which in turn is proportional to $\boldsymbol{j} \times \boldsymbol{B}$. In the case shown, $\boldsymbol{j}$ points out of the page, so that $\boldsymbol{j} \times \boldsymbol{B}$ opposes the drag. In fact, the Lorentz force is the larger of the two and leads to the deceleration of the ions as the flow proceeds to the right.

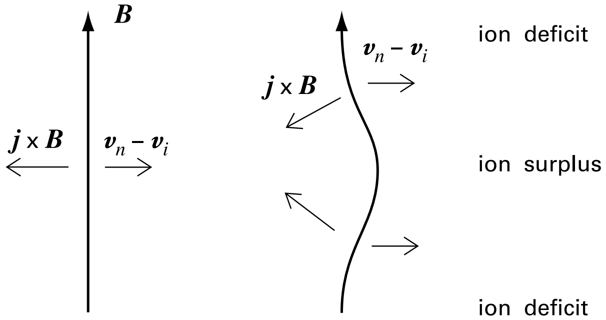


Figure 8.13 Origin of the Wardle instability. When the magnetic field is straight and perpendicular to the flow, the drag force on the ions and the Lorentz force are antiparallel. When the field is bent, the two forces cannot be collinear, and ions build up and deplete at alternate locations.

Suppose now that we impose a sinusoidal ripple on the magnetic field. Then, as illustrated in Figure 8.13, the Lorentz force must still be locally perpendicular to each field line and so changes direction spatially. Since the velocities $\boldsymbol{v}_n$ and $\boldsymbol{v}_i$ are as yet unaltered, the drag force is still horizontal. It is evident from the figure that $\boldsymbol{v}_n - \boldsymbol{v}_i$ may have a component *along the field line*, which cannot be opposed by $\boldsymbol{j} \times \boldsymbol{B}$. The result is that ions slide together along the field, increasing the density at some points and decreasing it at others. However, an increased density leads to greater drag on the neutrals, which in turn pull the ions farther in the flow direction, warping the field even more.

There are a number of mitigating factors that can stifle growth of the perturbation. If, for example, the initial ripple has too short a wavelength, the sharp rise in magnetic tension straightens the field again. Nevertheless, there exists a range of wavelengths for which even a tiny perturbation grows to large amplitude, provided the Alfvénic Mach number is sufficiently high. Numerical simulations of this *Wardle instability* find that gas tends to collect in thin sheets lying along the flow direction and oriented perpendicular to the magnetic field. In the final, steady-state pattern, the increased drag within the sheets is offset by bending of the distorted field. The density within a sheet is much higher than our estimate in equation (8.58). Despite this complexity, most excitation of molecular lines occurs upstream of the region where the sheets start to form. The observed flux, therefore, differs little from that of an idealized, planar C-shock.

Chapter Summary

Clouds exposed to interstellar radiation have an outer envelope of atomic hydrogen. Further inside, the atoms recombine on grain surfaces to form H_2. The newly formed molecules, along with ambient grains, absorb ultraviolet photons and thereby shield other molecules from dissociation. In the absence of nearby, massive stars, buildup of H_2 is essentially complete by a depth corresponding to $A_V = 2$.

Gas in the molecular interior is heated principally by cosmic rays. Even in the highest-density regions, this penetrating flux maintains a small ionization level. Molecular ions thus created transfer their positive charge to metal atoms, which in turn stick to grains. The equilibrium abundances of ions and free electrons may be calculated theoretically and agree with those obtained indirectly through the detection of species such as HCO^+.

Ultraviolet radiation from massive stars forms photodissociation regions in cloud gas. These are mainly seen through their far-infrared emission, produced by heated dust. Electrons ejected from grains also heat the gas, so that it radiates in the 63 μm and 158 μm lines of O I and C II, respectively. Any H_2 directly exposed to the ultraviolet photons emits a characteristic fluorescent spectrum of infrared lines.

Massive stars also produce strong winds. A wind impacting lightly magnetized cloud matter creates a J-type shock. Matter crossing the front is suddenly heated and emits radiation that can fully ionize the upstream gas. Grains may also be destroyed in the shock through their mutual collisions. Gas further downstream cools until molecules eventually reform. If the cloud is strongly magnetized, the wind creates a C-type shock, at least below some limiting speed. All properties of the incoming fluid now vary smoothly because of ion-neutral friction. This friction creates heat, but the temperature rise is relatively shallow. Any ripples in the shock front become amplified, so that gas collects in sheets oriented along the flow direction. Observationally, J- and C-type shocks coexist in some environments, including the Orion BN-KL region.

Suggested Reading

Section 8.1 The theory of hydrogen self-shielding in cloud envelopes was formulated by

Hollenbach, D. J., Werner, M. W., & Salpeter E. E. 1971, ApJ, 163, 155.

A useful later reference is

Federman, S. R., Glassgold, A. E., & Kwan, J. 1979, ApJ, 227, 466.

The buildup of CO is discussed in

Van Dishoeck, E. F. & Black, J. H. 1987, in *Physical Processes in Interstellar Clouds*, ed. G. E. Morfill and M. Scholer (Dordrecht: Reidel), p. 214.

Section 8.2 The temperature structure of cloud interiors has been calculated by

Le Bourlot, J., Pineau de Forets, G., Roueff, E., & Flower, D. 1993, AA, 267, 233.

The method of using HCO^+ to determine the electron fraction in dark clouds is due to

Wootten, A., Snell, R., & Glassgold, A. E. 1979, ApJ, 234, 876.

For a later evaluation of this fraction, see

Caselli, P., Walmsley, C. M., Terzieva, R. & Herbst, E. 1998, ApJ, 499, 234.

The theory of ionization balance in molecular clouds has been thoroughly reviewed by

Nakano, T. 1984, Fund. Cosm. Phys., 9, 139.

Section 8.3 The basic model of photodissociation regions is in

Tielens, A. G. G. M. & Hollenbach, D. J. 1985, ApJ, 291, 722.

The 63 µm line of O I was discovered by

Melnick, G., Gull, G. E., Harwit, M. 1979, ApJ, 227, L29,

and the 157 µm C II line by

Russell, R. W., Melnick, G., Gull, G. E., & Harwit, M. 1980, ApJ, 240, L99.

A concise theoretical review of fluorescent H_2 emission is

Sternberg, A. 1990, in *Molecular Astrophysics*, ed. T. W. Hartquist (Cambridge: Cambridge U. Press), p. 384.

Section 8.4 A detailed analysis of the physical processes and emission from fast shocks is in

Hollenbach, D. J. & McKee, C. F. 1979, ApJSS, 41, 555,

from which our numerical results are taken. Gas-phase chemistry is emphasized in the review of

Neufeld, D. A. 1990, in *Molecular Astrophysics*, ed. T. W. Hartquist (Cambridge: Cambridge U. Press), p. 374.

For the theory of grain destruction through sputtering, see

Draine, B. T. & Salpeter, E. E. 1979, ApJ, 231, 77,

as well as

Tielens, A. G. G. M., McKee, C. F., Seab, G., & Hollenbach, D. J. 1994, ApJ, 431, 321,

which incorporates later experimental findings.

Section 8.5 The discovery of C-shocks is due to

Mullan, D. J. 1971, MNRAS, 153, 145

Draine, B. T. 1980, ApJ, 241, 1021.

The first paper sets out the general concepts, but is limited in its applications to relatively weak shocks in HI gas. The second demonstrates that stronger shocks can exist in molecular clouds, because of enhanced cooling.

For the Wardle instability and its numerical simulation, see

Wardle, M. 1990, MNRAS, 246, 98

MacLow, M. & Smith, M. D. 1997, ApJ, 491, 596.

Two useful reviews are

Draine, B. T. & McKee, C. F. 1993, ARAA, 31, 373

Brand, P. W. J. L. 1995, ApSS, 224, 125.

The first is quite general, while the second addresses the issue of whether curved bowshocks can explain the infrared observations.

Part III

From Clouds to Stars

The Formation of Stars
Steven W. Stahler and Francesco Palla
© 2004 WILEY-VCH Verlag GmbH & Co.

9 Cloud Equilibrium and Stability

We now turn our attention to the dynamical theory of star formation. In doing so, we shift emphasis from the thermal properties of molecular clouds to their mechanical behavior, both as static entities prior to stellar birth (the present chapter), and during the collapse process itself (Chapter 10). The theoretical considerations presented here, in combination with our previous empirical survey, will make it clear that the formation of stars is *not* simply the result of giant clouds breaking apart into tiny, dense substructures. The onset of collapse is rather a highly localized occurrence within large complexes, and the character of that collapse dictates the structure of the nascent protostar (Chapter 11). Having said this, it is also true that individual collapses can occur over extensive regions of a complex. Chapter 12 examines the empirical data and main theoretical ideas concerning the formation of stellar groups.

Since most molecular gas is apparently not in a state of collapse, it is important to understand first the balance of forces allowing clouds to persist over long periods of time. Support against gravity arises partially from thermal pressure, but also from the interstellar magnetic field, especially on the largest scales. The latter portion of this chapter accordingly discusses aspects of magnetic support. The presentation in these sections is at a higher technical level than in previous chapters. Nevertheless, the reader equipped with a basic, working knowledge of electromagnetic theory should find no essential difficulty in following the various arguments.

9.1 Isothermal Spheres and the Jeans Mass

We begin by analyzing a simplified cloud that maintains equilibrium only through the forces of self-gravity and thermal pressure. We further ignore any internal temperature gradients, *i. e.*, we specify an isothermal equation of state. From our discussion in Chapter 8, this last condition is inappropriate for modeling the larger molecular clouds, but can serve as a useful first approximation in the case of dense cores and Bok globules (see Figures 3.21 and 8.6). One should also recall the empirical finding that mechanical support from MHD waves, as evidenced by enhanced molecular line widths, diminishes only at these smallest scales (Chapter 3). Nevertheless, we shall find that some of the lessons drawn from purely pressure-supported, isothermal configurations shed light even on the giant cloud complexes.

9.1.1 Density Structure

The reader has already encountered the mathematical expression of hydrostatic equilibrium (equation (2.4)), as well as the equation of state for an ideal, isothermal gas (equation (2.6)), both originally framed in the context of HI clouds. Generalizing to arbitrary chemical composition,

The Formation of Stars. Steven W. Stahler and Francesco Palla
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-40559-3

we have

$$-\frac{1}{\rho}\,\boldsymbol{\nabla} P \,-\, \boldsymbol{\nabla}\Phi_g \,=\, 0 \; , \tag{9.1}$$

and

$$P \,=\, \rho\, {a_T}^2 \; , \tag{9.2}$$

where a_T, the isothermal sound speed, is $(\mathcal{R}T/\mu)^{1/2}$. The gravitational potential Φ_g in (9.1) obeys Poisson's equation (2.8), with the righthand side now referring to the density of the cloud itself:

$$\nabla^2\Phi_g \,=\, 4\,\pi\, G\,\rho \; . \tag{9.3}$$

Initially, we limit ourselves to spherically symmetric clouds. Quite generally, equations (9.1) and (9.2) together imply that the sum $(\ln\,\rho \,+\, \Phi_g/a_T^2)$ is a spatial constant. For the spherical case, we thus write

$$\rho\,(r) = \,\rho_c \exp\left(-\Phi_g/a_T^2\right) \; . \tag{9.4}$$

Here, we have set Φ_g equal to zero at the cloud center ($r = 0$), where the density is denoted as ρ_c. Equation (9.3) now becomes

$$\frac{1}{r^2}\,\frac{d}{dr}\left(r^2\,\frac{d\Phi_g}{dr}\right) \,=\, 4\,\pi\, G\,\rho \tag{9.5a}$$

$$= 4\,\pi\, G\,\rho_c \exp\left(-\Phi_g/a_T^2\right) \; . \tag{9.5b}$$

It is useful to recast equation (9.5b) into dimensionless form. We define a new dependent variable ψ as Φ_g/a_T^2 and a nondimensional length ξ by

$$\xi \,\equiv\, \left(\frac{4\,\pi\, G\,\rho_c}{a_T^2}\right)^{1/2} r \; . \tag{9.6}$$

Equation (9.5b) then becomes the *isothermal Lane-Emden equation:*[1]

$$\frac{1}{\xi^2}\,\frac{d}{d\xi}\left(\xi^2\,\frac{d\psi}{d\xi}\right) \,=\, \exp\left(-\psi\right) \tag{9.7}$$

One boundary condition for this equation has already been specified: $\psi(0) = 0$. To derive a second condition, we note that the dimensional gravitational force per unit mass is $-GM(r)/r^2$, where $M(r)$ is the mass interior to r. Since this mass, in turn, approaches $(4\pi/3)\rho_c r^3$, it follows that the force, and therefore $\psi'(\xi)$, must vanish as ξ goes to zero.

The dashed curve in Figure 9.1 displays the function $\psi(\xi)$, as obtained by numerical integration of equation (9.7). Of greater interest is the ratio ρ/ρ_c (solid curve), which, from equation (9.4), is given by exp$(-\psi)$. Notice how the density, and hence the pressure, drop monotonically away from the center. This fall in pressure at every radius, a characteristic of all hydrostatic configurations whether isothermal or not, is necessary to offset the inward pull

---

[1] The generic Lane-Emden equation, which was of great historical importance in the development of stellar structure theory, governs the structure of spherical *polytropes*. These are hydrostatic configurations in which P is proportional to $\rho^{1\,+\,1/n}$, where n is a constant. Equation (9.7) applies in the limit $n \rightarrow \infty$.

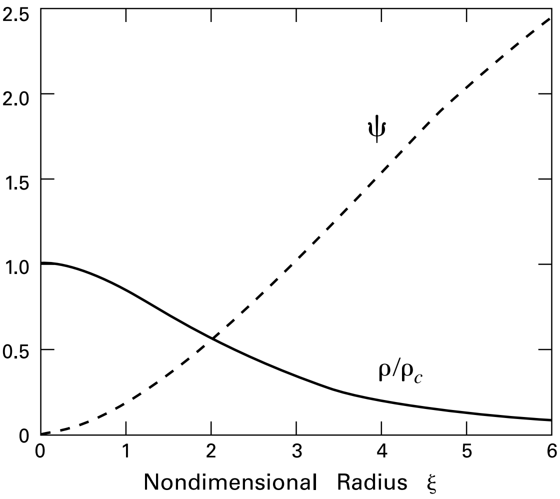


Figure 9.1 Nondimensional gravitational potential (*dashed curve*) and density (*solid curve*) in a spherical, isothermal cloud. Both quantities are plotted as a function of nondimensional radius.

of gravity. At large distances ($\xi \gg 1$), ρ/ρ_c asymptotically approaches $2/\xi^2$. The reader may verify that the corresponding potential $\psi = \ln{(\xi^2/2)}$ in fact satisfies equation (9.7) but not the boundary conditions at $\xi = 0$. The dimensional density in this *singular isothermal sphere* is

$$\rho(r) = \frac{a_T^2}{2\,\pi\, G\, r^2} \ , \tag{9.8}$$

and is frequently useful for estimating cloud properties.

In any actual cloud, the pressure does not fall to zero, but to some value $P_\circ$ characterizing the external medium. Suppose we fix $P_\circ$ and a_T. How does Figure 9.1 tell us the properties of the cloud, given these constraints? From equation (9.2), we know the density at the edge, $\rho_\circ$. Suppose we further specify that the cloud have a certain density contrast from center to edge, $\rho_c/\rho_\circ$. Then we may read off from Figure 9.1 the nondimensional radius $\xi_\circ$. The corresponding dimensional radius $r_\circ$ follows simply by inversion of equation (9.6). To summarize, Figure 9.1 actually describes an infinite *sequence* of models, conveniently parametrized by $\rho_c/\rho_\circ$.

Let us now consider the dimensional mass M for each model. Integration over spherical shells gives

$$\begin{aligned} M &= 4\,\pi \int_0^{r_\circ} \rho\, r^2\, dr \\ &= 4\,\pi\, \rho_c \left(\frac{a_T^2}{4\,\pi\, G\, \rho_c}\right)^{3/2} \int_0^{\xi_\circ} \mathrm{e}^{-\psi}\, \xi^2\, d\xi \ . \end{aligned} \tag{9.9}$$

Using equation (9.7) and the boundary condition $\psi'(0) = 0$, the last integral is equal to

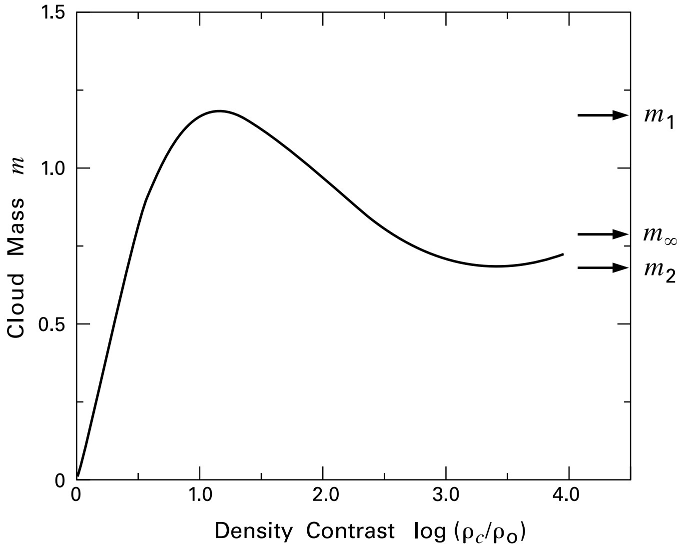


Figure 9.2 Nondimensional mass of pressure-bounded, isothermal spheres. The mass is shown as a function of the density contrast from center to edge.

$\xi^2\, d\psi/d\xi$, evaluated at $\xi_\circ$. If we define a nondimensional cloud mass m by

$$m \equiv \frac{P_\circ^{1/2}\, G^{3/2}\, M}{a_T^4}\ , \tag{9.10}$$

our final result is that

$$m = \left(4\pi\,\frac{\rho_c}{\rho_\circ}\right)^{-1/2} \left(\xi^2\,\frac{d\psi}{d\xi}\right)_{\xi_\circ}\ . \tag{9.11}$$

The value of $\xi_\circ$ is known for each $\rho_c/\rho_\circ$. Thus, the last factor on the right side of (9.11) may be read from the $\psi(\xi)$ curve in Figure 9.1. Figure 9.2 displays the function $m\,(\rho_c/\rho_\circ)$ obtained in this manner. At the beginning of the sequence, $\rho_c/\rho_\circ = 1$ and $\xi_\circ = 0$, implying that $m = 0$. With increasing density contrast, m first rises to a maximum value of $m_1 = 1.18$, attained at $\rho_c/\rho_\circ = 14.1$. The mass then drops to a minimum of $m_2 = 0.695$, eventually approaching in an oscillatory fashion the asymptotic limit $m_\infty = (2/\pi)^{1/2} = 0.798$. The reader may verify directly from equation (9.8) that m_∞ represents the nondimensional mass of the singular isothermal sphere.

9.1.2 Gravitational Stability

It should be apparent by now that all the physical characteristics of isothermal spheres follow from integration of the single equation (9.7). However, only a limited subset of the full model sequence is *gravitationally stable*. In all other clouds, an arbitrarily small initial perturbation in the structure grows rapidly with time, leading ultimately to collapse. The issue of gravitational stability is central in star formation theory, so we should examine it with some care.[2]

For a stable cloud, any increase of $P_\circ$ creates both a global compression and a rise of the internal pressure, where the latter acts to re-expand the configuration. Let us verify both these effects for clouds of relatively low density contrast, *i. e.*, those near the start of the curve in Figure 9.2. We begin with the internal pressure. If we hold the dimensional mass M fixed, then equation (9.10) implies that any increase in $P_\circ$ (at constant a_T) causes m to rise. According to Figure 9.2, $\rho_c/\rho_\circ$ must increase as well, provided this ratio does not initially exceed 14.1. Since P is proportional to ρ and decreases monotonically outward, we see that both the central and volume-averaged pressure rise above $P_\circ$ as the latter is increased.

Figure 9.1 shows that this increase in $\rho_c/\rho_\circ$ is accompanied by a rise in $\xi_\circ$. To see how the physical radius $r_\circ$ varies, we employ a Taylor series expansion for $\psi(\xi)$ in equation (9.7). We find that, for small values of ξ,

$$\psi(\xi) = \frac{\xi^2}{6} + \mathcal{O}\left(\xi^4\right) \ , \tag{9.12}$$

so that $\rho_c/\rho_\circ = \exp\left(-\psi\right) \approx 1 - \xi_\circ^2/6$ is the density contrast for a cloud of nondimensional radius $\xi_\circ$. Using these results in Equations (9.11) and (9.12) leads to

$$M \approx \frac{\xi_\circ^3}{6\,\pi^{1/2}} \frac{a_T^4}{P_\circ^{1/2}\,G^{3/2}} \ . \tag{9.13}$$

Equation (9.6) can be recast as

$$\xi_\circ^3 = (4\,\pi)^{3/2} \left(\frac{\rho_c}{\rho_\circ}\right)^{3/2} \frac{P_\circ^{3/2}\,G^{3/2}\,r_\circ^3}{a_T^6} \ , \tag{9.14}$$

where we may set $\rho_c/\rho_\circ$ to unity in the regime of interest. After making this approximation and eliminating $\xi_\circ^3$ between (9.13) and (9.14), we have finally

$$r_\circ^3 \approx \frac{3\,M\,a_T^2}{4\,\pi\,P_\circ} \ . \tag{9.15}$$

We conclude that $r_\circ$ indeed shrinks as $P_\circ$ increases. Moreover, the product of $P_\circ$ and the cloud volume $(4\pi/3)\,r_\circ^3$ remains constant. Equation (9.15) is thus a restatement of Boyle's law for an ideal, isothermal gas.

The fact that the constant G does not appear in equation (9.15) means that clouds of low density contrast are mainly confined by the external pressure, and not self-gravity. This situation changes as we progress along the curve in Figure 9.2 to models of higher $\rho_c/\rho_\circ$. In such gravity-dominated configurations, it is more difficult for the central regions to expand after application

---

[2] A variety of instabilities occur in astrophysical fluids. The gravitational type is one example of *dynamical* instability, characterized by rapidly growing internal perturbations. Other examples we shall soon encounter are *rotational* and *convective* instability. In Chapter 2, we discussed *thermal* instability of diffuse interstellar gas, a wholly different sort *not* involving dynamical motion. Finally, we will see in Chapter 10 that magnetized molecular clouds are *secularly* unstable, since they slowly evolve to lower-energy configurations through frictional dissipation.

of an enhanced $P_\circ$. Referring again to Figure 9.2, all clouds with $\rho_c/\rho_\circ > 14.1$, *i. e.*, those to the right of the first maximum, are gravitationally unstable. The critical value of M is known as the *Bonnor-Ebert mass*:

$$M_{\rm BE} \,=\, \frac{m_1\, a_T^4}{P_\circ^{1/2}\, G^{3/2}} \ . \tag{9.16}$$

To understand better the significance of $M_{\rm BE}$, let us broaden our perspective of the stability issue. The perturbation of any cloud in equilibrium creates internal oscillations. In a *normal mode*, the sinusoidal variation of any physical quantity at a given location occurs with the same frequency and phase throughout the volume; only the *amplitude* of the disturbance varies spatially. Mathematically, one writes each dependent variable as the sum of its static, equilibrium value and a small oscillatory component. Thus, if we consider spherically symmetric oscillations of our isothermal clouds, the density is the real part of

$$\rho\,(r,t) \,=\, \rho_{\rm eq}\,(r) \,+\, \delta\rho\,(r)\exp\,(i\,\omega\, t) \ , \tag{9.17}$$

where $\rho_{\rm eq}(r)$ is the unperturbed function. The amplitude $\delta\rho\,(r)$ is complex, to allow for a relative phase between the oscillations of different variables. The frequency ω, on the other hand, is identical for all variables. One introduces such perturbations into the basic Equations (9.2) and (9.3), as well as the mass continuity relation (3.7) and the non-magnetic version of the momentum equation (3.3):

$$\rho\,\frac{D\boldsymbol{u}}{Dt} \,=\, -\boldsymbol{\nabla} P \,-\, \rho\,\boldsymbol{\nabla}\Phi_g \ . \tag{9.18}$$

Retaining only terms that are linear in the various amplitudes, one solves the perturbation equations to obtain both the eigenfunction $\delta\rho\,(r)$ and the eigenvalue ω^2 for each mode of interest. If $\omega^2 > 0$, the density and all other physical quantities undergo oscillations of fixed amplitude. On the other hand, $\omega^2 < 0$ implies that the perturbation can grow exponentially.[3]

For any equilibrium cloud model, there exists an infinite sequence of normal modes, each with its own value of ω^2. These may be ordered by the number of nodes, *i. e.*, the radii where the amplitude of the perturbed fluid displacement is zero. The first, or *fundamental*, mode has no displacement nodes. Here the perturbed cloud "breathes" in and out as a whole. The fundamental mode also has the lowest associated value of ω^2. The *first harmonic* has one node and the next higher (more positive) ω^2-value.

Suppose that we again fix $P_\circ$ and a_T, and consider, as in Figure 9.3, two clouds of the same mass m near one of the extrema of the curve in Figure 9.2. These models have differing radii and central densities, as sketched here. However, since their masses are identical, the small displacements that connect each fluid element in one model to the corresponding element in the other can be regarded as a normal mode of zero frequency. By this reasoning, each time we pass a maximum or minimum along the mass curve, *some* normal mode undergoes a stability transition.

---

[3] Our discussion assumes that ω^2 is real. In the detailed analysis, this fact follows from the radial symmetry of the perturbations. Modes that are asymmetric about some central axis have complex values of ω^2, corresponding to oscillations that grow or decay in time. Growing oscillations can also occur in the outer regions of stars, as we shall discuss in Chapter 18.

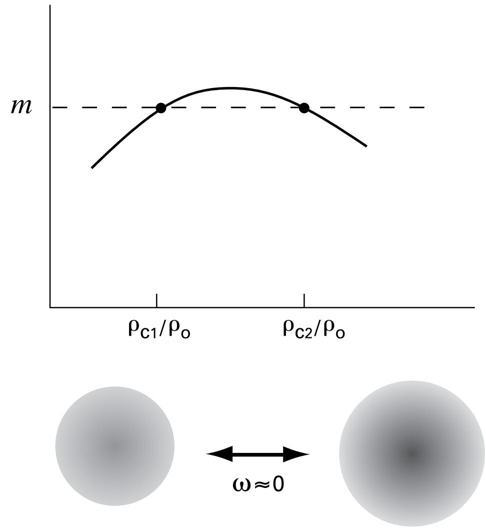


Figure 9.3 Stability transition in isothermal clouds. The sketch shows a turnover in cloud mass as a function of density contrast. Models on either side of the peak can be viewed as the extremal states in a zero-frequency oscillation, as sketched below.

The fact that clouds of the lowest density contrast are stable means that all their normal modes have positive values of ω^2. Just at the Bonnor-Ebert mass, the fundamental mode becomes unstable. Successively higher modes undergo this transition at the other extrema of the mass curve. The singular isothermal sphere is thus unstable to *all* spherically symmetric normal modes. However, since perturbations are inevitable in a realistic interstellar environment, the presence of even one unstable mode guarantees that the underlying equilibrium model is not viable. Linear theory allows either expansion or contraction of the unstable cloud. In practice, an initially expanding cloud can never draw in sufficient energy to disperse itself to infinity, so the presence of an unstable mode always leads to gravitational collapse.

9.1.3 Critical Length Scale

We have focused thus far on spherical clouds, but the phenomenon of gravitational stability is more general. For example, one reinterpretation of the result embodied in equation (9.16) is that a certain *size* scale of isothermal gas is prone to collapse, regardless of the specific, three-dimensional configuration. To illustrate this latter viewpoint, let us follow Jeans' classic analysis of self-gravitating waves propagating through a uniform, isothermal gas of density $\rho_\circ$. In place of equation (9.17), which prescribes a *standing-wave* perturbation, we employ a plane *traveling wave*:

$$\rho\,(x,t) \;=\; \rho_\circ \,+\, \delta\rho\,\exp\left[i(kx \,-\, \omega t)\right]\;, \tag{9.19}$$

where x is the direction of propagation and $k \equiv 2\pi/\lambda$ is the wave number. By assumption, the small velocity induced by the perturbation is also in this direction. We substitute analogous traveling-wave forms for all variables into Equations (3.7), (9.2), (9.3), and (9.18). After linearizing in the amplitudes and canceling the exponentials, we find

$$-i\,\omega\,\delta\rho \,+\, i\,k\,\rho_\circ\,\delta u \,=\, 0 \tag{9.20a}$$

$$\delta P \,=\, \delta\rho\,a_T^2 \tag{9.20b}$$

$$-k^2\,\delta\Phi_g \,=\, 4\,\pi\,G\,\delta\rho \tag{9.20c}$$

$$-i\,\omega\,\rho_\circ\,\delta u \,=\, -i\,k\,\delta P \,-\, i\,k\,\rho_\circ\,\delta\Phi_g \ . \tag{9.20d}$$

Multiplying (9.20a) by $-i\omega$, (9.20d) by $+ik$, and subtracting yields

$$\begin{aligned} -\omega^2\,\delta\rho \,&=\, -k^2\,\delta P \,-\, k^2\,\rho_\circ\,\delta\Phi_g \\ &=\, -k^2\,a_T^2\,\delta\rho \,-\, k^2\,\rho_\circ\,\delta\Phi_g \ , \end{aligned} \tag{9.21}$$

where we have further used (9.20b) for δP. If we now substitute from equation (9.20c) for $\delta\Phi_g$ and cancel $\delta\rho$ throughout, we find

$$\omega^2 \,=\, k^2\,a_T^2 \,-\, 4\,\pi\,G\,\rho_\circ \ . \tag{9.22}$$

Equation (9.22), the *dispersion relation* governing the propagation of the waves, is displayed as Figure 9.4. Here we plot the dimensionless variables $\omega/\omega_\circ$ and $k/k_\circ$, where $\omega_\circ \equiv (4\pi G\rho_\circ)^{1/2}$ and $k_\circ \equiv \omega_\circ/a_T$. For sufficiently short wavelengths (large k), $\omega \approx k a_T$. In this limit, the disturbance behaves like a sound wave, traveling at the phase velocity $\omega/k = a_T$. This is the usual isothermal sound speed associated with the background medium. However, both ω^2 and the phase velocity pass through zero when $k = k_\circ$. The corresponding wavelength $\lambda_J \equiv 2\pi/k_\circ$ is

$$\begin{aligned} \lambda_J \,&=\, \left(\frac{\pi\,a_T^2}{G\,\rho_\circ}\right)^{1/2} \\ &=\, 0.19\ \text{pc}\,\left(\frac{T}{10\ \text{K}}\right)^{1/2}\left(\frac{n_{\text{H}_2}}{10^4\ \text{cm}^{-3}}\right)^{-1/2} \ . \end{aligned} \tag{9.23}$$

Perturbations with wavelength exceeding this *Jeans length* have exponentially growing amplitudes. To compare with our previous discussion, the reader may verify that a uniform sphere of diameter λ_J contains about twice the mass given by equation (9.16), provided we identify $\rho_\circ$ as $P_\circ/a_T^2$. Indeed, when written in terms of density and sound speed (or temperature), M_{BE} is more commonly known as the *Jeans mass*, a nomenclature we shall also follow. Adopting the new symbol M_J, we recast equation (9.16) into the form

$$\begin{aligned} M_J \,&=\, \frac{m_1\,a_T^3}{\rho_\circ^{1/2}\,G^{3/2}} \\ &=\, 1.0\ M_\odot\,\left(\frac{T}{10\ \text{K}}\right)^{3/2}\left(\frac{n_{\text{H}_2}}{10^4\,\text{cm}^{-3}}\right)^{-1/2} \ . \end{aligned} \tag{9.24}$$

Our numerical versions of Equations (9.23) and (9.24) show that typical dense cores and Bok globules are close to the edge of gravitational instability. In fact, those with internal stars

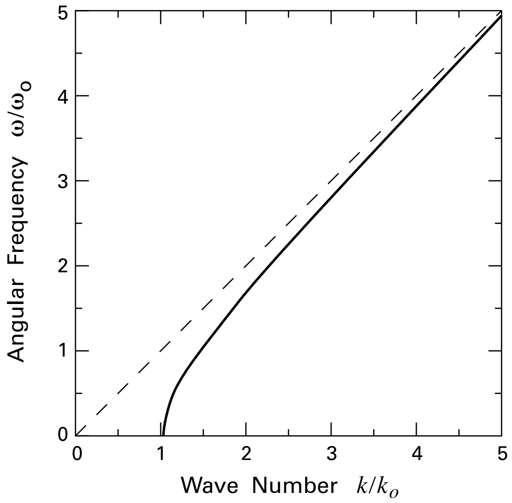


Figure 9.4 Dispersion relation (*solid curve*) for plane waves in a self-gravitating, isothermal gas. The dashed curve is the relation for sound waves.

have already crossed this threshold, and at least some of these ought to show signs of ongoing collapse. We shall revisit this observational issue in Chapter 11. Conversely, structures with measured masses substantially less than M_J or with sizes less than λ_J *could* be stable, given sufficient external pressure. Alternatively, they could be temporary configurations of positive total energy that will soon cool or disperse. Material passing through shocks provides an example of such transient structures. Thus, consider the patch of H_2 emission near IRc2 in Orion (§ 8.2). Using $T = 2000$ K and $n_{H_2} = 10^6$ cm^{-3}, we find $M_J = 280\,M_\odot$ and $\lambda_J = 0.41$ pc. In contrast, the observations show the emission arising from a sheet-like structure whose thickness is estimated at only 10^{13} cm, or 3×10^{-6} pc.

At the other extreme in both mass and size are giant molecular clouds. As we have seen, the complexes themselves appear to be self-gravitating swarms of more coherent clumps. Employing typical clump parameters of $n_{H_2} = 10^3$ cm^{-3} and $T = 10$ K in equation (9.24), we find that $M_J = 3\,M_\odot$, two orders of magnitude below the actual masses. To put the matter another way, the internal temperature of a 200 $M_\odot$ clump would have to be about 100 K for it to be supported entirely by thermal pressure. Since measured temperatures are much lower, at least throughout most of the interior, and since the clumps are apparently *not* undergoing global collapse, an extra source of support is necessary. The most plausible source is the interstellar magnetic field, whose dynamical effect we shall explore starting in § 9.4.

9.2 Rotating Configurations

As the next step toward a more complete theoretical description of molecular clouds, we allow for internal rotation. We saw in Chapter 3 that many dense cores have measured rotation rates, as determined from the relative Doppler shifts of the 1.27 cm line of NH_3 and other tracers.

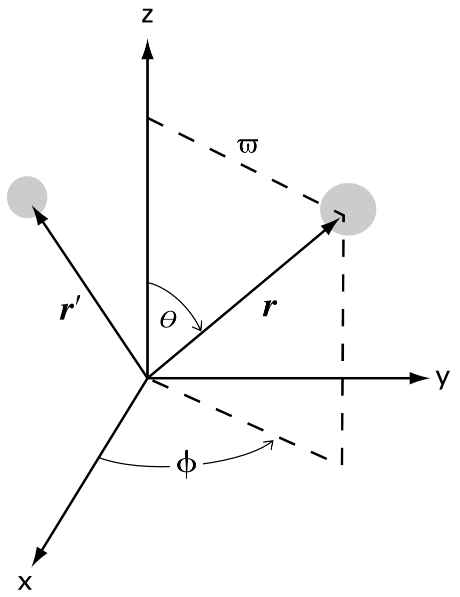


Figure 9.5 Coordinate system used for studying rotating clouds. The origin lies at the center of the cloud, and the z-axis points along the direction of rotation. Two fluid elements are shown – one at the point $\boldsymbol{r}$ of interest, and a second at some other point $\boldsymbol{r}'$ that contributes to the total gravitational potential at $\boldsymbol{r}$.

On the scale of giant molecular clouds, CO mapping at high spatial and spectral resolution frequently reveals a systematic gradient in radial velocity. The southernmost portion of Orion A, for example, is blueshifted with respect to the northern tip near the Trapezium. The inferred mean velocity gradient of 0.1 km s^{-1} pc^{-1} is twice the typical observed value in other complexes. In this and most other examples, the direction of this gradient coincides with the cloud's long axis, which itself lies roughly parallel to the Galactic plane. This fact suggests that the complexes spun up during condensation from the differentially rotating gas in the disk.

9.2.1 Poincaré-Wavre Theorem

How does rotation affect cloud morphology? In the preceding section, we were able to obtain spherical structures because the supporting thermal pressure is inherently isotropic. If we now consider the cloud to be rotating about a fixed axis, the associated centrifugal force on each fluid element points away from that axis, distending the equilibrium structure accordingly. It is natural, then, to erect a cylindrical coordinate system whose z-axis lies in the direction of rotation (see Figure 9.5). We suppose that each fluid element has a certain steady velocity u_ϕ about this axis, but has no motion in the z- or ϖ-directions. Here, $\varpi \equiv r \sin\theta$ denotes the cylindrical radius. We further suppose that the cloud is axisymmetric, so that all ϕ-gradients vanish, and that it has reflection symmetry about the $z = 0$ equatorial plane.

A physical quantity of key importance is $j \equiv \varpi u_\phi$, the specific angular momentum about the z-axis. Written in terms of j, the centrifugal force per unit mass on each element is j^2/ϖ^3. This term must be incorporated into equation (9.1) for hydrostatic equilibrium. If we again posit

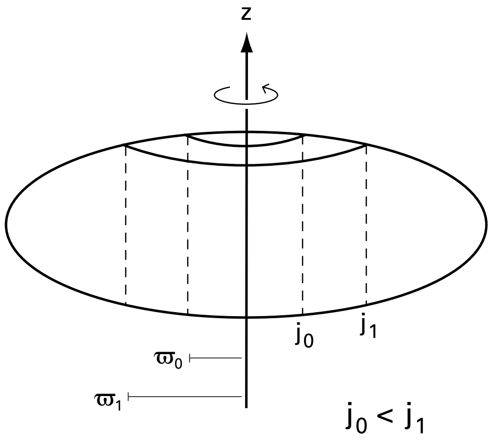


Figure 9.6 Illustration of Poincaré-Wavre theorem. The specific angular momentum j is constant along cylinders that are centered on the rotation axis. For rotational stability, the value of j must increase outward, as indicated.

an isothermal equation of state, then force balance in the ϖ- and z-directions requires

$$-\frac{a_T^2}{\rho}\frac{\partial\rho}{\partial\varpi} - \frac{\partial\Phi_g}{\partial\varpi} = -\frac{j^2}{\varpi^3} \tag{9.25a}$$

$$-\frac{a_T^2}{\rho}\frac{\partial\rho}{\partial z} - \frac{\partial\Phi_g}{\partial z} = 0 \ . \tag{9.25b}$$

An important consequence of these equations is that the specific angular momentum can only be a function of ϖ. That is, both j and u_ϕ must be constant along cylinders centered on the z-axis (Figure 9.6). To see why, we differentiate equation (9.25a) with respect to z, equation (9.25b) with respect to ϖ, and subtract the two. After canceling cross-derivatives, we find

$$\begin{aligned} -\frac{1}{\varpi^3}\frac{\partial j^2}{\partial z} &= \frac{a_T^2}{\rho^2}\left[\frac{\partial\rho}{\partial z}\frac{\partial\rho}{\partial\varpi} - \frac{\partial\rho}{\partial\varpi}\frac{\partial\rho}{\partial z}\right] \\ &= 0 \ , \end{aligned}$$

so that the assertion is proved. The reader may check that the same result would follow if the gas pressure were an *arbitrary* function of density. This more general statement constitutes the *Poincaré-Wavre theorem.*

The fact that j is independent of z allows us to reformulate force balance in a more concise manner. We define a *centrifugal potential* $\Phi_{\rm cen}$, also a function of ϖ alone:

$$\Phi_{\rm cen} \equiv -\int_0^\varpi \frac{j^2}{\varpi^3}\,d\varpi \ . \tag{9.26}$$

We may then express both parts of equation (9.25) as

$$-\frac{a_T^2}{\rho}\boldsymbol{\nabla}\rho - \boldsymbol{\nabla}\Phi_g - \boldsymbol{\nabla}\Phi_{\rm cen} = 0 \ . \tag{9.27}$$

Extending the non-rotating case, we see that $(\ln\rho + \Phi_g/a_T^2 + \Phi_{\rm cen}/a_T^2)$ is now a spatial constant. We thus write

$$\rho\,(\boldsymbol{r}) = \rho_c \exp\left[\frac{-\Phi_g - \Phi_{\rm cen} + \Phi_g(0)}{a_T^2}\right] \ . \tag{9.28}$$

Here, $\Phi_g(0)$ is the central value of the gravitational potential, which we no longer set to zero. Instead, we obtain Φ_g at any position $\boldsymbol{r}$ by integrating over all source points $\boldsymbol{r}'$ (Figure 9.5):

$$\Phi_g(\boldsymbol{r}) = -G \int \frac{\rho(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|}\, d^3\boldsymbol{x}' \,. \tag{9.29}$$

In writing equation (9.26), we have allowed the specific angular momentum j to be an arbitrary function of ϖ. However, clouds in which j decreases outward are *rotationally unstable*. To see the origin of this dynamical instability, we first note that the cloud's axisymmetry precludes any internal torques. Thus, if we imagine displacing a fluid element outward by a small amount, say from ϖ_0 to ϖ_1, j is a constant of the motion. The new centrifugal force, j^2/ϖ_1^3, is therefore *greater* than that acting on an unperturbed element already located at ϖ_1. The perturbed element feels a net force that impels it to move outward even more, rather than return to ϖ_0. Avoiding this instability and the rapid internal motions it generates requires that $j(\varpi)$ be monotonically increasing (see Figure 9.6).

9.2.2 Numerical Models

Beyond this general constraint, there is little guide to the proper form of $j(\varpi)$. Certainly, existing measurements of cloud rotational velocities are too crude to be of much help. This ignorance should not deter us from exploring the main characteristics of rotating equilibria. Suppose, for simplicity, that we imagine our cloud to have contracted from a larger, homogeneous sphere of the same mass M and initial radius $R_\circ$. Suppose further that this spherical cloud rotates with angular speed $\Omega_\circ$. Then, if M_ϖ is the mass contained within cylindrical radius ϖ, j is given by

$$j(M_\varpi) = \Omega_\circ\, R_\circ^2 \left[1 - \left(1 - \frac{M_\varpi}{M}\right)^{2/3}\right] \,, \tag{9.30}$$

within the uniform-density sphere. By angular momentum conservation, the same functional form must hold in the equilibrium state of interest. We fix $R_\circ$ by requiring that the spherical cloud have an internal pressure $P_\circ$ that matches the background value.

Construction of the equilibrium cloud now proceeds through the simultaneous solution of Equations (9.28) and (9.29), along with the auxiliary relations (9.26) and (9.30). The most widely used numerical approach is the *self-consistent field method*. One initially guesses a density distribution $\rho(\boldsymbol{r})$ and immediately obtains $\Phi_g(\boldsymbol{r})$ and $\Phi_{\rm cen}(\boldsymbol{r})$ from (9.29) and (9.26), respectively. Substituting these potentials into equation (9.28), one obtains a new density distribution which differs from the initial guess. This new density then becomes the source for updated potentials, which in turn generate yet another density distribution. Iterations continue until the density used to compute the potentials differs negligibly from that emerging from the hydrostatic balance equation (9.28).

As was the case for spherical clouds, it is best to formulate the problem in terms of nondimensional variables constructed from a_T, $P_\circ$, and G. Having done this, every equilibrium model falls into a continuous sequence, but one now characterized by *two* independent parameters. One parameter is again the density contrast from center to edge, $\rho_c/\rho_\circ = \rho_c\, a_T^2/P_\circ$. The second, traditionally denoted β, pertains to the degree of rotation. It is convenient to define β as

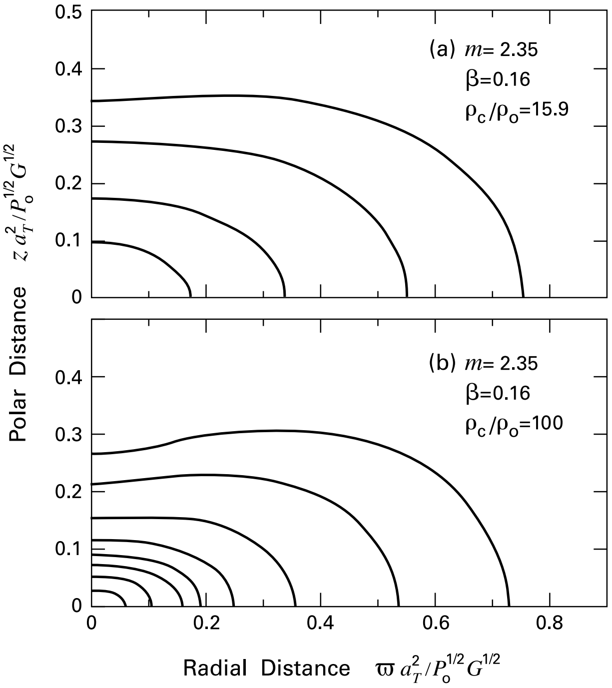


Figure 9.7 (a) Isodensity contours in one quadrant of a rotating cloud that is gravitationally stable. (b) Contours in an unstable cloud with the same nondimensional mass m and rotational parameter β.

the ratio of rotational kinetic to gravitational potential energy *in the spherical reference state*. Thus, we have

$$\beta \equiv \frac{\Omega_\circ^2\, R_\circ^3}{3\,G\,M} \;. \qquad (9.31)$$

The reader may verify that $\beta = 1/3$ corresponds to breakup speed in the spherical cloud. Hence β may range, at least in principle, from zero to that upper limit.

Figure 9.7a shows the isodensity contours for a model with a β-value of 0.16, *i. e.*, about half the maximum. The figure displays only one quadrant of the entire cloud, which is evidently flattened as a result of rotation about the z-axis. This particular model has a density contrast of 15.9 and a nondimensional mass m (as defined in equation (9.10)) of 2.35. Recall that spherical ($\beta = 0$) clouds are unstable to collapse for a density contrast in excess of 14.1. Nevertheless, the model in Figure 9.7a is actually stable, as we will show. Throughout the cloud volume, the main supporting force is the gradient of the internal pressure, while self-gravity supplies most of the confinement. In models of smaller m, gravity fades in importance, and the force balance is largely between the internal and external pressures. Such low-mass models have small density contrasts (as was also true for $\beta = 0$) and closely resemble *Maclaurin spheroids*, *i. e.*, configurations of uniform density in solid-body rotation.

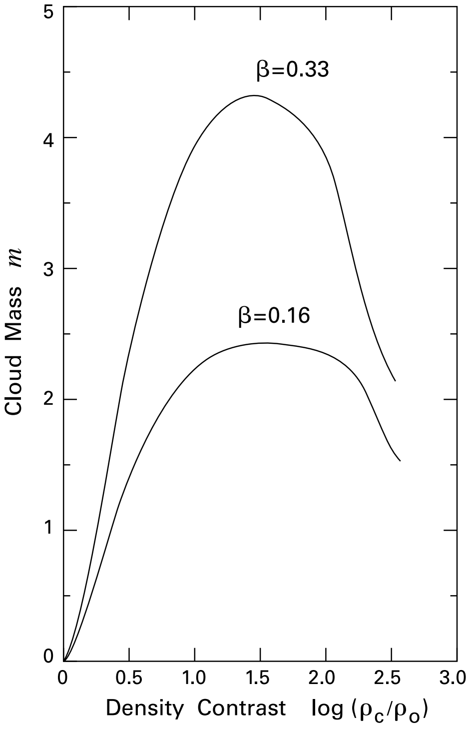


Figure 9.8 Nondimensional cloud mass as a function of density contrast in rotating, isothermal clouds. The two curves correspond to the indicated values of the rotational parameter β.

9.2.3 Susceptibility to Collapse

If we keep β fixed at 0.16 and consider models of increasing density contrast, the mass m first rises and then turns over (Figure 9.8). As before, the peak marks the onset of gravitational instability. To see why, we again select two models with the same m-value close to, but on either side of, the maximum. At fixed a_T and $P_\circ$, these clouds have the same dimensional mass. Hence, the spherical reference states from which they contracted had the same radius $R_\circ$ and, from equation (9.21), the same $\Omega_\circ$. It follows from equation (9.30) that their j-distributions are also identical. The two clouds are thus the extreme states in an oscillation of zero frequency. We see from Figure 9.8 that the model with $\rho/\rho_c = 34.0$ and $m = 2.42$ is just on the edge of instability. Hence the one pictured in Figure 9.7a is indeed stable, although not by much. Figure 9.7b shows the *unstable* model with the same β and m, but with the higher density contrast of $\rho_c/\rho_c = 100$. If we were to use this configuration as the initial state in the full dynamical equations, rather than just those for hydrostatic balance, the cloud would collapse within a few free-fall times.

Figure 9.8 also shows the mass variation for $\beta = 0.33$. It is apparent that all such curves exhibit a similar stability transition. A useful means of summarizing these results is to plot the peak mass, which we will denote m_{crit}, as a function of β (Figure 9.9). The quantity m_{crit}, a generalized, nondimensional Jeans mass, starts at the Bonnor-Ebert value $m_1 = 1.18$ for $\beta = 0$ and monotonically rises with β. This steady increase demonstrates quantitatively how rotation tends to stabilize a cloud against gravitational collapse. The most massive stable cloud, that

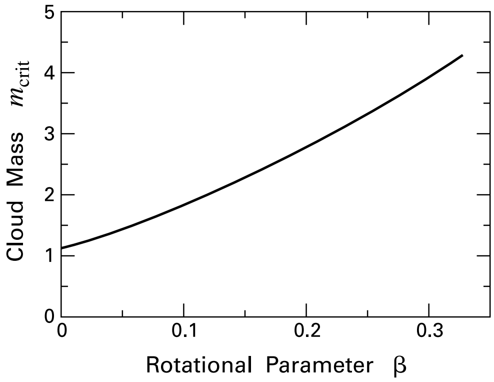


Figure 9.9 Critical, nondimensional cloud mass in rotating equilibria. The mass is shown as a function of the rotational parameter β.

with a β of 1/3, has a mass $m = 4.60$, a density contrast near 200, and an equatorial radius about three times the polar value.

Although we constructed clouds under the rather artificial assumption that they contracted from homogeneous, rigidly rotating spheres, the main characteristics of the models are not sensitive to the detailed form of $j(m_\varpi)$. In particular, significant flattening in the polar direction does not occur unless $\mathcal{T}_{\rm rot}/|\mathcal{W}|$, the ratio of rotational kinetic to gravitational potential energy, is greater than about 0.1. In our model sequence, $\mathcal{T}_{\rm rot}/|\mathcal{W}|$ is approximately equal to the parameter β, which measures the same ratio in the spherical reference state.

This link between cloud energetics and geometry shows unequivocally that the cloud elongations observed in molecular line maps do *not* arise principally from rotation. Consider first dense cores, which have an aspect ratio of about 0.6 (recall § 3.3). In equation (3.31), we estimated the typical $\mathcal{T}_{\rm rot}/|\mathcal{W}|$ as only 10^{-3}. Additionally, we adduced statistical evidence that the cores are actually *prolate* configurations. Turn now to giant molecular clouds. If their average velocity gradient of 0.05 km s^{-1} pc^{-1} represents solid-body rotation, then a 200 $M_\odot$ clump of diameter 2 pc also has $\mathcal{T}_{\rm rot}/|\mathcal{W}| \sim 10^{-3}$, according to equation (3.31). Once again, centrifugal force cannot account for the observed departures from spherical symmetry.

Negative results can be useful. Here, they lead to further questions that probe more deeply into star formation dynamics. For example, the low rotation rates of dense cores prompt us to ask if angular momentum is indeed conserved during their initial condensation, as our naïve model assumed. Unfortunately, neither the theoretical account of cloud formation nor empirical measurements of rotation are adequate to address this issue. We indicated earlier that larger clouds must be supported against collapse mostly by internal magnetic fields. These fields, we will see, exert torques that inhibit spinup during cloud contraction. Thus the observed Ω-values could actually be the result of rotational locking of dense cores to the surrounding gas.

9.3 Magnetic Flux Freezing

We have seen that neither their internal, thermal pressure nor centrifugal force are strong enough to support most clouds against collapse. We therefore turn our attention to the interstellar magnetic field. A promising indication that magnetic forces are generally adequate for the task comes from a simple virial theorem analysis. As we noted in Chapter 3, the typical fields in giant complexes are sufficiently strong that the terms $\mathcal{M}$ and $\mathcal{W}$ in equation (3.18) are comparable in magnitude. It is interesting that the same rough equality appears to hold in *all* clouds thought to be self-gravitating objects.

9.3.1 Observed Field Strengths

To elaborate this last point, we summarize in Table 9.1 the magnetic field measurements in a number of different environments, ordered by increasing gas density. Besides the standard cloud types covered in Chapter 3, entries include the dense clump of molecular gas sandwiched between the outflow lobes in S106 (Plate 2), a disk at the Galactic center (Sgr A West), and a powerful source of OH maser emission in Cygnus (W75 N). In all cases, it is the line-of-sight component $B_{\|}$ that is actually observed, through Zeeman splitting of either the 21 cm HI line or the hyperfine OH lines near 18 cm. With the exception of W75 N, all the measurements refer to lines seen in absorption against background continuum sources. When estimating the virial term $\mathcal{M}$ for each object, we may multiply $B_{\|}$ by a factor of two in order to obtain the median value of the total B for a randomly oriented field.

The general tendency, apparent from Table 9.1, for B to rise with density is an indication that the field lines in a cloud can be compressed along with the gas. We will shortly provide the theoretical basis for this observation. At the low-density end, the cloud in Ursa Major is an elongated object, some 3 pc in width, with a mean $n_{\rm HI}$ of only 15 cm^{-3}. Such a rarefied region is not expected to be self-gravitating. Indeed, careful integration across the cloud surface shows that $|\mathcal{W}|$ is about a factor of 60 lower than $\mathcal{M}$, which in turn is comparable to the kinetic energy term $\mathcal{T}$ in the virial theorem. The kinetic term, derived observationally from the broadening of molecular lines, represents disordered motion of the gas. The rough equality of $\mathcal{T}$ and $\mathcal{M}$ hints at the presence of MHD waves in the cloud interior.

As we consider denser objects, the terms $\mathcal{M}$, $\mathcal{T}$, and $|\mathcal{W}|$ all become comparable. In the central region of B1, to which the value in Table 9.1 refers, the observed line widths contributing to $\mathcal{T}$ are largely *thermal*. With its density of $n_{\rm H_2} = 2 \times 10^4$ cm^{-3} and diameter of 0.24 pc, this region resembles other dense cores, for which thermal and magnetic pressure both resist the strong self-gravity. The physical situation is less clear in the exotic environment of Sgr A West, part of a clumpy disk or ring of high density ($n_{\rm H_2} \sim 10^6$ cm^{-3}) and small radius (1.5 pc) that surrounds the radio source Sgr A*. Here $\mathcal{T}$ represents distinctly *non*thermal motion, including

Table 9.1 Zeeman Measurements of Magnetic Fields

Object	Type of Region	Diagnostic	$B_{\|}$(μG)
Ursa Major	Diffuse Cloud	HI	$+10$
L204	Dark Cloud	HI	$+4$
NGC 2024	GMC Clump	OH	$+87$
B1	Dense Core	OH	-27
S106	HII Region	OH	$+200$
Sgr A/West	Molecular Disk	HI	-3000
W75 N	Maser	OH	$+3000$

rapid streaming that could result from disruption of clouds by a central black hole. The molecular gas in W75 N is perhaps a factor of ten higher in density and is associated with a compact HII region containing the observed maser lines.

On energetic grounds, therefore, internal magnetic fields are important in counteracting gravity, wherever the latter actually provides the main cohesive force. In many environments, the magnetic force derives at least partially from dynamic waves in both the field and the gas. However, such waves may sometimes be subdued or even absent, as in the most quiescent dense cores. It is an important result of MHD theory that a purely static field, with no associated waves, can also support a self-gravitating cloud, although not for an indefinite period of time.

The physical basis for this mechanical support, as well as for the observed increase of B with density, is the phenomenon of *flux freezing*. Qualitatively, magnetic field lines behave as if they were tethered to the gas. Hence, any clumping of the latter brings together adjacent field lines, resulting in a larger B-value. Mathematically, the two magnetic terms in the force equation (3.6), representing field tension and the magnetic pressure gradient, both rise. In concert with an increased gradient in the thermal pressure, these forces can effectively oppose self-gravity.

9.3.2 MHD Equation

To see how flux freezing arises, we depart from equilibrium models *per se* to consider a cloud with internal motion. Any magnetic field threading this object is generated by a current density $\boldsymbol{j}$, related to $\boldsymbol{B}$ through Ampère's law:

$$\boldsymbol{\nabla} \times \boldsymbol{B} = \frac{4\pi}{c}\,\boldsymbol{j}\ . \tag{9.32}$$

Note that we have neglected the displacement current term on the right, $c^{-1}\partial\boldsymbol{E}/\partial t$, for the low frequencies of interest. The true current remaining in (9.32) is carried by the minor fraction of charged species in the cloud. These, as we saw in Chapter 8, include free electrons, ions, and grains. Although the grains are negatively charged, their number density is so low that they contribute negligibly to the current, which is therefore given by

$$\begin{aligned} \boldsymbol{j} &= n_i\, e\, \boldsymbol{u}_i - n_e\, e\, \boldsymbol{u}_e \\ &= n_e\, e\, (\boldsymbol{u}_i - \boldsymbol{u}_e)\ . \end{aligned} \tag{9.33}$$

Here, n_e and n_i are the electron and ion number densities, while $\boldsymbol{u}_i$ and $\boldsymbol{u}_e$ are the respective velocities. Equation (9.33) utilizes the fact that most ions are singly charged species (such as Fe^+ and Mg^+), so that charge neutrality demands equality of n_i and n_e.

We next invoke a phenomenological relationship describing how the current is actually generated by the local field. The familiar Ohm's law, stating that $\boldsymbol{j}$ is proportional to $\boldsymbol{E}$, applies only in a medium at rest. (Moving a copper wire through a magnetic field creates a current even with zero applied voltage.) To obtain the proper generalization, we temporarily shift to a reference frame moving at the local velocity of the neutral matter, $\boldsymbol{u}$, where we have dropped the subscript used in Chapter 8. Adding primes to all quantities in this frame, we have

$$\begin{aligned} \boldsymbol{j}' &= \sigma\, \boldsymbol{E}' \\ &= \boldsymbol{j}\ , \end{aligned} \tag{9.34}$$

where σ is the electrical conductivity. The second equality in (9.34) follows from equation (9.33), once we see that neither n_e nor the relative velocity $\boldsymbol{u}_i - \boldsymbol{u}_e$ can change in the new frame. Here we are neglecting relativistic corrections of order $(u/c)^2$. To the same accuracy, the new field $\boldsymbol{E}'$ is

$$\boldsymbol{E}' = \boldsymbol{E} + \frac{\boldsymbol{u}}{c} \times \boldsymbol{B} \, . \tag{9.35}$$

Equations (9.34) and (9.35) imply that the generalized Ohm's law *in the original reference frame* is

$$\boldsymbol{j} = \sigma \left(\boldsymbol{E} + \frac{\boldsymbol{u}}{c} \times \boldsymbol{B} \right) \, . \tag{9.36}$$

Using this result, we can now replace $\boldsymbol{j}$ in equation (9.32), obtaining a relation between the fields alone:

$$\boldsymbol{\nabla} \times \boldsymbol{B} = \frac{4\,\pi\,\sigma}{c} \left(\boldsymbol{E} + \frac{\boldsymbol{u}}{c} \times \boldsymbol{B} \right) \, . \tag{9.37}$$

Finally, we may eliminate $\boldsymbol{E}$ by using Faraday's law:

$$\boldsymbol{\nabla} \times \boldsymbol{E} = -\frac{1}{c}\frac{\partial \boldsymbol{B}}{\partial t} \, . \tag{9.38}$$

Multiplying (9.37) through by $c/4\pi\sigma$ and taking the curl of both sides, we combine the result with (9.38) to yield the fundamental *MHD equation* for the magnetic field:

$$\frac{\partial \boldsymbol{B}}{\partial t} = \boldsymbol{\nabla} \times (\boldsymbol{u} \times \boldsymbol{B}) - \boldsymbol{\nabla} \times \left(\frac{c^2}{4\pi\sigma} \boldsymbol{\nabla} \times \boldsymbol{B} \right) \, . \tag{9.39}$$

9.3.3 Estimating the Conductivity

The second righthand term in equation (9.39) represents the effect of Ohmic dissipation and vanishes as the conductivity becomes very large. In the present context, suppose L is the characteristic length over which $\boldsymbol{B}$ varies. This will be some appreciable fraction of the cloud diameter. We can neglect Ohmic dissipation– and obtain flux freezing– only if the quantity $\sigma L^2/c^2$ is much greater than any time scale of interest. We now show, by deriving an approximate expression for σ, that this is indeed the case for most molecular clouds.

Let us first consider in more detail the motion of the conduction electrons. The quantity $\boldsymbol{u}_e$ includes both the tight, helical gyrations about $\boldsymbol{B}$ and a much smoother drift component. The latter changes abruptly and stochastically due to collisions with other species, chiefly H_2 molecules. Such collisions, which occur only after many helical orbits, exert an effective drag on the electrons. This drag opposes the acceleration from the ambient electric field, establishing the steady-state drift velocity in a relatively brief time.

With this physical picture in mind, we return to our previous comoving reference frame and write an approximate equation of motion for the electrons, valid over time scales longer than that of individual collisions. Under steady-state conditions, the acceleration is small, and we have

$$0 = -e\,n_e\,(\boldsymbol{E}' + \frac{\boldsymbol{u}_e'}{c} \times \boldsymbol{B}') + n_e\,\boldsymbol{f}_{\rm en}' \, . \tag{9.40}$$

The first term represents the Lorentz force, while $\boldsymbol{f}'_{\rm en}$ in the second term is the drag exerted on each electron by the sea of neutrals. For simplicity, we have ignored the relatively rare electron-ion encounters. In any single collision with a neutral, the massive atom or molecule barely moves. If the electron scatters in a random direction, then it must, on average, transfer its full, original momentum to the neutral. That is, the electron gains the negative momentum $-m_e\,\boldsymbol{u}'_e$. In terms of the neutral density n, the number of collisions per unit time is $n\langle\sigma_{\rm en}u'_e\rangle$, where $\sigma_{\rm en}$ is the collision cross section, and where the angular braces denote an average over magnitude and direction in the thermal distribution of electron velocities. Note that the braces include $\sigma_{\rm en}$ because the cross section also depends on velocity.[4] We thus have

$$\boldsymbol{f}'_{\rm en} \,=\, -n\,m_e\,\langle\sigma_{\rm en}\,u'_e\rangle\,\boldsymbol{u}'_e\ . \tag{9.41}$$

Turning to the ions, we may write, after using $n_i = n_e$, an equation of motion analogous to (9.40):

$$0 \,=\, +e\,n_e\left(\boldsymbol{E}' + \frac{\boldsymbol{u}'_i}{c}\times\boldsymbol{B}'\right) + n_e\,\boldsymbol{f}'_{\rm in}\ . \tag{9.42}$$

It is now the ions themselves, with their average mass of $28\,m_H$, that are the heavier species in collisions with neutrals. In the *ion* rest frame, the momentum transfer must be the mass of a neutral, m_n, times its incoming velocity. In the present frame comoving with the neutrals, each collision thus imparts $-m_n\,\boldsymbol{u}'_i$, on average, to an ion. The drag force is therefore

$$\boldsymbol{f}'_{\rm in} \,=\, -n\,m_n\,\langle\sigma_{\rm in}\,u'_i\rangle\,\boldsymbol{u}'_i\ . \tag{9.43}$$

The components of $\boldsymbol{u}'_e$ and $\boldsymbol{u}'_i$ which concern us lie along the common direction of $\boldsymbol{E}'$ and the drag forces. (The other components do exist and give the conductivity a directional dependence that need not concern us.) Projecting (9.40) and (9.42) in this direction, the cross products involving $\boldsymbol{B}'$ both vanish. After applying (9.41) and (9.43), we generate two expressions relating $\boldsymbol{u}'_e$ and $\boldsymbol{u}'_i$, respectively, to $\boldsymbol{E}'$. Subtracting these and using (9.33) in our comoving frame, we find

$$\sigma \,=\, \frac{n_e\,e^2}{n}\left[\frac{1}{\langle\sigma_{\rm en}\,u'_e\rangle\,m_e} + \frac{1}{\langle\sigma_{\rm in}\,u'_i\rangle\,m_n}\right]\ . \tag{9.44}$$

Here the term $\langle\sigma_{\rm en}u'_e\rangle$ has the value 1.0×10^{-7} cm^3 s^{-1} in a molecular cloud of solar composition, while $\langle\sigma_{\rm in}u'_i\rangle$ is 1.5×10^{-9} cm^3 s^{-1}.

Returning to the dissipative term in equation (9.40), we can now evaluate $\sigma L^2/c^2$. Consider a clump within a giant molecular cloud, where $L \sim 1$ pc and $n \sim 10^3$ cm^{-3}. At this density, n_e/n is about 4×10^{-7}, according to equation (8.32). Despite the very low level of ionization, $\sigma L^2/c^2$ is of order 10^{17} yr. This time is enormous, of course, and will remain so even after a more refined treatment of σ. In the cloud environments of interest, therefore, Ohmic dissipation can certainly be neglected.

---

[4] Recall the discussion in § 5.1 of ion-molecule reactions, for which the cross section varies inversely with relative velocity.

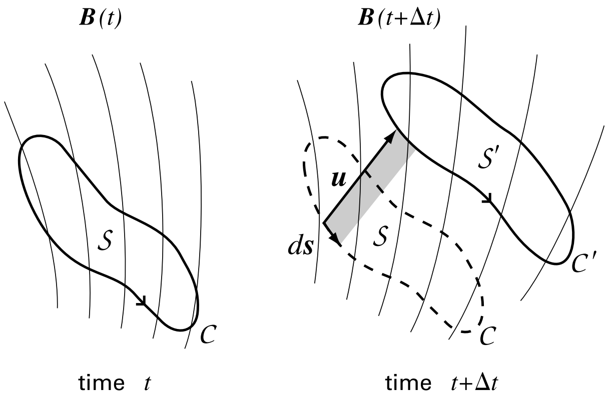


Figure 9.10 Proof of flux freezing. The closed loop $\mathcal{C}$ moves to become $\mathcal{C}'$ over the time interval Δt. Its enclosed surface $\mathcal{S}$ becomes $\mathcal{S}'$ during this time. Meanwhile, the small length $d\boldsymbol{s}$, moving at velocity $\boldsymbol{u}$, sweeps out the shaded area $d\boldsymbol{s} \times \boldsymbol{u}$. The area vector's direction coincides with the outward normal direction of the patch.

9.3.4 Field Transport in Ideal MHD

If we drop the second righthand term, equation (9.39) reduces to the *ideal MHD equation*:

$$\frac{\partial \boldsymbol{B}}{\partial t} = \boldsymbol{\nabla} \times (\boldsymbol{u} \times \boldsymbol{B}) \ . \tag{9.45}$$

This equation is the mathematical expression of flux freezing. To see why, imagine a closed loop $\mathcal{C}$ comoving with the gas (Figure 9.10). At time t, the magnetic flux Φ_B through this loop is given by a two-dimensional integral over an interior, comoving surface $\mathcal{S}$:

$$\Phi_B(t) = \int_{\mathcal{S}} \boldsymbol{B}(t) \cdot \boldsymbol{n}\, d^2\boldsymbol{x} \ ,$$

where $\boldsymbol{n}$ is a unit vector normal to the surface. At time $t + \Delta t$, both the loop and surface change, becoming $\mathcal{C}'$ and $\mathcal{S}'$, respectively. Our goal now is to demonstrate that $\Delta\Phi_B$, the change in flux through the moving surface, vanishes if equation (9.45) holds. Consider first the total flux, evaluated at $t + \Delta t$, through the *closed* surface consisting of $\mathcal{S}$, $\mathcal{S}'$, and the side walls generated by the loop's motion. Maxwell's equation $\boldsymbol{\nabla} \cdot \boldsymbol{B} = 0$, in conjunction with Gauss' theorem, implies that the flux through *any* closed surface vanishes. Referring to the figure, we have

$$\int_{\mathcal{S}} \boldsymbol{B}(t+\Delta t) \cdot \boldsymbol{n}\, d^2\boldsymbol{x} \,-\, \int_{\mathcal{S}'} \boldsymbol{B}(t+\Delta t) \cdot \boldsymbol{n}\, d^2\boldsymbol{x} \,-\, \Delta t \int_{\mathcal{C}} \boldsymbol{B}(t+\Delta t) \cdot (d\boldsymbol{s} \times \boldsymbol{u}) = 0 \ .$$

Note that the last term, a line integral around $\mathcal{C}$, represents the contribution from the side walls. Here, $|d\boldsymbol{s} \times \boldsymbol{u}|\, \Delta t$ is the elemental area indicated by the shading in Figure 9.10.

Supplied with this result, we may evaluate $\Delta\Phi_B$ as

$$\begin{aligned}\Delta\Phi_B &\equiv \int_{\mathcal{S}'} \boldsymbol{B}(t+\Delta t)\cdot\boldsymbol{n}\,d^2\boldsymbol{x} \;-\; \int_{\mathcal{S}} \boldsymbol{B}(t)\cdot\boldsymbol{n}\,d^2\boldsymbol{x} \\ &= \int_{\mathcal{S}} \boldsymbol{B}(t+\Delta t)\cdot\boldsymbol{n}\,d^2\boldsymbol{x} \;-\; \int_{\mathcal{S}} \boldsymbol{B}(t)\cdot\boldsymbol{n}\,d^2\boldsymbol{x} \;-\; \Delta t\int_{\mathcal{C}} \boldsymbol{B}(t+\Delta t)\cdot(d\boldsymbol{s}\times\boldsymbol{u}) \\ &\approx \Delta t\left[\int_{\mathcal{S}} \frac{\partial\boldsymbol{B}}{\partial t}\cdot\boldsymbol{n}\,d^2\boldsymbol{x} \;-\; \int_{\mathcal{C}} \boldsymbol{B}(t)\cdot(d\boldsymbol{s}\times\boldsymbol{u})\right]\;,\end{aligned}$$

where the approximation neglects errors of order Δt^2. If we now assume that equation (9.45) holds throughout the cloud, then we may invoke Stokes' theorem to transform the last line of the above equation. We thus obtain

$$\begin{aligned}\Delta\Phi_B &= \Delta t\left[\int_{\mathcal{C}} (\boldsymbol{u}\times\boldsymbol{B})\cdot d\boldsymbol{s} \;-\; \int_{\mathcal{C}} \boldsymbol{B}\cdot(d\boldsymbol{s}\times\boldsymbol{u})\right] \\ &= 0\;,\end{aligned}$$

where we have used invariance of the triple product under cyclic permutation.

Flux freezing signifies *both* that the magnetic field is tied to the motion of the fluid *and* that the gas itself is constrained by the field configuration. Which is the more apt description of a given cloud environment depends on the relative energy densities in magnetic field and fluid motion. Since the virial terms $\mathcal{M}$ and $\mathcal{T}$ are comparable in most clouds, the field and gas exert a strong mutual influence. Thus, the stratified appearance of many giant cloud complexes could well indicate alignment along the large-scale ambient field, as could the prolate shapes of dense cores. There is no doubt that flux freezing plays a key role in the structure and evolution of clouds on all scales.

It is equally certain, however, that the ideal MHD approximation must break down before stars are formed. Consider, for example, a 1 $M_\odot$ sphere within the central region of a dense core with $n_{\mathrm{H}_2} = 10^4$ cm^{-3}. Suppose this sphere, of radius $R_0 = 0.07$ pc, is threaded by a uniform magnetic field of strength $B_\circ = 30\,\mu$G. Suppose further that the sphere is destined to become a T Tauri star. Its radius will then decrease to $R_1 \approx 5\,R_\odot = 3\times 10^{11}$ cm. If flux freezing held during stellar formation, the product BR^2 would remain constant. Here we assume, for simplicity, that the final field is also uniform throughout the star's spherical volume. Thus, the stellar field would be $B_1 = 2\times 10^7$ G, a figure that exceeds by at least four orders of magnitude the observed values or upper limits in T Tauri stars (Chapter 17).

This simple example illustrates the *magnetic flux problem* in the theory of protostellar collapse, a puzzle which has not yet been completely solved. The difficulty really lies in the details, since it *is* clear that two well-studied processes can reduce the flux by a large amount. Prior to the onset of collapse itself, a dense core undergoes *ambipolar diffusion*. Here, electrons and ions remain tied to the field lines, while the dominant neutral species slip past and allow contraction to proceed. Once collapse is underway, entanglement of the field can lead to strong Ohmic dissipation through *magnetic reconnection*. In essence, the kinking of $\boldsymbol{B}$ reduces the length scale L in equation (9.45) until the dissipative term in the MHD equation (9.39) becomes substantial. We shall return to consider both these effects in Chapter 10.

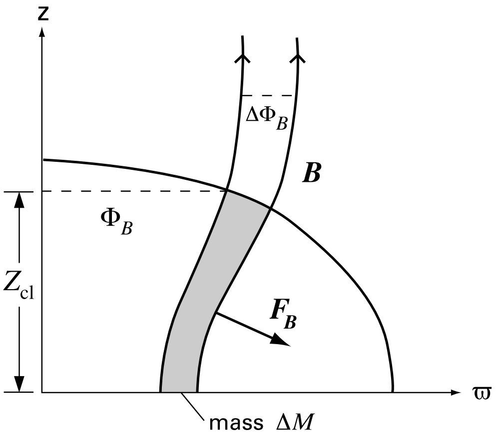


Figure 9.11 One quadrant of a magnetically supported cloud (*schematic*). The inner field line, when rotated about the axis, generates a surface that encloses flux Φ_B and penetrates the cloud surface a height $Z_{\rm cl}$ above the midplane. A neighboring line encloses an additional flux $\Delta\Phi_B$ and mass ΔM. The field curvature creates an outward magnetic force, as shown.

9.4 Magnetostatic Configurations

The freezing of field lines during cloud contraction means that the investigation of magnetic equilibria can proceed in a manner similar to that for rotating, nonmagnetized structures. Thus, the internal distribution of magnetic flux is inherited from an earlier, more rarefied environment, just as the run of specific angular momentum was previously. Suppose we now picture the cloud condensing from a state where the magnetic field $\boldsymbol{B}_\circ$ is everywhere straight and parallel. For convenience, we will employ a coordinate system whose origin is at the cloud center and whose z-axis lies in this direction. During contraction, matter can slide freely along field lines, until it is halted by the buildup of thermal pressure. Gas moving in the orthogonal direction, toward the axis, tugs on the field and is retarded by pressure (both thermal and magnetic), as well as by magnetic tension, *i. e.*, bending of the field lines. Thus, within the final equilibrium state, there is an additional outward force roughly similar to that arising from rotation (Figure 9.11).

In more detail, however, the magnetic-rotational analogy breaks down in a significant way. The centrifugal force in a rotating, equilibrium cloud points exactly in the ϖ-direction, implying that the specific angular momentum is constant along cylinders. If the analogy were exact, the equilibrium magnetic flux would also be constant within cylinders, which is *not* the case. As we have noted, field lines are curved by the inward tug of the gas. The new equation for force balance is

$$0 = -a_T^2\,\boldsymbol{\nabla}\rho - \rho\,\boldsymbol{\nabla}\Phi_g + \frac{\boldsymbol{j}}{c}\times\boldsymbol{B}\ , \tag{9.46}$$

where we again specify an isothermal equation of state. If the cloud does not rotate while contracting, the equilibrium $\boldsymbol{B}$-vector is poloidal, *i. e.*, it lies within the ϖ-z plane. Ampère's law then implies that the current $\boldsymbol{j}$ is toroidal, pointing in the ϕ-direction. It follows that the magnetic force per unit volume, $\boldsymbol{j}\times\boldsymbol{B}/c$, must be a poloidal vector with *both* ϖ- and z-components, unlike the centrifugal force. In summary, the fact that $\boldsymbol{B}$ can remain distorted in the equilibrium state precludes a magnetic analogue to the Poincaré-Wavre theorem and renders the actual calculation of equilibria more difficult.

9.4.1 Construction of Models

To proceed, we assume once more that our configuration has both reflection symmetry about the $z = 0$ plane and is axisymmetric about the central axis. Our first task is to change from the cumbersome vector variable $\boldsymbol{B}$ to a more convenient scalar quantity, the magnetic flux Φ_B. As illustrated in Figure 9.11, Φ_B is the flux contained within the surface generated by rotating any field line about the axis. The relation $\boldsymbol{\nabla} \cdot \boldsymbol{B} = 0$ allows us to recast $\boldsymbol{B}$ in terms of the magnetic potential $\boldsymbol{A}$:

$$\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A} \ .$$

Let $\mathcal{S}$ be any two-dimensional surface centered on the axis and bounded by a circle of radius ϖ. Then

$$\begin{aligned} \Phi_B &= \int_{\mathcal{S}} (\boldsymbol{\nabla} \times \boldsymbol{A}) \cdot d^2\boldsymbol{x} \\ &= 2\,\pi\,\varpi\,A_\phi \ . \end{aligned}$$

The vector field $\boldsymbol{A}$ need have *only* a nonzero ϕ-component in order to generate an arbitrary poloidal $\boldsymbol{B}$-field. Employing the unit vector $\hat{\boldsymbol{e}}_\phi$, we write

$$\begin{aligned} \boldsymbol{B} &= \boldsymbol{\nabla} \times (A_\phi\,\hat{\boldsymbol{e}}_\phi) && \text{(9.47a)} \\ &= -\frac{\hat{\boldsymbol{e}}_\phi}{\varpi} \times \boldsymbol{\nabla}\,(\varpi\,A_\phi) && \text{(9.47b)} \\ &= -\frac{\hat{\boldsymbol{e}}_\phi}{2\,\pi\,\varpi} \times \boldsymbol{\nabla}\Phi_B \ . && \text{(9.47c)} \end{aligned}$$

The reader may wish to check (9.47b) through component expansion. From Equation (9.47c), we confirm that $\boldsymbol{\nabla}\Phi_B \cdot \boldsymbol{B} = 0$, so that Φ_B cannot vary along a field line.

Our next task is to construct the analogue of equation (9.28), *i. e.*, to use the force balance equation (9.46) to find a relation between ρ and Φ_g. Since the magnetic force cannot be expressed simply as the gradient of a potential, a single, global relation does not exist. However, hydrostatic balance along field lines implies that the product $\rho\,\exp\left(\Phi_g/a_T^2\right)$ is conserved in that special direction. The actual value of this quantity must change from one field line to the next. To quantify the change, we define a new scalar q, having the dimensions of a pressure:

$$q \equiv \rho\,a_T^2 \exp\left(\Phi_g/a_T^2\right) \ . \qquad (9.48)$$

Recalling that $\boldsymbol{j}$ has only a ϕ-component, we then use equation (9.47) to rewrite (9.46) as

$$\exp\left(-\Phi_g/a_T^2\right) \boldsymbol{\nabla} q = \frac{j_\phi}{2\,\pi\,\varpi\,c}\,\boldsymbol{\nabla}\Phi_B \ .$$

Dotting both sides of this equation with $\boldsymbol{B}$, we find that $\boldsymbol{B} \cdot \boldsymbol{\nabla} q = 0$. Thus q is constant along field lines and is a function of Φ_B alone. The equation of force balance reduces to the scalar relation

$$\frac{j_\phi}{2\,\pi\,\varpi\,c} = \frac{dq}{d\Phi_B} \exp\left(-\Phi_g/a_T^2\right) \ . \qquad (9.49)$$

An additional relation between j_ϕ and Φ_B comes from Ampère's law, equation (9.32). Expanding the curl of equation (9.47c) and utilizing (9.49), we arrive at a field equation for Φ_B:

$$\frac{\partial}{\partial \varpi}\left(\frac{1}{\varpi}\,\frac{\partial \Phi_B}{\partial \varpi}\right) + \frac{1}{\varpi}\,\frac{\partial^2 \Phi_B}{\partial z^2} \;=\; -32\,\pi^3\,\varpi\,\frac{dq}{d\Phi_B}\exp\left(-\Phi_g/a_T^2\right) \ . \tag{9.50}$$

We now seek the analogous relation governing the gravitational potential Φ_g. This quantity is, of course, generated from the mass distribution through Poisson's equation. Replacing the density ρ in equation (9.3) by q through equation (9.48), we have

$$\frac{1}{\varpi}\frac{\partial}{\partial \varpi}\left(\varpi\,\frac{\partial \Phi_g}{\partial \varpi}\right) + \frac{\partial^2 \Phi_g}{\partial z^2} \;=\; \frac{4\,\pi\,G}{a_T^2}\,q\,\exp\left(-\Phi_g/a_T^2\right) \ . \tag{9.51}$$

Equations (9.50) and (9.51) are two of the fundamental relations of the problem, to be solved for Φ_B and Φ_g. Doing so requires that we specify the function $q(\Phi_B)$, which entails knowledge of the magnetic flux distribution. We also need a description of the medium surrounding the cloud. Here the common practice, as before, is to imagine the cloud immersed in a hypothetical gas of zero density and finite pressure $P_\circ$. We again require the surface pressure of our configuration to match $P_\circ$. Since the sound speed is infinite in the external medium, equation (9.48) implies that q becomes the constant $P_\circ$. From equation (9.49), the current j_ϕ therefore vanishes. Thus, we continue to solve Equations (9.50) and (9.51) outside the cloud, but replace both right-hand sides by zero. At some appropriately large outer boundary, we impose the conditions that the magnetic field have its original value $\boldsymbol{B}_\circ$ and that Φ_g be that due to the total cloud mass.

Our next task is to derive an expression for $q(\Phi_B)$ in terms of the flux distribution. We first let $M(\Phi_B)$ denote the cloud mass contained within the field line associated with flux Φ_B. From Figure 9.11, the small amount of mass ΔM between Φ_B and $\Phi_B + \Delta\Phi_B$ is

$$\Delta M \;=\; 2\int_0^{Z_{\rm cl}(\Phi_B)} dz \int_{\varpi(Z_{\rm cl},\Phi_B)}^{\varpi(Z_{\rm cl},\Phi_B+\Delta\Phi_B)} d\varpi\,2\,\pi\,\varpi\,\rho \ ,$$

where $Z_{\rm cl}\,(\Phi_B)$ is the cloud boundary. The second integration is trivial once we replace the independent variable ϖ by Φ_B. After further substituting for ρ in terms of q, we obtain an integral expression for the latter:

$$q \;=\; \frac{a_T^2}{4\,\pi}\,\frac{dM}{d\Phi_B}\left[\int_0^{Z_{\rm cl}(\Phi_B)} dz\,\varpi\,\frac{\partial \varpi}{\partial \Phi_B}\,\exp\left(-\Phi_g/a_T^2\right)\right]^{-1} \ . \tag{9.52}$$

Once this expression has been evaluated, the term $dq/d\Phi_B$ in equation (9.50) follows by numerical differentiation. Note that, since ϖ in the integrand of (9.52) is a function of both z and Φ_B, the function $q(\Phi_B)$ depends on the detailed spatial configuration of the field. Thus, one must solve (9.50), (9.51), and (9.52) simultaneously, through an iterative procedure analogous to the self-consistent field method.

9.4.2 Flattened Equilibria

It remains only to select the flux distribution, $dM/d\Phi_B$. In principle, one could utilize observational results on individual clouds and their embedded fields, but the existing data are much too sparse for this purpose. As an alternative, we resort to an admittedly simplified physical model and study the dependence of the equilibrium configurations on the model parameters. Proceeding as in the rotational problem, we imagine the cloud to have contracted from a uniform-density sphere threaded by the background field $\boldsymbol{B}_\circ$. If ρ_i and $R_\circ$ are the density and radius of this sphere, respectively, then

$$\begin{aligned} \frac{dM}{d\Phi_B} &= \frac{2\,\rho_i\,R_\circ}{B_\circ}\left(1\,-\,\frac{\Phi_B}{\Phi_{\rm cl}}\right)^{1/2} && \Phi_B \le \Phi_{\rm cl} \\ &= 0 && \Phi_B > \Phi_{\rm cl}\ , \end{aligned} \tag{9.53}$$

where $\Phi_{\rm cl} \equiv \pi B_\circ R_\circ^2$ is the total flux threading the cloud. Assuming that flux freezing holds during contraction of the reference sphere, equation (9.53) also gives the function $dM/d\Phi_B$ to be used in equation (9.52) for $q(\Phi_B)$.

It is convenient, for numerical solution, to recast all variables into nondimensional form, using the four basic quantities a_T, $P_\circ$, G, and $B_\circ$. If the reference sphere had an internal pressure equal to $P_\circ$, the problem would have two free parameters, just as in the rotational case. One is the density contrast $\rho_c/\rho_\circ$, where the cloud surface density is again $P_\circ/a_T^2$. A second parameter, the equivalent of the rotational β, is the ratio of magnetic to thermal pressure in the background medium:

$$\alpha \,\equiv\, \frac{B_\circ^2}{8\,\pi\,P_\circ}\ . \tag{9.54}$$

In practice, it is customary to let the reference sphere's internal pressure vary as well. Thus, we require a third free parameter, which could be $\rho_i/\rho_\circ$, but is generally taken to be the nondimensional radius of the initial sphere:

$$\xi_\circ \,\equiv\, \left(\frac{4\,\pi\,G\,\rho_\circ}{a_T^2}\right)^{1/2} R_\circ\ . \tag{9.55}$$

Note the difference in this definition from equation (9.6), which employed the *central* density instead.

Figure 9.12a displays the isodensity contours and field lines for a representative model, with parameters $\xi_\circ = 2.4$, $\alpha = 1.0$, and $\rho_c/\rho_\circ = 10$. Note that the unit of distance here is $R_\circ$. The flattening of the cloud in the polar direction is evidently quite similar to that obtained through rotation. In the present example, the equatorial radius is still 90 percent of its value in the parent sphere, while the polar radius has shrunk by some 40 percent. If we consider a model of much greater density contrast ($\rho_c/\rho_\circ = 10^3$; Figure 9.12b) but identical values of $\xi_\circ$ and α, the equatorial radius shrinks a bit more, while the central region contracts by a large amount, pulling the field with it. The polar radius decreases as well, giving the whole structure a concave appearance. This central dip, found also in rotating clouds, results from the higher gravity along the pole, where the force is unopposed by magnetic tension.

Flatter clouds also result if we raise either α or $\xi_\circ$. In the first case, we have increased the ambient magnetic pressure relative to the thermal value. With a higher retarding force in the horizontal direction, the cloud needs more mass to achieve a given central density. The resulting increase of gravity draws in the polar zone. In the second case, the larger parent cloud contains

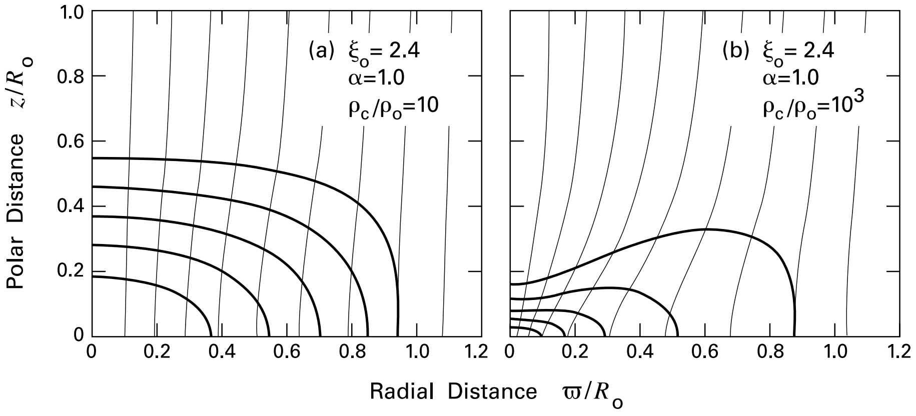


Figure 9.12 (a) Isodensity contours (*thick solid curves*) for one quadrant of a gravitationally stable, magnetized cloud. (b) Contours for an unstable cloud, with identical $\xi_\circ$ and α, but higher density contrast. The thin, solid curves are magnetic field lines.

more magnetic flux, again leading to a higher mass in the equilibrium state. Thus, a cloud with $\xi_\circ = 2.4$, $\alpha = 50$, and $\rho_c/\rho_\circ = 10$ has a ratio of equatorial to polar radii of 2.0, while this ratio is 4.5 for $\xi_\circ = 4.8$, $\alpha = 1.0$, and the same density contrast.

9.4.3 Critical Mass and Surface Density

Given a set of calculated models, we are in a position to examine the run of nondimensional cloud mass m versus density contrast, this time at fixed $\xi_\circ$ and α. Figure 9.13 shows three such mass curves. In all cases, m first rises monotonically from zero, when $\rho_c/\rho_\circ = 1$, then reaches a maximum before declining at greater density contrast. The turnover indicates, as before, that the fundamental mode of oscillation is becoming dynamically unstable, *i. e.*, it reveals the onset of gravitational collapse. The actual value of the peak mass, which we again denote as $m_{\rm crit}$, is higher with a stronger internal field, since the latter helps stave off collapse. Thus, in Figure 9.13, we can see that $m_{\rm crit}$ goes from 2.0 at $\xi_\circ = 2.4$ and $\alpha = 1.0$, to 5.1 at $\xi_\circ = 4.8$ and $\alpha = 1.0$, and reaches 8.9 for $\xi_\circ = 4.8$ and $\alpha = 5.0$. These results prove that the model depicted in Figure 9.12a is gravitationally stable, since its density contrast is lower than that for $m_{\rm crit}$ at the appropriate $\xi_\circ$ and α. Conversely, the model in Figure 9.12b is unstable. A good fit to the numerical values of $m_{\rm crit}$ over a wide range of $\xi_\circ$– and α–values is

$$m_{\rm crit} \approx 1.2 + 0.15\, \alpha^{1/2}\, \xi_\circ^2 \, . \qquad (9.56)$$

The physical meaning of equation (9.56) becomes apparent once we convert it back into dimensional form. Multiplying through by $a_T^4/P_\circ^{1/2} G^{3/2}$, we recognize the first term as the Bonnor-Ebert mass from equation (9.16). Recall that this is the marginally stable value in the absence of any magnetic field, *i. e.*, with $\alpha = 0$. Thus, we may rewrite equation (9.56) more suggestively as

$$M_{\rm crit} \approx M_{\rm BE} + M_\Phi \, . \qquad (9.57)$$

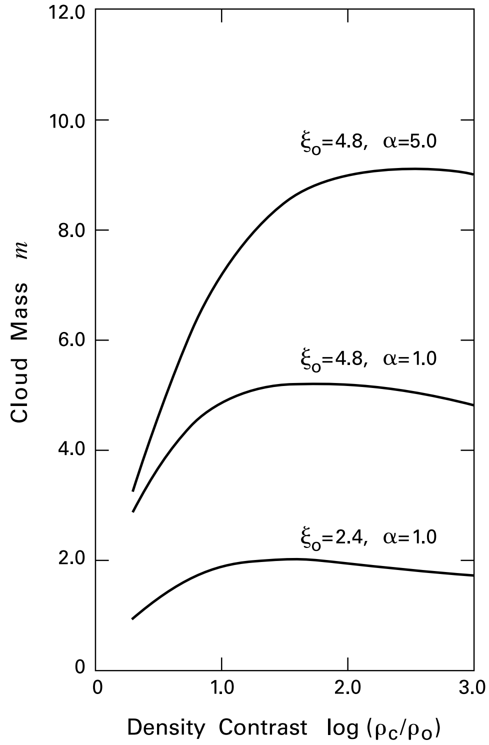


Figure 9.13 Critical, nondimensional mass in magnetized equilibria. The mass is plotted as a function of the density contrast, for the three indicated combinations of $\xi_\circ$ and α.

Here, the second term is

$$M_\Phi = 0.15\,\frac{\alpha^{1/2}\,\xi_\circ^2\,a_T^4}{P_\circ^{1/2}\,G^{3/2}} \tag{9.58a}$$

$$= 0.15\,\frac{2}{\sqrt{2\pi}}\,\frac{B_\circ\,\pi\,R_\circ^2}{G^{1/2}} \tag{9.58b}$$

$$= 0.12\,\frac{\Phi_{\rm cl}}{G^{1/2}}\ . \tag{9.58c}$$

The quantity M_Φ represents the critical mass for a cloud so cold that its thermal energy is negligible compared to its gravitational and magnetic contributions. We have derived the important fact that this mass is simply proportional to the total magnetic flux threading the cloud.

The gravitational stability of magnetized clouds differs qualitatively from the purely thermal case covered in § 9.1. Consider slowly squeezing an initially stable configuration of mass M by increasing $P_\circ$. In the absence of magnetic fields and rotation, the spherical cloud will naturally shrink. Assuming the temperature remains fixed, equation (9.16) implies that the critical mass $M_{\rm BE}$ will decline. This decline will continue until $M_{\rm BE} = M$, at which point the cloud will collapse. Now imagine the same process for a stable cloud supported additionally by a magnetic field. Any increase in $P_\circ$ will also lead to a smaller configuration. In equation (9.57) for $M_{\rm crit}$, the term $M_{\rm BE}$ still falls, but M_Φ remains constant, as long as flux freezing applies. Thus, if the cloud has such low mass that $M < M_\Phi$ initially, it will always remain stable and cannot be driven into collapse.

Any magnetized cloud of sufficiently low mass has a unique equilibrium configuration and is always gravitationally stable. Here, we are assuming a fixed internal flux distribution, such as that given by equation (9.53). Figure 9.13 shows that, for given values of α and $\xi_\circ$, two equilibria may share the same mass but have different central densities. It is only for masses below some transition value (dependent on α and $\xi_\circ$) that there is a unique state. In dimensional terms, the transition mass is given approximately by $0.59\, M_{\rm BE} + M_\Phi$. Here, the first term is the dimensional version of m_2, the first minimum in the thermal mass curve of Figure 9.2. For masses above the transition, the state of higher ρ_c is always gravitationally unstable. Finally, clouds with $M > M_{\rm crit}$ have *no* equilibria and are collapsing from the start.

It is worth remarking that the numerical coefficients 0.15 and 0.12 in Equations (9.56) and (9.58) refer to the rather arbitrary flux distribution posited in equation (9.53). In models constructed with other forms for $dM/d\Phi_B$, these numbers can increase or decrease significantly. However, for $M_{\rm BE} \ll M_\Phi$, all equilibrium configurations resemble flat slabs, at least in their densest, inner regions. In this limit, the mass range for which two equilibria are possible is relatively small, and the criterion for stability becomes simply $M < M_\Phi$. Given the slab-like geometry, it makes more sense to recast the stability criterion in terms of Σ_c, the cloud's surface density along the axis and B_c, the magnetic field value at the center. Detailed numerical calculations show, in fact, that these highly magnetized clouds are stable as long as

$$\frac{\Sigma_c}{B_c} < 0.17\, G^{-1/2} \ . \qquad \textit{flat–slab limit} \tag{9.59}$$

The numerical coefficient is now independent of the assumed flux distribution.

The criterion in equation (9.59) is so simple that we should try to understand it in a more direct, physically appealing manner. Figure 9.14 depicts a slab of surface density $\Sigma_\circ$, pierced by a uniform, perpendicular field $\boldsymbol{B}_\circ$. Consider now a disk of radius $\varpi_\circ$, located within the slab; this disk has mass $\pi \Sigma_\circ \varpi_\circ^2$. If its radius contracts by the fractional amount ϵ, then the additional inward gravitational force per unit mass at the disk edge is $F_G \approx 2\pi\epsilon G \Sigma_\circ \varpi_\circ^2/\varpi_\circ^2 \approx 2\pi\epsilon G\Sigma_\circ$. This force is opposed by the magnetic tension created by pulling in the initially straight field lines. The associated outward force per unit mass is $F_B = \mathcal{J} B_z / c\Sigma_\circ$, where the surface current $\mathcal{J}$ is the usual volume current density j integrated over the vertical thickness. Note that this current, which arises from the bending of the field lines, flows azimuthally within the disk. Integrating Ampère's law, equation (9.32), over the disk height, we find $\mathcal{J} = c|B_\varpi|/2\pi$, where B_ϖ is the small, radial field component generated by the contraction. As seen in Figure 9.14, B_ϖ actually reverses sign across the disk; our expression for $\mathcal{J}$ refers to the value just above the surface. To within a factor of order unity, we thus have $F_B \approx |B_\varpi| B_z / 2\pi\Sigma_\circ$.

The task now is to evaluate B_ϖ. Assuming a very low material density outside the slab, the Lorentz force associated with the magnetic field must vanish, implying also that the current is zero. That is, $\boldsymbol{\nabla} \times \boldsymbol{B} = 0$ above and below the slab. Thus we have $|\partial B_\varpi/\partial z| = |\partial B_z/\partial \varpi|$. Let Δz be the vertical distance above the disk over which the field relaxes back to its original,

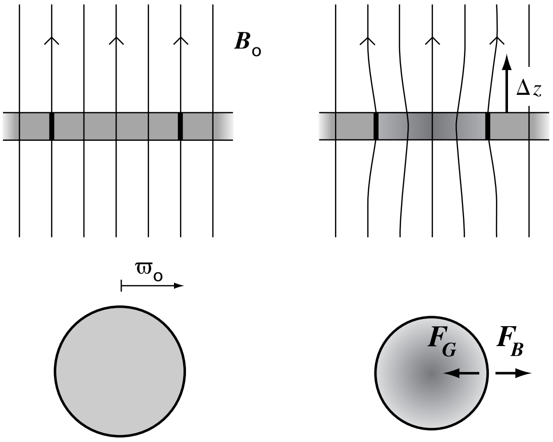


Figure 9.14 Gravitational stability of a cold slab embedded in a uniform magnetic field $\boldsymbol{B}_\circ$. The lateral compression of a small disk, with initial radius $\varpi_\circ$, generates extra gravitational and magnetic forces, as indicated. The external field relaxes to its uniform, vertical configuration by a height Δz above the slab.

straight configuration. As a result of flux freezing, B_z at the disk center increases by about $2\epsilon B_\circ$. We therefore have $B_\varpi \approx 2\epsilon B_\circ(\Delta z/\varpi_\circ)$. On the other hand, $\boldsymbol{\nabla}\cdot\boldsymbol{B} = 0$ implies that $|\partial B_\varpi/\partial\varpi| = |\partial B_z/\partial z|$. Hence we also have $B_\varpi \approx 2\epsilon B_\circ(\varpi_\circ/\Delta z)$. Combining these statements, we find that $\Delta z \approx \varpi_\circ$ and $B_\varpi \approx 2\epsilon B_\circ$, so that $F_B \approx \epsilon B_\circ^2/\pi\Sigma_\circ$. For the magnetic tension to exceed gravity, we need $F_B > F_G$, or $\Sigma_\circ^2/B_\circ^2 < 1/(2\pi^2 G)$. This result reproduces, to within a factor of two, the criterion of equation (9.59).

9.4.4 Comparison with Observations

How do the theoretical models we have presented compare with actual molecular clouds? First, let us revisit the basic problem of gravitational stability, introduced at the end of § 9.1. There, we noted that the larger cloud fragments, such as the clumps within giant complexes, have observed masses that greatly exceed the critical value $M_{\rm BE}$ (or M_J). On the other hand, numerical evaluation of equation (9.58c) tells us that the maximum stable mass for a cold, magnetized cloud is

$$M_\Phi = 70\,M_\odot \left(\frac{B}{10\ \mu{\rm G}}\right)\left(\frac{R}{1\ {\rm pc}}\right)^2 , \qquad (9.60)$$

where B is here the mean field threading the cloud of radius R. Since the critical mass for a magnetized cloud with finite temperature is approximately $M_{\rm BE} + M_\Phi$, we see that the basic problem has been largely resolved (recall Table 3.1). That is, most dark clouds *can* plausibly be stabilized against collapse by their internal magnetic fields alone. Dense cores, on the other hand, have $M \approx M_\Phi \approx M_{\rm BE}$. These statements are consistent with our previous discussion based on the virial theorem. The reader may verify that the empirical relation $\mathcal{M} \approx |\mathcal{W}|$ implies that cloud masses over a substantial range do not fall much below M_Φ.

Unfortunately, more detailed comparison between theory and observation is less encouraging. As we have noted, the spectral line widths of larger clouds all indicate the presence of MHD waves, in contrast to the smooth, static fields assumed in the models. With this significant caveat in mind, we first ask what the theory predicts for clouds with $M_{\rm BE} \ll M \lesssim M_\Phi$. By substituting for $\rho_\circ$ and $R_\circ$ in terms of $P_\circ$, a_T, and M, we may rewrite the nondimensional radius $\xi_\circ$ as $(M/M_{\rm BE})^{2/3}$ times a numerical coefficient, equal to 2.3. Similar manipulation allows us to recast the parameter α in equation (9.54) as $2.1\,(M/M_{\rm BE})^{2/3}\,(M/M_\Phi)^{-2}$. Thus, a marginally stable dark cloud with $M/M_{\rm BE} \approx 10$ would have $\xi_\circ \approx 5$ and $\alpha \approx 10$. A more stable cloud would have higher α. From our description of the numerical results, such a model would have an equatorial radius *at least* 5 times its polar value. Only a small fraction of dark clouds exhibit such extreme aspect ratios. Moreover, there is no convincing evidence that these are indeed flattened structures seen edge on, as opposed to elongated spindles.

Turning to dense cores, the near equality of M, $M_{\rm BE}$, and M_Φ implies that the appropriate values of $\xi_\circ$ and α are closer to unity. In addition, the observed line widths are nearly thermal, so that a static magnetic field is more plausible. Thus, the model depicted in Figure 9.12a, for which $M/M_{\rm BE} = 1.1$ and $M/M_\Phi = 1.5$, should serve as a representative choice. The model density contrast of $\rho_c/\rho_\circ = 10$ is in accord with observations, and the equatorial-to-polar aspect ratio of 1.6 is also within the range of those found from maps in NH_3 and other high-density tracers. Unfortunately, the statistical argument presented in § 3.3 shows that the cores are more likely to be prolate configurations.

Assuming that future studies confirm this finding, we are faced with a direct contradiction between theory and observation. Clearly, the weakest element in the theoretical framework is the flux distribution $dM/d\Phi_B$. The specific form adopted, equation (9.53), was obtained from a picture in which the cloud of interest contracted from a larger, spherical parent, itself embedded in a zero-density medium of finite pressure. No such medium exists, of course. More realistically, the present dense cores grew in both mass and size as they accreted surrounding material of only slightly lower density. To obtain prolate configurations, the function $dM/d\Phi_B$ needs to be more strongly peaked near $\Phi_B = 0$ than in equation (9.53). Whether a revised picture of core formation can naturally yield such a result is still an open question. We will return to this issue in the next chapter, when we discuss cloud evolution through ambipolar diffusion.

9.5 Support from MHD Waves

Optical and near-infrared polarization studies indicate that cloud magnetic fields are well ordered but have significant local deviations. Thus the magnetostatic structures we have considered, while useful as first approximations, need refinement. The observed substructure in $\boldsymbol{B}$ could be a manifestation of internal waves. Because of flux freezing, any impulsive disturbance to a fluid element is transmitted to its embedded magnetic field. Such kinking of the field is resisted by magnetic tension. In acting to remove the local distortion in $\boldsymbol{B}$, tension also causes the disturbance to propagate *along* the field line, much as elastic tension sends traveling waves down a plucked string. These MHD waves have long been detected, through spacecraft observations, in the solar wind. They can also be seen propagating away from sunspots and large flares. Indeed, the entire solar photosphere is in a continual state of turbulent motion as a result of subsurface convection. This motion generates MHD waves that may aid in heating the outer gas to temperatures in excess of 10^6 K. Similar activity is at least partially responsible for the strong optical, ultraviolet, and X-ray emission from T Tauri stars (Chapter 17).

Within molecular clouds, the gas motion associated with MHD waves is not seen directly, but may well underly the super-thermal broadening in spectral lines of CO and other molecular tracers. As we will see, the inferred amplitudes are large enough that the waves could, at least in principle, provide substantial support against gravitational collapse. The fundamental question of how these disturbances ultimately arise is still unanswered. We mentioned in § 3.3 that individual dense cores containing stars, while largely quiescent, have greater line widths than their starless counterparts. It is unlikely, however, that stellar outflows generate the much larger-scale turbulence in cloud complexes. An important counterexample is Maddalena's Cloud. This complex, at a distance of some 2 kpc, has spawned no massive stars, while its infrared luminosity limits star formation quite generally to a relatively low level. Nevertheless, CO observations show high internal velocities consistent with other clouds of its size and mass. Other tracers reveal a rich substructure, including both massive clumps and embedded dense cores.

Our understanding of how wave-like motions help to counteract the self-gravity of a large cloud is still schematic. Quantifying this idea will probably require such basic data as the wavelength dependence of the internal energy. Such information is not yet accessible to observation. It is hardly surprising, then, that there are no detailed cloud models incorporating wave support. We therefore limit ourselves in this section to a derivation of the essential properties of the waves and to a general description of their dynamical effect. We shall briefly revisit the issue of wave generation at the end.

9.5.1 Perturbation Analysis

We begin by considering small perturbations of a hypothetical, uniform gas, much as we did in § 9.1 when deriving the Jeans criterion. The background medium now contains a uniform magnetic field $\boldsymbol{B}_\circ$. We also take the wavelength of the perturbation to be much smaller than λ_J, so that we can ignore self-gravity in the waves themselves. Finally, we assume that the *period* of the disturbance is long enough that the gas remains isothermal. [5] Imagine, then, a traveling-wave perturbation characterized by the vector wave number $\boldsymbol{k}$, which points in the direction of propagation. If we take this direction to be arbitrary, then the generalization of equation (9.19) for the perturbed density is

$$\rho(\boldsymbol{r}, t) \,=\, \rho_\circ \,+\, \delta\rho \exp[i(\boldsymbol{k}\cdot\boldsymbol{r} \,-\, \omega t)] \ . \tag{9.61}$$

We adopt similar forms for the perturbed velocity $\boldsymbol{u}(\boldsymbol{r}, t)$ and magnetic field $\boldsymbol{B}(\boldsymbol{r}, t)$, where the amplitudes $\delta\boldsymbol{u}$ and $\delta\boldsymbol{B}$ are also constant. As before, the velocity is taken to be zero in the background state.

---

[5] This last assumption breaks down for MHD waves in stars, where an adiabatic approximation is more appropriate. The practical result is that the isothermal sound speed a_T should be replaced by the adiabatic speed a_S in such formulas as equation (9.63).

Beginning with mass continuity, equation (3.7), the generalization of equation (9.20a) is

$$-i\,\omega\,\delta\rho \,+\, i\,\rho_\circ\,\boldsymbol{k}\cdot\delta\boldsymbol{u} \,=\, 0\ . \tag{9.62}$$

The perturbed equation of state (9.20b) remains the same. For momentum conservation, we employ equation (3.3), but drop the gravitational term. Using the isothermal assumption for P and equation (9.32) for the current $\boldsymbol{j}$, we have

$$\rho\,\frac{D\boldsymbol{u}}{Dt} \,=\, -a_T^2\,\boldsymbol{\nabla}\rho \,+\, \frac{1}{4\,\pi}\,(\boldsymbol{\nabla}\times\boldsymbol{B})\times\boldsymbol{B}\ . \tag{9.63}$$

Linear perturbation of this last equation gives

$$-i\,\omega\,\rho_\circ\,\delta\boldsymbol{u} \,=\, -i\,a_T^2\,\delta\rho\,\boldsymbol{k} \,+\, \frac{1}{4\,\pi}\,(i\,\boldsymbol{k}\times\delta\boldsymbol{B})\times\boldsymbol{B}_\circ\ . \tag{9.64}$$

Finally, we incorporate the condition of flux freezing, equation (9.45). After application of the perturbation, we have

$$-i\,\omega\,\delta\boldsymbol{B} \,=\, i\,\boldsymbol{k}\times(\delta\boldsymbol{u}\times\boldsymbol{B}_\circ)\ . \tag{9.65}$$

9.5.2 Alfvén and Sonic Waves

If we dot $\boldsymbol{k}$ into equation (9.65), we find that $\boldsymbol{k}\cdot\delta\boldsymbol{B} = 0$. Thus, the disturbance in the magnetic field is normal to the direction of propagation. The other properties of the MHD wave depend on the relative directions of $\boldsymbol{k}$ and $\boldsymbol{B}_\circ$. Suppose first that these two vectors are parallel, *i. e.*, that the wave propagates along the background field lines. Within this framework, we also wish to examine separately *transverse* and *longitudinal* waves. In the former case, the velocity disturbance $\delta\boldsymbol{u}$ is perpendicular to $\boldsymbol{k}$, while it is parallel in the latter. Taking the transverse mode first, equation (9.62) tells us that $\delta\rho = 0$, *i. e.*, that the density is unaltered. This property of transverse waves is true generally, for any propagation direction.

To proceed with the case at hand, we expand the triple vector product in equation (9.65) to find

$$-\omega\,\delta\boldsymbol{B} \,=\, (\boldsymbol{k}\cdot\boldsymbol{B}_\circ)\,\delta\boldsymbol{u} \,-\, (\boldsymbol{k}\cdot\delta\boldsymbol{u})\,\boldsymbol{B}_\circ\ . \tag{9.66}$$

The product $\boldsymbol{k}\cdot\boldsymbol{B}_\circ$ is positive, by assumption, while $\boldsymbol{k}\cdot\delta\boldsymbol{u}$ again vanishes for a transverse wave. Thus, $\delta\boldsymbol{u}$ is antiparallel to $\delta\boldsymbol{B}$, and their magnitudes are related by

$$\delta u \,=\, -\frac{\omega\,\delta B}{k\,B_\circ}\ . \tag{9.67}$$

The vector relations are illustrated in Figure 9.15a, where we have taken $\boldsymbol{k}$ to lie along the x-axis, and $\delta\boldsymbol{u}$ to point in the positive y-direction. Turning to the momentum equation (9.64), the first righthand term vanishes, while the triple product also lies along the y-axis. We find, therefore,

$$\delta u \,=\, -\frac{k\,\delta B\,B_\circ}{4\,\pi\,\omega\,\rho_\circ}\ . \tag{9.68}$$

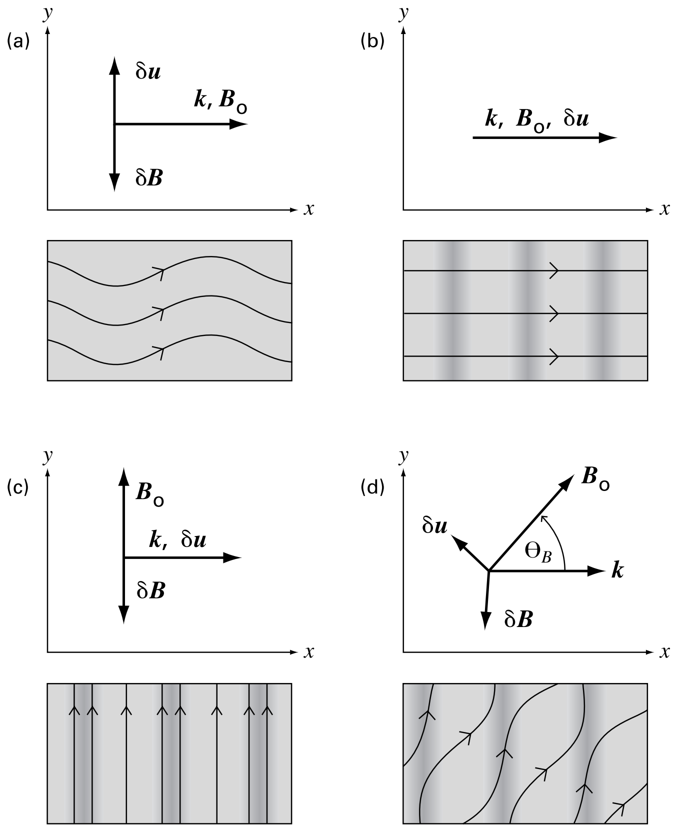


Figure 9.15 Relationships between the propagation vector $\boldsymbol{k}$, the underlying field $\boldsymbol{B}_\circ$, its perturbation $\delta\boldsymbol{B}$, and the induced fluid velocity $\delta\boldsymbol{u}$ in MHD waves. For each mode, the density and field perturbations are also sketched. If $\boldsymbol{k}$ and $\boldsymbol{B}_\circ$ are parallel, the mode can be (a) transverse or (b) longitudinal. (c) If $\boldsymbol{k}$ and $\boldsymbol{B}_\circ$ are perpendicular, the mode must be longitudinal. (d) In the most general case, the wave is neither purely transverse nor longitudinal.

Eliminating δu between (9.67) and (9.68) yields the dispersion relation:

$$\frac{\omega^2}{k^2} = \frac{B_\circ^2}{4\,\pi\,\rho_\circ} \, . \tag{9.69}$$

For any plane wave, ω/k gives the phase velocity. In our particular case ($\boldsymbol{k}$ parallel to $\boldsymbol{B}_\circ$, transverse mode), this quantity is V_A, the Alfvén velocity introduced in Chapter 8:

$$\begin{aligned} V_A &\equiv \frac{B_\circ}{\sqrt{4\,\pi\,\rho_\circ}} \\ &= 0.5 \text{ km s}^{-1} \left(\frac{B_\circ}{10\ \mu G}\right) \left(\frac{n_{\mathrm{H}_2}}{10^3 \text{ cm}^{-3}}\right)^{-1/2} \, . \end{aligned} \tag{9.70}$$

This disturbance is known as an *Alfvén wave*. Since ω and k are linearly related, V_A also equals $d\omega/dk$, the *group* velocity governing the motion of wave packets. Furthermore, the absence of the sound speed from the dispersion relation shows that the restoring force in this case is simply magnetic tension. Thus, the transverse disturbance corresponds to a sinusoidal rippling of the background field line (Figure 9.15a).

Longitudinal waves propagating along the background field have $\delta\boldsymbol{u}$ parallel to both $\boldsymbol{k}$ and $\boldsymbol{B}_\circ$ (see Figure 9.15b). The product $\delta\boldsymbol{u} \times \boldsymbol{B}_\circ$ thus vanishes in equation (9.65), so that $\delta\boldsymbol{B} = 0$. Since only the gas is effected, these perturbations must be indistinguishable from ordinary sound waves. Equation (9.64) gives, after projection along the x-axis,

$$\delta u \,=\, \frac{a_T^2\, k\, \delta\rho}{\omega\, \rho_\circ} \ . \tag{9.71}$$

Similarly, we deduce from equation (9.62) that

$$\delta u \,=\, \frac{\omega\, \delta\rho}{k\, \rho_\circ} \ . \tag{9.72}$$

Elimination of δu between (9.71) and (9.72) now gives

$$\frac{\omega^2}{k^2} \,=\, a_T^2 \ . \tag{9.73}$$

As anticipated, this is the usual dispersion relation for sound waves, in which thermal pressure alone constitutes the restoring force.

9.5.3 Magnetosonic Waves

We next consider waves propagating orthogonally to the ambient field. In Figure 9.15b, we keep $\boldsymbol{k}$ in the x-direction, but let $\boldsymbol{B}_\circ$ point up the y-axis. The first thing to notice is that transverse waves cannot exist under these conditions. We have already seen that all such waves have $\delta\rho = 0$, so that gas pressure cannot contribute to the restoring force. On the other hand, motion of the gas along $\boldsymbol{B}_\circ$ also encounters no opposition from magnetic pressure or tension. We reach the same conclusion mathematically by noting, from equation (9.65), that the vanishing of $\delta\boldsymbol{u} \times \boldsymbol{B}_\circ$ implies $\delta\boldsymbol{B} = 0$. Equation (9.64) then gives $\delta\boldsymbol{u} = 0$, as well.

Longitudinal waves, however, do still exist. In fact, the compression is now resisted by *both* thermal and magnetic forces. Looking at equation (9.66), $\boldsymbol{k} \cdot \boldsymbol{B}_\circ = 0$, while $\boldsymbol{k} \cdot \delta\boldsymbol{u} > 0$, so that $\delta\boldsymbol{B}$ is antiparallel to $\boldsymbol{B}_\circ$, and related in magnitude to $\delta\boldsymbol{u}$ by

$$\delta u \,=\, -\frac{\omega\, \delta B}{k\, B_\circ} \ . \tag{9.74}$$

Equation (9.62) also implies that

$$\delta u \,=\, \frac{\omega\, \delta\rho}{k\, \rho} \ . \tag{9.75}$$

In equation (9.64), we now have $(\boldsymbol{k} \times \delta \boldsymbol{B}) \times \boldsymbol{B}_\circ = +k\, B_\circ\, \delta B\, \hat{\boldsymbol{e}}_x$. Projecting (9.64) along the x-direction and using (9.74) and (9.75) gives the dispersion relation

$$\frac{\omega^2}{k^2} = a_T^2 + \frac{B_\circ^2}{4\,\pi\,\rho_\circ} \ , \tag{9.76}$$

in which the combined effect of thermal and magnetic pressure is evident. As illustrated in Figure 9.15c, the disturbance – known as a *magnetosonic wave* – consists of alternating compression and rarefaction of the ambient field and gas, with no bending of $\boldsymbol{B}$ and hence no magnetic tension. We shall denote the phase velocity obtained from equation (9.76) as $V_{\rm max}$, since it is the maximum attainable by an MHD wave:

$$V_{\rm max} = \left(a_T^2 + V_A^2\right)^{1/2} \ . \tag{9.77}$$

In the most general case, $\boldsymbol{k}$ is tilted with respect to $\boldsymbol{B}_\circ$ at some angle θ_B, neither 0 nor $\pi/2$ (see Figure 9.15d). Three types of waves are possible, as we now show. After substituting for $\delta\rho$ from (9.62) and $\delta\boldsymbol{B}$ from (9.65), equation (9.64) contains only the small amplitude δu:

$$\omega^2\, \delta\boldsymbol{u} = a_T^2\, (\boldsymbol{k}\cdot\delta\boldsymbol{u})\, \boldsymbol{k} + \frac{1}{4\,\pi\,\rho_\circ}\{\boldsymbol{k}\times[\boldsymbol{k}\times(\delta\boldsymbol{u}\times\boldsymbol{B}_\circ)]\}\times\boldsymbol{B}_\circ \ .$$

Expansion of the second righthand term leads to a rather lengthy expression for $\delta\boldsymbol{u}$ in terms of $\boldsymbol{k}$ and $\boldsymbol{B}_\circ$:

$$\begin{aligned}\left(\omega^2 - \frac{|\boldsymbol{B}_\circ\cdot\boldsymbol{k}|^2}{4\,\pi\,\rho_\circ}\right)\delta\boldsymbol{u} = &\left[a_T^2\,(\delta\boldsymbol{u}\cdot\boldsymbol{k}) - \frac{(\delta\boldsymbol{u}\cdot\boldsymbol{B}_\circ)(\boldsymbol{B}_\circ\cdot\boldsymbol{k})}{4\,\pi\,\rho_\circ} + V_A^2\,(\delta\boldsymbol{u}\cdot\boldsymbol{k})\right]\boldsymbol{k} \\ &- \frac{(\delta\boldsymbol{u}\cdot\boldsymbol{k})(\boldsymbol{B}_\circ\cdot\boldsymbol{k})}{4\,\pi\,\rho_\circ}\,\boldsymbol{B}_\circ\end{aligned} \tag{9.78}$$

Dotting this equation with $\boldsymbol{k}$ leads to the simpler scalar relation

$$(\omega^2 - V_{\rm max}^2\, k^2)\,(\delta\boldsymbol{u}\cdot\boldsymbol{k}) = -\frac{(\delta\boldsymbol{u}\cdot\boldsymbol{B}_\circ)(\boldsymbol{B}_\circ\cdot\boldsymbol{k})\,k^2}{4\,\pi\,\rho_\circ} \ , \tag{9.79}$$

while dotting equation (9.78) with $\boldsymbol{B}_\circ$ gives

$$a_T^2\,(\boldsymbol{B}_\circ\cdot\boldsymbol{k})\,(\delta\boldsymbol{u}\cdot\boldsymbol{k}) = \omega^2\,(\delta\boldsymbol{u}\cdot\boldsymbol{B}_\circ) \ . \tag{9.80}$$

One solution of Equations (9.79) and (9.80) has both $\delta\boldsymbol{u}\cdot\boldsymbol{k} = 0$ and $\delta\boldsymbol{u}\cdot\boldsymbol{B}_\circ = 0$. In this transverse mode, a generalized Alfvén wave, $\delta\boldsymbol{u}$ is perpendicular to the plane defined by $\boldsymbol{B}_\circ$ and $\boldsymbol{k}$. From equation (9.78), the dispersion relation may be written:

$$\begin{aligned}\omega^2 &= \frac{|\boldsymbol{B}_\circ\cdot\boldsymbol{k}|^2}{4\,\pi\,\rho_\circ} \\ &= k^2\,V_A^2\,\cos^2\theta_B \ ,\end{aligned} \tag{9.81}$$

so that the phase velocity is $V_A\,\cos\,\theta_B$. From equation (9.66), $\delta\boldsymbol{B}$ is again antiparallel to $\delta\boldsymbol{u}$, and the restoring force is solely magnetic tension.

The other two modes have $\delta\boldsymbol{u}$ lying in the $\boldsymbol{B}_\circ$–$\boldsymbol{k}$ plane. The field perturbation $\delta\boldsymbol{B}$ also lies in this plane and, as always, is normal to $\boldsymbol{k}$. However, $\delta\boldsymbol{u}$ is neither perpendicular nor parallel to $\boldsymbol{k}$, resulting in waves that are neither purely transverse nor compressional, but partake of both characteristics. To derive the dispersion relation, we divide equation (9.79) by (9.80), obtaining

$$\omega^4 - \omega^2 k^2 V_{\rm max}^2 + a_T^2 V_A^2 k^4 \cos^2\theta_B = 0 . \qquad (9.82)$$

This gives for the phase velocity

$$\frac{\omega}{k} = \frac{1}{2}\left[V_{\rm max}^2 \pm \sqrt{V_{\rm max}^4 - 4\,V_A^2\,a_T^2\,\cos^2\theta_B}\right]^{1/2} . \qquad (9.83)$$

Thus, the two additional modes propagate at distinct phase (and group) velocities, and are designated *fast* and *slow magnetosonic waves*. For any angle θ_B, the Alfvén velocity V_A lies between these two speeds. As $\boldsymbol{k}$ approaches $\boldsymbol{B}_\circ$ in direction, the fast mode becomes an Alfvén wave, with $\delta\boldsymbol{u}$ lying orthogonal to that in the transverse mode. In this same limit, the slow mode degenerates into a sound wave. As $\boldsymbol{k}$ becomes orthogonal to $\boldsymbol{B}_\circ$, the fast mode becomes a magnetosonic wave, while the slow mode does not propagate.

Which of these three wave modes actually exist within molecular clouds? Inside a hypothetical, nonmagnetized cloud supported only by thermal pressure, any fluid element in supersonic motion eventually forms a shock. Indeed, shocks also arise from ordinary sound waves, through the process of nonlinear *steepening*. Consider a solitary, low-amplitude pressure disturbance, traveling through the cloud at speed a_T. As this wave passes any point, the associated compression raises the gas temperature slightly, an effect we have ignored in our isothermal approximation. The sound speed increases with temperature as $T^{1/2}$. Hence, any disturbances within the temporarily heated region propagate slightly faster than the original waveform. The resulting pileup increases the pressure gradient until the traveling front becomes a true shock. Heating of matter as it crosses a shock front leads to radiative loss of energy. Thus, both supersonic motion and internal sound waves rapidly dissipate.

This situation changes with the addition of an internal magnetic field. The extra source of pressure makes shocks more difficult to form. For $\boldsymbol{B}_\circ$ oriented perpendicular to a flow, the criterion is now that the fluid element must be traveling faster than $V_{\rm max}$ in equation (9.77). From equation (9.70), larger clouds typically have $V_{\rm max}$ greatly exceeding a_T, which is only 0.2 km s^{-1} at a temperature of 10 K. The supersonic velocities deduced from molecular line widths are sometimes *sub-Alfvénic*, at least if the telescopic beam lies within a single clump (recall equation 3.21). Such motion can survive longer than the characteristic crossing time of L/V_A, which is 2×10^6 yr for $L = 1$ pc, $n_{\rm H_2} = 10^3$ cm^{-3}, and a mean field of 10 μG. However, both the fast and slow magnetosonic waves undergo steepening, as they involve compression of the gas. If the observed motion indeed represents MHD waves, it is the transverse, Alfvén mode that survives, at least until it damps through mode conversion (see below) or ambipolar diffusion (Chapter 10).

9.5.4 Effective Pressure

Let us now see how MHD waves create extra pressure. At any fixed location traversed by such a wave, the physical variables undergo rapid sinusoidal fluctuations. Averaging over a cycle, it is not obvious how a steady force could result. The key is that the extra force depends on *products* of oscillating variables, and such products can have a non-zero average. To develop this point quantitatively, we start again from the momentum equation (9.63). Expanding both the convective derivative on the left and the triple vector product on the right, we recast this equation in component notation:

$$\rho \frac{\partial u_i}{\partial t} + \rho\, u_j \frac{\partial u_i}{\partial x_j} = -a_T^2 \frac{\partial \rho}{\partial x_i} - \frac{1}{8\,\pi} \frac{\partial}{\partial x_i} (B_i\, B_j) + \frac{1}{4\,\pi} B_j \frac{\partial B_i}{\partial x_j} \ , \tag{9.84}$$

where we have utilized the Einstein summation convention for repeated indices. The continuity equation (3.7) and $\boldsymbol{\nabla} \cdot \boldsymbol{B} = \partial B_j / \partial x_j = 0$ enable us to write the foregoing equation in "conservation form":

$$\frac{\partial}{\partial t} (\rho\, u_i) = -\frac{\partial \Pi_{ij}}{\partial x_j} \ . \tag{9.85}$$

Here the *pressure* (or *momentum flux*) *tensor* Π_{ij} is

$$\Pi_{ij} = \rho\, u_i\, u_j + a_T^2\, \rho\, \delta_{ij} + \frac{1}{8\,\pi} B_k\, B_k\, \delta_{ij} - \frac{1}{4\,\pi} B_i\, B_j \ . \tag{9.86}$$

To obtain the force due to the waves, we substitute into the right side of (9.85) the perturbation form (9.61) for all variables, but generalized to include amplitudes and wave vectors that vary smoothly in space. We then integrate over a wave cycle to obtain the average incremental force. Within Π_{ij}, only equilibrium terms and those with a quadratic dependence on fluctuating variables survive the integration. The cycle-average of $\cos^2 \varpi t$ is 1/2, so we find the perturbed part of the pressure tensor to be

$$\Pi_{ij}^{\rm wave} = \frac{1}{2} \rho_\circ\, \delta u_i\, \delta u_j + \frac{1}{16\,\pi} \delta B_k\, \delta B_k\, \delta_{ij} - \frac{1}{8\,\pi} \delta B_i\, \delta B_j \ . \tag{9.87}$$

The force per unit volume due to the waves is then

$$F_i^{\rm wave} = -\frac{\partial}{\partial x_j} \Pi_{ij}^{\rm wave} \ . \tag{9.88}$$

For the generic MHD wave, Equations (9.87) and (9.88) indicate that the wave pressure tensor is anisotropic, so that the force in any one direction partially depends on gradients in orthogonal directions. Extending our earlier, mechanical analogy, this is the type of behavior one would expect from an elastic membrane, in which stresses arise from both stretching and shearing deformations. For an Alfvén wave, however, we recall that $\delta \boldsymbol{u}$ and $\delta \boldsymbol{B}$ are antiparallel. From Equations (9.67), (9.69), and (9.70), their components are related by

$$\frac{\delta u_i}{V_A} = -\frac{\delta B_i}{B_\circ} \ . \tag{9.89}$$

Substitution of this result into equation (9.87) leads to cancellation of the off-diagonal elements of Π_{ij}. Reverting to vector notation and changing to subscripts, we may write the force as gradient of a scalar:

$$\boldsymbol{F}_{\rm wave} = -\boldsymbol{\nabla} P_{\rm wave} \ , \tag{9.90}$$

where the *wave pressure* is now

$$P_{\rm wave} \,=\, \frac{1}{16\,\pi}\,|\delta\boldsymbol{B}|^2 \ . \tag{9.91}$$

The term $\boldsymbol{F}_{\rm wave}$ may be added to the right side of equation (9.46) to give a more complete account of force balance. If waves are present over a range of wavelengths, their individual pressures should be calculated according to (9.91) and then summed to give the total contribution.

As mentioned previously, our knowledge of the wave amplitudes is still too imprecise for the construction of detailed models. Generally speaking, however, the existence of the extra pressure alleviates the problem of cloud structure, at least for the larger clouds exhibiting nonthermal motion. If the amplitude of a wave were to decrease in the direction of $\boldsymbol{B}_\circ$, equation (9.91) shows that the associated force would also point in that direction. Thus, the prediction of severe flattening for relatively cold, magnetostatic clouds need no longer hold. In essence, the clouds are warmer than their gas temperatures would indicate.

Let us pursue this idea a bit further. In order for the extra pressure to be effective against gravity, the equivalent "sound speed" must be comparable to the virial velocity. Is this the case? The measured line widths of the larger clouds give internal velocities that are roughly Alfvénic. From equation (9.89), this means that $|\delta B| \approx B_\circ$, the result already indicated by the polarization studies. This latter relation implies, from equation (9.91), that the wave "sound speed," given by $(P_{\rm wave}/\rho_\circ)^{1/2}$, is close to V_A. Finally, the empirical finding that $\mathcal{M} \approx |\mathcal{W}|$ means that $V_A \approx V_{\rm vir}$. Assuming that the disturbances seen are typical of those occurring throughout the interior, this chain of reasoning implies that Alfvén waves do exert a force comparable to self-gravity and hence should be capable of altering cloud shapes.

Throughout this discussion, we have utilized plane waves as a convenient tool for deriving important physical results. However, a wave front within a clumpy molecular cloud will actually advance faster in some regions than others, depending on the local Alfvén velocity. The resulting distortion of the front is well documented in laboratory plasmas, where it is called *phase mixing*. Inhomogeneities in cloud density also lead to reflection of waves and to *mode conversion*, *i. e.*, the transfer of energy into both fast and slow magnetosonic waves.

These inevitable complexities to the simple wave picture suggest that it may be more appropriate to consider a model of MHD *turbulence*. Many numerical simulations have explored this idea in some detail. One introduces random velocity perturbations into a computational box enclosing magnetized gas. The results are a dramatic illustration of mode conversion. Whatever the character of the initial disturbance, energy soon appears in magnetosonic waves. These steepen and dissipate through shocks. Additional losses occur through viscous heating.

The volumetric rate of energy decline for turbulent eddies of characteristic size L and velocity V is of order $\rho\, V^3/L$, where ρ is the region's mass density. Each eddy is therefore losing its entire kinetic energy in about one turnover time, L/V. All told, the simulations find that the decay of impressed turbulence in a fluid does not depend sensitively on whether the fluid is magnetized at all.

There are at least two ways to interpret these findings. On the one hand, they may indicate that the real perturbations are *not* of a fully turbulent nature. Perhaps the motion is more organized, so that internal shocks are effectively avoided. If, on the other hand, turbulence is indeed present, then it must be continually pumped by some external energy source to prevent rapid decay. The nature of that source is unknown. But even if disturbances exist in the more diffuse gas just outside the cloud, it is not obvious that this energy will efficiently penetrate, instead of being reflected at the cloud surface. It does appear, in any case, that neither random turbulence nor an ensemble of plane waves provides a fully satisfactory description of the interior state.

Chapter Summary

A simple theoretical model for a molecular cloud consists of a uniform-temperature sphere supported by internal pressure against self-gravity. Such an object has a maximum possible mass, the Bonnor-Ebert (or Jeans) value. Clouds with masses exceeding this limit, either because they are too cold or too dense, are not in force balance, and undergo gravitational collapse.

Rotation is a stabilizing influence that raises the Jeans mass at fixed temperature and background pressure. Numerically constructed cloud models are flattened along the central axis if their rotational kinetic energy is an appreciable fraction of the gravitational potential energy. However, actual molecular clouds are rotating too slowly for their aspherical shapes to arise by this means.

All observed clouds larger than dense cores have masses exceeding the Jeans limit, yet are generally not collapsing. Additional support comes from the interstellar magnetic field, which is effectively frozen into the cloud matter. The field thus increases at higher density, providing both magnetic pressure and tension. Detailed magnetostatic models are flattened structures that become thin slabs in the strong-field limit. Again, the model shapes do not match those inferred from radio mapping.

Magnetized gas can sustain periodic disturbances of various kinds. Fast and slow magnetosonic waves alternately expand and compress both the field and gas, and are eventually dissipated as shocks. Alfvén waves are transverse disturbances in the field that do not compress the gas. Wave pressure in the gas, especially from the Alfvén mode, can provide additional mechanical support and alleviate the problem of cloud shapes.

Suggested Reading

Section 9.1 This chapter relies heavily on the "static method" of using equilibrium sequences to diagnose instability. A lucid explanation of the method, presented in the stellar context, is in

Tassoul, J.-L. 1978, *Theory of Rotating Stars*, (Princeton: Princeton U. Press), Chapter 6.

Historically, the gravitational instability of isothermal spheres was discovered independently by

Ebert, R. 1955, ZAp, 37, 222

Bonnor, W. B. 1956, MNRAS, 116, 351.

The original 1902 derivation of the Jeans length is reprinted in

Jeans, J. H. 1961, *Astronomy and Cosmogony* (New York: Dover), Chapter 13.

Section 9.2 Our discussion of rotating equilibria is based on

Stahler, S. W. 1983a, ApJ, 268, 155

———— 1983b, ApJ, 268, 165.

A numerical study that explores configurations of higher angular momentum and density contrast is

Kiguchi, M., Narita, S., Miyama, S. M., Hayashi, C. 1987, ApJ, 317, 830.

Section 9.3 For a derivation of the ideal MHD equation and its physical interpretation in terms of flux freezing, see

Parker, E. N. 1979, *Cosmical Magnetic Fields*, (Oxford: Clarendon Press), Chapter 4.

A detailed calculation of the electrical conductivity in molecular clouds is apparently still lacking. The general formulation of the problem for any lightly ionized plasma is in

Braginskii, S. L. 1965, in *Reviews of Plasma Physics*, Vol. 1, ed. M. A. Leontovich (New York: Consultants Bureau), p. 205.

Section 9.4 Our treatment of magnetized equilibria largely follows the work of

Mouschovias, T. Ch. 1976a, ApJ, 206, 753

———— 1976b, ApJ, 207, 141.

We have also utilized the studies of

Tomisaka, K., Ikeuchi, S., & Nakamura, T. 1988, ApJ, 335, 239

———— 1989, ApJ, 341, 220,

which emphasize the issue of gravitational stability.

Section 9.5 The significance of MHD waves in molecular clouds was first recognized by

Arons, J. & Max, C. E. 1975, ApJ, 196, L77.

Much of the research in this field comes from the solar physics community. For more discussion of the various wave modes, see

Priest, E. R. 1982, *Solar Magneto-hydrodynamics* (Dordrecht: Reidel), Chapter 4.

The pressure from Alfvén waves is derived in

Hollweg, J. V. 1973, ApJ, 181, 547.

For the decay of MHD turbulence in numerical simulations, see

MacLow, M.-M., Klessen, R. S., Burkert, A., & Smith, M. D. 1998, Phys. Rev. Lett., 80, 2754.

The Formation of Stars
Steven W. Stahler and Francesco Palla

10 The Collapse of Dense Cores

The detection of infrared point sources within many dense cores shows unequivocally that these cloud structures form stars. In this chapter and the next, we confront the central issue of how this transition actually occurs. Here we study, again from a theoretical perspective, the collapse process itself. Chapter 11 will then examine the properties of the nascent star and its infalling envelope of gas and dust.

In our study of equilibria, we remarked that no single model has self-consistently incorporated the combined effects of gravity, thermal pressure, rotation, and magnetic fields. The same statement applies to the problem of cloud collapse, apart from calculations that assume a disk-like geometry (see § 10.1 below). Even if more general results were available, they would be difficult to appreciate without the necessary background. Accordingly, we follow our previous approach of building up the reader's intuition through simplified models that include only a subset of the relevant effects. Our first order of business, however, is to see how an initially stable cloud, like the observed starless dense cores, can evolve to the point where collapse becomes inevitable.

10.1 Ambipolar Diffusion

The magnetostatic equilibria we have studied resist self-gravity through both thermal pressure and magnetic forces. More precisely, it is the charged species within a cloud that actually sense the magnetic field. Electrons and ions, gyrating rapidly around $\boldsymbol{B}$, collide with the neutrals, and the resultant drag on the latter helps to counteract gravity. As we saw, the drag force per unit volume between any charged component and the neutrals involves the product of the density and relative velocity of that species (recall equations (9.41) and (9.43)). If the cloud's level of ionization is sufficiently low, these relative speeds become appreciable. The neutrals then gradually drift across magnetic field lines in response to gravity. With the consequent loss of magnetic flux, the cloud contracts until it eventually becomes unstable to true collapse, *i. e.*, to the attainment of free-fall speeds. In the notation of Chapter 9, the mass $M_{\rm crit}$ decreases until it becomes lower than the actual cloud mass M.

10.1.1 Ion-Neutral Drift

In order to follow the flux loss in detail, we first note that the electrons and ions must be moving at very nearly the same velocity. That is, if L is a mean cloud diameter, then $L/|\boldsymbol{u}_i - \boldsymbol{u}_e|$ is much longer than the time over which the cloud evolves. We may see this by taking the magnitude of the vector equation (9.33), which expresses $\boldsymbol{j}$ in terms of the relative velocity. Using Ampère's

The Formation of Stars. Steven W. Stahler and Francesco Palla

ISBN: 3-527-40559-3

law, equation (9.32), for $\boldsymbol{j}$, we have

$$\begin{aligned}\frac{L}{|\boldsymbol{u}_i - \boldsymbol{u}_e|} &= \frac{n_e\, e\, L}{|\boldsymbol{j}|} \\ &= \frac{4\,\pi\, n_e\, e\, L}{c\,|\boldsymbol{\nabla} \times \boldsymbol{B}|} \\ &\sim \frac{4\,\pi\, n_e\, e\, L^2}{c\, B} \ .\end{aligned} \tag{10.1}$$

For a dense core with $L = 0.1$ pc, $B = 30\,\mu$G, $n_{\rm H_2} = 10^4\ {\rm cm}^{-3}$, and $n_e = 5 \times 10^{-8}\, n_{\rm H_2} = 5 \times 10^{-4}\,{\rm cm}^{-3}$, the time $L/|\boldsymbol{u}_i - \boldsymbol{u}_e|$ is of order 10^{10} yr. We may therefore consider the electrons and ions to be effectively a single plasma drifting relative to the neutrals.

Our determination of $\boldsymbol{v}_{\rm drift} \equiv \boldsymbol{u}_i - \boldsymbol{u}$ starts with the equations of motion for the electrons and ions. Recall that (9.40) and (9.42) are written in the reference frame comoving with the neutrals. Adding the two equations together, we find

$$\begin{aligned}0 &= e\, n_e\, (\boldsymbol{u}_i' - \boldsymbol{u}_e') \times \boldsymbol{B}'/c + n_e\, (\boldsymbol{f}_{\rm in}' + \boldsymbol{f}_{\rm en}') \\ &= \boldsymbol{j} \times \boldsymbol{B}/c + n_e\, (\boldsymbol{f}_{\rm in}' + \boldsymbol{f}_{\rm en}') \ .\end{aligned} \tag{10.2}$$

Here we have used the results that $\boldsymbol{j} = \boldsymbol{j}'$ (equation (9.34)) and $\boldsymbol{B} = \boldsymbol{B}'$ for nonrelativistic neutral velocities. We next apply the expressions for the drag forces in equations (9.41) and (9.43). The near equality of $\boldsymbol{u}_e'$ and $\boldsymbol{u}_i'$, together with the fact that $\boldsymbol{u}_i' \equiv \boldsymbol{u}_i - \boldsymbol{u} = \boldsymbol{v}_{\rm drift}$, allows us to write (10.2) as an expression for $\boldsymbol{v}_{\rm drift}$ in terms of $\boldsymbol{j} \times \boldsymbol{B}/c$. After again using Ampère's law for $\boldsymbol{j}$, we find

$$\begin{aligned}\boldsymbol{v}_{\rm drift} &= \frac{(\boldsymbol{\nabla} \times \boldsymbol{B}) \times \boldsymbol{B}}{4\,\pi\, n\, n_e\, [m_n \langle \sigma_{\rm in}\, u_i' \rangle + m_e \langle \sigma_{\rm en}\, u_e' \rangle]} \\ &\approx \frac{(\boldsymbol{\nabla} \times \boldsymbol{B}) \times \boldsymbol{B}}{4\,\pi\, n\, n_e\, [m_n \langle \sigma_{\rm in}\, u_i' \rangle]} \ .\end{aligned} \tag{10.3}$$

In making the last approximation, we have used the fact that the numerical values for $\langle \sigma_{\rm in} u_i' \rangle$ and $\langle \sigma_{\rm en} u_e' \rangle$ indicate a dominance by ion-neutral collisions, just the reverse of the situation for the conductivity σ (recall equation (9.44)). Our previous estimate of σ, in fact, needs correction, since a finite $\boldsymbol{v}_{\rm drift}$ alters $\boldsymbol{j}$ for given external fields. More significantly, the existence of $\boldsymbol{v}_{\rm drift}$ forces us to reevaluate the physical meaning of flux freezing. Reviewing the derivation in § 9.3, we note that the first form of Ohm's law, equation (9.34), really holds in the rest frame of the plasma. Thus the velocity appearing in equations (9.35)–(9.39) should be $\boldsymbol{u}_i$. In the modified form of (9.39), it is still true that the final term, representing Ohmic dissipation, is small. Using $\boldsymbol{u}_i = \boldsymbol{u} + \boldsymbol{v}_{\rm drift}$, equation (9.45) is now replaced by

$$\begin{aligned}\frac{\partial \boldsymbol{B}}{\partial t} &= \boldsymbol{\nabla} \times (\boldsymbol{u}_i \times \boldsymbol{B}) \\ &= \boldsymbol{\nabla} \times (\boldsymbol{u} \times \boldsymbol{B}) + \boldsymbol{\nabla} \times (\boldsymbol{v}_{\rm drift} \times \boldsymbol{B}) \ .\end{aligned} \tag{10.4}$$

In summary, flux freezing still holds if the conductivity is large enough, but implies that the electrons and ions are tied to $\boldsymbol{B}$, while the neutral atoms and molecules in the cloud slip past.

How important quantitatively is this slippage? Let us apply our previous technique of estimating the relevant time scale, here $L/|\boldsymbol{v}_{\rm drift}|$. Using equation (10.3), we have

$$\begin{aligned} \frac{L}{|\boldsymbol{v}_{\rm drift}|} &\approx \frac{4\,\pi\,n\,n_e\,m_n \langle \sigma_{\rm in}\,u_i' \rangle\,L}{|(\boldsymbol{\nabla}\times\boldsymbol{B})\times\boldsymbol{B}|} \\ &\approx 3\times10^6\,{\rm yr}\,\left(\frac{n_{\rm H_2}}{10^4\ {\rm cm^{-3}}}\right)^{3/2}\left(\frac{B}{30\ \mu{\rm G}}\right)^{-2}\left(\frac{L}{0.1\ {\rm pc}}\right)^2 \ , \end{aligned} \tag{10.5}$$

where we have incorporated equation (8.32) for the ionization fraction. Comparing this result to our estimate for cloud lifetimes in Chapter 3, we see that $\boldsymbol{v}_{\rm drift}$ is indeed significant. It is believed, in fact, that ambipolar diffusion is the main process setting the rate at which dense cores evolve prior to their collapse.

We now see that the magnetized structures discussed in § 9.3 are actually *quasi-static*. Both the gas and magnetic field move, but so slowly that the clouds can be viewed as progressing along a sequence of equilibria. Thus, in the full momentum equation (3.3), the velocity $\boldsymbol{u}$ is subsonic (and therefore sub-Alfvénic), so that the term $\rho D\boldsymbol{u}/Dt$ is relatively small and can safely be ignored. As such a cloud secularly evolves, it becomes more centrally condensed. The influence of gravity increases until the fluid velocity becomes substantial, at least within the deep interior. Once $|\boldsymbol{u}|$ approaches the sound speed a_T, the quasi-static description fails, and the cloud begins hydrodynamic collapse.

10.1.2 Magnetic Flux Loss

More precisely, the sequence of cloud equilibria is characterized by a continuous alteration of the function $dM/d\Phi_B$. The drift velocity is proportional to, and lies in the same direction as, the drag force exerted by ions on neutral matter. As illustrated in Figure 10.1, $\boldsymbol{v}_{\rm drift}$ points outward wherever the field lines are compressed toward the central axis. At just those places, $\boldsymbol{u}-\boldsymbol{u}_i = -\boldsymbol{v}_{\rm drift}$ points *inward*, *i. e.*, neutral matter crosses field lines from the outside. The result is that $M(\Phi_B)$ at fixed Φ_B increases with time. If $\boldsymbol{n}$ is the outward normal vector to a flux tube (as in Figure 10.1), then

$$\left(\frac{\partial M}{\partial t}\right)_{\Phi_B} = \int \rho\,\boldsymbol{v}_{\rm drift}\cdot\boldsymbol{n}\,d^2\boldsymbol{x} \ . \tag{10.6}$$

Here the surface integral is over the area of the tube.

For numerical calculations, we may use the definition of $q(\Phi_B)$ in equation (9.48), together with the force balance equation (9.46), to write

$$\boldsymbol{v}_{\rm drift} = \frac{\exp(-\Phi_g/a_T^2)\,\boldsymbol{\nabla}q}{n\,n_e\,m_n\,\langle\sigma_{\rm in}\,u_i'\rangle} \ , \tag{10.7}$$

where we have again neglected the electron-neutral collision term. We then recast $\boldsymbol{\nabla}q$ as $(dq/d\Phi_B)\boldsymbol{\nabla}\Phi_B$. Note that $\boldsymbol{\nabla}\Phi_B$ is parallel to $\boldsymbol{v}_{\rm drift}$ and $\boldsymbol{n}$, which are themselves parallel vectors (see Figure 10.1). We may write the element of area entering equation (10.6) in terms

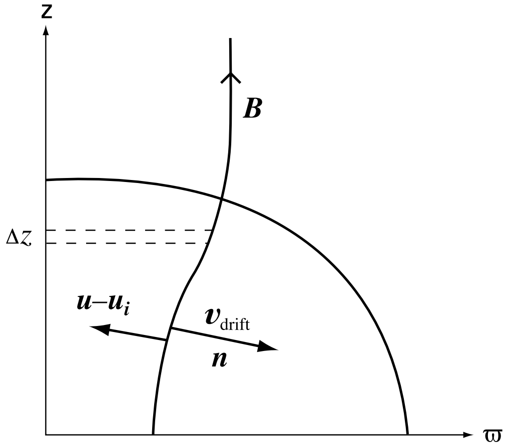


Figure 10.1 Ambipolar diffusion in a magnetized cloud. The drift velocity of ions relative to neutrals points away from the axis, in the same direction as the outward normal vector to the flux tube. Neutrals drift across the field in the opposite direction. Calculating the area of the flux tube requires summation of small distance increments Δz.

of the increment dz and the field-line slope $(\partial\varpi/\partial z)_{\Phi_B}$. We may also use this slope to express the magnitude of $\boldsymbol{\nabla}\Phi_B$. In this way, we find

$$\boldsymbol{\nabla}\Phi_B \cdot \boldsymbol{n}\, d^2\boldsymbol{x} \;=\; 2\,\pi\,\varpi\, dz\, \frac{\partial\Phi_B}{\partial\varpi}\left[1 + \left(\frac{\partial\varpi}{\partial z}\right)^2_{\Phi_B}\right] \;.$$

Finally, we use $m_n = \mu\, m_H$, together with equation (8.32) for n_e/n, to find

$$\begin{aligned}\left(\frac{\partial M}{\partial t}\right)_{\Phi_B} &= \frac{4\,\pi\,\mu\, m_H}{C\,\langle\sigma_{\rm in}\, u_i'\rangle}\,\frac{dq}{d\Phi_B} \\ &\cdot \int_0^{Z_{\rm cl}(\Phi_B)} \frac{dz\,\varpi}{\rho^{1/2}}\,\frac{\partial\Phi_B}{\partial\varpi}\left[1 + \left(\frac{\partial\varpi}{\partial z}\right)^2_{\Phi_B}\right]\exp\left(-\Phi_g/a_T^2\right) \;.\end{aligned} \tag{10.8}$$

The integral in this relation is to be evaluated at each value of Φ_B. Recall that we obtain $dq/d\Phi_B$ by numerical differentiation of equation (9.52). As time progresses, one must continually update $M(\Phi_B)$ using equation (10.8). The range of Φ_B containing the total mass diminishes, as flux gradually leaks out of the cloud (Figure 10.2). At each time, moreover, we may numerically differentiate equation (10.8) with respect to Φ_B in order to obtain $\partial^2 M/\partial\Phi_B\partial t$, which is also the temporal derivative of $dM/d\Phi_B$. After an interval Δt, we then have an updated form for this function that can be used in constructing the next equilibrium model, via the procedure in § 9.4.

One important feature of ambipolar diffusion is that it proceeds fastest toward the cloud center. To see why, note first that gravity and the magnetic force largely balance each other in the ϖ-direction, particularly in the outer regions. At a point with coordinates $(\varpi,\, z)$, we thus have

$$\frac{1}{4\,\pi}\,|(\boldsymbol{\nabla}\times\boldsymbol{B})\times\boldsymbol{B}| \;\approx\; \frac{G\,M\,\rho}{\varpi^2}\;, \tag{10.9}$$

where M is here the cloud mass interior to the point. If we model this interior region as an oblate spheroid, then $M \;\approx\; (4\pi/3)\,G\,\bar\rho\,\varpi^2\,z$, where $\bar\rho$ is the mean interior density. The diffusion time

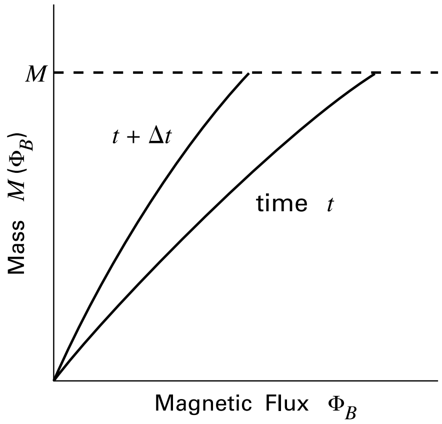


Figure 10.2 Flux loss from a quasi-statically contracting cloud (*schematic*). The amount of mass enclosed by a given flux tube increases with time. Since the cloud mass is constant, the total flux diminishes.

$\varpi/|\boldsymbol{v}_{\rm drift}|$ is then

$$\frac{\varpi}{|\boldsymbol{v}_{\rm drift}|} \approx \frac{3\,n_e\,\langle\sigma_{\rm in}\,u_i'\rangle}{4\,\pi\,G\,\bar{\rho}} \left(\frac{\varpi}{z}\right) \; .$$

Since n_e is proportional to $\rho^{1/2}$, this time varies as $\rho^{-1/2}\,(\rho/\bar{\rho})\,\varpi/z$. For an initially modest density variation, the characteristic time increases outward, since $\rho^{-1/2}\,\varpi$ climbs in that direction. At this stage, the contrast in diffusion rates is not yet strong.

On the other hand, a compact inner region eventually forms. If $M_{\rm inner}$ denotes the mass of this region, we now have

$$\frac{1}{4\,\pi}\,|(\boldsymbol{\nabla}\times\boldsymbol{B})\times\boldsymbol{B}| \approx \frac{G\,M_{\rm inner}\,\rho}{\varpi^2} \; . \tag{10.10}$$

In this case, the diffusion time becomes

$$\frac{\varpi}{|\boldsymbol{v}_{\rm drift}|} \approx \frac{n_e\,\langle\sigma_{\rm in}\,u_i'\rangle\,\varpi^3}{G\,M_{\rm inner}} \; .$$

Since the product $n_e\,\varpi^3$ increases outward rather steeply, the segregation of time scales is greater. These admittedly crude arguments suggest that the cloud's increase in central density should be an *accelerating* process.

10.1.3 Contraction of Flattened Clouds

Numerical simulations, which follow the motion of the fluid from the outset, have confirmed this effect and provided many more interesting details. Note that such calculations generally neglect the acceleration of the ions and electrons, as we implicitly did in equation (10.2). One only solves for the momentum of the neutrals, through equation (3.3) in conjunction with Ampère's law, equation (3.5). We stress that the Lorentz force that helps determine the neutral motion depends only on the instantaneous configuration of the magnetic field. To update that configuration, we employ the drift speed from equation (10.3) and the evolution equation (10.4). Poisson's equation (9.3) then supplies the gravitational potential.

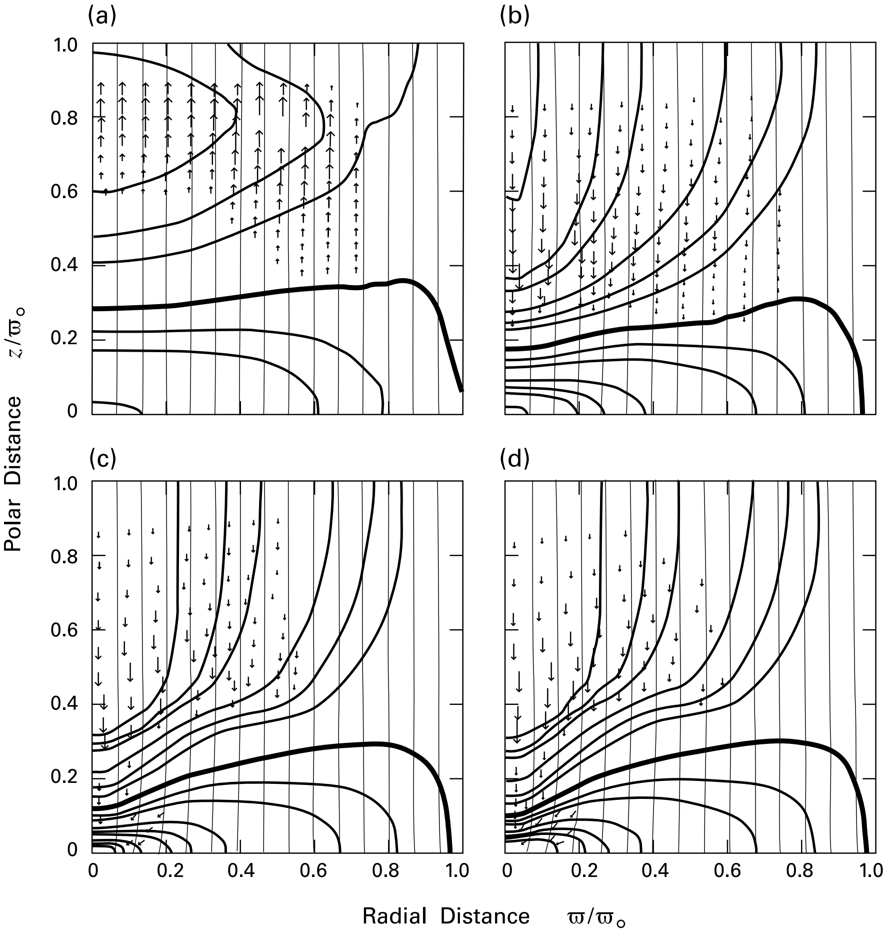


Figure 10.3 Numerical simulation of a contracting, magnetized cloud. The four panels correspond to times (a) 1.02×10^7 yr, (b) 1.51×10^7 yr, (c) 1.60×10^7 yr, and (d) 1.61×10^7 yr after the initial collapse. Light curves are magnetic field lines, heavier ones are isodensity contours, and arrows indicate relative fluid velocities. The thickest solid curve in each panel corresponds to $n_{\rm tot} = 300 \text{ cm}^{-3}$.

The four panels of Figure 10.3 display results from one evolutionary study. Here the initial reference state is a uniform-density cylinder, instead of the spheres discussed in § 9.4. During the calculation, the cylindrical boundary is held fixed, and gas is not allowed to enter or leave the interior volume. The initial, uniform field strength is 30 μG, the number density is 300 cm^{-3}, and the half-height and radius of the cylinder are both 0.75 pc. In the terminology of Chapter 9, the parameter α is 87, while $\xi_\circ$ in equation (9.55) is 1.4, provided we use the cylindrical radius $\varpi_\circ$ in place of $R_\circ$.

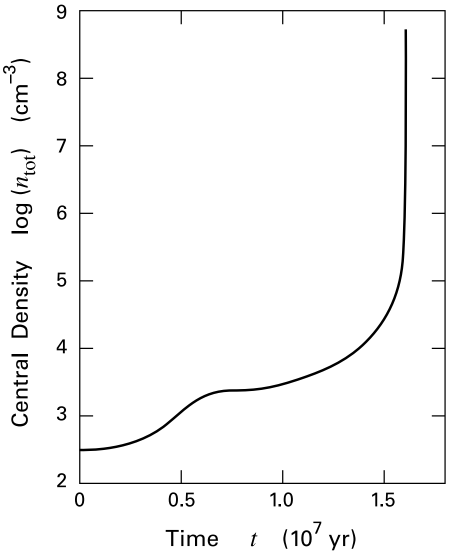


Figure 10.4 Rise of central density in the contracting cloud of Figure 10.3.

Because the height of the cylinder exceeds the Jeans length, the uniform gas immediately goes into collapse. According to equation (3.15), the free-fall time in this case is 4×10^6 yr. The initial ratio of column density to field strength is less than the critical value in equation (9.59). Hence gravity cannot squeeze $\boldsymbol{B}$ toward the axis during this early phase, and the flow occurs down essentially rigid field lines toward the $z = 0$ plane. By 6×10^6 yr, nearly all of the cloud has settled into the oblate configuration shown in the figure. Note that the thickest contour in each panel corresponds to the original $n_{\rm tot}$ of 300 cm^{-3}. During the subsequent evolution, remnant material of lower density continues to rain down on the surface but has negligible dynamical effect. Note also that the *upward* velocity vectors in panel (a) represent a transient pulse generated by the cloud bouncing at the equatorial plane during its collapse.

When the stable equilibrium structure first forms, its central density is only 8 times the value in the uniform cylinder. The density thereafter increases as the result of ambipolar diffusion in the inward, lateral direction. As seen in Figure 10.4, this rise is initially very gradual. By 1.5×10^7 yr, however, Σ/B along the central axis has surpassed the critical value, and the condensation process picks up speed. From then on, the contracting deep interior effectively separates from the more slowly evolving outer portion of the cloud, where ambipolar diffusion is much slower. Figure 10.3 shows how the thickness of the cloud near the axis diminishes with the rise in central density. At the same time, the partially frozen magnetic field is dragged inward, and the central value of B increases by several orders of magnitude.

Such explicit tracking of the vast increase in density possible through ambipolar diffusion is impressive, and the numerical calculations have provided a wealth of information. We defer until the next section any further consideration of the innermost material undergoing collapse. First we return to the issue of the large-scale evolution of dense cores and ask how well the current simulations accord with existing observations.

Referring again to Figure 10.3, the total mass of the cloud is 46 $M_\odot$, so the calculation actually follows the *formation* of a dense core within a larger structure and the approach to collapse of that interior object. The relatively high mass of the parent cloud explains why it is so much flatter than the magnetostatic configurations we encountered before (recall Figure 9.12). Since there is nothing observationally or theoretically that singles out the larger mass scale, the calculated properties of the parent cloud, as well as its history prior to stabilization, must be viewed as rather arbitrary.

We focus, then, on that portion of the cloud with $n_{\rm tot} \gtrsim 10^4$ cm^{-3}. The progressive narrowing of that structure toward the central axis is *not* an artifact of this particular calculation, but is an inevitable outcome of gravitational contraction, *assuming the cloud has a slab-like geometry.* To see why, note that the thickness of a slab at any point is equal, to within a factor of two, to the Jeans length λ_J, where we should use the local midplane density in equation (9.23). Thus, *any* process that allows such a cloud to settle toward higher densities must result in structures that become thinner, both temporally at a fixed radius (*e. g.*, at the center), and spatially, since the midplane density always increases toward the pole. The calculation depicted in Figure 10.3 follows the evolution to a central number density of 3×10^8 cm^{-3}. From equation (9.23), the central thickness of the dense core is then about 0.0017 pc (500 AU), while its equatorial radius (*i. e.*, where $n_{\rm tot} = 10^4$ cm^{-3}) is 0.15 pc. Do extraordinary structures like these, pinched to essentially zero thickness at the center, actually exist?

At the 140 pc distance of Taurus-Auriga, a length of 500 AU subtends an angle of $4''$, within the resolution of current millimeter-wave interferometers. Hence, our question could be answered by direct mapping in the future, but for now must be addressed through statistical and other indirect means. We presented, in fact, a statistical argument in Chapter 3 that dense cores are most likely elongated objects. We also noted in Chapter 9 that theoretical cloud models of even modest density contrast are already oblate. We now see that the discrepancy is only exacerbated once the cloud contracts through ambipolar diffusion.

One shortcoming of current theoretical calculations is that they artificially restrict mass flow *along* magnetic field lines. Thus, the cloud in Figure 10.3 is confined within a rigid cylindrical boundary. In the static models of § 9.4, the cloud has a flexible boundary, but is surrounded by a medium of zero density. For a real molecular cloud, however, there is a vast reservoir of low-density gas that can potentially be drawn in during contraction. Evolutionary calculations that utilize a more porous outer boundary may well achieve elongated structures. Flow along field lines must also play a key role in the approach to collapse, as we discuss in § 10.3 below.

10.1.4 Damping of Alfvén Waves

The temporal sequence in Figure 10.3 depicts the quiescent contraction of gas over a total distance exceeding 1 pc. However, we know from molecular line studies that clouds of this size are traversed by MHD waves. The wave amplitudes and associated kinetic energies are so large that the quasi-static assumption is surely violated. On the other hand, the same studies indicate that dense cores are *not* greatly agitated by waves. Apparently, the waves are unable to affect gas over lengths of roughly 0.1 pc or less. It is encouraging that MHD theory, once it includes ambipolar diffusion, predicts that Alfvén waves are rapidly damped below a critical wavelength. Moreover, this wavelength is indeed of order 0.1 pc, as we now show.

Recall that Alfvén waves represent the periodic, transverse displacement of a magnetic field line together with its attached fluid element. Of course, it is the gyrating ions and electrons that actually move with the field, and these drag along the neutrals through collisions. If the field displacement is too rapid for such collisions to occur, the neutrals respond only sluggishly and with reduced amplitude. In effect, the wave has been damped through ambipolar diffusion.

To obtain this result qualitatively, we rederive the wave dispersion relation $\omega(k)$, but now account for the coupling between the neutrals and charged plasma. We again consider perturbations of the form $\exp[i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)]$ imposed on a static, uniform background, and we include both the neutral velocity amplitude $\delta\boldsymbol{u}$ and its ion counterpart $\delta\boldsymbol{u}_i$. As before, our basic equations include mass continuity, equation (3.7), and the isothermal equation of state (9.2). Neglecting gravity, the momentum equation for the neutral component is

$$\rho\,\frac{D\boldsymbol{u}}{Dt} \,=\, -a_T^2\,\boldsymbol{\nabla}\rho \,-\, n_i\,\boldsymbol{f}_{\rm in}\ . \tag{10.11}$$

Here the drag force $\boldsymbol{f}_{\rm in}$ is identical to $\boldsymbol{f}_{\rm in}'$ in equation (9.43), since we can ignore the relative *acceleration* (as opposed to the velocity) of the ion-electron plasma. The equation of motion for this plasma, from (10.2), is

$$\begin{aligned} 0 &= \boldsymbol{j}\times\boldsymbol{B}/c \,+\, n_i\,\boldsymbol{f}_{\rm in} \\ &= \frac{1}{4\,\pi}\,(\boldsymbol{\nabla}\times\boldsymbol{B})\times\boldsymbol{B} \,+\, n_i\,\boldsymbol{f}_{\rm in}\ , \end{aligned} \tag{10.12}$$

where we have ommitted the relatively small electron-neutral drag. Finally, the magnetic field obeys the MHD equation, in the form (10.4).

After applying the perturbations and subtracting off the equilibrium conditions, mass continuity and the equation of state again follow equations (9.62) and (9.20b), respectively. Substitution of (9.43) for $\boldsymbol{f}_{\rm in}$ into (10.11) leads to the perturbed neutral momentum equation:

$$-i\,\omega\,\rho_\circ\,\delta\boldsymbol{u} \,=\, -i\,a_T^2\,\delta\rho\,\boldsymbol{k} \,+\, (n_i\,\rho)_\circ\,\langle\sigma_{\rm in}u_i'\rangle\,(\delta\boldsymbol{u}_i \,-\, \delta\boldsymbol{u})\ . \tag{10.13}$$

For the product of the ion and neutral densities, we retain only its equilibrium value $(n_i\,\rho)_\circ$, as its perturbation generates terms of order δ^2. To derive the perturbed momentum equation for the plasma, we again write $\boldsymbol{f}_{\rm in}$ in terms of velocities, obtaining

$$0 \,=\, \frac{1}{4\,\pi}\,(i\,\boldsymbol{k}\times\delta\boldsymbol{B}\times\boldsymbol{B}_\circ) \,-\, (n_i\,\rho)_\circ\langle\sigma_{\rm in}u_i'\rangle\,(\delta\boldsymbol{u}_i \,-\, \delta\boldsymbol{u})\ . \tag{10.14}$$

Finally, application of the perturbation to the first form of equation (10.4) leads to

$$-i\,\omega\,\delta\boldsymbol{B} \,=\, i\,\boldsymbol{k}\times(\delta\boldsymbol{u}_i\times\boldsymbol{B}_\circ)\ . \tag{10.15}$$

If we dot the last equation with $\boldsymbol{k}$, we again see that $\delta\boldsymbol{B}\cdot\boldsymbol{k}=0$. Moreover, expansion of the triple product leads to the analogue of equation (9.66):

$$-\omega\,\delta\boldsymbol{B} \,=\, (\boldsymbol{k}\cdot\boldsymbol{B}_\circ)\,\delta\boldsymbol{u}_i \,-\, (\boldsymbol{k}\cdot\delta\boldsymbol{u}_i)\,\boldsymbol{B}_\circ\ . \tag{10.16}$$

Rather than consider the most general wave mode, we specialize immediately to a "pure" Alfvén wave, in which $\boldsymbol{B}_\circ$ is parallel to $\boldsymbol{k}$, while both $\delta\boldsymbol{u}$ and $\delta\boldsymbol{u}_i$ are perpendicular. Under these conditions, equation (10.16) leads to a relation between the amplitudes δu_i and δB:

$$\delta u_i \,=\, -\frac{\omega\,\delta B}{k\,B_\circ}\ . \tag{10.17}$$

The equation of state and mass conservation together imply that $\delta P = \delta\rho = 0$ for this transverse mode. Hence, equation (10.13) becomes

$$-i\,\omega\,\rho_\circ\,\delta u \,=\, (n_i\,\rho)_\circ\,\langle\sigma_{\rm in}u_i'\rangle\,(\delta u_i \,-\, \delta u)\ , \tag{10.18}$$

which we recast as

$$\delta u \,=\, \delta u_i \left(1 \,-\, \frac{i\,\omega}{n_i\langle\sigma_{\rm in}u_i'\rangle}\right)^{-1}\ . \tag{10.19}$$

We recognize $n_i\langle\sigma_{\rm in}u_i'\rangle$ as the frequency with which a given neutral atom or molecule is struck by ions. Equation (10.19) shows that the neutral velocity amplitude δu only equals δu_i for ω much smaller than this collisional frequency, and that δu falls as ω climbs much higher. The fact that the denominator in equation (10.19) is complex means that the neutral and ion velocities also differ in phase.

Turning to the perturbed plasma momentum equation (10.14), we expand the triple product and use equation (10.18) to write the ion-neutral velocity difference in terms of δu alone. Algebraic manipulation then leads to

$$\delta u_i \,=\, -\left(1 \,-\, \frac{i\,\omega}{n_i\langle\sigma_{\rm in}u_i'\rangle}\right)\frac{k\,B_\circ\,\delta B}{4\,\pi\,\omega\,\rho_\circ}\ . \tag{10.20}$$

Equations (10.17) and (10.20) are two expressions for δu_i in terms of δB. After eliminating δu_i and canceling δB in the resulting equation, we arrive at the desired dispersion relation:

$$\frac{\omega^2}{k^2} \,=\, \frac{B_\circ^2}{4\,\pi\,\rho_\circ}\left(1 \,-\, \frac{i\,\omega}{n_i\langle\sigma_{\rm in}u_i'\rangle}\right)\ . \tag{10.21}$$

The new relation is a clear generalization of equation (9.69) and demonstrates how the latter is only valid at relatively low frequencies. If we let k be a real number, then equation (10.21) implies that ω is complex. Writing $\omega \,=\, \omega_R \,+\, i\omega_I$, where ω_R and ω_I are both real, we have

$$\exp[i(\boldsymbol{k}\cdot\boldsymbol{r} \,-\, \omega\,t)] \,=\, \exp(\omega_I\,t)\ \exp[i(\boldsymbol{k}\cdot\boldsymbol{r} \,-\, \omega_R\,t]\ .$$

The perturbation consists of a plane wave that travels at phase velocity ω_R/k and has an amplitude that increases or decays exponentially in time, depending on the sign of ω_I.

These considerations motivate us to solve equation (10.21) explicitly for ω. Introducing the Alfvén velocity V_A via equation (9.70), we recast (10.21) as a quadratic equation in ω:

$$\omega^2 \,+\, \frac{i\,V_A^2\,k^2}{n_i\langle\sigma_{\rm in}u_i'\rangle}\,\omega \,-\, V_A^2\,k^2 \,=\, 0\ .$$

We thus obtain

$$\omega \,=\, -\frac{i\,V_A^2\,k^2}{2\,n_i\langle\sigma_{\rm in}u_i'\rangle} \,\pm\, \frac{V_A\,k}{2}\sqrt{4 \,-\, \frac{V_A^2\,k^2}{n_i^2\langle\sigma_{\rm in}u_i'\rangle^2}}\ . \tag{10.22}$$

In order for the perturbation to have any propagating component, the quantity inside the square root must be positive. If we rewrite this condition in terms of the wavelength $\lambda \equiv 2\pi/k$, we find that propagation requires $\lambda > \lambda_{\rm min}$, where

$$\begin{aligned}\lambda_{\rm min} &\equiv \frac{\pi\, V_A}{n_i \langle \sigma_{\rm in} u_i' \rangle} \\ &= 0.06\,{\rm pc} \left(\frac{B_\circ}{10\,\mu{\rm G}}\right) \left(\frac{n_{\rm H_2}}{10^3\ {\rm cm^{-3}}}\right)^{-1} \ .\end{aligned} \tag{10.23}$$

The critical length $\lambda_{\rm min}$ is intriguingly close to the observed dimensions of dense cores. This result lends support to the idea that the cores form in relatively quiescent, wave-free environments, but does not by itself aid in understanding the formation process. Note also that even waves with $\lambda < \lambda_{\rm min}$ have changing amplitudes. Inspection of equation (10.22) reveals that, in this case, both ω-values are negative and imaginary, corresponding to decay. The characteristic damping time is the reciprocal of $|\omega_I|$, which is roughly

$$\begin{aligned}\tau_{\rm damp} &\approx \frac{n_i \langle \sigma_{\rm in} u_i' \rangle}{V_A^2\, k^2} \\ &= 1 \times 10^4\,{\rm yr} \left(\frac{\lambda}{0.06\ {\rm pc}}\right)^2 \left(\frac{B_\circ}{10\,\mu{\rm G}}\right)^{-2} \left(\frac{n_{\rm H_2}}{10^3\ {\rm cm^{-3}}}\right)^{3/2} \ .\end{aligned} \tag{10.24}$$

We see that waves with λ up to about $30\,\lambda_{\rm min}$ must be periodically regenerated if they are to survive for parent cloud lifetimes of order 10^7 yr.

10.2 Inside-Out Collapse

One major difficulty in modeling gravitational collapse is that the process spans an enormous range of distances. Physical understanding is best achieved by considering regimes where different effects predominate. Thus, the interstellar magnetic field strongly influences cloud morphology on the largest relevant scales ($L \sim 10^{17} - 10^{18}$ cm). Much closer to the relatively tiny protostar, *i. e.*, for $L \sim 10^{11} - 10^{14}$ cm, cloud material is moving at high speed, but is being diverted by the centrifugal force (§ 10.4). In between lies the regime where gas is being released from thermal and magnetic support, and is entering a state of free fall. The collapse here has an *inside-out* (or *nonhomologous*) character. That is, the region of infall slowly spreads into static, more rarefied gas.

10.2.1 The Spherical Problem

We may elucidate this spreading process, as well as other key elements of collapse, by focusing on a highly simplified model. Ignoring momentarily both magnetic and rotational effects, let us follow the evolution of a perfectly spherical cloud supported solely by thermal pressure. It goes without saying that such objects exist only in the theorist's imagination. But we can learn much by probing their dynamics, just as we did through analysis of their hydrostatic structure and stability in § 9.1.

Within a spherical cloud, a fluid element at radial distance r from the origin only senses the gravitational force from the mass interior to it. Suppose we designate this interior mass as M_r, a function of both r and t. That is, we define

$$M_r \equiv \int_0^r 4\pi r^2 \rho \, dr \ . \tag{10.25}$$

Our goal is to use M_r as a dependent variable in a system of differential equations. From (10.25), it follows that

$$\frac{\partial M_r}{\partial r} = 4\pi r^2 \rho \ . \tag{10.26}$$

To evaluate $\partial M_r/\partial t$, we differentiate under the integral sign in equation (10.25) and utilize the spherical version of the continuity equation (3.7):

$$\frac{\partial \rho}{\partial t} = -\frac{1}{r^2}\frac{\partial(r^2 \rho\, u)}{\partial r} \ , \tag{10.27}$$

to find

$$\frac{\partial M_r}{\partial t} = -4\pi r^2 \rho\, u \ . \tag{10.28}$$

Finally, we need the full momentum equation (3.3) with zero magnetic field. Adopting the usual isothermal approximation and noting that the gravitational force per unit mass is $-GM_r/r^2$, we have

$$\frac{\partial u}{\partial t} + u\,\frac{\partial u}{\partial r} = -\frac{a_T^2}{\rho}\frac{\partial \rho}{\partial r} - \frac{G\,M_r}{r^2} \ . \tag{10.29}$$

For a given set of initial and boundary conditions, we may numerically solve equations (10.26)–(10.29) for the variables M_r, ρ, and u as functions of r and t. Note that positive u-values indicate expansion, and negative ones contraction. Choosing the initial state to be one of the spherical equilibria of § 9.1 would only give the uninteresting result that $u = 0$ for all time, *i. e.*, the cloud would simply remain in force balance. Of course, we can always specify a cloud that is slightly perturbed from equilibrium, *e. g.*, one with small (subsonic) velocities throughout its interior. If the equilibrium were stable, the cloud would proceed to oscillate in a superposition of normal modes. If unstable, the object would either collapse or disperse. An unstable initial state is difficult to realize in practice, since such a cloud could not have existed for any substantial time *prior* to its collapse. Although it is instructive to consider a range of starting states, the most relevant one physically is the equilibrium model that is *marginally stable*. Within the sequence presented in § 9.1, this is the cloud with the Bonnor-Ebert mass, given by equation (9.16).

The simplest and most widely used boundary conditions are either *constant pressure* at the edge or *constant volume*. In the first case, the cloud shrinks in time to compensate for any density (and therefore pressure) drop in its outer region. In the second, one sets the fluid velocity to zero at a *fixed* radius, where the density and pressure do eventually fall. Neither choice is very satisfactory. The treatment of the boundary should ideally reflect physical conditions beyond the realm of the calculation. Since there is no adequate way to describe, in a spherical model, the anisotropic force of magnetic support or its loss through ambipolar diffusion, one is forced to these rather *ad hoc* prescriptions. However, once the inside-out collapse is established, the boundary condition has little effect until the end of the calculation.

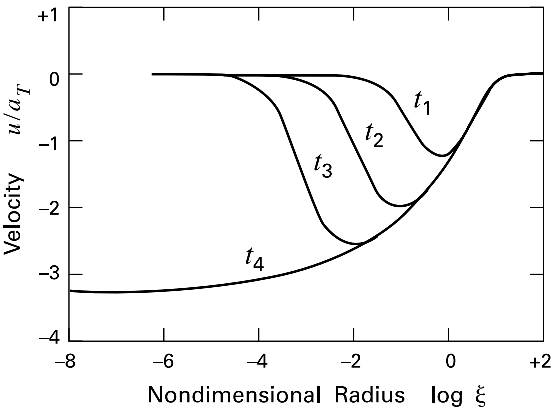


Figure 10.5 Velocity profiles in a collapsing, isothermal sphere, shown as a function of nondimensional radius ξ. The four times, measured in units of $t_{\rm ff}$ and relative to the instant of protostar formation, are $t_1 = -0.0509$, $t_2 = -0.0026$, $t_3 = -0.0001$, and $t_4 = 0$. Here, $t_{\rm ff}$ is the free-fall time associated with the cloud's initial central density.

On the other hand, numerical results for the earliest phase *are* heavily influenced by both initial and boundary conditions. Figure 10.5 shows the evolution of the velocity $u(r)$, as obtained in a calculation of the constant-pressure type. Notice that the radial distance is the nondimensional one defined by equation (9.6). The initial configuration is the marginally stable (Bonnor-Ebert) state, but with the density increased everywhere by 10 percent. The cloud is therefore too massive to be in force balance and begins to collapse. Starting at the outer edge, the velocity profile grows until every mass shell is accelerating inward. The figure tracks this change over a time interval of 0.05 $t_{\rm ff}$, where $t_{\rm ff}$ is the free-fall time associated with the cloud's initial central density (recall equation (3.15)). At the last instant shown, 44 percent of the cloud mass is moving supersonically.

This rapid, inward motion leads to shock formation and strong compression of the central region. Following this process is technically challenging and demands continual refinement of the spatial grid used in the numerical computation. An alternative is simply to collect all infalling matter into a central "sink cell." One keeps track only of the mass in this cell, recognizing that a separate, interior calculation is needed to resolve the details within this small volume. The increasing mass accurately represents the total being accumulated by the protostar, even though the size of the cell vastly exceeds stellar dimensions.

10.2.2 Mass Accretion Rate

A quantity of prime importance, then, is the increase per unit time of the sink-cell mass. This *mass accretion rate*, denoted $\dot{M}$, largely determines the properties of the growing protostar (Chapter 11). Mathematically, we want the inner limit of $\partial M_r/\partial t$:

$$\dot{M} \equiv \lim_{r \rightarrow 0} -4\pi r^2 \rho u \ . \tag{10.30}$$

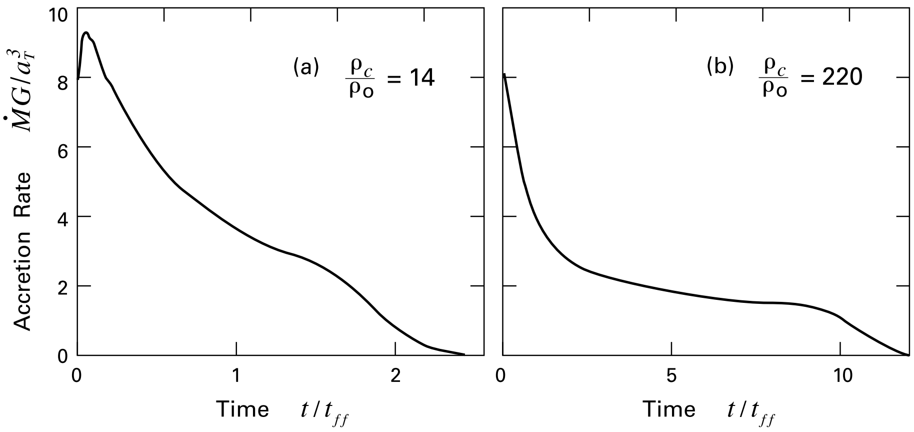


Figure 10.6 Evolution of mass accretion rate in two collapsing, isothermal spheres. Time is measured after the start of protostar formation, and is again in units of $t_{\rm ff}$. Each cloud has the initial density contrast indicated.

Figure 10.6a shows the evolution of $\dot{M}(t)$ for the same calculation as the previous figure, but for times *after* protostar formation. Notice that the rate is written in units of a_T^3/G, which is the only combination of relevant constants with the proper dimensions. The time is again relative to $t_{\rm ff}$ at the initial central density.

Soon after the protostar forms, $\dot{M}$ is relatively high, as the material previously set into motion reaches the center. The rate then quickly falls and starts to level off before it abruptly drops to zero. This final truncation represents the collapse of the cloud boundary into the sink cell. Since the surrounding matter is assumed to have zero density, $\dot{M}$ vanishes. Such a prescription for the environment is as unrealistic here as it was for magnetized configurations. In both cases, there is no reason why additional mass should not be drawn in by the falling pressure.

Just prior to the termination of collapse, the function $\dot{M}$ is approaching

$$\dot{M} = m_\circ \frac{a_T^3}{G} , \tag{10.31}$$

where $m_\circ$ is a number of order unity. This asymptotic behavior is highlighted if we start with a more centrally condensed equilibrium model. The curve in Figure 10.6b shows $\dot{M}(t)$ for an initial state with density contrast $\rho_c/\rho_\circ = 220$. Since the central $t_{\rm ff}$ is now shorter, the elapsed nondimensional time $t/t_{\rm ff}$ increases. The accretion rate still drops rapidly at first, but maintains a modest $m_\circ$-value over a longer period before the final truncation. The more condensed cloud is dynamically unstable, so we must view this result with caution. Nevertheless, the comparison demonstrates that, while the *initial* behavior of $\dot{M}$ is sensitive to the precise starting configuration (including the manner of its perturbation), the final leveling off is not.

Many other calculations, although differing in detail, give essentially the same result. For example, one may start with a *uniform*-density sphere. If one applies a constant-volume boundary condition, the density just inside the rigid outer surface immediately drops. The consequent fall in pressure retards the flow of gas in this region, while the density at the center builds up more quickly. In this manner, the cloud attains a peaked density profile similar to the marginally stable case, but bypasses the large velocities prior to protostar formation. The mass accretion rate then settles to a roughly steady value without the early, transient burst.

How, then, are we to interpret equation (10.31)? Once the protostar forms, nearby gas is in a state of free fall. That is, the retarding effect of thermal pressure is small compared to gravity, and the speed of a fluid element is nearly equal to $V_{\rm ff} \equiv (2GM_*/r)^{1/2}$. Here, M_* is the growing protostellar mass, equal to $\dot{M}t$ for a steady accretion rate.[1] With increasing distance from the protostar, the gravitational force weakens until the effect of pressure becomes appreciable. The transition to free-fall collapse thus occurs at the radius $R_{\rm ff}$ where $V_{\rm ff} \approx a_T$, *i. e.*, where $G\dot{M}t/R_{\rm ff} \approx a_T^2$. We may rewrite this latter condition as

$$\dot{M} \approx \frac{a_T^2\,\dot{R}_{\rm ff}}{G} \ , \tag{10.32}$$

where we have used $R_{\rm ff}/t \approx dR_{\rm ff}/dt \equiv \dot{R}_{\rm ff}$.

From this last equation, we deduce that $\dot{R}_{\rm ff} > 0$ as long as $\dot{M} > 0$. That is, the region of free fall spreads in time as the mass and gravitational influence of the protostar increase. Beyond the infalling region, the cloud is still in hydrostatic balance. A given mass shell only starts to fall after it loses pressure support. Once this occurs, inward motion of the shell weakens its support of the next shell beyond it. The initiation of collapse through pressure erosion is thus a progressive, wave-like phenomenon (see Figure 10.7). Indeed, the boundary of the infall region constitutes a *rarefaction wave*, a phenomenon encountered generally in fluid dynamics. Both rarefaction and sound waves represent the propagation of small disturbances in the ambient pressure, and both travel at the sound speed, here a_T. Setting $\dot{R}_{\rm ff}$ equal to a_T in equation (10.32), we recover equation (10.31) for $\dot{M}$.

10.2.3 Thermal Effects

Equation (10.31) states that, to within a factor of order unity, the asymptotic accretion rate depends only on the ambient temperature. To some extent, this simple result reflects our omission of magnetic support prior to collapse. However, the effect of the additional force is subtle. As we shall discuss in § 10.3, the cloud portion that finally collapses is that which was *not* magnetically supported, and equation (10.31) remains our best quantitative estimate. This important relation thus sets the basic time scale for the protostar phase. Supplying numerical values, we have

$$\dot{M} \approx 2 \times 10^{-6}\, M_\odot\ {\rm yr}^{-1} \left(\frac{T}{10\ {\rm K}}\right)^{3/2} \ . \tag{10.33}$$

Thus a protostar of 1 $M_\odot$ accumulates its mass in about 5×10^5 yr. Within the scope of stellar evolution, this period is exceedingly brief, even compared to the 3×10^7 yr pre-main-sequence contraction time for a star of the same mass.

---

[1] The free-fall velocity $V_{\rm ff}$ should not be confused with the similar virial velocity $V_{\rm vir}$, introduced in equation (3.20). The former refers to a test particle falling toward a *point mass*, while the latter is the characteristic speed within a *distributed volume* of self-gravitating gas.

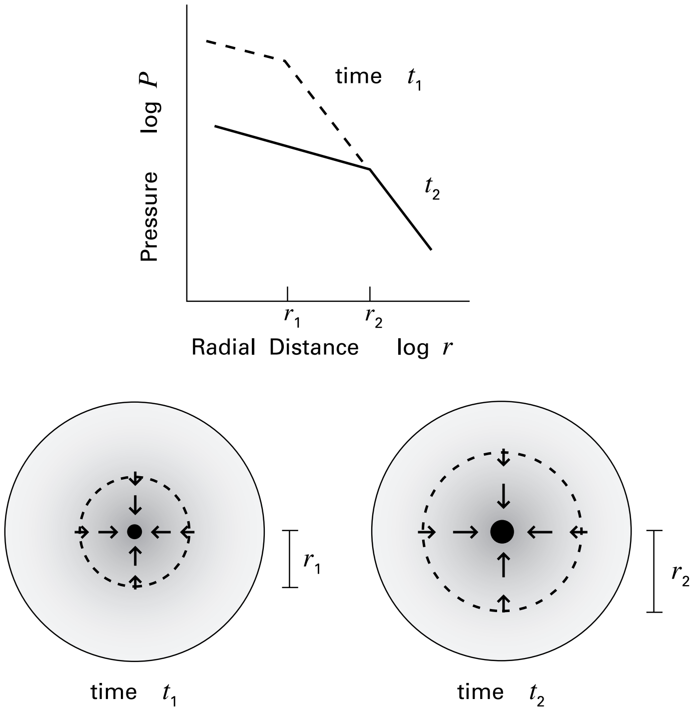


Figure 10.7 Rarefaction wave in inside-out collapse (*schematic*). An interior region of diminished pressure advances from radius r_1 at time t_1 to r_2 at t_2. Within this region, gas falls onto the central protostar of growing mass.

The freely falling portion of the cloud has a structure determined by the strong gravity of the protostar, rather than by conditions prior to collapse. If we focus on some fixed volume relatively close to the star, gas crosses this region in an interval brief compared with the evolutionary time scale, $M_*/\dot{M}$. Since there is no chance for appreciable mass buildup in any such volume, we may ignore the left-hand side of the continuity equation (10.27) and conclude that $r^2 \rho u$ is a spatial constant throughout the collapsing interior. Setting $u = -V_{\rm ff}$ and utilizing equation (10.30), we solve for the density to find

$$\rho = \frac{\dot{M}\, r^{-3/2}}{4\pi\, \sqrt{2GM_*}} \ . \qquad (10.34)$$

For comparison, we recall from § 9.1 that all spherical *equilibria* have ρ declining as r^{-2} in their outer regions. Thus, both the density and pressure profiles become flatter in the region of collapse, as illustrated in Figure 10.7.

The numerical calculations we have discussed assume spherical symmetry not only in the initial configuration, but at all subsequent times. It is not difficult to relax the second restriction. That is, one still begins with a spherical, thermally supported cloud, but now follows its collapse with the full, three-dimensional equations of mass continuity (3.7) and momentum conservation

(3.3), together with Poisson's equation (9.3) for the gravitational potential. The result is that the cloud evolution is virtually unchanged. In three dimensions, small, localized density enhancements inevitably arise. Once these enter the collapsing region, however, the straining motion induced by the protostar's gravity (*i. e.*, the increase of $|V_{\rm ff}|$ with decreasing r) tends to pull them apart. Inside-out collapse is thus stable against *fragmentation*. The situation is quite different for clouds that are initially far out of force balance, as we discuss in Chapter 12.[2]

One reason that pressure is ultimately ineffective in halting collapse is that the gas temperature has been assumed constant. Building up an adverse pressure gradient thus requires a steep inward rise of density. High density, on the other hand, only enhances the effect of self-gravity. It is for this reason, of course, that isothermal equilibria can only tolerate a modest density contrast before they are unstable to collapse.

How realistic, though, is the isothermal assumption? In the case of hydrostatic configurations, we have seen that the temperature responds rather sluggishly to cosmic-ray heating because of efficient cooling by CO and dust grains (recall Figure 8.6). A fluid element within a collapsing cloud has two additional sources of energy input. One is the compressional work performed by the surrounding gas. Here, the power input per unit volume is

$$\begin{aligned}\Gamma_{\rm comp} &= \frac{P}{\rho}\,\frac{D\rho}{Dt} \\ &= \frac{P\,u}{\rho}\,\frac{\partial\rho}{\partial r}\ ,\end{aligned} \qquad (10.35)$$

where we have assumed steady-state flow in the second form of this relation.

Suppose we now utilize $\rho(r)$ from equation (10.34) to evaluate $\Gamma_{\rm comp}$. Then, at radii where this rate is appreciable, we find it is overwhelmed by the second new source of energy, the radiation from the protostar and its surrounding disk. This luminosity stems from the kinetic energy of infall and is generated at the stellar and disk surfaces (Chapter 11). It is the dust grains within the flow that are actually irradiated and they respond, as usual, by emitting their own infrared photons. The temperature of the infalling envelope does not climb steeply until the ambient density is large enough to trap this cooling radiation. As we will see in § 11.1, such trapping occurs at a radius of roughly 10^{14} cm. The gas at this distance is already traveling at such high speed that the infall process cannot be impeded. Thus, the departures from isothermality, while both interesting physically and critical observationally, do not affect the overall dynamics of inside-out collapse.

10.3 Magnetized Infall

Within the sequence of spherical equilibria discussed in § 9.1, there is only one marginally unstable model. We have argued that a dense core approaches the point of collapse quasi-statically, *i. e.*, without ever being far removed from force balance. The implication is that the object *must*

---

[2] Detailed analysis reveals that nonspherical perturbations within the free-fall region *do* grow, but very weakly. The density contrast over the background increases as $(t_\circ - t)^{1/3}$, where $t_\circ$ is the time at which the unperturbed fluid element would reach the origin.

eventually reach the unique, Bonnor-Ebert configuration just before it collapses. This statement is true regardless of the core's prior history. Unfortunately, this simplicity vanishes with the addition of a magnetic field.

10.3.1 Origin of Dense Cores

The reason for the new complication, from a mathematical perspective, is that an additional function, $dM/d\Phi_B$, now enters the equations. Instead of one critical state, we now have an infinite number, each characterized by its own flux distribution. Since, moreover, $dM/d\Phi_B$ changes as a result of ambipolar diffusion, the problem of selecting the correct marginally stable state cannot be separated from the issue of prior history. Strictly speaking, one must trace the origin of dense cores in order to understand their collapse.

Stated in this manner, the collapse problem sounds virtually intractable. There is little observational evidence bearing on dense core origins, although the current millimeter-wavelength maps of low-mass, starless configurations are clearly a step in the right direction. On the theoretical front, we will need to understand better the mechanical support provided by MHD waves before we can calculate the growth of quiescent substructures. The word "quiescent" is the operative one here. A viable theory must explain the appearance and survival until collapse of regions that are distinct, in this important sense, from their environment.

As an aside, we note that computer simulations of fully turbulent clouds, mentioned briefly at the end of Chapter 9, often find clumpy substructure. In a typical calculation, one again imposes a random velocity field on an initially uniform gas. Self-gravity is at first omitted from the simulation. Unless the stochastic velocity disturbance is maintained continually, the turbulence rapidly decays. If it persists, density inhomogeneities develop throughout the magnetized fluid.

At this point, self-gravity is introduced within the dynamical equations. Compressed sites immediately go into collapse. That is, their density quickly climbs, as their size shrinks below the spatial resolution of the code. A snapshot of the computational box shows a collection of small lumps within a turbulent background. The visual impression is similar to what one thinks may be occurring in a cluster-forming molecular cloud.

Even the prolate structure of observed cores can be obtained in these numerical studies. However, the resemblance is only superficial. The simulated entities are fully dynamical from the start. That is, they have internal velocities comparable to the local free-fall value. To what state are they collapsing? With even higher resolution, we would witness opposing gas streams colliding and shocking. The resulting cooling would indeed lead to denser structures. But the buildup of a true star, with its *vastly* higher density, requires inside-out collapse, which we do not expect under these circumstances.

10.3.2 Flux Loss During Collapse

Return now to the case of a structure initially closer to force balance. Its evolution must embody certain key features. One is ambipolar diffusion. This drift of cloud matter across field lines sets the time scale for contraction, allowing gravity to compress the gas without excessive buildup of opposing magnetic forces. The inevitable result is that the function $dM/d\Phi_B$ rises near the central axis ($\Phi_B = 0$), regardless of its initial form. This development proceeds until the cloud becomes gravitationally unstable, and it continues during the subsequent collapse. The actual form of $dM/d\Phi_B$ at the point of marginal stability is hardly likely to be the illustrative one in equation (9.53). If the current numerical models were correct and dense cores become increasingly flatter structures, then equation (9.59) for thin slabs would apply. The sequence in Figure 10.3 indeed exhibits the onset of rapid, central contraction once $dM/d\Phi_B$ at the center (which equals Σ_c/B_c) approaches the limit in equation (9.59). The thin-slab approximation is not valid, however, for a more spherical or prolate configuration. Here, $M \gtrsim M_{\rm BE}$ prior to collapse, so that equality of M and M_Φ does not mark the stability transition.

Another general result from theory is that the nature of the support against gravity varies in different regions of the cloud. Thermal pressure is strongest in the dense, central part, where the magnetic flux has diffusively leaked out. In the farthest reaches of the cloud, *i. e.*, at distances greater than $\lambda_{\rm min}$ in equation (10.23), Alfvén-wave support is significant, while ordinary gas pressure is not. Also important here are tension and pressure from the static, ambient field. This static force must predominate throughout much of the intermediate region.

Finally, the collapse itself should proceed in an inside-out fashion, at least after the formation of the central protostar. The reason is that the force of gravity toward the star increases too steeply to be opposed by thermal pressure. Thus, as we have seen, the growth in stellar mass must be accompanied by a spreading of the infall region. This spreading sets into collapse that part of the cloud originally supported by pressure.

Figure 10.8 sketches a conception of dense core evolution that incorporates these three basic elements. Panel (a) depicts a region of relatively uniform gas and magnetic field, presumably the environment out of which the cores originate. Any incipient density enhancement, such as the one shown, attracts additional mass through its gravity. The incoming matter either diffuses across the field or else slides down it. Both types of motion occur. However, the gas will tend to accumulate along the field, *i. e.*, in the direction where it meets the least resistance. Thus, it forms the elongated structure shown in panel (b). Meanwhile, the cross-field drift also distorts $\boldsymbol{B}$, in the manner depicted. This distortion occurs through the collisional drag exerted on the gyrating ions and electrons by the inflowing neutral gas, and results in a buildup of magnetic tension.

As the central density grows, the field continues to pinch together until it reaches a configuration like the one sketched in panel (c). Here we are approaching a "split monopole" structure. The field lines diverge nearly radially from the center, but their direction reverses across the equatorial plane. Thus, the net flux through any surface enclosing the origin still vanishes, as required by $\boldsymbol{\nabla} \cdot \boldsymbol{B} = 0$. The split monopole has effectively divided the cloud into two types of regions. In the pinched columns above and below the center (labeled $\mathcal{A}$ and $\mathcal{A}'$ in the figure), the pull of gravity is counteracted almost entirely by the thermal pressure gradient. In the extended equatorial zone $\mathcal{B}$, which wraps around the axis by azimuthal symmetry, magnetic tension and pressure take up this role, with the former increasing toward the origin. Finally, wave support dominates in the outermost regions, not depicted here.

In the course of time, the columns $\mathcal{A}$ and $\mathcal{A}'$ grow as mass settles from the more turbulent, wave-supported exterior. Rapid contraction, and ultimately collapse, begin once the linear extent of a column exceeds the Jeans length λ_J, evaluated at the unperturbed density and temperature. As the protostellar mass grows, the zone of infall works its way out in the columns,

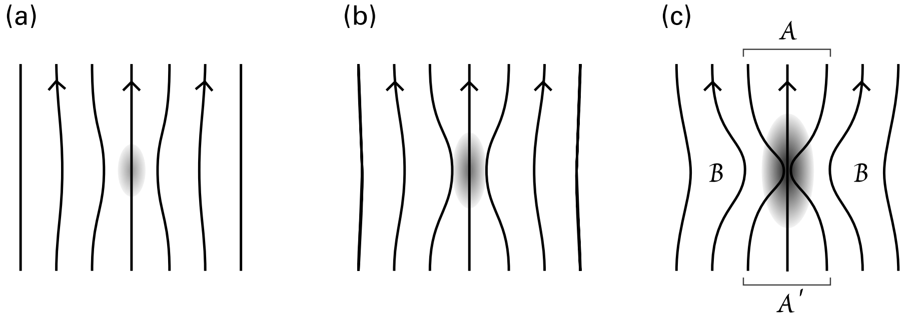


Figure 10.8 Origin of dense cores (*schematic*). A small region of enhanced density slowly grows by accruing external matter, simultaneously pulling in the ambient magnetic field. At the last epoch shown, severe field pinching has created two columns separated from a broad equatorial zone.

with the front of the associated rarefaction wave moving at the local sound speed.[3] Thus, equation (10.31) is still a reasonable approximation for the mass accretion rate $\dot{M}$, despite the fact that the geometry is substantially altered from spherical. The protostar's gravity also pulls on the matter in the equatorial region. This tug can be resisted in part by increased field curvature and tension, so that inward mass flow is not as large as from the two columns.

Existing collapse calculations generally start with unstable clouds of restricted mass that quickly flatten, as exemplified by Figure 10.3. Such prompt, global collapse to the equatorial plane is quite different from the sequence we have sketched, which involves slow buildup along field lines followed by inside-out collapse. Nevertheless, these more detailed studies continue to provide insight. One point, already mentioned, is that ambipolar diffusion must continue even during the collapse. Figure 10.9 shows the evolution of $M(\Phi_B)$ from another numerical investigation of a collapsing cloud. The initial flux distribution corresponds to equation (9.53), *i. e.*, it represents a uniform field threading a spherical cloud. As the collapse proceeds, $M(\Phi_B)$ increases for relatively small Φ_B-values. That is, the innermost flux tubes gain mass at the expense of the outer ones. Notice that the flux here is measured relative to the total cloud value $\Phi_{\rm tot}$, which itself diminishes with time, as sketched in Figure 10.2.

Within the framework of our evolutionary picture, ambipolar diffusion occurs chiefly in the equatorial region $\mathcal{B}$. One should also bear in mind that Figure 10.9 was obtained by assuming that the electron fraction everywhere declines as n_H^{-1}. As we saw in Chapter 8, such a falloff only occurs at relatively high densities, where ions and electrons recombine on grain surfaces. For lower densities, the $n_H^{-1/2}$ falloff in equation (8.32) applies, and ambipolar diffusion is less efficient. The flux in the outermost part of the equatorial zone cannot decrease greatly during collapse, even though diffusion proceeds rapidly in the deeper interior.

To quantify matters, we may compare the magnitude of $\boldsymbol{v}_{\rm drift}$ with the local speed of the neutrals. Once collapse is underway and the protostar has formed, we may crudely approximate the latter as the free-fall speed $V_{\rm ff}$. We assume, moreover, that the magnetic force is comparable (though less than) gravity. Note that the first approximation becomes more accurate as we

---

[3] Strictly speaking, the traveling front is an MHD wave in this case. Since the disturbance is longitudinal and $\boldsymbol{k}$ is nearly parallel to $\boldsymbol{B}$, the phase velocity is close to a_T, as in equation (9.73).

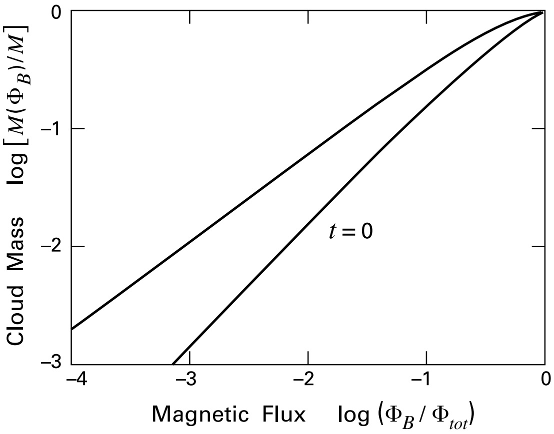


Figure 10.9 Magnetic flux loss in a collapsing cloud. As in Figure 10.2, the mass enclosed within a given flux increases. The total flux threading the cloud, $\Phi_{\rm tot}$, diminishes with time.

approach the star. On the other hand, the second assumption eventually breaks down in this same limit, so we are restricted to some intermediate region of the cloud. Equations (10.3) and (10.10) then imply that

$$\begin{aligned}\frac{|\boldsymbol{v}_{\rm drift}|}{V_{\rm ff}} &\approx \frac{V_{\rm ff}}{2\,n_e\langle\sigma_{\rm in}u_i'\rangle\varpi}\\ &= 0.02\left(\frac{n_H}{10^4\ {\rm cm}^{-3}}\right)^{-1/2}\left(\frac{\varpi}{0.1\ {\rm pc}}\right)^{-3/2}\left(\frac{M_*}{1\ M_\odot}\right)^{1/2}\end{aligned} \qquad (10.36)$$

within the appropriate region. Here we have substituted M_* for $M_{\rm inner}$ in equation (10.10), and have replaced r by ϖ in the definition of $V_{\rm ff}$. The numerical estimate further employs the ionization law of equation (8.32).

Equation (10.36) confirms that $|\boldsymbol{v}_{\rm drift}|/V_{\rm ff}$ is less than unity at large distances. Here, where wave-like motion is still prevalent, the matter and field are tied together. However, the velocity ratio is appreciable even in this region and increases as the gas approaches the star, since the product $n_H\varpi^3$ declines. Once the two speeds become comparable, the neutral material effectively decouples from the field. In the absence of a reliable density distribution in the equatorial region, it is difficult to be more quantitative. However, the conclusion seems inescapable that the contracting gas in this zone soon leaves the magnetic field behind. [4]

10.3.3 Magnetic Reconnection

Despite this slippage, the infalling matter retains *some* magnetization. What becomes of this residual field? Thus far, our discussion has neglected the role of Ohmic dissipation. That is, we have continued to assume that the ion-electron plasma acts like a perfectly conducting fluid, even as this fluid and its embedded field move relative to the neutral matter. This assumption,

---

[4] This slippage occurs even *before* a true protostar forms. As soon as the cloud becomes substantially condensed, the interior fluid velocity is approximately $V_{\rm ff}$, where the relevant mass is that of the central lump. Such early loss of magnetic support is consistent with the decoupling of Alfvén waves over distances comparable to the sizes of dense cores; recall equation (10.23).

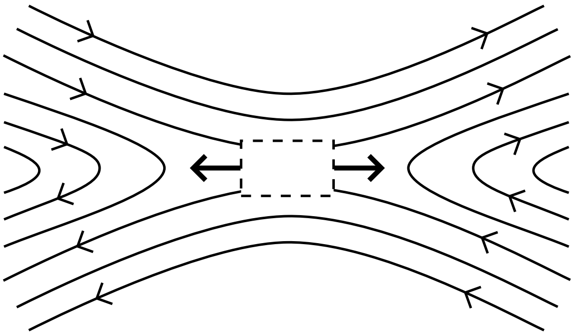


Figure 10.10 Topology of magnetic reconnection (*schematic*). Field lines of opposite direction press together, creating a small region (*dashed rectangle*) of high Ohmic dissipation. Pressure on this region expels fluid laterally, as indicated.

too, eventually breaks down. At sufficiently high densities, the field being dragged in undergoes *magnetic reconnection*. This process effectively destroys magnetic flux before it can be transported onto the star and its disk.

Reconnection occurs whenever field lines of opposite direction are pressed together. Historically, the phenomenon was first invoked to explain the large luminosities seen in impulsive solar flares. Here, as in all reconnection events, the outburst represents energy liberated from the magnetic field. Since $\boldsymbol{B}$ reverses sign, its magnitude must go through zero (see Figure 10.10). The magnetic pressure thus has a local minimum, and the field on either side of the interface presses inward. Concurrently, fluid is expelled sideways, as indicated in the figure. The field lines then crowd in more tightly, and the local gradient continues to rise. Magnetic energy is dissipated as heat within a restricted region, indicated by the dashed rectangle in the figure. Here, antiparallel lines annihilate one another, changing the field topology. From a mathematical perspective, the final, Ohmic term in the MHD equation (9.39) increases until it becomes dominant, regardless of the specific conductivity value.

These considerations are clearly applicable to protostellar collapse, where any residual field eventually develops large gradients. Figure 10.8, for example, indicates that the equatorial $\boldsymbol{B}$ becomes severely pinched as matter flows toward the protostar. Such pinching cannot continue indefinitely without bringing opposing field lines so close that reconnection ensues. The resulting change in topology is shown in Figure 10.11. Here we see the interior field being pulled to the left before breaking off into a closed loop. This latter configuration, known as an “O-type neutral point,” gradually shrinks as the enclosed current dies away. Outside the loop is a reconnecting “X-type neutral point,” similar to that depicted in Figure 10.10. Even farther to the right, $\boldsymbol{B}$ reattaches to the largescale equatorial field.

The analytical study underlying Figure 10.11 does not include the back-reaction of the stressed field on the contracting matter. It also neglects rotation of the gas, which further tangles $\boldsymbol{B}$ (see § 10.4). Accounting for such complications, along with the slippage due to ambipolar diffusion, has thus far prevented any detailed calculations of protostellar reconnection. In fact, the underlying theory of reconnection itself is still incomplete. For example, it is unclear whether the opposing fields near an X-type point annihilate within a single region or over a patchy network of smaller zones. Referring again to Figure 10.10, the second possibility arises if the fields being pressed together have small wiggles, so that dissipation is greater at some locations than at others. In the end, however, such issues may have little bearing on the net energy dissipation rate, which could be determined solely by the inward transport of magnetic flux.

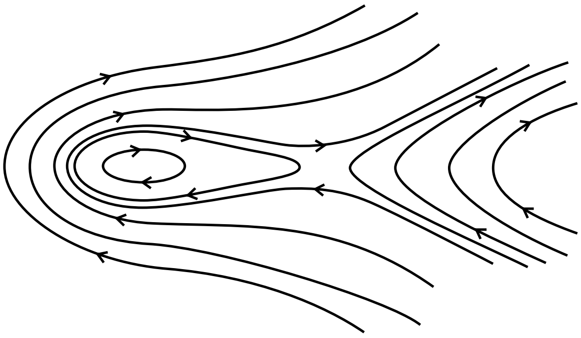


Figure 10.11 Distortion of field lines in a contracting cloud. Matter flowing to the left pulls on the field until it pinches off, creating both an O-type neutral point and a reconnecting X-type point.

As we have indicated, increased magnetic tension in the equatorial region should retard infall from this direction. What terminates collapse *along* field lines, *i. e.*, from the columns $\mathcal{A}$ and $\mathcal{A}'$ in Figure 10.8, is far less clear. Eventually the rarefaction wave must enter the more turbulent region supported by Alfvén waves. The drop in thermal pressure created by passage of the wave would then no longer trigger fresh infall. Lacking a more detailed picture of wave support, we can say little about how this transition actually occurs. In Chapter 13, we will adduce evidence from stellar evolution theory that supports equation (10.31) as a reasonable estimate for the time-averaged $\dot{M}(t)$. The same calculations indicate that the decline in accretion cannot occur too gradually, as the resulting, optically revealed stars would then have radii smaller than those inferred from observations. In this sense, the cutoff in new cloud material for the star must be rather efficient.

10.4 Rotational Effects

Thus far, we have neglected entirely the centrifugal force. We previously considered, but ultimately discounted, its effect in cloud *equilibria*, citing the observed low rotation rates for dense cores. During collapse, however, the initial angular rotation speed $\Omega_\circ$ of a fluid element originally located at cylindrical radius $\varpi_\circ$ increases to $\Omega_\circ(\varpi_\circ/\varpi)^2$ once the element reaches the smaller radius ϖ. The centrifugal force therefore increases as ϖ^{-3}, faster than the ϖ^{-2} rise in gravitational attraction toward any fixed interior mass. If the ratio of the two forces is initially of order 10^{-3} (recall equation (3.31)), it will reach unity once ϖ has decreased by the same factor. Thus, rotation must play a key role in the deep interior of a collapsing cloud.

This reasoning is qualitatively correct but overly simplified. First, the mass interior to an infalling fluid element is *not* fixed, apart from the idealized case of spherical collapse. Second, we have tacitly assumed that every element strictly conserves its specific angular momentum. This statement is true in an axisymmetric, *nonmagnetic* fluid medium, where the absence of azimuthal gradients precludes any torques about the central axis. Indeed, we used angular momentum conservation in Chapter 9 to generate models of rotating clouds. An embedded magnetic field, however, introduces torques and angular momentum transfer even if the configuration remains perfectly axisymmetric at all times. The reason is that $\boldsymbol{B}$ is anchored in the more rarefied medium outside the cloud proper. Any spinup during collapse therefore twists the field and

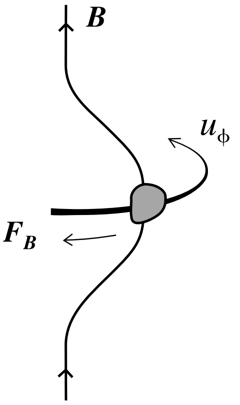


Figure 10.12 Rotational twisting of a field line (*schematic*). The azimuthal fluid velocity, pointing into the page, bends the field and creates a resisting tension force $\boldsymbol{F}_B$.

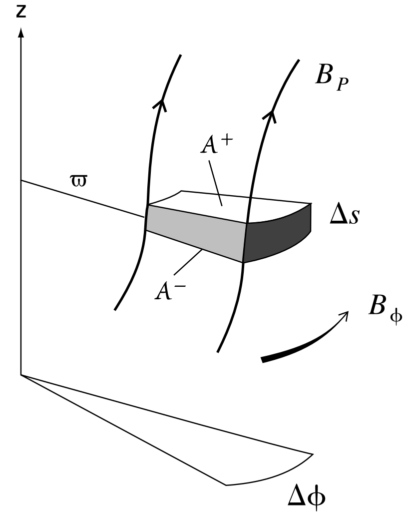


Figure 10.13 Derivation of magnetic torque. An elemental slab, of thickness Δs and top and bottom areas $\mathcal{A}^+$ and $\mathcal{A}^-$, respectively, is located a distance ϖ from the axis, and subtends the azimuthal angle $\Delta\phi$. Two neighboring, poloidal field lines are tangent to the edges of the slab. The azimuthal field component points into the page, as indicated.

increases the local magnetic tension (see Figure 10.12). This tension in turn creates a braking torque on the element that counteracts the spinup and lowers the specific angular momentum.

10.4.1 Magnetic Braking

To assess magnetic braking more quantitatively, we adopt the cylindrical coordinate system from Figure 9.5. We consider a magnetic field with components in all three directions and designate as $B_p \equiv \sqrt{B_\varpi^2 + B_z^2}$ the poloidal contribution, *i. e.*, the projection of the full $\boldsymbol{B}$ onto a meridional plane of constant ϕ. Figure 10.13 shows two neighboring poloidal field lines. Between them, lying a distance ϖ from the axis, is a small fluid element in the form of a patch of thickness Δs. Here, s measures the distance along a poloidal field line. The element is wedged between two meridional planes that span the small angle $\Delta\phi$, also indicated in the figure.

Assuming that both the fluid configuration and the field are axisymmetric, the only nonvanishing component of magnetic torque on the element is in the z-direction. This component, as measured per unit volume, is $\gamma_z = \varpi f_\phi$, where the volumetric magnetic force is given by

$$\begin{aligned} f_\phi &= \frac{1}{4\,\pi}\,[(\boldsymbol{\nabla}\times\boldsymbol{B})\times\boldsymbol{B}]_\phi \\ &= \frac{1}{4\,\pi}\left[\frac{B_\varpi}{\varpi}\,\frac{\partial(\varpi B_\phi)}{\partial\varpi} + B_z\,\frac{\partial B_\phi}{\partial z}\right] \; . \end{aligned} \tag{10.37}$$

Here, we have again invoked axisymmetry to set all ϕ-derivatives equal to zero. The torque is therefore

$$\begin{aligned} \gamma_z &= \frac{1}{4\,\pi}\left[B_\varpi\,\frac{\partial(\varpi B_\phi)}{\partial\varpi} + B_z\,\frac{\partial(\varpi B_\phi)}{\partial z}\right] \\ &= \frac{1}{4\,\pi}\,\boldsymbol{B}\cdot\boldsymbol{\nabla}_p\,(\varpi\,B_\phi) \; , \end{aligned}$$

where $\boldsymbol{\nabla}_p \equiv (\partial/\partial\varpi, \partial/\partial z)$ is the poloidal gradient. It is convenient to recast the last equation as

$$\gamma_z = \frac{1}{4\,\pi}\,B_p\,\frac{\partial(\varpi B_\phi)}{\partial s} \; . \tag{10.38}$$

The actual torque Γ_z exerted on the element is given by the product of γ_z and the small volume:

$$\Gamma_z = \frac{1}{4\,\pi}\,B_p\,\mathcal{A}\,\frac{\partial(\varpi B_\phi)}{\partial s}\,\Delta s \; , \tag{10.39}$$

where $\mathcal{A}$ is the surface area of the patch. In Figure 10.13, there are actually *two* relevant areas, $\mathcal{A}^+$ and $\mathcal{A}^-$, which differ slightly. A similar consideration applies to B_p, which also varies across Δs. However, $\boldsymbol{\nabla}\cdot\boldsymbol{B} = 0$ implies that the total magnetic flux entering the small volume must equal that leaving it. Axisymmetry guarantees that this condition is met for the flux penetrating the side walls, *i. e.*, that contributed by B_ϕ. For the outer and inner faces, we need

$$(B_p\,\mathcal{A})^+ = (B_p\,\mathcal{A})^- \; .$$

Hence, we may rewrite equation (10.39) as

$$\Gamma_z = \frac{1}{4\,\pi}\left[(\varpi\,B_\phi\,B_p\,\mathcal{A})^+ - (\varpi\,B_\phi\,B_p\,\mathcal{A})^-\right] \; . \tag{10.40}$$

10.4.2 Torsional Alfvén Waves

This form of the equation suggests another interpretation of magnetic braking. We view the twisted field itself as transporting a certain flux of angular momentum $\mathcal{F}_J$ along the poloidal direction. The net torque exerted by the field on our patch is then the rate at which angular momentum enters the inner face minus its efflux from the outer one. This interpretation corresponds to equation (10.40) provided we identify

$$\mathcal{F}_J = -\frac{\varpi}{4\,\pi}\,B_\phi\,B_p \; . \tag{10.41}$$

Remember that it is the z-component of angular momentum which is actually being transported. The minus sign in equation (10.41) means, for example, that this component is flowing upward in Figure 10.12, where B_ϕ is negative near the fluid element.

Suppose we now take some closed, two-dimensional surface $\mathcal{S}$ to represent the boundary of the rotating cloud. Then the total outflow of angular momentum due to magnetic braking is found by integrating equation (10.41) over this surface. Since $B_p = \boldsymbol{B} \cdot \boldsymbol{n}$, where $\boldsymbol{n}$ is the outward normal vector, the angular momentum *gain* is

$$\dot{J} \,=\, \frac{1}{4\,\pi} \int_{\mathcal{S}} \varpi\, B_\phi\, \boldsymbol{B} \cdot \boldsymbol{n}\, d^2\boldsymbol{x} \ . \tag{10.42}$$

This is again negative if B_ϕ points in the opposite direction as the rotation. The angular momentum leaving the cloud flows into the surrounding medium. That is, the twist imposed by the rotating cloud propagates along field lines through the associated magnetic tension. The propagation speed is the local value of V_A. The evolving field configuration is thus known as a *torsional Alfvén wave*, although it is quite distinct from the waves we have previously discussed.[5]

The net effect of magnetic braking on a dense core is to enforce corotation with the surroundings. As we have emphasized, one cannot cleanly distinguish the formation of a dense core from its later contraction once magnetic fields are involved. The braking phenomenon must operate from the earliest times, as gas condenses along field lines and begins to create the equatorial pinch illustrated in Figure 10.8. Over the bulk of the cloud, corotation should be a good approximation even during collapse, *i. e.*, the internal, local Ω should equal $\Omega_\circ$ in the more rarefied exterior. The reason is that the braking process propagates at V_A, a quantity close to the sound speed a_T. Hence, the time required to enforce corotation over a characteristic radius $\varpi_\circ$ is roughly equal to the sound-crossing time, which, in turn, is close to $t_{\rm ff}$ for the cloud as a whole. (In virial-theorem language, this latter statement is equivalent to $U \approx |\mathcal{W}|$; recall § 3.3.)

The actual degree of twisting required to maintain corotation is rather small. To see this, consider a highly idealized cylindrical cloud of constant density $\rho_\circ$, oriented along the z-axis and embedded in an initially *static* external medium (see Figure 10.14). A uniform magnetic field $\boldsymbol{B}_\circ$ also lies in the z-direction. If, at $t = 0$, the cloud is set rotating uniformly with angular speed $\Omega_\circ$, the braking action will propagate downward, as sketched by the shading in the figure. Meanwhile, a torsional Alfvén wave travels upward, setting the medium into rotation. Suppose that, after an interval Δt, a cloud segment of length Δz has ceased rotating. If $\varpi_\circ$ is the cloud radius, the change in angular momentum up to that time is

$$\Delta J \,=\, -\pi\, \rho_\circ\, \varpi_\circ^4\, \Omega_\circ\, \Delta z \ .$$

---

[5] The *plane-polarized* Alfvén waves introduced in § 9.5 are small-amplitude disturbances in which $\delta\boldsymbol{B}$ and $\boldsymbol{k}$ remain fixed in direction, at least over many wavelengths. In *torsional* waves, $\delta\boldsymbol{B}$ can be arbitrarily large, but must have a component that twists around the propagation direction. Note finally that one can form *circularly polarized* Alfvén waves by adding together two plane-polarized waves whose $\delta\boldsymbol{B}$-vectors are spatially orthogonal and oscillate 90° out of phase. These disturbances, like the plane-polarized ones, carry no angular momentum.

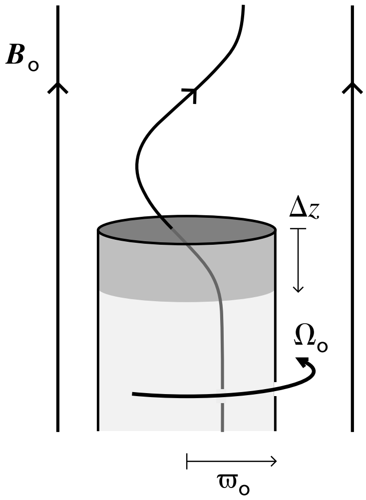


Figure 10.14 Magnetic braking of a rotating cylinder. A straight magnetic field $\boldsymbol{B}_\circ$ is frozen in the cylinder of radius $\varpi_\circ$, when the cylinder begins to rotate with angular speed $\Omega_\circ$. A torsional Alfvén wave propagates upward, while a portion of the cylinder with height Δz (*shaded region*) has ceased rotating.

Next consider $\dot{J}$ from equation (10.42). Letting B_ϕ be the radius-averaged azimuthal field component at the cloud face, we have

$$\dot{J}\,\Delta t \,=\, \frac{B_\circ\,B_\phi}{6}\,\varpi_\circ^3\,\Delta t\ .$$

We now equate ΔJ to $\dot{J}\,\Delta t$. Noting that $\Delta z/\Delta t \,=\, V_A \,=\, B_\circ/\sqrt{4\pi\rho_\circ}$, we find

$$\begin{aligned}\frac{|B_\phi|}{B_\circ} &= \frac{3\,\varpi_\circ\,\Omega_\circ}{2\,V_A}\\ &= 0.1\left(\frac{\varpi_\circ}{0.1\ \mathrm{pc}}\right)\left(\frac{\Omega_\circ}{10^{-14}\ \mathrm{s}^{-1}}\right)\left(\frac{B_\circ}{30\ \mu\mathrm{G}}\right)^{-1}\left(\frac{n_{\mathrm{H}_2}}{10^4\ \mathrm{cm}^{-3}}\right)^{-1/2}\ .\end{aligned} \tag{10.43}$$

10.4.3 Centrifugal Radius

These considerations indicate that magnetic braking is both rapid and efficient. On the other hand, the braking *must* fail within the deeper interior of a more realistic, collapsing cloud. As the density climbs, the matter and field decouple because of the drop in ionization fraction. In addition, much of the remaining field is left behind once it is severely pinched and reconnects. Thus, the matter inside some volume of the equatorial region indeed conserves angular momentum and spins up as it approaches the central protostar. Since the centrifugal force rises faster than gravity, each fluid element inevitably veers away from the geometrical center of the cloud. Whether or not this element lands on the protostar or misses it and joins the equatorial disk depends in part on its specific angular momentum j.

In fact, a *range* of specific angular momenta is present in the infalling matter at any time. Those elements with the highest j-values depart from radial trajectories earliest and ultimately land farthest from the center. The maximum impact distance in the equatorial plane is known as the *centrifugal radius*, here denoted ϖ_{cen}. As we will see in § 11.3, ϖ_{cen} also sets the scale

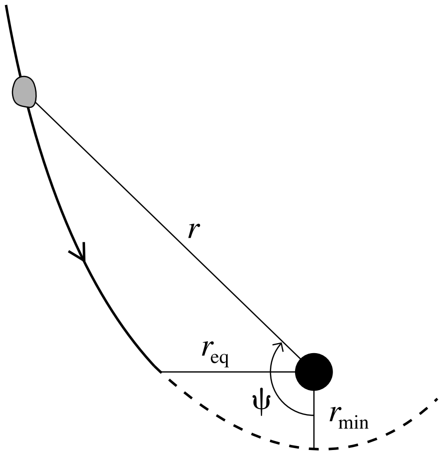


Figure 10.15 Parabolic orbit in rotating infall. A fluid element, with instantaneous polar coordinates $(r,\,\psi)$ in the orbital plane, falls into the radial distance $r_{\rm eq}$, where it impacts the disk. If the disk were absent, the element would have reached the smaller distance $r_{\rm min}$ before swinging back out.

of the circumstellar disk. The inside-out nature of collapse implies that fluid elements arriving at some later epoch begin falling from more distant locations within the cloud. These regions include higher j-values (recall Figure 9.6), so that some new elements cross the plane even farther from the center than before. In other words, $\varpi_{\rm cen}$ increases with time.

Exploring this idea more quantitatively requires that we determine the actual fluid trajectories. Suppose the magnetic field has effectively decoupled and the velocity is supersonic, so that thermal pressure is no longer significant. Then only gravity and rotation affect the infall. The former is supplied mostly by the protostar and its disk, which we may lump together as a point mass for our purposes. Under these conditions, the trajectory must be an ellipse, *i. e.*, the conic section corresponding to a bound orbit of negative energy. In practice, both the gravitational potential and kinetic energies at the start of infall are tiny compared to their magnitudes when the element nears the star or equatorial plane. The trajectory is thus closely approximated as a *zero*-energy conic section, *i. e.*, a parabola.[6]

Figure 10.15 shows a typical parabolic orbit, with the protostar located at the focus. Here, we have specified the instantaneous location of the fluid element through the radius r and angle ψ. The angle starts at π and decreases to $\pi/2$ by the time the element reaches the plane, when it is a distance $r_{\rm eq}$ from the protostar. At this point, the fluid velocity normal to the plane abruptly goes to zero, as the element either collides with a pre-existing disk or else with a streamline approaching from the opposite direction. We will consider both these possibilities further in Chapter 11. In any case, the dashed portion of the parabola, corresponding to $\psi < \pi/2$, is never actually traversed.

The functional relation between r and ψ is

$$r \,=\, \frac{r_{\rm eq}}{1\,+\,\cos\,\psi}\ . \tag{10.44}$$

---

[6] Notice that we are neglecting any increase in the protostar or disk mass during the relatively brief time a fluid element crosses the distance of interest. The same qualification applies to equation (10.34), which pertains to spherical collapse. In both cases, we are making a *steady-state* approximation for the interior region of the flow. Mathematically, we ignore the explicit time derivative $\partial/\partial t$ in the fluid equations and solve for the spatial variation of all variables.

In order to write r_{eq} in terms of constants of the motion, we pretend for a moment that the orbit actually extends to the minimum radius $r_{\rm min}$, also shown in Figure 10.15. At this point, the zero-energy condition implies that

$$V_{\rm max}^2 = \frac{2\,G\,M_*}{r_{\rm min}} \ .$$

Here, $V_{\rm max}$ is the maximum orbital speed, *i. e.*, that which would be attained at $r_{\rm min}$, while M_* is the combined mass of the protostar and disk. If j_n denotes the specific angular momentum normal to the orbital plane, then

$$j_n^2 = r_{\rm min}^2\,V_{\rm max}^2 = 2\,G\,M_*\,r_{\rm min} \ ,$$

which we can solve for $r_{\rm min}$. But the element would have reached $r_{\rm min}$ at $\psi = 0$, so equation (10.44) implies that $r_{\rm min}$ also equals $r_{\rm eq}/2$. From these facts, we deduce that

$$r_{\rm eq} = \frac{j_n^2}{G\,M_*} \ . \qquad (10.45)$$

Within the larger parent cloud, the orbital plane of any element is tilted at some angle $\theta_\circ$ with respect to the rotation axis (see Figure 10.16a). Knowing this angle helps us determine where the element originated, and therefore the proper j_n to use in equation (10.45). As explained previously, the specific angular momentum must increase with time, even at a fixed $\theta_\circ$. The centrifugal radius $\varpi_{\rm cen}$, also shown in Figure 10.16a, is then the maximum value of $r_{\rm eq}$ at any time, corresponding to the highest j_n.

For the limited purpose of supplying a function $j_n(\theta_\circ, t)$, let us employ the simplest cloud model exhibiting both rotation and inside-out collapse. Thus, we choose the density distribution before collapse to be that of the singular isothermal sphere (equation (9.8)) and impose a uniform angular speed $\Omega_\circ$. Aside from the issue of dynamical stability, such a model is not even self-consistent, since a rotating configuration cannot be spherically symmetric. Nevertheless, the pattern of infall close to the protostar and disk should not change greatly with a more realistic parent cloud.

Figure 10.17a depicts a side view of the orbital plane inside the collapsing sphere. Each element begins its infall at the rarefaction wave, a distance R from the center, and at the angle $\theta_\circ$ from the axis. The orbital velocity around the axis prior to the wave's arrival is $\Omega_\circ\,R\,\sin\,\theta_\circ$. Thus, the normal angular momentum vector, whose direction is indicated, has magnitude

$$j_n = R^2\,\Omega_\circ\,\sin\,\theta_\circ \ . \qquad (10.46)$$

Since the wave travels at the local sound speed, we also have

$$R = a_T\,t' \ . \qquad (10.47)$$

Here, t' is the time at which the element begins to fall, as measured after the initial accumulation of the central protostar.

The time t' is naturally less than that elapsed when the element crosses the equatorial plane. Denoting the latter by t, we can deduce its relation to t' by first noting that the rarefaction wave

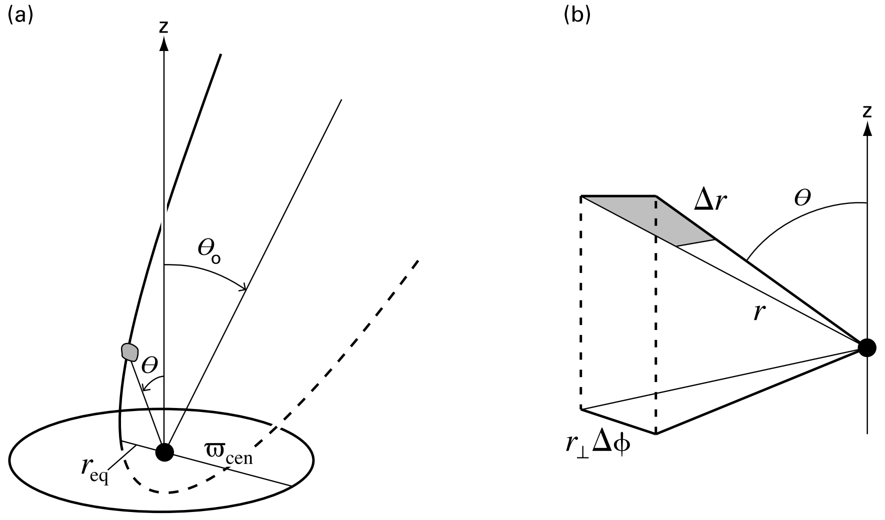


Figure 10.16 (a) Orientation of the orbital plane within a collapsing cloud. The plane is tilted from the rotation axis by the angle $\theta_\circ$. A fluid element has the instantaneous polar angle θ relative to the axis and crosses the equatorial plane at a radial distance $r_{\rm eq}$, less than the maximum value $\varpi_{\rm cen}$. (b) Near the position of the fluid element, mass continually flows downward through the shaded patch.

engulfs an incremental volume $4\pi R^2 a_T \Delta t'$ during the small interval $\Delta t'$. Using equation (9.8), the corresponding mass swept up is

$$\Delta M \;=\; \frac{2\, a_T^3\, \Delta t'}{G} \ .$$

All the matter in this shell eventually reaches the star and disk, over an interval Δt. We now use equation (10.31) for the mass accretion rate, where we take $m_\circ$ to be a strict constant. Then we also have

$$\Delta M \;=\; \frac{m_\circ\, a_T^3\, \Delta t}{G} \ .$$

Comparing these two expressions for ΔM, we deduce that $\Delta t' = m_\circ \Delta t/2$. Since both times are zero at protostar formation, we find

$$t' \;=\; \frac{m_\circ}{2}\, t \ . \tag{10.48}$$

We have finally assembled all the ingredients for evaluation of $r_{\rm eq}$. Combining equations (10.45)–(10.48), we conclude that

$$r_{\rm eq} \;=\; \frac{m_\circ^3\, a_T\, \Omega_\circ^2\, t^3\, \sin^2\,\theta_\circ}{16} \ . \tag{10.49}$$

Here we have also replaced M_* in equation (10.45) by $\dot{M} t$, where $\dot{M}$ is again taken from equation (10.31). We see that fluid elements falling at greater inclination angles $\theta_\circ$ land farther

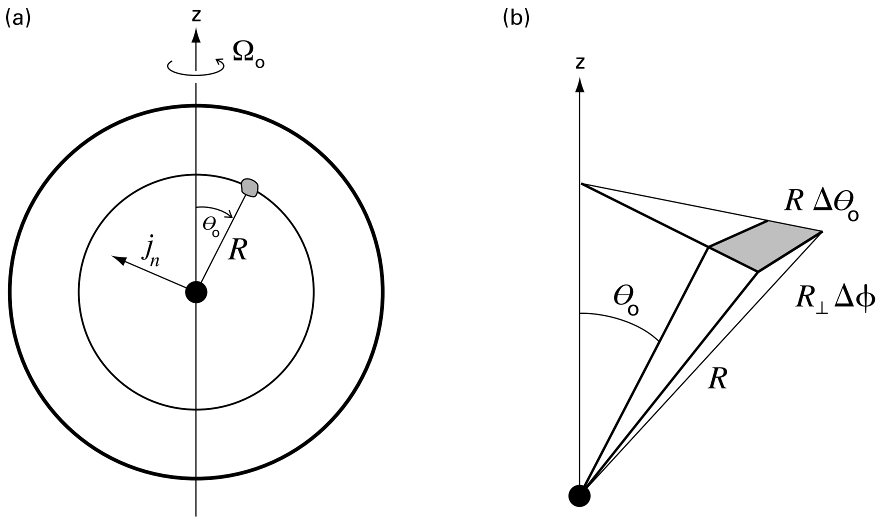


Figure 10.17 (a) Side view of orbital plane within the rotating cloud. Gas begins to fall after passage of the rarefaction wave, shown by the interior circle. This falling gas conserves its specific angular momentum vector $\boldsymbol{j}_n$, which lies normal to the plane. (b) Near the head of the rarefaction wave, mass flows inward through the shaded patch.

from the center, since their j_n-values are larger. We obtain the centrifugal radius by setting $\theta_\circ = \pi/2$ in equation (10.49). The result is

$$\begin{aligned}\varpi_{\rm cen} &= \frac{m_\circ^3\, a_T\, \Omega_\circ^2\, t^3}{16} \\ &= 0.3\ {\rm AU}\left(\frac{T}{10\ {\rm K}}\right)^{1/2}\left(\frac{\Omega_\circ}{10^{-14}\ {\rm s}^{-1}}\right)^2\left(\frac{t}{10^5\ {\rm yr}}\right)^3 ,\end{aligned} \tag{10.50}$$

where we have used $m_\circ = 1$ in the numerical expression. As expected, $\varpi_{\rm cen}$ is an increasing function of time.

10.4.4 Interior Structure

It is also of interest to obtain the spatial distribution of density and velocity within the deep interior of the cloud. The former should be a generalization of equation (10.34) for spherical collapse, and the latter some angle-dependent modification of $V_{\rm ff}$ for each of the velocity components. We will want to express all physical quantities as functions of the usual spherical polar coordinates (r, θ, ϕ), rather than the set $(r, \theta_\circ, \psi)$ we have been using until now. First, we must establish a geometric relation between three angles: the spherical coordinate θ, the tilt of the orbital plane $\theta_\circ$, and the angular displacement within that plane, ψ (recall Figure 10.15). By projecting the fluid element in Figure 10.16 onto both the z-axis and the axis of the parabola, the reader may show that

$$\cos\theta = -\cos\theta_\circ\,\cos\psi\ . \tag{10.51}$$

We insert this relation into the orbital formula (10.44) and use the subsequent equations to find

$$\frac{r}{\varpi_{\rm cen}} = \frac{\sin^2\theta_\circ\,\cos\theta_\circ}{\cos\theta_\circ - \cos\theta} \tag{10.52}$$

along a fluid trajectory. This last equation may be solved numerically to give the orbital inclination angle $\theta_\circ$ at any point (r,θ) within the collapsing portion of the cloud. Note that the resulting $\theta_\circ$ also depends on time, through the term $\varpi_{\rm cen}$.

In order to derive the velocity components, we first observe, from Figure 10.17a, that the z-component of specific angular momentum is $j_n\,\sin\,\theta_\circ$. Equating this to $r\,\sin\,\theta\,u_\phi$ and using (10.44), (10.45), and (10.51), we find

$$u_\phi = \left(\frac{G\,M_*}{r}\right)^{1/2}\frac{\sin\theta_\circ}{\sin\theta}\left(1 - \frac{\cos\theta}{\cos\theta_\circ}\right)^{1/2} . \tag{10.53}$$

This equation must be supplemented by (10.52) for $\theta_\circ$ in order to give the spatial dependence of u_ϕ. Turning to the other components, we employ $u_r = \dot{r}$, where the temporal derivative is applied to equation (10.44), considering $r_{\rm eq}$ to be effectively constant. Similarly, we use $u_\theta = r\,\dot{\theta}$ and thus establish the ratio

$$\frac{u_\theta}{u_r} = \frac{\cos\theta\, - \,\cos\theta_\circ}{\sin\theta} . \tag{10.54}$$

The zero-energy condition for the parabolic orbit implies that the sum of the squared components must be $V_{\rm ff}^2$. Combining this fact with equations (10.53) and (10.54), we find

$$u_r = -\left(\frac{G\,M_*}{r}\right)^{1/2}\left(1 + \frac{\cos\theta}{\cos\theta_\circ}\right)^{1/2} \tag{10.55}$$

$$u_\theta = \left(\frac{G\,M_*}{r}\right)^{1/2}\left(\frac{\cos\theta_\circ\, - \,\cos\theta}{\sin\theta}\right)\left(1 + \frac{\cos\theta}{\cos\theta_\circ}\right)^{1/2} . \tag{10.56}$$

Finally, the density distribution follows by invoking mass conservation. Referring to Figure 10.17b, we see that the area of a small patch normal to the rarefaction wavefront is $R\Delta\theta_\circ \cdot R_\perp\Delta\phi$, where $R_\perp \equiv R\,\sin\,\theta_\circ$. The mass per unit time entering the patch is this area times the flux $\dot{M}/4\pi R^2$. Similarly, Figure 10.16b demonstrates that the patch area normal to the parabolic trajectory is $\Delta r \cdot r_\perp\Delta\phi$, where $r_\perp \equiv r\,\sin\,\theta$. Here, the relevant mass flux is $\rho\,u_\theta$. If we equate the two rates of mass transfer, we find

$$\frac{\dot{M}}{4\,\pi}\,\frac{\sin\theta_\circ}{\sin\theta} = \rho\,u_\theta\,r\left(\frac{\partial\,r}{\partial\,\theta_\circ}\right)_\psi ,$$

where we note that the derivative is at fixed orbital angle ψ. Applying this derivative to equation (10.44) and utilizing (10.45), (10.46), and our expressions for the velocity components, we solve for the density and find

$$\rho = -\frac{\dot{M}}{4\,\pi\,r^2\,u_r}\left[1 + \frac{2\,\varpi_{\rm cen}}{r}\,P_2(\cos\theta_\circ)\right]^{-1} . \tag{10.57}$$

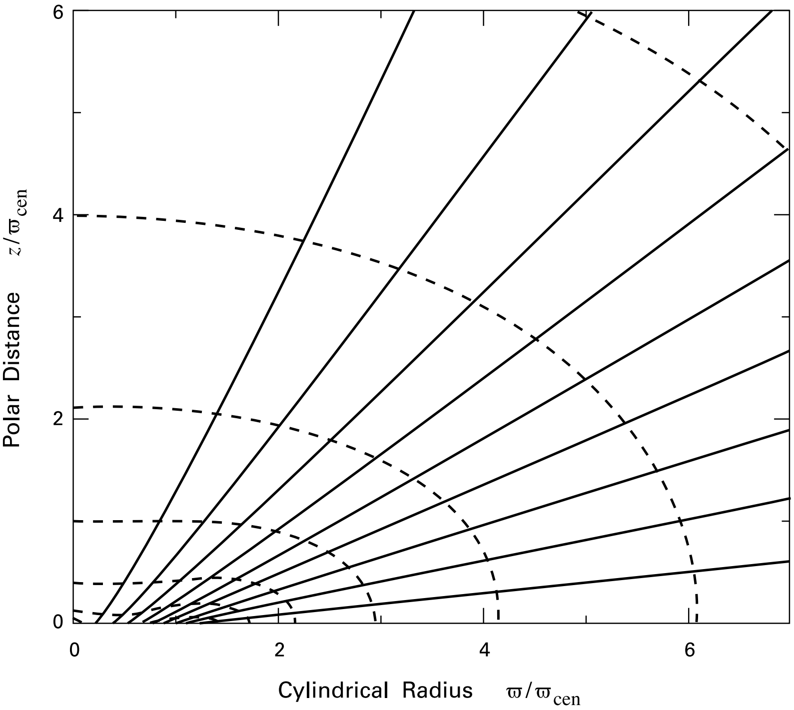


Figure 10.18 Fluid streamlines (*solid curves*) and isodensity contours (*dashed curves*) within a collapsing, rotating cloud. The values of density on successive contours differ by a factor of 2, and adjacent stream lines enclose equal mass flux.

Here, we have introduced the Legendre polynomial

$$P_2(\cos\theta_\circ) \equiv \frac{3}{2}\cos^2\theta_\circ - \frac{1}{2} .$$

Figure 10.18 plots the streamlines and density contours in a meridional plane for this rotating accretion flow. Note that both spatial coordinates are measured relative to $\varpi_{\rm cen}$. Thus, as time progresses, the entire set of curves expands as t^3, without suffering any distortion. Since the angle $\theta_\circ$ is constant on a streamline, each trajectory can be obtained by solving equation (10.52) for $r/\varpi_{\rm cen}$ as a function of θ. The value of $r d\theta/dr$ at each point is then equal to the velocity ratio u_θ/u_r. For display purposes, the $\theta_\circ$-values in Figure 10.18 were selected so that any two adjacent streamlines enclose an equal mass flux onto the origin.

At distances such that $r \gg \varpi_{cen}$, the density contours gradually become spherical, corresponding to purely radial infall. Notice, from equations (10.55) and (10.57), that ρ varies as $r^{-3/2}$ in this regime, as in spherical collapse (recall equation (10.34)). Conversely, the density has an $r^{-1/2}$ radial variation deep inside $\varpi_{\rm cen}$. In this region, the streamlines bend and the density contours flatten. From equation (10.57), the density formally diverges on the equatorial plane ($\theta_\circ = \pi/2$) at $r = \varpi_{\rm cen}$. The figure shows that this blowup results from the crowding of streamlines, as the infalling gas is repelled by the centrifugal force. We shall describe in the next chapter how $\varpi_{\rm cen}$ becomes the boundary of a disk that builds up around the star.

Chapter Summary

Magnetic forces within a cloud are transmitted by collisions from charged particles to the neutral gas. Dense cores have such a low ionization fraction that there is appreciable slip between ions and neutrals, and the cloud gradually loses magnetic flux. Quasi-static evolution mediated by this ambipolar diffusion is the likely precursor to gravitational collapse. However, the cloud shape problem remains. Magnetostatic structures initially flattened perpendicular to the field become even flatter with time, contrary to observation. Alfvén waves cannot disturb clouds on the scale of dense cores, since these perturbations are damped through ambipolar diffusion.

The fundamentals of protostellar collapse are well illustrated by the idealized case of a spherical, nonmagnetic cloud. Infall occurs first at the center, then spreads out at the sound speed. The rate at which mass accrues onto the central protostar depends mostly on the temperature in the parent dense core. How in detail collapse proceeds in a more realistic, magnetized cloud is less clear. While the outermost material is perturbed strongly by MHD waves, the field decouples from the matter in the more quiescent, infalling region. Any residual field undergoes violent reconnection closer to the nascent star.

A rotating, magnetized cloud emits torsional Alfvèn waves that efficiently brake the cloud's angular speed. This mechanism explains the observed slow rotation of dense cores. Interior matter decoupled from the field enters parabolic trajectories. As infall continues, this gas impacts the equatorial plane at increasingly large distances from the central star. The result is a rapidly growing circumstellar disk.

Suggested Reading

Section 10.1 The importance of ambipolar diffusion in dense, lightly ionized clouds was recognized by

Mestel, L. & Sptizer, L. 1956, MNRAS, 116, 503.

The first detailed evolutionary calculations incorporating this effect are reviewed in

Nakano, T. 1984, Fund. Cosm. Phys., 9, 139,

while subsequent work is exemplified by

Mouschovias, T. Ch. & Fiedler, R. A, 1993, ApJ, 415, 680,

which is the source for the quantitative results of this section. A more recent calculation, in a similar vein, is

Tomisaka, K. 2002, ApJ, 575, 306.

The theory of Alfvén wave damping is due to

Kulsrud, R. & Pearce, W. P. 1969, ApJ, 156, 445.

Section 10.2 Inside-out collapse first came to light through numerical simulations, such as those of

Bodenheimer, P. & Sweigart, A. 1968, ApJ, 152, 515

Larson, R. B. 1969, MNRAS, 145, 271.

These investigations have continued, a later example being

Foster, P. N. & Chevalier, R. A. 1993, ApJ, 416, 303.

Equally instructive have been semi-analytic calculations that describe collapse in an unbounded medium:

Penston, M. V. 1969, MNRAS, 144, 425

Shu, F. H. 1977, ApJ, 214, 488.

This last study derives our equation (10.31) for the mass accretion rate. The stability of inside-out collapse against fragmentation is demonstrated numerically by

Boss, A. P. 1987, ApJ, 319, 149.

Section 10.3 For the appearance of dense clumps in simulations of turbulent clouds, see

Klessen, R. S. 2001, ApJ, 556, 837.

The effect of ambipolar diffusion on magnetized collapse has been studied by

Black, D. C. & Scott, E. H. 1982, ApJ, 263, 696

Safier, P., McKee, C. F., & Stahler, S. W. 1997, ApJ, 485, 660

Li, Z.-Y. 1998, ApJ, 493, 230.

For the physics of magnetic reconnection, see

Parker, E. N. 1979, *Cosmical Magnetic Fields*, (Oxford: Clarendon Press), Chapter 15.

Section 10.4 The braking of rotating clouds through torsional Alfvèn waves is explored in

Mouschovias, T. Ch. & Paleologou, E. V. 1980, ApJ, 237, 877

Nakano, T. 1989, MNRAS, 241, 495.

A detailed study of collapse incorporating this effect has yet to be done. The accretion flow in nonmagnetic, rotating clouds was discovered independently by

Ulrich, R. K. 1976, ApJ, 210, 377

Cassen, P. M. & Moosman, A. 1981, Icarus, 48, 353.

Our derivation more closely follows the second article.

The Formation of Stars
Steven W. Stahler and Francesco Palla

11 Protostars

Thus far, we have viewed the protostar within a collapsing cloud simply as a mass sink and source of gravity for the surrounding, diffuse matter. We now examine more closely the properties of this object, utilizing the tools of stellar evolution theory. We shall also want to consider that inner portion of the cloud significantly heated by the luminous, central body. It is this region which holds the most promise for observational detection, both through its thermal emission and inward motion. When we then delve into the structure of protostars *per se*, we emphasize the role of deuterium fusion in subsolar masses. The steady release of energy from this reaction, while adding little to the star's surface luminosity, nevertheless exerts a powerful and lasting influence.

One cannot properly discuss the evolution of protostars without including their surrounding disks. Such disks have been observed around older, optically visible stars, and are the sources of planetary systems. Section 3 of this chapter concerns the theory of their origin and early growth. Returning to stars, we then extend the previous structural analysis to the intermediate-mass regime, thereby laying the groundwork for a theoretical understanding of Herbig Ae/Be stars. Finally, we depart from theory to assess the ongoing effort by infrared and millimeter observers to detect protostars in nearby star-forming regions.

11.1 First Core and Main Accretion Phase

How does a protostar initially form? In answering this question, we should bear in mind that the cloud environment at this time is one characterized by slow contraction and not violent collapse. We have seen how ambipolar diffusion mediates this contraction by gradually eroding the cloud's internal, magnetic support. We have also noted that the leakage of flux proceeds more quickly in denser regions. The accelerating density increase exemplified by Figure 10.4 is thus bound to occur, even if the quantitative details are not fully known. However, the structure that arises is not yet a bona fide protostar, but a temporary configuration known as the *first core*. Let us briefly trace its growth and rapid demise. Of necessity, our treatment is based on spherically symmetric calculations that omit the important elements of rotation and magnetic support. We accordingly limit ourselves to describing general features of the evolution that should not change markedly even after more complete studies become available.

11.1.1 Early Growth and Collapse

A key point of departure from our previous analysis of clouds is that the isothermal approximation, which served us well in describing larger-scale equilibrium and dynamics, now breaks down entirely. As its density climbs, the central lump quickly becomes opaque to its own

The Formation of Stars. Steven W. Stahler and Francesco Palla

ISBN: 3-527-40559-3

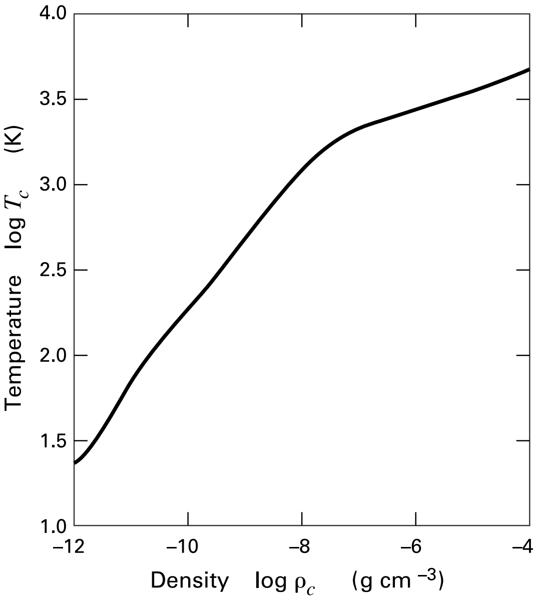


Figure 11.1 Evolution of central temperature in the first core. The temperature is plotted as a function of central density.

infrared, cooling radiation. Further compression then causes its internal temperature to rise steadily. The enhanced pressure decelerates material drifting inward, which settles gently onto the hydrostatic structure. The settling gas can still radiate rather freely in the infrared, at least before it is smothered by successive layers of incoming matter. This energy loss from the outer skin then further enhances compression. The calculations show, in fact, that the core eventually stops expanding and begins to shrink, even as fresh material continues to arrive. The total compressed mass is still small at this stage, about $5 \times 10^{-2}\ M_\odot$, but the radius is large by stellar standards, roughly 5 AU (8×10^{13} cm).

The interior of the central object, like its surroundings, consists mostly of molecular hydrogen. This fact alone seals the fate of the first core and ensures its early collapse. To see why, let us first estimate the mean internal temperature, utilizing the virial theorem in the version of equation (3.16). The object builds up from that portion of the parent cloud that was least supported rotationally and magnetically. We therefore tentatively ignore both the bulk kinetic energy $\mathcal{T}$ and the magnetic term $\mathcal{M}$ in the virial theorem. We further approximate the gravitational potential energy $\mathcal{W}$ as $-GM^2/R$, for a core of mass M and radius R. The internal energy becomes

$$\begin{aligned} U &= \frac{3}{2}\int P\, d^3\boldsymbol{x} \\ &= \frac{3}{2}\frac{\mathcal{R}T}{\mu}M \ , \end{aligned} \tag{11.1}$$

where T and μ are the volume-averaged temperature and molecular weight, respectively. Applying equation (3.16) and solving for the temperature, we find

$$T \approx \frac{\mu}{3\mathcal{R}} \frac{GM}{R} \tag{11.2a}$$

$$= 850\,\mathrm{K} \left(\frac{M}{5 \times 10^{-2}\,M_\odot}\right) \left(\frac{R}{5\,\mathrm{AU}}\right)^{-1} . \tag{11.2b}$$

Here we have set μ equal to 2.4, the value appropriate for molecular gas.

The internal temperature, while very low compared to true stars, is higher than in quiescent molecular clouds, as is the average mass density, which is now of order 10^{-10} g cm^{-3}. With the addition of mass and shrinking of the radius, T soon surpasses 2000 K, and collisional dissociation of H$_2$ begins. At this point, the temperature starts to level off. The effect is evident in Figure 11.1, which tracks the temperature as a function of density at the center. Viewing the situation energetically, we note that the number of H$_2$ molecules in the core is $XM/2m_H$, where $X = 0.70$ is the interstellar hydrogen mass fraction. From equation (11.1), the thermal energy per molecule is therefore $3k_BT/X$, or 0.74 eV when $T = 2000$ K. This figure is small compared to the 4.48 eV required to dissociate a single molecule. During the transition epoch, therefore, even a modest rise in the fraction of dissociated hydrogen absorbs most of the compressional work of gravity, without a large increase in temperature.

As the density of the first core keeps climbing, the region containing atomic hydrogen spreads outward from the center. We recall from § 9.1 that purely isothermal configurations can tolerate only a modest density contrast before they become gravitationally unstable. The reason is that the compression arising from any perturbations can no longer be effectively opposed by a rise in the internal pressure, once the temperature is held fixed. The interior temperature of the first core is *not* a fixed constant, but its rise is severely damped by the dissociation process. Hence, the partially atomic region can only spread and increase its mass by a limited amount before the entire configuration becomes unstable and collapses. This event marks the end of the first core.

11.1.2 Accretion Luminosity

The collapse of the partially dissociated gas takes the central region to much higher density and temperature. Indeed, the latter is now sufficient to collisionally ionize most of the hydrogen. The true protostar that emerges is not susceptible to another internal transition and remains dynamically stable. With a radius of several $R_\odot$, a protostar of 0.1 $M_\odot$ has, from equation (11.2a), a mean internal temperature above 10^5 K. Such a value, coupled with a mass density of order 10^{-2} g cm^{-3}, places the object within the stellar regime.

Gas that approaches the protostellar surface is now traveling essentially at free-fall velocity, which is considerably greater than the local sound speed. The steady rise in the protostellar mass gradually inflates this supersonic infall region, so that the cloud collapse proceeds in the usual inside-out manner. By this point, the protostar is said to have entered the *main accretion phase*. For now, we will continue to describe the main characteristics of this period as if the collapse were spherically symmetric. We will soon indicate, in varying detail, the alterations introduced by rotation and magnetic fields. To date, however, the spherical calculations have provided by far the most complete information and are still the only studies to have evolved the protostar to interestingly high masses.

Let us first consider the gross energetics of the main accretion phase. A protostar of mass M_* and radius R_* forms out of cold, nearly static cloud material whose dimensions are enormous compared to R_*. Thus, we may effectively set the initial energy, both mechanical and thermal, to zero. The star itself, however, is a gravitationally bound entity with a *negative* total energy. Some of the energy difference is radiated into space during collapse, while most of the rest goes into dissociating and ionizing hydrogen and helium. We denote this latter, internal component as $\Delta E_{\rm int}$, where

$$\Delta E_{\rm int} \equiv \frac{X\,M_*}{m_H}\left[\frac{\Delta E_{\rm diss}(H)}{2} + \Delta E_{\rm ion}(H)\right] + \frac{Y\,M_*\,\Delta E_{\rm ion}(He)}{4\,m_H}\ .$$

Here, $\Delta E_{\rm diss}(H) = 4.48$ eV is the binding energy of H_2, $\Delta E_{\rm ion}(H) = 13.6$ eV is the ionization potential of HI, and $\Delta E_{\rm ion}(He) = 75.0$ eV is the energy required to fully ionize helium. The protostar's thermal energy U is equal to $-\mathcal{W}/2$ by the virial theorem. Approximating $\mathcal{W}$ as $-GM_*^2/R_*$, we may write

$$0 = -\frac{1}{2}\,\frac{G\,M_*^2}{R_*} + \Delta E_{\rm int} + L_{\rm rad}\,t\ , \tag{11.3}$$

where $L_{\rm rad}$ is the average luminosity escaping over the formation time t.

Suppose that we first take the extreme step of ignoring $L_{\rm rad}$ entirely. Then equation (11.3) yields the *maximum* radius $R_{\rm max}$ which the protostar could have at any mass M_*. We readily find

$$\begin{aligned} R_{\rm max} &= \frac{G\,M_*^2}{2\,\Delta E_{\rm int}} \\ &= 60\ R_\odot\left(\frac{M_*}{M_\odot}\right)\ . \end{aligned} \tag{11.4}$$

This numerical estimate would change slightly with a more careful treatment of the gravitational energy $\mathcal{W}$. In any case, we know that $R_{\rm max}$ is considerably greater than the true radii of solar-type protostars. Their immediate descendants, the youngest T Tauri stars, are smaller by an order of magnitude, as we shall see in Chapter 16.

Since $R_* \ll R_{\rm max}$, the final term in equation (11.3) is actually comparable to the first one. Setting $\dot{M} = M_*/t$, we conclude that $L_{\rm rad}$ is close to the *accretion luminosity*, given by

$$\begin{aligned} L_{\rm acc} &\equiv \frac{G\,M_*\,\dot{M}}{R_*} \\ &= 61\ L_\odot\left(\frac{\dot{M}}{10^{-5}\ M_\odot\ {\rm yr}^{-1}}\right)\left(\frac{M_*}{1\ M_\odot}\right)\left(\frac{R_*}{5\ R_\odot}\right)^{-1}\ . \end{aligned} \tag{11.5}$$

We will later justify our representative numerical values for $\dot{M}$ and R_*. The important quantity $L_{\rm acc}$ is the energy per unit time released by infalling gas that converts *all* its kinetic energy into radiation as it lands on the stellar surface. Despite the approximations entering our derivation, $L_{\rm rad}$ is very nearly equal to $L_{\rm acc}$ throughout the main accretion phase, regardless of the detailed time dependence of $\dot{M}$. Moreover, this equality holds even if the gas first strikes a circumstellar disk, then subsequently spirals onto the star (see § 11.3 below). The only stipulation is that each fluid element's thermal plus kinetic energies be relatively small once it joins the protostar. For example, the star cannot be rotating close to breakup speed. The T Tauri observations indicate, in fact, that this latter condition is safely met (Chapter 16).

The accretion luminosity, although a product of cloud collapse, is mostly generated close to the protostar's surface. Additional radiated energy comes from nuclear fusion and the quasi-static contraction of the interior. However, these contributions are minor compared to $L_{\rm acc}$ for low and intermediate masses. It is therefore conventional to *define* a protostar as a mass-gaining star whose luminosity stems mainly from external accretion. This radiation is able to escape the cloud because it is gradually degraded into the infrared regime as it travels outward. Infrared photons can traverse even the large column density of dust lying between the stellar surface and the outer reaches of the parent dense core. Observationally, then, protostars are optically invisible objects that should appear as compact sources at longer wavelengths.

Figure 11.2 shows in more detail how the radiation diffuses outward. The figure also indicates the major physical transitions in the cloud material that is freely falling onto the protostar. Most of the radiation is generated at the *accretion shock*. Since matter further inside is settling with relatively low velocity, the shock front itself constitutes the protostar's outer boundary. Note how the figure suggests a turbulent state for the deeper interior. Such turbulence is induced by nuclear fusion at the center, as we will describe shortly.

11.1.3 Dust Envelope and Opacity Gap

The gas raining down on the protostar originates much farther away, in the *outer envelope*. This is the infalling region where, as we noted in § 10.2, the gas temperature rises sluggishly with density as a result of efficient cooling by dust. Despite the nomenclature, we recall that the matter here does not fall until it is inside the rarefaction wave gradually spreading throughout the cloud. Most of this expanding volume is nearly transparent to the protostellar radiation. However, as the infalling gas continues to be compressed, the radiation eventually becomes trapped by the relatively high opacity from the grains. Inside the *dust photosphere*, located at $R_{\rm phot} \sim 10^{14}$ cm, the temperature rises more quickly. The sphere with radius $R_{\rm phot}$ is the effective radiating surface of the protostar, as seen by an external observer.

We define the *dust envelope* to be the region bounded by $R_{\rm phot}$ that is opaque to the protostar's radiation. Once the temperature here climbs past about 1500 K, all the hot grains vaporize. The precise temperature depends on the adopted grain model, but the qualitative effect is always the same. Inside this *dust destruction front* ($R_d \sim 10^{13}$ cm), the opacity is greatly reduced. The infalling gas, which also collisionally dissociates above 2000 K, is nearly transparent to the radiation field. The region of vaporized grains is therefore known as the *opacity gap*. Even further inside, collisional ionization of the gas, and an attendant rise in the opacity, occur in the *radiative precursor*, immediately outside the accretion shock itself. We recall from Chapter 8 that such layers are ubiquitous features of high-velocity, J-type shocks.

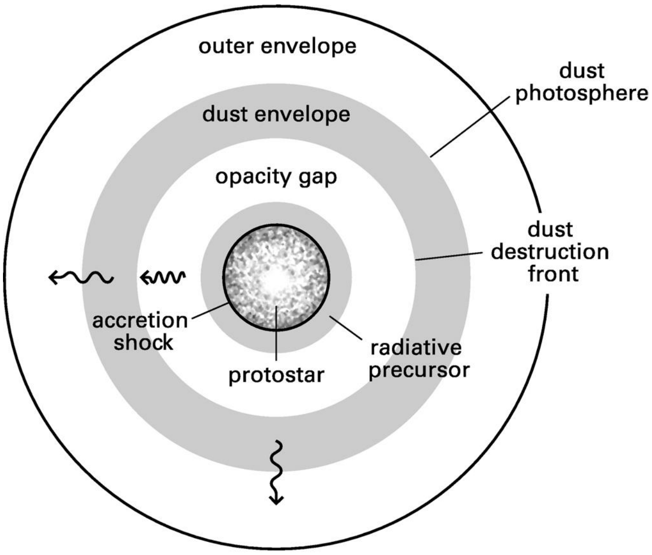


Figure 11.2 Structure of a spherical protostar and its infalling envelope. The relative dimensions of the outer regions have been greatly reduced in this sketch. Note the convection induced by deuterium burning in the central, hydrostatic object. Note also the conversion of optical to infrared photons in the dust envelope.

Simple arguments suffice to demonstrate the vast difference in the character of the radiation field near the shock and at the dust photosphere. Gas approaches R_* with speeds that are close to the surface free-fall value $V_{\rm ff}$. This is

$$\begin{aligned} V_{\rm ff} &= \left(\frac{2\,G\,M_*}{R_*}\right)^{1/2} \\ &= 280\ {\rm km\ s^{-1}} \left(\frac{M_*}{1\ M_\odot}\right)^{1/2} \left(\frac{R_*}{5\ R_\odot}\right)^{-1/2} \ . \end{aligned} \tag{11.6}$$

Setting $V_{\rm ff}$ equal to $V_{\rm shock}$ in equation (8.50), we see that the immediate postshock temperature (called T_2 in Chapter 8) exceeds 10^6 K. Such hot gas emits photons in the extreme ultraviolet and soft X-ray regimes ($\lambda \approx hc/k_B T_2 \lesssim 100\,\text{Å}$). The emission here is mainly in lines from highly ionized metallic species, such as Fe IX. In any case, the material in both the postshock settling region and the radiative precursor is opaque to these photons. The protostar therefore radiates into the opacity gap almost as if it were a blackbody surface. The effective temperature of this surface, $T_{\rm eff}$, is found approximately from

$$4\,\pi\,R_*^2\,\sigma_B\,T_{\rm eff}^4 \approx L_{\rm acc} \ . \tag{11.7}$$

Substituting for $L_{\rm acc}$ from equation (11.5) and solving for the temperature, we obtain

$$\begin{aligned} T_{\rm eff} &\approx \left(\frac{G\, M_*\, \dot{M}}{4\,\pi\,\sigma_B\, R_*^3} \right)^{1/4} \\ &= 7300\ {\rm K} \left(\frac{\dot{M}}{10^{-5}\ M_\odot\ {\rm yr}^{-1}} \right)^{1/4} \left(\frac{M_*}{1\ M_\odot} \right)^{1/4} \left(\frac{R_*}{5\ R_\odot} \right)^{-3/4} . \end{aligned} \tag{11.8}$$

The quantity $T_{\rm eff}$ characterizes, at least roughly, the spectral energy distribution of the radiation field. We see that the opacity gap is bathed by optical emission similar to that emanating from a main-sequence star of similar mass. Throughout this volume, the characteristic temperature of the radiation does not vary markedly, although the outward, frequency-integrated flux $F_{\rm rad}$ falls off as r^{-2}. The gas temperature also declines slowly, from a value at the precursor that is not far below $T_{\rm eff}$.

11.1.4 Temperature of the Envelope

This situation changes dramatically upon crossing the dust destruction front. The infalling matter is now highly opaque to optical radiation, and the dominant photon frequencies shift downward through multiple absorptions and reemissions. In such environments, the temperature, which is identical for both matter and radiation, is related to $F_{\rm rad}$ through the radiative diffusion equation (see Appendix G). Setting $F_{\rm rad}$ equal to $L_{\rm acc}/4\pi r^2$ and changing the temperature gradient to $\partial T/\partial r$ (where the differentiation is at fixed time), equation (G.7) becomes:

$$T^3\, \frac{\partial T}{\partial r} = -\frac{3\,\rho\,\kappa\, L_{\rm acc}}{64\,\pi\,\sigma_B\, r^2} . \tag{11.9}$$

This relation governs the fall of the temperature throughout the dust envelope. Here, the density ρ follows equation (10.34). The Rosseland-mean opacity κ is dominated by the dust contribution. In the temperature regime of interest, from about 100 to 600 K, this quantity may be represented approximately by a power law:

$$\kappa \approx \kappa_\circ \left(\frac{T}{300\ {\rm K}} \right)^{\alpha} , \tag{11.10}$$

where $\kappa_\circ = 4.8\ {\rm cm}^2\ {\rm g}^{-1}$ and $\alpha = 0.8$. Note that the power-law behavior stems from the fact that the *monochromatic* opacity varies as $\lambda^{-\alpha}$ (see equation (G.9)). In any case, dimensional analysis of equation (11.9) then tells us that $T(r)$ falls off as $r^{-\gamma}$. Here, γ is another constant:

$$\gamma \equiv \frac{5}{2\,(4-\alpha)} , \tag{11.11}$$

which has a value near 0.8 in the present case.

The steady temperature decline continues until the gas becomes transparent to the infrared radiation. Roughly speaking, this transition occurs when the mean free path of the "average" photon, given by $1/\rho\kappa$, becomes comparable to the radial distance from the star. At this point, the entire dust envelope emanates as if it were a blackbody of radius $R_{\rm phot}$ and temperature $T_{\rm phot}$. Our two conditions are therefore

$$\rho\,\kappa\,R_{\rm phot} = 1 \tag{11.12a}$$

$$L_{\rm acc} = 4\,\pi\,R_{\rm phot}^2\,\sigma_B\,T_{\rm phot}^4\ . \tag{11.12b}$$

With ρ given by equation (10.34), $L_{\rm acc}$ by equation (11.5), and κ by (11.10), these constitute two equations in the unknowns $R_{\rm phot}$ and $T_{\rm phot}$. For $\dot{M} = 10^{-5}\,M_\odot$ yr^{-1} and $M_* = 1\,M_\odot$, numerical solution yields $R_{\rm phot} = 2.1 \times 10^{14}$ cm and $T_{\rm phot} = 300$ K.

We emphasize that the last two relations are rather crude approximations, even within the context of our idealized, spherical protostar. Strictly speaking, there is no unique, photospheric boundary, since the medium becomes transparent to photons of varying wavelength at different radii. The same is true, of course, in a stellar atmosphere, but there a much sharper falloff in density pinpoints the boundary. It is best to visualize $R_{\rm phot}$ as the radius where a photon carrying the mean energy of the spectral distribution escapes the cloud. Our numerical result indicates that the wavelength of this photon is typically $\lambda \approx hc/k_B T_{\rm phot} = 49$ μm, falling within the far-infrared regime.

Moving beyond our simplified description to a more precise determination of the radiation field and the matter temperature is technically demanding, since one must include both highly opaque and nearly transparent regions. Deep inside the dust photosphere, the specific intensity I_ν is nearly isotropic and close to $B_\nu(T)$. Within the opacity gap, however, I_ν is highly anisotropic and peaks in the direction away from the central protostar. The intensity also becomes outwardly peaked in the more tenuous region close to the dust photosphere.

The most accurate numerical calculations solve for the radiation and matter properties in an iterative manner. Within the dust envelope, for example, one might first guess the spatial distribution of the temperature T_d, which nearly equals T_g. This guess provides, through equation (2.30), the emissivity j_ν at every grid point. Knowledge of this function allows one to integrate the radiative transfer equation (2.20) for the specific intensity I_ν, both as a function of radial distance and angle. Given the radiation field, equation (7.19) yields the heating rate of the grains. Equating this rate to the cooling (equation (7.36)) then gives a new estimate for T_d. One repeats the procedure until the calculated and guessed temperatures throughout the dust envelope agree to sufficient accuracy. The situation is more complicated once one includes the opacity gap, where photons arrive from both the radiative precursor and the hot dust just outside the destruction front.

In spite of these difficulties, theorists are providing increasingly accurate descriptions of the protostellar environment. Figure 11.3 shows the temperature profile from one numerical study incorporating a detailed radiative transfer calculation. Here, the central protostar is represented as a point source of luminosity, where $L_{\rm rad} = 26\,L_\odot$. The density in the dust envelope follows equation (10.34), with $M_* = 1\,M_\odot$ and $\dot{M} = 2 \times 10^{-6}\,M_\odot$ yr^{-1}. Note that this equation gives the *total* density; the dust fraction is taken to be 1 percent by mass. To model the opacity gap, the envelope encloses a central, evacuated cavity, whose radius of 0.2 AU is the position where $T_d = 1500$ K. Although the temperature falls swiftly just beyond this point, its subsequent decline is rather shallow and roughly follows a power law.

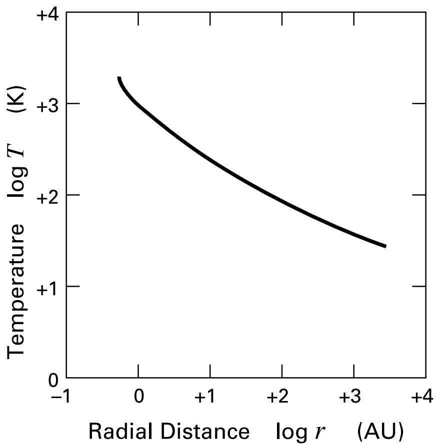


Figure 11.3 Temperature in the dust envelope of a spherical protostar of $1\,M_\odot$. The independent variable is radial distance from the star.

As expected, the envelope becomes transparent to outgoing radiation once its temperature falls below several 100 K. In this regime, the behavior of T_d follows from a simple energy argument. The radiative flux from the star falls off as r^{-2}. In addition, the emissivity of the dust grains varies as T_d^6, according to equation (7.39). It follows that T_d declines as $r^{-1/3}$. This optically thin profile is generally useful for modeling the observed emission at far-infrared and millimeter wavelengths from any dust cloud with an embedded star. (Recall the discussion of reflection nebulae in § 2.4.)

Returning to protostars *per se*, it is interesting to gauge the effect of rotation on the temperature distribution. Here, we may utilize the rotating infall model of § 10.4 and assign the nonspherical density distribution of equation (10.57). We replace both the protostar and its disk by a single point source whose radiation propagates through this envelope. The point-like representation is now more suspect, as the disk radius can easily extend past the dust destruction front (see § 11.3). On the other hand, most of the accretion luminosity still originates either on the stellar surface or within the inner region of the disk.

Figure 11.4 shows the temperature contours from a calculation of this type. Here, the parameters are $L_{\rm rad} = 21\,L_\odot$, $M_* = 0.5\,M_\odot$, and $\dot{M} = 5 \times 10^{-6}\,M_\odot$ yr^{-1}, while the adopted cloud rotation rate is $\Omega_\circ = 1.35 \times 10^{-14}$ s^{-1}. The figure also displays the appropriate isodensity contours. The reader may verify, using equation (10.50) with $m_\circ$ set equal to unity, that the centrifugal radius is $\varpi_{\rm cen} = 0.4$ AU, which indeed extends past the dust destruction front. The latter, shown by the innermost contour in the figure, corresponds to $T_d = 1050$ K, the sublimation temperature for the silicate grains that predominate in this model. Notice the slight oblateness of the inner cavity wall. The broadening stems from the infall density buildup near the centrifugal radius, which partially blocks outgoing radiation and increases the local dust temperature. Conversely, this ring-like enhancement in dust shadows the outer region and leads to modestly prolate temperature contours. It will be interesting to see how these results change once a circumstellar disk is included in the calculation. In any case, the contours in temperature should remain elongated in the polar direction, where the optical depth is relatively low.

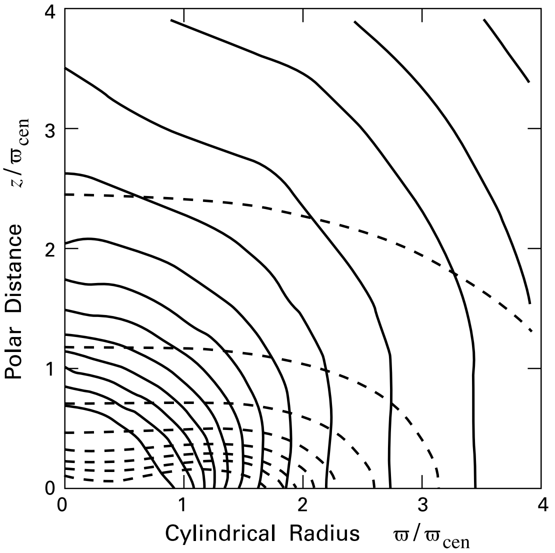


Figure 11.4 Two-dimensional distribution of temperature (*solid contours*) in a rotating, protostellar envelope. The temperature contours decrease outward from 1050 K in 50 K intervals. The dashed contours represent the density and are similar to those in Figure 10.18.

11.2 Interior Evolution: Deuterium Burning

Let us now shift our attention from the matter surrounding a protostar to the structure and evolution of the central object itself. As before, the emphasis is on theoretical, rather than observational, results. We want to present the underlying arguments with some care, as they will recur often in other contexts. Here, we will see that theory has provided a view of protostars almost as detailed as those for later evolutionary stages.

11.2.1 Stellar Structure Equations

The last point bears repeating. Protostars can, and should, be examined with the same computational techniques as red giants, white dwarfs, or main-sequence stars themselves. Since they accumulate at the center of a collapsing cloud, it is possible to include protostars as part of a numerical solution for the large-scale hydrodynamic flow. However, this procedure is both cumbersome and, given the realities of finite-difference calculations, prone to inaccuracy. One gains both physical insight and numerical precision by utilizing the *stellar structure equations* on the material inside the accretion shock front. One may then match conditions at the protostellar surface with those in the infalling envelope, where the latter is obtained through a hydrodynamic calculation.

The *mechanical* stellar structure equations are simply a restatement of hydrostatic balance. If we continue to neglect internal rotation, then a convenient spatial variable in our spherical protostar is the mass coordinate M_r, defined by equation (10.25). The radius r then acts as a dependent variable, whose variation with M_r is governed by the inversion of equation (10.26):

$$\frac{\partial r}{\partial M_r} = \frac{1}{4\,\pi\,r^2\,\rho} \; . \qquad (11.13)$$

As before, we take the derivative at fixed time. Equation (9.1) now expresses force balance and becomes

$$\frac{\partial P}{\partial r} = -\frac{G\,\rho\,M_r}{r^2} \qquad (11.14)$$

in spherical symmetry. Dividing (11.14) by (11.13), we derive the alternate form

$$\frac{\partial P}{\partial M_r} = -\frac{G\,M_r}{4\,\pi\,r^4} \; , \qquad (11.15)$$

which again utilizes M_r as the independent variable. The pressure itself obeys the equation of state for an ideal gas:

$$P = \frac{\rho}{\mu}\mathcal{R}\,T \; . \qquad (11.16)$$

Here the mean molecular weight μ depends on the state of ionization and dissociation of the gas, and is therefore a function of ρ and T. One may either calculate this function *ab initio*, by invoking statistical equilibrium, or obtain it from prior tabulations.

We turn next to the *thermal* stellar structure equations. Since the protostar's interior is highly opaque, the transport of radiation is governed by the diffusion equation (11.9). Replacing the luminosity variable by the internal value $L_{\rm int}$ and again utilizing (11.13), we have

$$T^3\,\frac{\partial T}{\partial M_r} = -\frac{3\,\kappa\,L_{\rm int}}{256\,\pi^2\,\sigma_B\,r^4} \; . \qquad (11.17)$$

The Rosseland mean opacity κ is again a function of ρ and T. Like μ, it is available in numerical form (see Figure G.2).

Finally, we consider the spatial variation of $L_{\rm int}$. Recall that the latter is the surface integral over a spherical shell of $|\boldsymbol{F}_{\rm rad}|$, where the flux points radially outward. In general, a fluid element gains heat either by being irradiated externally or from internal nuclear reactions. Let $\epsilon\,(\rho, T)$ represent the rate of nuclear energy release per unit mass. Both this rate and $\boldsymbol{F}_{\rm rad}$ enter as source terms in the *heat equation*:

$$\rho\,T\,\frac{\partial s}{\partial t} = \rho\,\epsilon - \boldsymbol{\nabla}\cdot\boldsymbol{F}_{\rm rad} \; , \qquad (11.18)$$

where s is the entropy per unit mass of the fluid. For a spherical star, it is convenient to recast this relation as

$$\frac{\partial L_{\rm int}}{\partial M_r} = \epsilon - T\,\frac{\partial s}{\partial t} \; . \qquad (11.19)$$

Equations (11.13), (11.15), (11.17), and (11.19) are the desired stellar structure equations for the dependent variables r, P, T, and $L_{\rm int}$. They must be supplemented by the equation of state (11.16) and by knowledge of μ, κ, ϵ, and s as functions of ρ and T. In practice, tabulations of the entropy utilize the Second Law of Thermodynamics:

$$T\,\Delta s \,=\, c_v\,\Delta T \,-\, P\,\Delta\rho/\rho^2 \ , \tag{11.20}$$

which governs small changes of the state variables. Here, c_v is the specific heat at constant volume. Knowing this latter quantity (or, equivalently, the specific internal energy) as a function of ρ and T allows one to integrate equation (11.20) along some convenient path in the $\rho - T$ plane and numerically obtain the function $s(\rho, T)$. Note that a purely monatomic gas would have $c_v = (3/2)\mathcal{R}/\mu$, where μ is now a constant. In this case, we may integrate equation (11.20) analytically to obtain

$$s \,=\, \frac{\mathcal{R}}{\mu}\,\ln\!\left(\frac{T^{3/2}}{\rho}\right) \,+\, s_\circ \ , \tag{11.21}$$

where $s_\circ$ is an arbitrary constant. Equation (11.21) is often a useful approximation in stellar interiors, where the gas is fully ionized and $\mu = 0.61$.

11.2.2 Boundary Conditions

Solution of the stellar structure equations requires specification of four boundary conditions. Two of these are the statements that both $r\,(M_r)$ and $L_{\rm int}\,(M_r)$ vanish at the center of the configuration:

$$r\,(0) \,=\, 0 \tag{11.22a}$$

$$L_{\rm int}\,(0) \,=\, 0 \ . \tag{11.22b}$$

A third condition is that $P\,(M_r)$ must equal the appropriate postshock value when $M_r = M_*$. This value is the ram pressure due to infalling matter, given by ρu^2. Using equation (10.34) for ρ just outside the shock and (11.6) for $u = -V_{\rm ff}$, we find

$$P\,(M_*) \,=\, \frac{\dot{M}}{4\,\pi}\left(\frac{2\,G\,M_*}{R_*^5}\right)^{1/2} \ . \tag{11.23}$$

The fourth boundary condition concerns the surface value of the temperature and its relation to the luminosity. For a main-sequence star, this relation would be the standard photospheric one given by equation (1.5). The total luminosity of a protostar, however, is that released by accretion plus the amount radiated from the interior:

$$L_* \,=\, L_{\rm acc} \,+\, L_{\rm post} \ . \tag{11.24}$$

Here, $L_{\rm post}$ is the value of $L_{\rm int}$ obtained by integration of equation (11.19) from $M_r = 0$ to M_*, and refers to the inner border of the postshock relaxation region (recall Figure 8.10). Let $T_{\rm post}$ denote the corresponding temperature, as found by outward integration of (11.17). Note that this value is much higher than that indicated in Figure 8.10, which concerns the surface layers of molecular clouds rather than optically thick stars. In any case, our task now is to relate $T_{\rm post}$ to $L_{\rm post}$ and $L_{\rm acc}$.

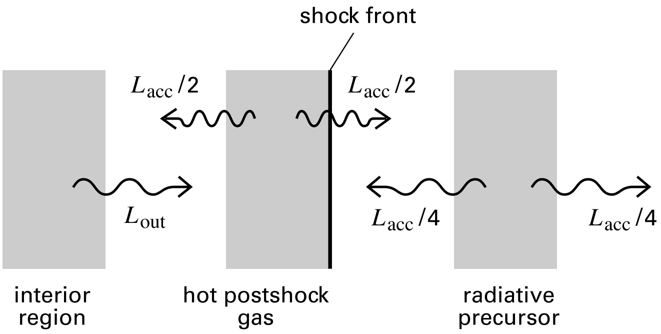


Figure 11.5 Contributions to the luminosity of a low-mass protostar. The two regions in front of and behind the accretion shock front (*thick vertical line*) have been separated for clarity. The net surface luminosity, called $L_{\rm post}$ in the text, is the sum of $L_{\rm out}$, the outward contribution from the interior, and inward contributions in X-rays and optical photons from the hot, postshock gas and radiative precursor, respectively.

We first remark that $L_{\rm post}$ is actually the sum of inward and outward contributions, as illustrated in Figure 11.5. It is only the outward luminosity, stemming from the deeper interior, that is given by the blackbody formula $4\pi R_*^2 \sigma_B T_{\rm post}^4$. This contribution is denoted $L_{\rm out}$ in the figure. The hot gas immediately behind the shock emits its soft X-rays isotropically. Thus, the inward contribution to $L_{\rm post}$ includes $L_{\rm acc}/2$ in X-rays. Additionally, there are optical photons stemming from the precursor. Assuming that this layer, which completely absorbs the X-rays entering it, also radiates equally in both directions, the postshock point receives an additional $L_{\rm acc}/4$ in optical photons.[1] In summary, we find that

$$L_{\rm post} = 4\pi R_*^2 \sigma_B T_{\rm post}^4 - 3 L_{\rm acc}/4 \ , \tag{11.25}$$

which is the desired boundary condition.

11.2.3 Mass-Radius Relation

The final ingredient needed to construct protostar models is the mass accretion rate $\dot{M}$. This quantity enters the boundary conditions of equations (11.23) and (11.25), and tells us, of course, how much to increase M_* from one time step to the next. Ideally, one should take this rate directly from collapse calculations, such as the ones depicted in Figure 10.6. Thus far, however, the most detailed studies of protostar interiors have considered only constant rates, effectively treating $\dot{M}$ as a free parameter. The range of this parameter follows by inserting plausible cloud temperatures (or, equivalently, sound speeds), into equation (10.31). For cloud temperatures from 10 to 20 K, one finds that $\dot{M}$ should span an order of magnitude, from about 10^{-6} to $10^{-5}\ M_\odot\ {\rm yr}^{-1}$.

---

[1] In low-mass protostars, the precursor is opaque to X-rays generated behind the shock, but transparent to the optical photons emitted locally. Thus, equation (11.8) provides only a rough approximation to the gas temperature in this region. The precursors in higher-mass protostars are opaque to *both* the X-rays and to their own cooling radiation.

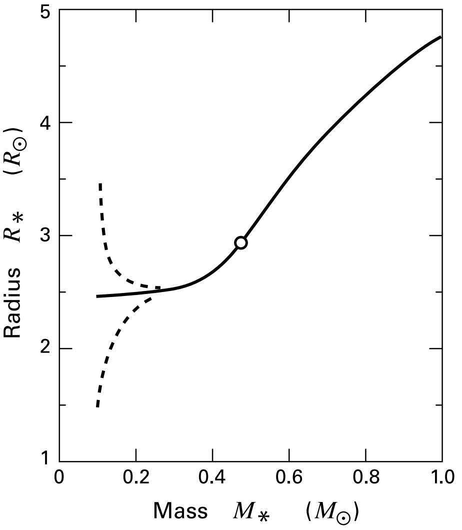


Figure 11.6 Mass-radius relation for spherical protostars with different initial radii. The accretion rate in all cases is $1 \times 10^{-5}\, M_\odot$ yr^{-1}. The open circle marks the onset of a fully convective interior.

In practice, we may solve the four structure equations at any time t by guessing central values for T and P, using equation (11.22) for $r\,(0)$ and $L_{\rm int}\,(0)$, and integrating outward. Along the way, we keep track of the specific entropy $s\,(M_r)$. We subtract, at each M_r-value, the corresponding entropy in the previous model for time $t - \Delta t$. We thus generate the temporal derivative ds/dt on the right side of (11.19). In general, equations (11.23) and (11.25) will not be satisfied when the integration reaches $M_r = M_*$. We therefore alter our guesses for the central temperature and pressure until these two conditions are met.

The procedure just outlined allows us to follow in detail the evolution of spherical protostars. The initial state, corresponding to the beginning of the main accretion phase, is somewhat arbitrary, reflecting our meager knowledge of this earliest epoch. Fortunately, the choice here has little impact on subsequent evolution. Figure 11.6 illustrates this point graphically, by showing the protostellar radius as a function of mass for $M_* \leq 1\, M_\odot$, assuming a constant accretion rate of $1 \times 10^{-5}\, M_\odot$ yr^{-1}. The three curves each have different values for R_* at the initial mass, here taken to be $0.1\, M_\odot$. By the time M_* has doubled, the curves are nearly identical.

To understand the rapid convergence of $R_*(M_*)$ as well as its subsequent, steady rise, it is helpful to view the accreting protostar as a collection of nested mass shells. Each new shell represents matter that has just passed through the shock front and settled onto the hydrostatic surface. As was the case for the first core, a settling fluid element is quickly smothered by overlying layers, *i. e.*, we have $L_{\rm post} \ll L_{\rm acc}$. In the absence of nuclear burning, equation (11.18) tells us that the specific entropy s for that mass shell ceases to fall and reaches a constant value. The protostar is thus characterized by its *entropy profile* $s\,(M_r)$, which in turn reflects the changing conditions at the accretion shock front.

Since the specific entropy represents the heat content of the associated mass shell, an increase of s with M_r results in a protostar that swells as mass is added. An entropy profile of this character arises naturally, as the accretion shock generally becomes stronger with time. Thus, any increase in the gravitating mass M_* raises the incoming velocity $V_{\rm ff}$. Consequently, the

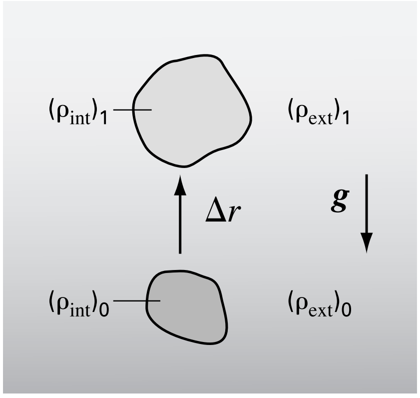


Figure 11.7 Criterion for radiative stability. A fluid element, initially with the same density as its surroundings, moves a distance Δr opposite to the gravitational acceleration $\boldsymbol{g}$. In the case shown, the interior density $\rho_{\rm int}$, while decreasing, is nonetheless greater than the exterior value $\rho_{\rm ext}$, so that the element will fall back down.

postshock temperature T_2 also goes up, as does the settled value $T_{\rm post}$ and the corresponding entropy. In summary, a stronger accretion shock leads to a monotonically increasing $s\,(M_r)$ for the interior. The tendency for such a profile to develop accounts for the ultimate rise of $R_*\,(M_*)$ in Figure 11.6.

Suppose, however, that the *initial* radius were very large, for whatever reason. Then $V_{\rm ff}$ would be correspondingly low, and the shock would be weaker. With the addition of mass shells, the entropy profile would dip, and the radius shrink. Conversely, a very small initial radius would result in such a strong shock that $s\,(M_r)$ would rise steeply. The resultant swelling of the radius is also evident in Figure 11.6. Despite our ignorance of the initial state, these calculations provide strong evidence that the natural evolution consists of a gentle increase of radius with mass, as shown by the middle curve in the figure.

11.2.4 Onset of Convection

The rising entropy profile is also of fundamental importance for *convective stability*. Imagine, as in Figure 11.7, that we displace a fluid element with internal density $\rho_{\rm int}$ by a small distance Δr in the direction opposite to the local gravitational acceleration $\boldsymbol{g}$, *i. e.*, toward the surface of the protostar. We know from our study of hydrostatic equilibria that the pressure always declines in this direction. Hence our fluid element, if it is to maintain pressure balance with its surroundings, must expand, and its density falls from $(\rho_{\rm int})_0$ to the lower value $(\rho_{\rm int})_1$. The salient question is whether this drop in density makes the element so buoyant that it continues to travel upward. If so, the protostar is *convectively unstable* in that region, and the upward motion becomes an important means of transporting heat. If, on the other hand, displacement makes the element denser than its surroundings, it will sink back down. The protostar is then *radiatively stable*. We now prove that this latter condition holds as long as $s\,(M_r)$ is an increasing function.

We proceed by comparing $(\rho_{\rm int})_1$ to $(\rho_{\rm ext})_1$, the external, background density at the same location. If the small displacement occurs so quickly that the element loses negligible heat through radiation, its specific entropy cannot change. Then a rising entropy profile in the external medium implies that $(s_{\rm int})_1 < (s_{\rm ext})_1$. Recall that the internal and external pressures are equal at this location. It is a property of ordinary gases that the density falls with increasing specific entropy at fixed pressure, *i. e.*, that $(\partial\rho/\partial s)_P < 0$. [2] For a monatomic gas, the reader may verify this fact by combining equation (11.21) with the equation of state (11.16). In any event, we have now demonstrated that $(\rho_{\rm int})_1 > (\rho_{\rm ext})_1$, so that *a rising entropy profile implies radiative stability*. Our condition for stability, $\partial s/\partial M_r > 0$, is one expression of *Schwarzschild's criterion*, which figures prominently in stellar structure theory.

Returning to our protostar, we now see that the strengthening accretion shock leads to a structure in which heat is transported outward by radiation, rather than the convective motion of fluid elements. We already *assumed* such a state, in fact, when we adopted the radiative diffusion equation (11.17). However, this situation cannot last forever. As is evident from Figure 11.6, the initial swelling of the radius does not keep pace with the growth of M_*. That is, the ratio M_*/R_* inexorably rises. The interior temperature also climbs, as can be seen from equation (11.2) after an appropriate change of μ. Nuclear reactions eventually begin near the center, where the temperature peaks. Their effect is to increase the central entropy until the profile overturns, *i. e.*, until $\partial s/\partial M_r$ becomes negative. Mathematically, the term ϵ in equation (11.19) grows large near $M_r = 0$. Since $L_{\rm int}$ is kept relatively small by the high optical depth, the derivative $\partial s/\partial t$ increases, as well. The important, physical consequence is that the protostar becomes convectively unstable.

The first nuclear fuel to ignite is a small admixture of the hydrogen isotope deuterium (^{2}H). Consisting of a proton bound to a single neutron, the deuteron is a product of primordial nucleosynthesis, forged during the universal cooling and expansion minutes after the Big Bang. Once incorporated into stars, the isotope is destroyed by fusion with protons:

$$^2{\rm H} + {}^1{\rm H} \rightarrow {}^3{\rm He} + \gamma \, . \qquad (11.26)$$

This reaction is exothermic by $\Delta E_D \equiv 5.5$ MeV. As in many nuclear processes, the associated rate is highly sensitive to the local temperature. Deuterium fusion first becomes appreciable near 10^6 K. Close to this temperature, calculations show that the energy generation term in equation (11.19) may be approximated by

$$\epsilon_D = [{\rm D/H}]\,\epsilon_\circ \left(\frac{\rho}{1\ {\rm g\ cm^{-3}}}\right) \left(\frac{T}{1 \times 10^6\ {\rm K}}\right)^{n_D} \, . \qquad (11.27)$$

Here, $\epsilon_\circ = 4.19 \times 10^7$ erg g^{-1} s^{-1}, $n_D = 11.8$, and [D/H] is the interstellar number density of deuterium relative to hydrogen. This latter ratio must be obtained from observation and has been the object of considerable study. Analysis of the absorption lines of stars lying behind diffuse clouds gives a mean [D/H] of 2×10^{-5}, with the apparent variation being about a factor of two.

Convection begins in protostars because deuterium fusion produces too much luminosity to be transported radiatively through the highly opaque interior. Instead, discrete "cells" of fluid rise buoyantly toward the surface. Since $\partial s/\partial M_r < 0$, the rising fluid is now both underdense and hot relative to its surroundings, and eventually transfers the excess heat to this medium.

---

[2] We may write the partial derivative as $(\partial\rho/\partial s)_P = -\rho T \kappa_P/c_P$, where $\kappa_P \equiv [\rho\,\partial(1/\rho)/\partial T]_P$ is the thermal expansion coefficient, and where $c_P \equiv T(\partial s/\partial T)_P$ is the specific heat at constant pressure. Under all conditions encountered in stars, both κ_P and c_P are positive, so that $(\partial s/\partial \rho)_P$ is indeed negative.

Having done so, the cooler and denser cell then sinks back down, and the cycle repeats. Such rising and falling motion must render protostellar interiors both turbulent and chemically well mixed.

From a computational perspective, the onset of convection modifies the thermal stellar structure equations. The luminosity $L_{\rm int}$ still follows from integration of the heat equation (11.19). However, the diffusion equation (11.17) is no longer valid and must be replaced by another relation between $L_{\rm int}$ and the temperature (or entropy) gradient. Because of the turbulent nature of convection, there is still no generally accepted relation derived from first principles. Instead, theorists have traditionally relied on a semi-empirical model known as *mixing-length theory*. Here, every cell at a given radius r expels its heat after traversing the same distance. This "mixing length" is usually taken to be some factor of order unity times the local pressure scale height, *i. e.*, the radial distance over which $P(r)$ would drop by e^{-1}. The cell velocity, and hence the heat flux, follows by assuming a steady, buoyant acceleration over this distance.

The detailed formulation of mixing-length theory need not concern us, but its general consequences are of interest. The main result is that the transport of heat through interior motion is highly efficient. That is, even when their average upward velocity is well below the local sound speed, the rising cells need only a slight gradient in the background entropy profile to carry any reasonable $L_{\rm int}$. To good accuracy, therefore, we may replace equation (11.17) by the simpler relation

$$\frac{\partial s}{\partial M_r} = 0 \ . \tag{11.28}$$

We emphasize that convective instability is a *local* phenomenon, in the sense that some regions of the star may be unstable while others remain radiatively stable. The computational procedure must account for this fact by testing for stability at each M_r-value. A convenient method is to compare $L_{\rm int}$ with $L_{\rm crit}$, the maximum value that can be carried at that location by radiative diffusion. To obtain this important quantity, we first divide (11.17) by (11.15), in order to express the radiative luminosity in terms of $\partial T/\partial P$. Recall that the differentiation here is performed at constant *time*. Just at the stability limit, it is the *specific entropy* which is held fixed, *i. e.*, $\partial T/\partial P$ becomes the thermodynamic derivative $(\partial T/\partial P)_s$. This latter term may be found numerically from the equation of state and entropy tabulations. We thus obtain

$$L_{\rm crit} = \frac{64\,\pi\,G\,M_r\,\sigma_B\,T^3}{3\,\kappa}\left(\frac{\partial T}{\partial P}\right)_s \ . \tag{11.29}$$

In any convection zone, we use equation (11.28) until $L_{\rm int}$ falls below $L_{\rm crit}$. At that point, we revert to (11.17). Note that such a transition *always* occurs at least once before the surface is reached, if only because the falling density eventually renders convection inefficient. Thus, the outermost layers of any star are radiatively stable.[3]

---

[3] Just prior to the onset of radiative stability, convection may become so inefficient that the (negative) entropy gradient is no longer small. We must then replace equation (11.28) by the full mixing-length relation between $\partial s/\partial M_r$ and $L_{\rm int}$. Such *super-adiabatic* regions are important in the outer portions of pre-main-sequence stars, as we will discuss in Chapter 16.

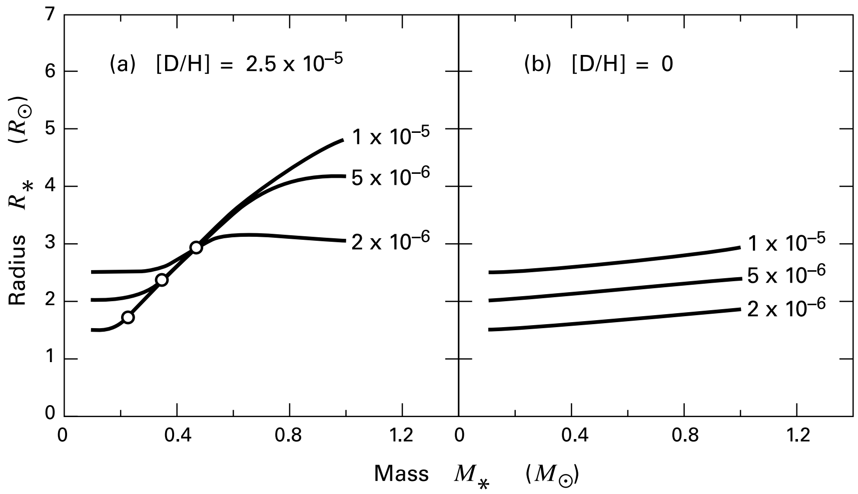


Figure 11.8 (a) Protostellar mass-radius relation for three different accretion rates, where the latter are written in units of $M_\odot$ yr^{-1}. A standard interstellar concentration of deuterium is assumed. (b) Mass-radius relation for the same accretion rates, but in the absence of deuterium.

11.2.5 Deuterium Thermostat

Once deuterium ignites near the center of a low-mass protostar, the induced convection quickly spreads. Soon, the whole interior is unstable, apart from a thin, outer settling region of negligible mass. The open circle in Figure 11.6 marks the onset of full convection. By this point, deuterium burning has significantly increased the protostar's radius. The degree of swelling depends somewhat on the assumed accretion rate. Figure 11.8a displays $R_*(M_*)$ for three different $\dot{M}$-values. The increase in radius is a rather transient effect for the lowest $\dot{M}$ shown, while the curves would change slowly for rates above 1×10^{-5} $M_\odot$ yr^{-1}. Once again, the open circles indicate where the protostar becomes fully convective.

To emphasize the structural change created by deuterium, Figure 11.8b shows an analogous set of radius curves, obtained by artificially setting [D/H] equal to zero in the models. The radius now continues its previous, gentle rise, and the curves are well-separated by their $\dot{M}$-values. In reality, such a modest radius increase is impossible once deuterium ignites. As we have seen, the associated climb in M_*/R_* would raise the central temperature T_c. However, ϵ_D is so temperature-sensitive that it would increase substantially, pumping enough heat into the protostar to swell the radius and lower T_c. In this manner, deuterium acts as a *thermostat*, trying to maintain the central temperature close to the ignition value of 1×10^6 K.

The thermostat is only effective if the protostar has a steady supply of the nuclear fuel. In addition, the temperature-sensitivity of ϵ_D implies that active burning is confined to the central region. The original deuterium residing here is quickly consumed after ignition. While additional fuel continually lands on the surface through infall, it cannot reach the deep interior as long as the protostar is radiatively stable. Once the interior is fully convective, however, turbulent eddies drag this deuterium toward the central furnace. Since this transport time is relatively

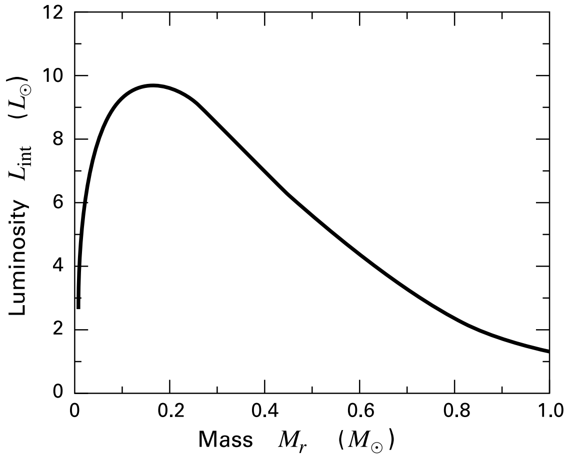


Figure 11.9 Luminosity within a protostar of 1 $M_\odot$. The independent variable is the mass enclosed within a spherical shell. An accretion rate of 1×10^{-5} $M_\odot$ yr^{-1} has been assumed.

brief, the consumption rate approaches that supplied from infall. Thus, the total luminosity L_D generated by deuterium is close to its steady-state value:

$$\begin{aligned} L_D &\equiv \int_0^{M_*} \epsilon_D \, dM_r \\ &\approx \dot{M}\, \delta \ . \end{aligned} \tag{11.30}$$

Here, δ is the energy available in deuterium per gram of interstellar gas:

$$\delta = \frac{[\mathrm{D/H}]\, X\, \Delta E_D}{m_H} \ . \tag{11.31}$$

The steady-state L_D is 12 $L_\odot$ for an accretion rate of 1×10^{-5} $M_\odot$ yr^{-1}.

The thermostat would *still* be ineffective if most of this deuterium-produced luminosity escaped the protostar. We have already seen, however, that $L_{\rm post}$ in the radiatively stable post-shock layers is relatively small, since this material is smothered by the freshly accreting gas. At any time, therefore, $L_{\rm int}(M_r)$ does rise at first to a high level close to L_D but then eventually falls (see Figure 11.9). Equation (11.19) shows that this outward, *spatial* decline is accompanied by a *temporal* rise in the nearly uniform, interior entropy. Finally, the protostar's luminosity jumps upward across the shock to the accretion value $L_{\rm acc}$, which is 60 $L_\odot$ in our sample calculation. In summary, we now see that the entropy increase and swelling of the protostar stem from two distinct sources – backheating from the accretion shock, providing $(3/4)\, L_{\rm acc}$ in luminosity, and internal heating from deuterium burning, which gives a smaller, but still important contribution.

11.3 Protostellar Disks

One key simplification in our analysis has been the assumption that fluid elements from the cloud envelope directly impact the protostar. However, we saw in Chapter 10 that rotation distorts the infalling trajectories. Material with sufficiently large specific angular momentum lands first in a disk and only later spirals to the stellar surface. The protostellar disk is of considerable interest itself, since it is the medium out of which planets ultimately form. We therefore want to trace the main theoretical ideas concerning its origin and growth. Along the way, we discuss the still problematic issue of mass transport onto the central star.

11.3.1 First Appearance

A disk is born when infalling gas begins to miss the protostellar surface. The analysis of rotating collapse in § 10.4 allows us to pinpoint this time, along with the associated stellar mass. Recall that the centrifugal radius $\varpi_{\rm cen}$ marks the greatest distance where fluid elements cross the equatorial plane. Equating this quantity, as given by equation (10.50), to the protostar radius R_*, we solve for the critical time:

$$\begin{aligned} t_0 &= \left(\frac{16\,R_*}{m_\circ^3\,\Omega_\circ^2\,a_T}\right)^{1/3} \\ &= 3\times 10^4\ {\rm yr}\left(\frac{R_*}{3\,R_\odot}\right)^{1/3}\left(\frac{\Omega_\circ}{10^{-14}\ {\rm s}^{-1}}\right)^{-2/3}\left(\frac{a_T}{0.3\ {\rm km\ s}^{-1}}\right)^{-1/3}\ , \end{aligned} \tag{11.32}$$

where we have set $m_\circ = 1$. The mass of the protostar at this time is then $M_0 \equiv \dot{M} t_0$. Using equation (10.31) for $\dot{M}$, this is

$$\begin{aligned} M_0 &= \left(\frac{16\,R_*\,a_T^8}{G^3\,\Omega_\circ^2}\right)^{1/3} \\ &= 0.2\,M_\odot\left(\frac{R_*}{3\,R_\odot}\right)^{1/3}\left(\frac{\Omega_\circ}{10^{-14}\ {\rm s}^{-1}}\right)^{-2/3}\left(\frac{a_T}{0.3\ {\rm km\ s}^{-1}}\right)^{8/3}\ . \end{aligned} \tag{11.33}$$

Equation (11.33) gives the reasonable result that the protostar mass at the time of disk formation increases with either a greater a_T (*i. e.*, higher infall rate) or a lower $\Omega_\circ$. Taken at face value, the numerical figure indicates that many stars could accumulate entirely by direct infall. However, our uncertainty regarding the proper values of infall and rotation rates is great enough that we should continue to explore the more general formation mechanism, even for stars at the lowest masses.

The third factor entering equations (11.32) and (11.33) is the radius R_*, and here the reader may have already noticed a problem. The "typical" value chosen is indeed compatible with M_0, but only for the *spherical* protostars considered previously. Protostars forming out of *rotating* gas should spin up and be centrifugally distorted, much like the clouds we considered in § 10.2. However, we know empirically that even the youngest pre-main-sequence objects – those near the birthline – have low rotation speeds, at least in their outer layers (Chapter 16). This fact is a strong indication that their antecedent protostars did *not* simply accumulate the angular momentum brought in by collapsing cloud matter. Instead, protostars must experience a strong, braking torque, presumably associated with the shedding of an MHD wind (Chapter 13). We are therefore justified in picturing them as almost spherical objects, even while discussing the rapid rotation that characterizes their near environments.

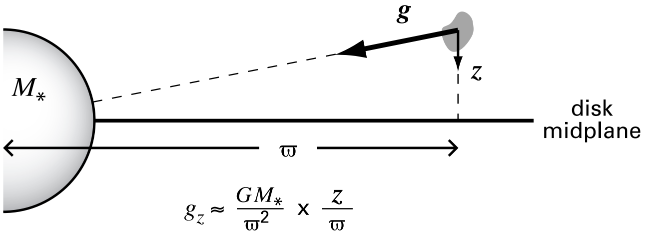


Figure 11.10 Physical basis of the thin-disk approximation. A fluid element within a low-mass disk is pulled toward the midplane by the z-component of the gravitational force exerted by the star.

Another key distinction between stars and disks is their geometrical thickness in the z-direction. In both cases, the internal force balance is between the upward thermal pressure gradient and the downward pull of gravity. In the younger disks of relatively low mass, the latter is simply the z-component of the radial acceleration from the central star (see Figure 11.10). Vertical force balance for an element located a height z above the midplane thus reads

$$-\frac{\partial P}{\partial z} = \frac{\rho G M_* z}{\varpi^3} .$$

In conjunction with the equation of state (11.16), this relation allows us to estimate the scale height Δz of the disk, *i. e.*, the distance over which P falls appreciably:

$$\Delta z \approx \left(\frac{a_T}{V_{\mathrm{Kep}}}\right) \varpi . \qquad (11.34)$$

Here, a_T is the average sound speed within the disk interior at radius ϖ, while $V_{\mathrm{Kep}} \equiv \sqrt{GM_*/\varpi}$ is the Keplerian orbital speed. Note that V_{Kep} is close to V_{ff} at the radius in question. We already saw, in our discussion following equation (11.6), that typical values of V_{ff} at R_* correspond to temperatures of at least 10^6 K. This is far above the effective temperature of the precursor radiation field (see equation (11.8)), which largely sets the inner-disk value of a_T. With increasing distance from the star, both the disk temperature and a_T drop even further, and the inequality $a_T \ll V_{\mathrm{Kep}}$ is always well satisfied. Hence, we may conveniently adopt the *thin-disk approximation*, $\Delta z \ll \varpi$. The reader should verify that the condition $a_T \ll V_{\mathrm{Kep}}$ also implies that the *radial* pressure gradient $|\partial P/\partial \varpi|$ is much smaller than the radial force per unit volume due to either stellar gravity or rotation.

For times greater than t_0, an increasing portion of the collapsing gas misses the star. We stress that this shift has no bearing on the *total* mass infall rate onto the equatorial plane. The latter is still given by $\dot{M}$ from equation (10.31) and is set by the equilibrium cloud structure at a much greater distance scale. In any event, the changing pattern of collapse is symmetric above and below the central plane. A trajectory that misses the star encounters its mirror image with reversed z-velocity, as depicted in Figure 11.11. The two opposing streams collide at supersonic speed. Consequently, the accretion shock now covers more than just the protostar and has effectively broadened to include the whole region inside the expanding centrifugal radius ϖ_{cen}.

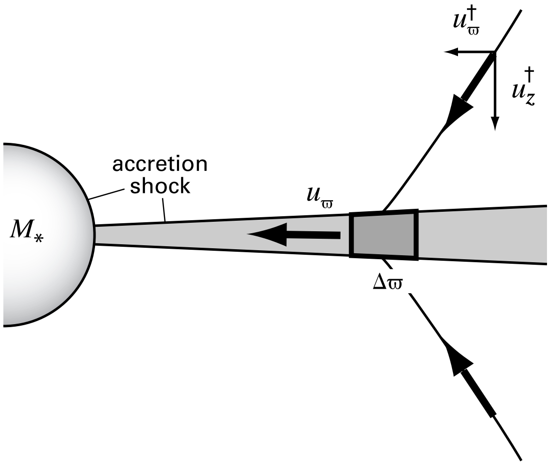


Figure 11.11 Collision of streamlines during rotating, protostellar collapse. The velocity of any infalling element consists of vertical and horizontal components. An element within the disk acquires a horizontal velocity u_ϖ from impact of the external streams. The accretion shock now covers both the protostellar and disk surfaces.

Gas entering this extended front preserves the component of its momentum lying in the equatorial plane. The accretion shock thus serves to deflect this portion of the infall back toward the protostar (Figure 11.11). Indeed, the mass of the central object continues to grow at the rate $\dot{M}$, as long as this fluid promptly reaches the protostellar surface. Since there is no time for matter to build up appreciably in the equatorial plane, the disk during this epoch is a structure of very low surface density. Note that the infalling material still has a finite angular momentum about the cloud rotation axis after passing through the shock. Once $\varpi_{\rm cen}$ grows large enough, the outermost fluid elements can no longer penetrate to R_*, and the character of the disk radically changes.

11.3.2 Evolutionary Equations

We may quantify these ideas by first deriving the properties of the flow that impacts the disk from above and below. Returning to our rotating infall model of Chapter 10, we set θ equal to $\pi/2$ in equation (10.52) to find where any trajectory lands relative to the disk's outer edge:

$$\frac{\varpi}{\varpi_{\rm cen}} = \sin^2\theta_\circ \ . \qquad (11.35)$$

Similar specialization of equations (10.53), (10.55), and (10.56) yields the three velocity components immediately above the disk, while (10.57) gives the density. After transforming to cylindrical coordinates, we obtain

$$u_\varpi^\dagger = -\left(\frac{G\,M_*}{\varpi}\right)^{1/2} \tag{11.36a}$$

$$u_z^\dagger = -\left(\frac{G\,M_*}{\varpi}\right)^{1/2} \cos\,\theta_\circ \tag{11.36b}$$

$$u_\phi^\dagger = +\left(\frac{G\,M_*}{\varpi}\right)^{1/2} \sin\,\theta_\circ \tag{11.36c}$$

$$\rho^\dagger = -\frac{\dot{M}}{8\,\pi\,\varpi^2\,u_\varpi^\dagger}\tan^2\theta_\circ\ . \tag{11.36d}$$

Here, the † notation distinguishes infall quantities from those within the disk. Note from (11.35) and (11.36c) that $u_z^\dagger$ declines with increasing ϖ. Thus, the accretion shock gradually weakens as the disk spreads.

We next employ conservation principles to derive the three basic evolutionary equations obeyed by the disk matter. Referring again to Figure 11.11, we first note that a total mass per unit time of $-4\pi\varpi\,(\rho u_z)^\dagger\Delta\varpi$ enters the small annulus of radius ϖ and width $\Delta\varpi$. If $\Sigma(\varpi)$ denotes the disk surface density, the *internal* rate of mass transport across any radius in the direction of the star is $\dot{M}_d \equiv -2\pi\varpi\Sigma u_\varpi$. This quantity is generally a function of ϖ, and any change $\Delta\dot{M}_d$ represents a small increase in the mass of our annulus. Since the latter is given by $2\pi\varpi\Sigma\Delta\varpi$, we find the appropriate expression of mass continuity:

$$\frac{\partial\Sigma}{\partial t} + \frac{1}{\varpi}\frac{\partial(\varpi\,\Sigma\,u_\varpi)}{\partial\varpi} + 2\,(\rho\,u_z)^\dagger = 0\ , \tag{11.37}$$

where we have again omitted all subscripts on the partial derivatives.

We may derive, in a similar fashion, the conservation laws governing radial and azimuthal momentum. In the former case, we must generally account for the gravitational force both from the protostar and from the disk itself. However, for this early phase, we may neglect the disk self-gravity. Recall that the internal pressure gradient can also safely be ignored. After some manipulation, we find the radial momentum equation to be

$$\frac{\partial u_\varpi}{\partial t} + u_\varpi\,\frac{\partial u_\varpi}{\partial\varpi} = \frac{j^2}{\varpi^3} - \frac{G\,M_*}{\varpi^2} - \frac{2\,(\rho\,u_z)^\dagger}{\Sigma}\left(u_\varpi^\dagger - u_\varpi\right) = 0\ . \tag{11.38}$$

The first term on the right side represents the centrifugal acceleration, which depends on the specific angular momentum $j \equiv \varpi u_\phi$. The final righthand term is the radial force per unit mass created by impacting gas. The infall also exerts an effective force in the ϕ-direction that enters the equation for azimuthal momentum conservation. Written in terms of j rather than u_ϕ, this latter relation is

$$\frac{\partial j}{\partial t} + u_\varpi\,\frac{\partial j}{\partial\varpi} = -\frac{2\,(\rho\,u_z)^\dagger}{\Sigma}\left(j^\dagger - j\right)\ . \tag{11.39}$$

At early times, the velocity components u_ϖ and u_ϕ are both of order $V_{\rm Kep}$ in magnitude. The internal crossing time for fluid elements is then brief compared to that required for $\varpi_{\rm cen}$

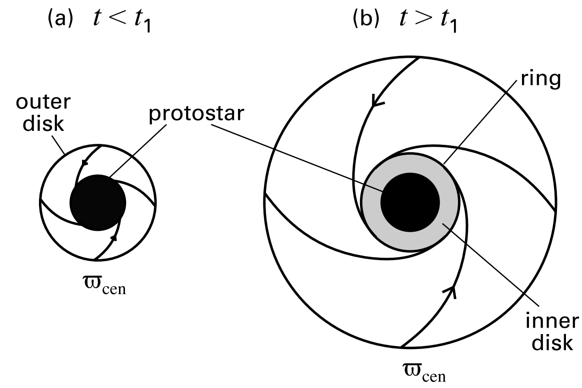


Figure 11.12 Early expansion of a protostellar disk. (a) Before time t_1, curved streamlines from the outer disk impact the protostar directly. (b) After t_1, these streamlines converge to form a dense ring, which transfers mass to an inner disk surrounding the star. At all times, the outer disk boundary is the growing centrifugal radius $\varpi_{\rm cen}$.

to increase significantly. We may thus apply a steady-state approximation to our derived relations. Dropping all explicit time derivatives from (11.37)–(11.39), these become three *ordinary* differential equations in the dependent variables Σ, u_ϖ, and j, and may be solved by standard techniques.

11.3.3 Inner and Outer Disks

Figure 11.12a, based on a numerical solution of the steady-state equations, shows several representative streamlines in the disk at this earliest epoch. These streamlines trace the inward motion of fluid elements which, we should remember, are continually gaining mass from the infalling envelope. The axisymmetry of the infall implies that any one curve can be rotated about the origin to generate all the others. Note also that the entire spiral pattern expands as t^3, reflecting the similar expansion of $\varpi_{\rm cen}$ and the infall trajectories (recall Figure 10.18). Thus, as shown in Figure 11.12b, the streamlines eventually miss the central protostar. This transition is an inevitable consequence of the disk's buildup of angular momentum.

Figure 11.12b indicates how each streamline collides with its neighbor inside the tangent circle, *i. e.*, the radius where u_ϖ vanishes. The numerical integration gives a value for this radius of $0.34\,\varpi_{\rm cen}$. Thus, the streamline crossing occurs at the time t_1, where

$$\begin{aligned} t_1 &= (0.34)^{-1/3}\, t_0 \\ &= 1.43\, t_0 \ . \end{aligned} \tag{11.40}$$

For $t \gtrsim t_1$, the disk contains a highly dissipative, turbulent region just outside the protostellar surface. Once the disk expands even more, the fluid well inside the tangent circle should relax to nearly circular orbits, *i. e.*, to a configuration that eliminates streamline crossing and minimizes energy dissipation. This *inner disk* is therefore separated from the low-density outer portion by a *ring* of turbulent gas. Figure 11.12b depicts the ring as a relatively narrow region, although its

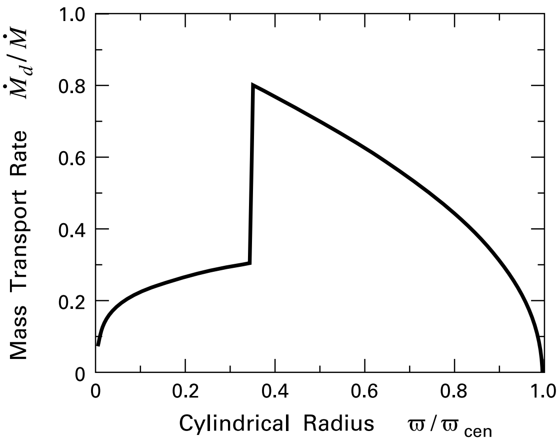


Figure 11.13 Mass transport rate within a protostellar disk. The rate is displayed as a fraction of the full $\dot{M}$ from the infalling cloud. The independent variable is the distance to the origin, measured relative to the expanding centrifugal radius $\varpi_{\rm cen}$.

true width has yet to be determined. The boundaries of all three components – the inner disk, ring, and outer disk – continue to expand as t^3.

Within the high-density inner disk, u_ϕ is close to the Keplerian value $(GM_*/\varpi)^{1/2}$. However, the orbits here cannot be precisely circular. We must still account for the azimuthal force arising from the impact of cloud matter, as seen in equation (11.39). Equation (11.36c) indicates that $u_\phi^\dagger$ is less than the Keplerian speed at any radius. Hence, the infall exerts a *drag* on the disk, causing the orbits to spiral inward. Because the surface density is now high, this drag force is relatively small, and the spirals are tightly wrapped. The product of Σ and u_ϖ entering $\dot{M}_d$ is nonetheless significant, *i. e.*, mass continues to flow across any radius at a substantial fraction of the total infall rate.

Mathematical analysis of the inner disk proceeds from the same evolutionary equations (11.37)–(11.39). Because of the slow radial drift, we cannot employ a steady-state approximation. However, the first two righthand terms in (11.38) dwarf all others in this equation, so we have an accurate estimate for $j \approx (GM_*\varpi)^{1/2}$. Substitution of this expression into equation (11.39) yields an algebraic relation between u_ϖ and Σ. We are thus able to write the partial differential equation (11.37) in terms of the surface density alone. The actual solution for $\Sigma(\varpi, t)$ requires proper matching to the outer disk across the ring interface.

Figure 11.13, again taken from the detailed integration, shows the variation within the disk of the mass transport rate. Notice that the radial coordinate here is actually $\varpi/\varpi_{\rm cen}$ and that $\dot{M}_d$ is displayed as a fraction of the infall rate $\dot{M}$. At any epoch, $\varpi/\varpi_{\rm cen}$ extends from some finite value up to unity at the disk edge. The innermost value of this ratio steadily falls as $\varpi_{\rm cen}$ grows. Hence, the figure is also useful for tracing the *temporal* evolution of the mass transport rate at this edge.

At $t = t_0$, ϖ is identically equal to $\varpi_{\rm cen}$ (which in turn equals R_*), and $\dot{M}_d$ is zero. Thereafter $\dot{M}_d$ at the inner disk edge monotonically rises, as the spreading disk incorporates more infalling gas, so that less impacts the star directly. A sharp downward plunge in $\dot{M}_d$ occurs at time t_1 and reflects the sudden appearance of the turbulent ring. The ring contains a finite, and increasing, amount of mass, so that $\dot{M}_d$ is spatially discontinuous for $t > t_1$. At even later times, the rate of mass transport onto the protostar's surface steadily falls. The drag from infall

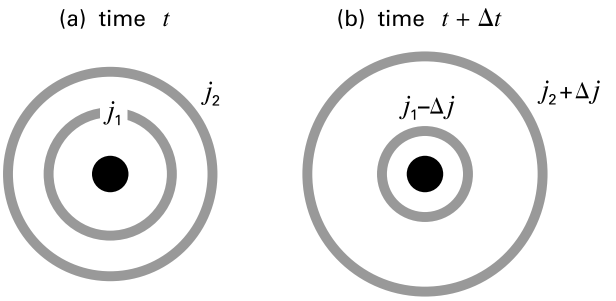


Figure 11.14 Torquing within a circumstellar disk. Over the time interval Δt, an inner annulus transfers angular momentum to an outer one. The first annulus thus contracts, while the second expands.

is now too weak to affect the inner disk, whose mass climbs nearly linearly with time at the rate $\dot{M}$. Concurrently, the protostar's mass rises only slowly, as nearly all the infalling trajectories fall far outside R_*.

11.3.4 Internal Torquing

The steady increase of the disk mass M_d for $t > t_1$ presents a serious problem. Should this rise continue unabated, M_d would soon outstrip M_*. Observed pre-main-sequence stars, however, always have disks with relatively low masses, typically a few percent of the stellar value. The theory just outlined makes it clear that such an extreme central concentration of matter in the star-disk system is *not* simply the outcome of gravitational collapse, which by itself would always spread mass more uniformly. This dilemma is one manifestation of the *angular momentum problem* that pervades star formation theory in much the same way as the magnetic flux issue. In both cases, naïve application of conservation laws leads to results squarely at odds with observation.

To avoid the rapid buildup of a massive disk, some process stronger than the drag from infall must allow material to spiral continually onto the protostar. Consider, as in Figure 11.14, two neighboring annuli at time t, where the inner annulus has a specific angular momentum j_1 and the outer has j_2. (Here, j_1 must be less than j_2 in order for the disk to be rotationally stable; recall § 9.2.) During a subsequent time interval Δt, the inner annulus can contract only if its specific angular momentum is reduced, say to $j_1 - \Delta j$. One way for such a reduction to occur would be if material with lower specific angular momentum were mixed into the region. Infall, as we have seen, supplies such material, and the lowering of j_1 constitutes the drag from this source. But a reduction in j_1 also happens when the annulus exerts a torque on its neighbor, raising j_2 to the higher value $j_2 + \Delta j$. The radius of the outer annulus consequently increases while the inner annulus is shrinking. Thus, after the drag from impacting matter has diminished, further inward flow of mass is generally accompanied by *outward* flow of angular momentum and concurrent spreading of the disk.

The simplest way to achieve this internal torquing would be through some kind of friction. Referring again to Figure 11.14, we note that the inner annulus has the higher rotational speed, even though it has the lower specific angular momentum. Thus, the presence of any *shear viscosity* would indeed cause this annulus to torque up its neighbor, just as required. The idea is elegant in its simplicity, and quantitative models confirm that inward mass flow is readily attainable, with the rate proportional to the assumed viscosity. What, then, is the physical source of this internal friction?

Even after decades of intensive research, there is still no generally accepted answer to this question, either in star formation or in other studies where disks figure prominently, such as the accretion in galactic nuclei. For our present purposes, two lessons may be drawn from this cumulative experience. The first is that disks in isolation have *no* clearly demonstrated tendency either to spread or transfer matter to their central bodies. Note the operative phrase *in isolation*. Disks in binary systems, for example, evidently *do* facilitate the mass transfer from one star to its companion. The second lesson is that the description of internal torquing as an effective viscosity may be inappropriate. In particular, there is no compelling reason why the torque should be proportional to the local gradient in azimuthal velocity.[4] Of course, a protostellar disk is *not* isolated, but gains matter continually from its parent cloud. It is thus sensible to explore, without being burdened by the viscous analogy, how mass accumulation itself might promote internal torquing.

11.3.5 Gravitational Instability

Let us return to our evolutionary picture of the growing protostellar disk. We first make the technical point that any substantial rise in M_d invalidates one of our previous, underlying assumptions. We have taken the gravitational force on a fluid element to be that from the protostar alone, and it is clear that this approximation breaks down. In part, the correction is simple enough. For any distribution of surface density in the inner disk, we may integrate over all annuli to obtain the extra force at a fixed radius. The deeper issue, however, is dynamical. Once the disk's own mass provides significant binding, the possibility arises that the entire structure becomes *gravitationally unstable.* Indeed, this occurrence is inevitable, given our working hypothesis of a virtually unlimited angular momentum and mass supply from the cloud. The instability must profoundly affect subsequent evolution and certainly impacts the mass transfer issue. However, the current understanding of disks during this phase is far from complete.

Before delving into the main issues, let us examine more carefully how the instability arises in the first place. It is helpful to recall our analysis of magnetized equilibria in Chapter 9 and, in particular, the discussion of highly flattened configurations in § 9.4. Consider, as before, a small disk-like element of surface density $\Sigma_\circ$ and radius $\varpi_\circ$, embedded within the much larger protostellar disk (see Figure 11.15). If we squeeze the small element so that its radius shrinks fractionally by ϵ, the additional gravitational force per unit mass toward its center is again given by $F_G \approx \epsilon G \Sigma_\circ$. Conservation of angular momentum (rather than of magnetic flux) now dictates that the angular rotation speed increase fractionally by about ϵ, from its initial value $\Omega_\circ$. It follows that the extra centrifugal force is $F_R \approx \epsilon \varpi_\circ \Omega_\circ^2$. Finally, the increase in thermal pressure, integrated over the disk height, is about $\epsilon \Sigma_\circ a_s^2$, where a_s is the internal sound speed,

---

[4] Nevertheless, the literature is still replete with "α-disk" models. Here, α is a nondimensional parameter that measures the strength of the viscosity. The traditional practice has been to assign this parameter *ad hoc* and follow the resulting evolution of the disk.

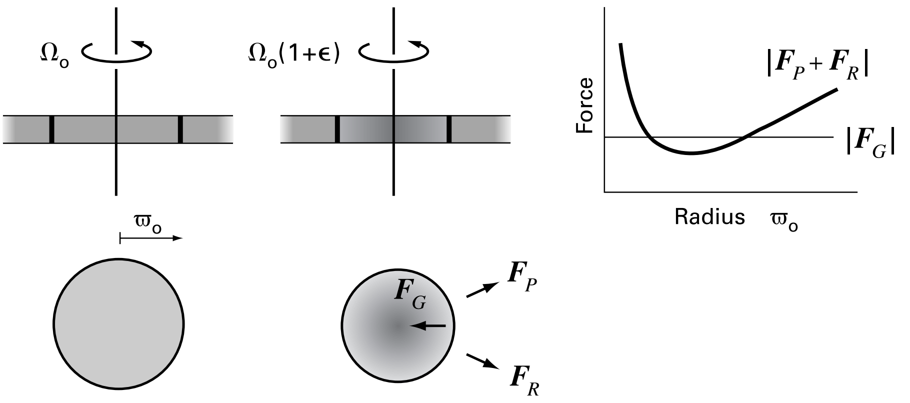


Figure 11.15 Gravitational stability of a rotating disk. A small, interior element of initial radius $\varpi_\circ$ contracts and spins up, thereby inducing extra gravitational, pressure, and rotational forces. In the case shown, the gravitational force exceeds the sum of the pressure and rotational forces over a finite range of $\varpi_\circ$-values. The parent disk is therefore unstable.

calculated assuming that the squeezing occurs adiabatically. Thus, the rise in the associated outward force is $F_P \approx \epsilon a_s^2/\varpi_\circ$.

The total repulsive force induced by squeezing the elemental disk is thus

$$F_R + F_P \approx \epsilon\,\varpi_\circ\,\Omega_\circ^2 + \frac{\epsilon\,a_s^2}{\varpi_\circ} ,$$

while the incremental gravitational force F_G is independent of $\varpi_\circ$. For fixed a_s and $\Omega_\circ$, $F_R + F_P$ diverges at both small and large $\varpi_\circ$. The sum therefore has a minimum value, which is $2\epsilon a_s \Omega_\circ$. If this minimum force is smaller than F_G, there exists a finite range of $\varpi_\circ$ over which the gravitational force dominates and the larger disk is unstable; this case is illustrated in Figure 11.15. Conversely, we have stability if the repulsive force is larger at all $\varpi_\circ$. This happens when $\epsilon G \Sigma_\circ < 2\epsilon a_s \Omega_\circ$, *i. e.*, when $2\Omega_\circ a_s/G\Sigma_\circ > 1$. A careful perturbational analysis of the disk yields substantially the same result. After dropping all subscripts, the more exact criterion is

$$Q \equiv \frac{\kappa\,a_s}{\pi\,G\,\Sigma} > 1 \qquad \textit{for stability} . \tag{11.41}$$

Here, Q is the *Toomre Q-parameter*. If $\Omega(\varpi)$ is the rotation speed as a function of cylindrical radius, then the quantity κ on the righthand side of (11.41) is defined through

$$\kappa^2 \equiv \frac{1}{\varpi^3}\,\frac{d(\varpi^4\,\Omega^2)}{d\varpi} , \tag{11.42}$$

and is known as the *epicyclic frequency*.[5] The reader may verify that κ approaches Ω if the latter has a radial falloff close to $\varpi^{-3/2}$, the result for Keplerian orbits.

---

[5] Physically, κ represents the frequency of small oscillations for elements displaced slightly from circular orbits. The terminology is thus a reference to Ptolemaic cosmogony. In modern astrophysics, κ plays a fundamental role in the theory of galactic structure. Toomre's criterion, equation (11.41), was first derived in this context.

This stability criterion, like the one governing convection in stars, is local, since Q is a function of radius within any particular disk. The actual variation of this parameter is not very broad in the model corresponding to Figure 11.4, since κ, Σ, and a_s all decline outward at comparable rates. Here the fall in a_s reflects the weakening of both heating agents – the protostellar flux and the surface shocks – with greater distance. In general, we expect Q at early times to be high enough everywhere that the relatively hot and small disk is unequivocally stable. Subsequent cooling and spreading eventually causes the parameter to drop below unity at some interior radius.

The disk properties at this transition epoch are still not known in detail, but may be crudely estimated. To obtain the mass, we first let $\Omega \approx V_{\rm Kep}/\varpi$, where $V_{\rm Kep}$ is again the local orbital speed. After further approximating Σ as M_d/ϖ^2, we find that

$$\frac{M_d}{M_*} \approx \frac{a_s}{V_{\rm Kep}} \, , \tag{11.43}$$

when $Q \approx 1$. But the righthand side, from equation (11.34), is a measure of the relative disk thickness, which we have seen is small. Hence, the structure begins its self-gravitating phase with a mass less than M_*; current calculations indicate a fraction of roughly 10 percent. The rapid growth of both M_d and $\varpi_{\rm cen}$ for $t > t_1$ implies that the latter cannot greatly exceed R_* when the transition occurs, certainly by no more than an order of magnitude. But the disks around pre-main-sequence stars are thought to have dimensions of about 100 AU, or $10^4\,R_*$. The outer radius must therefore spread enormously by the time the star is optically revealed. Meanwhile, the total disk mass has either climbed only slightly or else fallen.

11.3.6 Spiral Waves

Near the location where Q first drops near unity, any transient density rise is no longer damped out by pressure and rotational forces, but grows rapidly. Instead of separating from the background, the incipient clump is sheared apart by differential rotation. Within a few orbital periods, the sheared strands reorganize into a coherent *spiral wave* pattern. Such waves, whether they occur in galaxies or protostellar disks, are thus the natural products of self-gravity and rotation.

Figure 11.16 shows this remarkable process of wave formation in a particular numerical simulation. Here, the mass of the central star is 0.60 $M_\odot$, while that of the disk is 0.40 $M_\odot$. The outer edge is held rigidly fixed at 230 AU, and the boundary absorbs any fluid elements impinging on it. The initial distributions in Σ and a_s are such that Q varies between 1 and 3, with a minimum near 80 AU.[6] Note that the time unit displayed is $T \equiv 480$ yr, about half the orbital period at the disk edge. By $t = 3\,T$, a strong, two-armed spiral has developed and begins to move outward, rotating all the while at a fixed pattern speed. This wave is later absorbed at the edge, but another is regenerated deeper inside. Several cycles of instability appear over the total elapsed time of 10 T.

---

[6] The stability criterion in equation (11.41) actually pertains to axisymmetric (ring-like) perturbations. Disks become unstable to non-axisymmetric modes at somewhat higher Q-values.

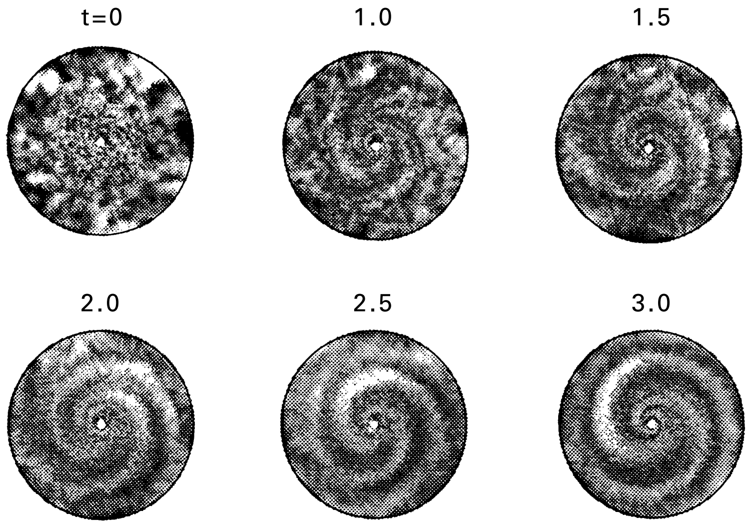


Figure 11.16 Development of a two-armed spiral wave in a protostellar disk, shown at 6 representative times. The grey scale indicates the perturbed values of surface density. Initially random perturbations quickly organize into a coherent, expanding wave pattern.

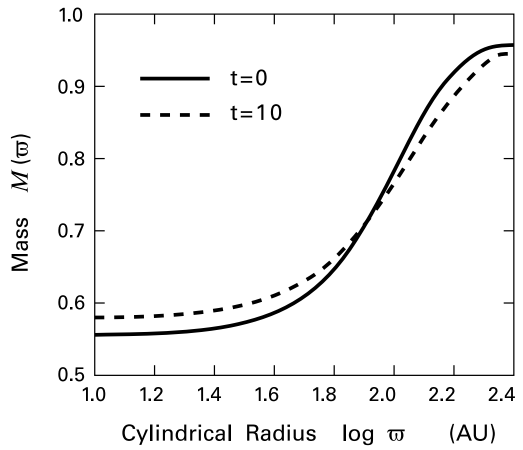


Figure 11.17 Mass transport in a protostellar disk. For the two times indicated, the mass contained within a cylindrical radius ϖ is plotted as a function of ϖ. This mass is scaled to an arbitrary, fixed value.

From our perspective, the most intriguing aspect of spiral waves is their ability to transfer angular momentum. Note that *any* non-axisymmetric density perturbation would create internal torques. The spatial distribution of these torques is especially smooth in a "grand design" spiral. In the particular example shown here, the net flow of angular momentum is outward, and interior annuli indeed migrate toward the central star. Figure 11.17 demonstrates this shift by displaying the distribution of $M(\varpi)$, the total mass interior to any radius ϖ. Notice how $M(\varpi)$ has risen toward the center by the end of the calculation.

Results such as these have bolstered the idea that "gravitational torques" can both spread the disk (perhaps well beyond $\varpi_{\rm cen}$) and simultaneously drain infalling matter onto the protostar. Certainly, these torques can be much stronger than those created by infall. However, we are still far from a complete, or even self-consistent, evolutionary account. The simulation shown in Figure 11.15 neglects infall entirely and *assumes* a disk radius of the same magnitude as those to be explained. Can gravitational torques actually spread a much smaller initial disk by the requisite amount? Specifically, does the spiral wave lead to a sustained, *outward* flow of angular momentum? Does the wave truly dissipate and regenerate itself periodically, or is this cycling an artifact of the chosen background state and conditions at the edge? Note that disks with much lower Q-values grow such strong spiral waves that they break apart into self-gravitating pieces. What physical effect guarantees that Q will actually hover close to unity, as implicitly assumed here?

From a practical viewpoint, one impediment to answering these questions is that numerical simulations can track only a modest number of orbital periods. Recall that the period around a 0.3 $M_\odot$ star is only 21 days at a radius of 0.1 AU; these should be representative values at the start of the self-gravitating epoch. Protostellar infall, however, lasts some 10^5 yr. Such a vast discrepancy between the *dynamical* and *secular* time scales appears to doom any direct simulation of the full evolution. The alternative would be an approximation scheme that averages over many orbits, concentrating instead on the relatively slow changes in the spiral-wave amplitude.

Finally, we return to the protostar itself. Even if gravitational torques do prove capable of redistributing disk mass over a large area, we must still understand how this material joins onto the stellar surface. Rapid spinup of the gas demands a highly efficient braking mechanism and energy sink to allow continued accretion. It is here that twisting of the stellar magnetic field and the generation of MHD winds probably enter the picture (Chapter 13). We will also see in Chapter 17 how the innermost disk, coupled to the stellar field, is susceptible to clumping and turbulent motion that may aid in the accretion process.

We may ask, finally, how a central object accreting partially from a disk differs from the spherical models considered earlier. The answer is not known in detail, but protostellar evolution is surely more sensitive to the *rate* of mass addition than to the detailed spatial pattern of the infall. Chapter 17 will present evidence that the disks around older, pre-main-sequence stars have an inner edge at several stellar radii. Disk matter could traverse this gap by flowing inward along magnetic flux tubes. If protostars also receive mass this way, the total $\dot{M}_d$ is apportioned into discrete "hot spots" on the surface. In this case, the radiated accretion *flux* has strong, localized peaks, but the surface-integrated *luminosity* $L_{\rm acc}$ remains the same as in a spherical model. So does, at least approximately, the specific entropy just beneath the surface. This entropy rises, as before, with increasing stellar mass, and our evolutionary picture should remain largely intact.

11.4 More Massive Protostars

The accretion rates used to build up low-mass protostars are those appropriate to parent clouds that are marginally stable prior to collapse and largely thermally supported, with internal temperatures of roughly 10 to 20 K. We have identified these structures with the well-studied dense cores. As we seek to extend the theory to higher masses, the observational situation grows murkier. Do these stars arise from the collapse of similar clouds, or rather from massive analogues to dense cores? Radio surveys do find systematically larger and warmer molecular fragments in regions such as Cepheus and Orion, already known for their production of high-mass stars. Some of these cloud entities contain entire infrared clusters while others appear to be starless, but all tend to have relatively large nonthermal molecular line widths. Those in the first category contain a larger fraction of high-density gas ($n \gtrsim 10^5\,\mathrm{cm}^{-3}$), but some of this compression may stem from shocks induced by stellar winds. Furthermore, it is unclear to what extent the higher velocity dispersion simply reflects a blending of substructures in relative motion that have not been spatially resolved. In the case of the Cepheus star formation region, its distance of 730 pc does severely limit resolution. The massive cores found in Orion A and B cannot be so easily dismissed. Whatever substructuring exists there should be revealed in the near future.

For now, our theoretical deliberations on massive protostars must retain this basic uncertainty. We will tentatively assume that the picture of collapse we have developed, including the range of $\dot{M}$-values, is applicable to *all* stellar masses, and let the observations indicate when this hypothesis fails. Here and in Chapter 18, we will see that the resulting models are largely successful at explaining intermediate-mass objects, *i. e.*, those from about 2 to 10 $M_\odot$. We will also see that the admittedly idealized spherical collapse is truly untenable for O stars, because of their extremely large luminosities. Finally, Chapter 12 revisits the topic of clusters and explores their possible role in stellar formation at the highest masses.

11.4.1 Return to Radiative Stability

We continue, then, to construct evolutionary models by solving the four stellar structure equations, regarding $\dot{M}(t)$ as a freely specifiable function. Provided we limit ourselves to the same range of accretion rates as before, we find that M_* cannot increase far above 1 $M_\odot$ before an important change occurs. The protostar's interior, which was fully convective from deuterium burning, now reverts to a radiatively stable state. It does so because of a decline in the average opacity, which makes it easier for the interior luminosity to reach the surface. If the protostar continues gaining mass, it maintains radiative stability until it eventually begins to fuse ordinary hydrogen.

To see quantitatively why convection must cease, we turn again to the luminosity L_{crit} in equation (11.29). This quantity varies spatially within any given star, but it is useful to gauge how its overall magnitude scales with stellar mass and radius. Thus, we replace M_r in the numerator of (11.29) by the full mass M_* and note, from (11.2a), that the average interior temperature should be proportional to $M_* R_*^{-1}$. In the same spirit, we set $(\partial T/\partial P)_s$ equal to T/P. The pressure itself should scale as $M_*^2\, R_*^{-4}$, according to equation (11.15). Finally, the opacity κ is proportional to $\rho T^{-7/2}$ in the relevant density and temperature regime (see Appendix G). After replacing ρ by $M_* R_*^{-3}$, we combine these results to find that L_{crit} scales as $M_*^{11/2} R_*^{-1/2}$. The steep mass-dependence largely reflects the sensitivity to temperature of the Kramers-law opacity.

The lesson here is that the average L_{crit} increases so sharply during protostellar accretion that it eventually surpasses the actual interior luminosity. It is at this point that convection

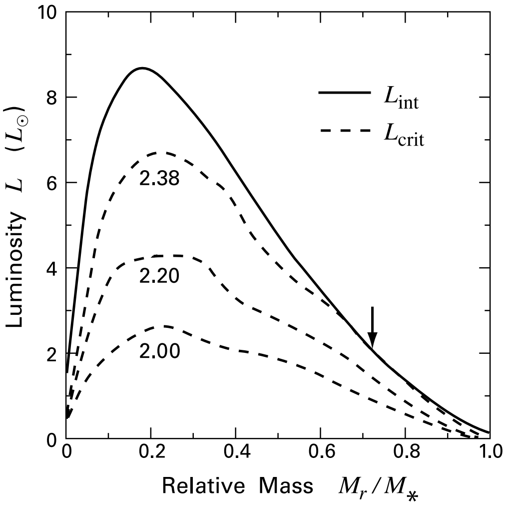


Figure 11.18 End of convection in protostars. The profiles of critical luminosity (*dashed curves*) gradually approach the true interior luminosity (*solid curve*). Numbers associated with each critical luminosity profile are the protostellar mass in solar units. The independent variable is the interior mass relative to the total protostellar value. The short, vertical arrow indicates where the radiative barrier first appears.

disappears. The true luminosity stems mostly from deuterium fusion, and its peak value is proportional to $\dot{M}$ (recall equation (11.30)). We see, then, that the value of the transition mass is *insensitive* to the accretion rate. For the $\dot{M}$-values of interest, this mass falls near 2 $M_\odot$.

Figure 11.18 shows in more detail how the alteration occurs. Here, the numerical calculation employs a constant accretion rate of $1 \times 10^{-5}\, M_\odot\, {\rm yr}^{-1}$. The dashed curves, representing $L_{\rm crit}(M_r)$, rise quickly as M_* advances through the values indicated. Meanwhile, the true interior luminosity (*solid curve*) has barely changed. In this particular sequence, $L_{\rm crit}(M_r)$ intersects $L_{\rm int}(M_r)$ when M_* reaches 2.38 $M_\odot$; the two curves actually touch at $M_r = 1.70\, M_\odot$. Just at this mass shell, indicated by the vertical arrow in the figure, $L_{\rm int}$ drops below $L_{\rm crit}$, and Schwarzschild's criterion is satisfied. A *radiative barrier* has now appeared and will quickly alter the protostar's thermal structure.

11.4.2 Deuterium Shell Burning

The stable region near 1.70 $M_\odot$ in Figure 11.18 constitutes a barrier because it prevents freshly accreted deuterium from reaching the center through the turbulent transport associated with convection. We recall that the convection itself is driven by nuclear fusion. Once the barrier is established, the residual deuterium inside it is consumed rapidly, and convection disappears throughout the interior volume. Referring again to Figure 11.18, the peak value of $L_{\rm int}(M_r)$ now declines until the luminosity profile falls below $L_{\rm crit}(M_r)$ everywhere inside 1.70 $M_\odot$.

As long as cloud collapse and infall persist, deuterium accumulates in a thick mantle outside the original radiative barrier. The central temperature of the protostar has been slowly but steadily rising with time. With the exhaustion of the interior nuclear fuel, the increase of R_* that began at lower masses virtually stops. Thus, the ratio M_*/R_* grows faster than before, and the temperature rise accelerates. Soon, even the base of the deuterium mantle reaches 10^6 K. The fuel ignites and induces convection out to the surface. Once again, fresh deuterium arrives

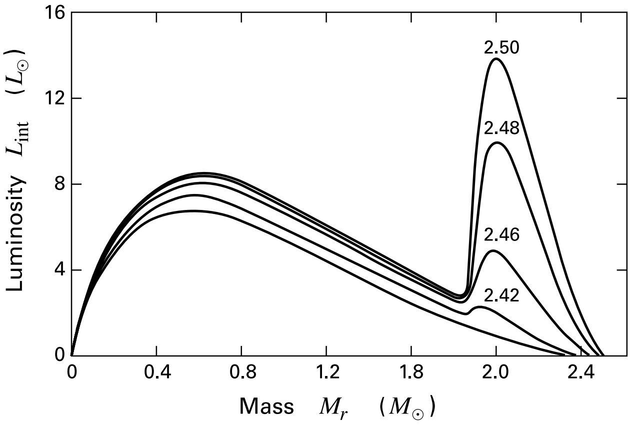


Figure 11.19 Onset of deuterium shell burning. The luminosity as a function of interior mass is shown for the indicated values of the total protostellar mass, in solar units. The lowest profile corresponds to $2.38\,M_\odot$, the protostar mass at which the radiative barrier first appears.

through infall and is quickly advected to the burning layer. Figure 11.19 shows the onset of this deuterium shell burning, as reflected in the change of interior luminosity. Here we see the rapid rise of the peak $L_{\rm int}$-value, which again approaches the steady-state level given by equation (11.30).

Figure 11.20 summarizes pictorially the stages of deuterium fusion in protostars. The shift in the site of active burning from the depleted central region to an overlying shell is reminiscent of the progression of hydrogen fusion during post-main-sequence evolution. In both cases, the establishment of the shell source is accompanied by major structural changes. Here, the injection of heat raises the specific entropy of the outer layers, and the protostar swells dramatically. This increase is apparent in Figure 11.21, which plots the protostar radius as a function of mass. The first open circle again signifies the onset of full convection, while the second marks the appearance of the radiative barrier. In this expanded view, we notice how the first rise in $R_*(M_*)$ from central deuterium burning occupies a relatively narrow mass range. Nevertheless, our knowledge of the initial mass function tells us that the majority of protostars are actually in this interval, while only a small fraction ever attain the second swelling due to shell burning.

11.4.3 Contraction and Hydrogen Ignition

If we continue adding mass to the protostar, both the convection and swelling gradually disappear. The inexorable rise of $L_{\rm crit}$ drives both the actively burning layer and its associated convection zone toward the surface. Now that the star is almost fully radiatively stable, our derived scaling relation for $L_{\rm crit}$ should apply to its actual luminosity, as well. Averaging over the stellar interior, we find

$$\langle L_{\rm int}\rangle \approx 1\,L_\odot \left(\frac{M_*}{1\,M_\odot}\right)^{11/2} \left(\frac{R_*}{1\,R_\odot}\right)^{-1/2} \quad . \tag{11.44}$$

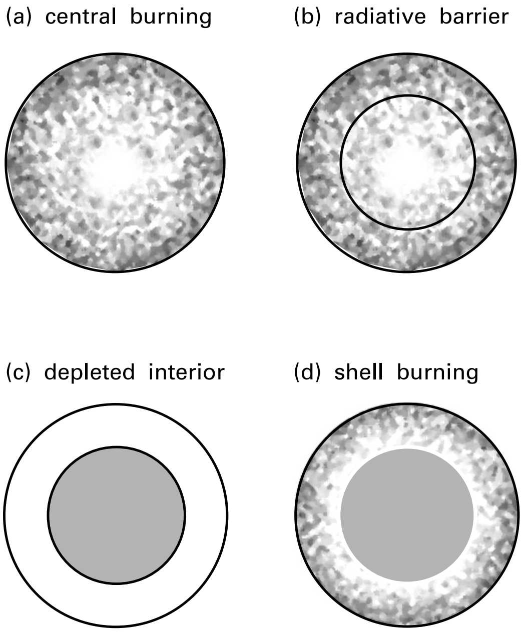


Figure 11.20 The four stages of deuterium burning in protostars. Active burning begins at the center and continues until a radiative barrier appears. The entire protostar is then radiatively stable, and its interior depleted of deuterium. The fuel later re-ignites in a thick shell, inducing convection in the outermost region.

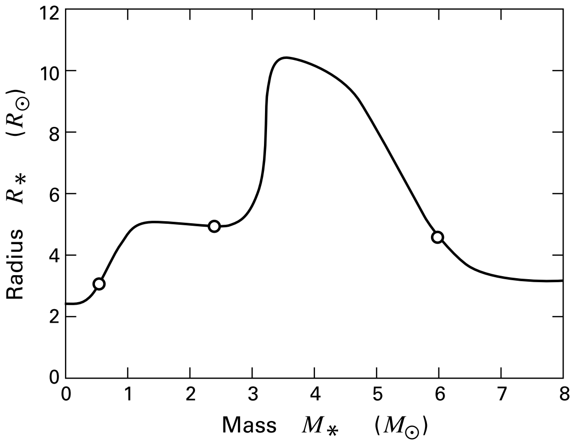


Figure 11.21 Mass-radius relation in a spherical protostar, accreting at the rate $1 \times 10^{-5}\,M_\odot$ yr^{-1}. The three open circles indicate the onset of full convection, the appearance of the radiative barrier, and the second initiation of central convection, as a result of hydrogen fusion.

We stress that this approximate (and very useful) relation is valid for *any* radiative star, regardless of its evolutionary state. Indeed, we have been able to assign the numerical coefficient from the fact that the Sun itself is stable against convection throughout most of its volume.

Equation (11.44) indicates that the protostar's interior luminosity soon outstrips that produced by steady-state shell burning. For $M_* \gtrsim 3\,M_\odot$, the interior contribution amounts to several 100 $L_\odot$ and even dominates $L_{\rm acc}$. Between 5 and 6 $M_\odot$, the luminosity surpasses $10^3\,L_\odot$. What accounts for this remarkable climb? The answer is gravitational contraction of the bulk interior. This energy source has been available from the earliest times, but does not truly come into play until after the cessation of central deuterium burning. From then on, the influence of self-gravity keeps building until it becomes paramount. Figure 11.21 shows how even the large swelling due to shell ignition is soon reversed, and the star begins a rapid, overall contraction.

Consider the observational and evolutionary status of the star during this period. Our assumption of continuing infall means that the brightening object is obscured at optical wavelengths by its dusty envelope. The same holds, of course, for its low-mass counterparts. However, the fact that most of the luminosity now stems from internal gravitational contraction violates the definition of "protostar" offered earlier. On the other hand, the star is certainly not in the pre-main-sequence phase, where any mass addition occurs at a much slower rate than the stellar contraction. We will continue to refer to the objects of interest as "intermediate-mass protostars," while recognizing their rather unique, hybrid nature.

It is important to realize that the accreting star has *not* entered a state of dynamical collapse like that which terminates the first core. Velocities interior to the accretion shock remain subsonic, as gravity slowly squeezes the configuration against thermal pressure. The time scale for R_* to decrease significantly is therefore much longer than $t_{\rm ff}$, and is set by the magnitude of the radiative losses. (Recall the definition of $t_{\rm KH}$ in Chapter 1.) We shall examine more carefully such quasi-static contraction in Chapter 16, when we study pre-main-sequence evolution. For the present, we simply note two generic features of the process. First, the internal luminosity, whose average value is given roughly by equation (11.44), now increases outward monotonically. The local maxima of $L_{\rm int}(M_r)$ seen in Figures 11.9 and 11.19 have vanished, and the specific entropy of the deep interior falls with time. Secondly, the decline of R_* implies an even faster rise of the interior temperature. Once the central value T_c surpasses 1×10^7 K, the protostar begins to fuse ordinary hydrogen.

The creation of ^{4}He from four hydrogen nuclei releases sufficient energy (ΔE = 26.7 MeV) to halt the stellar contraction. At this point, the reader may well ask why this reaction only begins at 10^7 K, while the prior fusion of deuterons and protons occurred at a much lower temperature. The answer is that the creation of ^{4}He requires two of the four protons to become neutrons. This transformation involves the weak nuclear interaction. Consequently, the particle kinetic energies must be higher for the fusion to proceed at significant rates.

Hydrogen burning in protostars commences when the total mass reaches about 5 $M_\odot$. Initially, pairs of protons begin to combine:

$$^1{\rm H} + {^1{\rm H}} \rightarrow {^2{\rm H}} + {\rm e}^+ + \nu \; . \qquad (11.45)$$

The emitted positron annihilates with an ambient electron, while the neutrino escapes the star. The deuterium produced here almost immediately fuses with another proton to create ^{3}He, through the much faster reaction of equation (11.26). The subsequent conversion of ^{3}He to ^{4}He proceeds along a number of different pathways, known collectively as the *PP chains* (see Chapter 16). All of these reactions operate simultaneously. However, the temperature within our contracting protostar becomes so elevated that a different mode of fusion soon takes over. Heavier nuclei begin adding successive protons until they eventually decay, spitting out α particles (^{4}He nuclei). This process is known as the *CNO bi-cycle*. As we shall later see, it is the dominant reaction network in early-type main-sequence stars.

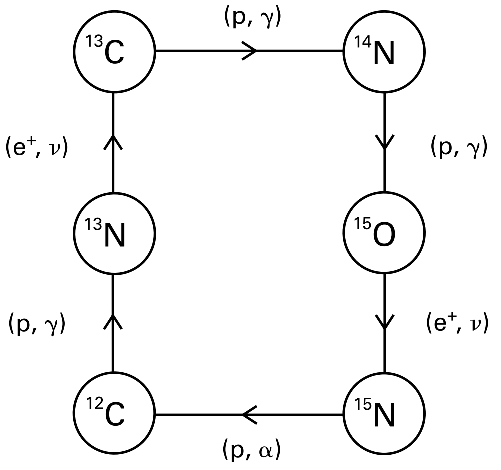


Figure 11.22 The CN cycle of hydrogen fusion. The reactions shown correspond to those in equation (11.46) of the text. Note, however, that protons are now designated as p instead of ^{1}H, and that we have employed a condensed notation for incident and final particles and photons.

Within protostars just beginning their hydrogen fusion, only a portion of the entire network is active. This *CN cycle* consists of the reactions

$$\begin{aligned} {}^{13}\mathrm{C} + {}^{1}\mathrm{H} &\rightarrow {}^{14}\mathrm{N} + \gamma \\ {}^{14}\mathrm{N} + {}^{1}\mathrm{H} &\rightarrow {}^{15}\mathrm{O} + \gamma \\ {}^{15}\mathrm{O} &\rightarrow {}^{15}\mathrm{N} + \mathrm{e}^{+} + \nu \\ {}^{15}\mathrm{N} + {}^{1}\mathrm{H} &\rightarrow {}^{12}\mathrm{C} + \alpha \\ {}^{12}\mathrm{C} + {}^{1}\mathrm{H} &\rightarrow {}^{13}\mathrm{N} + \gamma \\ {}^{13}\mathrm{N} &\rightarrow {}^{13}\mathrm{C} + \mathrm{e}^{+} + \nu \, , \end{aligned} \tag{11.46}$$

and is shown schematically as Figure 11.22. As the figure demonstrates, the sequence of reactions indeed forms a closed cycle. The incident ^{13}C (or any of the other heavy nuclei involved) acts merely as a catalyst and is regenerated at the end of each loop. Along the way, four protons are consumed to produce one ^{4}He nucleus. Thus, the net reaction is

$$4\,{}^{1}\mathrm{H} \rightarrow {}^{4}\mathrm{He} + 2\,\mathrm{e}^{+} + 2\,\nu \, , \tag{11.47}$$

for which the energy output is the same 26.7 MeV as in the PP chains.

The CN cycle predominates once the protostar's mass is about 6 $M_\odot$ and its central temperature has reached 2×10^7 K. It is clear from Figure 11.21 that the contraction begins to slow at just this point. While equation (11.44) still governs the *total* interior luminosity, the contribution from nuclear burning grows at the expense of that from gravitational contraction. The specific entropy near the center begins to rise, just as it did after the first deuterium ignition. Soon, the entropy profile overturns and a central convection zone appears; this event is marked by the third open circle in Figure 11.21. The zone extends to $M_r = 1.3\,M_\odot$ by the time contraction truly halts, at a protostellar mass near 8 $M_\odot$. The object is now more accurately described as an accreting main-sequence star. Infall makes only a minor contribution to the total luminosity of $3.5 \times 10^3\,L_\odot$, and continues to provide an obscuring screen for the emitted optical and ultraviolet photons.

The numerical results we have quoted so far were obtained using a constant accretion rate of $1 \times 10^{-5}\, M_\odot$ yr^{-1}. It is instructive to see the consequences of varying this rate. Suppose, for example, that we lower $\dot{M}$ by a factor of 10. The accretion luminosity bathing the star from the outside is now diminished, and there is more time for internally generated heat to escape. At any stellar mass, therefore, the radius is smaller than before. Both the ratio M_*/R_* and the internal temperature are consequently *higher*. Thus, the protostar begins main-sequence hydrogen fusion at a mass near 4 $M_\odot$. This consequence of the lower rate rules it out empirically, at least as a universal prescription. The well-studied Herbig Ae/Be stars often have masses that exceed 4 $M_\odot$. These objects must have originated from protostars that had not yet contracted and ignited hydrogen (Chapter 18).

Conversely, raising $\dot{M}$ to $1 \times 10^{-4}\, M_\odot$ yr^{-1} results in an evolutionary sequence with larger radii and lower internal temperatures at every mass. In particular, entry to the main sequence is now delayed to 15 $M_\odot$. This case is harder to dismiss out of hand but is probably also unrealistic. Equation (10.31) tells us that the enhanced $\dot{M}$ corresponds to a pre-collapse cloud temperature over 100 K. Such temperatures are far in excess of those encountered in isolated dense cores, but *are* routinely found in cloud fragments exposed to nearby (or embedded) O and B stars. There is no evidence, however, that Ae/Be stars are found only in regions that earlier produced even more massive objects. In summary, our adopted order of magnitude for the accretion rate is plausible, although a range in values and a temporal variation are both naturally expected.

11.4.4 Effect of Radiation Pressure

A far more problematic issue is the mechanical effect of the protostar's radiation on its infalling envelope. Heating of the gas and dust creates a retarding pressure, one that we have so far been able to ignore. As the star gains mass and its luminosity climbs swiftly, we can no longer afford this simplification. The new pressure source becomes so strong, in fact, that it must alter the pattern of collapse. Finding the new pattern is difficult (and has not yet been done), but we may at least hazard a guess as to the direction of this change.

First we should be more precise about the manner in which radiation impedes the flow. We have seen that the temperature in the dust envelope climbs steeply toward the protostar. The grains here emit copiously in the infrared, but these photons are soon absorbed. A grain that takes up the photon momentum quickly transfers it by collision with a gas-phase atom or molecule. The radiation and gas thus essentially constitute two components of a single fluid, each exerting its own partial pressure. The photon contribution is 1/3 of the associated energy density, as it would be for any gas of relativistic particles. Utilizing the formula for $u_{\rm rad}$ from equation (2.37), we find the radiation pressure to be

$$P_{\rm rad} = \frac{1}{3}\, a\, T^4 \ . \tag{11.48}$$

Both $P_{\rm rad}$ and the ordinary gas pressure rise with increasing temperature, but the former dominates close to the dust destruction front.

Suppose, as a first approximation, that we continue to ignore the gas pressure. Then the retarding force per unit volume is $-\partial P_{\rm rad}/\partial r$. The transfer of radiation within the dust envelope is governed by the diffusion equation. Changing independent variables in equation (G.7) from z to r, we see that the radiative force is also $\rho\kappa F_{\rm rad}/c$. This result is intuitively appealing, since the opacity κ represents the effective cross section of the matter for the stream of photons, each of which carries momentum $h\nu/c$. Retardation of the infall begins once $\rho\kappa F_{\rm rad}/c$ becomes comparable to the gravitational force per unit volume. Since the latter is $-\rho G M_*/r^2$, and since $F_{\rm rad}$ may be written as $L_*/4\pi r^2$, the condition becomes

$$\begin{aligned} \frac{L_*}{M_*} &= \frac{4\,\pi\,c\,G}{\kappa} \\ &\approx 1\times 10^3\,\frac{L_\odot}{M_\odot}\ . \end{aligned} \tag{11.49}$$

Evaluation of the righthand side of this equation requires an opacity, which in turn depends on the characteristic radiation temperature. Noting that the retarding force is highest just outside the destruction front, we have used the grain sublimation temperature of about 1500 K. Here, $\kappa \approx 10\ {\rm cm^2\ g^{-1}}$, as implied by the standard interstellar extinction curve at near-infrared wavelengths. Plausible variations in grain composition could alter this value by as much as a factor of three in either direction.

Even with this uncertainty, the numerical form of equation (11.49) confirms that radiation pressure is *not* significant for low-mass protostars. Indeed, the relevant value of L_*/M_* is high enough that the star must be one undergoing main-sequence hydrogen burning. Reference to Table 1.1 shows that the critical ratio occurs at a mass near 11 $M_\odot$. Thus, accretion should be largely unimpeded for most intermediate-mass protostars, but is significantly affected for central objects of greater mass.

Numerical simulations of collapsing spheres corroborate this finding. Here, one determines the temperature distribution and radiative intensity within the envelope self-consistently, through the iterative procedure sketched in § 11.1. Because this task is computationally intensive, one is forced to simplify the hydrodynamics through the assumption of spherical symmetry. The calculations also idealize the protostar itself to be a point source of luminosity. At early times, the luminosity stems mostly from accretion. During this epoch, infall cannot truly stop, even after $P_{\rm rad}$ starts to become dynamically significant. The reason is that any deceleration of the flow diminishes both $\dot{M}$ and $L_{\rm acc}$, and thus leads to a fall of $P_{\rm rad}$ itself. This lowering of the retarding force allows both $\dot{M}$ and $L_{\rm acc}$ to increase once more. In practice, one finds that $\dot{M}(t)$ displays oscillatory behavior, without actually reversing. However, once L_* climbs significantly above $L_{\rm acc}$, the feedback between $\dot{M}$ and $P_{\rm rad}$ is cut. Only an inner fraction of the cloud then collapses onto the star, while the rest disperses to large distances. The driving force behind this motion stems in part from the trapped infrared radiation. Also significant is the direct impact of stellar ultraviolet photons on the grains just before they vaporize.

These calculations provide dramatic evidence for the potentially destructive influence of radiation pressure. On the other hand, stars more massive than 11 $M_\odot$ certainly exist. How is this possible? There is no doubt that the rapidly increasing stellar luminosity carries enough energy to reverse infall and unbind the parent cloud fragment. Since L_* exceeds $L_{\rm acc}$, this dispersal could occur over a period less than the accretion time $M_*/\dot{M}$. The real issue is how efficiently the radiation transfers its energy to the envelope. Here, any departure from the idealized case of spherical collapse is likely to play a crucial role.

To see why, recall the asymmetry induced by cloud rotation. We have seen that infalling elements starting far off the central axis (at large $\theta_\circ$) follow highly nonradial trajectories. These elements may strike the equatorial plane at such great distances that their grains are never heated to the sublimation temperature. Once entering the disk, the fluid would be further shielded from radiation pressure by the high opacity along the midplane. The disk efficiently absorbs stellar photons and reradiates them from its faces. Internal torquing then promotes accretion toward the star. In this manner, a *portion* of the total infall might continue relatively unperturbed, even for high values of L_*. [7] The basic requirement is that the disk radius extend past the dust destruction front in the equatorial plane. This condition should not be difficult to meet, as Figure 11.4 illustrates.

What of the remainder of the infall? It is inevitable that some fluid elements closer to the pole will bear the full brunt of the elevated radiation pressure before their grains vaporize. In spherical geometry, this gas must brake and start to reverse its velocity. In the two-dimensional rotating collapse, however, the fluid can react to increased pressure by veering off laterally. Moreover, any strong deceleration of the highly supersonic infall creates a standing shock front. This would effectively replace the accretion shock in lower-mass protostars, but would be spatially removed from the star and disk, not far beyond the dust destruction front. We may picture incoming gas entering the shock obliquely, then turning toward the equator within a condensed shell of postshock matter (Figure 11.23).

Although this picture has appealing features, it is unlikely to provide a full solution to the problem of high-mass protostars. A shell supported against gravity by radiation pressure may well be dynamically unstable. Any small, incipient warping will tend to increase, until discrete portions break off and fall toward the star, while others are promptly blown outward. In addition, massive stars emit powerful winds from their surfaces. The associated ram pressure should be effective at clearing the infalling envelope, especially in the polar region.

Our discussion touches on another fundamental and longstanding question. Is there a theoretical limit to the mass of a star? Empirically, the upper bound is roughly estimated at 100 $M_\odot$. (The most reliable masses are derived spectroscopically, as we will describe in Chapter 12.) Since the factors delimiting *all* stellar masses are poorly understood, it is hardly surprising that the physical basis for this observation is even more uncertain. It may well be that the star formation process itself cannot yield objects beyond a certain mass. Equation (11.49) indicates that radiation pressure alone halts spherical collapse even at the intermediate-mass stage. Consideration of nonspherical infall may help the situation. However, for stars of the very highest masses, the combined effects of radiation and winds are so severe that the picture of collapse from a quiescent initial state may not apply at all. We will need to understand better any alternate production mode before we can assess its inherent limitations.

11.5 The Observational Search

Our theoretical account of protostar structure is largely based on physical processes that are well established from other areas of stellar evolution. We have accordingly been able to explore

---

[7] Note, however, that an *exposed* O star emits such strong ultraviolet radiation that it quickly evaporates its own disk; see Chapter 15.

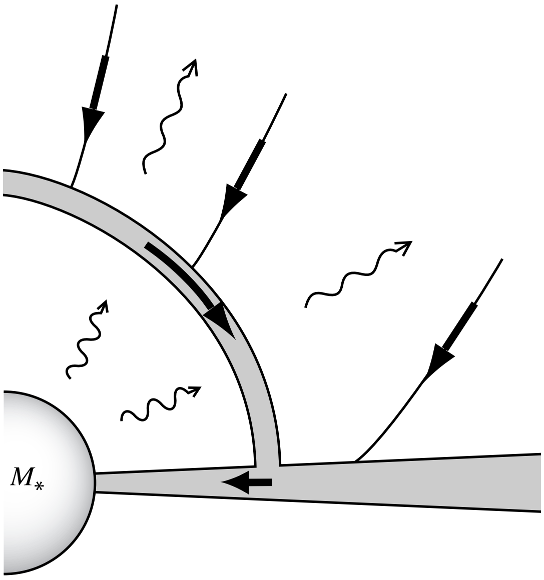


Figure 11.23 Possible deflection of rotating infall by radiation pressure (*schematic*). While some streamlines strike the protostellar disk directly, others first enter a shell, before veering off into the disk. Visible and ultraviolet radiation from the star is degraded to longer wavelengths after traversing the shell. The latter is probably dynamically unstable and will not maintain the coherent structure pictured in this idealized sketch.

several issues, such as the onset of convection and nuclear fusion, in some detail. While the picture is by no means complete, our *empirical* knowledge of protostars unfortunately lags far behind. The problem, of course, is that these objects are still buried in the dense cores that spawn them. The surrounding dust absorbs virtually all radiation emanating from the accretion shock and stellar interior. By the time this energy is reemitted at the dust photosphere, the resultant spectrum largely reflects properties of the grains themselves. Under these circumstances, it has proved challenging to verify even the presence of infall itself, much less the structural features of the underlying star or its disk. On the other hand, the detection of protostars remains a high priority, and several strategies have emerged. We describe here the basic ideas underlying the most widely employed search techniques, along with a sampling of results. Other topics of relevance, such as X-ray emission and near-infrared spectroscopy of deeply embedded stars, will be covered in subsequent chapters.

The most direct means of finding protostars is to compare the emitted fluxes from known sources with the results one expects in an infalling environment. One must first construct, in the process, a detailed, quantitative model that reproduces the data of interest. A major difficulty is that such models are rarely unique; alternative explanations may not involve protostellar collapse at all. Additionally, it is still not entirely clear which sources are the best candidates for investigation.

11.5.1 Nature of Class I Objects

We recall from Chapter 4 that young stars may conveniently be classified according to their broadband spectral energy distributions. Those in Class II and III have significant optical emission and are associated with classical and weak-lined T Tauri stars, respectively. Their visual spectra are modifications, but not radical departures, from those of main-sequence stars. Specifically, they contain the absorption lines that allow accurate determination of effective temperatures. For these sources, the radiation at infrared and millimeter wavelengths can legitimately be viewed as an “excess” with respect to main-sequence objects of comparable spectral type. In Chapter 17, we will describe the efforts to model this excess through appropriate distributions of circumstellar material.

The situation is quite different for the remaining, more embedded sources. In these, emission at longer wavelengths dominates the energy output, while spectral classification through optical absorption lines is rarely possible. The majority of the objects therefore cannot be placed in a conventional H-R diagram. Most common are those in Class I. Here we recall that the product λF_λ is still rising for wavelengths greater than 2.2 μm (Figure 4.3). These sources are invariably located near the centers of dense cores, but are they actually protostars?

We have seen that the infrared emission of a protostar stems from its dust photosphere. According to our analysis in § 11.1, this radiating surface is located roughly 10 AU from the star, with an effective temperature, $T_{\rm phot}$, of about 300 K. If the spectrum were a perfect blackbody characterized by $T_{\rm phot}$, λF_λ would peak near 10 μm. However, the emitting regions of the envelope actually include a broad range of temperatures. That portion with values lower than $T_{\rm phot}$ occupies a larger volume and is more heavily represented. The net result is that λF_λ is predicted to reach its maximum beyond 10 μm, but should be declining for $\lambda \gtrsim 100\,\mu$m.

This theoretical expectation is in general agreement with the observations of Class I sources. Consequently, a number of investigators have modeled their spectral energy distributions as arising from collapsing clouds, assumed to be either spherical or rotating. By adjusting such parameters as $\dot{M}$ and the cloud rotation rate $\Omega_\circ$, it is often possible to find a reasonable match to the spectrum *shape*, over all infrared wavelengths. Unfortunately, the *bolometric luminosity* predicted from the dynamical model, *i. e.*, the integral of F_λ over λ, is usually far too high, sometimes by more than a factor of ten.

Let us explore the discrepancy further. We first recall from equation (11.24) that L_* for an object accreting from a protostellar envelope is greater than that for a non-accreting object of the same mass. A typical Class I source should therefore be more luminous than one of Class II or III, *if* the former is indeed a protostar. In principle, one should be able to test this hypothesis from the data in young clusters, but the total number of embedded sources with measured bolometric luminosities is still small. The two best-studied cases are Taurus-Auriga and ρ Ophiuchi. In the first, observers have by now detected almost all stars with $L_{\rm bol} > 0.1\ L_\odot$. There is no tendency for the Class I sources to be brighter. Instead, the bolometric luminosity function for this group is quite similar to Class II (recall Figure 4.13a). Also similar is the typical luminosity of an individual source at millimeter wavelengths, which reflects the amount of dust warmed by the star. The implication is that Class I again represents pre-main-sequence objects, but those too heavily extincted to have optical spectra.

In the case of ρ Ophiuchi, the Class I population is indeed brighter than Class II (Figure 4.6). On the other hand, the average Class III luminosity is higher still. How are we to understand these facts? The first point to remember is that this system contains significantly more stars than Taurus-Auriga. The relatively small fraction of ρ Ophiuchi sources with measured $L_{\rm bol}$ represents the bright tail of a much broader distribution. Such higher-mass cluster members are simply absent from Taurus-Auriga.

A second distinguishing characteristic of ρ Ophiuchi is its greater column density in molecular gas. Toward the center of the L1688 cloud (Figure 3.17), A_V has been roughly estimated at 60 mag. Correspondingly, the extinction is about 6 mag at K (2.2 µm) and still exceeds unity at a wavelength of 7 µm. Under these conditions, one must carefully model the cloud environment in order to deredden the observed spectra and disentangle the influence of *interstellar* from *circumstellar* dust. What we can say at present is that the Class I objects represent a heavily obscured subset of the brighter stars. Members of even greater mass and luminosity have evolved further over the cluster age; these are now close to or on the ZAMS. Some are almost fully revealed optically and appear with Class III spectra. Further observations should find a significant fraction of lower-luminosity Class I objects, just as in Taurus-Auriga.

Some Class I sources in ρ Ophiuchi may turn out to be protostars, but there is no strong reason to believe that all are, either here or elsewhere. The data in Taurus-Auriga indicate that many of its Class I objects are not bright enough to be powered by accretion. In addition, the very fact that this category appears to be relatively common in embedded clusters makes it suspect in this regard. At an accretion rate of $1 \times 10^{-5}\ M_\odot\ {\rm yr}^{-1}$, a protostar of $1\ M_\odot$ builds up in only 10^5 yr. This same object has a pre-main-sequence lifetime of 3×10^7 yr. One would therefore expect the ratio of protostars to pre-main-sequence stars in this mass range to be quite small, on the order of 1 percent. (We will verify this conclusion through more careful reasoning in Chapter 12.) Such a figure is difficult to reconcile with the high populations of Class I sources – from 10 to 30 percent of the stars in Taurus-Auriga and ρ Ophiuchi. These fractions are bound to shift as observers continue to discover and analyze more embedded sources, but the basic message is already clear: Class I sources and protostars are not the same.

11.5.2 Modeling Spectral Energy Distributions

A more appropriate model for a Class I source of modest luminosity may thus be a pre-main-sequence star located inside a residual, dusty envelope. Beyond this, current theory provides little guidance, since we do not yet have a detailed picture of the remnant dense core that survives the main accretion phase. It is nevertheless interesting, from a strictly empirical perspective, to see the basic elements *any* model requires to match the observed infrared emission.

Figure 11.24 shows the spectral energy distribution of the IRAS source 04016+2610. This Class I object, with $L_{\rm bol} = 4\ L_\odot$, lies at the western edge of the L1489 dense core in Taurus-Auriga and is shown in the molecular line maps of the core in Figure 3.12. The IRAS satellite itself contributed the measured fluxes from 12 to 100 µm, while the data at other wavelengths come from a variety of groundbased and airborne telescopes. The average visual extinction toward the core, as obtained from ^{12}C^{18}O observations, is about 10 mag. Any field star that happens to lie behind such a cloud would surely be invisible optically and detectable only at in-

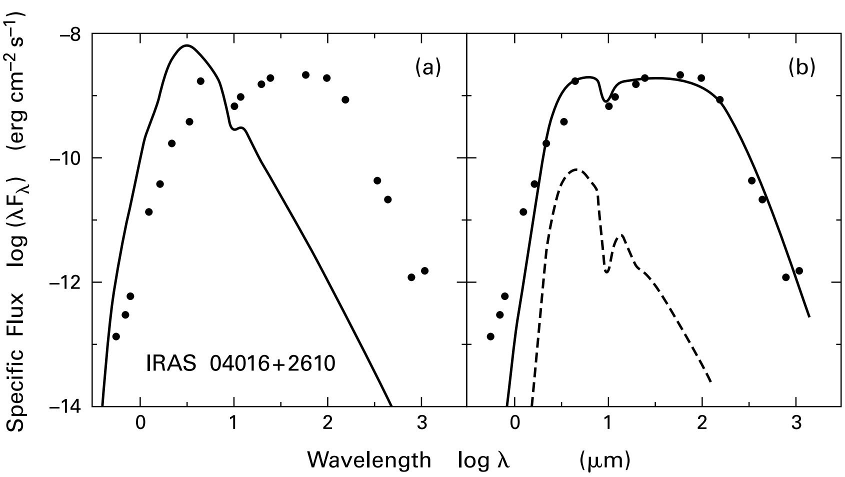


Figure 11.24 Spectral energy distribution of the Class I source IRAS 04016+2610 in L1489. (a) The solid curve shows the spectrum from an extincted, background star. (b) The spectrum from a star embedded within an optically thick dust shell. The dashed curve is the contribution from attenuated starlight alone.

frared and longer wavelengths. Could the observed source merely represent such a background star?

The answer is that it could not, and Figure 11.24a demonstrates why. By hypothesis, the dust within the foreground core is not appreciably heated by the star and thus contributes no thermal emission to the spectrum. The radiative transfer equation (2.20), with j_ν set to zero, implies that the specific intensity at each wavelength decreases exponentially through the cloud, with a total optical thickness $\Delta\tau_\lambda$ that varies with wavelength according to the interstellar extinction curve (see also equation (2.21)). Suppose that we represent the star's emitted spectrum as a perfect blackbody with $T_{\rm eff} = 4000$ K. The solid curve in the figure shows this spectrum after partial absorption by the cloud. Note that both the luminosity and distance of the star, as well as the cloud extinction (measured by A_V) are arbitrary, provided the combination gives a wavelength-integrated F_λ equal to the observed one.

It is clear that simply attenuating background starlight yields a broadband spectrum that is much too sharply peaked and narrow, and that fails in particular to give the relatively high level of emission observed at longer wavelengths. One may repeat the exercise with different stellar effective temperatures, but this basic difference remains. The breadth of Class I spectra indicates that they arise from a wide *range* of matter temperatures, rather than the single one characterizing a stellar photosphere. The most plausible origin for the long-wavelength component is heated dust grains. This dust must be located relatively close to the star, which is therefore physically embedded in the cloud.

We are thus led to consider another simplified model in which the star lies at the center of a spherical, optically thick dust shell. We now presume that the observed $L_{\rm bol}$ of 4 $L_\odot$ actually represents the star's full energy output. Transforming the original flux distribution (again assumed to stem from a blackbody with $T_{\rm eff} = 4000$ K) to infrared wavelengths requires solving the full radiative transfer equation (2.20). For a detailed study, one must add to the righthand side an extra emissive term that accounts for scattering of ambient radiation. As usual, one cannot specify j_ν without knowledge of the dust temperature, which in turn follows from energy balance in the grain component. A glance at Figure 3.12 shows that a spherically symmetric envelope can only serve as the crudest of approximations in this case. Nevertheless, if the density of the heated dust peaks close to the star, the displacement of the latter from more diffuse cloud material may not be critical.

For the spatial distribution of matter within our shell, we specify that the density vary inversely with radial distance to the star, *i. e.*, $\rho(r) = \rho_{\rm out}\,(r_{\rm out}/r)$. Here, $r_{\rm out}$ is the outer radius of the shell. This prescription for the density represents a simple compromise between the r^{-2} behavior in the outer regions of isothermal spheres and the much flatter inner profile expected in any hydrostatic body (recall Figure 9.1). Of course, a full theoretical solution would find both the density and temperature profiles self-consistently. Here, we treat $r_{\rm out}$ as a free parameter and also vary both the fiducial density $\rho_{\rm out}$ and the inner radius $r_{\rm in}$ until the resulting emitted spectrum matches the observations as closely as possible.

The solid curve in Figure 11.24b displays the best model fit, for which $r_{\rm out} = 1 \times 10^{17}$ cm, $r_{\rm in} = 2.5 \times 10^{12}$ cm, and $\rho_{\rm out} = 2 \times 10^{-19}\,{\rm g\,cm^{-3}}$ $(n_{\rm tot} = 5 \times 10^4\,{\rm cm^{-3}})$. With this choice of parameters, the dust temperature rises from 15 K at $r_{\rm out}$ to 1500 K at the inner border. Inclusion of heated dust as a source of emission has significantly broadened the model spectrum. Indeed, the match to the data is remarkably close at most wavelengths. The dashed curve shows the contribution of attenuated starlight alone, which now comprises only a portion of the near-infrared flux. The remainder stems from relatively hot grains and from the scattering that effectively reduces opacities at shorter wavelengths.

Since we did not derive our density profile from considerations of force balance, it should not be surprising that other models with very different characteristics fit the data equally well. However, all have certain features in common. First is the presence of heated dust grains, with a density that rises toward the star. The envelope in our model has an A_V of 30 mag along the central line of sight, far greater than the observed average extinction to the parent L1489 core. An enhanced inner density and A_V-value are necessary to block the star's direct radiation. On the other hand, if the extinction were *too* high, the resulting spectrum would be missing its near-infrared component. A second generic feature is therefore an interior region of relatively *low* optical thickness. Sublimation of the grains at high temperature clearly helps in this regard but is probably not enough, at least according to Figure 11.24b. The gas density itself may also slow its inward rise even before the dust vanishes. The reader may recall here our discussion in § 10.3 concerning the aftermath of magnetized infall. It is tempting (but entirely speculative at this point) to identify the inner "cavity" with the columns that have already collapsed onto the star, *i. e.*, with regions $\mathcal{A}$ and $\mathcal{A}'$ in Figure 10.8.

11.5.3 Dust Emission in Class 0

Whatever the nature of this interior region, it is apparently smaller for the objects designated as Class 0. In these, no flux at all is detectable for wavelengths shorter than 10 μm, despite the presence of a central star within the associated dense core. This star frequently appears

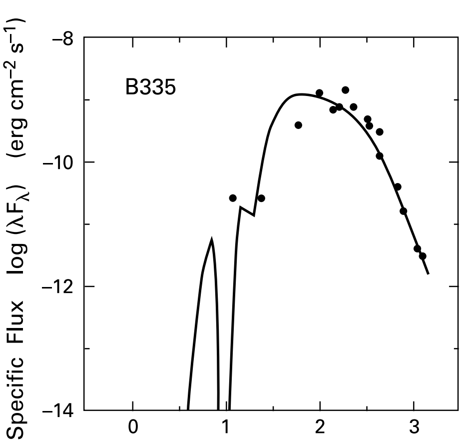


Figure 11.25 Spectral energy distribution of the Class 0 source in B335. The solid curve is the theoretical result for a star in the center of a spherical dust shell.

as a bright, spatially unresolved peak at centimeter wavelengths. In addition, Class 0 sources drive well-collimated molecular outflows, which clearly originate from the radio peaks. The measured velocities are much higher than those near less embedded stars. Searching molecular clouds for fast CO outflows and for peaks in continuum emission have been the main routes to discovering these heavily obscured stars.

Their properties suggest that Class 0 objects are in a very early and vigorous phase of evolution. The emitted spectra at infrared and millimeter wavelengths arise *entirely* from warmed dust, which must have an even higher column density than in Class I. Here again, millimeter continuum radiation, for which the dust is transparent, is valuable for probing the interior cloud structure (recall Figure 3.16). After integrating over the emitting region, one finds that the total luminosity for $\lambda > 350$ μm is indeed higher for Class 0 than for Class I. Specifically, the *ratio* of this long-wavelength luminosity to $L_{\rm bol}$ is typically 1 percent, while it is reduced by at least an order of magnitude in both Class I and II sources. A relatively larger amount of dust (and therefore gas) must be located close enough to the star to be warmed.

Let us examine more closely a typical Class 0 source, the embedded star in the Bok globule B335. Figure 11.25 displays the spectral energy distribution. We see immediately that λF_λ not only peaks at a longer wavelength than in the Class I case, but that the entire distribution is much narrower. The severe extinction of near- and mid-infrared flux cannot be interstellar, as the globule itself is a relatively isolated entity within a more extended molecular envelope of modest optical depth (recall Figure 3.20). Observations of the globule itself in NH_3 show that the star's location coincides with the peak in gas density.

It is again instructive to model the circumstellar matter as a spherical dust shell with an r^{-1} density variation. In this case, the prescribed density law is broadly consistent with the inner one found from molecular and dust observations (§ 3.3). To match the flux at the longest wavelengths, we now find that $r_{\rm out}$ should be increased to 2×10^{17} cm. The inner radius, on the other hand, can remain at 2×10^{12} cm. The biggest change is the increased outer density, which now corresponds to an $n_{\rm tot}$ of 3×10^5 cm^{-3} in the best-fit model. The emergent flux from this model is displayed as the solid curve in Figure 11.25. The value of A_V through the dust shell is now 320 mag, and the optical depth falls to unity only at $\lambda = 80$ μm. With such a high total column density, the grain absorption feature near 10 μm appears as a deep plunge in the emergent spectrum.

Our specific parameters for the shell cannot be assigned great weight, considering the present uncertainty regarding the dust opacity. We recall from Chapter 2 that the problem becomes especially severe at far-infrared and millimeter wavelengths. This, however, is the dominant spectral range for all Class 0 sources. In the present example, one could construct a shell model for B335 with significantly higher optical thickness. Reproducing the observed flux F_λ would then require a smaller solid angle $\Delta\Omega$ at each λ (recall equation (2.41)) and hence a more compact structure. Until the envelope is well resolved at all the relevant wavelengths, one cannot discount this possibility.

11.5.4 Self-Reversed and Asymmetric Lines

Even considering the range of acceptable shell models, there is no doubt that Class 0 sources are surrounded by denser envelopes than their Class I counterparts. But are they true protostars? In the case of B335, the bolometric luminosity is about 3 $L_\odot$, still low for a star undergoing accretion at the expected rate. Other sources are more luminous, but the total Class 0 sample currently numbers only a few dozen. Indeed, the very paucity of these sources argues in favor of their representing a relatively short-lived phase. The question of whether infall is occurring during this phase must be addressed through spectroscopic studies. That is, we need to examine the line profiles of molecular tracers to find any Doppler broadening induced by motion onto the protostar and its disk. Such broadening should be discernible in the quiescent gas of a dense core or Bok globule.

Figure 11.26 shows two emission spectra of the B335 infrared source. In the left panel, the line is the $J = 5 \rightarrow 4$ rotational transition of CS (λ = 1.2 mm), while the right panel shows the $J = 3 \rightarrow 2$ transition (λ = 2.0 mm) of the same molecule. The first line is optically thin, so that the brightness temperature at each velocity represents the contribution from all emitting molecules along the observed column. Here, some kind of bulk motion in the gas has symmetrically broadened the profile about line center, which itself corresponds to the mean cloud velocity of +8.4 km s^{-1}. This internal motion could conceivably be infall, rotation, or the outflow known to exist in the region. In any case, there is evidently an equal amount of material approaching and receding from the observer at each relative velocity.

The complex structure of the second profile signals the fact that the 2.0 mm line is optically thick to absorption by CS itself. In Chapter 6, we remarked that the optically thick profiles of $^{12}C^{16}O$ often have a flat-topped appearance that we associated with saturation broadening (Figure 6.1). The profile in Figure 11.26b instead shows a pronounced dip close to line center. This *self-reversal* is actually a common feature of optically thick lines, but only in the direction of embedded young stars. Previously, we noted a similar dip in the 21 cm HI profile of ρ Ophiuchi (Figure 3.7). A profile closely resembling Figure 11.26b, but in $^{12}C^{16}O$, is seen toward the Mon R2 association in NGC 2264. In all cases, the central depression reveals the presence of relatively cold, foreground gas that absorbs photons from warmer material behind it. This foreground gas does *not* participate in the bulk motion, but is static, thus accounting for the position of the dip at line center. The temperature is apparently climbing deeper into the cloud as a result of irradiation by the embedded star.

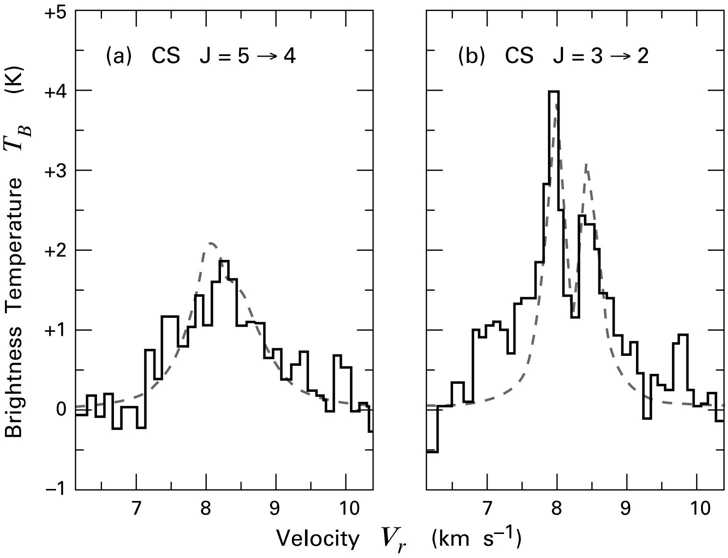


Figure 11.26 Two emission spectra in CS toward B335. The rotational transitions for each line are indicated. Dashed profiles are from a theoretical model assuming spherical infall.

To explore this picture more quantitatively, we start with equation (C.2) in Appendix C for the specific intensity of a line emerging from a slab-like cloud of thickness Δs. After further employing (C.8) for the source function and converting to optical depth as our independent variable, we find

$$I_\nu \;=\; \exp\left(-\Delta\tau_\nu\right) \int_0^{\Delta\tau_\nu} d\tau_\nu'\, B_{\nu_\circ}\left(T_{\rm ex}\right) \exp\tau_\nu' \;+\; I_\nu(0)\exp\left(-\Delta\tau_\nu\right) \;. \qquad (11.50)$$

Here, τ_ν is measured from the back face of the cloud, which has total optical thickness $\Delta\tau_\nu$ at each frequency (see Figure 11.27). The excitation temperature $T_{\rm ex}$ is assumed to fall with increasing τ_ν. This decline *could* stem from a corresponding drop in the gas kinetic temperature. It would also occur if the density were subcritical and falling toward the observer.

We next use equation (C.10) to replace I_ν by the brightness temperature $T_B(\nu)$. Since we are focusing on millimeter lines, we may employ the Rayleigh-Jeans approximation, $T_\circ \ll T_{\rm ex}$. We then find

$$T_B(\nu) \;=\; \exp\left(-\Delta\tau_\nu\right) \int_0^{\Delta\tau_\nu} d\tau_\nu'\, T_{\rm ex} \exp\left(\tau_\nu'\right) \;-\; T_\circ\, f(T_{\rm bg})\left[1 \;-\; \exp\left(-\Delta\tau_\nu\right)\right] \;. \qquad (11.51)$$

Here the function f in the last term was defined in equation (6.2). Note also that we have taken the background radiation field to be Planckian at the temperature $T_{\rm bg}$ and have neglected its frequency dependence over the line, *i. e.*, we have set $I_\nu(0)$ equal to $B_{\nu_\circ}(T_{\rm bg})$. We may simplify further by utilizing, as our independent variable, the optical depth measured from the *front* surface, $t_\nu \equiv \Delta\tau_\nu - \tau_\nu$ (Figure 11.27). Close to the line center, we suppose the cloud to

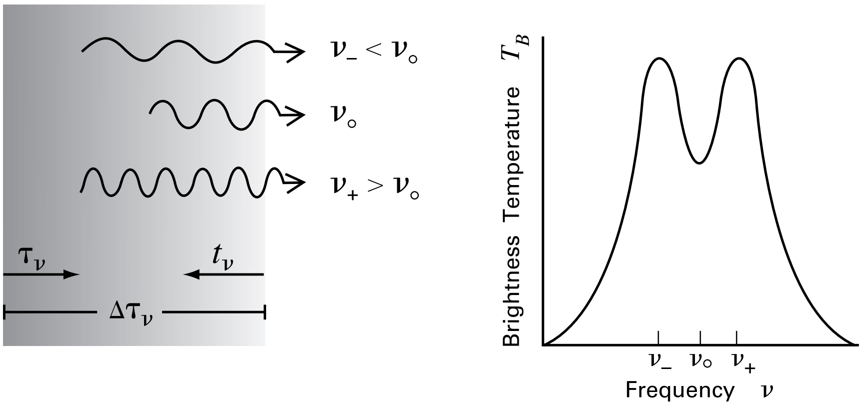


Figure 11.27 Origin of self-reversal in optically thick emission lines. Shading in the slab indicates the variation in excitation temperature. At each frequency ν, the slab has a total optical thickness $\Delta\tau_\nu$, which can be measured either from the rear (τ_ν) or the front (t_ν). A photon at line center $\nu_\circ$ originates relatively close to the front surface, while others come from deeper inside. The result is a self-reversed profile in the brightness temperature T_B.

be very optically thick, $\Delta\tau_\nu \gg 1$. Equation (11.51) then reduces to

$$T_B(\nu) \,=\, \int_0^\infty dt'_\nu \, T_{\rm ex} \exp\left(-t'_\nu\right) \,-\, T_\circ \, f(T_{\rm bg}) \qquad\qquad \Delta\tau_\nu \gg 1 \ . \tag{11.52}$$

The first term in this equation represents a mean value for $T_{\rm ex}$, evaluated between the cloud's front surface and the point where $t_\nu \sim 1$. Since the opacity peaks at line center, the depth of this point increases for ν either greater or less than $\nu_\circ$. As illustrated in Figure 11.27, the mean $T_{\rm ex}$ (and therefore $T_B(\nu)$) is lowest at $\nu_\circ$ and rises symmetrically on either side. This rise persists to frequencies where the optical thickness of the entire cloud is near unity. Even farther from line center, the cloud is optically thin, and equation (11.51) becomes

$$T_B(\nu) \,=\, [\bar{T}_{\rm ex} \,-\, T_\circ \, f(T_{\rm bg})] \, \Delta\tau_\nu \qquad\qquad \Delta\tau_\nu \ll 1 \ , \tag{11.53}$$

where $\bar{T}_{\rm ex}$ is now the mean value throughout the cloud. Since $\Delta\tau_\nu$ continues to fall with increasing departure of ν from $\nu_\circ$, the T_B-profile itself declines in this regime.

This model provides a basic account of self-reversed profiles, but it fails in several key respects. First, the slab is stationary, so that the predicted falloff in the line wings is much too steep compared to the observations. Even more intriguing is the manifest *asymmetry* of the optically thick profile, with the redshifted side depressed relative to the blue. This property appears not only in B335, but also in the other Class 0 sources exhibiting self-reversed emission profiles. One possible interpretation of the asymmetry is that the bulk motion in these cases is actually collapse onto the star.

Figure 11.28 illustrates the essential reasoning behind this idea. Here, the shading close to the star suggests the rise of $T_{\rm ex}$ in our idealized, spherical cloud. For simplicity, we have omitted the coldest, static material that gives rise to the self-reversal. Infalling gas from the

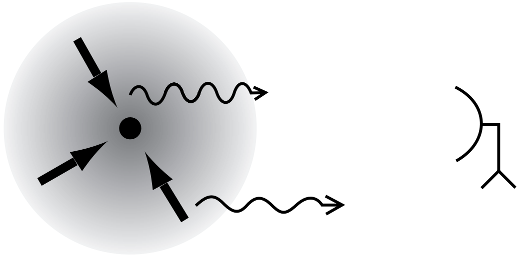


Figure 11.28 Contribution of infall to the asymmetry of molecular emission lines. The shading indicates the rise of T_{ex} toward the central protostar.

far side emits blueshifted photons, since it is approaching the observer. Conversely, the nearer side of the cloud contributes to the redshifted portion of the profile. Both types of photons still suffer extinction from other molecules that have the same velocities as the emitting species. Blueshifted photons have a greater chance of survival if they originate closer to the cloud's center. Here, T_{ex} is higher, so that the corresponding T_B is elevated in the profile. Most of the observed reddened photons, on the other hand, are emitted from the outer region of lower T_{ex}. The asymmetry in optically thick lines is thus neatly explained.

The dashed curves in Figure 11.26 are theoretical CS spectra derived from a radial infall model in the spirit of that described in § 10.2. The central portion of both profiles is well reproduced, but the predicted falloff in the line wings is evidently still too steep, at least in the righthand panel. It is tempting to associate the extended emission with the outflow. (The alternative of rapid rotation is inconsistent with the small gradient in radial velocity observed in this source.) Indeed, the relative contributions to the broadening from infall and outflow are far from certain. The mystery is deepened by the fact that some dense cores with no central star at all, such as L1544 in Taurus-Auriga, also exhibit asymmetric profiles. Here, neither collapse nor winds should be present. Although the effect shown in Figure 11.28 must operate at some level, the explanation of the presently observed asymmetries may be quite different.

In summary, current observations and theory have not yet established the protostellar nature of *any* objects, although the ones in Class 0 represent especially attractive candidates. The most dramatic observational feature in these sources, however, is the presence of strong outflows, and it will require more effort to demonstrate unequivocally an additional component of infalling gas. Improvement in our knowledge of dust opacities and continued high-resolution mapping at millimeter wavelengths are both necessary to elucidate the physical properties of the envelopes. Such observational studies, coupled with a better understanding of outflow structure itself, should eventually disentangle the various contributions to the molecular line profiles.

Chapter Summary

After a short-lived, transient phase, a stable protostar forms at the center of a collapsing dense core. Cloud matter that impacts the star does so in a strongly radiating accretion shock front. Shock-generated photons heat the incoming gas, destroying grains out to a radius of order

0.1 AU. The accretion luminosity diffuses through the dusty envelope to emerge as far-infrared continuum radiation.

The high energy loss from the shock leads to modest protostellar radii, not much larger than corresponding main-sequence values. As the object grows in both size and mass, deuterium eventually ignites and drives convection. Energy from deuterium fusion also tightly constrains the stellar radius as a function of mass. Convection ceases in intermediate-mass protostars, while deuterium continues to burn in an interior shell. If accretion continues, the protostar contracts and heats up until ordinary hydrogen ignites. A truly massive central object, *i. e.*, the precursor to an O star, would repel its infalling envelope through wind and radiation pressure, and therefore cannot form in this manner.

Disks arise during protostellar evolution from infalling material with too much angular momentum to impact the star itself. These geometrically thin structures expand rapidly with time. Streamlines from the outer disk collide and form a turbulent ring that feeds matter onto the central star. Eventually, the disk becomes gravitationally unstable. Spiral waves then create torques that may facilitate mass transfer and continuing protostellar accretion.

The observed spectral energy distributions of deeply embedded objects attest to the presence of heated dust. Simple dust-shell models reproduce the relative fluxes in both Class 0 and Class I sources, but give no clue as to which, if either, are true protostars. Molecular emission lines seen toward Class 0 objects and starless dense cores are frequently both self-reversed and asymmetric. The first feature indicates absorption by overlying, cold gas, while the second appears to signify inward motion.

Suggested Reading

Section 11.1 The establishment of the first core during spherical collapse is the subject of

> Masunaga, H., Miyama, S. M., & Inutsuka, S.-I. 1998, ApJ, 495, 346.

Our description of the main accretion phase follows

> Stahler, S. W., Shu, F. H., & Taam R. E. 1980, ApJ, 241, 637.

For a more recent calculation, see

> Masunaga, H. & Inutsuka, S.-I. 2000, ApJ, 531, 350.

A careful treatment of the envelope's thermal structure is

> Chick, K. M., Pollack, J. B., & Cassen, P. 1996, ApJ, 461, 956.

Section 11.2 The stellar structure equations, and the basic strategy for their solution, are presented in

> Clayton, D. D. 1983, *Principles of Stellar Structure and Nucleosynthesis* (Chicago: U. of Chicago), Chapter 6,

which also introduces the mixing-length theory of convection. A fuller exposition of the latter is in

Hansen, C. J. & Kawaler, S. D. 1994, *Stellar Interiors: Physical Principles, Structure, and Evolution* (New York: Springer-Verlag), Chapter 5.

The effect of deuterium burning on protostellar evolution is covered by

Stahler, S. W. 1988, ApJ, 332, 804.

Section 11.3 For a derivation of the equations governing early disk growth, see

Cassen, P. & Moosman, A. 1981, Icarus, 48, 353.

This paper also describes a number of different evolutionary scenarios. Our own account follows

Stahler, S. W., Korycansky, D. G., Brothers, M. J., & Touma, J. 1994, ApJ, 431, 341.

The original discussion of gravitational instability pertained to galaxies:

Toomre, A. 1964, ApJ, 139, 1217.

Later efforts at modeling disk evolution in various contexts are reviewed in

Lin, D. N. C. & Papaloizou, J. C. B. 1995, ARAA, 33, 505.

The emphasis here is on numerical simulations and on the possibility that disks are convectively unstable.

Section 11.4 The return to radiative stability in protostars and the onset of hydrogen burning are studied in

Palla, F. & Stahler, S. W. 1991, ApJ, 375, 288

____________________ 1992, ApJ, 392, 667.

For the CNO bi-cycle as a nuclear energy source, see

Rolfs, C. E. & Rodney, W. S. 1988, *Cauldrons in the Cosmos* (Chicago: U. of Chicago), Chapter 6.

Calculations that focus on the dynamical effect of radiation pressure in high-mass protostars are

Yorke, H. W. & Krügel, E. 1977, AA, 54, 183

Jijina, J. & Adams, F. C. 1996, ApJ, 462, 874.

Section 11.5 Two representative studies exploring dust envelope models for Class I sources are

Adams, F. C. & Shu, F. H. 1986, ApJ, 308, 836

Butner, H. M., Evans, N. J., Lester, D. F., Levreault, R. M., & Strom, S. E. 1991, ApJ, 376, 636.

The evolutionary status of Class 0 objects is discussed by

André, P., Ward-Thompson-D., & Barsony, M. 1993, ApJ, 406, 122.

The case for asymmetric profiles as signatures of collapse is presented in

Leung, C. M. & Brown, R. L. 1977, ApJ, 214, L73

Evans, N. J. 1999, ARAA, 37, 311.

The Formation of Stars
Steven W. Stahler and Francesco Palla
© 2004 WILEY-VCH Verlag GmbH & Co.

12 Multiple Star Formation

Our emphasis in the last two chapters has been on the buildup of individual protostars within dense cores. Star formation, however, is a group phenomenon. Each of the clusters and associations we surveyed in Chapter 4 results from the condensation of a large cloud or cloud complex. Hence, the transition to gravitational instability and protostellar collapse must occur in many locations at the same time, and over wide expanses of molecular gas. Moreover, a dense core does not normally give rise to just one star, but at least two. We know this because of the prevalence of binary systems both in the field and in young stellar populations. The members of binaries have similar ages, implying a common origin.

Any picture of single star formation is therefore a convenient fiction, the first approximation in a future, more comprehensive theory. Our findings are trustworthy only to the extent that interactions between young stars are relatively weak. If, as we shall argue, binaries form as individual protostars separated by many stellar radii, their mutual gravitational force is indeed minor compared to the self-gravity of either subcondensation. This is not to say that external forces play no role in star formation. Winds and supernovae from previously formed stars do exert a powerful influence. On a much larger scale, the collision of two galaxies triggers the dramatic starburst phenomenon. In both these cases, however, the probable effect is to alter the medium out of which new stars arise, rather than the basic collapse process itself.

Why, then, do stars tend to form in groups? Our understanding here is rather limited, at least compared to the evolution of single stars. The present chapter addresses this question by first reviewing theoretical results on cloud collapse. We next complete our earlier descriptive survey of stellar groups, summarizing empirical knowledge of the youngest binaries. This relatively new field already has a rich phenomenology that informs our subsequent discussion of binary formation. When turning to the origin of clusters in § 12.4, we delve into the evolution of luminosity functions and find empirically the stellar production rate within selected groups. We address, in a qualitative manner, the issue of why some clouds form T associations and others bound clusters. Finally, § 12.5 revisits OB associations. Again drawing on the observations, we sketch a view of their history and argue that high-mass stars are likely to form through a coagulation process near the associations' crowded centers.

12.1 Dynamical Fragmentation of Massive Clouds

Traditionally, the theory of gravitational collapse has largely concerned itself with the fate of objects that are too massive to undergo the inside-out evolution studied earlier. Such clouds break up during the course of their free-fall collapse. Let us first survey the main results concerning such *dynamical fragmentation*. As we shall see, the theory is well developed quantitatively and

The Formation of Stars. Steven W. Stahler and Francesco Palla
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-40559-3

has much to teach us about the physics of collapse. On the other hand, we will also see why this mode of evolution probably does not apply to most star-forming systems.

12.1.1 Role of the Jeans Length

We continue and extend our previous discussion of gravitational collapse by first remarking that the concept of the Jeans length has subtly entered the theory in two distinct ways. In the original analysis of Jeans, presented in § 9.1, λ_J characterizes the behavior of small perturbations. Specifically, one considers isothermal fluctuations in a medium of uniform density and temperature. Those disturbances with length scale less than λ_J oscillate periodically, while larger ones grow in amplitude.

On the other hand, the detailed construction of cloud models reveals a quite different aspect of λ_J. This quantity now represents an approximate *upper limit* to the size of possible equilibria. For example, the rotating cloud depicted in Figure 9.7a has equatorial and polar radii that are both of order λ_J. This configuration is also nearly the largest of any member of the $\beta = 0.16$ sequence. (Recall that β measures the relative amount of rotational energy.) Models with higher $\rho_c/\rho_\circ$, such as that in Figure 9.7b, are smaller as a result of their enhanced self-gravity. An even simpler example is provided by the sequence of isothermal spheres. Equations (9.6) and (9.23) imply that the dimensional cloud radius $r_\circ$ is given in terms of the Jeans length by

$$\frac{r_\circ}{\lambda_J} = \frac{\xi}{2\,\pi}\left(\frac{\rho_c}{\rho_\circ}\right)^{-1/2} . \tag{12.1}$$

For any density contrast $\rho_c/\rho_\circ$, we may read off the nondimensional radius ξ from Figure 9.1, which actually displays the inverse, $\rho_\circ/\rho_c$. We find, in this way, that $r_\circ/\lambda_J$ rises from zero at $\rho_c/\rho_\circ = 1$, reaches a maximum value of 0.29 at $\rho_c/\rho_\circ = 5.1$, and then gradually declines to an asymptotic value of 0.23.

These two roles of λ_J are not only logically independent but contradictory. Given the size limit from the model studies, it follows that the very large background state assumed by Jeans *cannot exist*. This fact becomes clearer once we realize that, in deriving equations (9.20a)–(9.20d), we implicitly ignored any gradient in the background gravitational potential Φ_g. The momentum equation (9.18) apparently justifies this step. In a static ($\boldsymbol{u} = 0$) medium of uniform density and pressure ($\boldsymbol{\nabla} P = 0$), it is indeed true that $\boldsymbol{\nabla}\Phi_g$ vanishes. However, Poisson's equation (9.3) implies that Φ_g *cannot* be spatially uniform, but must change appreciably, *i. e.*, by an amount a_T^2, over the characteristic distance $(a_T^2/G\rho_\circ)^{1/2} \approx \lambda_J$. The background pressure and density also vary over the same distance, which therefore sets the dimension of the entire system.[1]

While the derivation of Jeans remains instructive, it is only the second interpretation of λ_J, as the maximum spatial extent for equilibria, that has true physical significance. Larger structures cannot be in force balance, but must be collapsing *ab initio*. One can therefore be confident that static molecular clouds whose size greatly surpasses the Jeans limit, notably the

---

[1] We now see why the original derivation of λ_J is sometimes referred to as the *Jeans swindle*. Of course, the analysis can be saved if one considers only perturbations with size much smaller than λ_J. Such disturbances, however, are essentially sound waves, for which self-gravity plays no role.

clumps in giant complexes, are sustained by internal forces beyond gas pressure. We have identified these forces as arising from the interstellar magnetic field. The formation of small cores within such entities must proceed through the gradual loss of this extra support and not through the fluctuations envisaged by Jeans. Finally, the cores themselves grow until their sizes roughly exceed λ_J, at which point they collapse.

What we have said so far is familiar enough. We now wish to explore more deeply the collapse of clouds larger than λ_J (or, equivalently, more massive than M_J). Historically, this problem has been the focus of many theoretical studies. The issue has gradually diminished in importance with increased understanding of the star-forming environment. We have just argued that clouds with masses above M_J are never in force balance, so it is not obvious how they arise in the first place. We defer consideration of this important, general issue, as well as the specific relevance of our inquiry to multiple star formation.

12.1.2 Pressure-Free Collapse

It is natural to begin with the simplest case, a cloud with no internal pressure at all, *i. e.*, with $\lambda_J = 0$. If we further postulate that the cloud be initially spherical and of uniform density $\rho_\circ$, then its collapse can be followed analytically. We employ the momentum equation (10.29), setting $a_T = 0$ on the righthand side. The resulting equation, along with (10.27) for mass continuity, constitute an *Eulerian* description of spherical, pressure-free collapse, in which the independent variables are t and the usual spherical radius r. It will prove more convenient to employ a *Lagrangian* description, where the spatial coordinate is tied to the fluid motion. That is, we replace r by the interior mass M_r. The partial derivatives of M_r, as given by equations (10.26) and (10.28), can then be used to transform such terms as $(\partial u/\partial r)_t$. Note that, in the new description, r is a *dependent* variable. Its own temporal rate of change is simply the velocity u:

$$\begin{aligned}\left(\frac{\partial r}{\partial t}\right)_{M_r} &= -\frac{(\partial M_r/\partial t)_r}{(\partial M_r/\partial r)_t}\\ &= \frac{4\pi r^2 \rho u}{4\pi r^2 \rho}\\ &= u\ .\end{aligned} \tag{12.2}$$

The pressure-free version of the momentum equation (10.29) now becomes

$$\left(\frac{\partial^2 r}{\partial t^2}\right)_{M_r} = -\frac{G M_r}{r^2}\ , \tag{12.3}$$

while the continuity equation (10.27) transforms to

$$\frac{1}{\rho^2}\left(\frac{\partial \rho}{\partial t}\right)_{M_r} = -4\pi\left[\frac{\partial(r^2 u)}{\partial M_r}\right]_t\ . \tag{12.4}$$

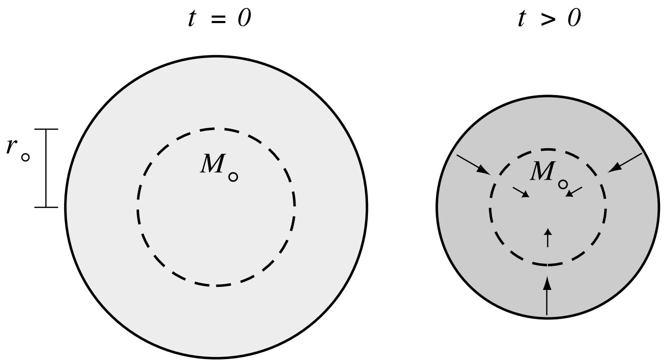


Figure 12.1 Collapse of a uniform-density, pressureless sphere. The interior sphere contains mass $M_\circ$ and has a radius $r_\circ$ initially. At a later time, this region shrinks, while the velocity vectors diminish to zero at the origin.

Imagine now a fluid element located at an initial radius $r_\circ$ (see Figure 12.1). During collapse, the mass $M_\circ$ interior to the element remains constant. After multiplying both sides of (12.3) by $(\partial r/\partial t)_{M_r}$, we integrate from $t = 0$ and obtain an expression for the velocity:

$$\left(\frac{\partial r}{\partial t}\right)_{M_r} = -\sqrt{\frac{2\,G\,M_\circ}{r_\circ}}\left(\frac{r_\circ}{r} - 1\right)^{1/2} . \tag{12.5}$$

Here, the negative sign signifies that the collapse is inward, *i. e.*, that r decreases with time.

To integrate equation (12.5), we convert to nondimensional variables, $\xi \equiv r/r_\circ$ and $\tau \equiv t/t_\circ$. If we choose the fiducial time $t_\circ$ to obey

$$\begin{aligned} t_\circ^2 &\equiv \frac{r_\circ^3}{2\,G\,M_\circ} \\ &= \frac{3}{8\,\pi\,G\,\rho_\circ} , \end{aligned} \tag{12.6}$$

then (12.5) simplifies to

$$\frac{d\xi}{d\tau} = -\left(\frac{1}{\xi} - 1\right)^{1/2} ,$$

where we have dropped the partial derivative notation. We further let $\xi \equiv \cos^2\,\alpha$, so that α goes from 0 to $\pi/2$ as ξ goes from 1 to 0. The last equation then becomes

$$\frac{d\alpha}{d\tau} = \frac{1}{2\,\cos^2\,\alpha} ,$$

which we integrate to find

$$\alpha + \frac{1}{2}\sin 2\,\alpha = \tau . \tag{12.7}$$

Equation (12.7) implies that our fluid element reaches the origin at $\tau = \pi/2$. Reverting to dimensional variables, we see that the collapse spans the interval $t_{\rm ff} \equiv (\pi/2)\, t_\circ$. From equation (12.6), this is also

$$t_{\rm ff} \,=\, \left(\frac{3\,\pi}{32\,G\,\rho_\circ}\right)^{1/2} \quad , \tag{12.8}$$

which motivates the definition of the free-fall time in equation (3.15). Since this result depends only on $\rho_\circ$, and not the value of the starting radius, *all* fluid elements reach $r = 0$ at precisely the same instant.

The behavior of the density during collapse follows from the continuity equation (12.4). We first use equations (12.5) and (12.6) to write

$$r^2 u \,=\, -\frac{r_\circ^3}{t_\circ}\,\cos^3\alpha\,\sin\alpha \ .$$

Here, only the term $r_\circ^3$ has a spatial dependence. Moreover, its derivative with respect to M_r is the same at all times:

$$\left(\frac{\partial\, r_\circ^3}{\partial\, M_r}\right)_t \,=\, \frac{3}{4\,\pi\,\rho_\circ} \ .$$

Suppressing the M_r-subscript on its left side, equation (12.4) now becomes

$$\begin{aligned}\frac{\partial\,(1/\rho)}{\partial\,\alpha} \,&=\, -\frac{6\,\cos^5\alpha\,\sin\alpha}{\rho_\circ} \\ &=\, \frac{1}{\rho_\circ}\,\frac{\partial\,(\cos^6\alpha)}{\partial\,\alpha} \ ,\end{aligned}$$

so that

$$\rho \,=\, \rho_\circ\,\sec^6\!\alpha \ . \tag{12.9}$$

The density thus remains spatially uniform as it increases without bound during the collapse. Figure 12.2 plots ρ as a function of time, where the latter is the nondimensional value $\tau^* \equiv \log\,[t_{\rm ff}/(t_{\rm ff} - t)]$. This "stretched" coordinate goes from 0 to ∞ as t itself progresses from 0 to $t_{\rm ff}$.

12.1.3 Growth of Perturbations

One interesting aspect of pressure-free collapse is its sensitivity to initial conditions. Rapid divergence of the evolution from the simple picture just described may occur in a number of ways. Imagine, for example, that the starting configuration were a very slightly oblate, uniform-density spheroid. The gravitational force acting on the pole would then be fractionally higher than that on the equatorial edge. No matter how small the difference, the spheroid would tend to flatten as a result. The force discrepancy would thereby increase, and the flattening accelerate. After a total elapsed time that is under $t_{\rm ff}$, the original object would collapse to a flat disk. Conversely, a cloud that is initially prolate, to whatever small degree, inevitably narrows to a thin needle along its central axis.

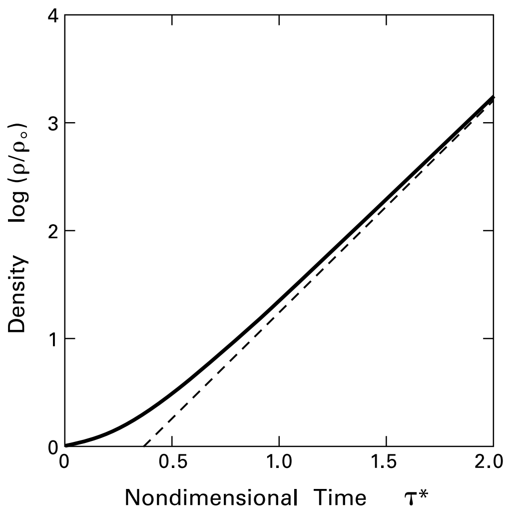


Figure 12.2 Density as a function of time in a collapsing, pressureless sphere. Plotted is $\rho/\rho_\circ$, the ratio of the density to its initial value. The temporal coordinate τ^* is defined in the text. The dashed curve is an analytic approximation for late times.

Even if the collapse were to remain spherically symmetric, any small perturbations within it would grow in magnitude. Such disturbances could arise either in the cloud's density or its velocity field. Their growth stems in part from the convergent nature of the main flow. Equation (12.5) may be recast as

$$\begin{aligned} u &= -\frac{r_\circ}{t_\circ}\tan\alpha \\ &= -\frac{r}{t_\circ}\sec^2\!\alpha\,\tan\alpha\ . \end{aligned} \tag{12.10}$$

Since α is only a function of time, the collapse velocity decreases linearly toward the origin, as indicated in Figure 12.1.[2] This flow pattern tends to compress internal perturbations. The situation is very different in a cloud undergoing inside-out collapse. Here, the fluid speed within the rarefaction wave is approximately the free-fall value $V_{\rm ff}$, which *increases* as $r^{-1/2}$ toward the center. Small lumps are therefore tidally stretched and prevented from rapidly growing.

An even stronger cause for perturbation growth in the pressure-free case is the self-gravity within the lumps. Ultimately, this effect causes them to reach high density ahead of the background flow. We term such runaway condensation of localized regions during collapse *dynamical fragmentation.* To describe the effect quantitatively, we first derive an explicit formula for the rise in parent cloud density, using equations (12.7) and (12.9). Toward the end of collapse, it is appropriate to replace α by another, small parameter:

$$\epsilon \equiv \frac{\pi}{2} - \alpha\ ,$$

---

[2] Pressure-free collapse is said to be *homologous*, since any fluid variable preserves the same spatial dependence at all times. By the same token, inside-out collapse is *nonhomologous*; recall § 10.2.

where $\epsilon \ll 1$. A Taylor-series expansion of equation (12.7) then tells us the time corresponding to any ϵ-value:

$$t \approx t_{\rm ff} \left(1 - \frac{4\,\epsilon^3}{3\,\pi}\right) .$$

Similarly, equation (12.9) gives the density as

$$\rho \approx \rho_\circ\,\epsilon^{-6} .$$

Combining the last two results, we find that

$$\rho(t) \approx \rho_\circ \left[\frac{3\,\pi\,(1 - t/t_{\rm ff})}{4}\right]^{-2} . \tag{12.11}$$

The dashed line in Figure 12.2 shows that this approximation is quite accurate by $\tau^* \approx 1$, or $t/t_{\rm ff} \approx 0.9$.

Imagine now that a central region of the sphere starts with a slightly enhanced density $\rho'(0) = \rho_\circ + \delta\rho(0)$. From equation (12.8), this portion will collapse to the origin a bit faster, within a time given approximately by

$$t'_{\rm ff} \approx t_{\rm ff} \left[1 - \frac{\delta\rho(0)}{2\,\rho_\circ}\right] . \tag{12.12}$$

Applying (12.11) to the enhanced density, we find its increase over the background:

$$\begin{aligned}\frac{\rho'(t)}{\rho\,(t)} &\approx \left(\frac{t_{\rm ff} - t}{t'_{\rm ff} - t}\right)^2 \\ &\approx \left[1 + \frac{\delta\rho(0)\,t_{\rm ff}}{2\,\rho_\circ\,(t'_{\rm ff} - t)}\right]^2 \\ &\approx 1 + \frac{\delta\rho(0)\,t_{\rm ff}}{\rho_\circ\,(t_{\rm ff} - t)} .\end{aligned}$$

In the last step, we have assumed that $\rho'(t)$ is still only slightly greater than $\rho(t)$. The density enhancement $\delta\rho(t) \equiv \rho'(t) - \rho(t)$ thus grows as

$$\frac{\delta\rho(t)}{\rho(t)} \approx \frac{\delta\rho(0)\,t_{\rm ff}}{\rho_\circ\,(t_{\rm ff} - t)} . \tag{12.13}$$

An initial perturbation of 1 percent, becomes a detached fragment ($\delta\rho \approx \rho$) only when $t/t_{\rm ff} \approx 0.99$, by which time the parent density has climbed by four orders of magnitude.

Disturbances within our collapsing sphere condense in a manner qualitatively different from the Jeans prediction in Chapter 9. First, the growth of $\delta\rho/\rho$ is *slower* than the exponential rise in the (fictitious) static medium. Second, pressure-free collapse is so prone to fragmentation that perturbations of all sizes grow at the same rate. That is, the small-scale cutoff predicted by equation (9.22) is absent. This last difference stems from our initial assumption that $\lambda_J = 0$, which has precluded any statement about the characteristic sizes or masses of fragments.

We can make further progress by considering the evolution of clouds that, while larger than λ_J, have finite pressures from the start. A detailed analysis for spheres of initially uniform density confirms that disturbances of length λ greater than λ_J grow relative to the background, while internal pressure supports those below this size limit. Even the largest perturbations are initially retarded to some degree by pressure. However, λ_J decreases as $\rho^{-1/2}$ in an isothermal medium (recall equation (9.23)), while λ in a lump of fixed mass falls roughly as $\rho^{-1/3}$. The ratio λ/λ_J therefore rises with time. The growth rate of the perturbation accelerates, until it approaches that given by equation (12.13).

12.1.4 Numerical Methods

These results apply only to the linear regime of small perturbations. Following collapse to the point of true fragmentation requires multidimensional numerical simulations. Starting with small, random perturbations, the analytic results indicate that one must follow the collapse to relatively high density before fragmentation occurs. This requirement severely taxes a conventional hydrodynamics solver that tracks fluid motion in a fixed, Eulerian coordinate grid. On the other hand, a purely Lagrangian scheme, such as the one used above, works best in spherical geometry and cannot easily follow severely distorted condensations.

Two innovative numerical techniques have proved effective in addressing the need to improve spatial resolution in a flexible manner. In *smoothed-particle hydrodynamics (SPH)*, one replaces the continuum fluid by a collection of moving points. The method derives its name from the fact that the physical variables associated with each point are smeared out over a small volume according to some prescribed smoothing function. These regions overlap, so that the value of a variable and its gradient at any spatial location must be obtained by summing contributions from a number of elements. The particles interact through both pressure, as calculated from the density and temperature gradient, and through their mutual gravitational attraction. The computer tracks all the individual orbits simultaneously, just as it would for stars in a simulated galaxy. In principle, any contracting region can be followed to arbitrarily high density, provided it contains a sufficient number of particles.

An alternative strategy is to retain the basic picture of continuum hydrodynamics, but to introduce additional coordinate zones as they are required. The most successful programs employ a flexible *nested grid* technique. If matter collects in a single cell of the original mesh, that cell is subdivided into an array of new ones. Each of these new cells may be further partitioned, or the zoning made coarser, as the evolution proceeds. Note that t_{ff}, the characteristic time scale for gravitational contraction, decreases as $\rho^{-1/2}$. A highly condensed region within a diffuse envelope thus evolves much more rapidly than its surroundings. Accordingly, one must track the fluid for several small time steps within a subgrid, while freezing the motion on larger scales. Changes accumulated during these short times are then transferred to the parent grid at specified intervals.

We saw in Chapter 10 how the material within collapsing clouds attains supersonic motion that can only be arrested in strongly radiating shock fronts. Locating this front is relatively simple in a spherical or axisymmetric collapse, where it constitutes the boundary of the protostar and its surrounding disk. In a massive cloud that is undergoing fragmentation, the accurate tracing of shocks represents a true challenge. The most widely used strategy is to employ some form of *artificial viscosity*. One introduces an internal, frictional force that decelerates the flow whenever the local velocity gradient becomes too steep. In this way, curved shock fronts can be identified automatically as they arise, at least within one or two zones. Material before and after the front obeys the correct shock jump conditions, so that one can also follow the associated energy losses. On the other hand, the technique is unsuitable for calculating detailed thermal and chemical relaxation processes close to the front itself.

12.1.5 Oblate and Prolate Configurations

Theorists employing these and other computational tools have been able to follow collapse in three dimensions over a wide range of initial conditions. Most simulations begin with uniform, spherical clouds that have a significant degree of *rotation*. The choice here is a practical one. Clouds that rotate while collapsing flatten along the rotational axis. For computational purposes, it is easier to follow the subsequent fragmentation in this extended configuration than in a much more compact, centralized region. The essential result, however, is the same as that indicated for the non-rotating sphere. Clouds with a total mass M exceeding M_J break apart into roughly M/M_J subcondensations.

Of course, this broad generalization glosses over a wealth of interesting details, which we cannot hope to cover adequately. A few representative studies must suffice. Let us first see what the simulations have taught us about the primary question whether a cloud undergoes dynamical fragmentation at all. If the initial configuration is a rigidly rotating sphere of uniform density, then it may be characterized by a nondimensional mass m and rotational parameter β. Here, we recall from § 9.2 that m closely approximates the number of Jeans masses. Once this cloud begins to collapse, it always flattens significantly, within about one free-fall time. Depending on the values of m and β, the evolution then follows one of two paths. Either the cloud rebounds to a *less* flattened shape, or else the flattening continues and the density in the plane rises at an accelerated pace. Earlier calculations required, for computational convenience, that the fluid preserve azimuthal symmetry about the original rotation axis. Under this artificial restriction, a cloud following the second path coalesces into a dense, equatorial ring. If the same collapse is followed with a fully three-dimensional code, any incipient ring quickly breaks up into two or more pieces.

The left panel of Figure 12.3 shows one example of such a fragmenting configuration. This SPH calculation utilized 1000 particles to simulate the isothermal collapse of a sphere with $m = 2.0$ and $\beta = 0.3$. Note that the latter value is close to the upper limit of $\beta = 1/3$, corresponding to breakup in the initial object. The figure displays the particle distribution in the equatorial plane at $t = 2.3\, t_{\rm ff}$ after the onset of collapse. By this time, the cloud has become highly flattened, with an aspect ratio of 6:1. Three distinct fragments are evident, with the maximum density being a factor of 4×10^3 greater than that of the parent sphere.

Figure 12.4 summarizes graphically the outcomes of several dozen collapses of rotating spheres. Each point represents a specific numerical simulation, located in the plane according to its values of m and β. The filled circles correspond to rebounding collapses. Open circles denote situations where the equatorial density increases without limit, either through ring formation or more direct breakup. Finally, the solid curve, taken from Figure 9.9, is $m_{\rm crit}(\beta)$, the locus of marginally stable *equilibrium* configurations. It is evident that nearly all the runaway collapses lie above this envelope. Conversely, the other clouds rebound about states that strongly resemble the stable equilibria with the appropriate values of m and β (see, *e. g.*, Figure 9.7a). We thus see that *collapsing clouds fragment if they have no accessible equilibrium states.* That is, fragmentation occurs only for $M > M_{\rm crit}$, where $M_{\rm crit}$ is the rotational generalization of M_J.

It is especially interesting to follow the evolution of clouds lying only slightly above the critical line in Figure 12.4. These objects exhibit a runaway growth in density, but *without fragmentation.* In both two- and three-dimensional calculations, they evolve to flattened structures with a steeply rising central concentration. Their collapse must continue in the nonhomologous, inside-out fashion that is well studied for the special case of $\beta = 0$. Since this type of infall is not prone to dynamical fragmentation, we now see what limits the process quite generally. Any cloud, no matter how massive, stops subdividing once individual pieces have significant thermal support. The further collapse of these pieces, each of mass about M_J, can proceed without additional breakup.

This is not to say that every fragment goes on to form a star. In the very smallest, thermal support may be high enough to prevent further condensation. The system of three orbiting bodies in the left panel of Figure 12.3 is probably unstable, in the same manner as a triplet of stars. Within a few orbital periods, one of the bodies (probably the least massive) is likely to be ejected, leaving the remaining pair more tightly bound. Fragments may also coalesce, particularly when a large number are produced. The right panel in Figure 12.3 shows one such situation, resulting from the collapse of a cloud with $m = 5.8$. Here, the initial sphere collapsed to a much thinner disk, with an aspect ratio of $14 : 1$. Once the disk had formed, material close to the center in the equatorial plane was flung out centrifugally, only to crash into additional matter approaching from farther out. It is at this point that the disk began to fragment, without

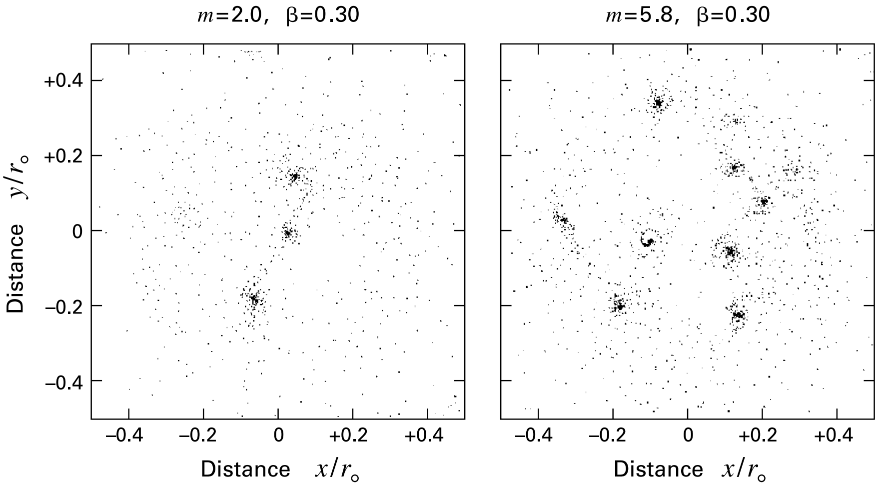


Figure 12.3 Collapse of rotating, spherical clouds. The two panels show the density distribution in the equatorial plane. They correspond to the indicated values of the rotational parameter β and nondimensional cloud mass.

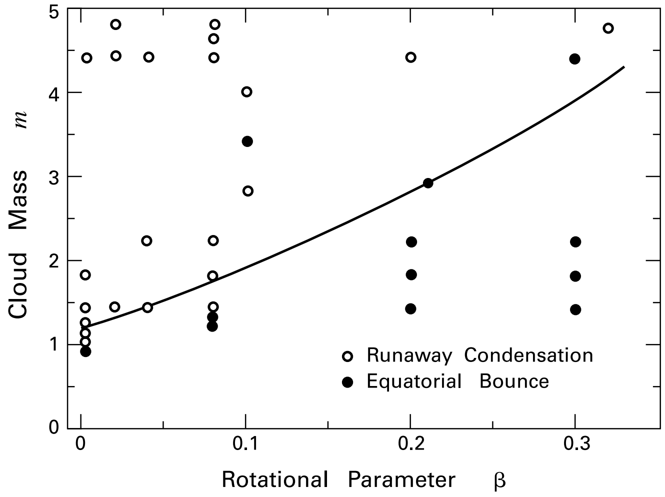


Figure 12.4 Outcomes for numerical collapses of rotating spheres. Within the m-β plane, open circles represent collapses where the central density increased without limit. For closed circles, the cloud underwent an equatorial bounce. The solid curve traces the maximum mass for equilibria.

ever forming a ring. For the time shown, $t = 2.2\,t_{\rm ff}$, eight well-defined fragments are orbiting one another. Some of these are expected to merge if the calculation were continued.

We mentioned earlier that pressure-free clouds which are even mildly elongated collapse to ever-narrowing filaments. Endowing the object with a relatively low temperature does not alter this result, provided the temperature is held strictly constant. More realistically, the deep interior must eventually heat up because of increasing optical thickness. The added pressure gradient slows the constriction, giving the cloud time to break up along its axis. Figure 12.5 shows what happens when the elongated object tumbles end over end. In this SPH calculation, the initial structure was a uniform-density ellipsoid, undergoing differential rotation in the x-y plane. The ellipsoid was fully triaxial, with axes in the ratio $2 : \sqrt{2} : 1$, and with a mass of about 20 M_J. Collapse again leads to a very narrow filament, but now exhibiting large-scale bending because of the rotation. By the time depicted, $1.6\,t_{\rm ff}$ after the onset of collapse, a chain of high-density knots has appeared.

Taken as a whole, our examples illustrate two central points. First, the detailed outcome of any specific collapse is sensitive to initial conditions. The cloud's shape, internal density variation, and state of rotation all play a role in determining such results as the spatial distribution of fragments and their subsequent interaction. On the other hand, the basic issue of whether a cloud fragments or not is largely *independent* of these factors. As long as thermal pressure is a major source of support prior to collapse, the cloud's fate in this regard is mainly set by its mass relative to M_J. If this ratio is large, the object is highly susceptible to fragmentation, no matter which route it takes. Thus, clouds which are initially flattened disks or slabs break up into pieces with size comparable to the original cloud thickness. Elongated filaments quickly separate into fragments of roughly the same length as the filament diameter. In either case, thermal support

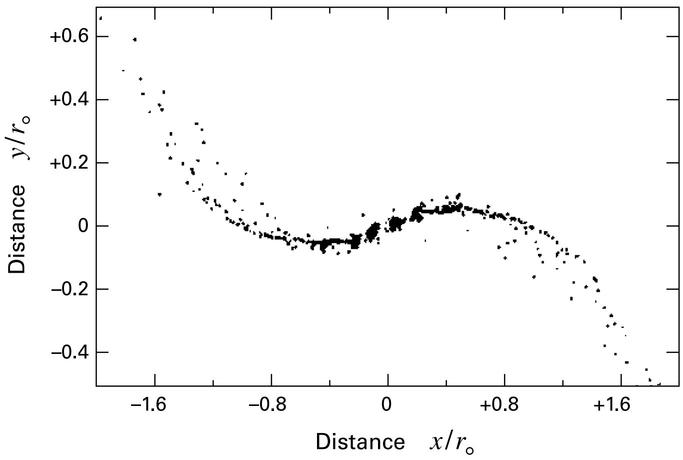


Figure 12.5 Collapse of a uniform-density, rotating ellipsoid. Displayed is the density in the equatorial plane, at a time when the cloud has started to fragment.

is much more important in the newly formed substructures than in the parent body, so that their subsequent evolution is also qualitatively distinct.

12.2 Young Binary Stars

The foregoing class of calculations helps to constrain the problem of how large clouds give rise to the many smaller structures destined to become stars. Let us pause in the theoretical development and turn to the empirical description of groups. Chapter 4 already described associations and bound clusters, both in terms of their intrinsic properties and in their relation to molecular clouds. Here, we complete the picture by treating the simplest of all groups, the binaries. It has long been known that a typical main-sequence field star is more likely to have an orbiting companion than to be isolated. The finding that pre-main-sequence stars in some regions have an even higher proportion of binaries is an exciting development whose implications are still unclear.

12.2.1 Basic Properties

Figure 12.6a illustrates the kinematics of a binary pair consisting of stellar masses M_1 and M_2. Each star follows an elliptical orbit about the system's center of mass, indicated as a cross in the figure. The two ellipses have identical eccentricity e, but semi-major axes, a_1 and a_2, that are inversely proportional to the associated masses. If u_i and r_i are the velocity and radial distance, respectively, of either star at an arbitrary time, then we have

$$\frac{M_1}{M_2} = \frac{u_2}{u_1} = \frac{r_2}{r_1} \, . \tag{12.14}$$

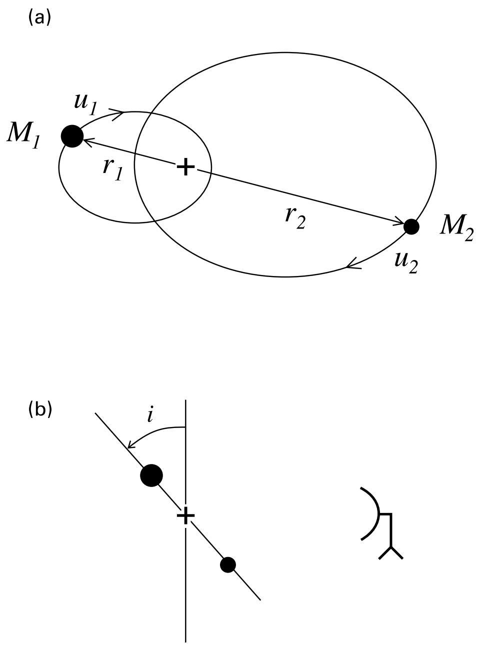


Figure 12.6 Motion of a binary's components (a) within the orbital plane, and (b) when that plane is seen edge-on. The stars travel on identical, but rescaled, ellipses about their center of mass, indicated by the cross in both panels. Also displayed are the radial distances and velocities in the center-of-mass reference frame, as well as the inclination angle between the orbital plane and the line of sight.

In the figure, we have assumed that M_1 exceeds M_2. For main-sequence objects, the more massive star is also brighter, *i. e.*, it is the primary in the system. Finally, note that both stars complete a single orbit in the same period P, given by the generalization of Kepler's third law for planetary motion:

$$\begin{aligned} P &= \frac{2\,\pi\,a_{\rm tot}^{3/2}}{(G\,M_{\rm tot})^{1/2}} \\ &= 1\ {\rm yr}\,\left(\frac{a_{\rm tot}}{1\ {\rm AU}}\right)^{3/2}\left(\frac{M_{\rm tot}}{1\,M_\odot}\right)^{-1/2} \quad . \end{aligned} \tag{12.15}$$

Here, $a_{\rm tot}$ and $M_{\rm tot}$ are

$$\begin{aligned} a_{\rm tot} &= a_1 + a_2 \\ M_{\rm tot} &= M_1 + M_2 \ . \end{aligned} \tag{12.16}$$

The most direct means of identifying a binary would be to observe the proper motion of both components. If two stars trace parallel paths across the sky, one can be sure they constitute a bound system. The orbital period may be so long, however, that both objects simply follow the straight line corresponding to their center-of-mass velocity. In more favorable cases, one sees the distortion of both paths from this linear motion. Such *visual* binaries have periods measured in decades or centuries. Relatively rare *astrometric* systems display modulated proper motion for a single star, whose partner is too faint for detection.

The orbital motion of stars in tighter systems with shorter periods is only discernible through the changing Doppler shift of spectral lines. These *spectroscopic* binaries are *double-lined* if one sees both sets of stellar absorption lines; otherwise, they are *single-lined.* Note that the Doppler shifts of either star do not track its full velocity u, but only the projection $u \sin i$. As illustrated in Figure 12.6b, i represents the inclination angle between the orbital plane (here shown edge-on) and the plane normal to the line of sight. Finally, repeated photometry of a star in some waveband often reveals periodic dips in its flux. Such minima may indicate rotating starspots, or else partial occultation by an orbiting companion. In the latter case, the binary is said to be *eclipsing*. Of course, these categorizations are not mutually exclusive. An eclipsing system, for example, usually displays spectroscopic variation, as well.

Binary observations date back to the earliest years of telescopic astronomy, but true surveys of stellar multiplicity are relatively recent. Here, one must contend with a number of selection effects that severely bias the raw data. Most apparent is the increasing difficulty of detecting relatively faint or low-mass companions. In practice, there is a maximum observable difference in magnitude at any waveband between the members of visual binaries. Companions to the surveyed stars become more difficult to find as this limit is approached. Similarly, spectroscopic binaries are only detectable above some threshold amplitude in the radial velocity variation. Their discovery is thus more problematic as one investigates stars of decreasing mass. On the other hand, high-mass stars tend to rotate more rapidly. Rotation broadens any absorption line and makes it difficult to measure a shift in the line center. Finally, a binary system whose orbit happens to lie in the plane of the sky exhibits no Doppler shift for either star, regardless of their masses.

Minimizing the bias from such effects requires first that a study be carefully designed. For example, "magnitude-limited" surveys, such as those observing all objects brighter than a prescribed m_V, tend to exaggerate the number of double-lined spectroscopic binaries. These have secondaries of nearly equal brightness and so are more likely to surpass the magnitude limit than single-line systems. Setting a maximum *distance* for the program stars overcomes this difficulty. It is also prudent to concentrate on a limited range of spectral types, in view of the special difficulties associated with various masses. Even such a restricted, distance-limited sample requires extensive analysis to correct the results for incompleteness.

12.2.2 Main-Sequence Systems

The best studied group is G-type main-sequence stars within the solar neighborhood. Out of 164 systems closer than 22 pc, the singles, doubles, triples, and quadruples are found to be in the proportion $57 : 38 : 4 : 1$. Thus, 43 percent of the stars have at least one companion. This figure only includes orbiting stars whose mass is more than 10 percent that of the primary. After allowing for incompleteness, the true fraction of systems that are multiple (within the same limiting mass ratio) is estimated to be 57 percent.

The double systems within this local population include spectroscopic, visual, and astrometric binaries, as well as pairs of stars with common proper motion. Observations of the latter do not directly yield periods. However, one may assign masses to both stars from their spectral types, convert (on a statistical basis) the observed separation to a semi-major axis $a_{\rm tot}$, and then use equation (12.15) to obtain P. Figure 12.7 displays the distribution of periods in the whole sample of binaries. The ordinate is the *differential binary frequency* f_B, defined as the number of binaries per logarithmic period interval divided by total number of systems, both single and

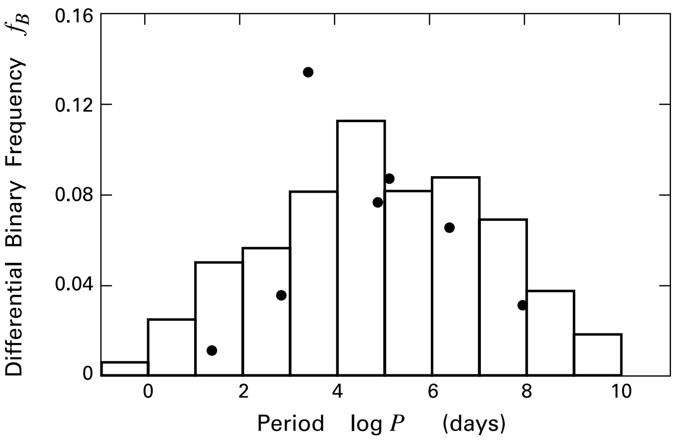


Figure 12.7 Period distribution of main-sequence binaries in the solar neighborhood. The solid histogram is the differential binary frequency for G dwarfs, while the filled circles are the results for M-type stars.

multiple. Note that a triple contributes two binaries, and so on. Figure 12.7 also includes estimated entries for the dozen or so systems that are undetected because of selection effects. It is apparent that the periods span a very broad range, from under one day to several thousand years. The distribution is also strikingly symmetric, with a broad maximum near 180 yr.

The left panel of Figure 12.8 shows the distribution of the secondary-to-primary mass ratio, a quantity conventionally symbolized as q. The observations themselves allow an unambiguous mass determination only for visual binaries and for double-lined spectroscopic binaries that also happen to be eclipsing (see below). The figure therefore incorporates both a statistical correction for the orbital inclination angle, as well as the usual extrapolation to account for undetected systems. Note that the primaries among the surveyed stars have an average mass of 1.3 $M_\odot$. Hence, the falloff at relatively high q seen in Figure 12.8 resembles the steep decline in the field-star initial mass function near this mass. We quantify the similarity by displaying, as the dashed curve, the analytic approximation to the IMF from equation (4.6). Here we recast the field-star function in terms of q by assuming a primary mass of exactly 1.3 $M_\odot$. The sequence of power laws indeed mimics the tail of the secondary mass distribution for $q \gtrsim 0.3$, but not the apparent turnover at lower q-values. On the other hand, a similar flattening *is* seen in the more detailed field-star IMF displayed in Figure 4.22. Finally, we note that the distribution of secondary masses remains essentially unchanged if we consider not the entire G-dwarf sample, but subgroups in restricted period intervals.

The observed or deduced orbits for these binaries vary from circular to highly eccentric. In general, there is a trend for systems of longer period to have larger eccentricity. Figure 12.9, which plots e as a function of P, shows that the long-period systems actually have a wide range of eccentricity. It is the *maximum* e-value that steadily increases with P. Interestingly, there is a well-defined cutoff near 11 days, below which all orbits are circular. Some physical mechanism must be at work that circularizes close binaries. We will later identify this effect as

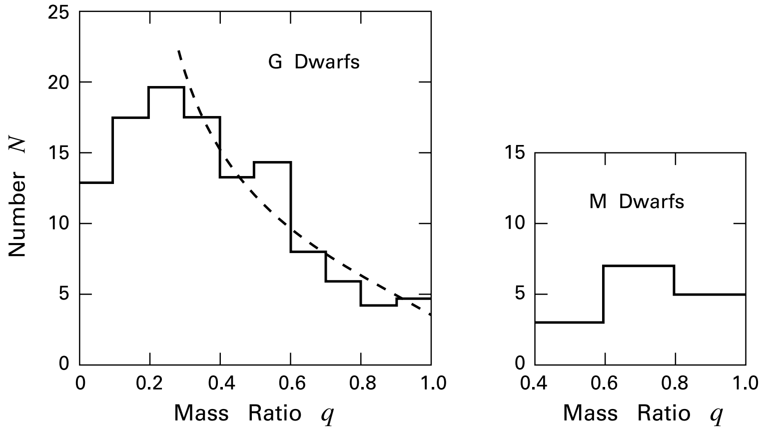


Figure 12.8 Distribution of primary-to-secondary mass ratio in G dwarfs (*left panel*) and M dwarfs (*right panel*). The dashed curve on the left represents the field-star IMF.

the tidal component of the gravitational force, which exerts an appreciable torque both during the pre-main-sequence phase and later.

About 60 percent of stars in the solar neighborhood are M dwarfs, so it is especially important to examine this population for multiplicity. A number of observational studies have searched for companions, employing techniques from visual and infrared imaging to high-resolution measurements of radial velocity. After correcting for incompleteness, the proportion of systems that are single, double, triple, and quadruple is $58 : 33 : 7 : 1$. In this case, 42 percent of observed stars have at least one companion. To understand the difference between this figure and the multiplicity of 57 percent for G dwarfs, consider that the highest mass among the surveyed M stars is only about 0.5 $M_\odot$. Any secondaries with $q \lesssim 0.2$ would be brown dwarfs that are especially hard to detect (and were not included in the above correction for incompleteness.) The true multiplicity of M stars could therefore be significantly higher, and closer to that of the G dwarfs.

Such a similarity in the populations is also suggested by the period distribution, shown by the filled circles in Figure 12.7. The points again exhibit a single, broad maximum, now lying between 10 and 200 yr. However, Figure 12.8 (*right panel*) shows that the profile of q is much flatter than for G stars. This apparent difference is again easy to reconcile. Even the highest q-values for an M-star primary correspond to secondary masses no greater than 0.5 $M_\odot$, and most of the masses are substantially lower. The relatively flat q-distribution is therefore at least broadly consistent with the low-mass turnovers seen in both the G-type binaries and in Figure 4.22. To summarize, the M- and G-dwarf studies point to the same conclusion. The masses of *both* the primaries and secondaries in binaries are apparently distributed according to the IMF.

This rule is violated in systems where the primary is an O star. If the secondary were indeed drawn randomly from the IMF, there would be very little chance of it also being an O- or even a B-type star. Yet such doubly massive pairs are not hard to find. Close examination shows that,

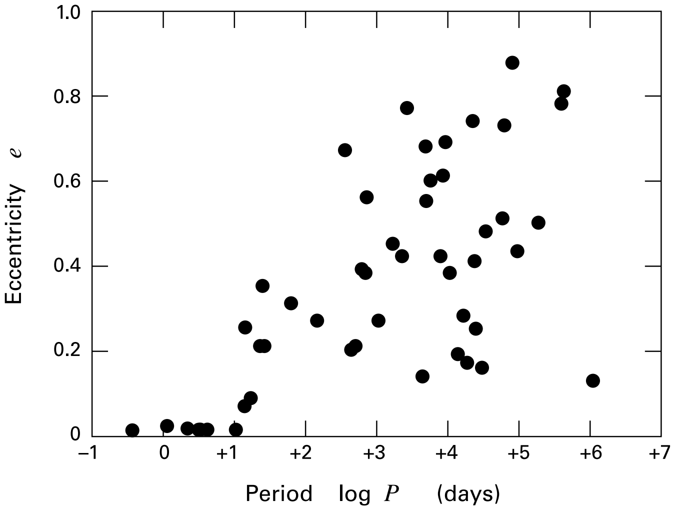


Figure 12.9 Orbital eccentricity as a function of period for G-dwarf binaries.

in many of the tighter systems amenable to spectroscopic study, the components are exchanging mass. However, in other massive pairs, the two stars remain separate. Consider S Mon, the O7 star that is the brightest member of the NGC 2244 cluster. Radial velocity measurements and interferometry reveal that a detached companion is orbiting with a period of 24 yr. Since the system is both a double-lined spectroscopic and astrometric binary, one may obtain both masses, which are 35 and 24 $M_\odot$.

Apart from the distribution of q, the binaries containing O stars resemble those of lower mass. The observed and estimated periods again span a very broad range, some 8 orders of magnitude. There is a severe lack of data for systems that are too wide for spectroscopic detection, yet too close for the components to be spatially resolved. In any case, the fraction of systems that are visible binaries appears normal, at least within OB associations. It is intriguing that binaries are uncommon among field O stars outside of associations and even less frequent among runaways. (Recall the discussion of § 4.3.) We must always bear in mind the severe difficulty of finding low-mass companions to O stars, considering both the luminosity contrast with the primary and the greater distance to these systems. It is even possible, in some cases, that both members in a well-separated pair of O stars are themselves binaries with low q-value. Sensitive infrared observations may eventually detect such hierarchical systems.

12.2.3 Imaging Methods

Let us now return to low-mass stars and explore multiplicity among the pre-main-sequence population. Occasional pairing of T Tauri stars has long been noted, ever since the initial classification of these objects in the 1940s. In undertaking a full census, one immediately faces a serious problem beyond the usual selection effects. The nearest star-forming regions are much farther than the limiting distances in the G- and M-dwarf surveys. Consider a typical binary in

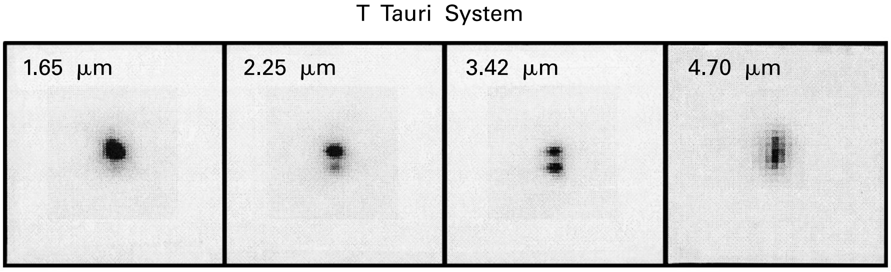


Figure 12.10 Near-infrared array observations of the T Tauri binary system. Each frame is an image at the indicated wavelength. The lower source is actually a tight binary, unresolved here.

Taurus-Auriga, composed of two solar-mass stars. If, as we shall verify, the period distribution for such systems resembles that of the dwarfs, then our example is likely to be a visual binary with a period of roughly 200 yr. From equation (12.15), $a_{\rm tot}$ would then be about 40 AU, or $0\overset{''}{.}4$ at the Taurus-Auriga distance of 150 pc. Achieving this spatial resolution is technically challenging.

Another difficulty stems from the fact that the emitted flux from most pre-main-sequence stars is highest in the near infrared. Optical instruments, while still useful, are therefore not ideally suited for systematic observations. A significant advance in this regard has been the introduction of near-infrared arrays. We saw in Chapter 4 how these detectors revolutionized the study of embedded clusters. The arrays have also led to the discovery of a large number of close companions to T Tauri stars, ranging in angular separation from about $0\overset{''}{.}7$ to $15''$. On a statistical basis alone, most of these systems must be pre-main-sequence binaries, rather than chance superpositions in the plane of the sky. Observers have been able to measure relative proper motions for some pairs; the calculated speeds appear to be consistent with those expected for gravitationally bound systems.

Figure 12.10 illustrates dramatically the power of near-infrared observations for revealing low-mass companions. The four frames all center on the star T Tau, the very prototype of the T Tauri class. Each of these images was obtained using an array detector at the wavelength indicated. The lower companion object, separated by $0\overset{''}{.}69$ or 97 AU, comes clearly into view at 3.42 μm, but is less apparent at either longer or shorter wavelengths. Additional study has revealed that this source is itself a tight double. One component is an obscured M-type star. The other, even more deeply embedded, has a spectral energy distribution that peaks in the near infrared, as in classical T Tauri stars, but no optical flux at all. Thus, the broadband spectrum does not comfortably fit into the standard classification scheme. One of these companion objects is powering an ionized wind and a molecular outflow. Careful observations have disentangled this outflow from a nearly orthogonal one driven by the upper star.

The two brightest components of the T Tauri system are barely resolvable using conventional photometry with near-infrared array detectors. Indeed, the companion source, then thought to be single, was originally found through another technique known as *speckle interferometry.* Here, one improves resolution by compensating for the blurring effect of turbulence in the Earth's atmosphere. The turbulence manifests itself as intermittent fluctuations in the refractive index along any line of sight. This "twinkling" distorts incoming wavefronts. Consequently, if one images any source over a sufficiently brief time interval, it breaks up into an irregular distribution of spots, or *speckles*. For longer integration times, typically 100 ms or greater in the near-infrared, the spots merge to produce the broader intensity pattern ordinarily encountered.

Each speckle has an angular size close to the diffraction limit of λ/D, where λ is the radiation wavelength and D the diameter of the telescope's aperture.[3] One may recover intrinsic structural features of the object at this resolution by comparing each brief exposure to that of a nearby point source, which displays a similar pattern of speckles. In practice, the observer makes hundreds of target and reference exposures in a single run, and then averages the results to produce the final image. When implemented with near-infrared arrays, speckle interferometry has allowed detection of binaries with projected separations from $0''.07$ to about $3''$. The upper bound is set by the array detector's total field of view.

Even more tightly bound systems are accessible under the right circumstances. The passage of the Moon's leading edge across a source creates a rapid falloff in flux that may be monitored using high-speed photometry. Conversely, there is a brightening associated with reappearance of the same object. In either case, the source's intrinsic luminosity distribution influences the precise character of the transition. Observers have employed this technique of *lunar occultation* in the near infrared to probe binary separations from about $0''.01$ to $1''$. Here, the measured separation is that perpendicular to the lunar limb at the point of occultation.

Figure 12.11 shows the variation in K-band flux from the pre-main-sequence star DF Tau during a reemergence event. The signal first jumps to a temporary plateau before rising again to its final, constant value. Such behavior signifies the presence of a binary system. The smooth, solid curve is the expected result for two point sources emerging 84 ms apart in time. Note the pronounced "ringing" after the final jump, a result of diffraction across the sharp lunar edge. The rate of passage of that edge across the sky was $0''.29\ \mathrm{s}^{-1}$ for this observation. Thus, the temporal delay in the companion's appearance translates to an angular separation of $0''.024$, or to a physical separation of 3.4 AU at the distance of Taurus-Auriga.

12.2.4 Spectroscopic Studies

No current technique can image binaries of much smaller separation. On the other hand, such pairs may have observable Doppler shifts in their optical spectra. Let us now see more explicitly how the changing radial velocities allow one to derive important physical properties. Figure 12.12 shows measurements of V_r from three pre-main-sequence binaries. The velocities are displayed not as a function of time, but of the orbital phase ϕ, *i. e.*, the fractional time elapsed within a single period. To effect this transformation, one first establishes P itself by examining the temporal record of velocities for periodic behavior.

Figure 12.12a shows data for the single-lined spectroscopic binary 155913-2233. The primary in this system, for which $P = 2.42$ days, is a weak-lined T Tauri star in Scorpius-Centaurus. The dashed horizontal line represents the mean V_r-value, here equal to –2.3 km s^{-1}.

---

[3] Recall the ring-like diffraction pattern of a plane wave impinging on an opaque screen with a circular hole. For hole diameter D, the angle from the central intensity peak to the first zero is $0.61\,\lambda/D$. Doubling this figure gives, by convention, the ideal, limiting resolution of a telescope with the same aperture diameter. Astronomers refer to the angular size of a point source obtained in normal photometry as the *seeing*.

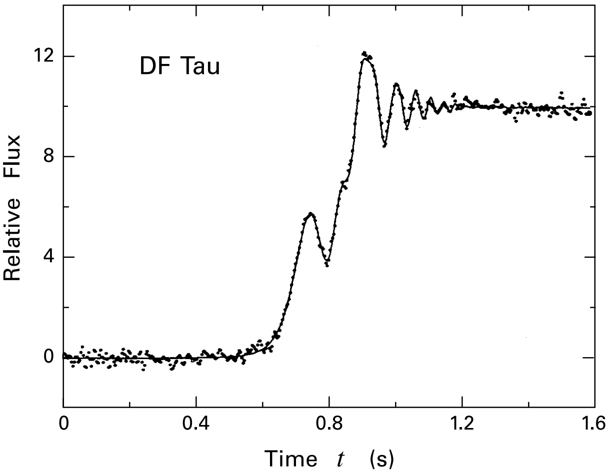


Figure 12.11 Lunar occultation of the binary DF Tau. Plotted is the temporal variation of the relative flux in the near infrared, as the system appears from behind the lunar limb. The smooth curve is a theoretical prediction for two point sources.

This quantity is the center-of-mass velocity along the line of sight and is consistent with other observed values for single stars in the region. The smooth solid curve is a theoretical fit to the phased variation of V_r. Its intrinsic shape depends on two parameters – the eccentricity e and the orientation of the primary's major axis within the orbital plane.[4] One varies both parameters to achieve the closest match to the data points. In the present case, the resulting curve is very nearly sinusoidal, corresponding to a circular orbit. Thus, the fitting procedure yields $e = 0$ to good accuracy.

This matching of theory to observation also requires that we specify the *amplitude* of the velocity change. More precisely, let K_1 be half the total range in the primary's V_r-values. This quantity is 63 km s^{-1} in our example. Knowledge of the period and velocity amplitude clearly gives information about the linear size of the orbit, *i. e.*, the semi-major axis a_1. We show in Appendix H that the *projected* value $a_1 \sin i$ obeys

$$a_1 \sin i = \frac{K_1 P}{2\pi} \left(1 - e^2\right)^{1/2} . \tag{12.17}$$

For the binary in Figure 12.12a, we find that $a_1 \sin i = 0.014\,\mathrm{AU}$.

We have not yet mentioned the stellar masses. Here one must invoke Kepler's third law, equation (12.15). If only the primary is observed, as in our example, we can deduce nothing about either M_1 or M_2 individually. Knowledge is limited to the composite quantity known as the *mass function* $f(M)$, defined as

$$f(M) \equiv \frac{M_2^3 \sin^3 i}{(M_1 + M_2)^2} . \tag{12.18}$$

---

[4] This orientation angle *within* the plane differs from the inclination i. As mentioned earlier, the latter is the angle *between* the orbital plane and that of the sky (recall Figure 12.6b).

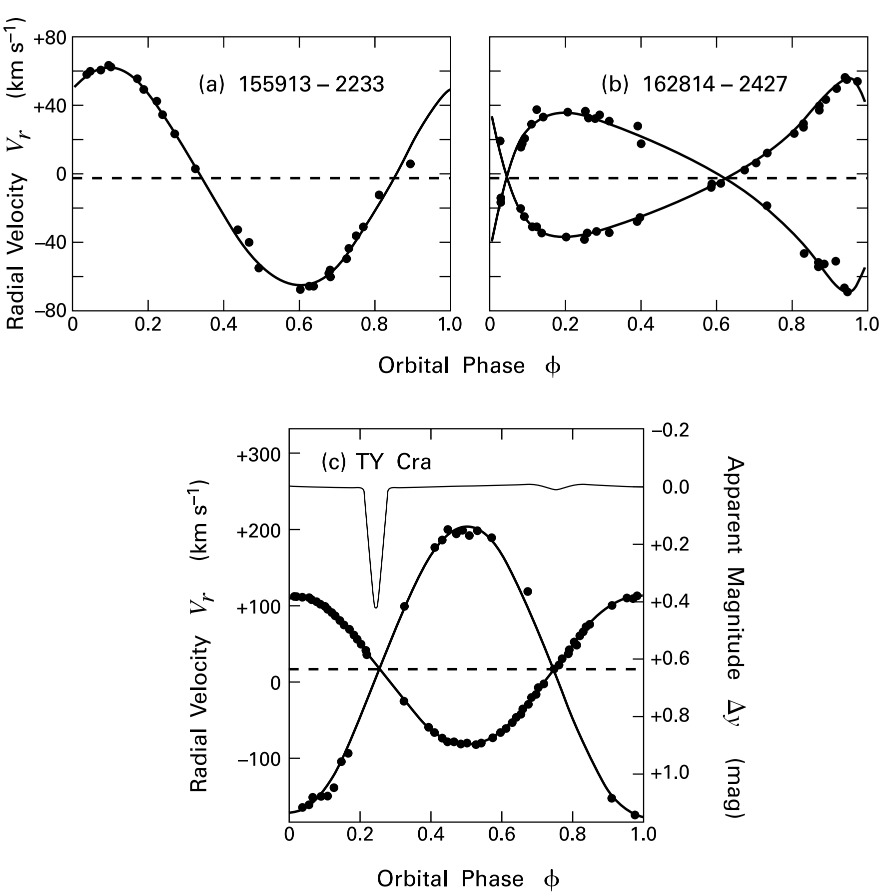


Figure 12.12 Radial velocity as a function of orbital phase for three pre-main-sequence binaries. The solid curves are the theoretical results for binaries of various eccentricities and orientations. The lower panel also shows the temporal variation of broadband, visual flux, as measured in the y-band.

Appendix H also derives an expression for this function in terms of known quantities:

$$f(M) = \frac{K_1^3 P}{2\pi G}\left(1 - e^2\right)^{3/2} . \tag{12.19}$$

Substitution of our present values gives $f(M) = 0.064\, M_\odot$.

More information is naturally available when there is also a spectrum for the secondary star. Figure 12.12b shows the double-lined spectroscopic binary 162814-2427 in ρ Ophiuchi. First detected through its X-ray emission, this pair consists of two weak-lined T Tauri stars orbiting each other with a period of 36.0 days. The phased radial-velocity points now fall along curves that are far from sinusoidal, so that a substantial eccentricity is present. The fitting procedure yields $e = 0.48$. Note that the center-of-mass velocity is –6.1 km s^{-1} in this case.

Each velocity curve has its own amplitude, here measured to be $K_1 = 44\,\mathrm{km\,s^{-1}}$ and $K_2 = 48\,\mathrm{km\,s^{-1}}$. Application of equation (12.17) to both orbits gives the projected semi-major axes as $a_1 \sin i = 0.128\,\mathrm{AU}$ and $a_2 \sin i = 0.139\,\mathrm{AU}$. These results make it clear that the primary and secondary masses are nearly equal, but ignorance of the inclination angle still hampers us. We may recast equation (12.15) as

$$M_1 \sin^3 i + M_2 \sin^3 i = \frac{4\,\pi^2}{G\,P^2}\,(a_1 \sin i + a_2 \sin i)^3 \ , \tag{12.20}$$

while equation (12.14) implies

$$\frac{M_1 \sin^3 i}{M_2 \sin^3 i} = \frac{K_2}{K_1} \ . \tag{12.21}$$

Solving these two relations simultaneously for $M_1 \sin^3 i$ and $M_2 \sin^3 i$ gives 1.02 $M_\odot$ and 0.94 $M_\odot$, respectively. Thus, we may obtain only *lower bounds* on the component masses.

These constraints are replaced by true mass determinations only when there is some independent method for fixing the inclination angle. The observation of eclipses in double-lined spectroscopic binaries provides such an opportunity. However, the orbital plane must be nearly edge-on for this effect to occur. Figure 12.12c shows one such fortunate case, the binary TY CrA. The primary here is a Herbig Be star in the Corona Australis dark cloud complex. This intermediate-mass object is one of the brightest members of a young cluster located at one end of the cloud (recall Figure 4.7).

With its period of 2.89 days, the TY CrA binary is another tight system that would escape direct imaging. The radial-velocity curves are closely sinusoidal, so that the orbital eccentricity is close to zero. The measured amplitudes, $K_1 = 85\,\mathrm{km\,s^{-1}}$ and $K_2 = 165\,\mathrm{km\,s^{-1}}$, now indicate that the component masses differ by almost a factor of two. Application of equations (12.17), (12.20), and (12.21) yields $M_1 \sin^3 i = 3.08\,M_\odot$, $M_2 \sin^3 i = 1.59\,M_\odot$, $a_1 \sin i = 4.86\,R_\odot$, and $a_2 \sin i = 9.38\,R_\odot$. Note again the very small values of the projected semi-major axes. Careful spectroscopy has revealed the presence of yet a third star orbiting at much greater distance, roughly 1 AU. This tertiary acts as a small perturbation on the inner binary and may be ignored for our purposes.

Also displayed in Figure 12.12c is a record of the broadband flux from the system. Here the apparent magnitude is measured in y, the visual band at 5500 Å in the Strömgren photometric sequence. The sharp plunge at the phase $\phi = 0.25$ marks the *primary eclipse*. This decrease occurs when the primary is directly behind its dimmer companion. Half a period later, at $\phi = 0.75$, the primary blocks the other star entirely. The flux decrease at this *secondary eclipse* is much shallower. Note also the slight rise in total flux just before and after this point. The distance between the stars is small enough that the primary's radiation significantly heats a portion of the secondary's surface. This exposed area comes into view just before the secondary is occulted, as well as a short time later, accounting for the temporary increases in flux. Careful modeling of the entire light curve gives the individual stellar radii, luminosities, and effective temperatures. Most importantly, this modeling also yields the inclination angle i, which is here 83°. Thus, we finally derive $M_1 = 3.16\,M_\odot$ and $M_2 = 1.64\,M_\odot$.

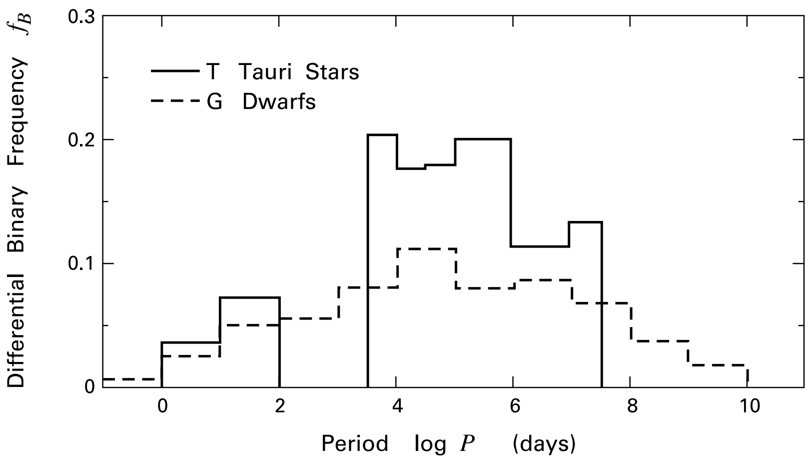


Figure 12.13 Period distribution of pre-main-sequence binaries (*solid histogram*). The dashed histogram, which is reproduced from Figure 12.7, shows the G-dwarf result for comparison.

12.2.5 Periods and Eccentricities

After collating the data supplied by spectroscopy and the various imaging techniques, observers have built up a census of the binary population among pre-main-sequence stars. The most complete information is available from the Taurus-Auriga and ρ Ophiuchi regions, but the surveys are rapidly expanding in scope. Figure 12.13 summarizes the presently known differential binary frequency. The systems included here are only those with a T Tauri star as the primary component. Thus the figure omits more massive binaries such as TY CrA. Also shown, for reference, is the distribution of f_B for main-sequence G stars.

Comparing the two groups, it is clear that T Tauri binaries cover a similarly broad range of periods. Note the apparent gap between 100 days and 10 yr, an interval that lies between the spectroscopic binaries and those amenable to direct imaging. The greatest fraction of pre-main-sequence binaries is again between 10^2 and 10^3 yr. A striking aspect of Figure 12.13 is that the overall frequency is *higher* than for G dwarfs. We should view any direct comparison of optical and infrared studies with some caution, taking into account such factors as the observational detection limits. Nevertheless, one possibility suggested by these data is that the binary population in a star-forming region may diminish with time. If so, it is not at all clear what might cause the dissolution of these systems. The alternative view is that the regions contributing to Figure 12.13 have a different binary population than the ones most heavily represented in the field.

The evolutionary influence on binaries should gradually become clearer as we continue to probe the detailed nature of the systems being found. For those discovered through imaging, it is often problematic to disentangle the individual contributions to the flux at various wavelengths. Even if this separation can be done, one must also observe the two stars spectroscopically to obtain accurate effective temperatures. In summary, there are not a great many instances where we can confidently place both components in an HR diagram. Mass determinations therefore

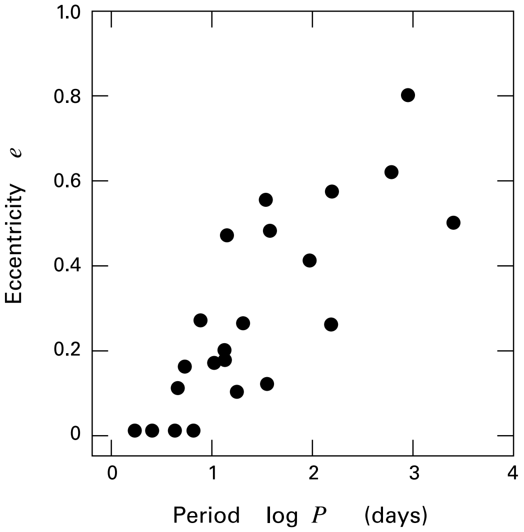


Figure 12.14 Orbital eccentricity as a function of period for pre-main-sequence binaries.

largely stem from the relatively small number of spectroscopic binaries. This same group provides the few dozen orbital eccentricities currently available.

Figure 12.14 displays these eccentricities as a function of period, again for T Tauri binaries. It is immediately clear that the wider systems of longer period tend to have higher e-values. This trend is strikingly similar to that among the G dwarfs (Figure 12.9). We again note the appearance of a lower cutoff in period, below which the orbits are circular. In this group of stars, whose ages cluster near 10^6 yr, the value appears to be about 4 days. It is interesting that the analogous figure for the Hyades cluster, with an age of 6×10^8 yr, is 8 days. Recall from Figure 12.9 that the cutoff was 11 days for our sample of main-sequence G stars and is 19 days for older binaries in the Galactic halo. We conclude that the efficacy of tidal torquing for circularizing orbits must grow steadily with time.

12.2.6 Infrared Companions and Protobinaries

In extending our search to the very youngest binaries, it is natural to focus on objects with the greatest infrared emission. We already noted that one star in the T Tau system has no optical component to its flux. Such pairing of a visible pre-main-sequence star with a highly embedded (specifically, infrared) one is relatively uncommon but not unique. Observers have found about a dozen such cases, comprising perhaps 10 percent of the T Tauri binary population. The nature of the embedded stars, termed *infrared companions*, is still quite uncertain.

On the one hand, there is little doubt that the heavy obscuration stems from a large column density of dust. The presence of the 10-µm silicate absorption feature in the T Tau example bolsters this view. On the other hand, the infrared emission also exhibits pronounced variability, even in the course of several years. The sharp gradient in extinction from the infrared companion to the lightly obscured T Tauri star further emphasizes the limited *spatial* extent of the intervening matter. In summary, there is little indication that the object in question has a younger age, only a very different circumstellar environment.

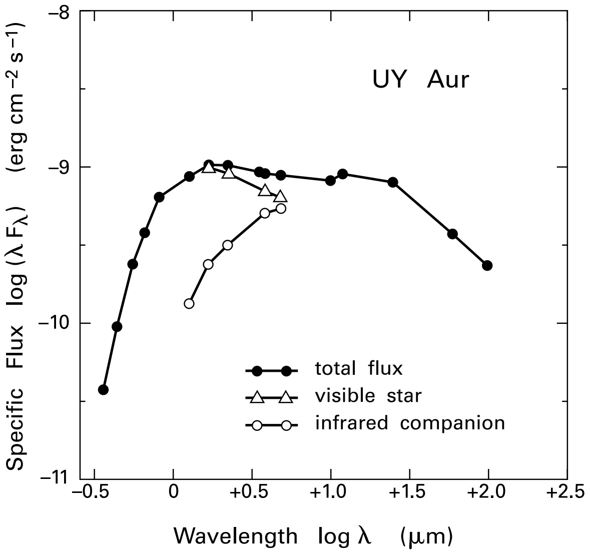


Figure 12.15 Spectral energy distribution for the binary UY Aurigae. In addition to the total flux, the figure also shows individual contributions from the visible star and the infrared companion.

Infrared companions have similar bolometric luminosities as their optically visible neighbors and provide most of the system's flux at wavelengths beyond 10 μm. Figure 12.15 shows the spectral energy distribution of the binary UY Aurigae. Here one may separate the individual fluxes through speckle interferometry. It is interesting to note that the current infrared companion was an optically visible object in 1944, when it appeared in the course of an early T Tauri survey. At present, the flux from this star rises to a broad maximum beyond 10 μm. Following our discussion of Class 0 and I sources, the breadth of the spectral energy distribution points to significant heating and reemission by the dust grains. Much of this material may lie within a disk that orbits this star. There is also direct evidence for a much larger disk surrounding the entire binary system, as we shall describe in Chapter 17.

Binaries in which *both* components are heavily embedded may well represent true precursors to the systems we have discussed so far. Radio and infrared studies conducted since the 1980s have begun to find examples. The total number of systems is still too low for any general characterization of their properties. In no instance, moreover, has it been possible to demonstrate convincingly gravitational binding of the stars, either through their proper motions or radial velocities. We shall nevertheless apply the term *protobinary* to any pair of visually obscured stars that appears likely to be bound.

One search strategy for protobinaries has been to examine the driving sources of molecular outflows. These extended lobes of moving gas appear to diverge from compact regions that are luminous in the infrared. Closer scrutiny of these locations with array detectors frequently reveals tight clusters of point sources. Occasionally the outflow stems from a relatively isolated stellar pair.

A case in point in IRAS 16293-2422 in the ρ Ophiuchi region. This source is located at the center of a complex and powerful outflow system, consisting of two pairs of red- and blue-shifted lobes. The flow velocities within the lobes are on the high end for molecular outflows, indicating that the driving star or stars are especially young. Infrared and submillimeter measurements of the central region confirm this suspicion. Despite a total luminosity exceeding $30\ L_\odot$, the source is undetectable for wavelengths less than 25 μm, and the spectral energy distribution fits rather neatly into the Class 0 category.

The binary nature of the system only becomes evident with higher spatial resolution. Interferometric maps at centimeter and millimeter wavelengths reveal two compact sources, separated by 840 AU in projection. The centimeter radiation probably stems from a wind-induced shock (Chapter 13). Winds might be expected here if each source drives a pair of the outflow lobes. Surrounding one of the stars is a tight grouping of H_2O masers, yet another sign of wind activity. Now the emitted flux in the millimeter continuum, like that in the infrared, originates from warm dust grains. The associated masses in circumstellar gas are a bit under $1\ M_\odot$ for each component. These figures represent, of course, only that portion of gas heated by the stars. There is, in addition, a larger parent dense core or slab, seen by its molecular line emission. Further study should elucidate the morphology of this object, as well as the linkage of each star to the molecular outflows extending well beyond the densest gas.

12.3 The Origin of Binaries

The available information on deeply embedded pairs is still too scant to justify any claim that they represent the precursors to all binaries. While the observational situation continues to improve, we may also look to the more exposed, pre-main-sequence systems for clues as to their origin. Later in this section, we will see what theoretical ideas concerning cloud evolution are likely to be relevant.

12.3.1 Ages of the Components

Any measure of stellar *ages* is clearly significant with regard to the issue of origins. If these ages are nearly the same for both components within a binary, then the two must have formed simultaneously as a gravitationally bound unit. Conversely, widely discrepant ages would imply that some kind of capture process joined two stars born at very different times and therefore presumably in different spatial locations. One might imagine such events occurring within a large parent cluster. However, the observational data has rendered this scenario unlikely, as we shall now illustrate through specific examples.

We begin with spectroscopic binaries. Obtaining the age of any young star requires placing it in the HR diagram. We thus need to determine the effective temperature and bolometric luminosity of each component. For eclipsing, double-lined systems, we have seen that reliable values for $T_{\rm eff}$ and $L_{\rm bol}$ follow from careful modeling of the composite light curve. After locating each star in the diagram, we may read off both its age and *mass* from the appropriate pre-main-sequence track. But we have also seen how both masses follow from analysis of the velocity curves and knowledge of the inclination angle i. There is no reason *a priori* why these *spectroscopic* values for M_1 and M_2 should match the *photometric* ones found from the pre-main-sequence tracks. Eclipsing binaries therefore present very favorable, and unfortunately very rare, opportunities both for measuring stellar ages and for gauging the accuracy of the evolutionary tracks themselves.

The upper panel of Figure 12.16 shows, within the $L_{\rm bol}$–$T_{\rm eff}$ plane, the primary and secondary of TY CrA, the eclipsing binary previously depicted in Figure 12.12c. In the current representation, the primary appears slightly below the ZAMS, near a mass of 3.0 $M_\odot$. The latter value compares reasonably well with the more precise spectroscopic determination of 3.16 $M_\odot$. The secondary, with a spectroscopic mass of 1.64 $M_\odot$, indeed falls just above the 1.5 $M_\odot$ evolutionary track. Thus, the two independent methods for obtaining the component masses are in essential agreement.

What of the stellar ages? Unfortunately, the location of the primary so close to the ZAMS makes it impossible to assign a reliable contraction age. However, the star illuminates a bright reflection nebula in the Corona Australis cloud, and the system's spectral energy distribution rises in the mid-infrared regime toward longer wavelengths. Thus, the primary is undeniably young, a typical Herbig Be star situated within both interstellar and circumstellar gas and dust. Suppose we tentatively assign it the pre-main-sequence age for a 3 $M_\odot$ star just arriving on the ZAMS, or 2×10^6 yr. Then the corresponding isochrone, shown as the dashed curve, passes close to the secondary's position. The components of TY CrA are therefore plausibly coeval, although the data do not permit a more quantitative assessment.

Age assignments are more trustworthy when both stars lie well above the ZAMS. A handful of eclipsing binaries of this kind are known, but none with secure properties. However, there are a number of non-eclipsing, double-lined systems. The lower panel of Figure 12.16 shows one of these, the weak-lined T Tauri star V773 Tau and its companion. In the absence of eclipses, one must obtain $T_{\rm eff}$ and $L_{\rm bol}$ for each star by careful analysis of both the composite spectral energy distribution and the shifting, narrow-band spectrum. One then finds that the components of V773 Tau have roughly the same age, which again happens to be about 2×10^6 yr. The photometric masses for the primary and secondary in this case are $M_1 = 1.7\ M_\odot$ and $M_2 = 1.2\ M_\odot$. We may compare their ratio $M_1/M_2 = 1.4$ to the value of 1.32 obtained spectroscopically, through application of equation (12.15). The similarity of these two estimates is gratifying and further bolsters our confidence in using the tracks to obtain both ages and masses.

Discerning the relative luminosities from an unresolved pair of stars is no simple matter, and the foregoing results are subject to some uncertainty. The situation further improves when we consider wider, spatially resolved binaries. In this case, however, we have no spectroscopic mass ratios for comparison. Figure 12.17 shows three systems taken from Taurus-Auriga and Orion. For two of them, the components lie close to theoretical isochrones, while the ages within WSB 18 apparently differ by about a factor of three. The sampling here is representative in a statistical sense. That is, roughly, a third of visual, pre-main-sequence binaries have component ages that do not match.

Some binaries in this category are proving, under closer scrutiny, to be triples or even quadruples. The nature of the remaining pairs with discrepant ages is still unclear. In any case, we may be confident that most binaries consist of stars that were born nearly simultaneously. No such coincidence would be expected from the random coupling of stars drawn from

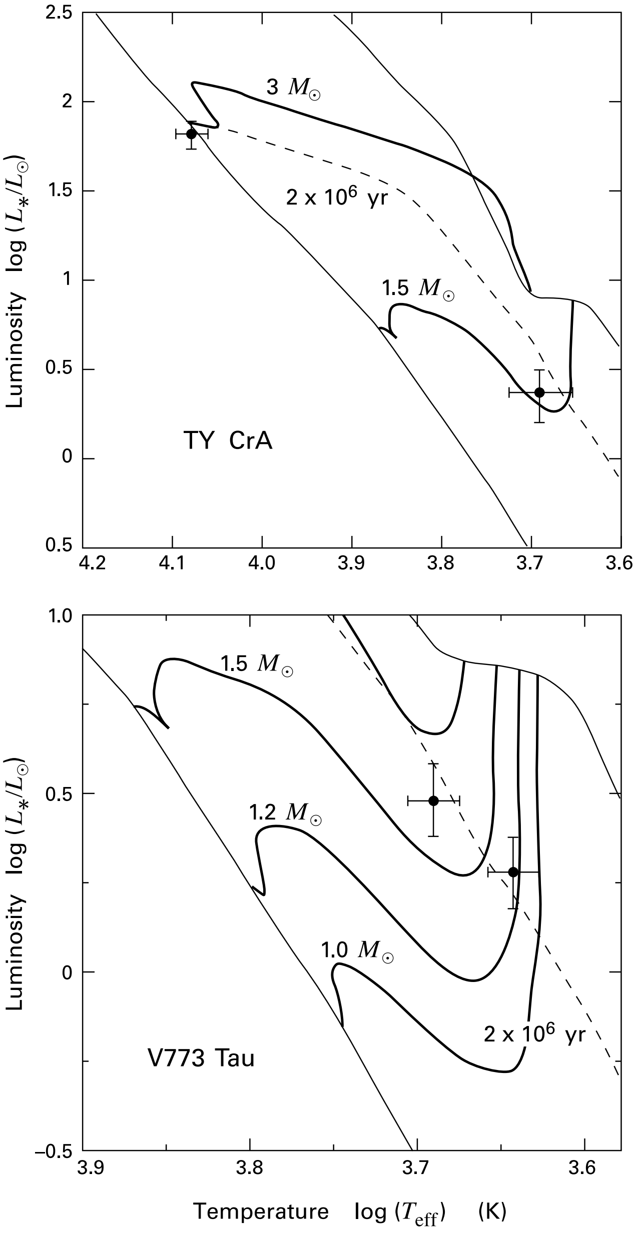


Figure 12.16 Placement in the HR diagram of the components in the spectroscopic binaries (a) TY CrA and (b) V773 Tau. The birthline, ZAMS, and selected evolutionary tracks are shown. The dashed curve is the theoretical isochrone corresponding to 2×10^6 yr.

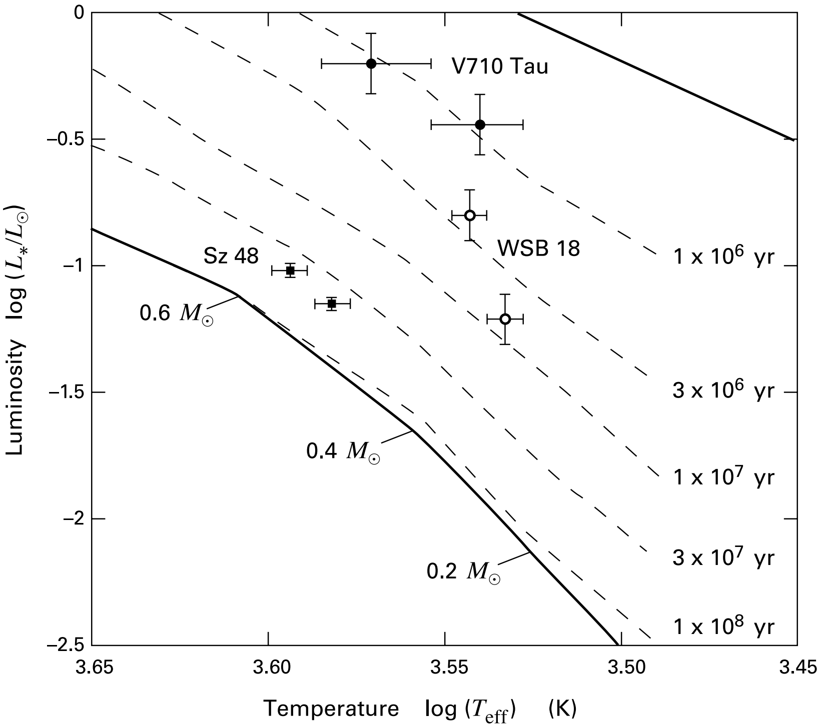


Figure 12.17 Placement in the HR diagram of the components in three visual, pre-main-sequence binaries. The ZAMS, birthline, and isochrones are shown, as well as the ZAMS positions for the indicated masses.

an extended cluster. The typical binary must therefore arise *in situ*, rather than through capture. Note, however, that a capture mechanism may still apply to binaries with truly differing ages, as well as to pairs of *massive* stars, as we shall discuss in § 12.5 below.

12.3.2 The Fission Hypothesis

What, then, is the actual formation mechanism? Figure 12.17 demonstrates how young binaries are found throughout the HR diagram, from the birthline down to the main sequence. The establishment of a bound pair must therefore occur not in the pre-main-sequence epoch, but in the earlier phase of protostellar infall. Could it be that the components were originally a single protostar that somehow split apart? One might imagine, for example, that the object gains so much angular momentum from infalling matter that its rapid rotation leads to gross deformation, followed by separation into distinct entities. This suggestion is a variant of the *fission hypothesis*, a venerable idea that was prominent for many decades before modern binary research.

It is still not feasible to measure the surface velocities for highly embedded stars, so one cannot dismiss out of hand the possibility of their rapid rotation. The viability of the fission hypothesis then rests on the expected physical behavior of stars endowed with increasing angular momentum. Advocates of fission have drawn inspiration from classic studies of incompressible fluid configurations. Such rotating liquid masses indeed deform bodily as they spin up.[5] Stars, on the other hand, are highly compressible objects with a large variation of internal density. Consequently, they undergo a very different kind of transition.

As we increase the angular momentum of a star, the most rapidly rotating fluid elements experience a centrifugal repulsion that rivals their gravitational attraction to the interior mass. The equatorial region then begins to detach. Figure 12.18 displays snapshots from a time-dependent numerical simulation. Here, the object was set immediately into rapid rotation at $t = 0$. Since angular momentum is conserved locally, the body becomes differentially rotating as it evolves. The panels show density contours in the equatorial plane, and the elapsed time is measured relative to the initial period along the central axis. We see that the object quickly sheds gas into two, trailing spiral arms. In calculations covering longer times, the ejecta form a detached ring around the remnant central core.

The ring has a relatively small amount of mass, but contains most of the system's angular momentum. This lopsided distribution is a result of efficient gravitational torquing. Equatorial shedding thus prevents the bulk deformation necessary for fission. In fact, it is doubtful that a protostar ever spins up enough for this shedding to begin. Even the youngest visible pre-main-sequence stars have rotation speeds well below breakup, probably as a result of magnetic winds. There is no reason why such winds should not also operate during the protostar phase. A portion of the infalling angular momentum would be redirected into the wind, while much of the rest would be incorporated into the growing protostellar disk.

Could binary companions result from the breakup of these disks? The situation here is analogous to that regarding stellar fission. Numerical calculations starting with highly unstable circumstellar disks do find them breaking apart into numerous fragments, each of a size comparable to the local disk thickness. Some of these pieces may subsequently merge, but there is no indication of their further collapse to stellar density. Moreover, it is unlikely that the assumed initial conditions ever apply. As we saw in Chapter 11, protostellar disks generate internal spiral waves as they become gravitationally unstable. The torquing from these waves probably transfers enough mass inward to suppress rapid growth of the instability.

12.3.3 Quasi-Static Fragmentation

One important clue to the origin of binaries is the distribution of secondary masses. We have seen how the current data, while incomplete, are consistent with the supposition that both the primary and secondary are drawn from the field star initial mass function. The implication is that the components form *independently* but in spatial proximity. The widest visual binaries have separations of order 0.1 pc, and the vast majority are considerably tighter. We should therefore explore the possibility that both protostars arise within a single dense core.

---

[5] Consider a sequence of uniform-density, self-gravitating equilibria that are in uniform rotation. Each member of the sequence may be characterized by the single, nondimensional parameter $\mathcal{T}/|\mathcal{W}|$, introduced in Chapter 9. For low values of this parameter, the stable configurations are the oblate Maclaurin spheroids. For $\mathcal{T}/|\mathcal{W}| > 0.27$, the equilibria are dynamically unstable to disturbances that transform them into tumbling, prolate configurations.

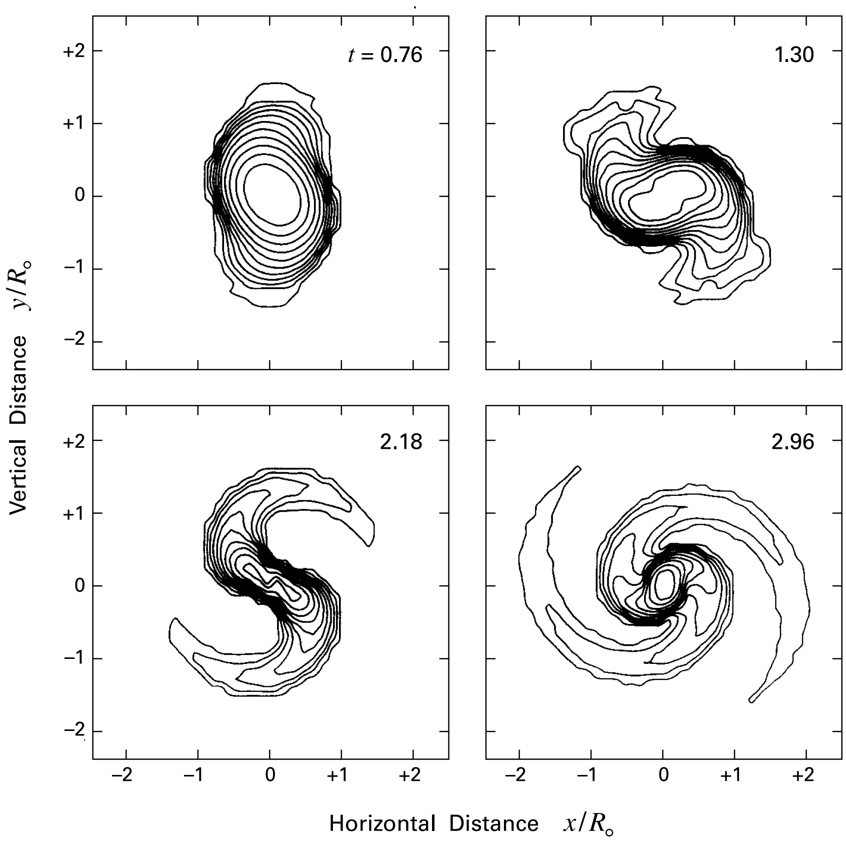


Figure 12.18 Breakup of a rapidly rotating cloud. Shown are density contours in the equatorial plane. Times are measured relative to the initial rotation period along the cloud axis.

Our previous treatment of dynamical fragmentation showed how a self-gravitating cloud can indeed break apart, but only if its mass exceeds M_J or its rotational analogue. We noted further that such bodies cannot themselves be equilibria. Since their internal forces are unbalanced, they immediately begin to collapse. While this property makes them convenient for theorists studying collapse dynamics, it also renders these objects implausible models of dense cores. To explain the existence of cores not yet containing stars, a phase of mass accumulation must occur *prior* to infall. The collapse event terminates the growth of these configurations, each of which is in dynamical equilibrium.

As we shift our focus to structures closer to force balance, the very concept of fragmentation changes. We are no longer concerned with prompt breakup accompanying free-fall collapse. Rather, we want to explore the more gradual emergence of separate density peaks within a slowly contracting parent body. Each peak evolves independently of its neighbor and eventually accumulates enough mass to undergo dynamical infall. We dub this entire process *quasi-static fragmentation*. While the idea is motivated by theoretical considerations, it is also consistent with the higher-resolution maps of dense cores, which display a lumpy substructure. Still lacking, however, is a detailed quantitative treatment. Hence we will need to base our discussion on plausible extensions of known results.

The new emphasis on equilibrium starting states makes the long tradition of collapse simulations less directly applicable. On the other hand, these still can supply valuable insight. Recall, for example, that uniform-density, rotating spheres only fragment if their nondimensional masses exceed $m_{\rm crit}(\beta)$, as depicted in Figure 12.4. Lower-mass structures rebound as stable equilibria, while objects near the critical condition evolve to higher density without breaking apart. Other collapse studies have demonstrated that spheres with substantial density contrasts, such as those encountered in true equilibria, remain intact. Thus, even the simplest theoretical description of quasi-static fragmentation must depart from spherically symmetric initial conditions.

12.3.4 Elongated Clouds

This last point may strike the reader as rather academic, since observed dense cores are far from perfect spheres. The real message is that the spatial configuration of a cloud plays a crucial role in its subsequent evolution. Having accepted that this configuration is nonspherical, we must also look beyond thermal pressure as the sole supporting mechanism. It is natural to consider *magnetized* structures. These, we have noted, may be flattened or elongated relative to the ambient field direction. We restrict ourselves to the latter case, since it is more favored by observations.

Theoretical study of elongated cloud equilibria has thus far been limited to the simplest case of infinite cylinders. Consider first non-magnetic, isothermal cylinders embedded within a background medium of pressure $P_\circ$. As in the spherical and oblate configurations we studied, these may be characterized by their density contrast from center to edge, $\rho_c/\rho_\circ$. Employing G, $P_\circ$, and the sound speed a_T, we may define a nondimensional cylinder radius $\varpi_\circ$ in terms of the fiducial length $(P_\circ\, G)^{1/2}/a_T^2$. Solution of the equations for hydrostatic equilibrium then shows that $\varpi_\circ$ increases from 0 at $\rho_c/\rho_\circ = 1$, reaches a maximum of $1/\sqrt{2} = 0.71$ at $\rho_c/\rho_\circ = 4$, and then gradually falls to zero at large density contrast. This behavior is reminiscent of the other types of equilibria. Structures with small $\rho_c/\rho_\circ$ are confined by the external pressure, while denser ones constrict under the influence of self-gravity.

Consider next the dynamical stability of a cylindrical cloud. Following Jeans, we may propagate a sinusoidal traveling wave down the configuration and ask for the minimum wavelength at which the wave amplitude grows exponentially. Numerical calculations show that the nondimensional wavelength, which we shall denote as $\lambda_{\rm cyl}$, again starts at zero for $\rho_c/\rho_\circ = 1$, reaches a value of order unity at modest density contrast, and thereafter declines. Note that the instability in lower-density models does not stem from bulk contraction; it is not present, for example, in isothermal spheres, which are dynamically stable in that regime. Destabilization of the low-density cylinders is caused by surface warping associated with the wave propagation.

Finally, let us introduce a uniform magnetic field $\boldsymbol{B}_\circ$ lying along the axial direction. The structure of the equilibrium models is unchanged, since a uniform field exerts no force. If we assume flux freezing, however, then the field is bent by passage of the wave. Nevertheless, we find that $\lambda_{\rm cyl}$ is virtually unchanged for models with $\rho_c/\rho_\circ \gtrsim 4$. For the cylinder with maximum radius, $\lambda_{\rm cyl}/\varpi_\circ$ increases from 3.7 in the non-magnetic case to 4.0 for an infinitely strong field. The reason for this insensitivity is that the waves cause the gas to clump along the axis, a motion that cannot be opposed by a purely longitudinal field. In contrast, the low-density cylinders confined by pressure exhibit a dramatic increase in $\lambda_{\rm cyl}$ even for modest field strengths. Here, magnetic tension effectively prevents the surface corrugation and buckling.

We have remarked that the real utility of the Jeans length or its analogue is not for diagnosing stability in fictitious infinite configurations, but for delimiting the spatial extent of bounded ones. In this case, we see that a magnetic field can support a long column, provided that the structure has relatively low density contrast. This result is broadly consistent with the observations of filamentary clouds, at least in those cases where the magnetic field appears to be axial. Moreover, the analysis indicates that any embedded, denser structures must have smaller aspect ratios. But how would any of these entities evolve with time?

The gas within magnetized clouds drifts slowly across the embedded field through ambipolar diffusion. This process sets the basic time scale for quasi-static contraction. We saw in Chapter 10 how initially flattened structures proceed in this manner to the point of gravitational collapse. For that geometry, the density maxima began and remained at the cloud center. The question now is whether two or more off-center peaks could arise in an elongated structure.

No studies have addressed this point directly, so we must turn to analogous, non-magnetic calculations involving dynamical collapse. We saw earlier how clouds of modest elongation but containing many Jeans masses constrict in the transverse direction to narrow spindles that subsequently break up along their major axes (recall Figure 12.5). At the other extreme are bodies of sufficiently low mass that they are only weakly self-gravitating. These simply deform into pressure-confined spheres. Thus a fluid element near one of the ends mainly experiences longitudinal motion either toward or away from the center. The case of most interest is the intermediate one, where both transverse and longitudinal motion occur.

Numerical simulations in this regime find a variety of outcomes, depending on the precise initial conditions. Here we restrict our attention to uniform-density objects, which are not predisposed to fragmentation by virtue of their initial structure. Thus, a collapsing cylinder promptly forms two subcondensations near its ends. These gather surrounding matter and fall toward one another. For a large enough total cloud mass, their local density exhibits runaway growth before the merger event. A tapered configuration of the same mass, such as a prolate ellipsoid, does not develop two separate condensation points during infall, but only one at the center. Here, a rudimentary spindle may emerge, but becomes spherical as the density climbs. Starting with a larger cloud mass leads to a more robust spindle. This later fragments, when pressure begins to retard axial constriction. Figure 12.19 illustrates the two basic collapse modes through snapshots from numerical calculations. Both studies began with identical prolate ellipsoids of 2:1 aspect ratio that just fit inside the cylindrical borders. The clouds had different masses, as indicated, and soon diverged markedly in their condensation.

The quasi-static evolution of magnetized equilibria plausibly follows similar routes. Suppose that the embedded magnetic field is tied to a uniform background, as sketched in Figure 10.8. In situations where the internal pressure is strong, ambipolar diffusion acts to straighten the field and relieve internal magnetic tension. A cloud of relatively low mass will thus relax to a sphere threaded by a uniform field. The single-star outcome in Figure 10.8 would occur with a starting mass that is greater, but too small for a spindle to be maintained. Finally,

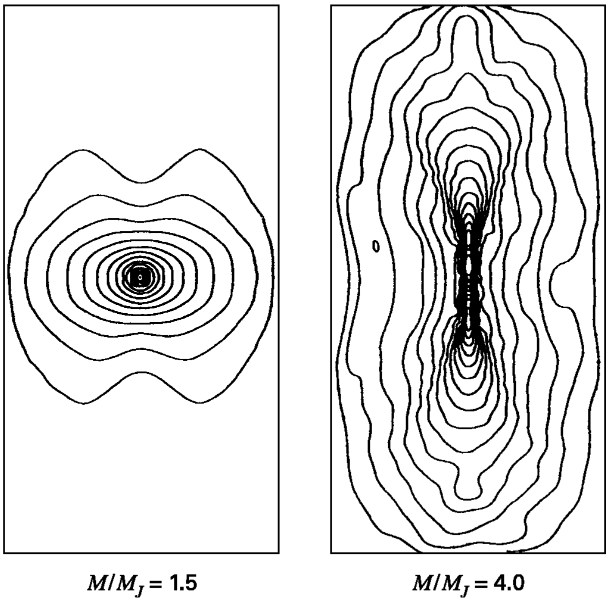


Figure 12.19 The two collapse modes for elongated clouds. The rectangular borders of each panel are the projections of the cylinders which just encompassed the initial, ellipsoidal configurations. Displayed are the density contours at similar times after the start of collapse, about $1.1\ t_{\rm ff}$.

a sufficiently massive cloud should again constrict along its axis and develop two or more high-density centers that eventually undergo collapse. The nascent stars will slowly drift together, over the diffusive time scale. Note that a direct merger can easily be averted if the cloud has even a modest degree of rotation.

12.3.5 Influence of Disks

Any primitive binary formed in this manner should be a highly eccentric system with a maximum separation exceeding that of most optically visible pairs. Thus, the orbital energy must somehow diminish. The total angular momentum, $J_{\rm tot}$, will also be too large initially, but not by a huge factor. If the original dense core of mass $M_{\rm tot}$ rotates with angular speed $\Omega_{\rm cloud}$, then

$$\begin{aligned} J_{\rm tot} &\approx \frac{2}{5}\, M_{\rm tot}\, R_{\rm tot}^2\, \Omega_{\rm cloud} \\ &= 2\times 10^{54}\ {\rm g\ cm^2\ s^{-1}} \left(\frac{M_{\rm tot}}{1\ M_\odot}\right) \left(\frac{R_{\rm cloud}}{0.1\ {\rm pc}}\right)^2 \left(\frac{\Omega_{\rm cloud}}{1\ {\rm km\ s^{-1}\ pc^{-1}}}\right) . \end{aligned} \qquad (12.22)$$

Here we have crudely modeled the cloud as a sphere of radius $R_{\rm cloud}$, and have taken the representative angular speed from Chapter 3. Suppose that this cloud eventually becomes a circular binary of period P, consisting of two identical stars with mass $M_{\rm tot}/2$. The angular momentum is now

$$J_{\rm tot} \,=\, \frac{\pi}{2}\,\frac{M_{\rm tot}\,a_{\rm tot}^2}{P}\ . \tag{12.23}$$

Substituting from (12.15) for $a_{\rm tot}$, we find

$$\begin{aligned} J_{\rm tot} \,&=\, \left(\frac{P}{128\,\pi}\right)^{1/3} M_{\rm tot}\,(G\,M_{\rm tot})^{2/3} \\ &=\, 1\times 10^{53}\ {\rm g\ cm^2\ s^{-1}} \left(\frac{M_{\rm tot}}{1\ M_\odot}\right)^{5/3} \left(\frac{P}{200\ {\rm yr}}\right)^{1/3}\ . \end{aligned} \tag{12.24}$$

Note that our numerical expression has utilized the mean observed period for G-dwarf binaries. Comparison of equations (12.22) and (12.24) indicates that $J_{\rm tot}$ must diminish by an order of magnitude. The orbital energy must decrease even more to account for the closest binaries observed.

Both the angular momentum and energy will indeed fall as the binary interacts with remnant gas. Consider, for example, the effect of circumstellar disks. It is plausible that the two stars approach each other to within a distance comparable to their disk radii. If so, material that was orbiting outside this periastron separation will be stripped away. Figure 12.20 shows that even smaller disks are greatly perturbed. Here, a diskless companion star follows the indicated parabolic trajectory about the star at the center of the frame. The central object's disk radius was initially 0.8 times the periastron distance. At the time shown, when the companion has just crossed the righthand border, over half the disk has been disrupted. Some of this material is ejected into the broad, expanding tail, while some is captured by the companion to form a new disk.

For this dispersal of circumstellar gas to affect the orbit significantly, two circumstances must hold. First, the disk should rotate in the same sense as the companion's relative velocity, in order to maximize the interaction. Such prograde motion is likely to occur in a forming binary. Second, the energy loss will be minor unless at least one of the disks has a mass comparable to that of its parent star. The observed disks around *pre-main-sequence* stars comprise only a small fraction of the stellar mass (Chapter 17). However, we have seen that theory does posit more massive structures, at least while the central object is in the protostar phase most relevant here. That is, any star-disk encounter actually occurs within a dense, collapsing envelope.

The continuing infall will preferentially collect near the more massive star, either directly onto its surface or in orbit around it. Infalling matter of still higher angular momentum lands farther out to form a *circumbinary disk*. The stars interact gravitationally with this larger disk, which can act as a sink for angular momentum and energy. Such torquing is especially strong near apastron, *i. e.*, when the binary separation is a maximum, and tends to *increase* the orbital eccentricity.

In response to the gravitational torquing, the circumbinary disk becomes largely evacuated close to the binary orbit. Calculations indicate that some disk material also accretes onto the secondary star. This material generally has higher specific angular momentum than the star itself, and its accretion *decreases* the eccentricity. Apparently, it is this decrease that eventually predominates, but there is still no theoretical explanation for the correlation of eccentricity and

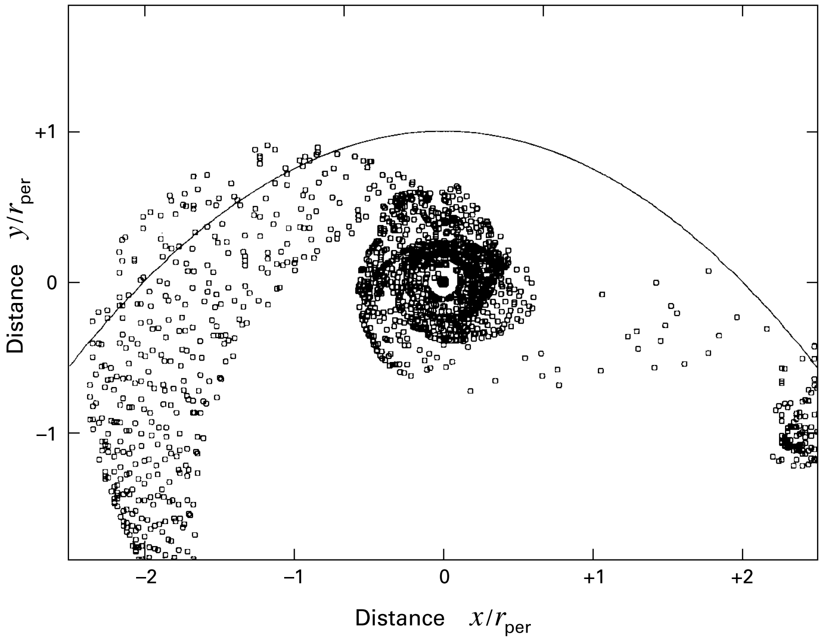


Figure 12.20 Disruption of a disk by a passing companion. The parabolic orbit of the point-mass companion is indicated. The particles represent the gas density in the equatorial plane.

period evident in Figure 12.14. Important aspects of the binary's evolution can be seen in Figure 12.21, which displays snapshots from an SPH simulation covering several orbital periods. Clearing of gas becomes pronounced within one revolution, as the stars generate a wavelike disturbance. We will see in Chapter 17 how high-resolution observations at infrared and millimeter wavelengths have actually furnished evidence for such centrally evacuated, circumbinary structures.

12.3.6 Tidal Circularization

We noted earlier that the minimum period for circular orbits within any group of binaries increases steadily with the age of that group. This trend demonstrates that even binaries with no circumstellar gas at all continue to undergo orbital evolution. The underlying physical effect is a mutual torquing of the stars created by the tides each raises in the other. The matter is of relevance for binary origins because the tidal effect is greatest when the stars themselves have their largest radii, *i. e.*, during the pre-main-sequence phase.

Figure 12.22 illustrates the basic process involved. Here, in a greatly exaggerated representation, we depict the two tidal bulges created in a primary star of mass M_1 by a secondary of mass M_2. The latter, which we assume moves in an eccentric orbit, is shown for simplicity as a point source of gravity. If the primary were rotating synchronously, so that its rotational period matched the orbital period of the secondary, the two bulges would point directly along the line

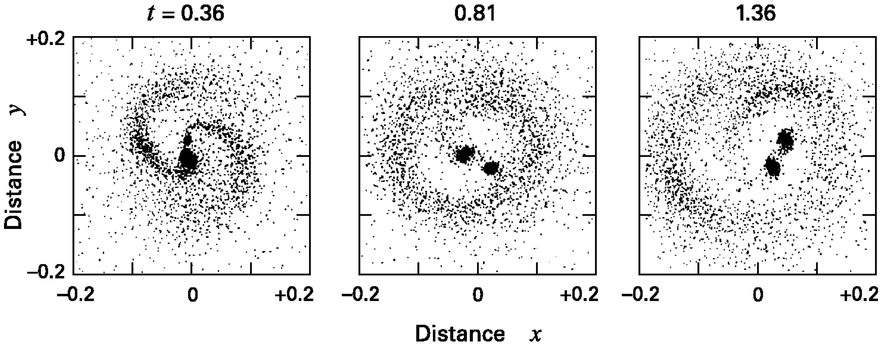


Figure 12.21 Clearing of a central gap in a circumbinary disk. The particles again represent density in the equatorial plane, and the time is in arbitrary units.

of centers. In this case, there would be no torque. Any internal friction, however, creates a phase lag in the primary's response to the tidal perturbation. The bulges then become misaligned by the angle α. The result, as depicted in Figure 12.22, is a torque that acts to pull the primary back into synchronism. Concurrently, the bulges retard the motion of the secondary and reduce the orbital eccentricity. The system evolves toward a state of lower energy, while conserving its total (spin plus orbital) angular momentum. The endstate is a binary with synchronous spins and circular orbits.

The actual rate of evolution varies with both the distance between stars and the magnitude of the internal friction. On the first point, we note that the bulge is created by the *difference* between the gravitational force exerted by the secondary locally and at the primary's center of mass. This tidal force, measured per unit mass, has the approximate magnitude $G\,M_2\,R_1/a_{\rm tot}^3$. Here, R_1 is the primary's unperturbed radius and $a_{\rm tot}$ is the interbinary separation, defined in equation (12.16). The potential energy per unit mass associated with this force is about $G\,M_2\,R_1^2/a_{\rm tot}^3$. In the presence of the primary's own gravitational field, such energy is sufficient

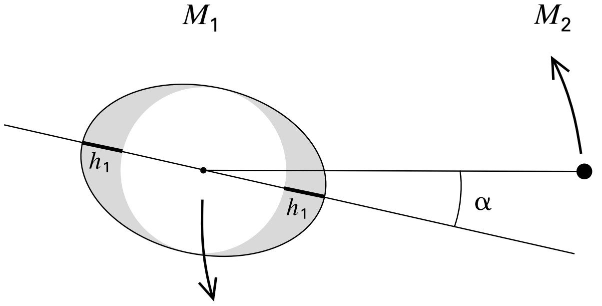


Figure 12.22 Tidal torquing of the primary star within a binary pair. The secondary, represented as a point mass, creates tidal bulges that are tilted with respect to the line of centers.

to lift an element from the surface to a height h_1, where

$$\frac{G\, M_1\, h_1}{R_1^2} \approx \frac{G\, M_2\, R_1^2}{a_{\rm tot}^3} \; . \tag{12.25}$$

Thus, the characteristic bulge height h_1, as depicted in Figure 12.22, is

$$h_1 \approx \left(\frac{M_2}{M_1}\right) \left(\frac{R_1}{a_{\rm tot}}\right)^3 R_1 \; . \tag{12.26}$$

The *mass* contained in the bulge, ΔM_1, is proportional to $(h_1/R_1)\, M_1 = (R_1/a_{\rm tot})^3\, M_2$. Finally, the magnitude of the synchronizing torque is

$$\Gamma_{\rm synch} \approx \frac{G\, \Delta M_1\, M_2\, R_1^2}{a_{\rm tot}^3} \sin \alpha \; , \tag{12.27}$$

which is proportional to $a_{\rm tot}^{-6}$. From Kepler's third law, equation (12.15), $\Gamma_{\rm synch}$ varies as P^{-4}. This sensitivity accounts for the rather sharp cutoff period for circular orbits.

Friction within the primary sets the magnitude of the lag angle α. Inside convection zones, the turbulent mixing of eddies provides an effective viscosity far greater than that arising from the random motion of atoms and molecules. Pre-main-sequence stars of low mass are either partially or fully convective (Chapter 16). Since these objects also have large radii, and since $\Gamma_{\rm synch}$ varies as R_1^5, the reduction in eccentricity proceeds efficiently. But what of young binaries composed of more massive stars that have no outer convection? Observations show that these also have circular orbits, at least for short enough periods. In this case, the system's energy loss stems from the damping of internal waves generated by the eccentrically orbiting companion. To summarize, while the tidal theory in both its forms still has significant gaps, it provides a solid basis for understanding both the early and longer-term evolution of binaries.

12.4 Formation of Stellar Groups

In seeking the origin of populous aggregates such as bound clusters, we turn to the larger cloud structures that spawn them. These are the clumps within giant molecular clouds or dark cloud complexes. The detection of numerous infrared sources within such entities has provided a glimpse of clusters in the making. Empirically, it is only the most massive clumps, generally those of $10^3\ M_\odot$ or greater, that harbor an embedded cluster. These bodies are centrally peaked in density and are gravitationally bound according to the virial theorem. That is, $|\mathcal{W}| \gtrsim |\mathcal{T}|$, where the kinetic energy term derives from the observed width of the tracer molecule, invariably a CO isotope. Lower-mass clumps for which this inequality is reversed must be confined by the ambient pressure within the complex. Systematic observations, particularly in the Rosette Molecular Cloud, have confirmed that star formation is never present in this latter type.

12.4.1 Parent Cloud Contraction

How does a massive clump evolve to the point of creating a multitude of dense cores and, ultimately, stars? It is tempting to view stellar groups as somehow arising from the breakup of the parent cloud, *i. e.*, from the process of dynamical fragmentation. Over the years, many have articulated this idea in various forms, with the breakup either occurring promptly or through a hierarchical series of steps. However, the cumulative results of both observation and theory cannot be reconciled with such a view. A more appealing picture, we shall argue, is that the parent cloud undergoes slower, quasi-static contraction. Dense cores accumulate their mass from this changing background structure.

Dynamical fragmentation fails here for essentially the same reason as it did for binary-producing dense cores. A massive cloud that is out of force balance *ab initio* has no precursor that is a coherent body. Yet giant complexes do contain gravitationally bound clumps with no infrared sources. These are presumably the analogues of starless dense cores. In the present context, a portion of these clumps should later evolve to produce internal clusters. Again, this development plausibly involves contraction, but *not* global collapse and breakup. We have seen in the numerical studies that the configurations which fragment by this latter route are either elongated spindles or flattened slabs, depending on the initial cloud shape. Neither outcome describes the present-day clumps forming stellar groups, nor the reconstructed configurations that gave rise to expanding OB associations (recall Figure 4.16).

In our earlier study of cloud equilibria, we noted that massive clumps owe their fully three-dimensional shapes to the mechanical support from MHD waves. The clouds are in near equilibrium between wave support and self-gravity, and therefore must contract because of a slight imbalance between the two. What ultimately drives any such quasi-static evolution is a net loss of energy, which plausibly results in this case from turbulent dissipation. How this dissipation arises in detail is not well understood. Certainly the ion-neutral drift we invoked for MHD wave damping is ineffective over typical clump dimensions. Nor is it clear, as discussed in § 10.3, how quiescent dense cores separate out from this turbulent medium on relatively small length scales. Finally, a complete theory must identify which properties of a clump dictate whether it produces a T association, bound cluster, or OB association.

12.4.2 Development of Luminosity Functions

Let us leave aside momentarily these basic issues associated with the condensation process itself, and focus instead on the resulting stellar groups. The very youngest of these are the embedded clusters discovered through near-infrared surveys. We saw in Chapter 4 how observations in multiple wavebands are beginning to establish the bolometric luminosity functions for these systems, *i. e.*, the number of objects per (logarithmic) unit of L_*. In any system, this function must change with time as member stars age. Suppose now that cluster formation indeed occurs through the collapse of dense cores scattered widely throughout a parent cloud. This basic picture, we will show, helps us to understand quantitatively the evolution of luminosity functions, and thus strengthens their role as potential observational tools.

Suppose that the cores, however created initially, are going into collapse at the rate $C(t)$. This rate starts at zero and is thereafter a smooth function of time. Presumably, the behavior here depends on the global contraction of the parent cloud. Once a core goes into collapse, it builds up a protostar at the mass infall rate $\dot{M}$. The cluster membership at any time t consists of both protostars and more evolved objects past their infall phase. Now theory supplies the luminosity $L_*(M_*, t)$ at any evolutionary stage of a star. For a specified formation rate $C(t)$, we may sum these contributions over mass to obtain the bolometric luminosity function, which we shall denote as $\Phi_*(L_*, t)$.

The most serious impediment to this program is our ignorance of how protostellar infall ceases. As long as a star is acquiring mass from its core, we may calculate its luminosity through equation (11.24). During its quasi-static contraction at fixed mass, L_* follows from pre-main-sequence theory (Chapter 16). At what point, however, does the transition occur? We cannot answer this equation from first principles, but observations provide a clue. As the cluster members continue to age, we know that they will ultimately appear on the main sequence with a mass distribution that resembles the field-star IMF. This empirical fact tells us when infall ends, at least in a statistical sense.

Figure 12.23 helps make our reasoning explicit. Let $p(M_*, t)\,\Delta M_*$ denote the number of protostars at any time t, with masses between $M_* - \Delta M_*/2$ and $M_* + \Delta M_*/2$. Similarly, let $s(M_*, t)\,\Delta M_*$ be the number of "post-infall" objects in the same mass interval; these include both pre-main-sequence and main-sequence stars. Referring to the figure, we see that the infall process will deplete the central bin of protostars over a time interval $\Delta t = \Delta M_*/\dot{M}$. Here we may take $\dot{M}$ to be a constant for simplicity. Not all of these protostars, however, become higher-mass objects. We suppose that a fraction $\nu(M_*)\,\Delta t$ cease their accretion. During the same time interval, the central bin is also partially refilled by protostars from the lower-mass bin. The fraction of these that shift into the central bin is $[1 - \nu(M_* - \Delta M_*)]\,\Delta t$. The net change in the protostar population thus obeys

$$\begin{aligned}&[p(M_*, t+\Delta t) - p(M_*, t)]\,\Delta M_* = \\ &\qquad p(M_* - \Delta M_*, t)\,[1 - \nu(M_* - \Delta M_*)\Delta t]\,\Delta M_* - p(M_*, t)\,\Delta M_* \ .\end{aligned}$$

Taking the limit of this equation for both small ΔM_* and Δt, and utilizing $\Delta M_* = \dot{M}\,\Delta t$, we find

$$\left(\frac{\partial p}{\partial t}\right)_{M_*} + \dot{M}\left(\frac{\partial p}{\partial M_*}\right) + \nu(M_*)\,p(M_*, t) = 0 \ . \qquad (12.28)$$

Equation (12.28) is subject to the boundary condition that there be no protostars initially:

$$p(M_*, 0) = 0 \ . \qquad (12.29a)$$

In addition, the ongoing collapse of new dense cores continually supplies protostars of zero mass. Since these are displaced to higher-mass bins through infall, we have

$$[p(0, t+\Delta t) - p(0, t)]\,\Delta M_* = C(t)\,\Delta t - p(0, t)\,\Delta M_* \ .$$

Invoking the same limit as before, we obtain a second boundary condition for equation (12.28):

$$p(0, t) = C(t)/\dot{M} \ . \qquad (12.29b)$$

It is now straightforward to derive a similar equation for the post-accretion stars. As depicted in Figure 12.23, such an object arises *only* from a protostar of the same mass that has ended infall. Thus, $s(M_*, t)$ is related to $p(M_*, t)$ through

$$[s(M_*, t+\Delta t) - s(M_*, t)]\,\Delta M_* = \nu(M_*)\,p(M_*, t)\,\Delta M_*\,\Delta t \ . \qquad (12.30)$$

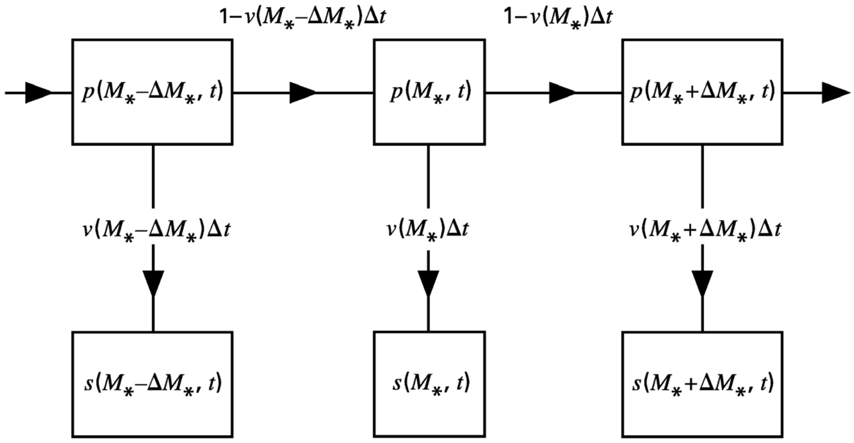


Figure 12.23 Transfer of mass among protostars (p) and pre-main-sequence objects (s) in a forming cluster. Stars are moved into neighboring mass bins with the indicated probabilities.

In the limit, we find

$$\left(\frac{\partial s}{\partial t}\right)_{M_*} = \nu(M_*)\, p(M_*,t) \ .$$

Finally, we also require that the post-infall population start out at zero:

$$s(M_*,0) = 0 \ . \tag{12.31}$$

Given the functions $C(t)$ and $\nu(M_*)$, along with $\dot{M}$, we may solve equations (12.28) – (12.31) for the populations $p(M_*,t)$ and $s(M_*,t)$. Whatever its explicit form, $C(t)$ must eventually fall to zero, by a time we shall denote as $t_{\rm cutoff}$. The function $\nu(M_*)$ is the probability per unit time that a protostar stops accreting new mass. Note that we have tacitly assumed that the relevant physical process depends only on M_* and not, for example, on the evolutionary state of the cluster. Such an assumption can only be justified *a posteriori*, through comparison of the theory with cluster data. Our procedure, then, is to choose $C(t)$, solve the equations for arbitrary $\nu(M_*)$, and then tune this latter function until the post-infall population follows the field-star IMF. Utilizing the notation for the latter from § 4.5, we demand that

$$\lim_{t\gg t_{\rm cutoff}} s\,(M_*,t) = N\,\xi(M_*) \ . \tag{12.32}$$

Here, $N \equiv \int_0^{t_{\rm cutoff}} C(t)\,dt$ is the total number of stars produced.

After obtaining the time dependence of the stellar populations, one multiplies by the appropriate luminosities to derive $\Phi_*(L_*,t)$. Figure 12.24 shows the results of a numerical calculation. Here, $C(t)$ was chosen, for simplicity, to be a strict constant from $t=0$ to $t_{\rm cutoff} = 1\times 10^7$ yr. The protostellar infall rate was set to $\dot{M} = 1\times 10^{-5}\,M_\odot\,{\rm yr}^{-1}$. At early times, the luminosity function has two distinct maxima. The lefthand one represents the steadily growing number of pre-main-sequence stars, while that on the right stems from accreting protostars. The protostellar contribution to the luminosity function is quite sharply peaked because

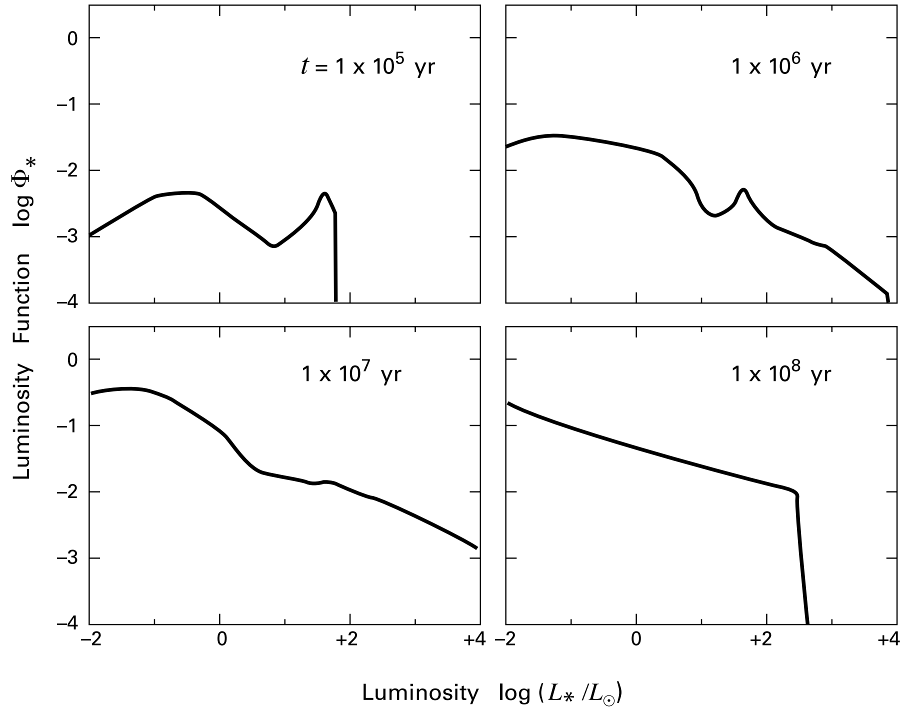


Figure 12.24 Evolution of the luminosity function in a model cluster.

$L_{\rm acc}$ is proportional to M_*/R_* (equation (11.5)). This ratio, in turn, is held nearly constant by the thermostatic action of deuterium burning. Hence, the luminosity varies little over a significant mass range.

The assumed constancy of $C(t)$ implies that the total cluster population increases linearly with time. After a brief transient period, this increase occurs entirely within the pre-main-sequence stars. That is, the total number of protostars saturates to a steady-state level. Recall here that protostars are both created by the collapse of new cores and effectively destroyed by infall, which ultimately transforms them into pre-main-sequence stars. If C now denotes an average of the global formation rate and $\bar{M}$ the typical stellar mass produced, then the steady-state protostar number is about $C\,\bar{M}/\dot{M}$. It follows that the *fractional* population is

$$f_{\rm proto} \approx \frac{\bar{M}}{\dot{M}\,t} \;, \tag{12.33}$$

over the time interval $\bar{M}/\dot{M} \lesssim t < t_{\rm cutoff}$. According to the IMF of equation (4.6), $\bar{M}$ is about $0.2\,M_\odot$; here we have assumed the IMF to be flat below $0.1\,M_\odot$. In our model cluster, therefore, $f_{\rm proto}$ drops to 0.02 by only 10^6 yr. Quite generally, the small values expected for this fraction are of clear significance in the observational search for protostars.

Returning to the luminosity function, Figure 12.24 shows how the pre-main-sequence contribution soon dominates. The curve becomes quite broad, reflecting the large spread in masses

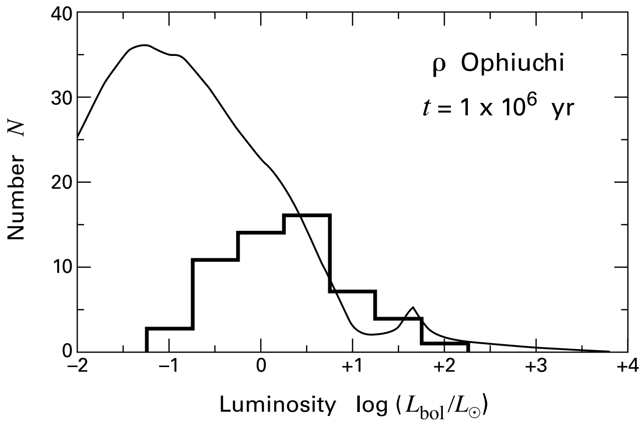


Figure 12.25 Comparison of the theoretical bolometric luminosity function (*lighter curve*) with the empirical one (*darker histogram*) for the ρ Ophiuchi cluster. The theoretical curve is for the indicated cluster age.

and contraction ages. Although no new protostars appear after 1×10^7 yr, $\Phi_*(L_*, t)$ keeps evolving because of this contraction. The most massive stars reach the main sequence first. Correspondingly, Φ_* approaches a featureless curve, beginning at relatively high L_*-values. This curve is $\Psi(L_*)$, the "initial" luminosity function discussed in § 4.5. Recall that Ψ is obtained by convolving the IMF with main-sequence luminosities. By $t = 1 \times 10^8$ yr, the upper portion of Φ_* is severely truncated, as the most luminous stars evolve off the main sequence. In the numerical calculation, these were simply deleted from the cluster.

How well does this theory accord with observations? For truly embedded clusters, the fraction of stars with measured bolometric luminosities is still small. Hence, a thorough comparison remains for the future. Nevertheless, Figure 12.25 shows that the available results are encouraging. Here, we have reproduced the luminosity function from the L1688 cluster of ρ Ophiuchi, as first presented in Figure 4.6. The smooth curve is Φ_* from the same sequence as in Figure 12.24. In this example, we obtain a best fit for a cluster age of 1×10^6 yr. This agreement, moreover, is only at the highest luminosities. At lower L_*, the empirical function declines steeply, revealing the finite sensitivity of the observations. The theoretical curve, meanwhile, rises to a much higher level. This hypothetical population consists almost entirely of pre-main-sequence stars relatively close to the birthline. Such a prediction is certainly reasonable, as near-infrared surveys indicate a much higher membership than is shown in the histogram.

12.4.3 Age Histograms

More complete luminosity functions are available from systems where a greater fraction of members are optically revealed. Consider Taurus-Auriga, whose luminosity function we showed as Figure 4.13a. There we noted that the empirical curve rises above $\Psi(L_*)$ at moderately low luminosities, but before the sensitivity falloff. This excess again stems from young pre-main-sequence stars and is evident in several of the theoretical curves in Figure 12.24. However, the evolutionary status of a T association may be gauged in a much more direct and precise

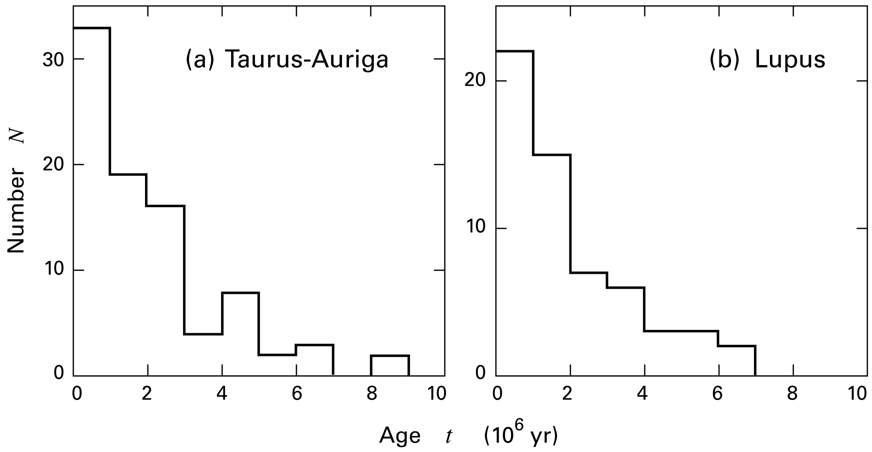


Figure 12.26 Age histograms for the T associations in (a) Taurus-Auriga, and (b) Lupus.

manner. Since we can place most stars in the HR diagram, it is feasible to read off their ages individually through comparison with pre-main-sequence tracks. The resulting distribution is a powerful diagnostic.

Figure 12.26a shows the ages of the Taurus-Auriga population, binned in 10^6 yr intervals. Here we have utilized the HR diagram of Figure 4.9a. The crowding of isochrones close to the main sequence makes it impossible to assign ages to some of the older, more massive stars. This limitation should always be borne in mind, but is not severe in the present case, where our figure includes 87 out of the 103 stars in the HR diagram. Scanning the age histogram from right to left tells us the star formation history of the region. It is apparent that stellar birth began nearly 10^7 yr ago and has been accelerating to the present epoch. A similar history is evident in Figure 12.26b, which shows the Lupus association (recall Figure 4.9b).

We saw in Chapter 4 how main-sequence turnons in the HR diagram allow us to quantify the onset of vigorous star formation in a region. Age histograms go further by showing in more detail how this activity progresses. The contraction of any individual dense core, eventually leading to its collapse, is mediated by self-gravity and local diffusion of the magnetic field. There is no known way that this process occurring in one location can stimulate contraction in another several parsecs distant. But Figure 12.26 shows, in effect, an orchestrated pattern of collapses in many cores. The cores must therefore be evolving in response to an alteration of their common environment. An accelerated rate of star formation, in particular, suggests again that the parent cloud is contracting, so that more sites are accumulating mass.

Only a handful of the closest T associations have been examined in enough detail to produce age histograms, and even these results are bound to change as observations improve. Nevertheless, the pattern shown in our two examples appears to be quite general. It is especially noteworthy that none of the systems is older than about 5×10^6 yr. That is, any star formation prior to that time is sporadic and at a low level compared to its current rate. If this feature also continues to hold, then the maximum inferred age should be a measure of the actual lifetime for such groups.

12.4.4 Demise of T Associations

What becomes of older T associations? Note that most of their individual members have contraction lifetimes well in excess of 10^7 yr. A commonly formed $0.3\ M_\odot$ star, for example, takes 2×10^8 yr to reach the main sequence. On a purely statistical basis, it would seem that objects near this age should outnumber the younger ones actually seen. The scarcity of this older population within T associations is known historically as the *post-T Tauri problem.* However, the issue ceases to be problematic if the groups themselves disperse at an earlier epoch. Once this occurs, the aging pre-main-sequence stars are mingled observationally with the general field and become much harder to identify.

The projected stellar density of associations does not greatly exceed that of the background. Hence, the group's physical size need only increase by a modest factor before it effectively loses its identity. Accompanying the expansion is dissipation of the parent cloud's gas. Indeed, the two phenomena are causally linked. To see this qualitatively, consider the total energy of a spherical cloud with mass M and radius R. The cloud is composed of both gas and embedded stars, where the latter comprise a minor fraction of the total mass. We write the energy as the sum of its kinetic and gravitational potential components:

$$E = \frac{1}{2} M V^2 - \frac{\eta G M^2}{R} \tag{12.34a}$$

$$= -\frac{\eta G M^2}{2 R} \ . \tag{12.34b}$$

Here, V is an average random velocity of the gas and stars, while η is a nondimensional factor of order unity, whose precise value depends on the distribution of mass. In this highly simplified model, we take the cloud to be wholly supported by the effective pressure from internal motion, neglecting both the thermal energy and that associated with a static magnetic field. Equation (12.34b) then follows by application of the virial theorem.

Suppose now that a small amount of gas is removed from the cloud. During this brief interval, we assume that neither V nor R changes. However, the energy of the remaining configuration is altered, and both V and R must secularly readjust to maintain force balance. Differentiation of (12.34a) yields

$$\begin{aligned} \frac{dE}{dM} &= \frac{1}{2} V^2 - \frac{2 \eta G M}{R} \\ &= \frac{3 E}{M} \ , \end{aligned} \tag{12.35}$$

where we have again employed the virial theorem result. Using subscripts to denote initial values, we integrate and find that

$$\frac{E}{E_\circ} = \left(\frac{M}{M_\circ}\right)^3 \ . \tag{12.36}$$

From equation (12.34b), we finally obtain

$$\frac{R}{R_\circ} = \frac{M_\circ}{M} \ . \tag{12.37}$$

As M falls below $M_\circ$, equation (12.37) tells us that the radius continually grows. This expansion results from weakening of the gravitational binding that the gas provides. From equation (12.34a), the diminution of M both lowers the kinetic energy and raises the (negative) potential energy. Since the latter varies quadratically with mass, the net result is that E rises toward zero, as may also be seen from equation (12.36). Of course, the system mass cannot truly vanish, since it eventually reaches that of the stellar component. By this point, the stars are so weakly bound that they are easily dispersed by such external forces as the tidal gravity from nearby clouds.

The system's gain in energy stems from whatever agent dispels the gas. Here we touch upon another critical, but poorly understood, aspect of the picture. It has long been supposed that the stars within the complex are responsible for its dissolution, presumably through their strong winds. We will see in Chapter 13 how these winds are collimated into narrow jets, which in turn stir up ambient gas to produce molecular outflows. In some complexes, such as NGC 1333 in Perseus, there are an impressive number of jets permeating the region. Other systems, such as Taurus-Auriga, appear comparatively quiescent, at least in terms of their gross morphology. Recall, in this latter example, that the molecular gas totals some $10^4\,M_\odot$, two orders of magnitude above the stellar contribution. Individual stars producing outflows only do so for a period of 10^5 yr, after which they are replaced by a comparable, new population. Can this relatively small number of objects disperse the entire complex within a few million years?

The alternative would be an *external* means of gas dispersal. The cloud morphology again rules out destruction by the even stronger winds from nearby massive stars. Another possibility is dissipation through the radiative heating from these same objects. Such photoevaporation indeed may occur, but only in a cloud that is within several parsecs of the massive star in question, so that it becomes engulfed in the spreading HII region (Chapter 15). Few, if any of the parent clouds of T associations are in this situation. More generally, any proposal involving external heating must face the issue of why the cloud did not evaporate *before* stars began forming.

We surmised earlier that the cloud first undergoes quasi-static contraction, both to produce dense cores initially and to create the subsequent accelerated pace of star formation. This contraction results from an energy *decrease* that we attributed to internal dissipation. Gas dispersal, whatever its underlying cause, effectively *increases* the system's energy. In reality, both effects occur simultaneously. We may picture the cloud slowly contracting and producing stars, all the while losing mass to the surrounding medium. Eventually, its mass falls to the point where contraction ceases. The increasingly rarefied cloud then expands outward until its stars disperse into the field.

Essential to this picture is our assumption that the total stellar mass is small relative to that of the gas. If this were *not* the case, *i. e.*, if the star formation efficiency were high, the cloud could not re-expand significantly through gas dissipation. As we have mentioned, the clouds producing T associations appear to do so with roughly 1 percent efficiency. Note also that a typical dense core of perhaps 3 $M_\odot$ generally creates a star or binary with total mass about 0.1 times that value. Since the ensemble of cores comprises some 10 percent of the parent cloud mass, our global estimate of the mass fraction is reasonable. On even larger scales, one may compare the extrapolated masses of OB associations with those of their parent clouds. This exercise also leads to fractions at the 1 percent level. The two types of associations provide a comparable total mass throughout the Galaxy, with T associations being less massive individually but more common. Our representative efficiency is therefore basic, empirical input for any general theoretical account of stellar groups.

12.4.5 Open Clusters

Let us now turn to open clusters and see if their origin may be understood through application and extension of the ideas presented thus far. Note first that most cluster members lie on, or very close to, the main sequence, so that construction of age histograms is not feasible. In addition, the present-day morphology of these systems gives little indication of their mode of production. Recall that the Pleiades, for example, consists of a tight core of relatively massive stars, surrounded by a distended halo of lighter members. This structure is a manifestation of dynamical relaxation, which has had ample time to operate even in this system of age 10^8 yr. The massive stars have gradually transferred energy, via repeated gravitational encounters, to lower-mass members, thereby sinking toward the center while inflating the halo. Moreover, the lightest stars are continually stripped away by the tidal gravitational force from the Galaxy as the outer region expands.

We might also hope to learn the circumstances of open cluster birth by probing even younger, embedded groups that seemed destined to be gravitationally bound. The most well known example is the cluster in ρ Ophiuchi. Here the very density of the cloud material frustrates attempts to study its internal structure. Thus, the conventional technique for measuring gas column density is to utilize an optically thin tracer such as $^{12}\mathrm{C}^{18}\mathrm{O}$. The detection equation (6.1) requires specification of T_{ex}, which is generally obtained using the optically thick $^{12}\mathrm{C}^{16}\mathrm{O}$ (recall equation (6.3)). Unfortunately, profiles of $^{12}\mathrm{C}^{16}\mathrm{O}$ are self-reversed, showing that much of the radiation emanates from a relatively cold, exterior sheath. As a result, both the measured T_{ex} and the resulting column density are only lower bounds. A more promising tool is the millimeter continuum emission from warmed dust. Observers have already used this technique to map a number of dense cores in the region.

Our analysis of the ρ Ophiuchi luminosity function indicated an age of 1×10^6 yr. Within this period, more stars have been produced than in the entire Taurus-Auriga complex. If this example is typical, the parent clouds of bound clusters form stars at a relatively high rate. We may suppose, therefore, that these structures contract more rapidly, though still in quasi-static fashion. If we further assume that star formation begins at a characteristic density in all clouds, then faster contraction could reflect a higher initial mass. A more massive cloud presumably has more vigorous internal motions and thus an elevated degree of dissipation.

As before, we expect that cloud contraction, star formation, and gas dispersal all proceed simultaneously. The decline in cloud mass is now unable to reverse the contraction, but does manage to slow it down and prevent runaway growth. Thus, by the time most gas has vanished, the structure is *more* compact than initially, but not by a huge factor. The requisite time is typically several million years, *i. e.*, greater than the inferred ρ Ophiuchi age, but less than the lifetime of OB associations. Over a much longer interval, the system of stars relaxes dynamically to form the more expansive open clusters seen today.

Furnished with this picture, we begin to see why open clusters are a distinct minority among stellar groups. In parent clouds of relatively low mass, gas dissipation is the dominant effect.

That is, contraction occurs initially, but is slow enough to be eventually reversed. In massive clouds, on the other hand, self-gravity is the strongest influence, and the contraction quickly accelerates. We shall argue shortly that this circumstance leads, somewhat paradoxically, to expanding OB associations. Only parent clouds within some intermediate range are able to contract and lose mass at roughly the same rate, so that they evolve to gravitationally bound stellar groups.

12.5 Massive Stars and Their Associations

We finally consider the origin of groups containing high-mass stars. As our starting point, we recall the central fact that nearly all OB associations are found near giant molecular complexes, which we may take to be their birthplaces. Systems containing high-mass stars therefore emerge from the greatest aggregates of diffuse material. Looking to even younger environments, it is also true that embedded clusters with more luminous infrared sources are generally located within larger and denser molecular clouds. If we are to place OB associations within the family that includes bound clusters and T associations, then each system must result from the contraction of an especially massive cloud clump.

12.5.1 The Highest Stellar Masses

How any *individual* O or B star forms is another matter entirely. We stressed in Chapter 11 the difficulty extending traditional protostar theory to this regime. In addition, observations have yet to provide convincing analogues to dense cores, *i. e.*, relatively quiescent structures that appear destined to form massive stars. Recall that the typical core does not have sufficient gas to form even a B2 star of 10 $M_\odot$. We will study in Chapter 15 the entities known as *hot cores*. As their name implies, these already contain luminous stars, which are rapidly destroying their host structures through the combined effects of radiative heating and winds. Younger, more pristine bodies of the appropriate mass are not yet evident and indeed may not exist.

An especially important characteristic of massive stars is their tendency to form in crowded environments. Extreme cases have already been noted. The cluster associated with the Trapezium in Orion is among the densest in the Galaxy. Starburst regions in other galaxies harbor many more O stars and have even greater stellar densities. If the highest-mass stars do *not* form through the collapse of isolated, quiescent structures, then this empirical trend may be of critical significance. Specifically, the stars in question might *require* high densities to form. They might even arise, as we will suggest, from the coalescence of previously formed cluster members. Whether or not this is true, the observations do indicate a more intimate relationship between the stars and their groups than for low-mass objects.

How massive are the stars within OB associations, and how is such information obtained? Chapter 4 covered the technique of spectroscopic parallax, which has established distances to many of these groups. For any star not in a binary, this is the first step toward a mass measurement, since it allows one to convert the object's apparent magnitude (typically m_V) to an absolute one. Hence one also obtains the total luminosity, given a reliable bolometric correction. One could then, at least in principle, derive $T_{\rm eff}$ from a broadband color index. For massive stars, however, standard indices such as $B - V$ change little with spectral type and are thus unreliable as temperature indicators. The practice instead is to fit a theoretical model of the stellar atmosphere to the observed narrowband spectrum. Such models contain two free parameters: T_{eff} and the specific gravity $g \equiv GM_*/R_*^2$. The star's mass may then be read off from its position within the main sequence in the $L_{\mathrm{bol}} - T_{\mathrm{eff}}$ plane. Alternatively, one determines R_* through the blackbody relation between L_{bol} and T_{eff}, equation (1.5). The M_*-value follows by using the best-fit determination of g from the atmosphere model.

This basic method has revealed extraordinarily high O-star masses, for surveys that cover sufficiently large volumes. Closest at hand is the complex of associations in Orion. Its most massive objects reside in the 1b subgroup, with ζ Ori topping the list at 49 $M_\odot$. In contrast, the brightest star in the Trapezium is θ^1 Ori C, an O6 object of 33 $M_\odot$. The very high density in the surrounding Orion Nebula Cluster is rivaled or surpassed in NGC 3603, an HII region in the Carina spiral arm, at a distance of 7 kpc (Plate 7). Here one finds six stars of roughly 50 $M_\odot$ within a volume less than 0.03 pc^{-3}. The clusters Tr 14 and Tr 16, some 3 kpc distant, each contain stars classified as O3, with $M_* \lesssim 100$ $M_\odot$. Even higher masses are indicated for a number of objects within the R136 region of 30 Doradus, in the Large Magellanic Cloud (Plate 8).

12.5.2 Clustering Systematics

An O3 star has a bolometric luminosity of 10^6 $L_\odot$ and drives a wind with a velocity of several thousand km s^{-1}. Given its ability to repel ambient material both thermally and mechanically, such an object may only be able to form through some rapid accumulation event. We have hinted that the key factor in any emerging theoretical picture will be the density of ambient material, whether in gas or stars. It is thus important to quantify, as much as possible, the clustering tendency of massive objects.

One fruitful approach to this question is to examine how often young stars in a given mass range are surrounded by other objects in close proximity. For O stars, the answer is that most lie near the centers of associations or dense clusters. Apparent exceptions to this rule are the relatively isolated objects in the field, which, we recall, constitute about 25 percent of the massive star population. (Many of these, however, could be runaways.) At the other extreme are T Tauri stars. Here, an individual object may be part of a loose aggregate of similar stars, as we found in Taurus-Auriga (Figure 4.8). There is no tendency, however, for the star in question to reside near the density peak of a crowded, and perhaps gravitationally bound, system.

The transition between these two cases must be at some intermediate mass value. Many Herbig Be stars are situated within groups of several dozen partially embedded objects. Here, we remember the example of BD+40°4124 from Chapter 4 (Plate 5). A correlation exists between the mass of the central Be star and the number of lower-luminosity neighbors. There is even a rather well defined breakpoint. Thus, Herbig Ae stars are never accompanied by conspicuous groups. Conversely, rich clusters with stellar densities exceeding 10^3 pc^{-3} only appear around stars earlier than B5, corresponding to a mass near 6 $M_\odot$.

Could this trend represent an evolutionary effect? The typical Herbig Ae star is at least several million years old, so the absence of clustering might conceivably reflect prior dispersal. Indeed, numerical experiments show that groups of modest population disintegrate within a few

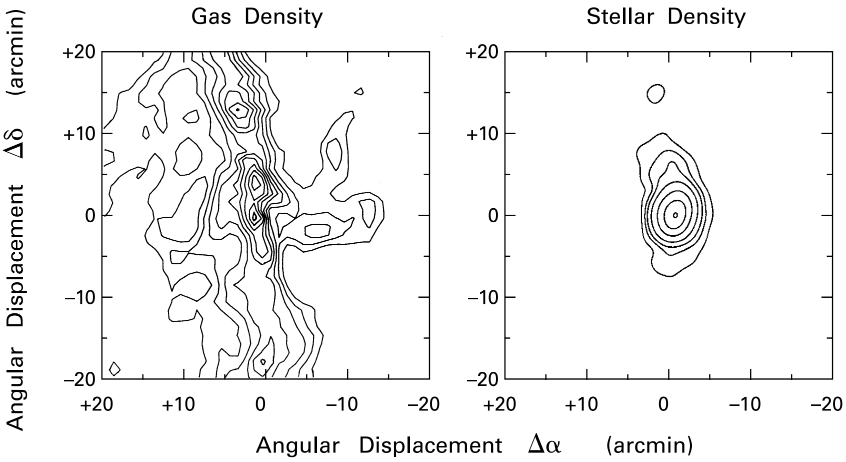


Figure 12.27 Depiction of the Orion Nebula Cluster in molecular gas (*left panel*) and stars (*right panel*). The gas column density is shown as contours of $^{13}C^{16}O$, while the stars are represented by a smoothed surface density. Both maps are centered on the star θ^1 Ori C.

crossing times. This is another example of dynamical relaxation, in which the most massive objects transfer enough energy to lighter members to actually unbind them. The observational studies, however, find no correlation between the *age* of a star and the presence of a cluster – only the mass seems to count.

12.5.3 Orion Revisited

The reason dynamical relaxation does *not* play a significant role in these systems is the presence of a sizable quantity of gas. This material can exert a gravitational force on any star comparable to or greater than that from the other cluster members. More generally, it is the internal cloud structure which largely sets the distribution of a young stellar population, including those with massive objects. This important point is well illustrated by the Orion Nebula Cluster. The left panel of Figure 12.27 shows the ridge-like distribution of molecular gas, as seen through contours of $^{13}C^{16}O$. Here the map covers a region about 4 pc across and is centered on the star θ^1 Ori C. The same area is depicted in the right panel, but now through the surface density of stars. The isodensity contours are clearly not spherical, as one would have expected for a swarm of point masses. Moreover, the nested shapes are generally oriented along the molecular ridge. Notice finally how θ^1 Ori C lies at the geometrical center of the stellar system.

It is not easy to discern the properties of individual stars in such a crowded region. Nevertheless, observers have managed to assign luminosities to over half of the cluster's 1600 optically visible members. The values of L_* are not directly integrated from fluxes, but rely on a bolometric correction to I-band observations. Figure 12.28 displays the resulting luminosity function. Also shown is the theoretical determination for Φ_* using the statistical method outlined earlier. For the best-fit cluster age of 2×10^6 yr, the two curves match over a broad range of luminosity,

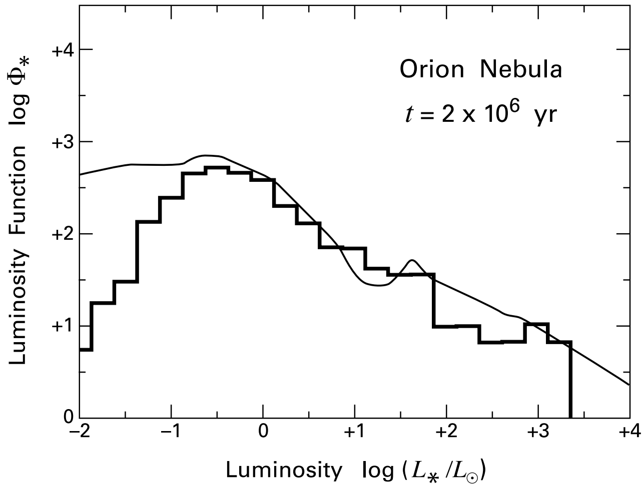


Figure 12.28 Bolometric luminosity function of the Orion Nebula Cluster. The lighter curve is the theoretical prediction for the indicated cluster age.

but depart for $L_* \lesssim 0.1\, L_\odot$. The falloff in the empirical function again reflects incompleteness of the survey.

As in the case of ρ Ophiuchi, which has a similar age, the vast majority of cluster members are predicted by the theory to be contracting pre-main-sequence stars of relatively low mass. Indeed, additional observations we will discuss shortly confirm that the stellar mass distribution is similar to the field-star IMF. (This result was assumed, of course, in the theoretical model for Φ_*.) As a matter of pure statistics, the cluster may be said to contain massive stars simply because its *total* population is large enough for the tail of the distribution to be represented. We stress, however, that such a viewpoint does not aid us in understanding the physical manner by which such objects form.

The stars indicated in Figure 12.28 are sufficiently revealed to allow study of their narrowband spectra at visual wavelengths. Thus, one may determine spectral types and effective temperatures for this large population, and place the stars in the usual HR diagram. We show the result as Figure 12.29. It is evident that the cluster membership spans over two orders of magnitude in stellar mass, from objects close to brown dwarfs up to O stars. Note the apparent displacement of the latter from the ZAMS. The offset here is not real, but reflects inaccuracy of the assigned T_{eff}-values in this regime.

One striking feature of the diagram is the extreme crowding at lower masses. Tallying up all the stars in the various mass bins confirms that the total distribution roughly follows the IMF. Of even greater interest are the stellar ages, obtained once more from theoretical isochrones. Figure 12.30 shows the age histogram for the cluster. Although the numbers involved are much greater than before, the pattern is a familiar one. The system displays a relatively low level of star formation activity 10^7 yr in the past, a gradual acceleration, and finally a steep rise toward the present epoch. It is now clear that the system "age" of 2×10^6 yr obtained from the luminosity function is simply a crude measure of the interval covering the most active formation. Thus,

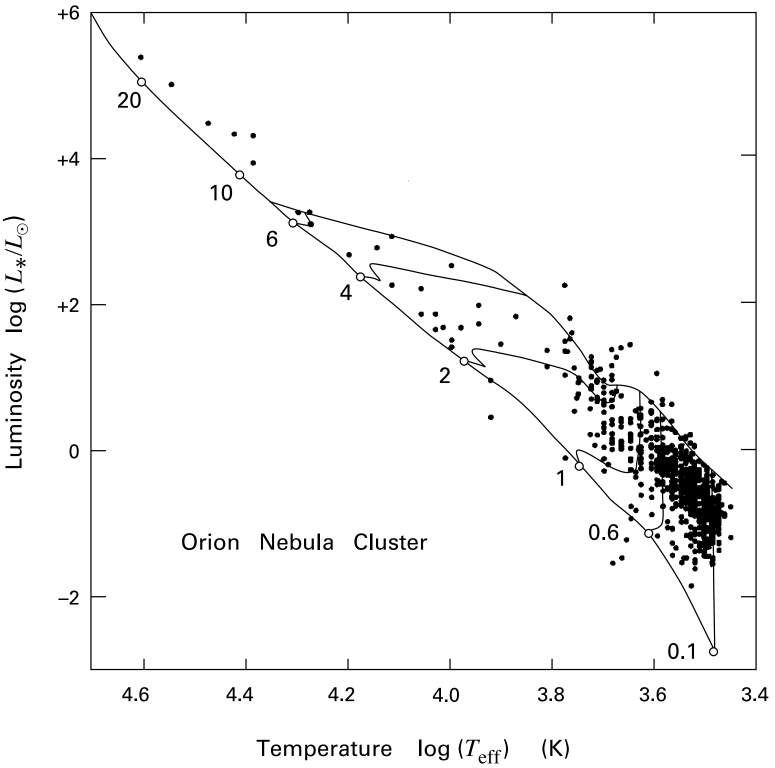


Figure 12.29 HR diagram of the Orion Nebula Cluster. The birthline, ZAMS, and selected pre-main-sequence tracks are also shown. Stellar masses for the tracks are given in solar units.

the birthrate function $C(t)$ employed in the statistical model should more properly increase with time.

The resemblance of Figure 12.30 to our other histograms actually hides an important difference. Consider again the rather remarkable fact that hundreds of stars close to the Trapezium are optically visible, despite the extraordinary density of the region. From an observational view, the extinction A_V actually declines as we approach to within a few parsecs of the cluster's center. The relatively low density in *gas* means that there are few, if any, deeply embedded stars, such as Class I sources. Thus, *star formation is no longer occurring.* That the age histogram does *not* show a corresponding decline is simply a reflection of its limited temporal resolution. Apparently, the falloff has happened within the last 10^6 yr.

It is the Trapezium O stars that are responsible for this turn of events. Ram pressure from their winds, along with ionization and heating from ultraviolet photons, have evacuated the gas in a rapidly expanding volume. Infrared observations show that vigorous star formation is continuing in the molecular gas *behind* the visible cluster. However, production has largely ceased within the foreground volume. Generalizing from Orion, the creation of objects which effect such a rapid clearing is clearly a pivotal event in the life of a stellar group.

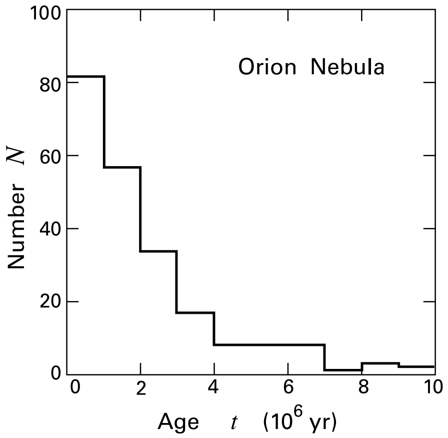


Figure 12.30 Age histogram of the Orion Nebula Cluster.

The rise of star formation prior to gas dispersal again indicates that the parent cloud has undergone bulk contraction, presumably through turbulent dissipation of its wave support. This contraction proceeds to higher densities than for the lower-mass clouds forming T associations or open clusters. At any time, moreover, the greatest density is found toward the central region. In our example of Orion, more detailed study reveals the presence of mass segregation, at least for stars with $M_* \gtrsim 5\,M_\odot$. Numerical N-body simulations confirm that the interior crowding of higher-mass objects *cannot* be the result of dynamical relaxation, which takes longer to act than a few million years. These calculations are highly simplified, as they include no background gas at all. Nevertheless, their essential conclusion seems unavoidable. The massive stars did not drift to their present locations, but must have formed *in situ.*

12.5.4 Birth of OB Associations

Creation of the O stars themselves may therefore be viewed as an event triggered by the attainment of some threshold in the central density. Gas dispersal then leads to expansion of the system through the loss of gravitational binding. It is instructive to view this process through our highly simplified model of a spherical cloud supported by turbulent motion. The new feature is that the dispersion now occurs *rapidly* compared to the system's dynamical time scale.

Consider again the cloud's total energy. Equation (12.34) still holds for the initial value $E_\circ$, if we also append subscripts to M, V, and R. As before, we assume that the latter two quantities are unable to change during the brief dispersal event. In place of (12.35), however, we must account for a finite change of mass, ΔM. The energy still increases, and its new value is

$$\begin{aligned} E &= \frac{1}{2}\,(M_\circ + \Delta M)\,V_\circ^2 - \frac{\eta\,G\,(M_\circ + \Delta M)^2}{R_\circ} \\ &= E_\circ + \frac{1}{2}\,\Delta M\,V_\circ^2 - \frac{2\,\eta\,G\,M_\circ\,\Delta M}{R_\circ} - \frac{\eta\,G\,\Delta M^2}{R_\circ}\ . \end{aligned}$$

Employing equation (12.34) for both $E_\circ$ and $V_\circ$, we find

$$E = -\frac{\eta\, G\, M_\circ^2}{2\, R_\circ} - \frac{3\, \eta\, G\, M_\circ\, \Delta M}{2\, R_\circ} - \frac{\eta\, G\, \Delta M^2}{R_\circ} \tag{12.38a}$$

$$= E_\circ \left(1 + \frac{\Delta M}{M_\circ}\right) \left(1 + \frac{2\, \Delta M}{M_\circ}\right) . \tag{12.38b}$$

The presence of the first factor after $E_\circ$ simply indicates that the energy goes to zero if all the mass were to vanish. (Recall that $\Delta M < 0$.) More interesting is the second factor, which tells us that E is positive when $|\Delta M| > M_\circ/2$. The system thus becomes gravitationally unbound if over half its mass disperses rapidly. Under the previous hypothesis of slow dispersal, the sphere would have remained bound, but doubled in size, under the same net mass loss (recall equation (12.37)).

To assess the change of radius in the present case, we again invoke mechanical equilibrium following the dispersal event. Utilizing the virial theorem, as in equation (12.34b), we find that the energy is proportional to the new value of M^2/R, so that

$$\frac{E}{E_\circ} = \frac{(1 + \Delta M/M_\circ)^2}{R/R_\circ} . \tag{12.39}$$

Combining equations (12.38b) and (12.39), we readily find that

$$\frac{R}{R_\circ} = \frac{M_\circ + \Delta M}{M_\circ + 2\, \Delta M} . \tag{12.40}$$

This relation confirms that the sphere again expands as a result of dissipation. Moreover, the loss of half the mass leads to the marginally bound state of infinite radius, as expected.

We noted earlier that the global efficiency throughout the Galaxy of converting molecular gas into stars is only a few percent by mass. We further argued that this figure represents a plausible estimate for most individual clouds forming stellar groups. Thus, when O stars rid a system of its gas, the fractional decline in M must generally exceed 50 percent. The configuration is then left with positive net energy. It expands into space as an OB association, gradually fading from view over the nuclear lifetime of its luminous members.

The gas dispersal we have invoked depends solely on the massive star content of the parent cloud. We have ignored the wholly different mechanism operating on those clouds that spawn T associations and bound clusters. This other form of gas erosion, which may stem from the *low-mass* stars, must still be at work, but it has less effect over the shorter contraction period. Here, we return to our central assumption that more massive parent clouds are subject to enhanced turbulent dissipation. Faster evolution to a denser central region is the logical extension of our general picture. The irony is that this circumstance soon leads to unbound, expanding stellar systems.

But is this indeed the fate of every group containing massive stars? In Chapter 4, we called attention to the ill-defined boundary between young open clusters and OB associations. There are certainly observed systems containing massive stars that have a high spatial density. Indeed, the Orion Nebula Cluster is a prime example. Only accurate measurement of the stars' velocity dispersion can answer the question of whether any of these groups are actually bound. Conversely, some low-mass systems already known to be bound, and residing within OB associations, may be remnants of incomplete dispersion. One example might be the α Per cluster within the Cas-Tau association (Figure 4.14).

12.5.5 Merging Cores and Stars

Let us conclude with some brief remarks on the actual dynamics of massive star formation. The challenge, once again, is to assemble the star so quickly that it has no opportunity to drive back its own reservoir of cloud matter. In the language of traditional protostar theory, we require that the mass accretion rate $\dot{M}$ be at least several orders of magnitude above the values appropriate for the low-mass regime. Equation (10.31) tells us, however, that the sole means to achieve this is to have a higher sound speed, and thus temperature, in the dense core *prior* to its collapse. For example, an $\dot{M}$ exceeding $10^{-4}\,M_\odot\,\mathrm{yr}^{-1}$ would require $T \gtrsim 150\,\mathrm{K}$. Such temperatures are indeed measured observationally, but only in cloud material being heated by one or more luminous stars, such as the hot cores already mentioned. This difficulty forces us to consider alternatives to the standard picture.

The environment from which massive objects arise is already packed with both stars and gas. Since low-mass star formation is highly active during this time, the gas component includes numerous dense cores. Global contraction of the background medium implies that the cores themselves become increasingly crowded. It is therefore plausible that these entities begin to touch and merge. Any such event does increase the mass of a given dense core. The new object, however, will face the same problem as before in forming a massive star through collapse. A promising alternative is that the merging entities are dense cores *already containing stars*, which themselves coalesce after their cores have joined. In this manner, we may avoid a prolonged buildup to stellar densities.

The stars in question are likely to be of low mass. They may either be accreting protostars, or else very young pre-main-sequence objects still retaining their mantles of molecular gas. Once two dense cores begin to join, the stars (and any attendant disks) will fall toward each other under the influence of their mutual gravity. This picture is reminiscent of our earlier description of close binary formation. In both cases, the stars should enter a highly elliptical orbit that decays through transfer of energy and angular momentum. The greater core mass in the present case should exert more dynamical influence and allow true coalescence of the spiraling stars, rather than formation of a close binary.

Judging from the data on clustering and mass segregation, such an initial merger may produce stars with $M_* \gtrsim 5\,M_\odot$. These intermediate-mass objects, in turn, combine to yield early-B and O-type objects. The very highest stellar masses would then arise through a rapid cascade of merging events near the very center of the parent cloud. Whether or not the process continues to such a point, this innermost region should contain a large proportion of intermediate and high-mass objects. Such also is the indication from the runaway stars that have left the area. (Again recall our discussion at the end of § 4.3.) In addition, some spiraling, massive stars do *not* complete their merger, but remain in tight orbits. Thus we may account for the doubly massive binaries observed with some frequency.

Turning this rough, and admittedly speculative, account into a bona fide theory will require a great deal of effort and ingenuity. A number of investigations have already attacked the problem of cloud merger, generally through SPH simulations. Studies of binary orbital decay are also of relevance. However, the most fruitful approach, at least for the near future, may be to focus on observational consequences of the picture. One would expect, for example, to find a rise in binary frequency toward the centers of forming clusters. Finally, the youngest massive stars may exhibit anomalous spectral or photometric features. These could be a signature of their unique origin.

Chapter Summary

A hypothetical cloud in which self-gravity overwhelms pressure support collapses promptly, breaking apart in the process. The specifics of this dynamical fragmentation are sensitive to initial conditions. Thus, a massive rotating sphere forms a ring-like configuration before breakup, while an elongated cloud first narrows to a spindle. In all cases, pressure is much more important in the daughter fragments, which therefore do not collapse in the same way.

Turning from theory to observation, most mature stars have orbiting companions. The statistical mass distribution of each component within a binary resembles that of field stars. Pre-main-sequence stars have at least as high a fraction of binaries. The prevalence of circular orbits in tighter systems demonstrates that the two stars interact tidally as they contract. Some visible young objects have infrared companions, but double infrared sources, the likely precursors to binaries in general, are still rare.

The components within a pre-main-sequence binary have similar ages and probably form within a single dense core. Since this cloud is nearly in force balance prior to collapse, it is *not* susceptible to dynamical breakup. A more likely, but still qualitative, scenario is quasi-static fragmentation, in which the cloud gradually forms two condensations that individually collapse to protostars. Once formed, these stars are initially surrounded by remnant material, including disks, that absorb energy and angular momentum from the system, driving it into a tighter orbit.

The existence of starless clumps within giant complexes argues that such clumps also evolve quasi-statically before producing stellar clusters. Even without access to the details of this process, one may predict the evolution of a cluster's luminosity function. Observations generally agree with this prediction, but also indicate that the rate of star formation accelerates in real clusters and associations. As the gas in an aging T association begins to dissipate, presumably through stellar outflows, the entire structure expands into space. A bound cluster may result when the parent cloud is more massive, so that it contracts even while losing gas. If the cloud is more massive still, the deeper contraction could lead to O and B stars arising from the merger of dense cores and previously formed stars. Once a few high-mass objects form, they quickly disperse the parent cloud, leading to free expansion of the stellar group.

Suggested Reading

Section 12.1 A classic analysis of collapse and fragmentation for pressure-free spheres is

Hunter, C. 1962, ApJ, 136, 594.

The SPH technique was introduced by

Lucy, L. B. 1977, AJ, 82, 1013

Gingold, R. A. & Monaghan, J. J. 1977, MNRAS, 181, 375.

Two reviews covering massive cloud collapse that remain useful are

Tohline, J. 1982, Fund. Cosm. Phys., 82, 1

Boss, A. 1987, in *Interstellar Processes*, eds. D. J. Hollenbach and H. A. Thronson (Dordrecht: Reidel), p. 321.

Section 12.2 The multiplicity of nearby main-sequence stars is studied in

Duquennoy, A. & Mayor, M. 1991, AA, 248, 485

Fischer, D. A. & Marcy, G. W. 1992, ApJ, 396, 178.

These focus on systems with G- and M-type primaries, respectively. Surveys of pre-main-sequence binaries include

Ghez, A. M., Neugebauer, G., & Matthews, K. 1993, AJ, 106, 2005

Reipurth, B. & Zinnecker, H. 1993, AA, 278, 81,

and the field has been well summarized in

Zinnecker, H. & Mathieu, R. D. (eds) 2001, *The Formation of Binary Stars*, (San Francisco: ASP).

Section 12.3 For a comparison of the component ages in pre-main-sequence binaries, see

Hartigan, P., Strom, K. M., & Strom, S. E. 1994, ApJ, 427, 961.

The fission hypothesis is discussed in

Tassoul, J.-L. 1978, in *Theory of Rotating Stars* (Princeton: Princeton U. Press), Chapter 11.

The gravitational stability of cylindrical clouds has been analyzed by

Nagasawa, M. 1987, Prog. Theor. Phys., 77, 635.

The interaction of young stars and their disks has been studied through numerical simulations, such as

Clarke, C. J. & Pringle, J. E. 1993, MNRAS, 261, 190.

For the tidal theory of binary orbital evolution, as applied to pre-main-sequence pairs, see

Zahn, J.-P. & Bouchet, L. 1989, AA, 223, 112.

Section 12.4 Our treatment of luminosity function evolution follows

Fletcher, A. B. & Stahler, S. W. 1994a, ApJ, 435, 313

__________________________1994b, ApJ, 435, 329.

The post-T Tauri problem was first articulated by

Herbig, G. H. 1978, in *Problems of Physics and Evolution of the Universe*, ed. L. V. Mirzoyan (Yerevan: Armenian Academy of Sciences), p. 171.

The acceleration of star formation activity in T associations is the subject of

Palla, F. & Stahler, S. W. 2000, ApJ, 540, 255.

The expansion of clusters due to mass loss was first analyzed by

Hills, J. G. 1980, ApJ, 225, 986,

while a numerical simulation of the effect is

Lada, C. J., Margulis, M., & Dearborn, D. 1984, ApJ, 285, 141.

Section 12.5 Observational studies of the most massive stars include

Massey, P., Lang, C. C., DeGioia-Eastwood, K., & Garmany, C. D. 1995, ApJ, 438, 188

Figer, D. F., McLean, I. S., & Morris, M. 1999, ApJ, 514, 202.

For the clustering around stars of intermediate mass, see

Testi, L., Palla, F., & Natta, A. 1999, AA, 342, 515.

Evidence bearing on the dynamics of the Orion Nebula Cluster is presented in

Hillenbrand, L. A. & Hartmann, L. 1998, ApJ, 492, 540.

A coalescence picture for massive star formation is advocated by

Bonnell, I. A., Bate, M. R., & Zinnecker, H. 1998, MNRAS, 298, 93.

Stahler, S. W., Palla, F., & Ho, P. T. P. 2000, in *Protostars and Planets IV*, ed. V. Mannings, A. P. Boss, and S. S. Russell (Tucson: U. of Arizona Press), p. 327.

For a contrary view, see

McKee, C. F. & Tan, J. C. 2003, ApJ, 585, 850.

Part IV

Environmental Impact of Young Stars

The Formation of Stars
Steven W. Stahler and Francesco Palla

13 Jets and Molecular Outflows

We now begin a series of chapters describing how newly created stars disturb their surrounding gas. The influence here is both mechanical and thermal. Cloud material is stirred into turbulent motion, expelled from the vicinity of a star, or heated to high temperatures. For regions that are either very dense or at a considerable distance, such activity may be the best, and indeed the only, means for revealing the presence of the stars themselves. The physical processes we will study are also of considerable interest in their own right.

One of the surprising discoveries in this field has been the disproportionate effect of low-mass objects. In the present chapter, we shall see how each such star generates, during its embedded phase, an energetic outflow extending well beyond its parent dense core. The star also emits a jet of much higher-speed gas that can travel even farther, entering regions nearly devoid of cloud material. These striking phenomena were wholly unanticipated by theorists, who are still struggling to understand the basic mechanisms of wind generation and jet propagation. We shall introduce the key concepts in both of these developing areas.

Jets are rendered visible by the shocks they produce. If the shocked gas has sufficiently high density, it may also generate beams of radiation that are intensified through the quantum phenomenon of stimulated emission. Such interstellar masers have been extensively studied, both for their intrinsic properties and for what they reveal about the dynamics of the regions producing them. Chapter 14 is accordingly devoted to this topic. Finally, we turn in Chapter 15 to the highly destructive effects of massive stars. Ionizing photons create HII regions that, along with stellar winds, disrupt entire cloud complexes. The fluorescent gas also serves as a beacon for star formation activity, both within our own Galaxy and in others far distant.

13.1 Jets from Embedded Stars

The emission-line surveys designed to locate T Tauri stars have, on occasion, yielded surprising dividends. In this section, we examine energetic flows that were first discovered in this serendipitous manner. It took several more decades, and wholly new detectors, before the remarkable, collimated nature of these stellar jets were appreciated. Observations by now have revealed very fine structural detail, which is helping to unravel the remaining mysteries in this field.

13.1.1 Herbig-Haro Objects

Around 1950, G. Herbig and G. Haro independently noted the presence in Orion of two nebulous patches bright in Hα. The two were located within the Orion A Molecular Cloud, close to the HII region NGC 1999 (see Figure 1.3), an area already known to contain many T Tauri stars. It was the unusual spectra of the patches that drew attention. In addition to Hα and the

The Formation of Stars. Steven W. Stahler and Francesco Palla

ISBN: 3-527-40559-3

other Balmer lines, the objects emitted a broad continuum and a number of forbidden optical transitions: [S II], [N II], and [Fe II], along with [O I], [O II], and even [O III]. The metastable states giving rise to such lines are collisionally deexcited even at modest densities, so the radiating medium was unlikely to be an interstellar cloud or clump. On the other hand, the presence of neutral and singly ionized species ruled out an HII region, where the level of ionization is uniformly high.

We recall from Chapter 8 that the layers just downstream from a strong shock comprise a wide range of physical conditions, as gas cools from the ionized to the atomic, and then the molecular state. It thus appeared that the newly discovered Herbig-Haro objects, of which many more were later found, might arise from the impact of high-velocity stellar winds on cloud matter. This hypothesis ultimately proved correct, but required another 25 years to verify, first by comparison of the observed emission spectra with those of known shocks (supernova remnants), and then through direct, numerical simulations. But what were the exciting stars? With the advent of infrared astronomy in the 1970s, it became clear that the objects powering the winds were *not* located within the original nebulae, but some distance away. By the early 1980s, sensitive CCD photometry had uncovered luminous strands of faint, optical emission that linked the glowing patches to their driving stars. The Herbig-Haro objects themselves were seen, in effect, to be tracers of these extended jets.

A modern image of the original HH 1/2 system from the Hubble Space Telescope encapsulates many of these developments (see Plate 9 at the end of this chapter). This photograph is actually a composite of three that were taken through a broadband red filter, and two narrowband filters isolating the Hα line at 6563 Å and the doublet [S II]$\lambda\lambda\,6716, 6731$, respectively. We see that the two Herbig-Haro objects are located on either side of their driving star VLA 1, invisible optically but indicated here by the cross. The irregular outer boundaries of the nebulae, each spanning some 10^3 AU, indeed resemble the bowshocks created when dense plugs of material plough supersonically into a more diffuse medium. The broader of the two regions, HH 2, has a more pronounced substructure. Apparently, the shockfronts are not progressing smoothly, but are in the process of breaking up, either through some inherent instability, or as a result of the clumpy nature of the background gas. Rapid evolution of the system is also indicated by its temporal variability. The optical flux from one prominent HH 2 knot has changed by 3 mag over a decade.

Little is known of the star VLA 1, which is deeply embedded in a slab-like molecular core. First discovered through its centimeter radio emission, it was located by carefully tracing back the proper motions of HH 1 and 2. These studies showed that the two shocked regions are both retreating from VLA 1 at great speed, with tangential velocities of 350 km s^{-1}. The *radial* velocities of the objects, as indicated by Doppler shifts in their emission lines, are much smaller. It follows, therefore, that the motion occurs nearly in the plane of the sky. Plate 9 shows a small, reddened jet emanating from VLA 1 and pointing directly toward HH 1; this alignment confirms the driving nature of the radio source. The less embedded Cohen-Schwartz star, situated about midway between this jet and HH 1, is thought to be unrelated.

We see in the HH 1/2 system a highly obscured and therefore presumably very young star creating a nearly symmetric *pair* of shocked regions. This *bipolarity* is a fundamental aspect of the systems containing Herbig-Haro objects. It is evident, on the other hand, that HH 1 and HH 2 have quite distinct morphologies. We may plausibly attribute such differences to spatial variation in the cloud environment over distances of order 0.1 pc. This detail should not distract us from the central feature of the bipolarity itself, a phenomenon reflecting the nature of the stellar jet that ultimately creates the emission.

Cloud matter in the vicinity of HH 1/2 has such a high column density that only the most luminous, outer shocks are visible. Additionally, the relatively faint jet pointing toward HH 1 lacks a counterpart on the other side of VLA 1. In all such "monopolar" cases, Doppler shifts in the emission lines show the visible portion of the jet to be *blueshifted*. This material is therefore advancing toward us, away from the background, obscuring dust. The redshifted portion, which is either fainter or wholly invisible optically, is retreating deeper into the molecular cloud.

13.1.2 Jet Dynamics and Morphology

Another interesting source from Orion is shown as Plate 10. Here we see the blueshifted portion of the HH 111 jet. The driving star, located below the jet base, is an infrared object (Class I) of about 25 $L_\odot$, embedded within the L1617 cloud of Orion B. At the top, there is a broader arc-like emission region that shares the bowshock appearance of HH 1 and 2. Connecting this region to the star is a series of bright knots, some of which appear to be miniature bowshocks. The average total width of the jet midsection is $0.''8$, or 370 AU at the distance of Orion.

Plate 10 actually depicts, through its color scheme, the spatial variation of the ratio of Hα flux to that in [S II]. Because of the different energies required to produce the lines, this ratio is a sensitive diagnostic of shock strength, at least for $V_{\rm shock} \lesssim 100$ km s^{-1}. Low values of Hα/[S II] occur in relatively weak shocks. Here, either the incoming speed itself is lower, or else material is entering a region that is already moving in the same direction, so that the relative velocity is reduced. Conversely, Hα emission dominates in higher-speed impacts.

The dark blue and black areas of Plate 10 indicate Hα-bright regions. These occur near the base of the jet, on the rim of each internal knot, and within the broad bowshock. Figure 13.1 also depicts the variation in shock strength through two long-slit spectra. In one of these, we have identified the main emission lines. The spectrum in the top panel was obtained by aligning the slit with the jet's bright midsection. We again see the overall dominance of the [S II] doublet to Hα emission, the characteristic signature of weak shock excitation. Conversely, Hα is stronger in the lower panel, where the slit crosses a wide bowshock occurring after the one in Plate 9.

While most jets are straight, as in this example, some exhibit strong curvature or display other interesting features. Plate 11 illustrates the variations in morphology. Each photograph includes only the blueshifted arm of the jet, although there are patches of fainter, redshifted emission in every case. The driving stars are represented by crosses near the bottom. The left panel again shows HH 111, now rendered as a composite image in Hα (*green*) and [S II] (*red*). We see how the previously depicted portion of the jet leads eventually to an even broader bowshock near the upper border of the frame; this is the structure whose spectrum is shown in Figure 13.1. On either side of the bright midsection, the jet is extremely faint, and its internal knots grow larger and more diffuse away from the star. In CCD images taken through filters that exclude the emission lines, the midsection disappears, but a faint image of the top bowshock remains.

The middle panel of Plate 11 shows the HH 46/47 system, located in the Gum Nebula, about 450 pc away. Here the driving star is only detectable in the infrared and at longer wavelengths,

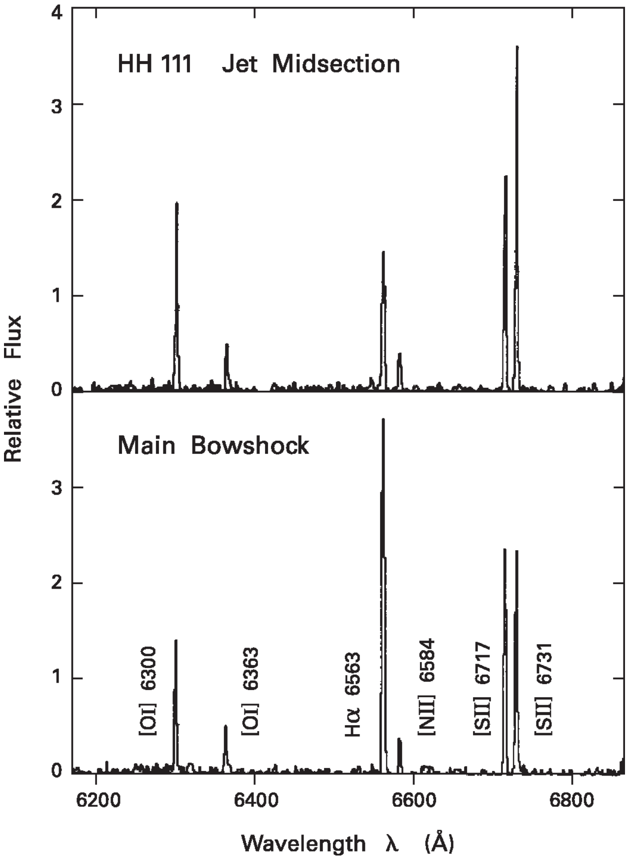


Figure 13.1 Two long-slit spectra of the HH 111 jet, with the main emission lines labeled. In the upper frame, the slit covers the jet's bright midsection, while the bottom frame corresponds to the topmost bowshock visible in Plate 11.

and has an estimated luminosity of 15 $L_\odot$. The bright, central knot (HH 47A) again has a bowshock morphology, and emits strongly in the [S II] lines. Above it is a much wider and fainter arc (HH 47D) that only appears here in Hα. The widening of successive bowshocks is reminiscent of HH 111.

Closer to the star, a broad, parabolic sheath of emission (HH 46) surrounds the jet base. The rather diffuse and featureless appearance of this region is typical of a reflection nebula. That is, much of the observed flux is not generated locally at a shock, but originates from the star, and is then scattered by dust toward the observer. This scattered emission from HH 46 alternately dims and brightens over intervals of several years. Such variations must originate in the central star. Quite generally, reflection nebulae in both the optical and near-infrared are common features near jets, where they also show up as enhancements in linear polarization.

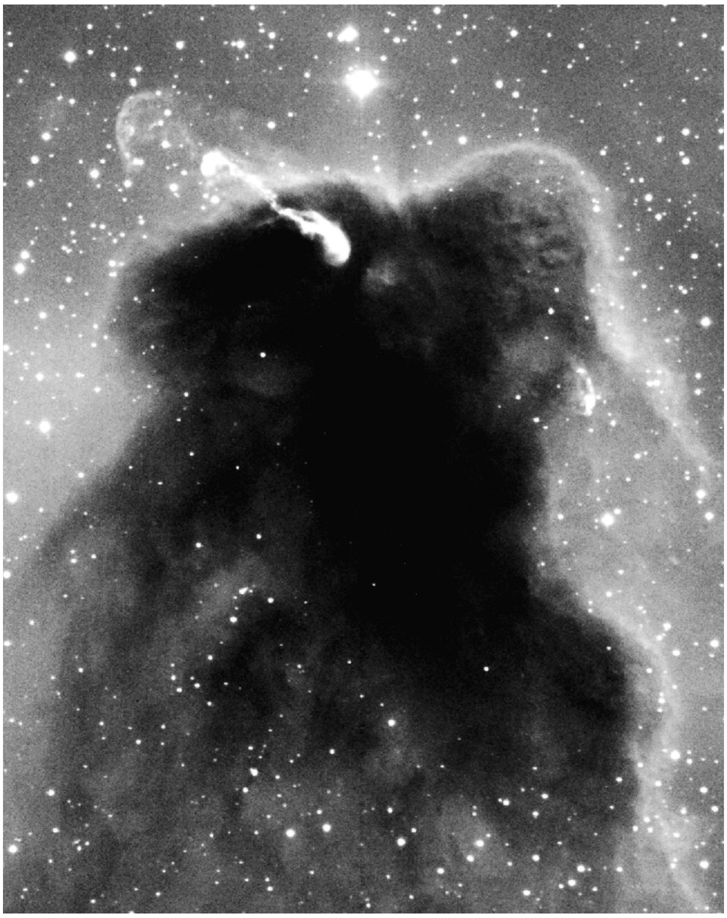


Figure 13.2 Optical photograph of the HH 46/47 jet and its parent Bok globule. The image shows emission in both [S II] and Hα. Note the multiple bowshocks in the brighter portion of the jet and the small counter-jet to the right of the globule.

Both the reflection nebula and the HH 47D arc are again visible in Figure 13.2. This black and white image, taken through a combination of [S II] and Hα filters, shows the location of the system within the star's parent Bok globule. We see how the blueshifted arm of the jet protrudes beyond the cloud edge. Aligned with this structure, but on the other side of the globule, is another isolated bowshock. This region marks the place where the receding, redshifted arm of the jet erupts from the cloud. This striking photograph should also remind us that many young stars *not* located near cloud edges may drive jets that are difficult to observe.

In the HH 34 jet, displayed as the right panel in Plate 11, the main body of the structure is remarkably narrow. There is no discernible gap between this chain of knots and the driving source, a faint, highly reddened star in the L1641 cloud of Orion A. The knots are also more tightly spaced than in the similarly spaced HH 111, while the terminal bowshock is the largest of our three examples. The apex of this structure emits in the 5007 Å line of [O III]. Here, the ionization state indicates a relatively high shock speed. The full pattern of emission lines is best matched by an impact velocity of 120 km s^{-1}, where this figure represents an average over the curved surface. By comparison, the relative line strengths in HH 47D, in which no [O III] is apparent, give a value closer to 70 km s^{-1}.

The numerical modeling for HH 34 also yields a preshock number density of 65 cm^{-3}, well below that surrounding the star. At a shock speed of 120 km s^{-1}, this preshock gas must be significantly ionized by radiation generated from the front itself. (Recall the discussion of § 8.4.) The copious emission in Hα, indicated by the bowshock's green color in Plate 11, stems from recombination of the ionized hydrogen further downstream. In contrast, modeling of the internal shocks associated with the main body of the jet indicates a much lower impact velocity, typically 30 km s^{-1}. The thin Hα rims seen in these weaker bowshocks stem from collisional ionization of largely *neutral* hydrogen encountered by the jet. Finally, we note, from Plate 11, that the terminal bowshock in HH 34 does contain isolated patches of [S II] emission. The largest of these is thought to be the *Mach disk* for the system. This is a separate, lower-velocity shock occurring behind the main front. Mach disks arise in all wind-driven shocks, and we shall explore their origin presently.

13.1.3 Spectroscopy of Bowshocks

The spectra of HH 111 shown in Figure 13.1 are valuable for discerning the relative strengths of emission lines and therefore the distribution of shock strengths. On the other hand, the wavelength resolution of the observations is too low for studying directly the internal velocities. Other instruments, such as echelle spectrometers, can provide much greater spectral resolution, albeit with a more limited total wavelength coverage. Observers may thus examine in great detail the features of individual lines.

A case in point is HH 32, a chain of knots driven by the T Tauri star AS 353A in Aquila. The system is unusual in that the driving star is optically visible. Moreover, the brightest Herbig-Haro object, designated HH 32A, is actually redshifted with respect to the stellar source. Figure 13.3 shows the Hα emission from this particular knot. Note that the telescope beam encompasses all of HH 32A, *i. e.*, the *spatial* resolution of the observation is not especially high. All of the emission occurs at positive radial velocity, so that the object is indeed retreating. The line itself is extremely broad, with a total width of about 300 km s^{-1}. The hydrogen emission occurs in postshock gas with a temperature of roughly 10^4 K. The corresponding Doppler width, from equation (E.26), is only 20 km s^{-1}. Thus, the broadened profile indicates a *range* of supersonic velocities in the emitting gas.

Another notable feature of the Hα profile is its double-peaked nature. Both this characteristic and the large total width are neatly explained if the emission stems from a curved bowshock. To see why, consider first an observer at rest with respect to the shock (Figure 13.4). Ambient material flows to the right with speed V_{shock}. After a fluid element crosses the front, its parallel velocity component $V_{\|}$ is preserved, while the orthogonal component $V_{\perp}$ drops by about a factor of 4, according to equation (F.10). The Hα emission does not occur until the postshock gas has cooled to 10^4 K. At this point, which is not far inside the bowshock, $V_{\perp}$ has dropped

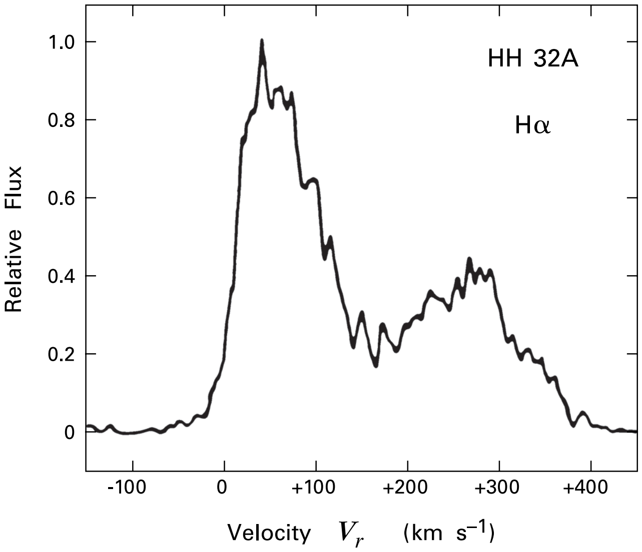


Figure 13.3 Profile of Hα emission from HH 32A, a bowshock driven by the visible star AS 353A. Radial velocity is relative to the star.

even further. The emitting gas thus has a speed close to $V_\parallel$ and is nearly moving along the sides of the curved shock front. Referring again to the figure, an element impacting at a distance ϖ from the axis is deflected by the angle θ.

Suppose now that the line of sight is straight down the central axis. Denote as V_r' the radial velocity of the emitting, postshock element, as detected by the observer. This speed is just the axial projection of $V_\parallel$, which is itself $V_{\rm shock}\ \cos\,\theta$. Thus,

$$V_r' = V_{\rm shock}\ \cos^2\,\theta\ . \tag{13.1}$$

The element in question is really part of an expanding, thin ring, whose current radius is ϖ. We need to sum the flux from all such rings to obtain the line profile. Those rings with largest surface area contribute the most. That is, the line intensity is proportional to $\Delta\mathcal{A} \equiv 2\pi\varpi\Delta s$. Here, as may be seen in the figure, the arc length Δs is given in terms of the ring width and height by $\Delta s \equiv (\Delta z^2 + \Delta\varpi^2)^{1/2}$.

The geometric contribution to the line profile is thus

$$\begin{aligned}\phi_{\rm bow} &= \frac{d\mathcal{A}}{dV_r'} \\ &= \frac{2\pi\varpi}{V_{\rm shock}}\,\frac{ds}{d\,\cos^2\,\theta} \\ &= \frac{\pi}{V_{\rm shock}\sqrt{1-\cos^2\theta}}\,\frac{d\varpi^2}{d\,\cos^2\theta}\ .\end{aligned} \tag{13.2}$$

To obtain the last form, we have also used the relation $dz/d\varpi = \cot\theta$.

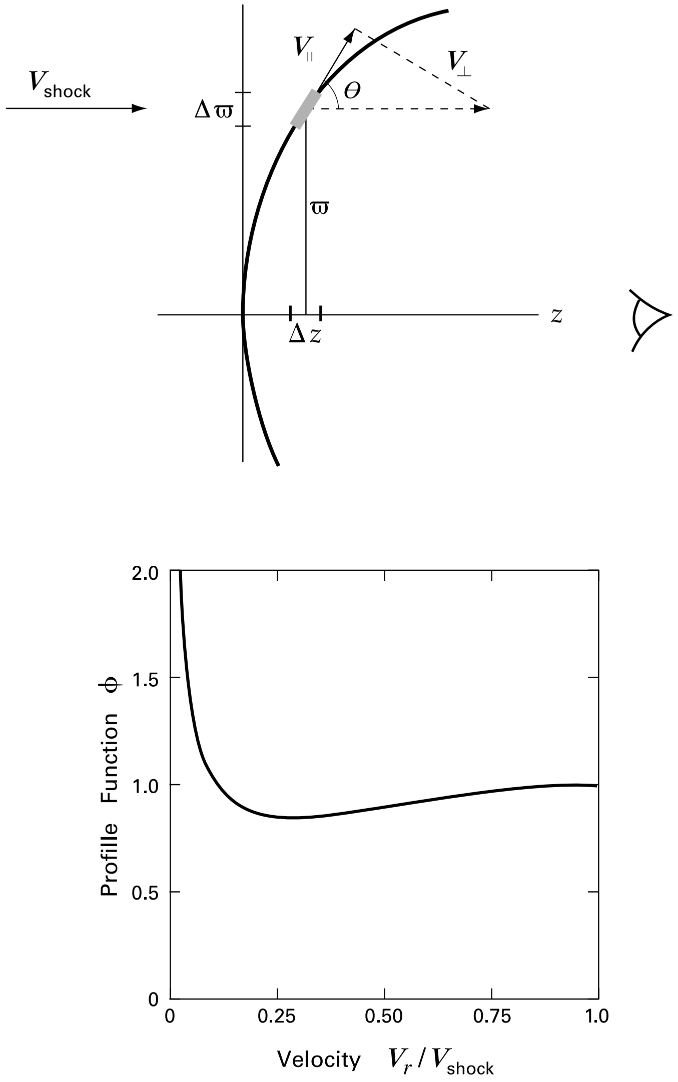


Figure 13.4 *Upper panel:* Deflection of jet ambient gas by a stationary bowshock. Only the postshock velocity component $V_{\|}$ is preserved. The bowshock arc is that described by equation (13.4). *Lower panel:* Theoretical emission profile for a line of sight down the jet axis. Here the shock is retreating from the observer at speed $V_{\rm shock}$. Note that the profile function vanishes for $V_r > V_{\rm shock}$.

The first factor in equation (13.2) indicates that the intensity rises toward the wings of the bowshock, as the corresponding surface area increases in any fixed interval of radial velocity. Evaluation of the second factor requires knowledge of the specific bowshock shape. Both hydrodynamic simulations and laboratory studies find that the the derivative falls with increasing obliquity, *i. e.*, as $\cos^2\,\theta$ increases. (The reader may verify that this property does *not* hold for a parabola, but only for more flared curves.) A simple prescription that mimics the detailed results reasonably well is

$$\frac{d\varpi^2}{d\,\cos^2\theta} \;=\; \varpi_\circ^2\,\exp(-\cos^2\theta)\ , \tag{13.3}$$

where $\varpi_\circ$ is some fiducial radius. To find the corresponding shape, we integrate this equation and obtain

$$\cos^2\theta \;=\; -\ln\left(1-\frac{\varpi^2}{\varpi_\circ^2}\right)\ .$$

Further rearrangement gives

$$\frac{dz}{d\varpi} \;=\; \left[\frac{-\ln(1-\varpi^2/\varpi_\circ^2)}{1+\ln(1-\varpi^2/\varpi_\circ^2)}\right]^{1/2}\ . \tag{13.4}$$

The bowshock depicted in Figure 13.4 is actually the curve resulting from numerical integration of this last equation. Substituting (13.3) into (13.2), we derive the line profile. To within a multiplicative constant, this is

$$\phi_{\rm bow} \;=\; \frac{1}{\sqrt{1-(V_r^\prime/V_{\rm shock})^2}}\ \exp[-(V_r^\prime/V_{\rm shock})^2]\ . \tag{13.5}$$

Let us now apply this result to the case of HH 32A. Here, the large values of radial velocity indicate that the jet axis is indeed close to the line of sight. The proper reference frame, however, is one in which the external medium is at rest, while the bowshock is retreating at speed $V_{\rm shock}$. Thus, the observed radial velocity is given by

$$V_r \;=\; V_{\rm shock}\;-\;V_r^\prime\ . \tag{13.6}$$

The bottom panel of Figure 13.4 shows the profile in equation (13.5) as a function of V_r. Note that the calculated profile plunges to zero for $V_r\;>\;V_{\rm shock}$, since $V_r^\prime$ cannot be negative. Evidently, our purely geometric considerations have captured the essential origin of the two maxima. The higher peak, as we already noted, stems from the increase of emitting area toward the wings. The true emission does not diverge for small V_r, of course, but plunges to zero when $\sin\theta\;=\;a_T/V_{\rm shock}$, *i. e.*, when the impact velocity falls to the sound speed a_T and the shock itself dies away. The second, and much weaker, maximum represents material near the apex. Rings here are of smaller area individually, but have very similar radial velocities ($V_r\;\approx\;V_{\rm shock}$). An important factor we have omitted is the change in postshock emissivity with impact speed $V_\perp$. The latter is highest at the apex, but declines by intermediate values of V_r. Comparison with Figure 13.3 indeed shows a pronounced dip in this region. Note finally that our predicted total line width is $V_{\rm shock}$ itself, which may indeed be about 300 km s^{-1} for the present case.

If we were instead to orient our observer *perpendicular* to the jet axis, the advancing and retreating elements of all rings would contribute equally to the emission. The three jets displayed in Plate 11 all happen to lie nearly in the plane of the sky. Emission lines from their principal bowshocks, while still broad, are therefore more symmetric than those in HH 32A. On the other hand, this face-on orientation does facilitate the measurement of bowshock proper motion. Using background stars as a reference, the observer compares CCD images over a number of years. The result is a snapshot of the system's transverse velocities.

13.1.4 Motion of the Knots

What one actually observes through this technique is the proper motion of individual emission knots. These are sites of shock excitation and may not follow the bulk motion of the underlying fluid. The shocks could even be stationary, if they arose wherever the fluid encountered clumpy, inert obstacles in its path. However, any such picture is ruled out by the observational studies, which typically find proper motions of several hundred km s^{-1}. These speeds, we shall see, are comparable to those in stellar winds. It is at least plausible, then, that the knots are indeed tracing the jet flow.

Returning to HH 46/47, one finds that the proper motion vectors lie along the jet's sinuous path. There is a gradual increase of speed from the base to the HH 47A bowshock, followed by a sudden drop and then a further rise before HH 47D is reached. Interestingly, the *radial* velocities, as obtained spectroscopically, exhibit a similar pattern. The latter are also systematically lower in magnitude, as one would expect in a nearly face-on orientation. Adding the two velocity components in quadrature, we find that the full, three-dimensional velocity is changing along the jet path. Thus, the alterations in radial motion are *not* simply a geometric effect, as would arise from bending of the structure in and out of the plane of the sky.

The case of HH 34 reinforces this conclusion. Here, the jet is manifestly straight, so that any change in projection angle should be minor. Yet the radial velocity steadily climbs outward. The higher, tangential speeds have the same trend, reaching some 300 km s^{-1} at the main bowshock. Figure 13.5 displays the proper motion vectors near the jet head. The isolated knots toward the bottom are too faint to be seen in Plate 11, but have relatively high velocities, oriented nearly along the main axis. Closer to the wide bowshock, there is evidently more scatter in both the direction and magnitude of the vectors. The pattern becomes clear in the righthand panel, which shows the same velocities *in the frame of the jet*, *i. e.*, after subtracting off a vector representing the mean, axial flow. With few exceptions, the knots are following the curved sides of the front, just as predicted by theory.

The reader may have noticed a discrepancy between the very large space velocity for the main HH 34 bowshock and the more modest shock speed deduced from the pattern of emission lines. Counting the radial component, the former is 330 km s^{-1}. The estimated shock speed at the apex, corresponding to $V_{\rm shock}$ in Figure 13.4, is 140 km s^{-1}, as gauged both from the relative strengths of the lines and from the widths of their profiles. The difference is even greater for the jets' internal knots. These also have spatial velocities of several hundred km s^{-1}, but values of $V_{\rm shock}$ lower by an order of magnitude.

Such reduced impact speeds mean that the shocks within HH 34 are not traveling into a stationary medium, but one already moving in the same direction. The shocks arise when a

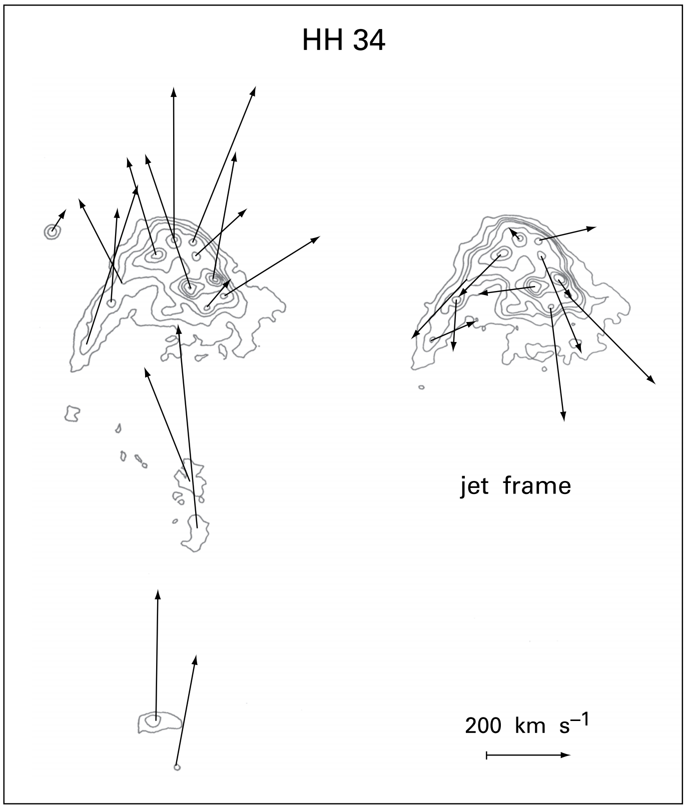


Figure 13.5 Observed proper motion vectors near the main bowshock of the HH 34 jet (HH 34S). In the jet frame (*right side*), the vectors lie nearly along the bowshock surface.

fluid element overtakes one ahead of it with smaller velocity. Apparently, the fractional velocity difference is only of order 10 percent in the main body of the jet. Such relatively weak shocks travel nearly at the sound speed in the fluid's rest frame. Since the bulk speed is highly supersonic, the knots indeed closely trace the jet's velocity. The change in velocity is much greater at the large bowshock, where the medium has a speed of $330 - 140 = 190$ km s^{-1}. As we shall discuss later, the spatial motion of the front in this situation depends on the density contrast between the pre- and postshock gas.

To summarize, a picture in which the velocity changes along the jet path fits well with both proper motion and radial velocity data. Most systems, including HH 46/47 and HH 111, display anomalously low shock speeds and so may be understood within this framework. (One notable exception is HH 32, where we have seen that $V_{\rm shock}$ is 300 km s^{-1} for the brightest knot.) What, then, is the cause of these velocity changes? Do they originate from interaction of the flow with the cloud material outside it? Or are they inherited from temporal variations in the driving star itself?

Although theorists have offered a variety of calculations addressing this issue, the observations themselves provide a convincing answer. Figure 13.6 is a near-infrared image of HH 212,

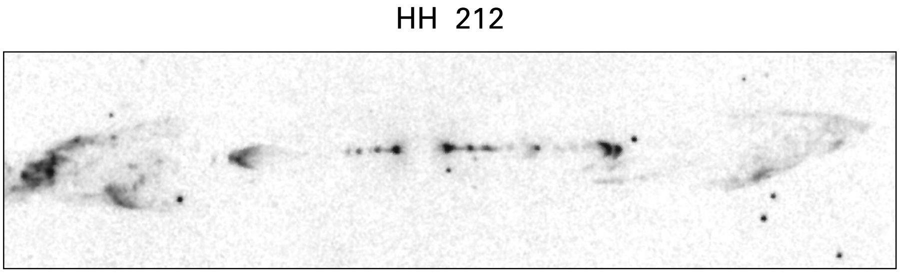


Figure 13.6 Near-infrared (K-band) image of the HH 212 jet. Note the mirror symmetry displayed by the innermost knots.

a jet system in the L1630 region of Orion B. The large, terminal bowshocks, separated by 0.6 pc, resemble those of HH 1/2. As in that case, the morphological difference between the two fronts reflects their propagation through cloud environments that vary on this length scale. A key distinction from HH 1/2, however, is the presence of a tight sequence of knots closer to the invisible, central star. These structures are remarkable for their high degree of symmetry. On either side of the central star, the chain is irregularly spaced, but essentially the mirror image of the other. The only plausible explanation for this circumstance is that both sequences were created by the driving star in an identical manner. Thus, the knots must arise from velocity (or, more generally, momentum) fluctuations in the underlying stellar wind.

We see that the internal structure of a jet provides a record of the wind activity powering it. Let us return to the example of HH 111. Plate 10 shows that the main body of the jet contains knots spaced quasi-periodically. Here, the typical separation is 3000 AU. For an average, measured space velocity of 300 km s^{-1}, the low-amplitude fluctuations creating these shocks must recur over an interval of 40 yr. Notice that several of the individual arcs are directed off the main axis. The implication is that the stellar wind altered its direction slightly while its velocity was changing. The large bowshock terminating the main jet is 0.1 pc from the infrared source. This structure was created during a larger velocity increase some 300 yr in the past. Finally, the even wider arc seen in Plate 11 is 0.3 pc from the star, and moving at 400 km s^{-1}. The event creating it must have occurred 500 yr before that forming the main bowshock in Plate 10.

The emerging picture is that the anisotropic winds powering jets undergo relatively minor variations in both speed and direction on a time scale measured in decades. Averaging over these short-term fluctuations, the velocity still rises or falls systematically, changing in amplitude by roughly a factor of two over a period of order 10^3 yr. Note that the sharp discontinuities in speed across the present-day bowshocks need not have been present in the material actually leaving the star. Even smooth changes in the highly supersonic wind velocity would steepen naturally in the course of propagation.

13.1.5 Giant Flows

The examples shown in Plate 11 make it clear that the smaller, internal knots fade relatively quickly, *i. e.*, over a few centuries. Larger bowshocks, on the other hand, are remarkably persistent and can be seen a considerable distance from the driving star. Beginning in the early

1990s, observers availed themselves of CCD detectors covering relatively large fields of view, approaching 1° on a side. The resulting images have revealed chains of Herbig-Haro knots that extend for several parsecs beyond the brighter, more continuous portions of the jets we have examined thus far. These *giant flows* trace the stellar wind history over much greater periods of time.

Figure 13.7 displays the complex of Herbig-Haro objects associated with the HH 34 jet. Toward the center of this optical photograph is the embedded driving star, HH 34/IRS. The short segment protruding from the star toward the south is the jet shown in Plate 11. We see again how the jet leads, after a break, to the "terminal" bowshock, here labeled HH 34S. Directly opposite the star is an analogous, but fainter structure, HH 34N, created by the invisible counter-jet. These bowshocks constitute only the innermost pair in a long sequence, ending with HH 88 to the north and the southern HH 33.

The remarkable alignment of the emission knots, stretching over 3 pc, makes it clear that they are physically related. Velocity measurements confirm the connection. Spectra of the northern objects reveal that they are all redshifted, while the southern knots are blueshifted. The same measurements also demonstrate, incidentally, that such features as the large arc north of HH 34/IRS are *not* part of the system. Proper motion studies find that each bowshock in the chain is moving away from the star in the expected direction.

Mapping of the region in CO indicates that the ends of the flow project into regions largely devoid of molecular gas. What, then, creates the shocks themselves? As before, the answer is that the shocks arise when gas overtakes slower *jet* material ahead of it. It is interesting that the radial velocities associated with these shocks decline systematically away from the star. The falloff is nearly identical on the two sides and amounts to a factor of three from the central star to the final knot.

Other patterns are also noteworthy. The linear size of each Herbig-Haro object increases outward. It may be that the jet is unconfined laterally, so that shocks are spreading out in a purely ballistic manner. Additionally, structures close to the driving star exhibit the characteristic bowshock morphology: a well-defined apex pointing along the main flow and wings that sweep back toward the star. Farther out, the emission regions lose their bowshock appearance and become more complex, with numerous internal subcondensations and filaments.

These characteristics are all seen, with more or less clarity, in the several dozen giant flows discovered thus far. Some jets protrude into regions so evacuated that external galaxies may be seen in proximity. The HH 34 flow in Figure 13.7 has an S-shaped configuration, a feature that is also common. In these cases, it is apparent that the bipolar jet leaving the star has undergone a change in direction. Similar examples include the complex driven by the highly variable star PV Cephei and that associated with the Herbig-Haro object RNO 43 in Orion. The chains leading from the HH 111 jet, on the other hand, stretch a full 8 pc from end to end and yet show little bending.

The measured velocities and spacing of the shocks in the HH 34 complex imply that the major alterations in wind speed recurred every 800 yr or so. Here, the estimate is necessarily crude, as the observations give us only the bowshock speed and not that of the underlying jet material. If the latter maintains a velocity of 300 km s^{-1} out to HH 88 and HH 33, then these regions were created 5×10^3 yr in the past. This time would be about twice as long if the jet speed is closer to that of the bowshock and declines outward. In any case, the interval is still

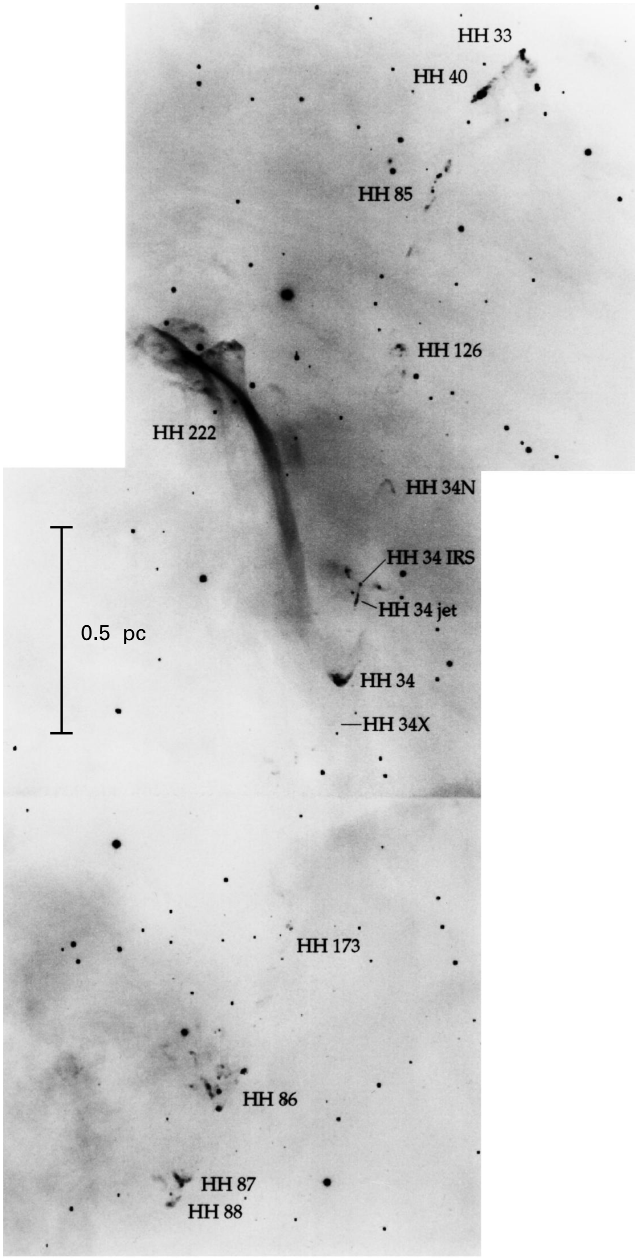


Figure 13.7 Giant Herbig-Haro flow associated with the HH 34 jet. The main bowshocks are labeled. The short, contiguous jet pointing toward the south (bottom of the page) is the one shown in Plate 11. Note the prominent arc labeled HH 222; this is not associated with the HH 34 jet.

brief compared to the evolutionary time scale of the driving star. Thus, the systematic trends we have noted must reflect intrinsic alteration of the flow itself, rather than any secular change in the driving source. For example, the complexity of the outer emitting regions is reminiscent of that seen in HH 1/2 (Plate 9) and plausibly arises from the intrinsic tendency of the shock front to break up through rapid cooling. Indeed, HH 1/2 itself drives a giant flow, in which the terminal regions are even more extended and fragmentary than any shown here. We shall return to these issues in § 13.5, when we discuss jet propagation from a more theoretical perspective.

13.1.6 Radio Observations

Our treatment of jets has thus far covered only their optical manifestations. Suppose, however, that we wish to study the region closer to the driving star. Then we must turn to wavelengths that can penetrate the intervening dust. We already saw one example, in Figure 13.6, of how near-infrared observations help in this regard. Another fruitful regime is the radio continuum. It is true that the total radio flux emitted by jets is much smaller than in the optical. On the other hand, interferometers such as the Very Large Array offer very high angular resolution, of order $0.''1$, at centimeter wavelengths. Let us close this section with a brief look at findings in this area.

A large fraction of optical jets and Herbig-Haro objects exhibit radio emission from the vicinity of their driving sources. Indeed, the extraordinary spatial resolution of these observations frequently provides the most accurate location for the star itself. We noted earlier the case of HH 1/2, where the central object was first identified through radio studies, in conjunction with proper motions. Centimeter maps of other systems have picked out the flows' true sources among a tight grouping of infrared stars.

Let us return to the example of VLA 1, the embedded object powering HH 1 and 2. Figure 13.8 is a map of the region at 6 cm. Here we have also indicated the effective beam area of the telescope. The map is oriented so that the line joining the centers of the Herbig-Haro objects is horizontal, with HH 1 off to the right. This line also runs through the faint optical jet visible in Plate 9. It is apparent that the radio-emitting region is elongated in the same direction. The true aspect ratio of the area cannot be discerned from this image, since the width is still unresolved. Assuming a distance of 460 pc, the physical length of the configuration is 390 AU. Jet material traveling at 350 km s^{-1}, the measured speed of the Herbig-Haro objects, thus crosses the region in under 3 yr.

Other Herbig-Haro complexes, including HH 111 and HH 34, have radio jets. What physical mechanism underlies this emission? We saw in Chapter 3 how warmed dust grains provide continuum radiation. (Recall Figure 3.16.) Here, the spectral energy distribution peaks at millimeter or infrared wavelengths, depending on the dust temperature, and would be negligible in the centimeter regime. Another possible source could be radiation from electrons gyrating about a local magnetic field. The required field strengths, however, vastly exceed any observed near young stars.

The remaining, and most plausible, mechanism is thermal radiation from the electrons in an ionized plasma. This *free-free emission* stems from impulsive acceleration of the electrons as they pass close to ions and is identical to the bremsstrahlung creating X-rays in visible T Tauri stars (Chapter 7). In that context, we noted that the mean intensity is nearly constant up to the

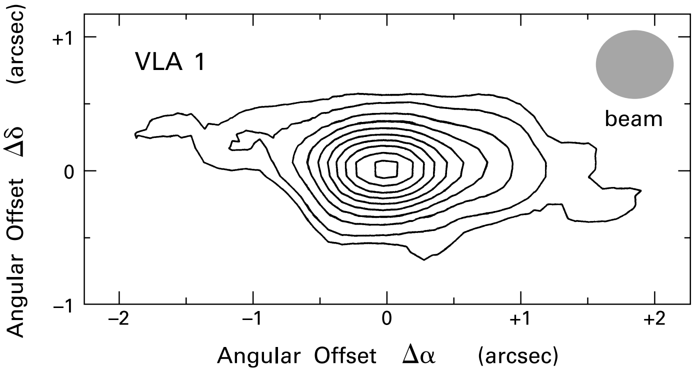


Figure 13.8 Radio interferometric map at 6 cm of the VLA 1 jet. The effective area of the beam is shown to the upper right.

cutoff frequency $\nu \approx k_B T/h$, where T is the gas temperature. For a plausible jet temperature of 10^4 K, this cutoff is in the optical range, so the radio flux should be almost independent of frequency. This is observed to be true of all sources. In the case of HH 1/2, for example, the total flux integrated over the central jet increases by only a factor of two from 2 to 20 cm.

How, then, does the star create the requisite ionization? Most sources of radio jets (also called *thermal* jets) have bolometric luminosities under 100 $L_\odot$ and are thus expected to be low-mass objects. Here, we must insert the significant caveat that these stars are too embedded for any direct assessment of their evolutionary status. Whether they are in the protostar or pre-main-sequence phase, however, the ultraviolet component of their emitted spectra would be relatively small. In particular, we do not expect significant photoionization of ambient gas out to 100 AU. The ionization must rather be created locally, through shock impact of supersonically moving gas.

For corroboration of this idea, we turn once again to our prototype system. Both HH 1 and 2 exhibit their own radio emission, with spectral characteristics similar to that from the central jet. Here, the flux is associated with the brightest visual knots in the Herbig-Haro objects and shares their high proper motions. Since the optical radiation is undeniably shock-generated, the radio emission plausibly derives from the same mechanism. The observations near VLA 1 indicate that any strong shock at the base of the jet must begin within a few tens of AU from the star. We will see in § 13.5 how a stellar wind may indeed encounter such a shock as it narrows to form a propagating jet.

13.2 Molecular Outflows

A jet represents gas from the star itself that burrows through the surrounding molecular cloud and even projects into the sparse region beyond. It is remarkable that the relatively narrow structure maintains itself out to such distances, and the discovery of jets is a triumph of CCD imaging. Even with a sensitive array detector, however, the brightest inner portion of a jet cannot be seen unless the driving star is close to the edge of its dense core. This may happen because the object migrates there, or because the cloud face has been eroded by nearby, massive stars. Such erosion, and the accompanying ionization, are evident in the illuminated rim of the Bok globule containing HH 46/47 (Figure 13.2).

Of course young stars that are not so fortuitously located still emit jets, whose presence must now be inferred from their effect on ambient matter. Molecular outflows are the turbulent expanses of cloud gas being pushed out from the vicinity of a stellar source. No special viewing circumstances are necessary to find them, as even the most deeply embedded object stirs up its cloud over a considerable distance. Thus, while the total number of contiguous optical jets is only a few dozen, the tally of molecular outflows is well into the hundreds.

13.2.1 Bipolar Lobes

The first outflow was discovered in 1980, within the L1551 dark cloud of Taurus-Auriga. Over the preceding decade, radio observers had been using the 2.6 mm line of CO to investigate the structure and kinematics of such regions. Evidence for rapid motion near young stars was already clear. The new flow in L1551, however, had a distinctive morphology not encountered before. The reader should turn back to Figure 1.13, which is a $^{12}C^{16}O$ map of the area. Notice how the two lobes are well separated spatially, and how their common axis runs through the central source IRS 5. Subsequent observations found that this source is a binary, with a total luminosity under 30 $L_\odot$. This fact, too, marked a change, as wind-generated motion had been exclusively associated with much more luminous stars.

The original outflow detection, as well as all subsequent ones, have been made through examination of CO line profiles. Figure 13.9 shows a portion of the data from L1551/IRS 5. The upper panel shows the $^{12}C^{16}O$ ($J = 1 \rightarrow 0$) emission at a central position within the lefthand lobe in Figure 1.13, while the lower panel is from the other side of the outflow. Both profiles have sharp maxima near $V_r = +6.7\,\mathrm{km\,s^{-1}}$, the average radial velocity of the L1551 cloud. Most of this peak emission stems from ambient gas either in front of or behind the lobes themselves. This material exhibits no systematic motion, but rather a more random velocity field with a typical speed of several km s^{-1}, judging by the width of the intensity spike.

Other gas, however, is moving in a directed fashion and at a speed well above the environmental level. This component shows up as the extra hump, or “pedestal” feature in each profile. The reversal of the pedestal relative to the main spike indicates that the motion itself changes direction. Thus, the leftmost lobe consists primarily of redshifted gas retreating into the plane of the sky, while the right lobe is mainly blueshifted. Of course, these radial velocity measurements capture only a part of the full spatial motion. One’s visual impression from the map is that gas is moving out from the central star. Support for this view comes from observations of two Herbig-Haro objects within the blue lobe. These optical patches have proper motion vectors that indeed point away from IRS 5.

The L1551 cloud itself is a relatively compact structure, several parsecs in extent, and contains several other active, young stars. The combined length of the outflow lobes from IRS 5 exceeds 1 pc, and each lobe appears to terminate just where the ambient density is falling off. Thus, the volume of the parent cloud occupied by this single outflow, while not large, is appreciable, roughly 6 percent of the total. Converting the CO measurements into column densities, one finds that the lobe masses sum to 3 $M_\odot$, a comparable fraction of the value for the cloud as

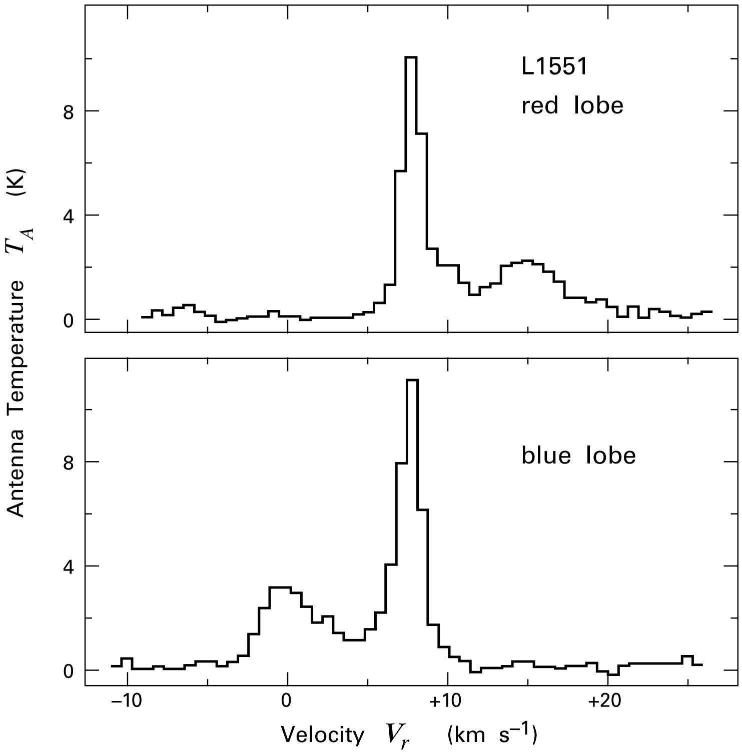


Figure 13.9 Profiles in $^{12}C^{16}O$ $(J = 1 \rightarrow 0)$ of the L1551/IRS 5 molecular outflow. The upper panel shows the redshifted lobe, while the lower one corresponds to the blueshifted portion.

a whole. It follows that the mean volume density of outflow material, about 800 cm^{-3}, is also similar to that of the environment. This agreement is no accident, but underscores the fact that outflows are indeed cloud gas, set into motion by a stellar jet buried within.

Observers have by now used radio telescopes, both single-dish instruments and interferometers, to map high-velocity CO in dozens of star formation regions. It appears that virtually all embedded stellar objects power molecular outflows at some level. The L1551/IRS 5 system may be regarded as fairly representative, but the observed properties span a broad range. Many outflows, for example, have lobes that are not as well delineated. In some cases, the outflow axis is directed more nearly along the line of sight. Here the telescope beam incorporates both red- and blueshifted material, and the two lobes become difficult to identify. In other cases, the system is favorably oriented, but the lobes flare out more broadly or less smoothly from the central star. A lobe may also terminate abruptly in places where the ambient cloud gas is depleted, perhaps by action of neighboring stars. In especially crowded regions, one observes numerous fragmentary shells and filaments of CO emission that cannot easily be traced back to their driving sources.

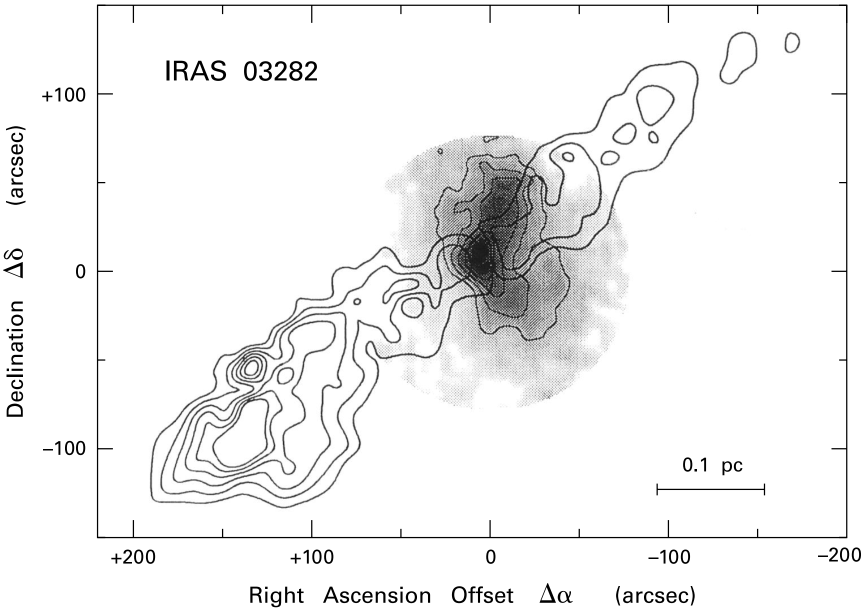


Figure 13.10 Map of the IRAS 03282 outflow in Perseus. The heavy contours represent emission in the $J = 2 \rightarrow 1$ line of $^{12}C^{16}O$, while the grey-scale image shows NH_3 $(1, 1)$ emission. Angular offsets are measured from the position of the infrared star.

13.2.2 Narrow Outflows

There are also flows narrower than the one seen in L1551. These tend to be produced by stars more obscured than the binary IRS 5, a Class I source. Figure 13.10 shows, as an example, the IRAS 03282 outflow in the Perseus Molecular Cloud. Here the exciting star, designated IRAS 03282+3035, lies in a molecular ridge near B1, the region whose magnetic structure we previously examined (Chapter 6). The spectral energy distribution of the low-luminosity star is similar to that of L1448 mm (Figure 4.4) and B335 (Figure 11.25), and identifies it as a highly embedded, Class 0 object.

The heavy contours in Figure 13.10 represent emission from the $J = 2 \rightarrow 1$ line of $^{12}C^{16}O$. Note that the map covers radial velocities less than 0 and greater than +14 km s^{-1}, while the mean velocity of the ambient gas is +7.1 km s^{-1}. Thus, only the lobes themselves show up. These have a total length of 0.8 pc and are again cleanly separated kinematically, with the left lobe now being blueshifted. The average (*i. e.*, mass-weighted) radial speed is 20 km s^{-1}, several times that of L1551/IRS 5. In fact, the speeds measured in CO attain values as high as 70 km s^{-1}. This rapidly moving gas is located close to the central axis, within a strip no wider than 0.02 pc.

Superimposed on the CO contours is a grey-scale image of the NH_3 (1,1) emission, along with several intensity contours. We see that IRAS 03282+3035, like other Class 0 and I sources, is situated within a dense core. This latter object has the familiar properties: a peak density near 10^5 cm^{-3}, a diameter between 0.1 and 0.2 pc, and a gas temperature close to 10 K. Notice how the high-velocity CO emission is at a *minimum* close to the star, just where the NH_3 peaks. Thus, the jet emitted by the central object stirs up ambient gas in a progressive manner. This physical influence, manifested as the CO outflow, gradually builds to a maximum well beyond the borders of the dense core.

Another noteworthy feature of the molecular outflow is its clumpy appearance. We may readily imagine such a non-uniform matter distribution arising, at least in part, from agitations produced by the central wind. In the case of L1551/IRS 5, much of the blueshifted lobe contains diffuse optical light. This relatively faint emission is linearly polarized, and the pattern of electric field vectors is symmetric about IRS 5. The entire lobe, in effect, serves as a giant reflection nebula. For this to occur, some visible photons from the star must be able to penetrate a considerable distance into the outflow region, before being scattered by dusty clumps.

Clumping is also indicated by the actual brightness of the CO emission. In a number of cases, it is possible to observe a system simultaneously in two isotopes of the molecule. Using the techniques outlined in Chapter 6, one determines both the excitation temperature $T_{\rm ex}$ and the optical thickness $\Delta\tau$. The former is usually close to the expected gas kinetic temperature of 10-15 K, at least for the broader outflows. As for $\Delta\tau$, both the $J = 1 \rightarrow 0$ and $J = 2 \rightarrow 1$ lines of $^{12}C^{16}O$ are generally optically thick at velocities not far from the ambient value. According to equation (6.1), the received brightness temperature, T_B, should nearly equal $T_{\rm ex}$ under these circumstances. Figure 13.9 shows that the *antenna* temperature for L1551/IRS 5 has the expected magnitude at its peak, but falls quickly below this on either side. Such dilution is almost certainly due to clumpiness of the emitting material (see equation C.12). Finally, we note that higher-resolution CO mapping of IRAS 03282 reveals a correlation between the degree of clumpiness and the outflow velocity. The fastest-moving gas mentioned earlier is confined to discrete parcels, or "bullets," situated close to the axis. These fragments, measuring only 0.01 pc in size, are grouped in symmetrical pairs about the origin. Thus, they plausibly arise from traveling pulses in the central jet.

Plate 12 shows an even more extreme case of a high-speed, narrow outflow, as imaged by an interferometer. Here we see the HH 211 system, buried within the IC 348 cluster of Perseus. The driving star has a total luminosity near 5 $L_\odot$ and has been detected only at far-infrared and millimeter wavelengths. White contours trace the $J = 2 \rightarrow 1$ line of $^{12}C^{16}O$, for V_r less than +2 km s^{-1} and greater than +18 km s^{-1}. Since the background material has a radial velocity of +9.2 km s^{-1}, we are again focusing on the fastest, innermost portion of the outflow. Note the relatively small extent of this structure, all of which would fit inside the dense core at the center of Figure 13.10.

It is clear that the molecular gas in the HH 211 outflow is quite close to the stellar jet and therefore strongly influenced by it. The red contours in this image represent continuum emission at 1.3 mm. Within this region, some 700 AU in diameter, dust grains are being warmed by the shock interaction. Even more dramatic manifestation of shocks occurs at the ends of the CO lobes. Here, the green patches indicate emission in the $1-0\ S(1)$ line of H_2, at 2.12 µm. This line is produced by gas with a temperature of roughly 2000 K. The morphology in the present case suggests that the hot gas traces bowshocks created by the jet. These are regions which would have been optical Herbig-Haro objects in a less dense environment. The jet speed must drop just ahead of each bowshock, accounting for the abrupt cutoff in high-velocity molecular gas.

13.2.3 Velocity Nesting

If we were to include lower-speed CO in Plate 12, we would find that it forms a thick sheath around the lobes displayed here. Such velocity nesting is not limited to the narrow structures produced by Class 0 sources, but is a universal feature of molecular outflows. The falloff in speed away from the central axis reflects the diminishing influence of the jet itself. To see the effect most clearly, we utilize *channel maps*, *i. e.*, images tracing the CO emission within narrow slices of radial velocity. Consider again the case of L1551/IRS 5. Figure 13.11 is a sequence of $^{12}C^{16}O$ channel maps of the blueshifted lobe, with the velocity interval noted in each panel. Although the elongated morphology is no longer apparent, there is still a well-defined kinematic structure. Material with radial speed close to the background value of +6.7 km s^{-1} occupies the space just within the lobe boundary, while the high-speed gas concentrates within progressively smaller interior regions.

The total CO emission observed in each channel of higher velocity diminishes rapidly. To compensate for this effect and reduce the total dynamic range, the relative brightness of the smaller maps of Figure 13.11 has been artificially enhanced. Suppose now that one's observations covered *only* the most intense, lowest-velocity component within an outflow. The resulting image would resemble the upper left panel of the figure and would give the impression of an expansive, shell-like structure with a nearly evacuated interior. Such indeed was the original interpretation of the CO observations. The idea was that the hollow interior was inflated by a broad stellar wind pushing out the supposed molecular shell. We now understand that this volume is actually filled with clumpy gas moving at higher speeds. The stellar wind itself is mainly confined to the narrow, central jet, as evidenced by the patches of shocked-induced emission often found along the lobe's axis. Finally, it should be stressed that the velocity of each cloud element, discounting the turbulent contribution, is directed *parallel* to this axis and has no significant component normal to the lobe boundary.

Besides occupying a smaller volume, the more rapidly moving gas also tends to be farther from the driving star. The effect is evident in the panels of Figure 13.11, where the coordinate origin marks the location of IRS 5. We may also explore this aspect of outflow structure through a *position-velocity diagram*. The CO spectrum at a given location is a two-dimensional plot of specific intensity (or antenna temperature) versus radial velocity V_r. The ensemble of spectra in a well-defined outflow gives the intensity as a function of three independent variables: V_r and two spatial coordinates. Suppose now we consider observations taken only along a specified line, such as the flow's central axis. Under this restriction, we may picture the intensity as a surface in three-dimensional space. Along the surface run curves of fixed elevation; these are the iso-intensity contours. By selecting a number of contours and plotting them in two dimensions, we create a position-velocity diagram.

Figure 13.12 offers two examples of this useful visualization technique. To the left is the now familiar L1551/IRS 5, plotted using the $J = 2 \rightarrow 1$ line of $^{12}C^{16}O$. The righthand panel shows $J = 1 \rightarrow 0$ data from the VLA 1623 system in ρ Ophiuchi. The latter is driven by a

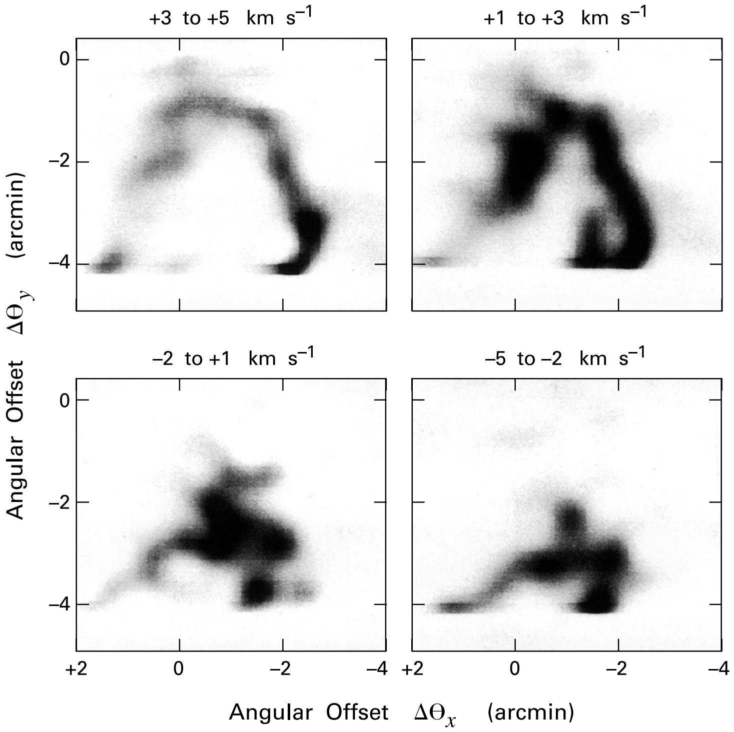


Figure 13.11 Channel maps of the blueshifted lobe within the L1551/IRS 5 outflow, constructed from observations in $^{12}C^{16}O$ ($J = 2 \rightarrow 1$). The velocity intervals are shown for each panel. Angular offsets are relative to the exciting source IRS 5.

Class 0 source of modest luminosity. In both diagrams, the vertical axis measures position along the outflow's central axis. The fan-shaped pattern of iso-intensity contours is nearly mirror-symmetric about some axial position, which marks the location of the driving star. For positions on either side of the star, the gas is predominantly red- or blueshifted. This separation again occurs because the outflows both lie nearly in the plane of the sky.

Also evident in each diagram is a thick, vertical ridge. Here the contours representing the strongest CO emission pack together. (For clarity, the severe crowding has been omitted from the VLA 1623 diagram.) The ridge occurs at that radial velocity corresponding to the parent cloud's bulk motion. Recall that the intense emission here stems from gas behind or in front of the outflow lobes themselves. The width of the ridge is thus a measure of the velocity dispersion in the undisturbed cloud environment.

The most intriguing aspect of the diagrams is the *spreading* of the contours. With increasing distance from the driving source, both the range of velocities and the maximum observable V_r initially rise. Eventually, however, the maximum velocity peaks and falls, so that the contours retreat back to the ridge. This transition, which evidently may be sudden or more prolonged, marks a general falloff in the CO intensity. That is, the axial position of peak velocity corre-

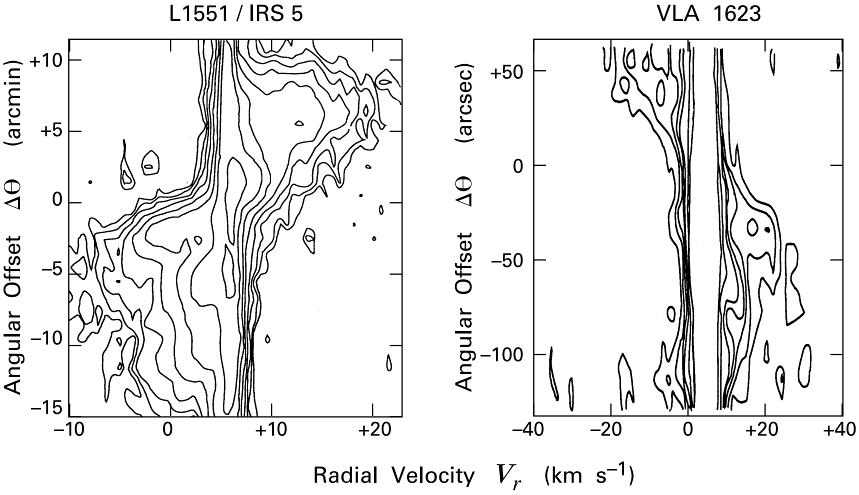


Figure 13.12 Position-velocity diagrams of the L1551/IRS 5 and VLA 1623 outflows. Both were constructed from observations in $^{12}C^{16}O$, with the first diagram using the $J = 2 \rightarrow 1$ line and the second $J = 1 \rightarrow 0$. Emission close to the background velocity has been omitted for VLA 1623. Note the very different linear scales for the two panels.

sponds roughly to the lobe boundary in conventional maps. Note finally how the velocities are uniformly smaller in L1551/IRS 5 than in VLA 1623, which is predictably a more vigorous system.

What we have found, in effect, is that outflows *accelerate* outward from the stars that generate them. Clearly, the molecular gas is *not* simply coasting on its own momentum, as would the material in an unconfined stellar wind. Nor is it decelerating by sweeping up ambient gas, as in an expanding shell. The acceleration is rather a sign that cloud material is being continually pulled forward by the underlying jet.

We mentioned earlier that most gas within a lobe is moving relatively slowly. Conversely, the high-speed component represents a minor fraction of the perturbed cloud gas. It is instructive to quantify this trend, *i. e.*, to measure dM/dV_r, the amount of outflowing mass allotted to each velocity slice. In practice, one must examine in detail each channel map, converting the specific intensity at each location into a gas column density. The latter process requires some care in accounting for the optical depth of the CO line.

Figure 13.13 shows the results of this exercise. The left panel includes both L1551/IRS 5 and the NGC 2071 outflows. The last system originates near the reflection nebula of the same name in the L1630 dark cloud of Orion B. Near the center of the flow is a knot of infrared sources, of total luminosity 600 $L_\odot$, which also power compact HII regions. The actual driving star is thus presumably more massive than the ones we have encountered so far. Morphologically, the outflow is at least as broad as L1551/IRS 5 and has extensive 2.12 μm emission near its central axis.

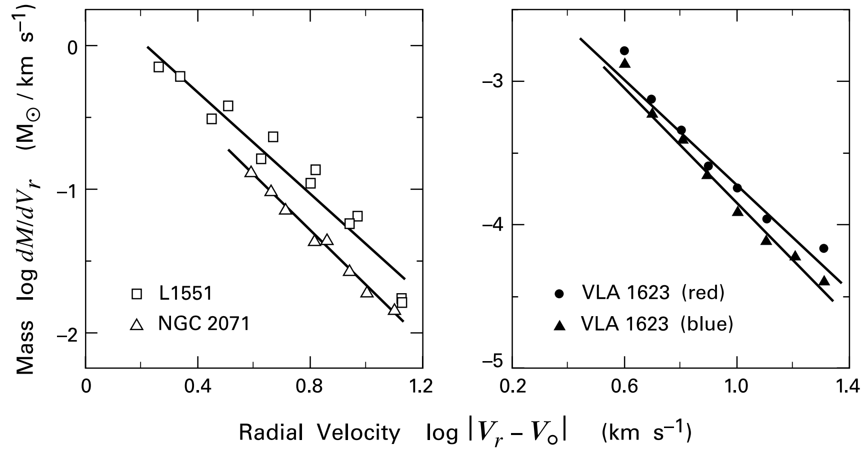


Figure 13.13 Total mass plotted as a function of radial velocity, shown for L1551/IRS 5, NGC 2071, and VLA 1623. All plots display approximately power-law behavior, with an exponent close to -2.

In the right panel of Figure 13.13 we display separately the red- and blueshifted lobes of VLA 1623. Note again the change in both mass and velocity in going from the broader outflows to this Class 0 system. Despite this difference, all three plots are strikingly similar. They reveal power-law behavior:

$$\frac{dM}{dV_r} = k\left(\frac{V_r}{V_j}\right)^{-\gamma} , \tag{13.7}$$

where the constant k evidently differs from one system to another. Both panels show least-squares fits to the data, from which we conclude that $\gamma \approx 2$. Note finally that the quantity V_j, which also varies, represents the time-averaged speed of the underlying jet. The empirical relation (13.7) is only established for the range $V_\circ \lesssim V_r \ll V_j$, where $V_\circ$ is the parent cloud's radial speed. In fact, additional measurements indicate that the power law breaks down as the outflow velocity approaches V_j itself. Nevertheless, we shall argue in § 13.5 that the existence of such a relationship provides a further clue to the nature of the jet-outflow coupling.

13.2.4 Energy and Momentum

Suppose that we know the angle by which a given outflow is tilted from the plane of the sky. This information might come from observing both proper motions and radial velocities of embedded HH objects. One may also infer the angle by matching a geometrical model of the outflow lobes to an empirical position-velocity diagram. In any case, the inclination allows one to convert each radial velocity V_r into a fully three-dimensional speed u. From the mass distribution, now written as dM/du, one may then obtain the outflow's *mechanical luminosity*:

$$L_{\rm mech} = \frac{1/2\int_{u_\circ}^{\infty} du\, u^2\, dM/du}{t_{\rm dyn}} . \tag{13.8}$$

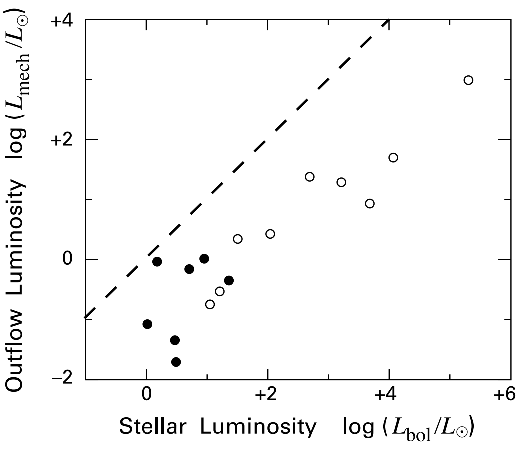


Figure 13.14 Mechanical outflow luminosity as a function of bolometric stellar luminosity for both Class 0 (*filled circles*) and Class I sources (*open circles*). The dashed line represents equality of the two luminosities.

Here, the integral covers all speeds greater than the background value $u_\circ$. The dynamical time $t_{\rm dyn}$ is $L/\langle u\rangle$, where L is a lobe length and $\langle u\rangle$ the mass-averaged speed within that lobe. For the prototypical L1551/IRS 5, the blueshifted lobe is 0.8 pc long and is inclined by 15° out of the plane of the sky. The mass-averaged speed is 18 km s^{-1}, yielding a dynamical time of 4×10^4 yr.

The mechanical luminosity is a measure of the rate at which the outflow is transporting kinetic energy. This energy, of course, must ultimately be supplied by the central jet. The transfer process is not a conservative one, *i. e.*, we expect considerable energy to be lost along the way. Observations support this view. For example, the L1551 cloud emits copiously in the far infrared. The relative fluxes at 60 and 100 µm indicate that the dust grains responsible have been heated to 25 K. Their diffuse radiation, while poorly resolved, follows roughly the orientation of the outflow lobes centered on IRS 5, which are the dominant ones in the cloud. The corresponding luminosity, $L_{\rm rad}$, is about 7 $L_\odot$ and exceeds $L_{\rm mech}$ for the outflow by about a factor of three. We may picture $L_{\rm rad}$ arising from dissipation of the turbulent motion within the outflow, although the details are far from clear. This random component of the velocity is not included, of course, in equation (13.8). The jet must therefore supply energy to the parent cloud at the total rate $L_{\rm mech} + L_{\rm rad}$.

Since the radiated luminosity is generally unknown, we must rely on $L_{\rm mech}$ to gauge the outflow power. It is interesting to compare this quantity with $L_{\rm bol}$ for the driving star. As seen in Figure 13.14, there is a clear correlation between the two. In this plot, closed and open circles represent systems driven by Class 0 and I sources, respectively. Apparently, both types drive outflows with mechanical luminosities roughly 1 percent of $L_{\rm bol}$. This fraction tends to be higher for Class 0, but the number of observed systems is still low. Note that we are really using $L_{\rm bol}$ as a surrogate for the energy transport rate within the jet itself, but this latter information is more difficult to ascertain (see § 13.5 below). Assuming the two rates are linked, the indication from Figure 13.14 is that Class 0 sources are more efficient in driving their outflows. We also have seen that the lobes themselves are narrower in this case, and that $\langle u\rangle$ is higher. Since Class 0 stars are thought to be the youngest, molecular outflows must broaden with time, while undergoing a progressively greater amount of dissipation.

We may also gauge the *momentum* content of outflows. The rate of momentum transport has the units of force, so the expression analogous to equation (13.8) is

$$F_{\rm out} = \frac{\int_{u_\circ}^{\infty} du \; u \; dM/du}{t_{\rm dyn}} \; . \tag{13.9}$$

Here, we again sum the contributions of the two lobes, although their motion is in opposite directions. Indeed, we expect the total vectorial momentum of the lobes to be nearly zero, reflecting the symmetry of the underlying jet. The transfer process from jet to outflow is now conservative, since momentum cannot be dissipated. Thus, the momentum gain in a lobe should just balance a decline in the outwardly traveling wind material. This effect is not yet evident in the data on jets. As we saw in § 13.1, the observed velocity changes within sequences of Herbig-Haro knots are probably dominated by temporal variations of the stellar wind speed. In addition, the best-observed jets have relatively weak outflows, rendering a detailed comparison of the momenta impractical.

It is nevertheless encouraging that typical values of $F_{\rm out}$ are in rough accord with estimates of the jet momentum transport rate, as we shall later demonstrate. The data cited previously for L1551/IRS 5 yield $F_{\rm out} = 1 \times 10^{-3} \, M_\odot \, {\rm km \, s^{-1} \, yr^{-1}}$. Surveying the ensemble of well-mapped outflows, the rates are generally lower than this and vary widely. There is again a tendency for stars with higher $L_{\rm bol}$ to produce flows with greater $F_{\rm out}$. In all cases, the transport rate exceeds, by one to two orders of magnitude, $L_{\rm bol}/c$. The last quantity is the force imparted by radiation pressure in an optically thin medium. Historically, this inequality was noted early on, and served to emphasize the fact that the driving agent was mechanical rather than radiative.[1]

13.2.5 Thermal and Chemical Effects

Returning to energetic considerations, we note that the perturbed cloud radiates over a broad range of temperatures. In L1551, we find not only the far-infrared continuum emission from warm dust, but also the 2.12 µm line from molecular hydrogen that has been heated to several thousand degrees. This near-infrared emission has also been detected in optical jets, sometimes in conjunction with the $2-1\,S(1)$ line at 2.25 µm. Examples include systems we have already studied: HH 1/2, HH 111, and HH 46/47. Even within molecular outflows containing no visible Herbig-Haro objects, near-infrared array detectors allow one to pick out embedded shocks with arcsecond resolution. These hydrogen lines probably account for much of the cloud's total energy dissipation.

Outflow gas that is further removed from the jet is also elevated in temperature. As we have seen, such heating is not apparent in the broader lobes, but it does show up in the narrow flows driven by Class 0 sources. Here, multi-line observations of CO give excitation temperatures near 40 K, significantly above the surrounding medium. The rotational lines of SiO and the inversion transitions of NH_3 have also proved useful in this regard.

---

[1] The outflow material is actually opaque to the radiation from the stellar photosphere. Photons therefore diffuse outward, imparting a force per unit volume of $\rho\kappa F_{\rm rad}/c$. (Recall the discussion following equation (11.48).) To a crude approximation, the total force is then $\Delta\tau L_{\rm bol}/c$, where $\Delta\tau$ is a representative optical thickness. Most of the radiation that reaches the lobe interior has been degraded to the infrared, where $\Delta\tau$ is well below the figure of 10 to 100 required for the radiative force to match $F_{\rm out}$.

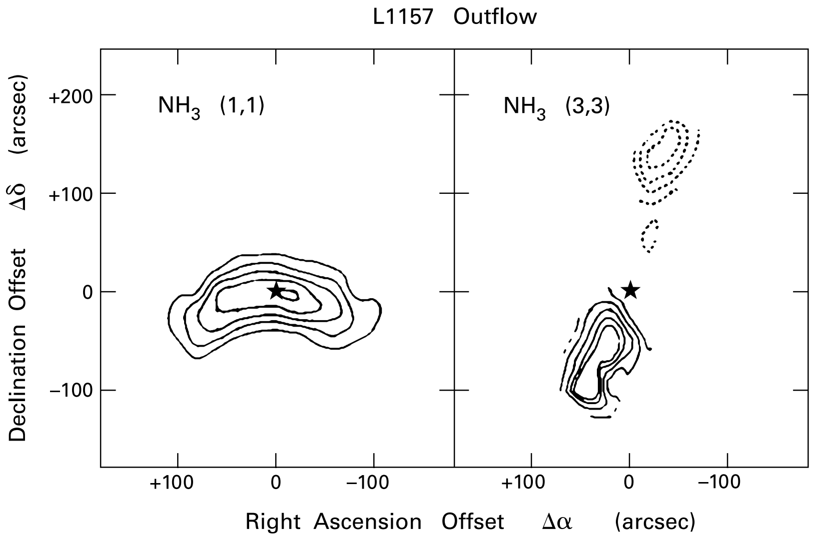


Figure 13.15 Map of the L1157 outflow in both the NH_3 $(1,1)$ (*left panel*) and NH_3 $(3,3)$ (*right panel*) lines. The solid and dashed contours on the right show red- and blueshifted emission, respectively. The Class 0 driving source is located at the central star in each panel.

The NH_3 measurements are especially interesting, as the molecule is a sensitive probe of physical conditions. Consider the high-velocity L1557 outflow, driven by a Class 0 source of 11 $L_\odot$. Here, there is strong emission both at 1.27 cm, stemming from the NH_3 (1,1) rotational level, and also at 1.26 cm, the inversion transition in the (3,3) state. The observed line width, both for the main line and satellites, is relatively narrow in the first case, and the peak emission occurs close to the ambient radial velocity. In contrast, the (3,3) line is both broad and shifted. Figure 13.15 shows that the lines also arise in different locations. The (1,1) emission originates within the cold, quiescent dense core surrounding the star itself. As expected, the intensity peaks near the star. This is just the region avoided by the (3,3) emission, which concentrates in the two lobes. Analysis of the relative fluxes yields temperatures near 10 K in the core, and between 50 and 100 K in the outflow.

The same analysis indicates that the fractional abundance of the NH_3 molecule is enhanced by an order of magnitude within the lobes. An even more dramatic increase, amounting to *at least* a factor of 10^4, is found for SiO, which is undetectable in the central core. Interferometric mapping of the lower rotational transitions reveals that the lobe emission arises from individual bowshocks centered on the jet axis. Other Class 0 outflows radiate strongly in these SiO lines. So do a number of massive star formation sites, including NGC 7538, the spectacular HII region in the Cas OB2 association (Plate 6).

As we mentioned, SiO is *not* seen toward the dense core from which the L1157 outflow emanates. Nor has it been detected in the well-studied Taurus core TMC-1 or the Bok globule B335. These facts together suggest that the molecule is actually *created* in outflows. There is good reason to believe that silicon in dense, quiescent environments is locked up in grain cores. After all, the 10 μm absorption feature in the interstellar extinction curve originates in the Si-O stretching mode (recall § 2.4). Within the shocks accompanying stellar jets, both sputtering and grain-grain collisions will release atomic silicon. We discussed both effects briefly in § 8.4, where we noted that shock speeds of 200 km s^{-1} are needed for significant grain disruption. In a C-shock, however, there is enhanced interaction between charged grains and the ambient neutrals, which have a different flow speed. Grain mantles are completely destroyed even for $V_{\rm shock} = 50$ km s^{-1}, while the cores are partially eroded. The released silicon reacts with molecular oxygen and hydroxyl (OH) to produce the observed SiO. The more modest increase of NH_3 could arise from heating and sublimation of the outer grain mantles, which then eject the molecule intact. Similar activity could account for the enhancement of methanol (CH_3OH) in L1157 and other systems. Note that methanol is the second most abundant species (after water) in grain ices.

13.2.6 The OMC-1 Outflow

We see that stellar jets and their attendant outflows not only perturb their host clouds dynamically, but even alter their chemical makeup. The shocks erode or destroy grains, while their heating drives gas-phase reactions that would otherwise be inoperative. All of these activities increase substantially in the vicinity of massive stars. The variety of ways by which such objects affect their surroundings will be the topic of Chapter 15. Anticipating this treatment, let us conclude the present discussion with a look at the well-studied OMC-1 outflow.

The very dense OMC-1 region of Orion A contains both the BN object and the KL infrared cluster. It is also the site of a vigorous molecular outflow. Within a volume about 0.2 pc in diameter, gas is moving outward from IRc2, the brightest source ($L_{\rm bol} \lesssim 10^5\, L_\odot$) in the KL Nebula. The radial velocity of the material approaches 150 km s^{-1}, but the morphology is only weakly bipolar. That is, at most positions, the CO profiles contain *both* red- and blueshifted material, and any "lobes" are extremely broad. Toward the outskirts of the flow are two regions bright in shocked molecular hydrogen. (Recall Figure 8.9 and the accompanying discussion.)

Thus far, our discussion *might* apply to a massive, but otherwise unexceptional, outflow in which the central axis happens to lie close to our line of sight. However, the extensive additional observations of this system cast doubt on this interpretation and indicate that the flow is not driven by a single jet. Toward IRc2 is a compact, hot core of relatively quiescent molecular gas. High-temperature dust provides copious infrared continuum radiation, while a plethora of molecular lines emanates from such species as SiO, HCO^+, and HCN. The ambient volume density here exceeds 10^6 cm^{-3}. From the presence of such high-order NH_3 inversion lines as (10,10) and even (14,14), the gas temperature is at least 400 K. This value falls to about 150 K within the larger flow.

The most remarkable aspect of OMC-1 is a spray of high-velocity streamers extending beyond the main CO outflow. Plate 13 is a color rendition constructed from a mosaic of 2.12 μm images. Here we see that the molecular hydrogen shoots out in discrete "fingers" from IRc2, which lies just below the bottom of the frame. The tip of each finger is especially bright in the 1.64 μm line of [Fe II], which is excited at even a higher temperature than the H_2 transition. Gas-phase iron, like silicon, is severely depleted in quiescent environments, and its presence suggests at least partial dissolution of grain material. Conspicuously absent here are the curved bowshocks we saw in other jets. The white contours in Plate 13 trace emission from NH_3 (1,1) and (2,2). These clumpy, dense filaments extend to about 0.5 pc.

At the present time, there is no generally accepted picture that accounts for these diverse phenomena. One's visual impression from Plate 13 is of a violent, explosive event, spewing out a fan of clumps while dispersing and shredding the ambient gas seen in NH_3. The shocked fingers actually lead back, not precisely to IRc2, but to an adjacent radio continuum source. Apparently, IRc2 itself is a cluster of small patches reflecting starlight from the central object. Perhaps both the wide-angle CO and the shower of dense fragments result from an especially high-velocity wind, which cannot cool efficiently after shocking. We shall revisit this issue later in this chapter, when we examine more closely the dynamics of jets, as well as in Chapter 15. There we shall find that massive outflows quite generally differ from their low-mass counterparts.

13.3 Wind Generation: Pressure Effects

Let us now turn our attention from the large-scale effects of stellar jets to the physical nature of the flows themselves. We defer momentarily the question of how these streams attain their impressive degree of collimation. Instead, we wish to probe the more basic issue of how young stars produce energetic winds in the first place. These objects are too deeply embedded in cloud material for direct observation. Hence, we will rely heavily on what is known about winds from visible, pre-main-sequence stars. Even here, as we shall see, the available information is sparse and the theory by no means fully developed.

13.3.1 Types of Winds

Stars of every mass and age exhibit winds, *i. e.*, shedding of their atmospheric layers. These flows vary considerably in observable properties, including both $\dot{M}_w$, the total rate of mass loss, and V_∞, the coasting velocity attained far from the stellar surface. Along the main sequence, the highest $\dot{M}_w$-values (of order $10^{-5}\,M_\odot\,\mathrm{yr}^{-1}$) stem from early-type stars. Terminal velocities here are also relatively large, from 600 to 3500 km s^{-1}. At the low end of this range, the values are close to the driving star's gravitational escape speed, *i. e.*, V_{ff} evaluated at the stellar radius R_*. The higher wind velocities are up to 4 times that value. The mass loss rate falls drastically for late-type dwarfs. It climbs again, however, for the giants and asymptotic branch stars. A K5 supergiant of mass 16 $M_\odot$ and radius 400 $R_\odot$, for example, emits a wind with $\dot{M}_w \approx 10^{-7}\,M_\odot\,\mathrm{yr}^{-1}$. Curiously, the terminal speeds from such cool, post-main-sequence objects are *less* than the corresponding escape value. The winds from the most luminous of these stars, like our K5 supergiant, show evidence for substantial dust and regular stellar pulsations. Finally, the far less brilliant pre-main-sequence stars of solar-type mass again exhibit massive winds ($\dot{M}_w \sim 10^{-8}\,M_\odot\,\mathrm{yr}^{-1}$), but have V_∞-values of several hundred km s^{-1}, *i. e.*, near the escape speed.

The broad range of wind characteristics throughout the HR diagram indicates that a number of processes must be at work. In all cases, material leaves the stellar surface with a speed well below the escape value. Since the initial speed is also less than the local *thermal* velocity, some mechanism is required to accelerate the gas until it becomes supersonic. This mechanism, which varies from one type of wind to another, largely determines $\dot{M}_w$. In contrast, the value of V_∞ is mainly set by whatever momentum addition occurs *beyond* the sonic transition. The most massive and highest-speed winds, those from O and B stars, owe their existence to the extreme luminosity of these objects. As we shall describe in Chapter 15, ultraviolet photons are absorbed and then reradiated by ionized atoms in the flow. Such scattering accelerates the gas to the sound speed and beyond. Stars along the asymptotic giant branch also drive their winds radiatively, this time through the pressure exerted on ambient dust grains. For all other stars, including the young, low-mass ones of most interest here, the winds are mechanically propelled.

By far the best studied wind is that from our own Sun, for which spacecraft have long provided *in situ* measurements. The mass loss rate associated with this flow is only $2 \times 10^{-14}\, M_\odot\ \mathrm{yr}^{-1}$, while its velocity at the Earth's orbit is about 450 km s^{-1}. This latter figure represents an average of the "quiet" speed of 350 km s^{-1} usually seen, and the 800 km s^{-1} that characterizes transient, high-speed streams. For comparison, the Sun's escape speed is 620 km s^{-1}.

To understand the solar wind, we must first examine the tenuous gas lying outside the photosphere. This region, by definition, is optically thin to the visible radiation that carries the bulk of the Sun's luminosity. On the other hand, it is susceptible to heating by shorter-wavelength photons and by acoustic and MHD waves propagating from the interior. Such heating explains why the gas temperature, while it initially declines beyond the photospheric value of 5800 K, reaches a minimum of 4300 K and then begins to rise. A relatively thin layer, known as the *chromosphere*, is bright in emission lines of HI and such ions as Ca II and Mg II. Once the temperature reaches about 10^4 K, it climbs much more steeply, increasing by two orders of magnitude in a very short distance. This *transition region* marks the beginning of the *corona*, a more rarefied and physically extended zone of very hot gas.

The solar wind itself starts near the base of the corona, at a heliocentric radius of $1.25\, R_\odot$. Here, the temperature is 1.5×10^6 K, some 250 times the photospheric value. It is the outward falloff in thermal pressure that accelerates the wind, or at least its quiet component. Note that the high-speed streams, along with the geomagnetic disturbances they create, recur in 27-day cycles. This interval is the mean solar rotation period as seen from the Earth. Satellite observations in the 1970s revealed that the streams come from *coronal holes*, *i. e.*, regions of diminished X-ray flux. Within a hole, the magnetic field topology changes from closed loops that leave and enter the solar surface to open field lines projecting far into space. The additional speed of the streams may be provided by MHD waves traveling along this open field component.

13.3.2 Thermal Flows

There is good reason to believe that all low-mass stars near the main sequence emit winds similar to that from the Sun. The evidence here is circumstantial, since the density in all such flows, including our own, is too low for direct observation of the emitted radiation. Despite this limitation, there is a solid understanding of how thermal pressure from a hot corona accelerates a solar-type wind. Let us now examine the underlying physics in more detail. We emphasize from the start that this thermal driving is *not* responsible for the stronger winds emanating from the youthful counterparts to these stars. The concepts we shall introduce, however, are useful for our later discussion.

To simplify matters, let us consider a purely radial wind in steady-state motion, with velocity $u(r)$ and density $\rho(r)$. Here, r measures the distance from the star's center. The equations of mass and momentum conservation are

$$\frac{1}{r^2}\frac{d(r^2\rho\,u)}{dr} = 0 \tag{13.10}$$

$$u\,\frac{du}{dr} = -\frac{a_T^2}{\rho}\frac{d\rho}{dr} - \frac{G\,M_*}{r^2}\ , \tag{13.11}$$

where M_* is the stellar mass. Here, we have supposed that the wind, which is fully ionized, conducts heat so efficiently that it remains strictly isothermal, with an associated sound speed a_T. (Recall equations (10.27) and (10.29) for the analogous infall problem.) After expanding the derivative in equation (13.10), the term $d\rho/dr$ may be eliminated in favor of du/dr. The transformed equation (13.11) then reads

$$\frac{\left(u^2 - a_T^2\right)}{u}\,\frac{du}{dr} = \frac{2\,a_T^2}{r} - \frac{G\,M_*}{r^2}\ . \tag{13.12}$$

We now integrate equation (13.12) outward from the initial radius $r_\circ$, which represents the base of the corona. If $u_\circ$ is the initial, subsonic velocity at this radius, we find an implicit relation for the general velocity $u(r)$:

$$\left(\frac{u}{a_T}\right)^2 - 2\,\ln\!\left(\frac{u}{a_T}\right) = \left(\frac{u_\circ}{a_T}\right)^2 - 2\,\ln\!\left(\frac{u_\circ}{a_T}\right) + 4\,\ln\!\left(\frac{r}{r_\circ}\right) + \frac{4\,R_s}{r_\circ}\left(\frac{r_\circ}{r} - 1\right)\ . \tag{13.13}$$

Here, R_s is the *sonic point*:

$$\begin{aligned} R_s &\equiv \frac{G\,M_*}{2\,a_T^2} \\ &= 7\,R_\odot\left(\frac{M_*}{M_\odot}\right)\left(\frac{T}{10^6\,\mathrm{K}}\right)^{-1}\ . \end{aligned} \tag{13.14}$$

Within the numerical expression for R_s, our representative wind temperature T is an average from the solar coronal base out to the sonic point. The observed temperature actually falls slowly with distance, so that it is lower by an order of magnitude at 1 AU (210 $R_\odot$). In comparison, the number density in the wind drops by seven orders of magnitude over the same distance.

Figure 13.16 displays the run of velocity for three choices of $u_\circ$. Here, we have set $R_s/r_\circ$ equal to 10. For the lowest value, $u_\circ/a_T = 7.0 \times 10^{-7}$, the velocity first climbs, peaks at $r = R_s$, and then falls, remaining wholly subsonic. As $u(r)$ becomes vanishingly small at large distance, we see from equation (13.11) that the density $\rho(r)$ rises linearly, contrary to observations. Equally unacceptable as a physical solution is that for the highest initial speed, $u_\circ/a_T = 1.2 \times 10^{-6}$. The acceleration now continues until du/dr becomes infinite at a certain radius, $r/r_\circ = 6.2$. Formally, the solution then doubles back toward smaller r with supersonic velocity, as shown. Referring to equation (13.12), it is clear that the only realistic

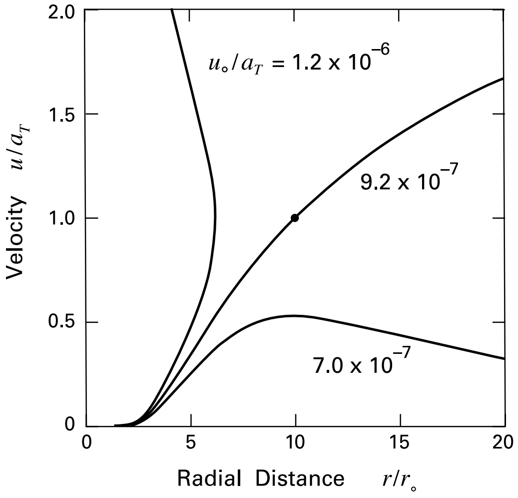


Figure 13.16 Velocity as a function of radius in a spherical, isothermal wind driven by thermal pressure. The initial values of velocity relative to the sound speed a_T are shown for each curve, while the radius is displayed in units of $r_\circ$, the value for the coronal base. The filled circle on the middle curve marks the sonic point.

situation is when the coefficient of du/dr vanishes just as the righthand side goes through zero, *i. e.*, for which $u(R_s) = a_T$. This special velocity profile, corresponding in the present case to $u_\circ/a_T = 9.2 \times 10^{-7}$, is also plotted. The flow now passes smoothly through the sonic point, represented by the filled circle, while $\rho(r)$ monotonically decreases.

Despite the simplifications in our model, the derivation reveals the essential mechanism of pressure-driven expansion. The inequality $R_s > r_\circ$ is equivalent to the statement that a fluid element at the coronal base is tightly bound by gravity, even at the thermal speed. Nevertheless, the element drifts gently and continually outward, impelled by the gradual drop in thermal pressure. This efflux lowers the pressure even more, inducing layers further inside to expand. The deeper issue, then, is not what drives the expansion, but what supplies thermal energy to the corona itself. Our understanding in this area is poor, again reflecting the lack of direct observations, even in the solar case. Note that the total luminosity required is only of order $10^{-4}\, L_\odot$. This figure includes not only the power to drive the wind and other, more impulsive mass ejections, but also to offset the corona's radiative losses. There is general agreement that the power originates from turbulent fluid motions within the Sun's outer convection zone. It is the transformation of this energy that remains puzzling. The dissipation of outwardly propagating MHD waves may play a role, as may reconnection in tangled or braided magnetic flux tubes.

13.3.3 Influence of Alfvén Waves

We mentioned earlier that MHD waves are also implicated in the high-velocity streams emanating from coronal holes. The existence of such waves is incontrovertible in the solar case. Spacecraft observations reveal that fluctuations in the wind velocity are correlated with those of the local magnetic field, in just the manner expected for Alfvén waves. We recall from Chapter 9 that the field may also carry fast and slow magnetosonic waves. These, however, create periodic compression of the plasma, which in turn leads to steepening and the formation of shocks. The dissipation of energy in these shocks adds heat to the corona, but only Alfvén waves survive well past the sonic point.

Alfvén waves propagating in a wind exert an additional force on the gas. From equation (9.91), the wave pressure at any point is proportional to $|\delta\boldsymbol{B}|^2$, the squared amplitude of the field perturbation, and thus to the volumetric energy density of the wave. Within an expanding flow, $|\delta\boldsymbol{B}|^2$ naturally declines, so that the net force is outward.[2] Mathematically, one may incorporate the wave pressure gradient as an extra term in the momentum equation (13.11). The net result is to increase the effective *temperature* of the wind. From equation (13.14), the generalized sonic point moves inward, closer to the stellar surface. The velocity profile passing through this point also has a higher V_∞.

The enhanced pressure from Alfvén waves provides a natural explanation for the high-speed component of the solar wind. This effect also plausibly accounts for the far more massive winds emanating from late-type giants. Even if the parent star has a mass greater than the Sun, its surface gravity may be smaller by 10^{-4}. Thus, a solar-type flux of Alfvén waves can indeed produce a very high value of $\dot{M}_w$. The problem is that the mechanism is *too* efficient. If the Alfvén waves indeed persist beyond the sonic point, as they do in the solar wind, they inevitably create a large V_∞. We have noted, however, that the observed terminal speeds in the winds of giants are relatively small, below the star's escape velocity.

Giant stars do not exhibit the copious X-rays that would be produced by a hot and massive corona. Thus, their winds *cannot* be thermally driven, and the temperatures are probably much lower than solar, closer to 10^4 K. Given the relatively high densities, ion-neutral friction should dampen Alfvén waves. (Recall the discussion of § 10.1.) As long as the waves do not survive far beyond the sonic point, they can still drive a massive wind, but with a reduced V_∞. This explanation for giant winds is promising, as least qualitatively. Calculations show, however, that if the damping is even slightly too great, the mass loss rate drops precipitously. Conversely, too weak a damping gives back unacceptably large values of V_∞. If Alfvén waves are indeed the driving agent, these results hint at the existence of some feedback mechanism that regulates their production or dissipation. We should also bear in mind that the underlying calculations are highly simplified. Most assume, for example, a perfectly spherical wind, whereas the solar example indicates that the flow may be at least partially confined to discrete flux tubes.

How applicable are these ideas to pre-main-sequence stars? Here, the winds comprise a much larger portion of the energy budget than for either giants or our own Sun. Consider, for example, a T Tauri star of $1\,M_\odot$ and $3\,R_\odot$, with a luminosity $L_* = 3\,L_\odot$. Suppose that this object emits a wind with $\dot{M}_w = 1 \times 10^{-8}\,M_\odot\,\mathrm{yr}^{-1}$. Assuming a terminal speed equal to the escape value of 360 km s^{-1}, the associated mechanical luminosity is $0.1\,L_\odot$, or 4 percent of the stellar value. Such a flow is assuredly not driven by a massive corona, whose X-ray output would far exceed the star's optical luminosity, in marked contrast to the observations.

The pressure from Alfvén waves could alleviate this difficulty, at least in principle. Calculations reveal, however, that generating a wind similar to that in our example would require a

---

[2] The energy of an MHD wave packet is not conserved, even in the absence of dissipation. A perturbation analysis, in the short-wavelength regime, shows that the conserved quantity is the *wave action*. This is the energy in the packet divided by the comoving frequency, *i. e.*, the frequency detected by an observer traveling at the mean wind velocity. As the wind picks up speed, the comoving frequency declines through the Doppler effect. To conserve wave action, the energy also falls, reflecting the work done by the wave on the background flow.

mechanical luminosity in the waves themselves close to L_*. This energy, along with that from radiation and the background wind, must originate from the star, and a significantly enhanced total would be difficult to reconcile with pre-main-sequence theory. The essential problem here is the relatively high gravity near the stellar surface, which makes the outward transport of wave energy more inefficient than for the bloated giants. Moreover, even if one were to posit a very high wave luminosity, the resulting V_∞ would generally exceed observational limits. One must then include dissipation of the waves beyond the sonic point, either through ion-neutral collisions or conversion to the fast and slow wave modes. The whole chain of reasoning is then subject to the same criticism as for giants, *i. e.*, that conditions must be fine-tuned to give reasonable results.

Although the Alfvén-wave mechanism does not provide the final answer, our discussion has introduced two key elements that are likely to remain. The first is the presence of a strong magnetic field emanating from the star itself. As we shall see in Chapter 17, this prediction is confirmed empirically, and the finding lends credence to the notion that the field plays an important role in wind generation. The second idea is that matter leaving the vicinity of the star is subject to wave-like or perhaps turbulent motion, in addition to its smooth, outward flow velocity. Observationally, emission lines from T Tauri stars are highly broadened and generally more complex than in a homogeneous wind.

13.4 Wind Generation: Rotation and Magnetic Fields

Another important clue that magnetic fields are involved with wind production comes from the extensive data on stellar rotation. As one proceeds down the main sequence toward lower surface temperatures, one finds an abrupt decrease in rotational speed at F-type stars. This is also the point at which main-sequence objects develop an outer convection zone. Quite generally, stellar magnetic fields are thought to arise from dynamo action, *i. e.*, the interplay of rotation and turbulence within convectively unstable regions. Suppose now that a late-type star ejects gas that travels out along rotating magnetic field lines anchored to the stellar surface. If the field maintains its rigidity for some distance, then the ejected matter picks up a high angular velocity before it is flung into space. Such a *centrifugal wind* can therefore provide efficient rotational braking of the star, accounting for the observations. More powerful versions of these winds may be occurring in younger stars. Indeed, the bipolar anisotropy of flows from embedded sources plausibly arises from the reflection symmetry of the underlying field.

13.4.1 Configuration of Magnetized Winds

Let us now explore the elements of centrifugal wind theory. Our goal, to be achieved in equation (13.36) below, will be to derive a relation governing the velocity analogous to equation (13.13) for pressure-driven flows. Along the way, we shall also gain a better understanding of wind braking. We assume from the outset that both the wind gas and magnetic field are azimuthally symmetric about the z-axis of a cylindrical coordinate system, and that this axis coincides with the direction of stellar rotation. It is then convenient to split both the field $\boldsymbol{B}$ and the velocity $\boldsymbol{u}$ into poloidal and toroidal (*i. e.*, azimuthal) components (recall Figure 10.13). Suppose that the wind material is sufficiently ionized that the ideal MHD equation (9.45) applies. Assuming the flow to be steady-state, this condition of flux freezing reduces to

$$\boldsymbol{\nabla} \times (\boldsymbol{u} \times \boldsymbol{B}) = 0 \; . \tag{13.15}$$

Matter subject to flux freezing can only slide along field lines or rotate with them around the z-axis. To develop this picture mathematically, we note that equation (13.15) implies that $\boldsymbol{u}\times\boldsymbol{B}$ is expressible as the gradient of a scalar field. But no quantities, including this scalar, vary azimuthally. Thus $\boldsymbol{u}\times\boldsymbol{B}$ must itself be a poloidal vector. If we now write $\boldsymbol{u} = \boldsymbol{u}_p + \boldsymbol{u}_\phi$ and $\boldsymbol{B} = \boldsymbol{B}_p + \boldsymbol{B}_\phi$, we quickly see that the only toroidal contribution to $\boldsymbol{u}\times\boldsymbol{B}$ is $\boldsymbol{u}_p\times\boldsymbol{B}_p$. For this to vanish, $\boldsymbol{u}_p$ must be everywhere parallel to $\boldsymbol{B}_p$. That is,

$$\boldsymbol{u}_p = \kappa \boldsymbol{B}_p \; , \tag{13.16}$$

where κ is a scalar function of position. Within any meridional plane, matter indeed slides along the unchanging field lines.

Having eliminated the toroidal component of $\boldsymbol{u}\times\boldsymbol{B}$, equation (13.15) is now

$$\boldsymbol{\nabla} \times (\boldsymbol{u}_p\times\boldsymbol{B}_\phi + \boldsymbol{u}_\phi\times\boldsymbol{B}_p) = 0 \; . \tag{13.17}$$

The velocity $\boldsymbol{u}_\phi$ may be recast as $\varpi\Omega\boldsymbol{e}_\phi$, where $\boldsymbol{e}_\phi$ is the unit toroidal vector. Using equation (13.16) for $\boldsymbol{u}_p$, equation (13.18) becomes

$$\boldsymbol{\nabla} \times (\boldsymbol{B}_p \times \alpha\varpi\boldsymbol{e}_\phi) = 0 \; . \tag{13.18}$$

Here, α is a new scalar quantity:

$$\alpha \equiv \Omega - \kappa B_\phi/\varpi \; . \tag{13.19}$$

This equation and (13.16) together imply that

$$\boldsymbol{u} = \kappa \boldsymbol{B} + \alpha\,\varpi\,\boldsymbol{e}_\phi \; . \tag{13.20}$$

To ascertain the physical meaning of α, we expand the triple vector product in equation (13.18).[3] In axisymmetry, Maxwell's equation $\boldsymbol{\nabla}\cdot\boldsymbol{B} = 0$ reduces to $\boldsymbol{\nabla}\cdot\boldsymbol{B}_p = 0$, while $\boldsymbol{\nabla}\cdot(\alpha\varpi\boldsymbol{e}_\phi)$ also vanishes. We thus find

$$\alpha\frac{\partial \boldsymbol{B}_p}{\partial\phi} - (\boldsymbol{B}_p\cdot\boldsymbol{\nabla})(\alpha\varpi\boldsymbol{e}_\phi) = 0. \tag{13.21}$$

Noting that $\partial\boldsymbol{e}_\varpi/\partial\phi = \boldsymbol{e}_\phi$ and $(\boldsymbol{B}_p\cdot\boldsymbol{\nabla})\boldsymbol{e}_\phi = 0$, this last equation simplifies to

$$\alpha\, B_\varpi - \boldsymbol{B}_p\cdot\boldsymbol{\nabla}(\alpha\varpi) = 0 \; ,$$

from which it follows that

$$\boldsymbol{B}_p\cdot\boldsymbol{\nabla}\alpha = 0 \; .$$

---

[3] The relevant identity is $\boldsymbol{\nabla}\times(\boldsymbol{C}\times\boldsymbol{D}) = (\boldsymbol{D}\cdot\boldsymbol{\nabla})\boldsymbol{C} - \boldsymbol{D}(\boldsymbol{\nabla}\cdot\boldsymbol{C}) - (\boldsymbol{C}\cdot\boldsymbol{\nabla})\boldsymbol{D} + \boldsymbol{C}(\boldsymbol{\nabla}\cdot\boldsymbol{D})$, for any two vector fields $\boldsymbol{C}$ and $\boldsymbol{D}$.

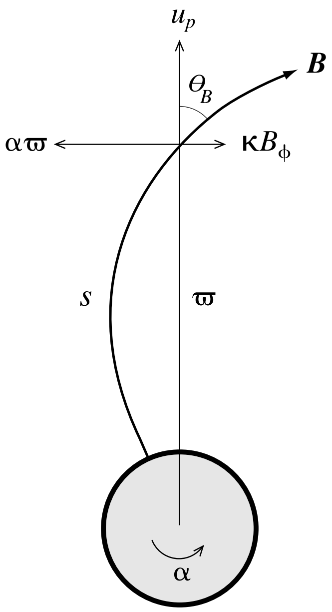


Figure 13.17 A stellar magnetic field is dragged in the counterclockwise direction at angular speed α. Depicted are the projection of the field and gas flow onto the equatorial plane. At radial distance ϖ, the material velocity has poloidal and azimuthal components, as indicated. The former makes an angle θ_B with the field vector $\boldsymbol{B}$. The coordinate s measures distance along the field line.

But $\boldsymbol{B}_\phi\cdot\boldsymbol{\nabla}\alpha = B_\phi\,\partial\alpha/\partial\phi = 0$ by axisymmetry, so that finally

$$\boldsymbol{B}\cdot\boldsymbol{\nabla}\alpha = 0\ . \tag{13.22}$$

The scalar α is thus unchanging in the direction of $\boldsymbol{B}$. This quantity represents the angular speed at which each field line rotates about the z-axis. Equation (13.20) states that the total azimuthal speed of a fluid element is $\alpha\varpi$ plus κB_ϕ, the slippage along the field itself. These two contributions are of opposite sign if the field line forms a trailing spiral, as illustrated in Figure 13.17.

13.4.2 Flow Dynamics

Within the constraints imposed by flux freezing, the wind is accelerated by the ambient pressure, gravitational and magnetic forces. The governing momentum equation (3.3) becomes, in steady state,

$$\rho\,(\boldsymbol{u}\cdot\boldsymbol{\nabla})\boldsymbol{u} = -\boldsymbol{\nabla}P - \rho\boldsymbol{\nabla}\Phi_g + \frac{1}{4\pi}(\boldsymbol{\nabla}\times\boldsymbol{B})\times\boldsymbol{B}\ . \tag{13.23}$$

Here we have utilized Ampére's law, equation (9.32) to replace the current $\boldsymbol{j}$ in the final right-hand term. The toroidal projection of this equation is

$$\rho\,[\boldsymbol{u}\cdot\boldsymbol{\nabla}(\varpi\Omega) + u_\varpi\Omega] = \frac{1}{4\pi}\,[(\boldsymbol{\nabla}\times\boldsymbol{B})\times\boldsymbol{B}]_\phi\ . \tag{13.24}$$

Since $\boldsymbol{u}\cdot\boldsymbol{\nabla}\varpi = u_\varpi$, the left side becomes

$$\frac{\rho}{\varpi}\boldsymbol{u}\cdot\boldsymbol{\nabla}(\Omega\varpi^2) = \frac{\rho\,\kappa}{\varpi}\boldsymbol{B}\cdot\boldsymbol{\nabla}(\Omega\varpi^2)\ , \tag{13.25}$$

while the right side of (13.24) expands to

$$\begin{aligned}\frac{1}{4\pi\varpi}\left[B_z\,\frac{\partial}{\partial z}(\varpi B_\phi)\,+\,B_\varpi\,\frac{\partial}{\partial\varpi}(\varpi B_\phi)\right] &= \frac{1}{4\pi\varpi}\boldsymbol{B}_p\boldsymbol{\cdot}\boldsymbol{\nabla}(\varpi B_\phi)\\ &= \frac{1}{4\pi\varpi}\boldsymbol{B}\boldsymbol{\cdot}\boldsymbol{\nabla}(\varpi B_\phi)\ .\end{aligned} \tag{13.26}$$

The combination $\rho\kappa$ appearing in equation (13.25) is also conserved along a field line. To see this, we invoke the steady-state version of mass continuity, $\boldsymbol{\nabla}\boldsymbol{\cdot}(\rho\boldsymbol{u}) = 0$. As usual, the divergence operates only on $\boldsymbol{u}_p$, so that

$$0\,=\,\boldsymbol{\nabla}\boldsymbol{\cdot}(\rho\boldsymbol{u}_p)\,=\,\boldsymbol{\nabla}\boldsymbol{\cdot}(\rho\kappa\boldsymbol{B}_p) = \rho\kappa\boldsymbol{\nabla}\boldsymbol{\cdot}\boldsymbol{B}_p\,+\,\boldsymbol{B}_p\boldsymbol{\cdot}\boldsymbol{\nabla}(\rho\kappa)\ .$$

Noting that $\boldsymbol{\nabla}\boldsymbol{\cdot}\boldsymbol{B}_p = 0$ and that the operator $\boldsymbol{B}_p\boldsymbol{\cdot}\boldsymbol{\nabla}$ is equivalent here to $\boldsymbol{B}\boldsymbol{\cdot}\boldsymbol{\nabla}$ (recall equation (13.22)), it follows that

$$\boldsymbol{B}\boldsymbol{\cdot}\boldsymbol{\nabla}(\rho\kappa)\,=\,0\ , \tag{13.27}$$

as was to be proved.

After combining (13.25) and (13.26), then dividing through by the constant term $\rho\kappa$, we see that the azimuthal momentum equation (13.24) is equivalent to

$$\boldsymbol{B}\boldsymbol{\cdot}\boldsymbol{\nabla}\left[\Omega\,\varpi^2\,-\,\frac{\varpi B_\phi}{4\pi\rho\kappa}\right]\,=\,0\ .$$

We may thus write

$$\Omega\,\varpi^2\,-\,\frac{\varpi\,B_\phi}{4\pi\rho\kappa}\,\equiv\,\mathcal{J}\ , \tag{13.28}$$

where $\mathcal{J}$ is yet another quantity conserved along each field line. To interpet $\mathcal{J}$, we let $\mathcal{A}$ be the varying cross section of any flux tube. Multiplying equation (13.28) through by the mass flow rate $\dot{\mathcal{M}}\,\equiv\,\rho u_p\mathcal{A}\,=\,\rho\kappa B_p\mathcal{A}$, we find that

$$\dot{J}\,\equiv\,\rho\,u_p\,\mathcal{A}\,\Omega\,\varpi^2\,-\,\frac{\varpi B_p B_\phi}{4\pi}\mathcal{A} \tag{13.29}$$

is also conserved. The quantity $\dot{J}$ is the rate at which the z-component of angular momentum is transported along any flux tube. The first righthand term in (13.29) is the transport rate due to axial rotation of wind material, while the second term, from equation (10.41), is the contribution from the magnetic stress, *i. e.*, from the twisting of the field. It follows that $\mathcal{J}$ in equation (13.28) is the total angular momentum per unit mass within the tube.

We may now use equations (13.19) and (13.28) to solve for both B_ϕ and Ω as functions of ϖ and ρ. We find

$$B_\phi\,=\,\frac{4\pi\rho\kappa\,(\alpha\varpi\,-\,\mathcal{J}/\varpi)}{1\,-\,4\pi\rho\kappa^2} \tag{13.30a}$$

$$\Omega\,=\,\frac{\alpha\,-\,4\pi\rho\kappa^2\mathcal{J}/\varpi^2}{1\,-\,4\pi\rho\kappa^2}\ . \tag{13.30b}$$

Both denominators vanish at the *Alfvén point*. At this radius, $\varpi \equiv \varpi_A$, the product $4\pi\rho\kappa^2$ rises to unity. From equation (13.16), this condition is equivalent to

$$u_p^2 \,=\, \frac{B_p^2}{4\pi\rho} \,\equiv\, V_{\rm Ap}^2 \;. \qquad\qquad at \;\; \varpi = \varpi_A \tag{13.31}$$

That is, the poloidal velocity equals the Alfvén speed associated with B_p. Both B_ϕ and Ω remain finite at this location provided that

$$\mathcal{J} \,=\, \alpha\,\varpi_A^2 \;. \tag{13.32}$$

The ensemble of Alfvén points for all flux tubes constitutes the *Alfvén surface* surrounding the star. Within this volume, the magnetic energy density exceeds that associated with fluid motion, *i. e.*, $B_p^2/8\pi > (1/2)\,u_p^2$. The interior field is stiff, despite the inertia of the gas. According to equation (13.32), we may picture the fluid elements in this region rotating at angular speed α, right up to the Alfvén surface. Beyond this point, the system sheds angular momentum, and the field lines are swept back. This is the essence of the braking action provided by centrifugal winds.

We next consider momentum conservation along the field lines. Starting with equation (13.23), we use a vector identity to rewrite the lefthand side as

$$\begin{aligned}\rho\,(\boldsymbol{u}\cdot\boldsymbol{\nabla})\boldsymbol{u} \,&=\, \rho\,\boldsymbol{\nabla}(u^2/2) \,+\, \rho\boldsymbol{u}\times(\boldsymbol{\nabla}\times\boldsymbol{u}) \\ &=\, \rho\,\boldsymbol{\nabla}(u_p^2/2) \,-\, \rho\,\boldsymbol{\nabla}(\varpi^2\Omega^2/2) \,-\, \rho\boldsymbol{u}\times(\boldsymbol{\nabla}\times\boldsymbol{u}) \;.\end{aligned} \tag{13.33}$$

We now take the dot product of equation (13.23) with $\boldsymbol{u}$. After recalling that $\boldsymbol{u}\cdot\boldsymbol{\nabla} = \boldsymbol{u}_p\cdot\boldsymbol{\nabla}$ and using (13.33) above, we find

$$\rho\boldsymbol{u}_p\cdot\boldsymbol{\nabla}(u_p^2/2 \,+\, \varpi^2\Omega^2/2 \,-\, GM_*/r \,+\, a_T^2\,\ln\rho) \,=\, \frac{1}{4\pi}\boldsymbol{u}\cdot[(\boldsymbol{\nabla}\times\boldsymbol{B})\times\boldsymbol{B}] \;. \tag{13.34}$$

Here, we have recognized that the gravitational potential is essentially that from the star alone, and we have assumed the gas to be isothermal in evaluating the pressure gradient. Equation (13.20) for $\boldsymbol{u}$ now shows that the righthand side of equation (13.34) is

$$\begin{aligned}\frac{\alpha\varpi}{4\pi}\,[(\boldsymbol{\nabla}\times\boldsymbol{B})\times\boldsymbol{B}]_\phi \,&=\, \frac{1}{4\pi}\boldsymbol{B}_p\cdot\boldsymbol{\nabla}(\alpha\varpi B_\phi) \\ &=\, \rho\boldsymbol{u}_p\cdot\boldsymbol{\nabla}(\alpha\Omega\varpi^2) \;,\end{aligned} \tag{13.35}$$

where we have also used equations (13.22), (13.26), and (13.28). In summary, equations (13.34) and (13.35) tell us that

$$\mathcal{E} \,\equiv\, u_p^2/2 \,+\, \varpi^2\Omega^2/2 \,-\, GM_*/r \,+\, a_T^2\,\ln\rho \,-\, \alpha\,\Omega\,\varpi^2 \tag{13.36}$$

is constant along a field line. This quantity is effectively the energy per unit mass of the wind matter, with the final term representing the work done by the twisted magnetic field. Of course, the true mechanical plus thermal energy is *not* conserved, even in the absence of the field, since heat must somehow be injected to maintain isothermality of the expanding gas.

The results we have obtained thus far are all useful, but still insufficient for the full description of any real system. One problem, of course, is that we have said nothing about conditions at the base of the wind, such as how the field lines rotate near the stellar surface. Of more fundamental concern is the balance of forces within the meridional plane, but *transverse* to the field lines. This latter issue is difficult enough that it has prevented any general, fully self-consistent solution to the problem of centrifugal winds. Theorists have been forced to adopt a variety of stratagems, all of them involving rather severe approximations or physical assumptions.

One route is to posit *ad hoc* a spatial configuration for the poloidal magnetic field. The simplest choice, and the first used historically, is the Weber-Davis model. Here one assumes the field to be a split monopole, and considers the flow only near the equatorial plane. Alternatively, one could assume a dipole field, at least close to the stellar surface. We shall not pursue any of these models in more detail, but simply note an interesting feature that they all share.

13.4.3 Critical Points

Suppose that we wish to determine the rate of change of the poloidal velocity in the direction of the local field. That is, if s represents the distance along any flux tube, we want du_p/ds (see Figure 13.17). This may be read from the momentum equation (13.34) or, alternately, by setting $d\mathcal{E}/ds$ equal to zero in (13.36). The latter operation still does not give us du_p/ds directly, since the equation contains other variables, such as Ω and ρ. To gauge their variation, we similarly demand that $d\alpha/ds$, $d\mathcal{J}/ds$, and $d\dot{\mathcal{M}}/ds$ vanish, using the appropriate definitions for each quantity. The net result may be cast as four equations in the four unknowns du_p/ds, $d\Omega/ds$, $d\rho/ds$, and dB_ϕ/ds.

We may now solve these equations for the derivative of interest. Omitting the rather lengthy manipulations involved, we may conveniently express the final result as

$$\frac{\varpi}{u_p}\frac{du_p}{ds} = \frac{c_1\, d\varpi/ds \,+\, c_2\, dr/ds \,+\, c_3\,(\varpi/\mathcal{A}) d\mathcal{A}/ds}{\mathcal{D}} \ . \tag{13.37}$$

The nondimensional terms $d\varpi/ds$, dr/ds, and $(\varpi/\mathcal{A})\, d\mathcal{A}/ds$ all quantify the degree of bending and flaring of the flux tube. Their functional forms may be taken from one of the phenomenological models, and will also result naturally from the eventual, self-consistent calculation. The three coefficients in the numerator are

$$c_1 \equiv (u_p^2 - V_{\rm Ap}^2)\, u_\phi^2 \,+\, 2\, u_p\, u_\phi\, V_{\rm Ap}^2 \tan^2\theta_B \,-\, 2\, u_p^2\, V_{\rm Ap}^2 \tan^2\theta_B \tag{13.38a}$$

$$c_2 \equiv -2\,(u_p^2 - V_{\rm Ap}^2)\, a_T^2\, \varpi\, R_s/r^2 \tag{13.38b}$$

$$c_3 \equiv (u_p^2 - V_{\rm Ap}^2)\, a_T^2 \,+\, u_p^2\, V_{\rm Ap}^2 \tan^2\theta_B \ . \tag{13.38c}$$

Here, θ_B is the angle between the poloidal magnetic field and the full vector $\boldsymbol{B}$ (Figure 13.17). The denominator in equation (13.37) is

$$\mathcal{D} \equiv u_p^4 \,-\, (V_A^2 + a_T^2)\, u_p^2 \,+\, a_T^2\, V_A^2 \cos^2\theta_B \ . \tag{13.39}$$

Note that V_A denotes the Alfvén velocity associated with $\boldsymbol{B}$ itself, as in equation (9.70).

The quantity $\mathcal{D}$ vanishes at certain points in the flow. We are reminded of the isotropic, thermally-driven wind, where the acceleration diverges unless the sound speed is reached precisely at R_s. We also recall the dispersion relation for magnetosonic waves, equation (9.82). Dividing the latter by k^4, we see that the phase velocity ϖ/k enters the dispersion relation in precisely the same way as u_p enters $\mathcal{D}$ in equation (13.39). We have found, in other words, that this denominator vanishes whenever u_p equals the velocity for fast or slow magnetosonic waves, where these waves are directed along $\boldsymbol{B}_p$. At both these points, the numerator of equation (13.37) must also go to zero.

We saw earlier, in equations (13.30), that Ω and B_ϕ exhibit similar behavior as u_p reaches the Alfvén velocity associated with the poloidal field. Centrifugal winds thus possess three *critical points*. These occur physically whenever the velocity matches one of the characteristic signal speeds in the fluid. The sensitivity of the flow to small disturbances prior to a critical point reflects the fact that disturbances traveling at the appropriate speed can still propagate back upstream. Once matter has crossed the point, however, such backward propagation is no longer possible. The sonic point in a thermally-driven wind is reached when the velocity is a_T, which is the only signal speed in a non-magnetized, isothermal gas. More generally, the velocity field must pass, in order, through the slow magnetosonic, Alfvénic, and fast magnetosonic critical points if it is to describe a true wind. Note that the acceleration of the gas is primarily thermal inside the slow magnetosonic point, while the combined magnetic and centrifugal forces take over until the Alfvén point is reached.

These considerations also lead to the conclusion that a strongly magnetized wind must be anisotropic. Even if matter within certain flux tubes accelerates smoothly through the critical point, gas elsewhere may not. Whether a region participates in the wind or not depends on the competition between the local field strength and the inertia of the gas. The overall topology of the flow is therefore sensitive to the magnetic field structure at the wind base.

To illustrate this last point, Figure 13.18 depicts schematically a star whose field is a perfect dipole close to the stellar surface. Let us ignore the temperature of the external matter, so that only the Alfvénic critical point applies. At all latitudes, u_p starts out much smaller than V_{Ap}. Thus, the field is relatively stiff, and gas flow initially follows the dipolar configuration. The field lines close to the pole diverge quickly enough that the inequality $u_p > V_{\mathrm{Ap}}$ is eventually met. Gas within these flux tubes crosses the Alfvén point and thereafter flows out more radially, dragging the field with it. Closer to the equator, however, the field lines loop back before the gas can attain the requisite velocity. The matter here remains in hydrostatic balance along the rigid field loops. The outer boundary between this *dead zone*, shaded in Figure 13.18, and the wind region is labeled A. This point represents the intersection of the wind's Alfvén surface with the equatorial plane.

The analogue of the dead zone in solar studies is, of course, the well-observed magnetic loop structure associated with the corona. Just as the solar loops are sites of enhanced X-ray emission, one may hope that wave energy pumped into the closed field lines of young stars helps to account for their own activity in this regime. Whatever the appeal of such ideas, we should recognize that the dead zone itself is still largely a theoretical construct, with scant empirical basis. Let us ask, more generally, to what extent the combination of rotation and magnetic fields is required to explain winds from young stars. With regard to the late pre-main-sequence history of our own Sun, estimates based on the Weber-Davis model and the current $\dot{M}_w$-value

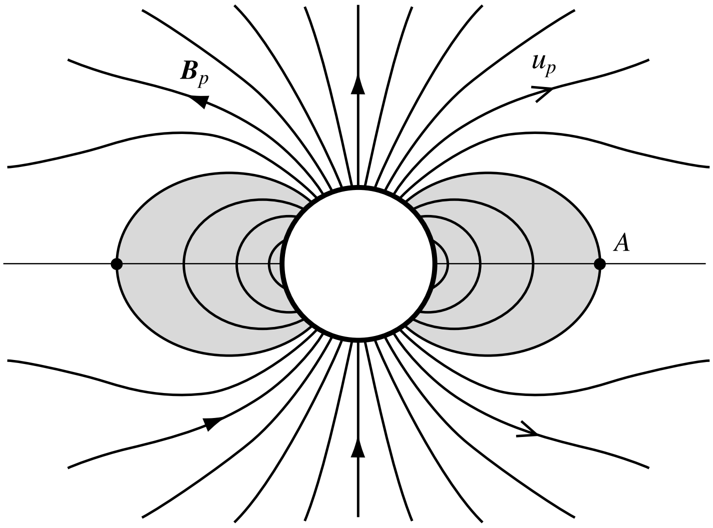


Figure 13.18 Schematic representation of wind flow from a magnetic field that is dipolar at the stellar surface. Filled arrowheads show the direction of the poloidal field, while open ones depict the same component of the fluid velocity. Closed field lines intersect the equatorial plane out to the point A. Within this region is the dead zone, which does not participate in the flow.

give reasonable values for the magnetic braking timescale, *i. e.*, comfortably less than the Sun's estimated age. This is an encouraging sign that the observed pattern of rotational speeds along the main sequence indeed arises from the same cause. However, solar-type winds are still essentially thermally driven. That is, the measured surface velocities are far too small to aid in the outward acceleration of gas. The same is true for younger pre-main-sequence stars. With its typical surface velocity of 10 km s^{-1}, a T Tauri star is rotating at less than 10 percent of breakup speed, so that the associated centrifugal force is at most a few percent of gravity.

13.4.4 The Role of Disks

What, then, provides the relatively high values of $\dot{M}_w$ for these objects? Faced with the inadequacy of thermal and wave pressure, as well as the low stellar rotation rates, a number of theorists have adopted the view that the lion's share of $\dot{M}_w$ does not originate from the star at all, but from its surrounding disk. The idea is that matter spirals in toward the star from this disk, most of which lies beyond point A in Figure 13.18. As the dead zone is approached, some of this incoming gas flows along the closed loops toward the stellar surface, while the rest is ejected as a massive wind. This ejection occurs along open field lines that are anchored in the disk. The acceleration here may indeed be centrifugal, since the disk gas is already rotating very close to breakup speed. There is also a relatively minor component of the wind stemming from the polar regions of the star itself.

For this picture to hold, pre-main-sequence stars must undergo simultaneous inflow *and* outflow of material. We will see in Chapter 17 that emission-line and continuum data indeed show that both processes may be occurring, at least in *classical* T Tauri stars. Even here, however, the observations of line profiles do not demonstrate a centrifugal flow *per se*. Nor is there any direct evidence that the stars' disks are threaded by strong magnetic fields. Thus far, the only sign that disks are involved at all is the intriguing fact that the diagnostic signs of winds and infall are *not* present in *weak-lined* stars. The latter, we recall, exhibit little or no infrared excess, the implication being that they lack circumstellar disks.

It is not our intention to review here the various detailed models that have been offered for disk-generated winds. We limit ourselves instead to several observations of a more general nature. Suppose that the central idea is correct, *i. e.*, that disks are indeed magnetized in the proper way and that pre-main-sequence winds originate mainly from these structures rather than from the stars. Then our original question has not really been answered, but rather transformed into another: What accounts for the required mass flow rates through the disks themselves? Despite much discussion of this point, there is still little understanding of how material orbiting a T Tauri star might steadily drift toward the central object. Ironically, the situation may be more favorable in the deeply embedded stars driving jets and molecular outflows. If some of these sources are true protostars, then a portion of the surrounding infall must strike their disks. As we saw in Chapter 11, continuing infall onto a disk is expected to create spiral disturbances that transfer gas inward. This disk would then act as an intermediary between the collapsing envelope of cloud gas and the ejected spray of matter within. If the same mechanism applies to classical T Tauri stars, then they, too, must be surrounded by envelopes that continually fall inward, albeit at a reduced rate. (Recall the discussion of Class I sources in Chapter 11.)

The transfer of matter to the star along closed field lines is one of the most appealing aspects of the picture, yet also raises a serious concern. To see how the accretion process might operate, imagine the nearly rigid loops of the dead zone rotating uniformly at the stellar rotation rate Ω_*. (In this case, the value of α on each closed flux tube would just be Ω_*.) Material far out in the disk has an angular speed much lower than this. As the gas spirals inward, however, $\Omega(\varpi)$ increases until the two rates match. It is within this *corotation radius* that accretion may occur. The magnetic field brakes the incoming gas so that its angular velocity no longer rises, but remains at Ω_* until that element lands on the star. Since this braking also reduces the fluid element's centrifugal support, the poloidal inflow velocity is a substantial fraction of free fall. Such high speeds are indicated observationally by the presence of hot spots on the stellar surface, as we shall discuss later.

The difficulty here is that the specific angular momentum of the gas, $\Omega\varpi^2$, initially exceeds that of the stellar surface. It is true that, once Ω becomes fixed, $\Omega\varpi^2$ declines during infall. However, equation (13.28) tells us that the *total* specific angular momentum is conserved along the flux tube. If B_ϕ is small initially, any subsequent braking of the gas is compensated by a rising magnetic stress that ultimately torques the stellar surface. In other words, Ω_* itself must increase as a result of accretion. But prolonged spinup would violate the observation of low rotational speeds. The resolution here may simply be that pre-main-sequence stars never acquire a great deal of additional mass through this accretion process.

Another general difficulty faced by the picture we have outlined is its implicit assumption of *steady-state* flow. All numerical simulations of the star-disk interaction via magnetic fields find that the gas moves both inward and outward in an intermittent, eruptive fashion. Consider, for simplicity, a purely dipole field emanating from the star and threading a conducting disk.

If the latter is initially truncated beyond the corotation point, then the field quickly wraps up. The resulting torque on disk material may well cause it to move inward, as in the steady-state models. On the other hand, the twisted field is also susceptible to violent reconnection.

Figure 13.19 shows in more detail the sequence of events. In this particular simulation, the central 1 $M_\odot$ star rotates with a period of 1.8 days, and the disk is truncated at 5.7 stellar radii. The panels display the poloidal field structure and density distribution at the indicated times. Even at 1.4 days, before one complete stellar rotation, twisting of the initially dipolar field has created severe, equatorial pinching. The stressed disk has moved inward toward the star in the second panel. By 4.2 days, reconnection of the pinched field has created two closed, toroidal loops. The centrifugal force acting on the rotating plasma impels it to move outward. During this expansion, some gas on the closed field lines is driven toward the central axis and forms a jetlike flow. Other gas skims along the disk surface. Following ejection of the toroids, the disk edge retreats, the now-relaxed stellar field wraps up again, and the process repeats itself.

Such calculations demonstrate how the presence of a disk (or any other circumstellar matter) greatly complicates the flow, even for a relatively simple stellar field. What happens outside the star and inner disk is unclear, apart from certain general features of the flow. Any ejected toroids must eventually decay, so that the wind resumes a spatially smooth, but still anisotropic, distribution. This outgoing matter drags the weakened field along with it. Since the interior field is still being wrapped up, the net effect is that $\boldsymbol{B}$ is swept backward, as in the highly simplified sketch of Figure 13.17. That is, a strong component B_ϕ is maintained outside the Alfvén surface, regardless of the precise interior conditions. Ampère's law then tells us that there must exist a current $\boldsymbol{j}$ directed along the central axis. The volumetric force acting on the gas, $\boldsymbol{j}/c \times \boldsymbol{B}$, thus has a radially inward component. In other words, the winding up of the field creates a *magnetic pinch*.

Whether the pinch impels most of the gas to turn toward the axis, forming a massive jet, depends on a number of factors. The inward force must compete against the outward centrifugal force associated with u_ϕ, as well as the gas and magnetic pressure gradients from fluid closer to the central axis. A strongly twisted field may also be intrinsically unstable, even if it is not anchored to a disk. After undergoing reconnection and Ohmic dissipation, the field would relax to a more open, untwisted structure.

Determining the asymptotic form of the wind launched by the star and inner disk may ultimately prove to be a rather academic exercise. The most active young stars are surrounded by copious amounts of matter. Indeed, we shall argue in Chapter 17 that the primary emission-line diagnostics of T Tauri winds arise not from the winds themselves but from their *interaction* with circumstellar gas. For the even more embedded stars that drive molecular outflows, it may suffice to know that the wind is launched anisotropically and in a bipolar pattern. We now explore the processes governing such a flow as it travels to much greater distances within the cloud medium.

13.5 Jet Propagation and Entrainment

A major question for theory is how jets maintain their remarkable, collimated structure. Additionally, we want to know how they interact with cloud matter to produce the much broader

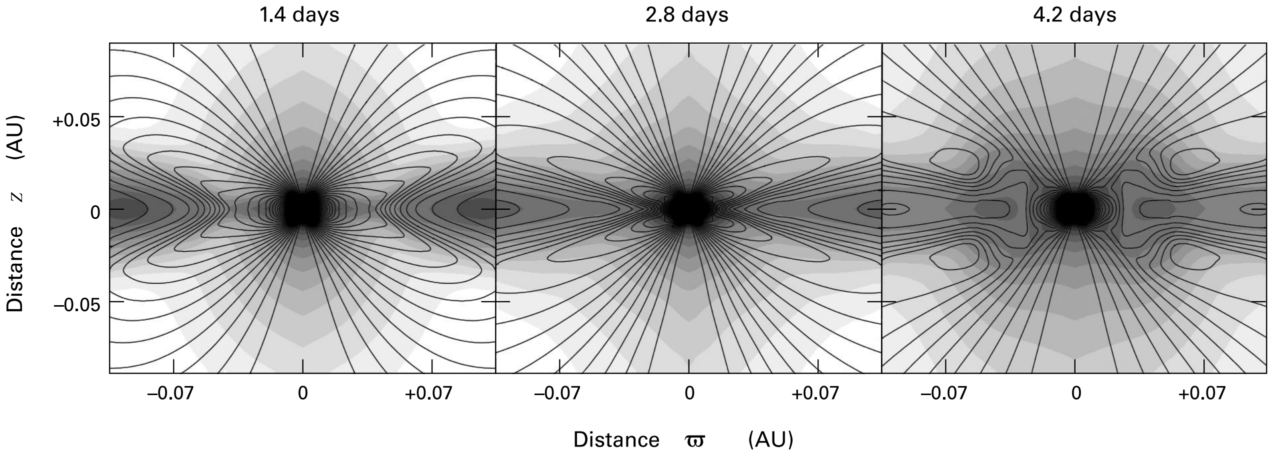


Figure 13.19 Numerical simulation of star-disk interaction. Contours trace the poloidal magnetic field, while greyscale shading represents the ambient density, shown in logarithmic intervals. The star, which has an initially dipolar field, is rotating with a period of 1.8 days. Results are shown at the three indicated times.

molecular outflows. Neither problem has a definitive, generally accepted solution. The situation is akin to that regarding the magnetized collapse of dense cores, as described in Chapter 10. In both cases, we can describe physical effects that almost certainly play a major role, but cannot yet tie them together into a fully coherent picture.

13.5.1 Mass Transport Rate

Before tackling these general issues, let us begin with a purely empirical matter: What is $\dot{M}_j$, the actual rate of mass transport along a typical jet? We have seen that the internal speed varies spatially, and that this change produces the traveling shocks observed. Hence $\dot{M}_j$ must change as well, and no single value will suffice in a complete description. A spatially averaged transport rate is still meaningful, however, even if it can only be ascertained with limited precision.

Let $V_\perp$ denote the velocity of material within some observed portion of the jet, as projected onto the plane of the sky. This quantity is obtained from the proper motion of emission knots. Similarly, let $\Delta L_\perp$ be the projected length of this segment. Then $\Delta L_\perp / V_\perp$ is the travel time, independent of the jet's inclination with respect to the plane of the sky. The transport rate is simply

$$\dot{M}_j \;=\; \Delta M_j \, V_\perp / \Delta L_\perp \ , \tag{13.40}$$

where ΔM_j is the jet mass contained within the segment. Most of this mass consists of hydrogen, of course, but we can still find it through the emission of trace species.

A good candidate is atomic oxygen, whose emission via the [O I] 6300 Å line is collisionally excited by ambient electrons. At the characteristic postshock emission temperature of 8000 K, the critical density for collisional deexcitation is 2×10^6 cm^{-3}. Since, as we shall see, the actual electron density is far less, each collisionally induced upward transition is promptly followed by photon emission. Hence we may apply equation (7.24) to obtain the volumetric cooling rate, after replacing n_H on the righthand side by n_e. Multiplying through by the relevant emitting volume, we obtain the luminosity in the line:

$$L_{[\mathrm{OI}]} \;=\; \frac{g_u}{g_l} N_l(\mathrm{O\ I})\, n_e\, \gamma_{\mathrm{ul}}\, \Delta E_{\mathrm{ul}}\ \exp(-T_\circ/T_g) \ . \tag{13.41}$$

Here, as in equation (7.24), u and l denote the upper and lower states, respectively, of the transition, $T_\circ$ is the equivalent temperature of the transition (2.2 × 10^4 K), and $N_l(\mathrm{O\ I})$ is the population in the lower state. Since the fraction within the upper state is relatively small, $N_l(\mathrm{O\ I})$ is essentially the total number of oxygen atoms. Thus, if we utilize the known fractional abundance of the element (see Table 2.1), we may invert equation (13.41) to obtain the desired mass ΔM_j in terms of the observed luminosity $L_{[\mathrm{OI}]}$ and the still unknown electron density n_e.

To obtain n_e, we consider emission in the doublet [S II]$\lambda\lambda\, 6716, 6731$. The upper levels here are both excited from the same electronic ground state of S II, again through collisions with ambient electrons. The critical densities at 8000 K, however, are much lower: 4×10^4 cm^{-3} for the 6716 Å transition, and 1×10^4 cm^{-3} for 6731 Å. If the actual n_e were far below either n_{crit}-value, equation (7.24) would predict that the *ratio* of line luminosities would be the corresponding ratio of the product $g_u\,\gamma_{ul}$, since the energies ΔE_{ul} are nearly identical. Conversely, the luminosity ratio would depend on $g_u\,A_{\mathrm{ul}}$ in the supercritical case, according to equation (7.25). One must solve the full set of equations for statistical equilibrium to obtain the line ratio in the general case, but the answer will depend only on the kinetic temperature and n_e. Conversely, knowing the temperature to be about 8000 K, we may use the observed ratio to infer n_e itself.

Consider, as an example, the main body of the HH 34 jet (Plate 11). The ratio of fluxes in the [S II] doublet yields an n_e of 650 cm^{-3}, while the knot proper motions indicate $V_\perp = 200\,\mathrm{km\,s^{-1}}$, over a projected length of $\Delta L_\perp = 1.4 \times 10^{16}\,\mathrm{cm}$. The observed luminosity in [O I] 6300 Å is $1.2 \times 10^{-4}\,L_\odot$, from which equations (13.40) and (13.41) give an $\dot{M}_j$ of $1.7 \times 10^{-7}\,M_\odot\,\mathrm{yr}^{-1}$. One obtains a similar value by comparing observed line ratios with those predicted by detailed numerical models of shock cooling. The latter technique shows, incidentally, that the *total* density in the emitting region is 6×10^4 cm^{-3}, so that the ionization fraction is only a few percent.

Such levels for the mass transport rate exceed, by an order of magnitude, those of optically visible T Tauri stars (obtained by precisely the same method; see Chapter 17). On the other hand, the jets from Class 0 sources, such as L1448, appear to have even higher $\dot{M}_j$-values, of order $10^{-6}\,M_\odot$ yr^{-1}. Here one cannot rely on optical lines as mass or velocity tracers. One turns instead to the infrared [O I] 63 µm emission that ultimately provides the bulk of postshock cooling for *all* the matter below about $T_{\rm cool} \equiv 5000$ K. (Recall the discussion in § 8.4.) That is, the luminosity in this line should roughly equal $\dot{M}_j\,\Delta E$, where $\Delta E = k_B T_{\rm cool}$. While individual values for $\dot{M}_j$ from either technique are naturally uncertain, their relative ordering bolsters the idea that the wind momentum output falls as the driving star becomes more revealed.

13.5.2 The Working Surface

Let us return to the issue of wind variability and to the character of the flow that results. Figure 13.20 consists of four snapshots from a one-dimensional simulation. Initially, the velocity profile is a single Gaussian perturbation superposed on a uniform flow. (This base speed is not shown.) The symmetric character of the profile soon changes, as more rapidly moving elements overtake slower ones. By $t = 1.98$ yr, the velocity rises almost linearly to a peak and then drops sharply, representing the creation of a shock front. By the last panel it is evident that *two* shocks have been created from the original smooth fluctuation.

This finding is quite robust. That is, the usual result of continually varying the input velocity is *not* to produce a train of single shocks, but of traveling shock pairs. To see why, consider first the idealized case of a jet impacting a perfectly stationary medium head on. The Rankine-Hugoniot jump conditions give us, from knowledge of the incoming density, velocity and temperature, the pressure arising just beyond the shock front. If the jet density is not greatly different from that of the environment, and if the incoming speed is highly supersonic, then this postshock pressure greatly exceeds the ambient one. Hence their difference drives a second shock downstream from the first.

The same conclusion holds for the more general and relevant case of a moving preshock medium, as may be seen by transforming to a reference frame in which that medium is stationary. Now in a truly one-dimensional flow, like the one in Figure 13.20, the excess pressure between the fronts eventually forces them to spread apart, and the two shocks concurrently weaken as they propagate downstream. In a jet of finite cross section, however, any overpressured material is driven sideways from the main flow direction, and the shocks maintain equal

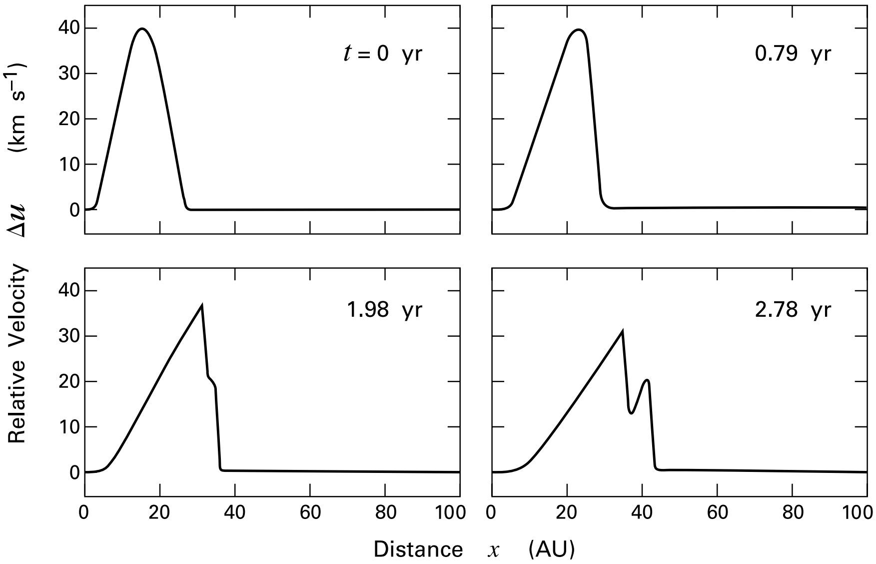


Figure 13.20 Formation of a double shock within a variable jet. The four frames depict successive times within a one-dimensional flow. In each, the velocity relative to the base jet speed is shown as a function of distance, where the latter is always reset to zero at the start of the pulse.

speed. Figure 13.21 illustrates this situation in a highly schematic fashion. We refer to the entire double-shock structure as the *working surface*. Incident jet gas of density ρ_1, moving at speed V_1, encounters directly not the downstream gas with density ρ_2 and lower speed V_2, but intershock material at the elevated pressure $P_{\rm high}$. Impact occurs at the *Mach disk*, also known as the *jet shock*. Gas traversing this front drives into the downstream flow and creates the curved bowshock on the right. In between the two fronts, material spills out laterally, as indicated in the figure. This cooling waste gas collecting along the side of the jet is called the *cocoon*.

The working surface is effectively a single pulse, created by the original change in velocity, that travels in the flow direction. What is $V_{\rm pulse}$, the velocity of the disturbance itself? Referring again to Figure 13.21, we see that $V_{\rm pulse}$ must be intermediate between V_1 and V_2. To an observer moving with the two shocks, matter flows across the Mach disk with speed $V_1 - V_{\rm pulse}$. Additionally, material crosses the bowshock into the high-pressure region with speed $V_{\rm pulse} - V_2$. There is thus a continual buildup of intershock material, a gain compensated by lateral ejection into the cocoon. The pressure between the shocks, $P_{\rm high}$, must match both the upstream ram pressure $\rho_1\,(V_1 - V_{\rm pulse})^2$ and the downstream value $\rho_2\,(V_{\rm pulse} - V_2)^2$. Here we are neglecting the relatively small effect of the finite gas temperatures outside the shocks. Equating the two ram pressures, we solve for $V_{\rm pulse}$ to find

$$V_{\rm pulse} = \frac{\beta V_1 + V_2}{1 + \beta} \,, \tag{13.42}$$

where $\beta \equiv \sqrt{\rho_1/\rho_2}$. Note that if $\beta \gg 1$, we have $V_{\rm pulse} \lesssim V_1$. That is, a dense jet plows into

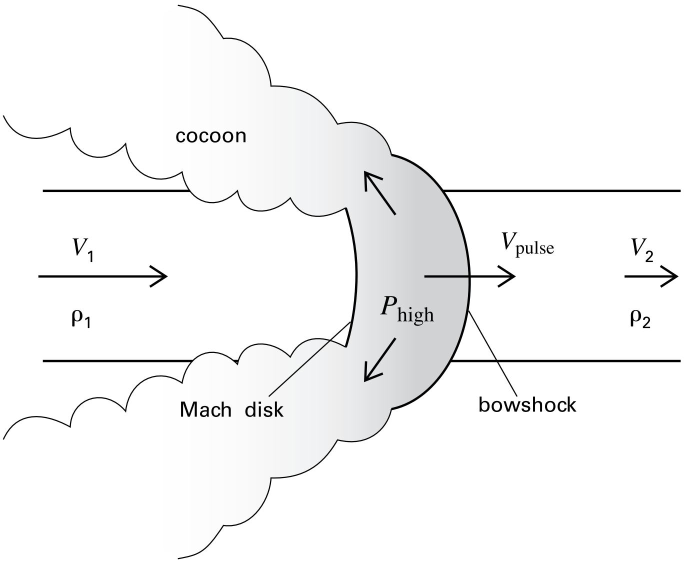


Figure 13.21 Structure of the working surface in a variable jet (*schematic*). Gas of density ρ_1 and speed V_1, emitted by a star on the left, encounters a flow with density ρ_2 and lower speed V_2. The large value of pressure P_{high} in the intershock region expels material into the cocoon. Both the Mach disk and the bowshock travel at the speed V_{pulse}, intermediate between V_1 and V_2.

the medium ahead of it with virtually undiminished speed. Conversely, the pulse in a rarefied segment of the jet ($\beta \ll 1$) advances only slowly into the upstream gas, since $V_{\text{pulse}} \gtrsim V_2$.

High-resolution images of jet knots often show distinctive bowshocks, with only faint Mach disks. (Recall again the case of HH 34, as seen in Plate 11.) Why should one front within a working surface be brighter than the other? Our derivation of equation (13.42) assumed equality of the *momentum* fluxes. The rate of *energy* dissipation at a front, however, is proportional to the incoming density times the *cube* of the relative velocity. For the case of a heavy jet (large β), we have $V_1 - V_{\text{pulse}} < V_{\text{pulse}} - V_2$. The luminosity of the Mach disk should indeed be smaller in this circumstance. We conclude that a jet displaying a series of relatively bright bow shocks and only faint Mach disks must also have an internal density that falls in the downstream direction.

The widest bowshocks observed have a clumpy appearance and display short-term temporal fluctuations. Here the cloud environment must play some role, particularly in explaining marked differences between the two lobes in a given system. There is also a thermal effect of significance. In our idealized description of the working surface, we have treated the intershock material as if it were a homogeneous fluid. One should bear in mind, however, that two flows with distinct properties enter the region and are unlikely to mix completely. That from the weaker shock enters at a lower temperature and cools more rapidly, tending to form a dense shell. Numerical simulations find that this shell breaks apart even as fresh material joins it. Such rapid, ongoing fragmentation plausibly underlies both the clumpiness and intermittency of the observed radiation.

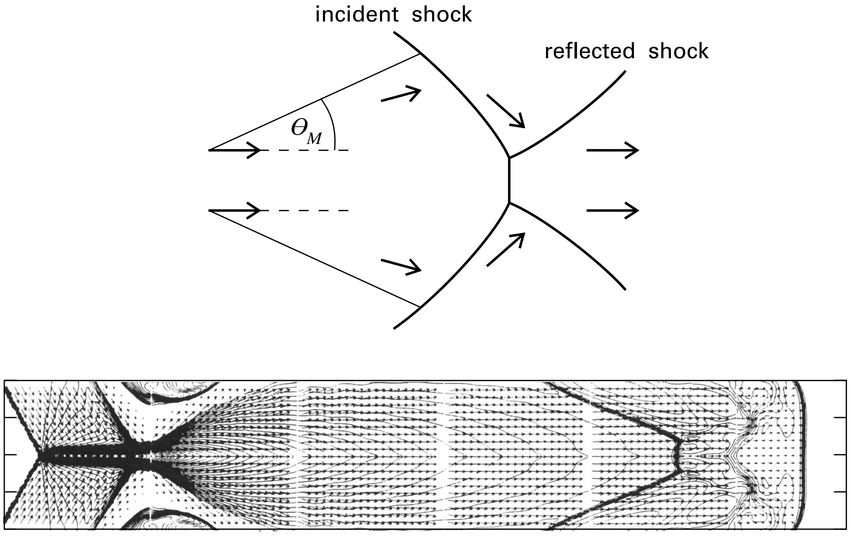


Figure 13.22 (*top*) The creation of incident and reflected shocks by a jet diverging at the Mach angle θ_M. (*bottom*) A jet of initial speed 100 km s^{-1} and number density 100 cm^{-3} converges at the left, and then flows along the central axis. By the time shown ($t = 2070$ yr), the jet has produced incident and reflected shocks at the right. Solid contours represent density and arrows velocity, for which only the direction is significant. Each tick mark on the boundary corresponds to 1×10^{16} cm.

13.5.3 Crossing Shocks

Although it is difficult to gauge the temperature *between* successive pulses, the value cannot fall much below 10^4 K in sections of the jet displaying more or less contiguous optical emission. Since, as we have noted, the internal density is comparable to that in a cloud core, the jet must be significantly overpressured with respect to its environment in these regions. We are thus led back to the collimation issue. What creates the narrow structures observed?

The mystery deepens once we realize that even an idealized, parallel jet beam must spread laterally at the local sound speed a_T. That is, a jet moving with average speed V_j diverges at least with the *Mach angle*, given by

$$\theta_M \equiv \sin^{-1}\left(a_T/V_j\right) . \tag{13.43}$$

The top portion of Figure 13.22 illustrates the geometric situation. Using a representative V_j of 300 km s^{-1} and temperature of 10^4 K yields a value for θ_M of 2°. Thus, a jet propagating for 0.1 pc should increase its full width by at least 1200 AU, contrary to observation.

Suppose, however, that the jet were actually to begin expanding in this manner. Then both laboratory results and theoretical calculations show what would happen. The initial flaring rapidly lowers the internal density, until the pressure falls to the external value. At this point, lateral expansion is halted by an *incident shock*, as sketched in Figure 13.22. A jet streamline crossing this oblique, curved shock front refracts so that it turns back toward the central axis.

This converging flow is still highly supersonic. Hence, a second oblique front, the *reflection shock*, arises to turn the flow parallel to the axis. In diverting the flow, however, this second shock elevates the pressure above the environmental value. Hence the initial condition at the injection point— a well-collimated jet entering at high pressure— is repeated. The jet again diverges with half-angle θ_M, and a new set of shocks appears farther downstream. In this way, a sequence of internal *crossing shocks* is generated as the jet bores its way into the ambient cloud.[4]

The bottom panel of Figure 13.22 is a snapshot from a time-dependent simulation of a supersonic jet penetrating an initially static medium. In this particular example, the initial high pressure was achieved by introducing, at the left, a flow converging toward the axis. The convergence sets up the complex pattern of shocks in this portion of the frame, but the net result is simply an overpressured jet moving to the right. A pair of incident and reflected shocks is evident just before the main bowshock, which marks the furthest advance of the fluid at this time.

Given their ubiquitous nature, it is difficult to escape the conclusion that crossing shocks indeed arise in the astrophysical setting, and that they help to collimate the jet. The theoretical calculations find that the first incident shock is the strongest, while the others progressively weaken. Any initial redirection of the wind may occur within an environment so dense that no optical emission can escape. A strong enough shock, however, would emit free-free radiation from hot, postshock electrons that *can* be detected. It is tempting to identify this energetic output with the radio jets seen at centimeter wavelengths.

Despite such tantalizing hints, there is still no clear identification of crossing shocks in the less obscured portion of the flow, where they might show up optically. These fronts should be stationary with respect to the driving star, a feature that distinguishes them from the traveling shocks generated by velocity fluctuations. There has been little theoretical study, in fact, of the crossing shocks arising from a time-dependent flow. Apparently, the actual fronts are sufficiently oblique that their emission is quite low. The pressure from the ambient magnetic field could also help to stem lateral expansion, in which case the crossing shocks would be weaker still.

13.5.4 Cloud Entrainment

Laboratory and theoretical investigations of steady jets also provide a clue to the origin of molecular outflows. We know that the latter consist of cloud material residing outside the jet proper. However, even if this environmental gas were initially static, the jet-cloud interface would be subject to the *Kelvin-Helmholtz instability*. Any small warping of the boundary between two fluids in relative motion grows rapidly with time. Here, the result is that streaming jet gas partially intermingles with the cloud medium. In the well-studied terrestrial jets, a turbulent sheath arises at this border. This *mixing layer* grows in thickness along the jet direction. The laminar

---

[4] As Figure 13.22 shows, the incident and reflected shocks may not intersect on the central axis, but at a third shock transverse to that axis. Fluid dynamicists refer to this situation as *Mach reflection* and to the transverse shock as the Mach disk. We will follow astronomical convention and reserve the latter term for the first shock within the traveling working surface created by a variable jet.

core of the jet concurrently narrows, until the entire flow is eventually turbulent. In this way, the injected stream of high-speed gas drags forward, or *entrains*, a portion of its environment.

The idea that molecular outflows represent cloud gas entrained by the central jet is an appealing one. As indicated in § 13.2, a jet and its outflow do tend to have comparable rates of momentum transfer. The jet associated with the deeply embedded star HL Tau, for example, has a momentum flow rate estimated at $2 \times 10^{-5}\ M_\odot\ \mathrm{yr}^{-1}\ \mathrm{km\ s}^{-1}$, a bit over twice the value of F_{out} for its CO outflow. Of course, such a relation would be expected to hold for *any* process that efficiently tapped the momentum of the jet. A more difficult question is whether the specific kind of turbulent entrainment found in steady jets indeed generates molecular outflows.

One problem with making this identification is that the mixing layers created by steady jets are narrower than the broad expanses seen in most (but not all) molecular outflows. In addition, there is no documented case astrophysically where the central jet is choked off, after having transferred *all* of its momentum to the cloud gas. The observations of giant Herbig-Haro flows indicate quite the opposite, that a jet may persist well beyond the borders of its parent cloud. We have noted that such protruding jets may be diverging ballistically. The most promising explanation for the width of outflows inside clouds is that, while they are indeed created by narrow jets, the underlying winds are *not* steady, but varying with time. A spray is ejected laterally from each working surface, as sketched in Figure 13.21. This transverse flow could stir up cloud gas to an appreciable distance. Once again, numerical simulations of variable jets have been useful for testing this idea.

Returning to the simplified case of steady jets, the entrainment hypothesis helps us to understand the observed systematics of outflow velocities. Within laboratory entrainment flows, the velocity (suitably time-averaged to eliminate the random, turbulent component) is always highest close to the jet. This situation is, of course, reminiscent of the astrophysical one. The spatial distribution of this velocity is such that it is constant on nested, conical surfaces centered on the jet axis. Furthermore, most of the volume within the flow is occupied by gas moving at the lowest speeds. Figure 13.23 illustrates this situation with three velocity values; here $V_1 < V_2 < V_3$ and $V_2 - V_1 = V_3 - V_2$. We must bear in mind that the actual velocity vectors point not along the cones, but in the jet direction.

Suppose that an observer positions the telescope beam a distance x_1 from the driving star. Assuming the relevant spectral line to be optically thin, the received intensity within any velocity interval is proportional to the column density. Then it may be that the outflow material with $V_1 < V_r < V_2$ is detectable, while that between V_2 and V_3 has too small a column density. Moving the telescope beam farther out to x_2 increases both signals and allows detection of the higher-velocity gas. A rise in both the maximum and average speeds along the central axis is just the pattern seen in the data. (Recall Figure 13.12.)

The increase of the top speed is only apparent, as all velocities are actually present at each position. What, then, is the true spatial distribution of the velocity within an outflow lobe? We know from equation (13.7) that the total *mass* traveling at speeds close to any V_r-value falls off systematically with that speed. Given this information, the spatial distribution must be

$$V_r \,=\, V_*\,\eta^{-2/(\gamma-1)}\ . \qquad (13.44)$$

Here V_* is a fiducial velocity that differs from one outflow system to another. The nondimensional variable η is defined to be $\varpi/\varpi_\circ(x)$, where ϖ is the transverse distance from the jet axis

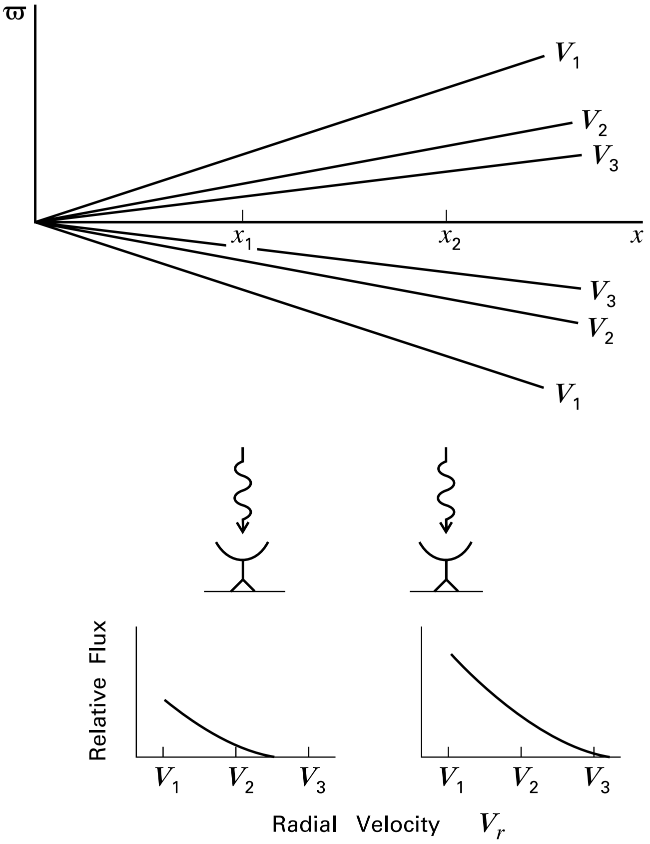


Figure 13.23 The nature of the acceleration within entrained outflows. The velocity is constant along cones centered on the jet axis; three representative values are shown. A telescope positioned at x_1 detects a finite intensity for $V_1 < V_r < V_2$, but none at the higher radial velocity interval $V_2 < V_r < V_3$. Both intervals become detectable at the downstream position x_2.

and $\varpi_\circ(x)$ is the lobe width at each point, *i. e.*, the ϖ-value at which V_r falls to the background value $V_\circ$. (This function is proportional to x itself in the case of a perfectly conical outflow.) Finally, γ is the empirical exponent introduced in equation (13.7).

To see why equation (13.44) should hold, consider an annular slice of an outflow lying perpendicular to the axis (see Figure 13.24). Let $m(V_r, x)\,\Delta V_r\,\Delta x$ be the amount of mass located between x and $x + \Delta x$ from the stellar source, and between V_r and $V_r + \Delta V_r$ in radial velocity. Let $\rho(x)$ be the cloud's mass density. Then we have

$$m(V_r, x) \;=\; -2\,\pi\,\varpi\,\rho(x)\,\left(\frac{\partial V_r}{\partial \varpi}\right)_x^{-1} \;. \tag{13.45}$$

The minus sign accounts for the fact that the partial derivative in radial velocity is negative.

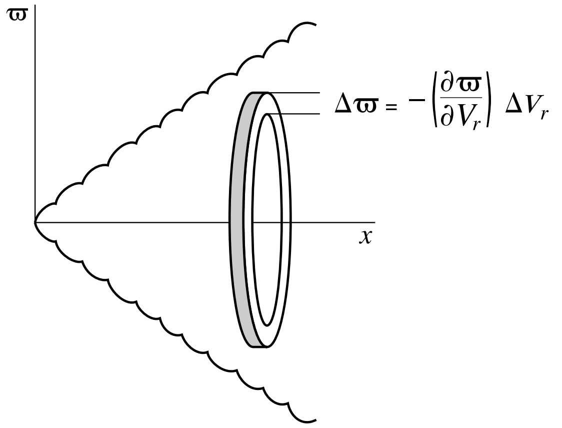


Figure 13.24 Annular slice of an outflow lobe. Note the minus sign in the relation between the width of the annulus, $\Delta\varpi$, and the corresponding interval of radial velocity ΔV_r.

After evaluation of this derivative using equation (13.44), we find

$$m(V_r, x) = \kappa(x) \left(\frac{V_r}{V_j}\right)^{-\gamma} , \tag{13.46}$$

where V_j is an appropriate mean jet speed. The coefficient $\kappa(x)$ incorporates all x-dependent factors. Integrating equation (13.46) with respect to x over the length L of the lobe, we recover equation (13.7), provided we identify the coefficient k in the latter as

$$k \equiv \int_0^L \kappa(x)\, dx \; . \tag{13.47}$$

Let us close this section by discussing, more generally, the role of stellar winds in clearing away molecular gas to reveal obscured young stars. Take first the dense cores that directly produce the stars themselves. Theorists have long entertained the notion that their disappearance may be attributable to the eruption of powerful winds. Empirically, however, we have ample evidence of dense cores and winds coexisting. All Class 0 and I sources reside within such cores, and most are associated with CO outflows. It is true that molecular line tracers have greater widths in those cores containing stars. (Recall Figure 3.13.) In some cases, one sees that the line shifts in velocity along the same direction as the outflow. Nevertheless, the actual disruption, whatever its cause, must be more of a gradual erosion than a violent dispersal.

Each molecular outflow extends far beyond the boundaries of the original dense core, so there is good reason to look for damage on a larger scale. The jet itself may extend even farther, beyond the individual dark cloud that spawned an entire cluster of cores and their stars. We should thus picture a molecular outflow as tracing only a portion of the stellar wind. The momentum in each outflow lobe certainly stems from the jet, perhaps through the intermediary of ejecta from the working surfaces. This same mechanism should, of course, drain some of the jet's momentum even in the absence of a surrounding cloud. The curious deceleration of giant Herbig-Haro flows may be a manifestation of this process. In any case, there are no plausible alternatives to jet-driven outflows as the means for removing the molecular gas containing stellar groups. Progress in understanding will come once we have documented more explicitly clouds in progressive stages of their demise.

Chapter Summary

The winds from embedded, low-mass stars create narrow jets with a bipolar morphology. Within a few hundred AU from the star, these jets are seen through radio continuum flux arising from strong shocks. Knotty strands of emission in Hα and other optical lines, also shock-generated, trace the flow on a scale of 0.1 pc. Each optical knot, or Herbig-Haro object, represents wind material overtaking slower, previously ejected gas. Wind velocity fluctuations of relatively low amplitude occur over decades, while the largest changes take millennia. The last sort manifest themselves as chains of discrete Herbig-Haro objects extending for parsecs from the driving star.

Far more common than optical or radio jets are another manifestation of winds, molecular outflows. These are broader regions of cloud gas stirred up and dragged forward by a central jet. They are mainly identified through millimeter emission lines of CO and again have a clumpy, bipolar appearance. Deeply embedded Class 0 sources drive outflows that are narrower and faster than those from Class I stars. The velocity within an outflow lobe is highest near the jet and falls off laterally, in a way that is characteristic of entrained matter. Shocks driven into the gas heat it and create new molecules, such as SiO. On a larger scale, the multiple outflows in a T association help disperse the parent cloud.

Stars of all masses and ages generate winds. The relatively weak flow from the Sun is driven by thermal pressure in its extended corona. This mechanism is unsufficient to account for pre-main-sequence winds, as is mechanical pressure from Alfvén waves. However, the magnetic field is still likely to play a key role. Rotation of a magnetized star bends the field, creating a torque on gas that flings it outward. Such a centrifugal wind could also arise from the inner region of a circumstellar disk.

Whatever its specific point of origin, a magnetized, pre-main-sequence wind is at least weakly bipolar when launched. Such a flow can drive lateral crossing shocks in the surrounding medium. A sequence of crossing shocks may collimate the jet, as long as it remains within the parent cloud. Additional shocks, corresponding observationally to Herbig-Haro objects, arise from wind fluctuations. These weaker fronts, transverse to the flow, eject material sideways from the jet, facilitating the entrainment of cloud gas into molecular outflows.

Suggested Reading

Section 13.1 The original discovery of Herbig-Haro objects was by

Herbig, G. H. 1951, ApJ, 113, 697

Haro, G. 1952, ApJ, 115, 572.

For a modern discussion of these regions and their connecting optical jets, see

Reipurth, B. & Bally, J. 2001, ARAA, 39, 403.

Giant Herbig-Haro flows are reviewed briefly by

Bally, J. & Devine, D. 1997, in *Herbig-Haro Flows and the Birth of Low-Mass Stars*, ed. B. Reipurth and C. Bertout (Dordrecht: Kluwer), p. 29, while radio observations of jets are summarized in

> Anglada, G., Villuendas, E., Estalella, R., Beltrán, M. T., Rodríguez, L. F., Torelles, J. M., & Curiel, S. 1998, AJ, 116, 2953.

Section 13.2 The first outflow mapped was that in L1551:

> Snell, R. L., Loren, R. B., & Plambeck, R. L. 1980, ApJ, 239, L17.

Two useful and broad-ranging reviews of the phenomenon are

> Bachiller, R. 1996, ARAA, 34, 111

> Richer, J. S., Shepherd, D. S., Cabrit, S., Bachiller, R., & Churchwell, E. 2000, in *Protostars and Planets IV*, ed. V. Mannings, A. P. Boss, and S. S. Russell (Tucson: U. of Arizona Press), p. 867.

For a study of the massive system in Orion A, see

> Chernin, L. M. & Wright, M. C. H. 1996, ApJ, 467, 676.

Section 13.3 There is a vast literature on stellar winds. A comprehensive text that includes both observation and theory is

> Lamers, H. J. G. L. M. & Cassinelli, J. P. 1999, *Introduction to Stellar Winds*, (Cambridge: Cambridge U. Press).

The original analysis of the solar wind as a thermally driven flow is due to

> Parker, E. N. 1958, ApJ, 128, 664.

For a lucid, physically oriented discussion of pressure effects in winds, see

> Holzer, T. E. 1987, in *Proceedings of the Sixth International Solar Wind Conference*, eds. V. J. Pizzo, T. E. Holzer, and D. G. Sime, (Boulder: NCAR), p. 3.

An exploration of Alfvén waves as a possible wind-driving mechanism in young stars is

> Hartmann, L., Edwards, S., & Avrett, E. 1982, ApJ, 261, 279.

Section 13.4 Our treatment of centrifugal winds follows the work of

> Mestel, L. 1968, MNRAS, 138, 359.

Another influential account of wind braking, using a split-monopole model for the poloidal field, is

> Weber, E. J. & Davis, L. 1967, ApJ, 148, 217.

The theory of disk-generated winds has been addressed in many papers, but key issues in this subject remain unresolved. One important analysis is

> Blandford, R. D. & Payne, D. G. 1982, MNRAS, 199, 883.

Two later reviews are

Lovelace, R. V., Ustyugova, G. V., & Koldova, A. V. 1999, in *Active Galactic Nuclei and Related Phenomena*, eds. Y. Terzian, E. Kachikian, & D. Weedman (San Francisco: ASP), p. 208.

Shu, F. H., Najita, J. R., Shang, H., & Li, Z.-Y. 2000, in *Protostars and Planets IV*, eds. V. Mannings, A. P. Boss, & S. S. Russell (Tuscon: U. of Arizona), p. 789.

The first reference is a unified account, covering both jets from young stars and galactic nuclei. The second describes the X-wind model, which has been widely used to interpret T Tauri infall and outflow signatures.

Section 13.5 The question of the ionization level and mass transport rate in jets is analyzed by

Hartigan, P., Morse, J., & Raymond, J. 1994, ApJ, 436, 125.

Bacciotti, F., Eislöffel, J. 1999, AA, 342, 717.

A representative numerical study of variable jets is

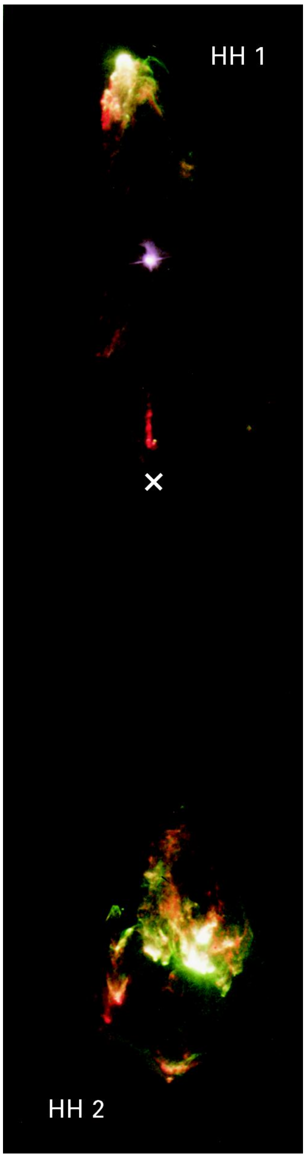
Stone, J. M. & Norman, M. L. 1993, ApJ, 413, 198.

For a classic treatment of crossing shocks and other issues in jet physics, see

Pai, S.-I. 1954, *Fluid Dynamics of Jets* (New York: Van Nostrand).

Our treatment of the origin of the acceleration in molecular outflows follows

Stahler, S. W. 1994, ApJ, 422, 616.



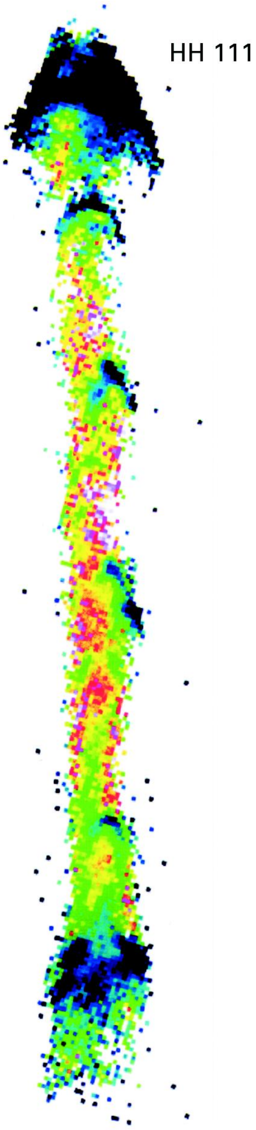


Plate 9 Optical photograph of the HH 1/2 jet. The image is a composite of Hα (shown in green), [S II] (red), and a broadband, continuum filter (blue). Note the driving star VLA 1, shown by the cross.

Plate 10 Color representation of the Hα/[S II] intensity ratio in the HH 111 jet. This ratio varies from 0.1 (shown as red), to 0.3 (green), to a maximum of 0.8 (dark blue).

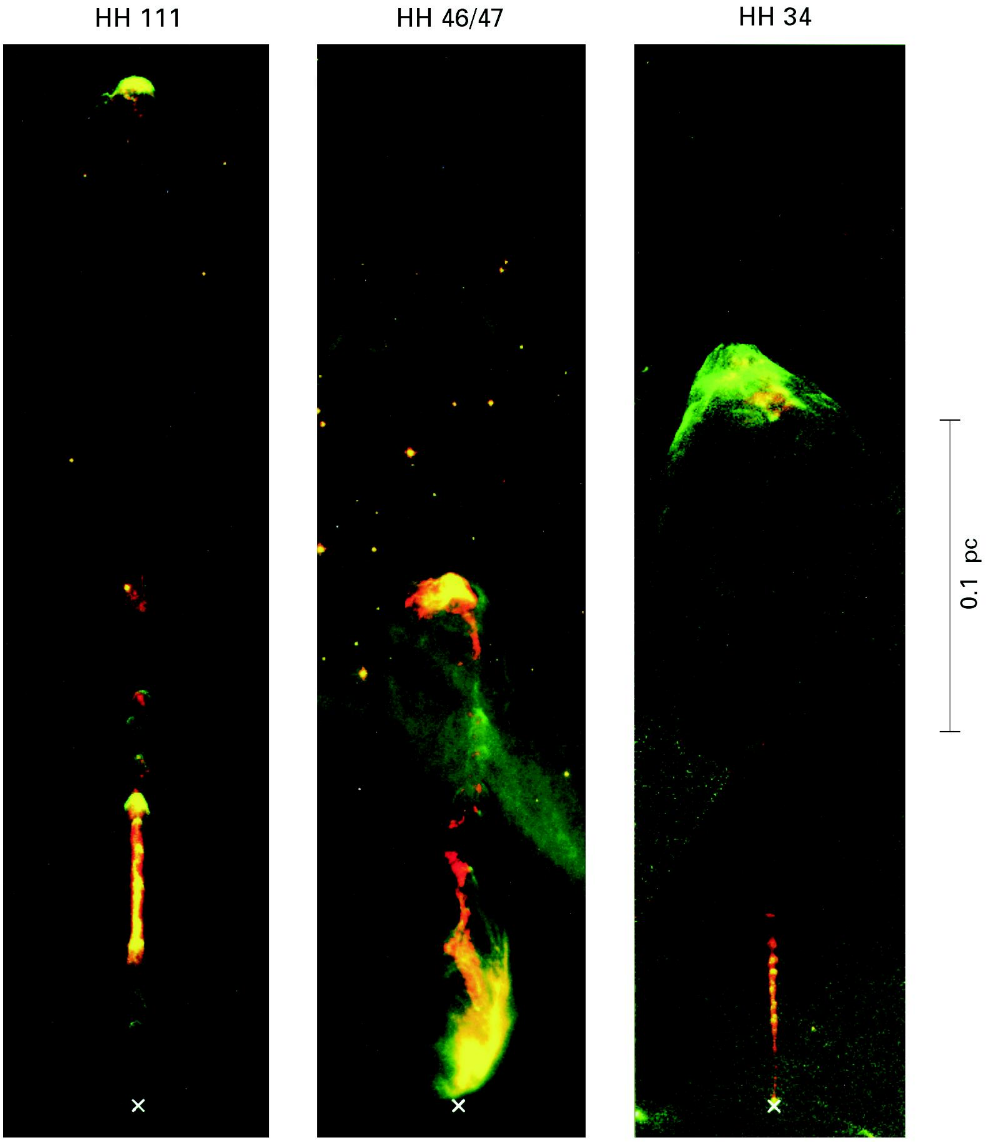


Plate 11 Comparison of the HH 111, HH 46/47, and HH 34 jets. The jets are shown at the correct relative scale, and their driving stars are indicated by crosses near the bottom of each frame. The images are composites of [S II] (red) and Hα (green).

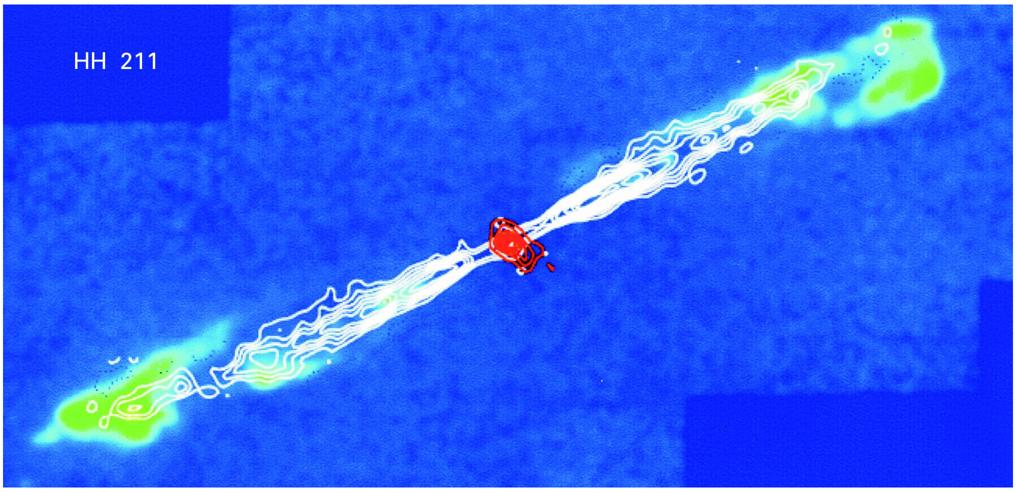


Plate 12 The jet-like HH 211 molecular outlow. The white contours represent $J = 2 \rightarrow 1$ emission of $^{12}C^{16}O$, while the central, red contours trace continuum emission at 1.3 mm. Green patches near the ends show emission from the 2.12 µm line of shocked H_2. These patches are separated by 2100 AU.

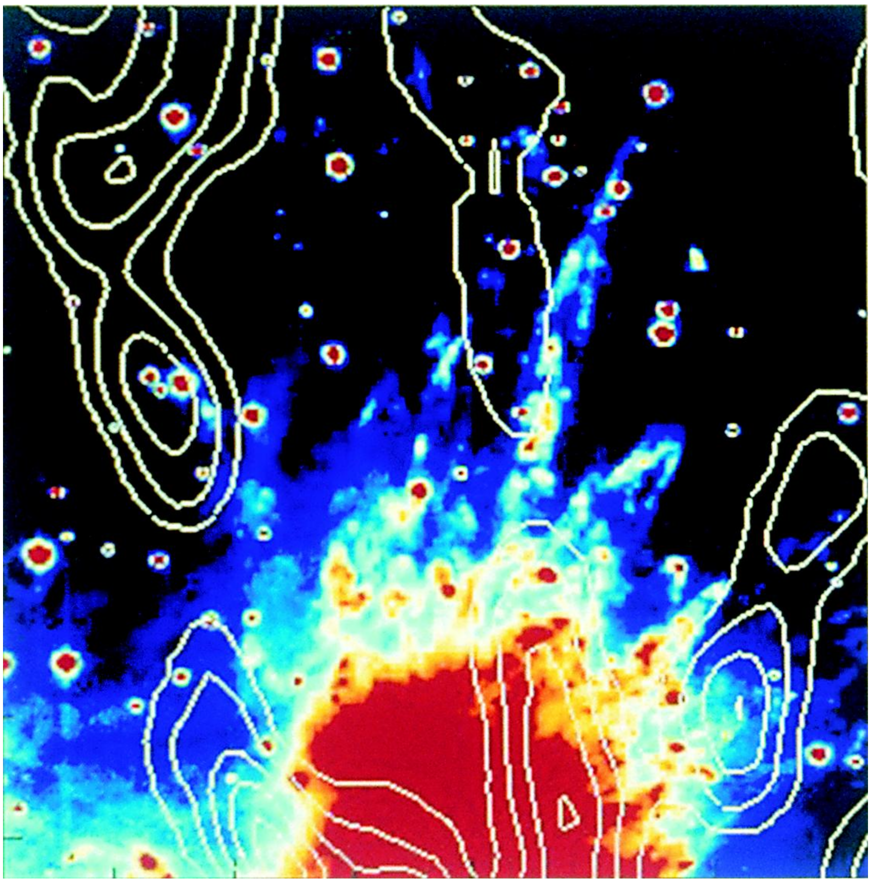


Plate 13 The molecular outflow in OMC-1. The color image shows combined emission in the 2.12 μm line of H_2 and the 1.64 μm line of [Fe II]. White contours represent 1.3 cm emission from the (1,1) and (2,2) inversion transitions of NH_3. The area covered is 0.4×0.4 pc.

The Formation of Stars
Steven W. Stahler and Francesco Palla
© 2004 WILEY-VCH Verlag GmbH & Co.

14 Interstellar Masers

One of the basic effects described by quantum theory is that a photon incident on an atom or molecule may cause it to produce another of identical energy and momentum. Under the proper conditions, this stimulated emission amplifies an initially weak source of radiation. Although firmly rooted in theory, the effect found no practical application until 1955, when J. Gordon, H. Zeiger, and C. Townes developed a method for focusing a beam of excited NH_3 molecules into a resonant cavity. Thus was born the laboratory maser, to be followed later by its optical analogue, the laser. It is a remarkable historical development that, only a decade after its laboratory implementation, the maser phenomenon was found to be operating in interstellar space. Here there are no cavities or mirrors to assist the amplification process. Instead, the radiation intensity is typically boosted by a factor of 10^{10} in a single pass through a dense region of molecular gas. Such regions, with their clusters of luminous spots, occur in the perturbed vicinity of young stars. They are also found in many other environments, including comets, the atmospheres of aging giant stars, supernova remnants, and the nuclei of external galaxies.

This chapter provides a brief overview of interstellar masers, *i. e.*, those in star-forming regions. The effect has now been detected in a variety of molecular lines, so that the groundwork laid in Chapters 5 and 6 will again prove useful to the reader. We start by describing the empirical properties that all interstellar masers share. The next two sections delve into the theory underlying the phenomenon. The essential requirement for amplification is a population inversion. Such a state is relatively easy to achieve in the low densities of interstellar clouds. Although many features of maser emission are well understood, the detailed mechanism through which the inversion is maintained remains unclear in some cases. In the chapter's final section, we describe how masers within outflows are used to trace the motion of the underlying gas.

14.1 Observed Characteristics

To begin our description of maser activity, we look first at a region of massive star formation. We examine in some detail the properties of the emission, before comparing with other locales. Along the way, we will see the variety of physical transitions that give rise to the intense lines.

14.1.1 Masers in W49

Consider W49, the Galaxy's most conspicuous source of maser emission. This region of prodigious infrared and submillimeter radiation is located in the constellation Aquila, near the Galactic plane, and lies at a distance of 11 kpc. It is completely invisible optically and is associated with one of the most massive cloud complexes. The complex harbors both a supernova remnant and a highly perturbed region of gas designated W49N, whose total luminosity is nearly

The Formation of Stars. Steven W. Stahler and Francesco Palla
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-40559-3

10^7 $L_\odot$. When W49N is imaged with an interferometer, the radio continuum flux arising from free-free emission reveals a striking pattern of ultracompact HII regions. These localized sites of ionization, each created by one or more O stars, are strung out along an elliptical ring some 2 pc in diameter. (See the upper panel of Figure 14.1.) Observations in millimeter lines show redshifted absorption dips, indicating that foreground gas is flowing toward the ring.

The origin of this motion is itself an intriguing problem, but we want to focus on an even smaller scale. Figure 14.1 shows that one of the small HII regions is significantly brighter than the others; the bottom panel is an expanded view of that area. This image was made at a wavelength of 2 cm and spans a linear extent of about 0.5 pc. Near the maximum of 2 cm emission is an unresolved source of mid-infrared flux. This presumably marks the location of the massive star.

If we now view the same ionized region at 18 cm, we find a number of strong peaks. These spots, represented by filled triangles in the figure, represent maser emission from the OH molecule. Aggregates of OH masers are found throughout the elliptical ring, where they concentrate in the outskirts of each ultracompact HII region. The region shown here, however, also has maser emission of another kind. These even smaller spots become apparent at a wavelength of 1.35 cm and cluster tightly near the O star. Each source in this case is maser radiation in a spectral line of H_2O.

Let us examine more closely these 1.35 cm (22.2 GHz) H_2O masers. The group shown here spans about $1''$, or 3×10^4 AU at the W49 distance. A single spot, on the other hand, is only about 1 AU in size. The emission is also highly localized in the frequency domain. Figure 14.2 is a spectrum in the maser line of the entire W49N region, obtained with a single-dish telescope that spatially averages over the emission distribution. There is a strong intensity maximum at a radial velocity V_r of $+6.5$ km s^{-1}, corresponding to the bulk motion of the molecular cloud. There are also very sharp peaks on either side, extending to several hundred km s^{-1} in V_r. With higher spectral and spatial resolution, one can observe individual components within a single peak, each of width $\Delta V_r \gtrsim 0.5 \text{ km s}^{-1}$. Note that there is a significant distribution of widths beyond this minimum figure. As a general rule, spectral features closer to the cloud's rest velocity tend to be narrower than the rarer spikes found with large V_r-values.

Returning to the spatial maps of W49, comparison of two or more images taken some time apart gives us the proper motion of individual maser spots. The exceptional resolution of interferometers allows us to track displacements as small as $0.''001$ per year. These data may then be combined with the Doppler shifts in Figure 14.2 to yield full, three-dimensional velocities. Figure 14.3 displays velocity vectors for a large number of the H_2O masers. Here, the angular offset is from a point that coincides with the peak of the radio continuum emission. It is evident that the crowd of spots is expanding away from that position. Moreover, the underlying flow is clearly *not* spherically symmetric, but confined to a limited solid angle. There is also a predominance of redshifted velocities to the left of the star, and blueshifted velocities to the right. Thus, the expansion is bipolar.

14.1.2 Intensity of the Radiation

The determination of both proper motions and radial velocities is facilitated by the masers' extraordinary intensity, which is their salient characteristic. To quantify matters, let us consider

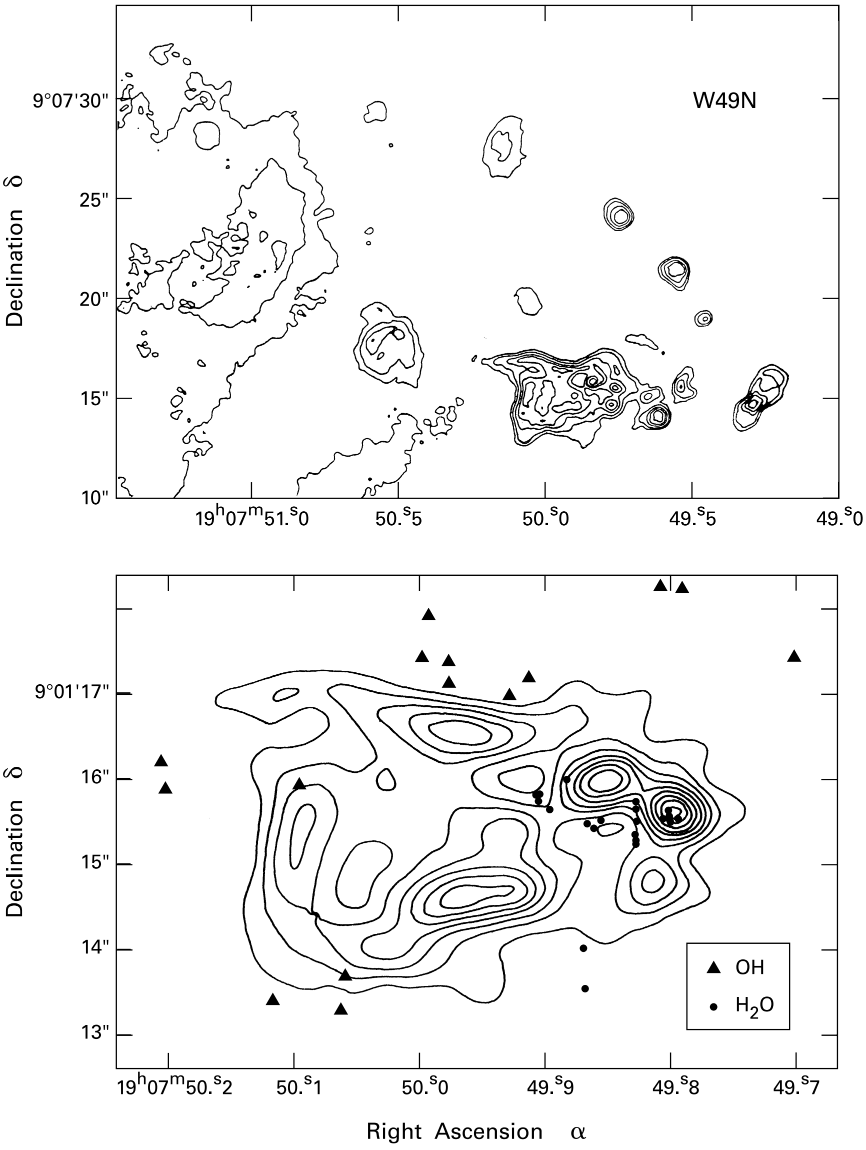


Figure 14.1 *Upper panel*: The ring of ultracompact HII regions in W49N, as seen through 6 cm continuum radiation. *Lower panel*: The largest of these regions, with H_2O masers (*filled circles*) and OH masers (*filled triangles*). Not shown is a clump of dimmer OH spots, close to the radio continuum peak.

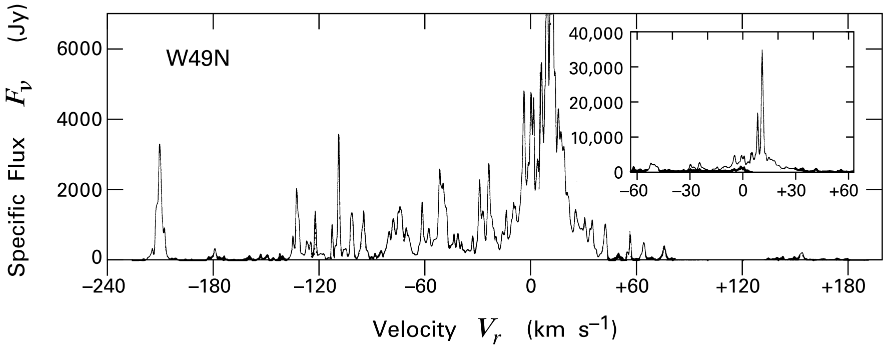


Figure 14.2 The 22 GHz spectrum of W49N.

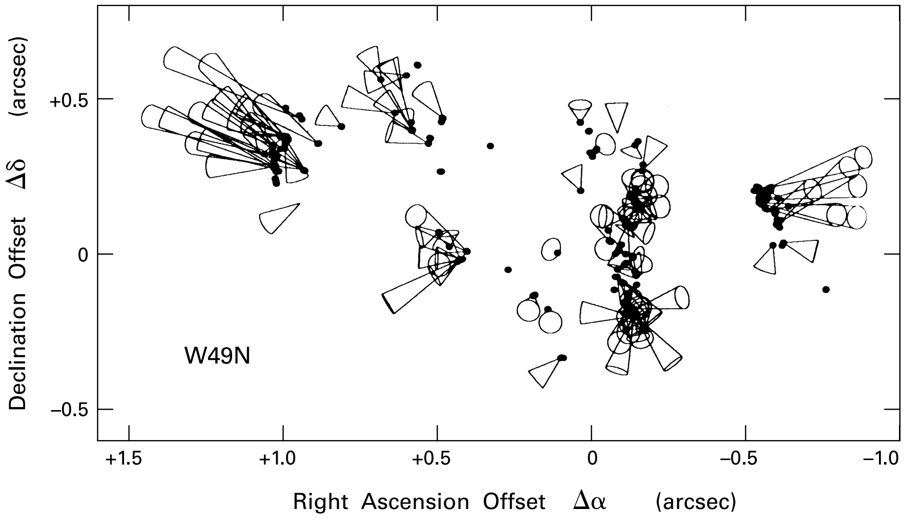


Figure 14.3 Three-dimensional motion of the H_2O masers in W49N. The cone lengths show how far each spot would travel in 150 yr. Angular offsets are measured from the dynamical center of the flow; this coincides with the peak of the 6 cm emission in the lower panel of Figure 14.1.

the brightness temperature of a typical H_2O maser in W49. We recall from equation (C.10) of Appendix C that this quantity is related to the specific intensity at the line center frequency $\nu_\circ$ by

$$T_B = \frac{c^2\, I_{\nu_\circ}}{2\,\nu_\circ^2\, k_B} \,, \tag{14.1}$$

where we have neglected any background emission. According to equation (2.17), $I_{\nu_\circ}$ is equal to $F_{\nu_\circ}/\Delta\Omega$, where $F_{\nu_\circ}$ is the specific flux and $\Delta\Omega$ the small solid angle subtended by the spot. The latter, in turn, is given by $\Delta\Omega = (d/D)^2$, where d is the spot diameter and D the distance to W49. Combining these terms, we have

$$\begin{aligned} T_B &= \frac{c^2\, F_{\nu_\circ}\, D^2}{2\, \nu_\circ^2\, k_B\, d^2} \\ &= 3 \times 10^{15}\,\mathrm{K} \left(\frac{F_{\nu_\circ}}{10^4\ \mathrm{Jy}}\right) \left(\frac{D}{11\ \mathrm{kpc}}\right)^2 \left(\frac{d}{1\ \mathrm{AU}}\right)^{-2} \ . \end{aligned} \tag{14.2}$$

Note that our representative specific flux of 10^4 Jy is consistent with the spectrum displayed in Figure 14.2.

How are we to interpret such extreme brightness temperatures? Recall that the definition of T_B comes from the Rayleigh-Jeans approximation to the blackbody formula, equation (2.28). That equation, in turn, describes the radiative emission from matter in thermal equilibrium. Whatever the physical processes occurring in W49, they assuredly do not heat regions to 10^{15} K, far above the point where all molecules would dissociate. The fact that T_B vastly exceeds any reasonable kinetic temperature signifies that the gas itself is out of thermal equilibrium. In other words, the relevant level populations of the H_2O molecules do not follow a Boltzmann distribution.

The brightness temperature from a maser is elevated both by the very small emitting size d and by the high value of $F_{\nu_\circ}$. The specific flux, in turn, is so large because energy is being emitted in an extraordinarily narrow interval in frequency $\Delta\nu$ or, equivalently, in radial velocity, ΔV_r. Both features are remarkable, and their origin is not readily apparent. The agent that excites H_2O masers in W49 does so over a region spanning 10^4 AU. How, then, can abrupt changes in physical conditions occur in just 1 AU? The answer is that interstellar shocks provide just such a sharp transition. We also recall from Chapter 8 that a number of molecules, including H_2O, are created in the cooling regions behind shocks.

The narrowness of the radial velocity interval ΔV_r is also puzzling. To form H_2O, the temperature behind a shock must fall to a plateau value of some 500 K. The radial velocity dispersion associated with this temperature is 0.5 km s^{-1}. To this figure, we should add (in quadrature) the dispersion arising from differential motion of the cloud along the propagation path of the radiation. Observations of relatively broad CO lines confirm that such motion occurs. Yet we have seen that the typical velocity width of the H_2O maser emission in W49 is only 0.5 km s^{-1}, just the thermal value. How is the additional line broadening suppressed?

These issues remain when we consider the 18 cm OH masers in W49. Here we find generally similar properties as the 1.35 cm sources, but also some systematic differences. The diameter of each maser spot, about 10^{14} cm, is larger by an order of magnitude. Velocity widths of the spectral features are again low, between 0.6 and 1.5 km s^{-1}. As in the case of H_2O, the strongest peaks in the spectrum occur near the bulk velocity of the cloud. This ubiquitous trend will prove important when we later consider the geometry of the propagating radiation. The larger emitting areas of OH masers result in lower T_B-values, although the actual figures, near 10^{13} K, still greatly exceed any plausible gas temperature. Finally, observation of proper motions in W3(OH), an HII region much closer than W49, again shows that the masers surrounding a massive star expand as a group.

There is one characteristic of maser radiation that is especially pronounced in the OH sources. This is the observed *polarization* of the beam. The 22 GHz H_2O masers are linearly polarized to some degree, with the fractional polarization ranging from about 1 to 10 percent within a given HII region. The 18 cm lines from OH, on the other hand, exhibit nearly total *circular* polarization. We recall from Chapters 5 and 6 that the polarization of emitted radiation from OH arises when an external magnetic field interacts with an unpaired electron. In fact, observations of Zeeman splitting for both types of masers allow us to probe this field within HII regions. One finds generally that the H_2O masers are associated with field strengths (typically 50 mG) higher by an order of magnitude than those in OH masers, a sign that the latter occur in lower-density portions of the cloud.

14.1.3 Variability and Spot Clustering

Another important, general feature of masers is that their intensity is strongly time-dependent. Studies of variability have concentrated mainly on the 22 GHz masers, owing to their higher brightness. A given spot can rise in intensity, reach its maximum, and then disappear entirely in a matter of weeks. On the other hand, there are some spots in W49 that have persisted for a number of years. It is generally true that the highest-velocity features tend also to have the shortest lifetimes. Even the most transient spots, however, are clustered in "centers of activity" that have a substantially longer duration. The H_2O masers of W49 contain at least three such centers, which are apparent in the clumping of velocity vectors in Figure 14.3. Note finally that the flaring of an individual spot does not occur in tandem with any increase of the stellar luminosity. Whatever creates these changes therefore does so locally.

In fact, the temporal variation in maser emission is broadly consistent with its origin in a turbulent environment. To make this idea more quantitative, consider that the typical emitting size d divided by the velocity width ΔV_r also yields a characteristic time. Using figures appropriate for one of the brighter H_2O masers in W49, $d = 1\,\mathrm{AU}$ and $\Delta V_r = 0.5\ \mathrm{km\,s^{-1}}$, this period is 10 yr, comparable to the longest observed spot lifetimes. The characteristic time also diminishes for the higher-velocity features with their greater line widths. This simple exercise hints that it is velocity gradients within the cloud that are responsible for disrupting maser activity.

Numerous other sites of massive star formation have H_2O and OH masers clustered near a common, luminous source. In both cases, the spots are found within a region about 10^{17} cm in diameter, with distinctive subclustering. There is usually some degree of overlap between the OH and H_2O groups. On the other hand, each maser type also has its own, characteristic distribution pattern. The 18 cm spots from OH are often located near the edges of HII regions, as delineated by the falloff in radio continuum flux. The HII region itself may have a swept-back, "cometary" appearance, indicating that the star is ionizing a medium with a strong density gradient. Here the OH spots tend to occur in the more compressed portion of the ionized gas.

Figure 14.4 shows, as an example, a combined radio continuum and spectral line image of NGC 6334F, situated at one end of a molecular cloud with high star formation activity. The OH masers, indicated again by filled triangles, all lie between the star, *i. e.*, the peak in radio continuum flux, and the main body of the parent cloud. We have already noted that H_2O masers require even higher densities. This may mean, as in Figure 14.1, that they are nested inside the OH masers. However, the very different situation in Figure 14.4 is also common. Here we indeed see a few H_2O sources near the OH masers, but many more clumped in a linear fashion well outside the HII region.

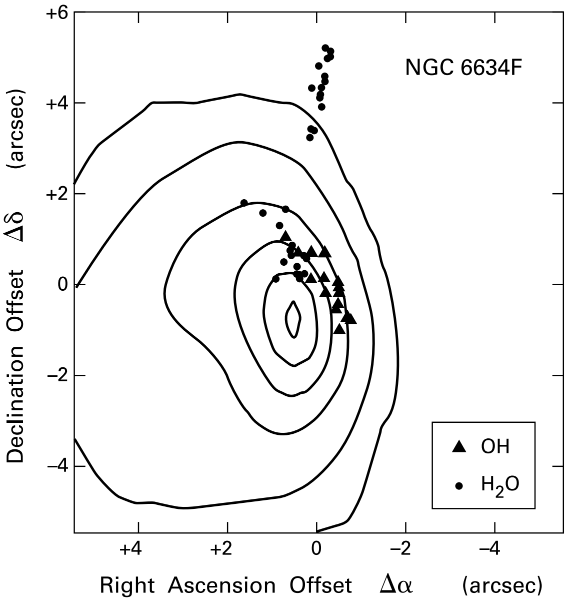


Figure 14.4 The 15 GHz continuum emission (*contours*) of NGC 6634F, together with H_2O masers (*filled circles*) and OH masers (*filled triangles*). Angular offsets refer to a point near the peak continuum flux.

Further observations of NGC 6334F reveal emission in heated NH_3 near the location of the external H_2O spots. The elevated temperature implies that there is a second massive star responsible for this maser activity, although this deeply embedded object has not yet created its own HII region. Such findings underscore the fact that the essential ingredient for H_2O maser activity is a high-density shock. The energized gas *may* be situated within a larger HII region, or it may arise through the impact of a stellar wind on cloud matter in a neutral environment.

An OH maser, on the other hand, appears to need conditions specific to HII regions or their immediate surroundings. Examination of low-mass, embedded stars strengthens this view. It is not uncommon to find H_2O masers associated with such objects, but OH masers are entirely absent. This does not mean that every OH maser appears within a detectable HII region. Maps at centimeter wavelengths occasionally show an apparently isolated group of OH spots. In such cases, the radio continuum emission from the ionized region may simply be too weak to be observed. The gas near the masers and their exciting source is only revealed through infrared and millimeter studies.

The H_2O masers associated with lower-mass stars are situated close to, but not coincident with, the stellar source. Note that the most accurate position of the star, which is always an embedded Class 0 or I object, is given by the peak of emission in a radio jet, if one is observed (§ 13.1). The masers themselves are typically displaced from 30 to 100 AU from the star, corresponding to a few arcseconds for an object 200 pc away. Emission in the 22 GHz line has mainly been found for sources that also drive CO outflows, so that the two phenomena – winds

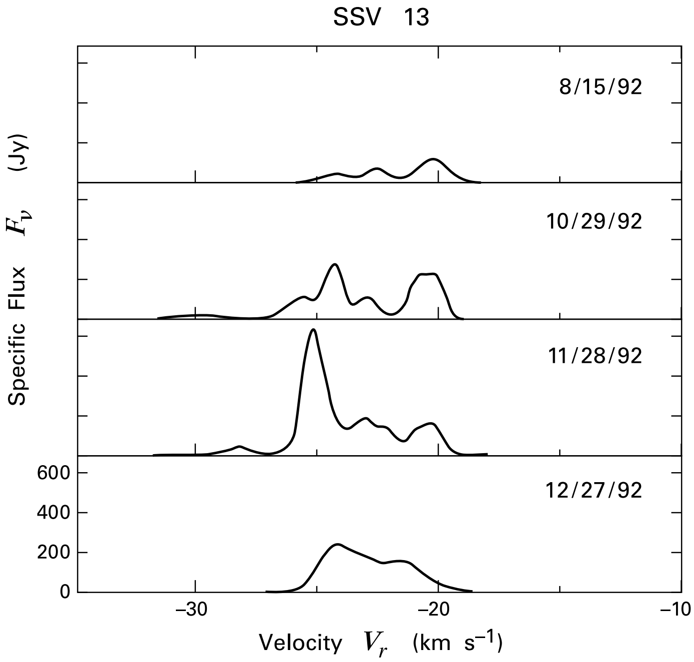


Figure 14.5 Four low-resolution spectra of H_2O masers in SSV 13. Each spectrum was recorded on the indicated date.

and masers – appear to be linked. As we shall discuss in § 14.3, the arrangement of the spots often implicates the central, stellar jet as the underlying agent of excitation.

Most of the properties we have discussed for the H_2O masers near high-mass stars also seem to hold, at least qualitatively, for sources of more modest luminosity, say $L_* \lesssim 100\,L_\odot$. The total intensities of the spot clusters are lower and approximately scale with L_*. This proportionality relation also applies to luminous stars, and with the same coefficient. Not all low- and intermediate-mass embedded stars have associated maser activity. However, the fraction that does exhibit 22 GHz emission, with a flux of at least 1.5 Jy, rises steeply with the stellar luminosity and is close to unity for $L_* \gtrsim 25\,L_\odot$.

Observations taken over year-long timescales have established that the H_2O masers from lower-mass stars are again highly variable. In some, usually weaker, sources, there are long periods of quiescence interspersed with outbursts of maser emission. Other, stronger, sources display a more continuous level of activity. Figure 14.5 shows one example of the latter type. Here we see four successive, low-resolution spectra of the source SSV 13. This infrared star in the NGC 1333 region of Perseus has a bolometric luminosity of about 25 $L_\odot$ and powers, in addition to the masers, a string of Herbig-Haro objects known as HH 7-11. It is apparent that the detailed spectral distribution of the 22 GHz emission changes qualitatively within periods of a month or longer. At all times, the peak intensity is significantly displaced from the ambient cloud velocity of $+7.0$ km s^{-1}.

14.1.4 Molecular Transitions

Thus far, we have said little about the actual transitions giving rise to maser lines. The 1.35 cm line of H_2O stems from the $6_{16} \rightarrow 5_{23}$ rotational transition, discussed in § 5.5. There is no reason, in principle, why other lines of the same molecule should not exhibit maser emission, provided the levels in question attain a population inversion. The difficulty is really an observational one. Following the 1969 discovery of the 1.35 cm line, nothing more was seen until the 1980s, when radio probes of the Orion Nebula found maser emission in the $4_{14} \rightarrow 3_{21}$ (789 µm) and $3_{13} \rightarrow 2_{20}$ (1.64 mm) lines; note that the latter transition occurs between the less populated para-rotational states (recall Figure 5.16). The ensuing decade saw the addition of yet more lines, all within the submillimeter range. These included, among others, $5_{15} \rightarrow 4_{22}$ (923 µm) and $10_{29} \rightarrow 9_{36}$ (934 µm). The last transition is rather unusual in that the upper state does not lie along the rotational backbone.

The high rotational levels involved in some of the submillimeter H_2O masers call for both elevated densities *and* temperatures at the site of formation. In our discussion of strong J-shocks in Chapter 8, we noted that all molecules are initially dissociated as gas passes through the front. However, H_2O and other species reform downstream, once the gas cools below about 500 K. Such conditions adequately account for the 1.35 cm maser line, but not for the ensemble of submillimeter transitions. For these, we must invoke the C-shocks that arise when a wind impacts magnetized cloud material. Molecules now pass intact through the extended shock region, where they are temporarily heated to temperatures that may exceed 2000 K.

Turning to OH masers, the 18 cm "line" is actually a group of four at 1612, 1665, 1667, and 1720 MHz. All are produced within the rotational ground state of the molecule, through the processes of Λ-doubling and magnetic hyperfine splitting. The underlying physics, as detailed in § 5.6, is that the single, unpaired electron is affected both by the molecule's overall rotation and by the magnetic field from the spinning hydrogen nucleus. Maser emission is most commonly observed in the main lines at 1665 and 1667 MHz, but the two satellite lines are seen, as well. Finally, observers have found OH masers in other radio lines, including a group near 6 cm. These arise from analogous transitions within rotationally excited states.

Many other molecules besides H_2O and OH display maser radiation. The best studied of these is methanol (CH_3OH), also first seen toward Orion. Observed lines arise from rotation of the OH bond about the symmetry axis of the CH_3 tetrahedron. There are two varieties of transitions (E-type and A-type), distinguished by the manner in which the OH tunnels through the complex potential barrier created by the CH_3. States are labeled by the notation J_K, where J is the total angular momentum and K its projection onto the axis of the tetrahedron. The known maser transitions all occur within regions of massive star formation.

The numerous maser lines associated with CH_3OH fall empirically into two mutually exclusive categories. Some transitions, designated "Class I," are always found to be offset from compact HII regions. Like the external H_2O masers in NGC 6334F, they must be powered by the impact of a wind on dense gas. Other lines of CH_3OH, known collectively as "Class II," only appear projected against the radio continuum emission from the ionized gas. These, like OH masers, must be energized by radiation from the massive star. Note that each of these two classes includes both E- and A-type transitions.

Table 14.1 summarizes the most easily observed maser lines from OH, H_2O, CH_3OH, and several other species. For each transition, we give $E_{\rm upp}$, the energy (in temperature units) of the upper state above ground. This provides a measure of the difficulty in exciting the molecule to maser emission. We also list characteristic ranges of density and temperature, n and T, over which emission occurs. These figures are taken from numerical modeling of the level populations.[1] For CH_3OH, the first line falls in the Class I category, the second in Class II. Note further that the maser lines of SiO stem from adjacent rotational levels. It is also not difficult to excite rotational transitions within higher vibrational states, so that a rich spectrum of lines results. Observations of SiO masers in star-forming regions are much less common than in the envelopes of late-type, evolved stars. Thus far, the only objects, whether young or old, that appear capable of exciting the emission from this molecule are those of high luminosity. The same holds for NH_3 and H_2CO. In the first case, the transition is the usual tunneling of the nitrogen atom through the plane of hydrogens. Maser emission of H_2CO stems from rotation of this asymmetric top. The level notation here is analogous to that for H_2O.

Table 14.1 Prominent Maser Lines in Star-Forming Regions

Molecule	Transition	Wavelength (cm)	$E_{\rm upp}/k_B$ (K)	n (cm^{-3})	T (K)
OH	$^2\Pi_{3/2}$ (J$= 3/2$; $\Delta F = 0, \pm1$)	18	0.08	$10^5 - 10^7$	$100 - 200$
H_2O	$6_{16} \rightarrow 5_{23}$	1.35	640	$10^7 - 10^9$	$300 - 1000$
CH_3OH	$4_{-1} \rightarrow 3_0 \;\; E$	0.83	29	$10^4 - 10^5$	$20 - 100$
	$5_1 \rightarrow 6_0 \;\; A$	4.49	49	$10^4 - 10^5$	$20 - 100$
SiO	$v = 1$; $J = 2 \rightarrow 1$	0.35	1774	$10^9 - 10^{10}$	$700 - 1000$
NH_3	$(J, K) = (3, 3)$	1.25	122	$10^4 - 10^5$	$60 - 150$
H_2CO	$1_{10} \rightarrow 1_{11}$	6.3	14	$10^4 - 10^5$	$20 - 40$

14.2 Maser Theory: Basic Principles

Let us now explore the actual mechanism by which radiation becomes amplified. We will see that a critical ingredient is maintenance of a certain population of excited molecules. This insight helps to constrain the excitation process itself, *i. e.*, the physical characteristics of the regions giving rise to such emission.

---

[1] Note the large disparity between the energy separation of the OH levels and the temperature needed for maser activity. To excite this activity, the molecule must first be elevated to the next higher rotational level. According to Figure 5.18, this level ($J = 5/2$) lies some 120 K above ground.

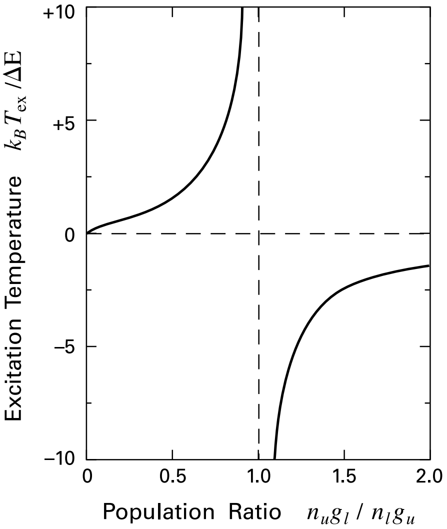


Figure 14.6 The excitation temperature in a two-level system, shown as a function of the population ratio $n_u g_l/n_l g_u$. Note how the temperature becomes negative once this ratio exceeds unity.

14.2.1 Level Populations

The high intensity of maser spots results when photons of a certain frequency stimulate a molecule to emit others of the same frequency and direction. This augmented group of photons then creates even more stimulated emission, so that the amplification is exponential. As we shall see, the entire process hinges on the fact that there are more molecules in the upper state of the transition than the lower. Recall that we have defined the excitation temperature of a two-level system through the relation

$$\frac{n_u}{n_l} = \frac{g_u}{g_l} \exp\left(-\frac{\Delta E}{k_B\, T_{\rm ex}}\right) . \tag{14.3}$$

Here, n_u and n_l are the number densities of the upper and lower levels, respectively, g_u and g_l are the degeneracies, and ΔE is the transition energy. In a *population inversion*, the necessary condition for maser amplification, we have $n_u/g_u > n_l/g_l$. The system therefore has a *negative* excitation temperature along the path of the emission buildup.

We see once again that the region in question cannot be in a state of thermal equilibrium, where the excitation temperature would equal the local, gas kinetic value. Figure 14.6 uses equation (14.3) to plot the actual value of $T_{\rm ex}$ (in units of $T_\circ \equiv \Delta E/k_B$) as a function of $n_u g_l/n_l g_u$. As the latter ratio approaches unity from below, $T_{\rm ex}$ goes to positive infinity and then plunges to large negative values for a slight inversion. If $n_u g_l/n_l g_u$ becomes very high, $T_{\rm ex}$ is still negative, but of diminished amplitude.

Maintaining a population inversion costs energy. In maser theory, the process that accomplishes this task is known as the *pump*. For microwave lines and typical kinetic temperatures of several hundred degrees, we have $T_\circ \ll T_{\rm kin}$. Thus, $n_u/g_u \approx n_l/g_l$ even in thermal equilibrium. The actual shifts in population required for maser emission are relatively minor and apparently not difficult to attain within interstellar clouds. Many pumping mechanisms have been proposed for the various types of observed masers. We shall briefly discuss two of them at

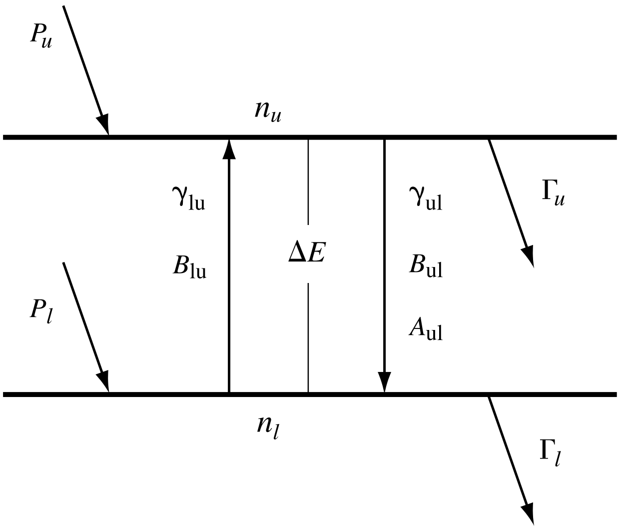


Figure 14.7 Physical processes governing the level populations in a maser transition.

the end of this section. First, however, we should quantify the degree of inversion as a function of ambient conditions.

Whatever its detailed nature, the pump takes molecules from other levels and adds them to the two which actually interact with the maser radiation.[2] We therefore define the *pump rate* P_u as the number of additional molecules brought to the upper level per unit time, and per unit volume. The analogous rate P_l applies to the lower level (see Figure 14.7). Having postulated a link between our two-level system and a reservoir of background states, we further recognize that transitions can occur both ways. That is, molecules in either of our two levels can decay back into the reservoir. The number of such transitions per unit time must be proportional to the current population of the upper or lower level. Hence, we let $n_u\Gamma_u$ and $n_l\Gamma_l$ be the respective *loss rates*, again measured per unit volume.

The level populations also change through collisions with ambient molecules. Employing the notation of Appendix B, we let $n_{\rm tot}$ be the number density of the background gas, while γ_{lu} and γ_{ul} are the collisional excitation and deexcitation coefficients. Finally, we must account for the presence of the radiation field, which we characterize through $\bar{J}$, the frequency-integrated mean intensity (see equation (B.1)). The steady-state population of each level follows by summing all the rates to zero:

$$0 = P_u - n_u\Gamma_u - (n_u B_{\rm ul} - n_l B_{\rm lu})\bar{J} - (n_u\gamma_{ul} - n_l\gamma_{lu})\,n_{\rm tot} - n_u A_{\rm ul} \qquad (14.4a)$$

$$0 = P_l - n_l\Gamma_l + (n_u B_{\rm ul} - n_l B_{\rm lu})\bar{J} + (n_u\gamma_{ul} - n_l\gamma_{lu})\,n_{\rm tot} + n_u A_{\rm ul}\ . \qquad (14.4b)$$

We also know that the two Einstein B-coefficients are related through equation (B.8b). There is a more complex relation, equation (B.4), between γ_{ul} and γ_{lu}. Since, however, $T_\circ \ll T_{\rm kin}$, the

---

[2] Appendix B shows that $T_{\rm ex}$ within an *isolated*, two-level system lies between $T_{\rm kin}$ and $T_{\rm rad}$, both of which are positive quantities. To obtain a negative $T_{\rm ex}$, the two levels must be connected physically to others.

latter reduces to $\gamma_{ul} g_u \approx \gamma_{lu} g_l$. We then have

$$0 \,=\, P_u \,-\, n_u \Gamma_u \,-\, (n_u - n_l g_u/g_l)(B_{ul}\bar{J} + \gamma_{ul}\, n_{\rm tot}) \,-\, n_u A_{\rm ul} \tag{14.5a}$$

$$0 \,=\, P_l \,-\, n_l \Gamma_l \,+\, (n_u - n_l g_u/g_l)(B_{ul}\bar{J} + \gamma_{ul}\, n_{\rm tot}) \,+\, n_u A_{\rm ul} \ . \tag{14.5b}$$

14.2.2 Degree of Inversion

Our goal is to derive an expression for the population difference. To this end, we now adopt two more simplifying assumptions. First, we neglect the spontaneous emission term in equation (14.5a), which is always relatively small in practice, *i. e.*, $A_{\rm ul}$ is negligible compared to either Γ_u or Γ_l. (We shall soon have to include the A-coefficient, however, when considering the radiative transfer.) Second, we shall assume that the two loss rates are equal, *i. e.*, that $\Gamma_u = \Gamma_l \equiv \Gamma$. Accounting for separate rates is not difficult, but only complicates the algebra without adding physical understanding.

After simplifying equations (14.5a) and (14.5b) as indicated, we now solve them for the level populations n_u and n_l. We find

$$n_u \;=\; \frac{(P_u - P_l) \,+\, (P_u + P_l)\,[1 \,+\, (2g_u/g_l)(B_{\rm ul}\bar{J} + \gamma_{\rm ul} n_{\rm tot})/\Gamma]}{2\Gamma \,+\, 2(1 + g_u/g_l)(B_{\rm ul}\bar{J} + \gamma_{\rm ul} n_{\rm tot})} \tag{14.6a}$$

$$n_l \;=\; \frac{(P_u + P_l)\,[1 \,+\, 2(B_{\rm ul}\bar{J} + \gamma_{\rm ul} n_{\rm tot})/\Gamma] \,-\, (P_u - P_l)}{2\Gamma \,+\, 2(1 + g_u/g_l)(B_{\rm ul}\bar{J} + \gamma_{\rm ul} n_{\rm tot})} \ . \tag{14.6b}$$

We may describe the population inversion through a quantity Δn:

$$\begin{aligned} \Delta n \;&\equiv\; n_u/g_u \,-\, n_l/g_l \\ &=\; \frac{P_u/g_u \,-\, P_l/g_l}{\Gamma \,+\, (1 + g_u/g_l)(B_{\rm ul}\bar{J} + \gamma_{\rm ul} n_{\rm tot})} \ . \end{aligned} \tag{14.7}$$

This equation shows that inversion requires a higher pump rate per sublevel into the upper state than the lower: $P_u/g_u \,>\, P_l/g_l$.

It is also convenient to consider Δn°, the inversion that would hold in the absence of any radiation field. This is

$$\begin{aligned} \Delta n^\circ \;&\equiv\; n_u^\circ/g_u \,-\, n_l^\circ/g_l \\ &=\; \frac{P_u/g_u \,-\, P_l/g_l}{\Gamma \,+\, (1 + g_u/g_l)\gamma_{\rm ul} n_{\rm tot}} \ . \end{aligned} \tag{14.8}$$

Here, n_u° and n_l° are the level populations obtained by setting $\bar{J} \,=\, 0$. Combining equations (14.7) and (14.8), we find

$$\Delta n \;=\; \frac{\Delta n^\circ}{1 \,+\, \bar{J}/\bar{J}_s} \ , \tag{14.9}$$

where

$$\bar{J}_s \;\equiv\; \frac{\Gamma \,+\, (1 + g_u/g_l)\gamma_{\rm ul} n_{\rm tot}}{(1 + g_u/g_l) B_{\rm ul}} \ . \tag{14.10}$$

Equation (14.9) tells us that, for $\bar{J} < \bar{J}_s$, the radiation hardly affects Δn. On the other hand, if $\bar{J}$ is ever amplified so much that it exceeds $\bar{J}_s$, then the population difference will decline significantly. But it is this very difference that increases $\bar{J}$, through stimulated emission from the upper level. Thus, maser amplification is a self-limiting process.

14.2.3 Gain and Saturation

To describe this quenching effect, known as *saturation*, as well as the amplification that precedes it, we turn to the equation of radiative transfer. In the notation of Appendix C, we have:

$$\frac{dI_\nu}{ds} = -\alpha_\nu \, I_\nu + j_\nu \; . \tag{14.11}$$

Utilizing equations (C.4) and (C.7) for the emission and absorption coefficients, respectively, the transfer equation becomes

$$\frac{dI_\nu}{ds} = \frac{h\nu_\circ}{4\pi} \, \Delta n \, g_u \, B_{\rm ul} \, \phi(\nu) \, I_\nu + \frac{h\nu_\circ}{4\pi} \, n_u \, A_{\rm ul} \, \phi(\nu) \; . \tag{14.12}$$

Here, $\nu_\circ$ is the frequency at line center. The normalized profile function $\phi(\nu)$ gives the proportion of molecules that emit or absorb a photon at each frequency. This function does *not* describe the frequency dependence of I_ν, which is generally different. We shall assume that $\phi(\nu)$ is the Doppler profile at the ambient gas temperature. From equation (E.23), we have an expression for $\phi(\nu_\circ)$:

$$\phi(\nu_\circ) = \frac{1}{\sqrt{\pi} \, \Delta\nu_D} \; , \tag{14.13}$$

which we shall use presently. Here $\Delta\nu_D$ is the Doppler width, given in terms of the kinetic temperature and molecular mass by equation (E.24). We now integrate equation (14.12) over frequency. Defining $I \equiv \int I_\nu \, \phi(\nu) \, d\nu$, we obtain

$$\frac{dI}{ds} = \frac{h\nu_\circ \, \Delta n^\circ \, g_u \, B_{\rm ul}}{4\pi \, \Delta\nu \, (1 + \bar{J}/\bar{J}_s)} \, I + \frac{h\nu_\circ \, n_u \, A_{\rm ul}}{4 \, \pi \, \Delta\nu} \; . \tag{14.14}$$

Note that we have used equation (14.9) to insert the dependence of Δn on the radiation intensity. We have also defined an effective bandwidth for the maser radiation $\Delta\nu$ by

$$\Delta\nu \equiv \frac{\int I_\nu \, d\nu}{I} \; . \tag{14.15}$$

If the line is narrower than the Doppler width, as is often the case, then it follows from equation (14.13) that $\Delta\nu \approx \sqrt{\pi} \, \Delta\nu_D$. For now, we shall neglect any spatial variation of $\Delta\nu$.

Our goal is to integrate equation (14.12) and thereby obtain $I(s)$. First, we require some prescription for $\bar{J}$, as well as for n_u, the upper-level population. On the first point, the small spot sizes of masers indicate that the radiation is tightly beamed. Let $\Delta\Omega$ be the small solid angle into which the photons propagate, taken to be constant along the beam. Then the definition of

J_ν, equation (2.19), tells us that $J_\nu = I_\nu \, \Delta\Omega/4\pi$. Evaluating $\bar{J}$ through equation (B.1), we find $\bar{J} = I \, \Delta\Omega/4\pi$, and thus write

$$\frac{\bar{J}}{\bar{J}_s} = \frac{I}{I_s} \, , \tag{14.16}$$

where $I_s \equiv 4\pi \, \bar{J}_s/\Delta\Omega$.

The second righthand term in equation (14.14) represents the effect of spontaneous emission. This contribution is quickly overwhelmed by the shower of photons arising from stimulated emission. However, there is no other initial source of radiation in the absence of a background source. Hence, our strategy is to keep the relevant term in equation (14.14), but replace the n_u multiplying $A_{\rm ul}$ by n_u°, the population at $I = 0$.

After incorporating this change, we may manipulate the transfer equation to read

$$\frac{dI}{ds} = \frac{I}{L\,(1 + I/I_s)} + \beta \, \frac{I_s}{L} \, . \tag{14.17}$$

Here L, the *unsaturated growth length*, is

$$L \equiv \frac{4\pi \, \Delta\nu}{h\nu_\circ \, \Delta n^\circ \, g_u \, B_{\rm ul}} \, . \tag{14.18}$$

The dimensionless quantity β is

$$\beta \equiv \frac{n_u^\circ}{g_u \, \Delta n^\circ} \left(\frac{\Delta\Omega}{4\pi}\right) \frac{A_{\rm ul}}{B_{\rm ul} \, \bar{J}_s} \, . \tag{14.19}$$

Note that the quantity $g_u \, \Delta n^\circ/n_u^\circ$ is a measure of the efficiency of the inversion process, before the level populations are changed by maser radiation. For viable pump models, this efficiency is not much less than unity, so that the inverse, which appears in equation (14.19), is not a very large number. The final term in this expression may be rewritten with the aid of equation (14.10) as

$$\frac{A_{\rm ul}}{B_{\rm ul} \, \bar{J}_s} = \frac{(1 + g_u/g_l) \, A_{\rm ul}}{\Gamma + (1 + g_u/g_l)\gamma_{\rm ul} n_{\rm tot}} \, , \tag{14.20}$$

and is always small. Since $\Delta\Omega$ is also small, we expect $\beta \ll 1$.

Equation (14.17) is readily integrated to yield an implicit expression for the intensity $I(s)$. In the limit of small β, we find

$$\frac{I(s)}{I_s} - \frac{I(0)}{I_s} + \ln\left[\frac{I(s)/I_s + \beta}{I(0)/I_s + \beta}\right] = \frac{s}{L} \, , \tag{14.21}$$

where $I(0)$ represents the background source. Figure 14.8 plots this result. At the beginning of the amplification process, we are in the *unsaturated* regime. Here, $I \ll I_s$, and the logarithmic term in equation (14.21) dominates the lefthand side. We thus have the approximate solution

$$I(s) = I(0) \exp{(s/L)} + \beta \, I_s \left[\exp{(s/L)} - 1\right] . \qquad \textit{unsaturated} \tag{14.22}$$

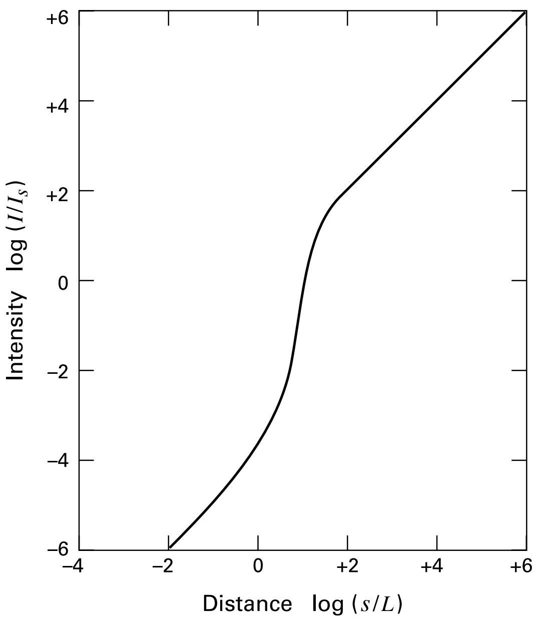


Figure 14.8 The amplification of intensity in a maser. Here, we have set $\beta = 10^{-4}$ and $I(0) = 0$ in equation (14.21). Note that $I(s)$ first rises linearly, before the exponential growth takes hold. The rise becomes linear again after saturation.

The growth becomes exponential in this regime, whether the initial flux is supplied by the background or just by spontaneous decay. In the former case, the quantity $\ln\,[I(s)/I(0)]$, commonly referred to as the maser *gain*, increases linearly with path length.

The rapid growth of intensity continues until the stimulated emission begins to drain the upper level. In this *saturated* regime, we have $I \gg I_s$ and the logarithmic term in (14.21) is small. Since the transition between the two regimes occurs at $I \approx I_s$, we have approximately

$$I(s) = I_s + I_s\left(\frac{s - s_1}{L}\right) . \qquad \textit{saturated} \tag{14.23}$$

Here the transition length s_1 may be found by setting $I(s) = I_s$ in equation (14.22).

Once saturation has begun, the intensity itself climbs only linearly with path length. Equivalently, the gain varies only logarithmically. Equation (14.5a), along with the definition of Δn in (14.7), tells us that the volumetric rate at which the upper level produces maser photons is $\mathcal{R} = g_u\,\Delta n\,B_{\rm ul}\,\bar{J}$. Using equation (14.9) for Δn, we see that $\mathcal{R}$ approaches a finite limit:

$$\begin{aligned}\mathcal{R} &\approx g_u\,\Delta n^\circ\,B_{\rm ul}\,\bar{J}_s \\ &= \frac{P_u - P_l\,g_u/g_l}{1 + g_u/g_l} .\end{aligned} \tag{14.24}$$

Thus, the asymptotic photon production rate is proportional to the difference of the two pump rates per sublevel. Every molecule injected into the upper level is potentially able to create a maser photon, while every new molecule added to the lower level can absorb one. Note, however, that the actual population inversion tends to zero in the saturated regime, whatever the difference in pump rates. That is, from equation (14.9), Δn vanishes for $\bar{J} \gg \bar{J}_s$.

14.3 Maser Theory: Further Considerations

Their high brightness temperature is but one property of observed masers. We now turn to other puzzling characteristics, beginning with the remarkably narrow line widths. We also need to address the small apparent sizes of the spots themselves. This last feature relates to the basic issue of how masers are pumped.

14.3.1 Line Narrowing

The relative narrowness of maser lines suggests that some process reduces the line width in the course of propagation. It is easy to see that the strong amplification in the unsaturated regime has just such an effect. Returning to the frequency-dependent transfer equation (14.11), we note that the absorption coefficient α_ν is proportional to $\phi(\nu)$, as explicitly shown in (14.12). Since we are assuming a Doppler absorption profile, we may write

$$\alpha_\nu \;=\; \alpha_\circ \exp\left[-\frac{(\nu-\nu_\circ)^2}{\Delta\nu_D^2}\right] \;, \tag{14.25}$$

where $\alpha_\circ$ is the coefficient at line center.

Suppose that we ignore the final term in equation (14.11) representing spontaneous emission. Then, after propagation through a path length s, the background specific intensity $I_\nu(0)$ amplifies to

$$I_\nu(s) \;=\; I_\nu(0)\,\exp\left(\alpha_\nu s\right) \;,$$

so that

$$\frac{I_\nu(s)}{I_\circ(s)} \;=\; \frac{I_\nu(0)}{I_\circ(0)}\,\exp\left[(\alpha_\nu-\alpha_\circ)s\right] \;. \tag{14.26}$$

The background is presumably some continuum source, for which $I_\nu(0)/I_\circ(0)$ is close to unity. Then equations (14.25) and (14.26) tell us that the propagated intensity falls to e^{-1} of its central value at a frequency shift $\Delta\nu^*$ such that

$$1 \,-\, \exp\left[-\left(\Delta\nu^*/\Delta\nu_D\right)^2\right] \;=\; \frac{1}{\alpha_\circ\, s} \;. \tag{14.27}$$

Expanding the exponential yields an approximate solution for $\Delta\nu^*$

$$\Delta\nu^* \;=\; \Delta\nu_D\;\left(\alpha_\circ\, s\right)^{-1/2} \;, \tag{14.28}$$

which succinctly describes the narrowing as a function of path length.

In more detail, the line can only narrow by a finite amount before the central portion begins to saturate. Since the wings are still growing exponentially, the line begins to broaden and eventually becomes fully saturated. Beyond this point, dI_ν/ds is given by $\alpha_\nu I_s$ rather than $\alpha_\nu I_\nu$. [3] Here the absorption coefficient α_ν still has the frequency dependence given by equation (14.25). Thus, the amplification is linear at all frequencies, and the profile retains a constant, Doppler width from then on.

---

[3] To prove this statement, one must consider not the full level populations, as in equation (14.5), but only that subset which responds to each frequency. One finds that the population difference relevant for the radiative transfer, which appears as $\Delta n\,\phi(\nu)$ in equation (14.12), declines as I_s/I_ν in the saturated regime. (Compare equation (14.9).)

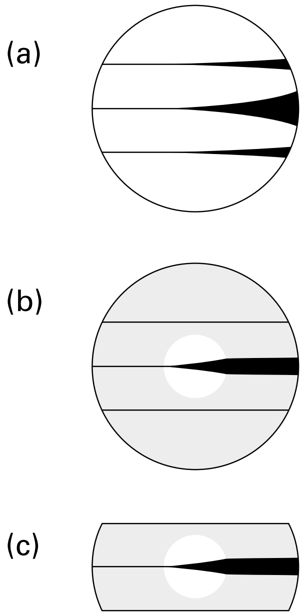


Figure 14.9 Maser beaming in (a) unsaturated spheres, (b) partially saturated spheres, and (c) filaments. The thickness of the horizontal lines indicates the amplification for rays moving to the right. Here we ignore the linear growth following saturation. Notice in (b) and (c) how strong amplification ceases once a ray leaves the unsaturated core, shown in white.

14.3.2 Geometric Beaming

We next consider the geometric pattern of the radiation field. The linear sizes of observed maser spots are so small that the intensity would not be amplified substantially along such a length. Thus, the spots represent the termini of narrow, elongated paths directed away from the observer. There is nothing physically constraining the photons to follow such a tube-like path. However, even radiation traveling through a broader region has a tendency to become more centrally peaked in the course of propagation. This process, analogous to the narrowing in the frequency domain, is known as *beaming*.

The simplest illustration of beaming is provided by a spherical region emitting maser radiation (see Figure 14.9a). Here, we suppose that the seed photons are supplied by spontaneous emission within the sphere. We further suppose, at least initially, that the region is wholly unsaturated. That is, the integrated intensity $\bar{J}$ reaching any point is not sufficient to affect the population inversion, which is maintained by an unspecified, internal pump. The only rays observed are those emanating from the near side of the sphere, and which have the proper direction. The intensity $I(s)$ of all such rays grows exponentially along their respective chords. As illustrated in the figure, the gain is greatest along the sphere's diameter and falls off severely to either side. The observed spot size is clearly less than the full angle subtended by the sphere.

Most interstellar masers are bright enough that their intensity must have saturated. To incorporate this element into our picture, we may imagine weakening the pump throughout the sphere. That is, we lower the rates P_u and P_l, both still presumed to be spatially uniform. The first region to become saturated is the outer boundary, where the radiation, originating from the whole interior, has been amplified the most. As we continue turning down the pump rates, the saturated region advances inward. The sphere now consists of a saturated mantle and an unsaturated, central core (Figure 14.9b). The strongest rays observed are now those which pass through this core, since they undergo at least some exponential growth. Moreover, the path along the core's diameter still provides more amplification than its neighbors. We see that the apparent angular extent of the maser is even smaller under these more general conditions.

Our example of a perfectly spherical maser has no observational or theoretical basis. However, the fully saturated paths in Figure 14.9b have such weak intensities that they hardly contribute to the observed pattern. The original sphere may thus be replaced by a cylindrical, or filamentary emitting region, as illustrated in Figure 14.9c. Such a geometry is, in fact, quite plausible. The filament is distinguished from its surroundings not by a contrast in density, but by its internal *velocity coherence*. Given the differential motion expected in all clouds, maser amplification is only possible if the emitting and absorbing molecules share essentially the same velocity, *i. e.*, if they have a relative Doppler shift no greater than $\Delta\nu^*$ in equation (14.28). Our idealized filament then represents a path along which the velocity vectors happen to have the correct orientation and magnitude. Even such limited coherence is only temporary. Under the action of the differential motion, the vectors become misaligned, and the maser spot fades away. As we remarked in § 14.1, such a view is consistent with the observed temporal fluctuations.

14.3.3 Pumping of H_2O and OH

Velocity coherence naturally plays a role in any detailed account of maser pumping. We noted earlier that the 1.35 cm H_2O masers seem to require a high density. This fact indicates that the masers could be *collisionally* pumped, a supposition that is verified by detailed calculations. Molecules attain their excited state through impact with ambient species, predominantly H_2. The ultimate energy source is the bulk motion of gas, specifically that associated with high-velocity stellar winds. In contrast, OH masers do not require as high a density, but only appear in the vicinity of massive, luminous stars. It is reasonable to suppose, therefore, that the inversion in this case is maintained by either stellar or ambient photons, *i. e.*, that the masers are *radiatively* pumped.

The basic tendency for population inversions in H_2O is not difficult to understand. Referring again to Figure 5.16, we recall the unique status of the backbone states, *i. e.*, the lowest levels at each J-value. The fastest radiative deexcitation and collisional transition rates are between these states, and between those sharing the same J. Thus, molecules in a level above the backbone quickly decay to it. For transitions along the backbone itself, the relatively large Einstein A-values mean that the emitted photons are easily trapped. Frequent upward and downward collisions are therefore more effective than radiation in facilitating these transitions. At any time, then, most of the H_2O molecules are found along the backbone, where they communicate with each other via collisions. Any other level, such as 5_{23}, is fed by the relatively infrequent radiative decays from the nearest backbone state (here 6_{16}) and has a lower population.

How, then, are molecules pumped to relatively high levels within the backbone? Table 14.1 shows that the 6_{16} state lies 640 K above ground. Strong J-shocks, as we noted in Chapter 8, produce H_2O in their postshock cooling regions. The temperature at which this formation occurs is about 400 K. Slower C-shocks also give H_2O emission. Here, somewhat paradoxically, the molecule may be at a significantly higher temperature. In any case, the interface between stellar winds and clouds is an attractive environment for collisional excitation of maser emission. Numerical models confirm that the cloud material must be compressed to densities of order 10^7 cm^{-3} for the excitation to occur. Note that the calculations assume a *preshock* density also of this order, significantly higher than typical cloud values. The actual pumping of H_2O is due to collisions with energetic H_2, which itself either comes from the cloud material (C-shocks), or is reformed downstream on grain surfaces (J-shocks).

Linear strings of H_2O maser spots, like those seen in Figure 14.4, indicate that the relevant shocks lie close to the central axis of a collimated jet. We also saw how, in W49N, the spots are organized into longer-lived centers of activity. The typical size of a center, 10^{14} cm, may represent the diameter of a discrete clump of density 10^8 to 10^9 cm^{-3}, heated and compressed by passage of the shock. Individual spots are the endpoints of transient, velocity-coherent paths within each clump. These paths reside in the relatively thin layer of heated gas created by the sheet-like, propagating shockfront. Indeed, the typical observed spot diameter of 10^{13} cm matches the expected thickness of the postshock region where H_2O is both produced and collisionally excited.

Turning to OH masers, the observed intensities again require that the excitation occur within a gas of elevated density, $n \sim 10^7$ cm^{-3}. This figure exceeds by two orders of magnitude the average internal density of ultracompact HII regions like those in W49N and W3(OH). We have noted that clusters of 18 cm spots appear toward the outer boundaries of their parent HII regions, whenever the latter are detectable through radio emission. These findings together imply that the masers arise in a condensed *shell* surrounding the ionized gas. Theory indeed predicts the existence of just such a zone. As we shall describe in the next chapter, an expanding HII region drives a shock front into the surrounding gas, compressing it in a relatively thin layer.

Figure 14.10 illustrates how the path of maser amplification lies along the tangent to the HII region, here idealized as a perfect sphere of radius R. Suppose we identify the observed spot size d with the thickness of the outer shell. Then the chord shown has a length l, where

$$\begin{aligned} l &= 2\,(R+d)\,\sin\theta \\ &= 2\,(R+d)\left[1-\left(\frac{R}{R+d}\right)^2\right]^{1/2} \\ &\approx 2^{3/2}\sqrt{R\,d}\;. \end{aligned} \tag{14.29}$$

Here the final approximation has utilized $d \ll R$. Substituting representative values for R and d of 5×10^{16} cm and 1×10^{14} cm, respectively, we find that l is 6×10^{15} cm. This figure provides an upper limit to the maser amplification path, which occurs within some velocity-coherent portion of the chord. That is, the emitting filament has an aspect ratio of at most 60:1. Clusters of spots presumably arise because a realistic HII region is significantly warped and so has a number of closely spaced tangent surfaces.

What pumps OH masers? For those associated with post-main-sequence giants, the answer is known with confidence. The stars classified as Mira variables fluctuate regularly in luminosity, over a time scale of about a year. Masers located outside the stars undergo identical variation in their intensity, proving that stellar radiation is the underlying energy source. A similar situation holds for the longer-period OH/IR stars, which are also losing mass at a prodigious rate. What actually excites the molecules in both cases is diffuse, far-infrared radiation

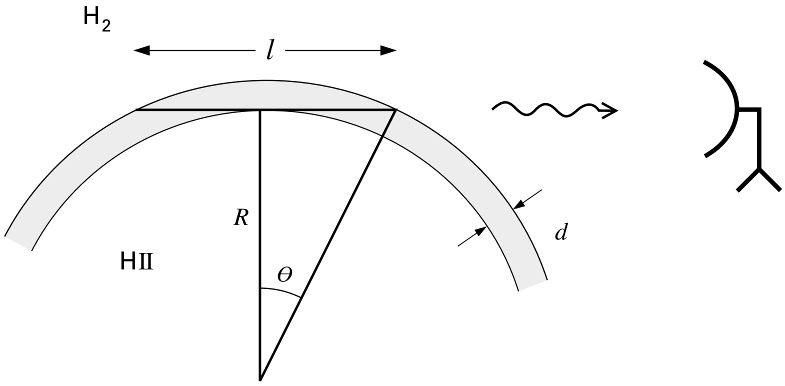


Figure 14.10 Amplification of OH masers within the postshock shell of an HII region. The HII region has radius R and thickness d, and is surrounded by molecular gas. The maser is amplified along the chord of length l, which subtends an angle 2θ at the region's center.

from heated dust grains. The photons promote OH to higher rotational levels, whose subsequent decay through both collisions and radiation leads to the ground-state inversion needed for 18 cm maser emission.

The same mechanism may well hold for OH masers associated with HII regions, but a fully self-consistent picture is not yet available. Unfortunately, the massive stars in question do not display regular variability. The actual fluctuations in maser brightness are also more sporadic and probably stem from turbulent motion, as in H_2O emission. One encouraging sign is that diffuse, far-infrared luminosity and the intensity of 1665 MHz spots are correlated over a large sample of HII regions. Level population calculations that produce inversions for all four transitions require dust and gas temperatures near 150 K. Whether these inversions, created by the ambient radiation field, can be maintained over sufficient paths to yield strong maser emission remains to be seen.

Finally, we emphasize that 18 cm maser emission is only found near *ultracompact* HII regions, *i. e.*, those of diameter 10^{17} cm or less. The older, and more commonly observed, regions of much greater size rarely exhibit maser activity, either from OH or any other species. This trend is also consistent with our theoretical expectation. As an HII region expands, its internal pressure falls. Once this pressure reaches the ambient value, the bounding shock front disappears. We see again how maser emission is a transient, relatively brief phenomenon in the early life of stars.

14.4 Tracing Jets and Outflows

Let us now examine more closely the observed link between masers and outflowing gas. We begin with jets, before turning to the broader molecular outflows associated with low- and high-mass stars. In the latter category, masers generated in the BN-KL region of Orion have been particularly well studied. Finally, we see how the geometric expansion of outflowing masers provides a useful technique for gauging the source distance.

Before launching into the descriptive survey, we remark that the velocity structure of masers is linked to their mode of production. If OH maser activity indeed occurs in the manner illustrated by Figure 14.10, then the observer should see no radial velocity offset of the emission spike relative to the systemic velocity of the source. Of course, the radiation from each filament diverges with a finite angle, so that the filament need not lie precisely on the tangent for its emission to be detectable. Paths that are more offset from the tangent yield lower-intensity spikes, but with finite radial velocity offsets. In summary, the highest-amplitude features within the full line profile should have radial velocities near that of the parent cloud, and the intensity should generally decrease for more red- or blueshifted spikes.

Such indeed is the character of observed OH maser profiles. Our description applies equally well to 1.35 cm H_2O emission, as is evident in Figure 14.2. Here, too, we are dealing with outflowing material whose emission is strongly beamed laterally, *i. e.*, in a plane perpendicular to the local velocity direction. For both types of masers, we have identified this plane with a geometrically thin layer of compressed gas.

14.4.1 Jets and HII Regions

We begin by probing the smallest accessible size scale, that of radio jets. Recall that these structures trace the innermost portions of collimated stellar winds and cover a distance less than 10^3 AU from the driving star. Their continuum luminosity represents the free-free emission from powerful, wind-induced shocks. This picture is corroborated by the finding that H_2O masers may be situated within the same region.

Consider W75N in Cygnus. This active center of massive star formation, about 2 kpc distant, is as embedded as the more extensive W49N, but has a much smaller infrared luminosity, closer to $10^5\ L_\odot$. There are several areas of especially strong radio continuum flux. Figure 14.11 is an interferometric map of one of these, designated VLA 1 (not to be confused with the driving source for HH 1-2 in Orion). The contours display flux at 1.3 cm. It is evident that the emitting area is elongated about the local peak, which marks the position of the massive, driving star. The centimeter radiation spans a length of some 800 AU. A line bisecting this radio jet roughly coincides with the central axis of a parsec-scale CO outflow.

The filled circles in Figure 14.11 represent the 1.35 cm H_2O masers. These are located close to, but not coincident with, the jet axis and have a mean separation from the line of about 100 AU. The elongated distribution of spots is reminiscent of Figure 14.4, where the masers are located much farther from the ionized gas. With a few exceptions, the radial velocities of the spots in Figure 14.11 are near the cloud value of +10 km s^{-1}. Thus, the jet in this case may lie close to the plane of the sky. It is also possible that the masers are not indicating the speed of the wind itself, as we shall discuss shortly.

The W75N area contains several other radio-emitting regions with associated H_2O masers. In one of them, designated VLA 2, the dozen or so spots are more isotropically distributed. Here, the radio continuum region is too small to be resolved. A third region, VLA 3, is more spherical than VLA 1. Both VLA 2 and VLA 3 are extreme examples of ultracompact HII

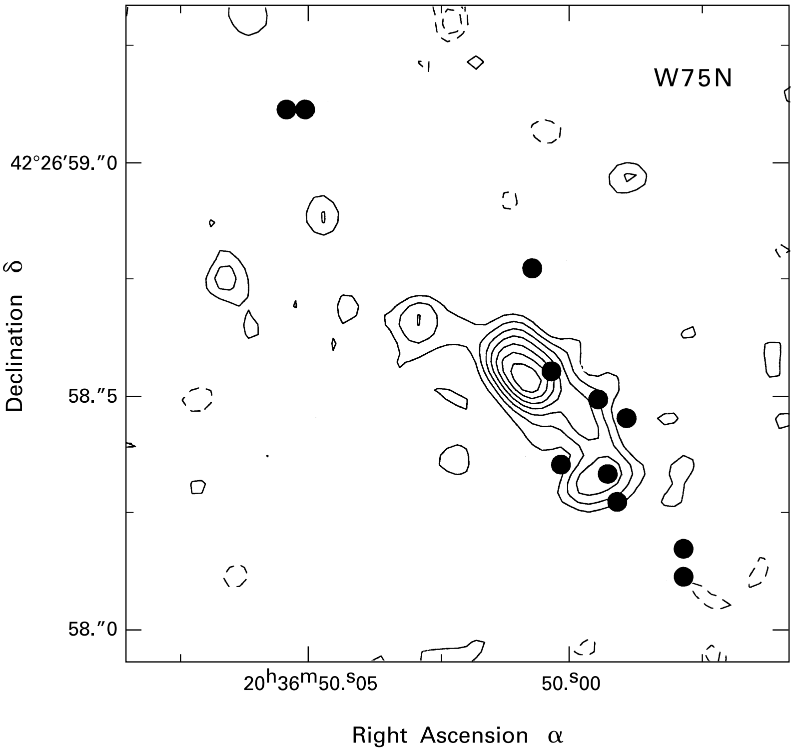


Figure 14.11 Jet-like configuration of H_2O masers in W75N. The contours represent 1.3 cm continuum emission, while the filled circles show the maser positions.

regions, with diameters less than 10^{16} cm. It is apparent that H_2O masers can also arise in such environments, where the ionization is spreading through especially dense gas.

Strings of H_2O masers are found not only outside HII regions, but are also generated by low-mass stellar winds. We previously discussed, in § 14.1, the variability of the spots outside the embedded star SSV 13. Consider also the Class 0 source L1448/mm, whose spectral energy distribution we showed in Figure 4.4. This star drives a vigorous molecular outflow, with relatively narrow CO contours and high internal velocity. As seen in Figure 14.12, patches of emission in the 2.12 μm line of shocked H_2 lie along the outflow's central axis. Radio observations reveal a handful of 1.35 cm masers situated on either side of the star and spanning a total distance of about 100 AU. A line joining the masers has the same orientation as the much larger (10^4-AU) jet delineated by the 2.12 μm emission. Moreover, the radial velocities of the maser spots are now offset from the cloud value, and in the expected sense. Thus, the upper two masers within the inset of Figure 14.12 are blueshifted, while the lower three are redshifted.

Yet another instructive example is the HH 212 outflow, whose remarkably symmetric knots in shocked H_2 were previously shown as Figure 13.6. Here, the several dozen observed H_2O masers lie within 70 AU of the star and close to the jet axis inferred from both H_2 and from a

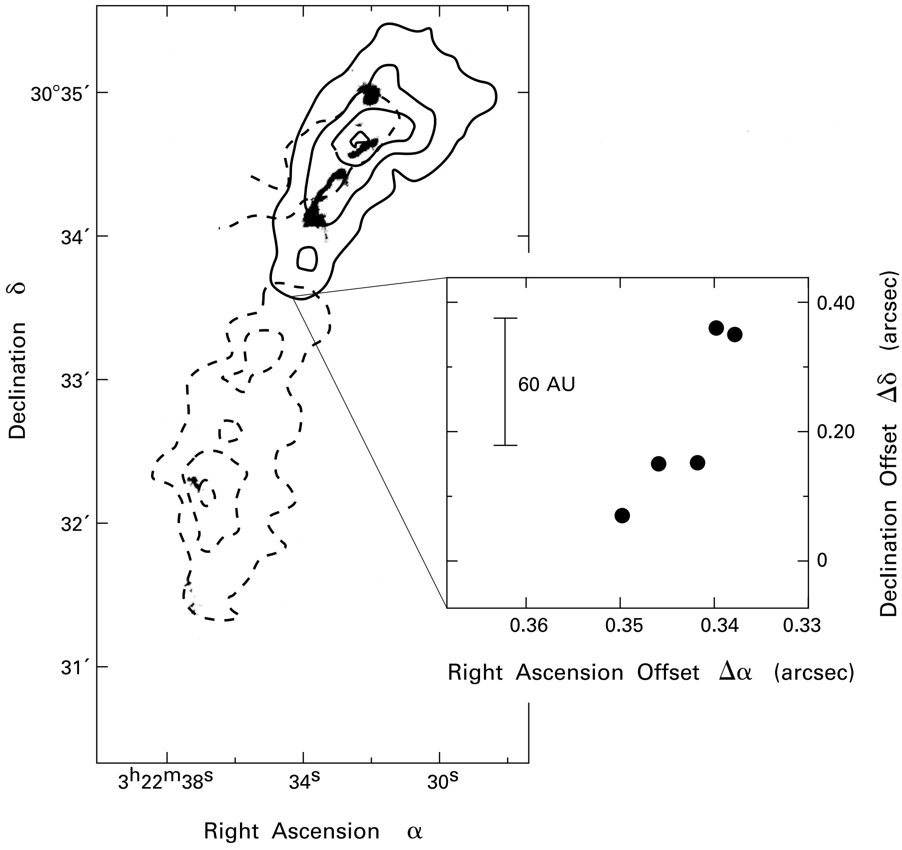


Figure 14.12 Molecular outflow and H_2O masers in L1448. The solid and dashed contours show, respectively, the blue- and redshifted emission from the $J = 2 \rightarrow 1$ emission of $^{12}C^{16}O$, while the dark patches are the 2.12 μm emission of H_2. The filled circles within the inset show positions for the 5 brightest H_2O masers. The angular offsets are measured from $\alpha = 3^h22^m34^s$, $\delta = 30°33'35''$.

broader CO outflow. Radio interferometric observations taken several months apart reveal that the maser spots are moving away from the central star. All these features are highly variable in their intensity. Some of them lie along an arclike structure that mimics the larger H_2 bowshocks. This latter observation is especially important, as it bolsters the idea that H_2O masers are energized through the shocking of wind gas.

14.4.2 Clumpy Outflows

The L1448 and HH 212 outflows are hardly unique, or even unusual, in displaying maser activity. Examination of over a hundred molecular outflows finds that at least 50 percent have 1.35 cm emission, above a flux level of 1.5 Jy. Moreover, the observed maser spots are invariably located close to the driving star. Some of the latter, like L1448/mm, are Class 0 sources, which have especially high cloud density in this region. Many of the outflows, however, are driven by less embedded objects. The preshock density of 10^7 cm^{-3} required for maser excitation is greater than that in cloud cores and must be associated with individual, wind-generated clumps.

Whatever process in the wind actually forms these clumps and shocks them only operates within 100 AU or so from a low-mass star. It is natural to relate this vigorous activity to that powering the free-free emission in the same region. There is, in fact, a good correlation between the total luminosity in the 1.35 cm maser line and the radio continuum. For massive stars, on the other hand, we have seen from the example of NGC 6334F that H_2O masers may be situated far from any ionized gas, although still relatively close to their own source of hot gas and wind momentum.

The study of such environments has made it clear that H_2O masers appear extremely early in the life of a massive star, before any HII region at all can be detected. The dense gas surrounding the star and containing the masers is known as a *hot core*; we shall be returning to this topic in Chapter 15. In older regions, ionized gas has spread outward from the star. Since no molecules can survive within the high temperature of an HII region, it may seem puzzling that H_2O masers persist, as in the example of W49N. High-resolution observations reveal, in this case, that the ionization is patchy, and that the maser activity occurs only within neutral regions that are presumably shielded from the destructive ultraviolet photons.

Returning momentarily to NGC 6334F and Figure 14.4, surveys have found this geometric situation to be quite common. That is, cometary HII regions frequently have clumps or linear strings of H_2O masers lying beyond the apex of the curved arc. This arc tells us where ionization has been stalled because of a steep density rise in the surrounding material (Chapter 15). On the opposite side of the star, ionization spreads unimpeded. The H_2O masers project from their own massive star, as we know from the presence of heated molecules in their vicinity. The associated hot core must stifle local ionization, preventing any detectable HII region from arising. On the other hand, this more deeply embedded star is clearly powering a jet-like flow that penetrates the hot core.

Precisely how dense gas actually collimates a wind is no better understood for massive stars than for their low-mass counterparts. In both situations, we have a much better hope of sorting out the dynamics if the shocked gas is actually moving at the local wind velocity. We argued in Chapter 13 for the plausibility of this assumption with regard to Herbig-Haro objects. In that case, the internal knots represent low-amplitude shocks that travel nearly at the jet speed. For masers, however, the precise nature of the shock interaction is not yet known. The high densities required by the pumping models indicate, as we have noted, that emission is activated by the collision of internal clumps, each of which could be moving more slowly than the wind that is pushing them.

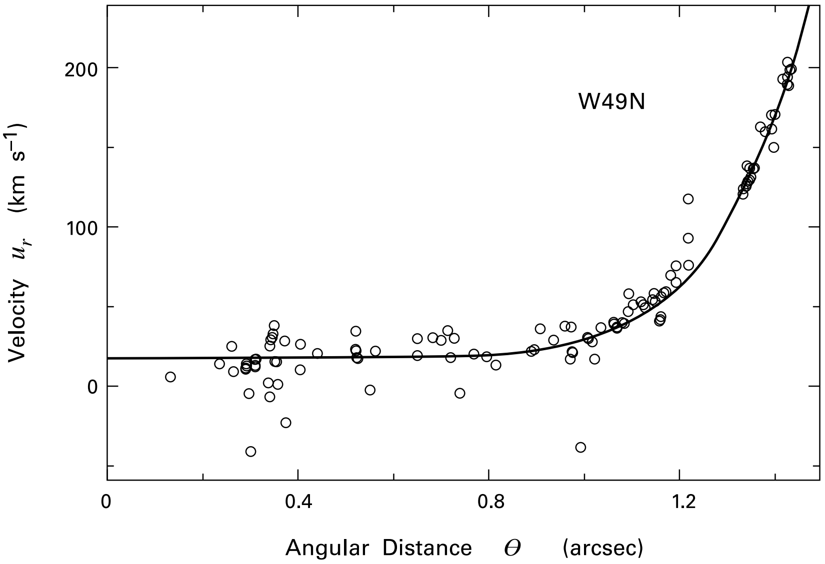


Figure 14.13 Velocity of H_2O masers in W49N as a function of angular distance from their common center of expansion. The velocity component shown is that directed along the radius to the center. The smooth curve is an analytic fit to the data.

An important clue in this regard is the magnitude of the observed, radial velocities themselves. For the broader flows of maser spots, these speeds are indeed much lower than in a plausible stellar wind. Figure 14.13 displays the H_2O maser speeds in W49 as a function of angular distance from the center of expansion. What is actually shown is u_r, the velocity projected onto the radial direction from that center. Out to an angle of about $1''$, corresponding to a physical distance of 1×10^{17} cm, the speed is nearly constant. Its average magnitude of 17 km s^{-1} is only a few percent of the wind speed from an O-type star. Equally striking is the dramatic acceleration for large radii. There is no generally accepted explanation for these peculiar features. They strongly suggest, however, that the observed velocities represent not the wind as it is launched from the star, but the turbulent and clumpy flow created when that high-speed gas enters the surrounding cloud.

14.4.3 The Orion Masers

Acceleration similar to that in W49 occurs within the ultracompact HII region W3(OH) and at other sites, as well. One particularly well-studied example is the BN-KL region of Orion. In this case, there is only a tiny HII region surrounding the powerful infrared source IRc2. (For the morphology of this area, recall Figure 1.7.) Yet there is strong evidence for an expanding flow centered on that position. In this case, the ambient density is apparently high enough to suppress the growth of ionization, but not to collimate the especially strong wind. Proper motions of H_2O masers, combined with Doppler velocities, show that the spots are moving away from IRc2 in an anisotropic, roughly bipolar fashion. The measured speed is nearly constant out to a radius of

2×10^{17} cm. Here, the magnitude is 18 km s^{-1}, remarkably close to the W49N value. Clumps of masers, analogous to the W49N centers of activity, are seen within the interior region. Farther out, the flow is more isotropic and faster, with the velocity eventually surpassing 100 km s^{-1}.

Observers have closely examined the BN-KL region in many other molecular lines. Their findings both support the picture of an accelerating outflow and lend credence to the supposition that H_2O masers are moving at the local gas speed, even if it is not the pure wind value. Activity in SiO masers is seen very close to IRc2, within a distance of order 10^{15} cm. Emission comes from vibrationally excited ($v = 1$ and $v = 2$) SiO, indicating an ambient temperature near 2000 K. The Doppler shift of the line yields a modest velocity, consistent with that found from H_2O. At a radius of some 10^{16} cm, SiO emission is thermal (*i. e.*, non-maser) and comes from the ground vibrational state. The measured radial velocity is again consistently low. So too is that derived from strong, narrow emission in several inversion lines of NH_3. At even greater distance from IRc2, the millimeter emission from CO indicates that the velocity has climbed by an order of magnitude. Toward the outskirts of the CO, at a distance exceeding 10^{17} cm, are the peaks in the 2.12 μm line of H_2. The high-speed gas in this region has an elevated temperature that signifies internal energy dissipation through shocks. (Recall Figure 8.9.) Some of these shocks appear as the protruding fingers in Plate 13.

The observations of SiO maser emission in Orion are an object lesson in how the prevalent physical interpretation of a phenomenon may change markedly with improvements in imaging capability. The $v = 1$ $(J = 2 \rightarrow 1)$ line from the BN-KL region is the strongest in the Galaxy and was detected as early as 1974. The line had two distinct peaks, well separated in radial velocity. Judging from similar observations of OH masers associated with post-main-sequence stars, it appeared that the radiation emanated from an expanding shell, whose front and back sides were being viewed simultaneously.

By 1980, interferometry had revealed that the SiO signal was associated with IRc2. Subsequent observations showed that the two peaks actually came from separate areas on either side of this source. Each region, in fact, appeared to trace part of an elliptical arc. From the 1980s through the late 1990s, this emission was generally ascribed to a rotating, expanding ring. However, with even higher, sub-milliarcsecond resolution, what had appeared to be arcs broke up into more dispersed collections of numerous spots (see Figure 14.14). All the spots on one side of the source are redshifted with respect to IRc2, while those on the opposite side are blueshifted. The overall impression is now of a wide-angle outflow that participates in the larger one seen through H_2O masers.

14.4.4 Expanding Cluster Parallax

The fact that the H_2O masers in Orion appear to be streaming away from a common source offers a practical means to determine the distance to that source, *i. e.*, to IRc2 itself. Suppose one constructs a three-dimensional model of the spot motion. That is, one writes down an analytic expression for the velocity components at any position as a function of the distance to the center of expansion, where the precise location of the latter is left momentarily as an unknown parameter. The observational data include coordinates in the plane of the sky for all maser spots, as well as their line-of-sight velocities, V_r. From the model, one can predict the *transverse* velocities, $V_\perp$, of these same spots. But one also observes proper motions of the

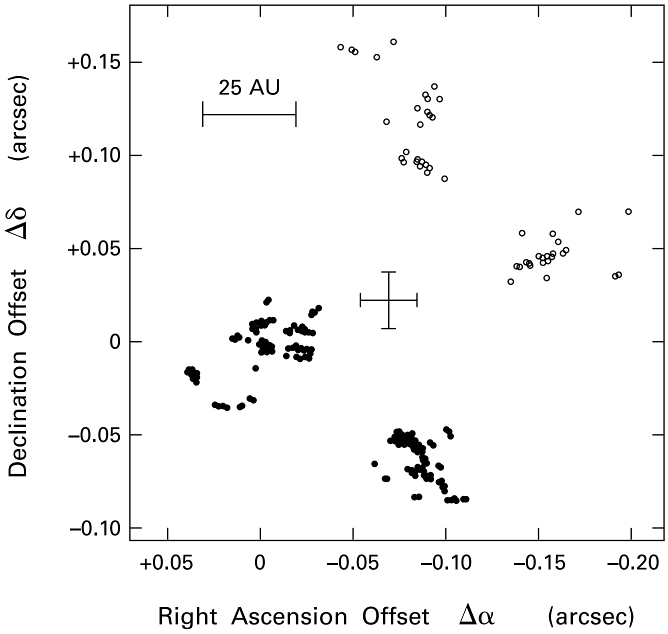


Figure 14.14 Clumps of SiO masers in the Orion BN-KL region. Filled circles represent masers that are blueshifted with respect to IRc2; open circles are redshifted. The crosshairs show the centroid and position uncertainty of a feature within IRc2 known as Source I. Angular offsets are measured from $\alpha = 5^{\rm h}35^{\rm m}14.^{\rm s}5,\ \delta = -05^\circ 22' 36.''3$.

masers. These angular displacements may also be converted to transverse velocities, *provided one knows the distance*. One therefore adjusts both the assumed center of expansion and the distance until the two sets of transverse velocities best agree.

It is apparent from our description that this technique of *expanding cluster parallax* is akin to the moving cluster method used to set the Hyades distance (Chapter 4). In the case of the H_2O maser stream in Orion, the parallax method yields a distance of 480 pc, close to the value obtained by other means. The expanding crowd of spots in W49N is also amenable to such study. Here, the derived distance to the system is 11.4 kpc. The method has also been applied to the H_2O masers in Sgr B2/North, the star-forming cloud situated close to the Galactic center (Figure 5.3). The inferred distance of 7.1 kpc is somewhat less than the accepted value of 8.5 kpc for the Galactocentric radius.

We close this chapter by stressing that the nature of the motion being revealed by masers is the subject of ongoing study and remains problematic in some cases. The situation is complicated by the variability of maser emission, which can make it difficult to discern a stable pattern of spots. On occasion, the string of H_2O masers close to a star is found *not* to be aligned with the axis of a radio jet or CO outflow, but is instead more orthogonal to that axis. If the masers also exhibit a significant gradient in radial velocity, one is tempted to place them in a circumstellar disk, presumably one seen nearly edge-on. Certainly the requisite high densities would be plausibly attained in such an environment. It is by no means clear, however, how the masers could be collisionally pumped, especially if the wind itself is predominantly normal to the disk plane. We must await future, high-resolution observations, including proper motion data, to elucidate the true conditions.

Chapter Summary

Embedded young stars create flickering spots of high-intensity radio emission. These masers arise when either the stellar wind or radiation field energizes molecules within a dense gaseous environment. The two most commonly observed types of masers are the 1.35 cm rotational transition of H_2O and the hyperfine OH lines near 18 cm. However, numerous other molecules exhibit the effect. In all cases, the spots are not only extraordinarily bright, but also small in size and narrowly confined in frequency.

Maser emission occurs when some external energy source overpopulates the upper level of a transition. Stimulated emission then causes the intensity to rise exponentially along some path where the molecules are aligned in velocity. The rapid intensity growth creates geometrical beaming into ever narrower spots and decreases the width of the line. Ultimately, the radiation intensity becomes so high that it begins to depopulate the upper level, and the amplification saturates.

Shocks associated with stellar winds provide the energy for H_2O masers. These shocks occur within dense clumps created by the wind. Strings of H_2O masers are found near both high- and low-mass stars, and are often aligned with radio continuum jets or shocked H_2 emission. The observed spatial velocities of the spots trace the flow pattern of the underlying wind. In contrast, OH masers arise in the outskirts of HII regions exclusively surrounding massive stars. The pump in this case may be far-infrared radiation from heated dust. Both types of mechanism power the various maser lines of CH_3OH, which trace dense gas in or near compact HII regions.

Suggested Reading

Section 14.1 The serious student of this topic will want to consult the comprehensive text by

Elitzur, M. 1992, *Astronomical Masers* (Dordrecht: Kluwer).

Ongoing research in a variety of astrophysical contexts is described in

Migenes, V. & Reid, M. (eds.) 2002, *Cosmic Masers: from Protostars to Black Holes* (San Francisco: ASP).

The location of OH and H_2O masers near massive stars was studied by

Forster, J. R. & Caswell, J. L. 1989, AA, 213, 339.

For the H_2O masers near low-mass stars, see

Wilking, B. A., Claussen, M. J., Benson, P. J., Wootten, A., Myers, P. C., & Tereby, S. 1994, ApJ, 431, L119

Brand, J., Cesaroni, R., Comoretto, G., Felli, M., Palagi, F., Palla, F., & Valdettaro, R. 2003, AA, 407, 573.

The second reference summarizes a long-term study of the variability of these sources, including those associated with massive stars. For the complex properties of CH_3OH masers, see

Menten, K. M. 1991, in *Skylines: Proceedings of the Third Haystack Conference on Atoms, Ions, and Molecules*, eds. A. D. Haschick & P. T. P. Ho (San Francisco: ASP), p. 119.

Section 14.2 The basic theory of masers is covered well in several older review articles. Two especially clear, and still useful, ones are

Moran, J. M. 1976, in *Frontiers of Astrophysics*, ed. E. H. Avrett (Cambridge: Harvard U. Press), p. 385

Cohen, R. J. 1989, Rep. Prog. Phys., 52, 881.

Section 14.3 An early and influential paper treating beaming effects in various geometries is

Goldreich, P. & Keeley, D. A. 1972, ApJ, 174, 517.

The basic argument for collisional pumping of H_2O masers was articulated by

de Jong, T. 1973, AA, 26, 297.

The subsequent elaboration of this mechanism using J-shocks is due to

Elitzur, M., Hollenbach, D. J., & McKee, C. F. 1989, ApJ, 346, 983.

For a discussion of OH maser pumping, see

Cesaroni, R. & Walmsley, C. M. 1991, AA, 241, 537.

Section 14.4 The spatial relation of H_2O masers and ultracompact HII regions is the subject of

Hofner, P. & Churchwell, E. 1996, AAS, 120, 283.

One example of the maser-jet connection for low-mass stars is the HH 212 system:

Claussen, M. J., Marvel, K. B., Wootten, A., & Wilking, B. A. 1998, ApJ, 507, L79.

The kinematics of the W49N H_2O masers was analyzed by

Gwinn, C. R., Moran, J. M., & Reid, M. J. 1992, ApJ, 393, 149,

while the flow in the Orion BN-KL region was first studied in

Genzel, R., Reid, M. J., Moran, J. M., & Downes, D. 1981, ApJ, 244, 884.

The Formation of Stars
Steven W. Stahler and Francesco Palla
© 2004 WILEY-VCH Verlag GmbH & Co.

15 Effects of Massive Stars

We now conclude Part IV by exploring more systematically the various ways in which stars of the highest mass influence surrounding matter. The salient characteristic of any massive star is its extreme energy output, much of which is at ultraviolet wavelengths. It is appropriate, therefore, to begin with a general discussion of HII regions. After presenting the essential theory of ionization fronts, we describe the observed features of both mature HII regions and the younger, ultracompact variety, whose nature is not as well understood. We next turn our attention to the hot cores surrounding the very youngest, massive stars. Here the gas density is so high that it slows the spread of ionization. While the detailed structure and dynamics of the cores themselves is not established, infrared observations and maser studies have nevertheless provided valuable information.

Powerful winds are another feature of massive stars and are intimately linked to their high luminosities. The winds, in fact, are driven outward by the momentum of ultraviolet photons, an acceleration mechanism wholly different from that pertaining to low-mass objects. After describing this physical process, we also show how the high-velocity gas generates both radio and X-ray emission. We then broaden our scope to encompass the large molecular outflows associated with individual massive stars and clusters. The outflow morphology broadly resembles that of lower-mass analogues, but there are interesting, qualitative differences.

Our last two topics concern the disruption and clearing of gas by high-mass stars. We start at the smallest scale, with the photoevaporation of circumstellar disks and the erosion of compact clouds within HII regions. The latter are bodily compressed by ambient radiation; the enhanced density may then lead to more stellar production. Finally, we describe how supernovae eject high-speed shells that both crush and vaporize nearby gas. We examine critically the possibility that the associated gas compression spawns a new generation of stars.

15.1 HII Regions

Consider a massive star buried within the cloud material from which it was born. For simplicity, let us suppose that the cloud is composed initially of pure molecular hydrogen. Energetic photons streaming from the stellar surface both dissociate the H_2 and ionize the resulting HI. At the same time, electrons and protons in this plasma recombine, creating new atomic hydrogen. Since each ionization event deletes a photon from the beam, a star with a fixed output in ultraviolet radiation can only ionize a limited region in the surrounding cloud. If this matter has uniform density, the ionization spreads isotropically and fills out a volume known as the *Strömgren sphere*. Let us first derive the properties of this simplest of HII regions.

The Formation of Stars. Steven W. Stahler and Francesco Palla
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-40559-3

15.1.1 The Strömgren Sphere

We begin by noting that the condition of ionization balance holds at each location within the region. That is, the volumetric rate at which any parcel of gas is being ionized just offsets the recombination rate of free electrons and protons in the same parcel. Integrating over the whole sphere, the total number of ionization events per unit time equals the total rate of recombinations.[1] But the former must be $\mathcal{N}_*$, the rate at which the star emits ionizing photons, *i. e.*, those with energy E above 13.6 eV. The minimum frequency here is the Lyman limit of $\nu_1 \equiv 3.29 \times 10^{15}\ \mathrm{s}^{-1}$, corresponding to $\lambda_1 = 912\,\text{Å}$. Table 15.1 lists values of $\mathcal{N}_*$ for stars spanning a range of spectral types. The table also gives these stars' far-ultraviolet output ($6 < E < 13.6$ eV), a figure that will enter our later discussion (§ 15.4).

Table 15.1 Ultraviolet Radiation from Massive Stars

Spectral Type	Mass ($M_\odot$)	log $\mathcal{N}_*$ (s^{-1})	log $\mathcal{N}_{\rm FUV}$ (s^{-1})
O4	70	49.9	49.5
O5	60	49.4	49.2
O6	40	48.8	48.8
O7	30	48.5	48.6
O8	23	48.2	48.4
O9	20	47.8	48.2
B0	18	47.1	48.1
B1	13	45.4	47.5
B2	10	44.8	47.1

To describe recombinations qualitatively, we introduce $\alpha_{\rm rec}(T)$, the *recombination coefficient*. This quantity, with units of $\mathrm{cm}^3\,\mathrm{s}^{-1}$, is a function only of the ambient temperature within the HII region. The volumetric rate at which free electrons and protons combine to produce atomic hydrogen is then written as

$$\begin{aligned} \mathcal{R} &= n_e\, n_p\, \alpha_{\rm rec}(T) \\ &= n_e^2\, \alpha_{\rm rec}(T)\ , \end{aligned} \tag{15.1}$$

where the last expression uses charge neutrality in the ionized plasma.

The total rate of recombinations within the sphere follows by integrating $\mathcal{R}$ over the volume. Assuming, for simplicity, that both n_e and T are spatial constants, we balance total ionizations and recombinations to find

$$\mathcal{N}_* = \frac{4\,\pi}{3}\, n_e^2\, \alpha'_{\rm rec}(T)\, R_s^3\ . \tag{15.2}$$

The quantity R_s is the *Strömgren radius*. Note that the recombination coefficient, written here with a prime, differs from the one in (15.1). The latter includes recombinations to all possible

---

[1] The assumption of global ionization balance fails if the ambient density is sufficiently low, or if the cloud gas is clumpy. Ionizing photons that escape from porous HII regions may be the energizing source for the warm neutral medium; recall § 2.1.

levels of HI, including the ground ($n = 1$) state. But in considering global ionization balance, we should neglect recombinations to $n = 1$. Each such event produces another ionizing photon that is promptly absorbed nearby.

We shall later show that the ionization spreads so quickly to the Strömgren radius that the original cloud density is not able to change appreciably. Thus, we may equate n_e in equation (15.2) with n_H^0, the number density of hydrogen atoms external to R_s. (Since the cloud is presumed to be molecular, n_H^0 is twice $n_{H_2}^0$, the number density of H$_2$.) We then solve equation (15.2) for R_s to find

$$\begin{aligned} R_s &= \left[\frac{3\,\mathcal{N}_*}{4\,\pi\,\alpha'_{\rm rec}\,(n_H^0)^2}\right]^{1/3} \\ &= 0.4\,{\rm pc}\,\left(\frac{\mathcal{N}_*}{10^{49}\ {\rm s}^{-1}}\right)^{1/3}\left(\frac{n_{\rm H_2}^0}{10^3\ {\rm cm}^{-3}}\right)^{-2/3} \quad . \end{aligned} \tag{15.3}$$

In the numerical expression, we have used an $\mathcal{N}_*$-value of $10^{49}\ {\rm s}^{-1}$, corresponding to an O6 star. We have also adopted a temperature of 10^4 K for the purpose of evaluating $\alpha'_{\rm rec}$, which is then $2.6 \times 10^{-13}\ {\rm cm}^3\ {\rm s}^{-1}$. The characteristic temperature of an HII region, in other words, is of the same order as, but lower than, the surface value for the central star. We shall indicate presently how this figure is derived.

When the hydrogen within the HII region is exposed to the flux of Lyman continuum photons from an O or B star, the probability per unit time of ionizing an atom is far greater than the probability for a proton to recombine with any of the free electrons in its vicinity. For the *volumetric* rates to match, therefore, the number density of neutral atoms has to be very much smaller than the density of protons (or electrons). In other words, the interior of the HII region is nearly completely ionized. The neutral density becomes significant only when the photon flux becomes attenuated by the ionizations it is creating. The transition to neutrality at the boundary of an HII region thus occurs over some multiple of the photon mean free path within the fully neutral medium, *i. e.*, over Δr, where

$$\Delta r = \frac{1}{n_H^0\,\sigma_{\nu_1}} \quad . \tag{15.4}$$

Here σ_{ν_1} is the photoionization cross section of a ground-state hydrogen atom, evaluated at the Lyman limit. Using $\sigma_{\nu_1} = 6.8 \times 10^{-18}\ {\rm cm}^2$ and our representative n_H^0 of $2 \times 10^3\ {\rm cm}^{-3}$, we find that $\Delta r = 7.4 \times 10^{13}$ cm, or about 5×10^{-5} of R_s itself. Thus, HII regions tend to be very sharp-edged.

15.1.2 First and Second Expansions

So far, we have treated the Strömgren sphere as if it were a static entity. In fact, the zone of ionization expands with time. At the earliest stages, no HII region at all exists around the star, a situation we shall presently explore. Later, the region spreads, first quickly to the Strömgren radius, and then more slowly beyond. Let us follow the sequence of events in more detail. For now, we retain our simplifying assumption that the parent molecular cloud has a spatially uniform density.

The moving edge of the HII region is known as the *ionization front*. From our earlier remarks, the true boundary has a finite width, but is narrow enough that we may treat the front as a sharp discontinuity. When, at early times, the front radius R is less than R_s, the total rate of recombinations within the HII region is insufficient to match the star's output of ionizing radiation. Thus, surplus photons reach the front itself; these ionize additional gas beyond it. In this way, the front expands.

Figure 15.1 shows in more detail how the spreading occurs. Let $F_*(t)$ be the flux of ionizing photons reaching R at time t. Over a subsequent interval Δt, the front advances a distance ΔR, engulfing $n^0_{H_2}\,\Delta R$ hydrogen molecules per unit area. Every one of these molecules is first dissociated during Δt, creating $2\,n^0_{H_2}\,\Delta R$ neutral atoms. Each of these atoms is then also ionized. The number of photons crossing the front is $F_*\,\Delta t$. We recall from Chapter 5 that direct dissociation of a hydrogen molecule requires a single photon with energy exceeding 14.7 eV. An additional two photons of energy greater than 13.6 eV are needed to ionize both hydrogen atoms. In summary, three photons create two ionizations. Thus

$$\frac{F_*\,\Delta t}{2\,n^0_{H_2}\,\Delta R} = \frac{3}{2}\ ,$$

so that

$$\begin{aligned}\frac{dR}{dt} &= \frac{F_*}{3\,n^0_{H_2}}\\ &= \frac{2\,F_*}{3\,n^0_H}\ .\end{aligned} \qquad (15.5)$$

Here, we have implicitly assumed that the fraction of the star's energy output above either 13.6 or 14.7 eV is essentially the same.

Consider next the recombinations occurring within the spreading HII region. The total rate of such events leading to $n \geq 2$ hydrogen levels is now $(4\pi/3)\,n_e^2\,\alpha'_{\rm rec}\,R^3$. Assuming that ionization balance still holds, this figure also represents the number of stellar photons lost to ionization per unit time. But the rate at which the surviving photons cross the ionization front is $4\pi R^2 F_*$. We thus find

$$\mathcal{N}_* = 4\pi R^2 F_* + \frac{4\pi}{3}\,(n^0_H)^2\,\alpha'_{\rm rec}\,R^3\ , \qquad (15.6)$$

where we have again equated n_e with n^0_H.

By combining equations (15.5) and (15.6), we arrive at an equation of motion for the ionization front:

$$\frac{dR}{dt} = \frac{\mathcal{N}_*}{6\,\pi\,n^0_H\,R^2} - \frac{2}{9}\,n^0_H\,\alpha'_{\rm rec}\,R\ . \qquad (15.7)$$

We define a nondimensional radius λ as R/R_s and a nondimensional time $\tau \equiv t/t_{\rm rec}$. Here, $t_{\rm rec}$ is the recombination time within the HII region:

$$\begin{aligned}t_{\rm rec} &\equiv \frac{1}{n^0_H\,\alpha'_{\rm rec}}\\ &= 61\ {\rm yr}\left(\frac{n^0_{H_2}}{10^3\ {\rm cm}^2}\right)^{-1}\ .\end{aligned} \qquad (15.8)$$

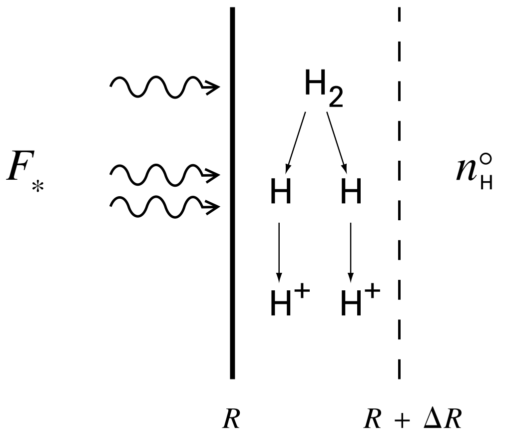


Figure 15.1 Spreading of the ionization front into a cloud with hydrogen number density n_H^0. For every three ultraviolet photons crossing the front at R, one dissociates a hydrogen molecule and two more ionize the resulting atoms.

Cast in these new variables, equation (15.7) becomes

$$\frac{d\lambda}{d\tau} = \frac{2}{9\,\lambda^2} - \frac{2\,\lambda}{9} \, . \tag{15.9}$$

The solution obeying $\lambda(0) = 0$ is then

$$\lambda = [1 - \exp(-2\tau/3)]^{1/3} \, . \qquad \textit{first expansion} \tag{15.10}$$

Equation (15.10) shows that the front races out at high speed (formally infinite at the start) and then slowly creeps toward the Strömgren radius, at which point the speed ostensibly vanishes. Specifically, R reaches a fraction $\left(1 - \mathrm{e}^{-1}\right)^{1/3} = 0.86$ of R_s in a time $3\,t_{\mathrm{rec}}/2$, or 91 yr for our canonical parameters. Over this interval, the average speed is very high, about 6×10^3 km s^{-1}.

Our analysis has proceeded under the assumption that the mass density within the HII region matches the exterior value. This simplification is justified at early times, when the front's velocity exceeds the isothermal sound speed a_1 within the ionized region. Note that this speed is 11 km s^{-1} in an ionized gas at 10^4 K. Since a_1 represents the velocity at which small pressure disturbances propagate, the density has no time to readjust. However, the *pressures* inside and outside the HII region are vastly different. With its temperature near 10^4 K, the interior has a pressure greater by three orders of magnitude. Once R approaches R_s and the front speed begins to diminish, the pressure difference is able to drive a second phase of expansion for the HII region. Moreover, because the difference is so large, the ionization front is preceded by a shock wave that first expands into the ambient cloud (see Figure 15.2).

The second phase begins when the velocity of the ionization front declines to the internal sound speed. At that point, pressure disturbances from within the HII region are able to cross the front and create an expanding shock. External matter traversing this shock is compressed

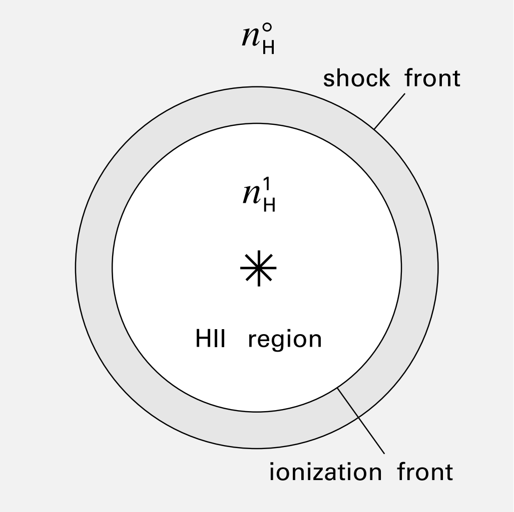


Figure 15.2 Second expansion phase of a spherical HII region. The ionized volume surrounding the massive star has the relatively low density n_H^1. It is bounded by a thin shell of neutral matter. This shell advances supersonically into the surrounding cloud of density n_H°, creating an expanding shock front.

into a relatively thin, neutral shell that bounds the HII region. Both the shock and the ionization front move out at a common speed that soon falls below a_1. The mass of ionized gas continues to rise, as ultraviolet photons impinge on the inner side of the shell. Meanwhile, expansion allows the interior density to fall. Thus, the system gradually approaches pressure equilibrium.

Gas within the shell has the low, cloud temperature at the leading edge and is heated close to 10^4 K approaching the HII region. Material toward this inner surface emits strongly in various infrared transitions, such as [C II] 158 µm and [O I] 63 µm. Observationally, these areas appear as the photodissociation regions we studied in § 8.3. The most famous example is the Orion Bar (Figure 1.5). Here, a variety of atomic and molecular lines are seen ahead of the visible ionization front, *i. e.*, away from the Trapezium stars.

To derive the motion of the shell, we use the fact that postshock cooling is so efficient that the shock may be regarded as isothermal. The reader may verify, using Appendix F, that the pressure inside the shell equals $n_H^0 m_H (dR/dt)^2$, where m_H is the hydrogen atom mass. But this pressure matches that inside the HII region, which is $n_H^1 m_H a_1^2$. Here, n_H^1 is the interior number density of hydrogen atoms, which now falls below n_H^0. We thus have

$$n_H^1\, a_1^2 \,=\, n_H^0\, (dR/dt)^2 \ . \tag{15.11}$$

During the expansion, nearly all the stellar ionizing photons are still consumed by the HII region, with only a small residual eroding the neutral shell.[2] We thus modify equation (15.3) to read

$$R \,=\, \left[\frac{3\,\mathcal{N}_*}{4\,\pi\,\alpha'_{\rm rec}\,(n_H^1)^2}\right]^{1/3} \ . \tag{15.12}$$

---

[2] Photons leaving the HII region with energy between 11.2 and 13.6 eV can dissociate H_2 by first exciting it into the Werner band. (Recall § 5.2.) A mantle of HI gas created by this means surrounds both the shock and ionization fronts for much of the second expansion; see Chapter 18.

Elimination of n_H^1 between (15.11) and (15.12) gives the equation of motion during the second expansion:

$$\left(\frac{dR}{dt}\right)^2 = \frac{a_1^2}{n_H^0}\left[\frac{3\,\mathcal{N}_*}{4\,\pi\,\alpha'_{\rm rec}\,R^3}\right]^{1/2} . \tag{15.13}$$

The velocity of the shell is thus $a_1(R/R_s)^{-3/4}$. To find the radius at all times, we again let $\lambda \equiv R/R_s$. However, we define a new dimensionless time as $\tau' = a_1 t/R_s$. Equation (15.13) then becomes

$$\frac{d\lambda}{d\tau'} = \lambda^{-3/4} . \tag{15.14}$$

Since the initial approach to the Strömgren radius takes relatively little time, we adopt the initial condition $\lambda(0) = 1$. Integrating equation (15.14) then yields

$$\lambda = \left(1 + \frac{7\,\tau'}{4}\right)^{4/7} . \qquad \textit{second expansion} \tag{15.15}$$

15.1.3 Champagne Flows

In the final, equilibrium state, n_H^1 must fall to $(a_0/a_1)^2 n_H^0$, *i. e.*, far below the external density. It follows from equation (15.12) that the final radius R_f is

$$R_f = (a_1/a_0)^{4/3}\,R_s . \tag{15.16}$$

Written in terms of temperature, the expansion factor is $(2T_1/T_0)^{2/3}$, so that the radius swells by two orders of magnitude. In practice, such a large inflation never goes to completion. The expansion time may become so long that the star itself dies before equilibrium is achieved. More realistically, the ionization front reaches the boundary of the parent molecular cloud. That is, the HII region is *density bounded* rather than *ionization bounded*. The pressurized HII gas bursts out of the cloud into the surrounding medium, creating a *champagne flow*.

The simplest case to consider is when the star is located within one Strömgren radius from the cloud edge. We suppose that beyond this edge is a medium of much lower density and correspondingly higher temperature. The initial expansion of the HII region occurs just as before. However, once a portion of the ionized volume crosses the edge into the lower-density gas, equation (15.7) shows that the front velocity increases. This part of the HII region, therefore, fans out rapidly into a plume-like structure (see Figure 15.3). The density within the plume initially matches that of the external medium. Hence, there is a large pressure discontinuity within the ionized gas. A shock rushes outward, accelerating material behind it to velocities exceeding the sound speed. Simultaneously, this efflux of gas creates a rarefaction wave, which travels inward from the now disrupted cloud edge, back toward the star.

The solid curves in Figure 15.3 represent isodensity contours in a simulated champagne flow. In this particular calculation, the parent cloud is cylindrical; its boundary appears as two vertical lines. The original ionization front has nearly traversed the cylindrical radius, while the low-density portion of the front has spread far beyond this. Note how the outflowing, ionized gas reaches its maximum speed about midway to the expanded front. This gas consists of both material from the cloud and the original, unperturbed medium.

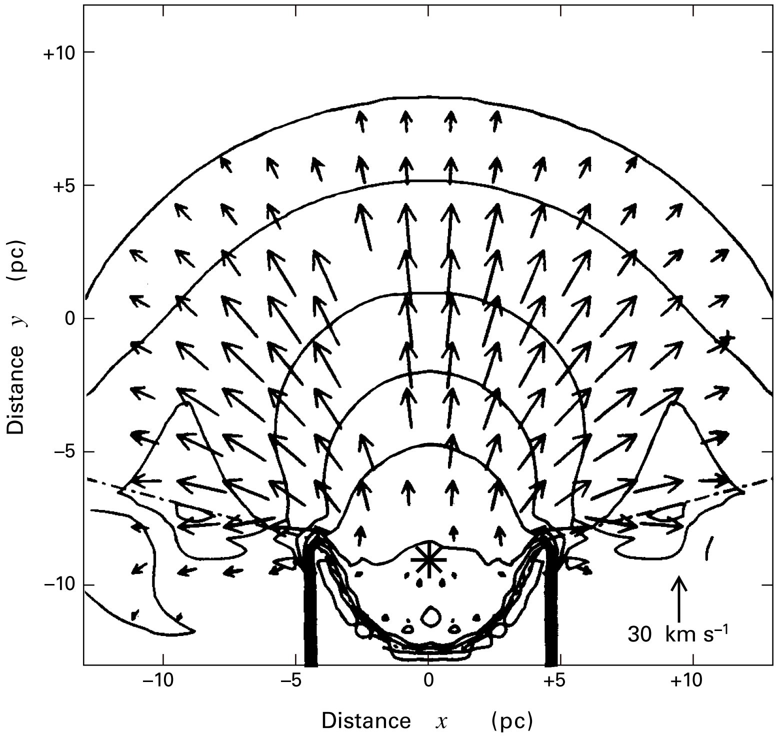


Figure 15.3 Numerical simulation of a champagne flow. The O star, with $\mathcal{N}_* = 8 \times 10^{48}$ s^{-1}, was originally located at the edge of a cylindrical cloud; the star's present location is indicated by an asterisk. Density contours and velocity vectors are shown 7×10^5 yr after the star turns on. The contour nearest the star corresponds to $\rho = 1 \times 10^{-22}$ g cm^{-3}. Successive contours fall by $\Delta\log\rho$ = -0.5 until the density levels off at $\rho = 3 \times 10^{-24}$ g cm^{-3}.

Figure 15.4 shows one observed case, the HII region S88B. The upper panel is an optical image taken in the Hα line at 6563 Å. The sharp boundary of the visible emission demarcates clearly the edge of the parent molecular cloud. The glowing gas has fanned out from this edge to the right. In this direction, the radial velocity of Hα increases beyond the cloud value by more than 7 km s^{-1}. Solid contours represent radio continuum emission at 6 cm. The hemisphere on the left is radiation from the cavity hollowed out of the cloud by the HII region; the emission here peaks at the star itself. Contours to the right of the edge again arise from the outward flow.

The lower panel of Figure 15.4 shows S88B as viewed in H92α, a radio recombination line of hydrogen. Such lines, like Hα itself, arise during the downward cascade following photoionization. In this case, the transition is between the $n = 93$ and $n = 92$ levels. The very high n-values mean that the resulting emission is in the radio regime, here 3.6 cm. Recombination lines, including also those from helium and carbon, are especially useful for mapping radial velocities. The isovelocity contours in the figure show how ionized matter picks up speed as it flows to

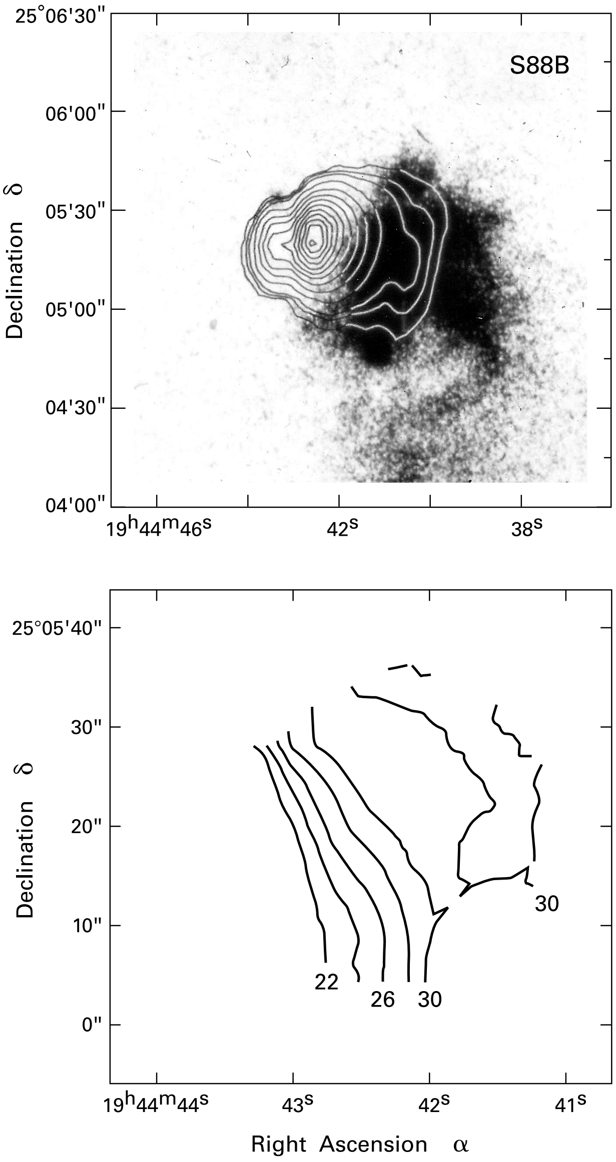


Figure 15.4 The HII region S88B. The upper panel is an Hα image, with 6 cm radio contours superimposed. Note how ionized gas flows to the right, away from the dense cloud. The contours in the lower panel show the radial velocity of the gas, as viewed by the hydrogen recombination line H92α at 3.6 cm. The velocity unit is km s^{-1}.

the right from the cloud, for which the radial velocity is 22 km s^{-1}. Other recombination line studies have effectively measured the distances to HII regions, by comparing their average radial velocities with the known pattern of Galactic rotation. The HII region map of Figure 2.3 was obtained by this means.

15.1.4 Emission Characteristics

The Hα line has traditionally been the principal tool for detecting optically visible HII regions. Radio continuum and line emission are also important as diagnostics of physical conditions, as we shall detail shortly. However, most of the region's thermal energy is not lost through these processes. Metallic ions of oxygen, nitrogen, and carbon provide the lion's share of the cooling. Even these most common ions are relatively low in abundance, but they all have metastable electronic states lying a few eV above ground. Impact by ambient electrons can readily generate forbidden line emission, the same cooling process we studied in Chapter 7.

The explanation of why this line emission is so strong is that, with an ambient temperature near 10^4 K, the typical electron within the nebula has an energy of about 1 eV, not much below that of the metastable transition itself. The cross section for excitation of the line is therefore relatively large, orders of magnitude above that for electron-proton recombination. For the same reason, emission during the recombination cascade of hydrogen, another cooling process one might think would dominate, is also not very significant.

All of the forbidden optical and ultraviolet lines stem from multiply ionized species, such as O II, O III, N II, and C IV. Conspicuously absent here are the neutral atoms that radiate from stellar jets. For stars with effective temperatures of about 4×10^4 K or less, the single strongest cooling line is the doublet [O II]$\lambda\lambda\,3726, 3729$. For hotter stars, the emission is primarily from [O III]$\lambda\lambda\,4959, 5007$. Note that the critical electron density for all these lines is 10^3 cm^{-3} or greater. Thus, at the lower densities encountered in visible HII regions, typically 10^2 cm^{-3}, each electron collision results in a downward, radiative transition.

It is straightforward to apply the results of Chapter 7 and derive a quantitative expression for the bulk cooling through forbidden line emission. Consider a case in which the [O II] lines dominate. Equation (7.24) gives the subcritical rate when the impact is due to ambient hydrogen. Hence, we should substitute n_e for n_H, but the two are nearly the same within the fully ionized nebula. The ratio of degeneracies g_u/g_l, is 1 for the 3726 Å line and 3/2 for 3729 Å, so we adopt an average value of 1.25. The transition energy is 3.4 eV, corresponding to a temperature of 3.9×10^4 K. Assuming that all of the oxygen is in the form of O II and using a fractional abundance relative to hydrogen of 7×10^{-4}, we find

$$\Lambda_{\mathrm{OII}} \,=\, 7 \times 10^{-7} \left(\frac{n_H}{10^2\ \mathrm{cm}^{-3}}\right)^2 \exp\left(-\frac{3.9 \times 10^4\ K}{T}\right)\ \mathrm{eV\ cm^{-3}\ s^{-1}}\ . \tag{15.17}$$

To maintain overall energy balance, this cooling from the HII region must offset heating from the central star. In fact, equating the two gives us the internal temperature of the nebula. Each photoionization of a hydrogen atom ejects an electron with some nonzero kinetic energy. Collisions between this electron and ambient species are the heating agent. If ΔE is the average electron energy, then the *total* heating rate throughout the nebula is just $\mathcal{N}_*\,\Delta E$.

Employing the condition of ionization balance, equation (15.2), we see that the *volumetric* rate is $n_H^2\,\alpha'_{\rm sec}(T)\,\Delta E$.

The energy of the ejected electron is that provided by the photon in excess of 13.6 eV. But the ionization cross section σ_ν falls off swiftly above this threshold. Thus, the extra energy is relatively small and depends on the spectral energy distribution of the star. Calculation reveals that $\Delta E \approx k_B\,T_{\rm eff}$, where $T_{\rm eff}$ is the stellar effective temperature. (Note that $k_B T_{\rm eff} \ll 13.6$ eV.) We thus derive an expression for the volumetric heating rate due to photoionization:

$$\Gamma_{\rm PI} \,=\, 9\times 10^{-9}\,\left(\frac{n_H}{10^2\ {\rm cm}^{-3}}\right)^2 \left(\frac{T_{\rm eff}}{4.1\times 10^4\ {\rm K}}\right)\ {\rm eV\ cm^{-3}\ s^{-1}}\ . \tag{15.18}$$

Here we have used the effective temperature of our canonical O6 star. Equating the heating and cooling rates, we find that the temperature of the HII region is 8900 K for this case, in good agreement with a more exact determination.

As we indicated earlier, HII regions also emit continuum radiation. Some of this is in the optical and represents light from the central star scattered by dust grains. Dusty cloud gas is conspicuous in photographs, where it partially blocks nebular light. Long, optically thick columns known as "elephant trunks" may protrude into the HII region from the outside. A spectacular example is M16 or the Eagle Nebula, shown as Figure 15.5. Most grains that are immersed within such ionized gas are promptly destroyed. An outer mantle of dusty gas, in which the grains are typically heated to some 30 K, often surrounds the HII region. The emission from this structure at mid- and far-infrared wavelengths has been another observational tool of value.

Within the radio regime, the recombination lines mentioned earlier are superposed on a broad continuum. The latter stems from free-free emission, *i. e.*, the radiation from thermally excited electrons accelerating in the vicinity of ambient protons. We have already encountered this process twice – in the jets close to embedded stars (Chapter 13) and in the X-ray production of pre-main-sequence objects (Chapter 7), where the mechanism is known as bremsstrahlung. As before, the emissivity j_ν is nearly independent of frequency up to the cutoff value, given by $k_B T/h$. Since $\nu \ll k_B T/h$ at centimeter wavelengths, j_ν indeed has a modest frequency dependence over the observed range.

A region generating free-free emission may become optically thick to its own radiation. It is a curious fact that this occurs at lower, rather than higher, frequencies. To see why, we recall from Kirkhoff's law governing thermal emission that the absorption coefficient, α_ν, is given by $j_\nu/B_\nu(T)$. (See equation (E.20), where α_ν is called $\rho\kappa_\nu$.) In the radio regime, the Rayleigh-Jeans approximation holds, and $B_\nu(T)$ is proportional to ν^2. (See equation (C.11) and below.) Since j_ν is nearly frequency-independent, α_ν varies as ν^{-2}. The frequency at which the HII region becomes optically thick often falls in the observable range. Below this value, the specific intensity I_ν is simply $B_\nu(T)$ and is thus proportional to ν^2. The observed slope provides a direct estimate for the temperature of the HII region. At higher frequencies, I_ν reverts to the optically thin form and actually declines slowly for increasing ν.

Within this optically thin regime, the flux observed must be proportional to the product of electron and proton densities, and also to the path length L, *i. e.*, the diameter of the HII region. In other words, the flux is proportional to $n_e^2\,L$, a quantity known as the *emission measure*. Here the constant of proportionality depends only on the temperature, which, we have seen, may be

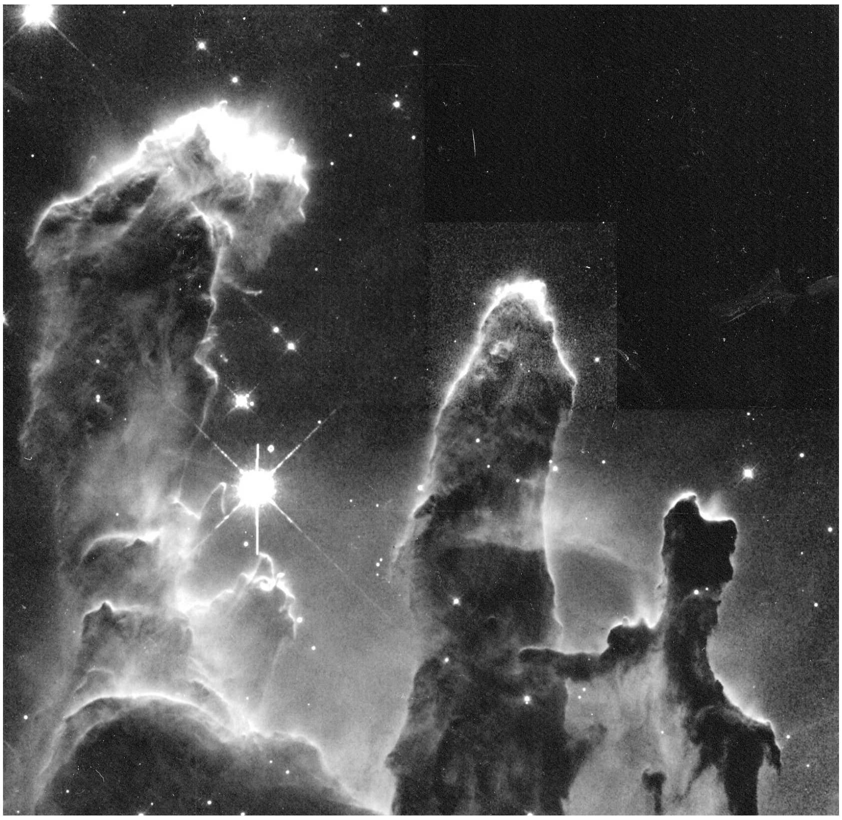


Figure 15.5 Elephant trunks in the HII region M16. The columns in this optical image represent dense gas that has not yet been dissipated by ultraviolet radiation. Note the glowing ionization fronts at the top of each structure.

extracted from the optically thick portion of the spectrum. If the HII region is spatially resolved, then one also obtains L, from which n_e follows.

15.2 Ultracompact HII Regions and Hot Cores

Optically visible, or *classical* HII regions, such as the Orion Nebula, have diameters of order 10^{18} cm or larger, and relatively low electron densities, $n_e \lesssim 10^4\ \mathrm{cm}^{-3}$. There is also considerable interest in their precursors – HII regions that are wholly invisible at optical wavelengths because of the high column density of surrounding gas and dust. The smallest of these, already introduced in Chapter 14, are dubbed *ultracompact*. These have diameters $L \lesssim 10^{17}$ cm and electron densities $n_e \gtrsim 10^5\ \mathrm{cm}^{-3}$. At far-infrared wavelengths, they are the most luminous objects in the Galaxy. This emission, as we have seen, stems from heated dust within the ionized gas.

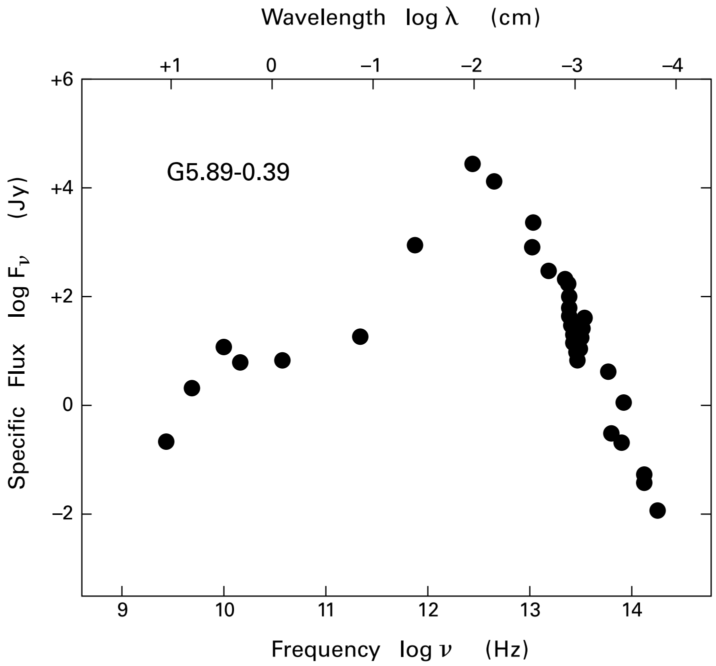


Figure 15.6 Spectral energy distribution of the ultracompact HII region G5.89-0.39. Note the turnover near $\lambda = 3$ cm and the sharp maximum near $\lambda = 100$ µm.

15.2.1 Morphology of the Radio Emission

Figure 15.6 is a combined radio and infrared spectrum of the ultracompact HII region G5.89−0.39, at a distance of 2 kpc. The ionized area has a diameter of 0.05 pc and is embedded within a molecular cloud of mass 30 $M_\odot$ and diameter 0.2 pc. Starting at the left, we see how the radio flux from free-free emission first climbs with frequency in the optically thick regime. After the turnover at a wavelength of about 3 cm, the flux remains nearly flat, as predicted by theory. The subsequent steep climb to the pronounced maximum near 100 µm represents the infrared contribution from dust. Note the sharp dip at 10 µm due to silicate absorption. The total luminosity, which stems almost entirely from the infrared contribution, is 3×10^5 $L_\odot$, indicating a central star of spectral type O6. Numerical modeling of the infrared spectral energy distribution suggests that the grains are distributed in a thick shell, with outer radius of about 0.5 pc and an inner hole with a radius smaller by an order of magnitude. The hole is necessary in the model to limit emission in the near infrared.

It is gratifying that this spatial structure, deduced entirely from the broadband spectrum, at least qualitatively matches that seen in the radio continuum. The upper left panel of Figure 15.7 is a map in 6 cm of the same region. Here we see clearly the shell-like morphology. Measurements spanning several years find the shell to be expanding at a velocity of about 30 km s^{-1}. Of the many ultracompact HII regions that have been spatially resolved at centimeter wavelengths, only about 5 percent have a similar structure. Roughly 25 percent exhibit a *core-halo* morphol-

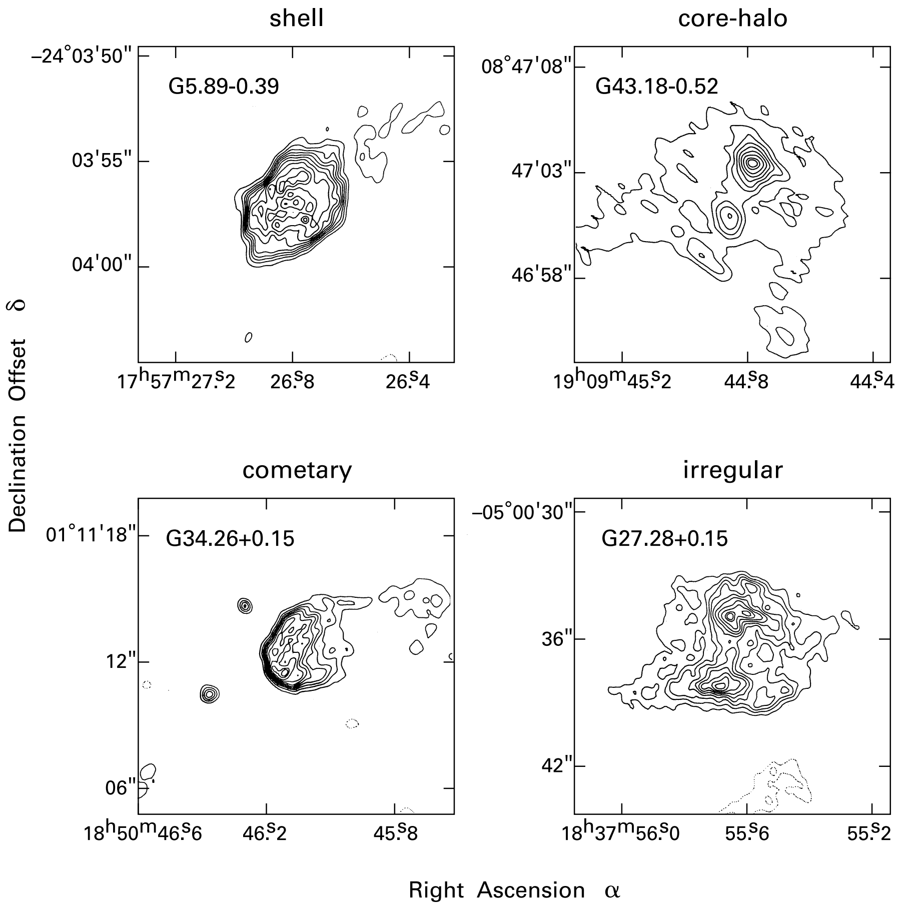


Figure 15.7 Morphology of ultracompact HII regions. Shown here are 6 cm radio continuum maps of representative cases. Most common are the cometary structures.

ogy, as seen by the example G43.18−0.52 in the upper right panel. The single largest group, about 30 percent of the total, are the *cometary* nebulae. We have already seen one example, NGC 6334F, in Chapter 14. The lower left panel in Figure 15.7 gives another, G34.26+0.15. The remainder of the ionized regions are classified as *irregular*. Some, like the example in the lower right panel, exhibit more than one peak in emission.

Radio interferometric maps, such as those in Figure 15.7, have allowed us to examine selected ultracompact HII regions in extraordinary detail. It is their powerful infrared emission, however, that signals their presence, even to the far ends of the Galaxy. In practice, one identifies such regions by their steeply rising spectrum from near- to mid-infrared wavelengths. With the possible exception of the double-peaked regions, each is presumably powered by a single O star. Surveying the Galaxy, we find that about 2000 O stars are in this deeply embedded phase. The figure represents some 10 percent of the *total* number of O stars. Most of these are found outside the confines of molecular clouds, although there is usually gas nearby.

This valuable comparison tells us that a typical O star spends about 10 percent of its main-sequence lifetime in the embedded phase. Since this lifetime is 4×10^6 yr, we conclude that an ultracompact HII region lasts for some 4×10^5 yr. The difficulty is then to understand their small sizes. If we assume that each region is in the second expansion phase, then the ionization front advances at roughly the internal sound speed, or about 10 km s^{-1}. To cover the maximum observed size of 0.1 pc, therefore, would require only about 10^3 yr. To put the matter another way, there are far too many ultracompact HII regions for their youthful dynamical state.

The problem is in part alleviated by the observation that a number of ultracompact HII regions reside within larger ionized structures. This broader emission, again seen in the radio continuum, may extend over a linear size of order 1 pc. The contribution to the total radio emission from this component dominates that from the central source, sometimes by as much as an order of magnitude. Such larger ionized volumes are, of course, more consistent with a longer lifetime for the system as a whole. However, it remains unclear how the compact, central region maintains its integrity over this period.

15.2.2 Stifling Ionization

One possibility is that there is a continuous supply of neutral matter being introduced to the region. Since the Lyman continuum photons are consumed through ionizing this fresh matter, the first expansion of the HII region is stalled. As an example, the O star may be surrounded by a massive disk. Photoevaporation of the disk, a process we shall later consider in more detail, would then eject material into the circumstellar environment.

The neutral matter may also enter through infall, *i. e.*, through the gravitational collapse of the parent cloud containing the nascent star. We discussed previously how the picture of inside-out collapse developed for low-mass stars may not be applicable to their massive counterparts. Additionally, halting the ionization through infall yields an unstable equilibrium, in which hot parcels of gas continue to escape. (Recall the discussion accompanying Figure 11.23.) It is interesting, nevertheless, that the mass accretion rate needed to (at least partially) stifle the HII region is not much greater than typical low-mass values.

Consider, then, an HII region within a spherical accretion flow. We recall from Chapter 10 that the density in such a region varies as $r^{-3/2}$. We may thus write for the electron density

$$n_e \,=\, n_* \,(r/R_*)^{-3/2} \ , \tag{15.19}$$

where n_* and R_* refer to the stellar radius. To gauge the size of the Strömgren sphere under these circumstances, we generalize equation (15.2) to read

$$\mathcal{N}_* \,=\, 4\,\pi\,\alpha'_{\rm rec}(T) \int_{R_*}^{R_s} n_e^2\, r^2\, dr \ , \tag{15.20}$$

from which we find

$$\mathcal{N}_* \,=\, 4\,\pi\,\alpha'_{\rm rec}(T)\, n_*^2\, R_*^3\, \ln\,(R_s/R_*) \ . \tag{15.21}$$

A slight modification of equation (10.34) gives us an expression for the mass infall rate:

$$\dot{M} \,=\, 4\,\pi\,\sqrt{2GM_*}\, m_H\, n_*\, R_*^{3/2} \ . \tag{15.22}$$

We may thus eliminate the product $n_*^2 R_*^3$ between equations (15.21) and (15.22). Solving the resulting expression for R_s, we find

$$R_s \,=\, R_* \,\exp\left[\frac{8\,\pi\, m_H^2\, \mathcal{N}_*\, G\, M_*}{\alpha'_{\rm rec}\, (\dot{M})^2}\right] \ . \tag{15.23}$$

In summary, there is a critical $\dot{M}$-value, above which R_s hovers close to the stellar radius. This is given by setting the argument of the exponential equal to unity:

$$\begin{aligned}\dot{M}_{\rm crit} \,&=\, \left[8\,\pi\, \mathcal{N}_*\, G\, M_*\, m_H^2\, (\alpha'_{\rm rec})^{-1}\right]^{1/2} \\ &=\, 6 \times 10^{-5}\, M_\odot\ {\rm yr}^{-1} \left(\frac{\mathcal{N}_*}{10^{49}\ {\rm s}^{-1}}\right)^{1/2} \left(\frac{M_*}{30\, M_\odot}\right)^{1/2} \ .\end{aligned} \tag{15.24}$$

The essence of the argument here is that increasing $\dot{M}$ raises the density within the HII region. Consequently, the volumetric recombination rate climbs. To maintain ionization balance with a fixed output of Lyman continuum photons, the entire volume must shrink. The actual dynamics of infall plays no direct role. Indeed, any increase of the *external* density will also limit the HII region's growth. Recall that the second expansion phase ends once the internal pressure falls to the ambient value. If this pressure is much greater than we have supposed, the endpoint will be reached at a correspondingly smaller value of R_s.

We shall see momentarily that the external density may in fact exceed our canonical $10^3\ {\rm cm}^{-3}$ by three orders of magnitude. The environmental temperature T_0 is also higher, typically 40 K. Using the density enhancement, equation (15.3) tells us that the Strömgren radius is reduced considerably, to only 1×10^{16} cm. To obtain the radius at the end of the second expansion, we rewrite equation (15.16) as

$$R_f \,=\, \left(\frac{2\,T_1}{T_0}\right)^{2/3} R_s \ . \tag{15.25}$$

The internal temperature T_1 is now closer to 5000 K, since the higher electron density gives greater cooling. We thus find that $R_f \,=\, 40\, R_s$, or about 0.1 pc. This figure is indeed within the observed range. The HII region still takes less than 10^4 yr to attain this radius, but is then presumably halted for an extended time.

15.2.3 Properties of the Densest Gas

The new assessment of physical conditions in the environments of the youngest massive stars has come from molecular line studies. Observers have utilized the high infrared luminosity of ultracompact HII regions as a signpost to guide their search for line emission. Water masers have also served this purpose. The detection of high-level rotational transitions in such species as ^{12}C^{34}S and CH$_3$CN attests to both elevated temperatures and densities.

From such studies, we know that many ultracompact HII regions are situated within molecular cloud clumps, roughly a parsec in diameter. The derived number density and temperature in such regions are close to the values just quoted, *i. e.*, $n_H^0 \sim 10^6\ {\rm cm}^{-3}$ and $T_0 \sim 40$ K. Note that

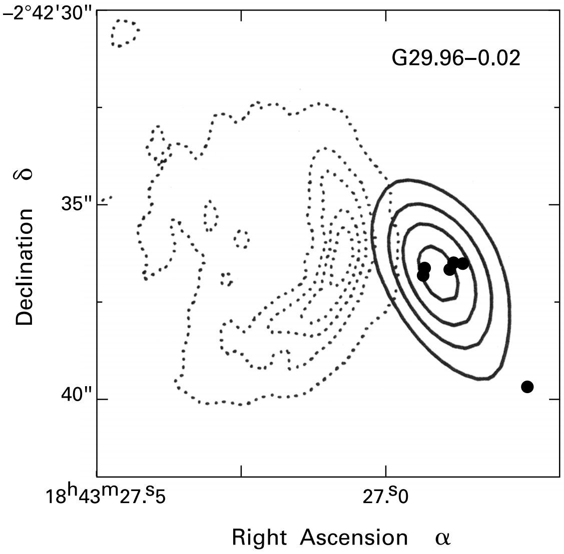


Figure 15.8 Hot core near the HII region G29.96-0.02. The dotted contours show 1.3 cm continuum emission from the HII region. Solid contours represent 2.7 mm line emission from a high rotational transition of CH_3CN. Filled circles are H_2O masers.

the density value is similar to that reconstructed for the gas surrounding the Trapezium cluster in Orion. It is becoming apparent, then, that massive stars are born into environments vastly different from those probed by CO measurements.

An even more dramatic illustration of this fact comes when we examine smaller length scales. The same molecular line observations reveal the existence of *hot cores*, compact, infrared-bright regions with densities of about 10^7 cm^{-3} and temperatures of 100 K or higher. Each hot core has a diameter of order 0.1 pc and a total mass of several hundred solar masses.

Given their means of discovery, it is not surprising that many of these objects are situated close to ultracompact HII regions. Figure 15.8 shows one example. Here, the HII region G29.96−0.02 appears to the left as the cometary structure seen in 1.3 cm emission. The hot core on the right is revealed through contours of CH_3CN ($J = 6 \rightarrow 5$) at 2.7 mm. Near the core's center is a cluster of H_2O masers, indicated by the filled circles. What is *not* seen, in this and numerous other cases, is conspicuous radio continuum emission within the molecular line contours. That is, many hot cores lack an internal HII region.

It is clear, on the other hand, that a hot core does contain its own massive star. Sensitive probes such as the inversion line of NH_3 (4,4) at 1.25 cm show that the gas temperature increases toward the center. In the example of Figure 15.8, the total infrared emission exceeds 10^5 $L_\odot$. The absence of a resolvable HII region again indicates that the ionization front has been stalled because of the extraordinary density. Finally, we notice, as we did in Chapter 14, that the cometary arc of the well-detected HII region directly faces the core, a clear sign that the ionization has difficulty penetrating this region.

The juxtaposition of two massive stars that we see in Figure 15.8 is not uncommon and illustrates the important fact that such entities are rarely born in isolation. In this case, the two stars are apparently at different evolutionary stages. That associated with the hot core is still surrounded by the dense gas that was present at its formation. Its neighbor, on the other hand, has had time to begin clearing its own remnant matter through ionization and heating, so that the average internal density has dropped by two orders of magnitude.

Investigation of hot cores through molecular lines has been aided by the distinctive chemical makeup of these regions. There is an abundance of hydrogenated species, such as H_2O, NH_3, CH_3OH, and H_2S. The enhancement of NH_3 is especially convenient, as it facilitates use of this valuable tracer, even at the higher-level inversion transitions. In addition to H_2S, sulfur-bearing molecules such as SO, SO_2, and CS have increased concentrations. Numerous additional species, including organics, are observed, and there are indications of differentiation in relative abundances between neighboring regions. The general pattern of molecular enrichment is reminiscent of that we encountered in the Orion BN-KL region (Chapter 5). Evaporation of grain mantles in these heated environments is almost certainly a major factor in their chemical history.

The H_2O masers associated with the cores act as tracers of wind-induced velocity. Figure 15.9 shows an interesting example, the infrared source IRAS 20126+4104. In the upper panel, the exciting star itself, which has a luminosity of about $10^4\ L_\odot$, is indicated by the central cross inside the contours of 3.6 cm emission. Three clumps of H_2O masers are also shown (*filled circles*). They span a total distance of some 2000 AU. The hot core itself is delineated by emission in CH_3CN $(J = 12 \rightarrow 11)$ at 1.3 mm, displayed here by the open circles. In this case, therefore, the jet erupts at an oblique angle with respect to the core. The bottom panel displays, on a larger scale, contours of $H^{13}C^{16}O^+$ $(J = 1 \rightarrow 0)$ at 3.46 mm, as well as the 2.12 µm emission from H_2. The contours delineate a bipolar structure extending over some 20″, or 0.2 pc at the 1.7 kpc distance of this source. The patches of excited H_2 are similarly aligned on an axis that must trace the internal jet.

Hot cores appear to represent the earliest environment of high-mass stars. However, to answer fully the question of how such stars arise, we shall eventually have to find gaseous structures that do not yet contain luminous, central sources. These presumably reside near the crowded centers of stellar clusters. The clouds in question may be the cold analogues of the presently known hot cores, *i. e.*, entities of similar density, but with relatively low temperatures. On the other hand, it may be that the formation proceeds so rapidly once the molecular density climbs, that one can only find less compact structures, perhaps already containing intermediate- or lower-mass stars. Molecular line studies of the densest regions within clusters should eventually clarify the issue.

15.3 Winds and Molecular Outflows

The foregoing example of a jet-like flow emanating from a young, massive star is hardly unique. On the other hand, one rarely finds strings of optically visible Herbig-Haro objects in these environments. The ambient density, and therefore the extinction, are simply too high, especially close to the star itself. We shall see in this section, nevertheless, that there is ample evidence for outflowing gas, both in the form of high-velocity winds and entrained cloud matter.

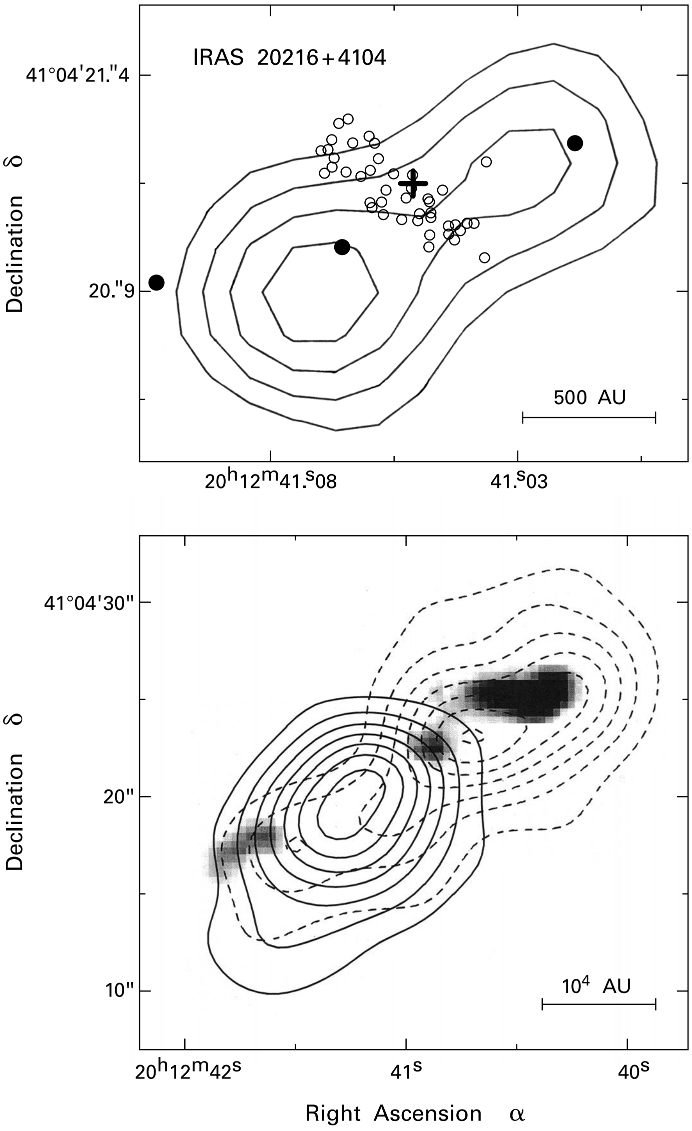


Figure 15.9 Hot core surrounding the embedded star IRAS 20216+4104. Within the upper panel, solid contours show 3.6 cm continuum emission, open circles are 1.3 mm line emission from CH_3CN, and filled circles are clumps of H_2O masers. The infrared star itself is shown by the cross. In the lower panel, solid and dashed contours are blue- and redshifted emission, respectively, at 3.46 mm from $H^{13}C^{16}O^+$. The greyscale patches show 2.12 µm emission from H_2. Note the different scales of the two panels.

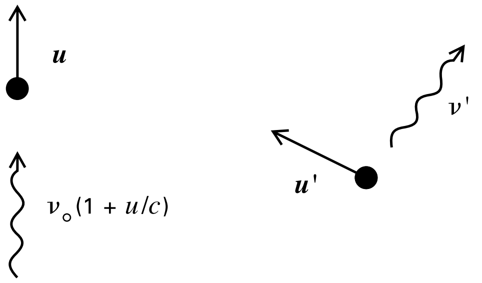


Figure 15.10 Resonant scattering of an ultraviolet photon from a moving ion. The photon initially has frequency $\nu_\circ(1 + u/c)$, where $h\nu_\circ$ is the transition energy and $\boldsymbol{u}$ is the ion velocity. After absorption and reemission, the new photon has frequency ν', while the ion has velocity $\boldsymbol{u}'$.

15.3.1 Radiative Acceleration

We discussed in Chapter 13 how the launching of winds from low-mass stars proceeds through the combined effect of centrifugal and magnetic forces. Massive stars also rotate, but are radiatively stable in their outer layers. Hence, they do not have the internal dynamo action needed to generate a magnetic field.[3] Instead, material is lifted off the surface and accelerated through radiation pressure exerted by ultraviolet photons. The momentum transfer is rather inefficient, in the sense that the wind momentum transport rate, $\dot{M}_w V_\infty$, falls below L_*/c. Let us now examine in more detail how the process occurs.

The atoms directly affected by the radiation are actually ions of the most common heavy elements, such as carbon, nitrogen, and oxygen. To impart momentum, a photon must first be absorbed. That is, its energy must coincide with that of an internal transition, which is an electronic one at ultraviolet frequencies. Especially important in this regard are *resonance* lines, in which the lower level is the ground state.

The full interaction is a scattering event. After absorbing the photon, the ion emits another in a random direction. Suppose the ion is initially moving radially outward and is struck by a photon moving in the same direction (Figure 15.10). Calculation shows that, after both absorption and reemission, the ion's outward momentum increases, on average, by $h\nu_\circ/c$. Here, $h\nu_\circ$ is the transition energy. It is important to realize that $\nu_\circ$ is *not* the frequency of the incident photon as seen by an outside observer. This is $\nu_\circ(1 + u/c)$, where u is the wind (and therefore ion) velocity. The ion sees the photosphere recede, so that the photon frequency becomes redshifted back to $\nu_\circ$. The star actually emits few photons at $\nu_\circ$ itself, because of prior absorption within deeper layers. Thus, the frequency shift exposes the ion to a larger flux and increases the acceleration.

The wind velocity $u(r)$ is some monotonic function of radius, starting near zero at the stellar surface, quickly becoming supersonic, and eventually reaching the terminal value V_∞. At each position r, the condition

$$\nu = \nu_\circ\,[1 + u(r)/c] \tag{15.26}$$

tells us which frequency ν emanating from the photosphere interacts with the ion through a transition of energy $h\nu_\circ$. Having gained the momentum $h\nu_\circ/c$, the ion then shares this added momentum with neighboring ions through mutual collisions. The species that provide most

---

[3] Nevertheless, some radiative stars are magnetic. A subclass of A-type stars, designated Ap, exhibit strong surface fields and anomalous chemical abundances. (The "p" in Ap stands for "peculiar.") The magnetic field in this case may be fossil, *i. e.*, trapped in stellar material during the formation epoch.

of the acceleration typically constitute only about 10^{-4} by mass of all the wind gas (recall Table 2.1).

Equation (15.26) assumes implicitly that the line in question has negligible width. In fact, there is a finite probability for absorption even if ν and $\nu_\circ$ do not obey the relation precisely. This probability is the familiar profile function $\phi(\nu)$, which governs the frequency dependence of the line opacity κ_ν (see, *e. g.*, equation (C.7)). Using the normalization of ϕ, we may write

$$\kappa_\nu \,=\, \phi(\nu) \int_0^\infty \kappa_{\nu'}\, d\nu' \ . \tag{15.27}$$

Note that $\phi(\nu)$ does not peak at $\nu = \nu_\circ$, but at $\Delta\nu \,=\, 0$, where $\Delta\nu \,\equiv\, \nu - \nu_\circ[1 + u(r)/c]$.

We now wish to determine the force due to a single line. Consider first the associated optical thickness, from the photosphere out to a radius r. This is

$$\begin{aligned} \tau_\nu(r) \,&=\, \int_{R_*}^{r} \rho(r')\,\kappa_\nu\, dr' \\ &=\, \int_0^\infty \kappa_{\nu'}\, d\nu' \int_{R_*}^{r} \rho(r')\,\phi(\nu)\, dr' \ . \end{aligned}$$

Within the second integral of the last line, we change independent variables from r to $\Delta\nu$ and find

$$\tau_\nu(r) \,=\, \int_0^\infty \kappa_{\nu'}\, d\nu' \int_{\Delta\nu_*}^{\Delta\nu} \rho(r')\,\phi(\Delta\nu') \left(\frac{\partial\Delta\nu'}{\partial r'}\right)^{-1} d\Delta\nu' \ ,$$

where $\Delta\nu_* \,\equiv\, \Delta\nu(\nu, R_*)$, and where r' is now a function of ν and $\Delta\nu'$. Evaluating the partial derivative, we have

$$\tau_\nu(r) \,=\, \frac{c}{\nu_\circ}\,\rho(r_1) \left(\frac{du}{dr}\right)_{r_1}^{-1} \Phi(\Delta\nu) \int_0^\infty \kappa_{\nu'}\, d\nu' \ . \tag{15.28}$$

Here, r_1 is the radius for which, at a given ν, $\Delta\nu$ vanishes. We have used the fact that ϕ sharply peaks when $\Delta\nu' = 0$ to extract from the integral the product of density and velocity gradient. The function $\Phi(\Delta\nu)$ in (15.28) is

$$\Phi(\Delta\nu) \,\equiv\, \int_{\Delta\nu}^{\Delta\nu_*} \phi(\Delta\nu')\, d\Delta\nu' \ , \tag{15.29}$$

and is positive since $\Delta\nu < \Delta\nu_*$ at fixed ν. For all frequencies of interest, the point r_1 is situated outside the stellar surface. That is, $\Delta\nu_*$ is positive. Then Φ is unity for negative values of $\Delta\nu$, zero for positive values, and makes a rather abrupt transition near $\Delta\nu \,=\, 0$. In summary, we may write

$$\tau_\nu(r) \,=\, \tau_1(\nu)\ \Phi(\Delta\nu) \ , \tag{15.30}$$

where

$$\tau_1(\nu) \,\equiv\, \frac{c}{\nu_\circ}\,\rho(r_1) \left(\frac{du}{dr}\right)_{r_1}^{-1} \int_0^\infty \kappa_{\nu'}\, d\nu' \ . \tag{15.31}$$

To calculate the force per unit mass at a given radius, we must account for the reduction of the flux $F_\nu(r)$ because of the optical thickness. Let $F_\circ(r)$ denote the unattenuated continuum flux in the neighborhood of $\nu_\circ$. Then, as in our previous treatment of H_2 dissociation (Chapter 8), we assume a simple exponential law:

$$F_\nu(r) \;=\; F_\circ(r)\exp[-\tau_\nu(r)]\ . \tag{15.32}$$

In any volume $\mathcal{V}$ of the gas, the total rate of *energy* absorption per unit frequency is $\rho\kappa_\nu F_\nu(r)\mathcal{V}$. From our previous discussion, we multiply by $1/c$ to obtain the corresponding rate of *momentum* deposition. Dividing by the mass $\rho\mathcal{V}$ and integrating over frequency yields the force:

$$f_{\rm rad} \;=\; \frac{1}{c}\int_0^\infty \kappa_\nu\, F_\nu(r)\, d\nu\ . \tag{15.33}$$

We now utilize equations (15.27), (15.30), and (15.32) within the integrand of (15.33). We find

$$\begin{aligned} f_{\rm rad} \;&=\; \frac{1}{c}\,F_\circ(r)\int_0^\infty \kappa_{\nu'}\,d\nu' \int_0^\infty d\nu''\,\phi(\nu'')\exp[-\tau_{\nu''}(r)] \\ &=\; \frac{1}{c}\,F_\circ(r)\int_0^\infty \kappa_{\nu'}\,d\nu' \int_0^\infty d\nu''\,\phi(\nu'')\exp\left[-\tau_1(\nu'')\Phi(\Delta\nu'')\right]\ . \end{aligned}$$

Since ϕ within the integral again peaks at $\Delta\nu'' = 0$, we may evaluate τ_1 at r itself. We then use equation (15.29) to obtain

$$\begin{aligned} f_{\rm rad} \;&=\; \frac{1}{c}\,F_\circ(r)\int_0^\infty \kappa_{\nu'}\,d\nu' \int_0^1 d\Phi\,\exp\left[-\tau_1\,\Phi\right] \\ &=\; \frac{1}{c}\,F_\circ(r)\int_0^\infty \kappa_{\nu'}\,d\nu'\ \frac{1-\exp(-\tau_1)}{\tau_1}\ . \end{aligned} \tag{15.34}$$

If the line is thermally broadened, then $\int_0^\infty \kappa_{\nu'}\,d\nu' \approx \kappa_\circ\,\Delta\nu_D$, where $\Delta\nu_D$ is the Doppler width and $\kappa_\circ$ is the opacity at line center. We then have

$$f_{\rm rad} \;\approx\; \frac{1}{c}\,\kappa_\circ\,F_\circ(r)\,\Delta\nu_D\ \frac{1-\exp(-\tau_1)}{\tau_1} \tag{15.35a}$$

$$\tau_1 \;=\; \kappa_\circ\,\rho\,\Delta V_{\rm therm}\left(\frac{du}{dr}\right)^{-1}\ . \tag{15.35b}$$

Here, $\Delta V_{\rm therm}$ is the one-dimensional velocity dispersion associated with thermal motion (see equation (E.24)).

Our derivation has assumed that the scattered ultraviolet photon promptly escapes from the wind region. This is a reasonable supposition, since the relative motion between neighboring ions suppresses further absorption. However, there is a finite probability that the photon is absorbed by another ion in a different spectral line. Including such multiple scattering enhances the force by at most a factor of two. In a full calculation of wind acceleration, one must sum the contributions from several thousand lines. One then adds the total $f_{\rm rad}$ to the right side of the momentum balance equation (13.11), where this equation applies within the isothermal approximation.

The solution to the equation of motion exhibits the same general character we saw in the purely thermal wind. That is, arbitrary choices of the initial velocity at the stellar surface cause $u(r)$ either to remain small and decline asymptotically or to have a divergent gradient at some radius. Only a single choice leads to a velocity that continuously and monotonically increases. For this to occur, the flow must pass smoothly through a critical point analogous to R_s in equation (13.14). The actual speed is well above a_T at this location, since the radiative force exceeds the thermal pressure gradient.

The total radiative force depends on the optical thickness τ_1, where the latter incorporates a generalization of the opacity $\kappa_\circ$. Since τ_1 is proportional to ρ, the critical solution also dictates both the density variation of the wind and the mass transport rate, $\dot{M}_w$. In the thermal case, the force per unit mass is independent of the density, and the equation of motion yields only $u(r)$ (recall equation (13.13)). One finds the density by first specifying it at the coronal base and then invoking mass continuity.

A final technical note is of interest. The density ρ varies as $1/(r^2u)$ for a steady-state wind. From equation (15.35b), τ_1 then depends on the combination $(r^2u)\ du/dr$. But $f_{\rm rad}$ is proportional to $F_\circ(r)$, which falls off as r^{-2}. The stellar gravity declines in the same way. Once the radiative force exceeds that from thermal pressure, we see that $(r^2u)\ du/dr$ depends on τ_1, which in turn is a function of the same quantity. The only possible solution is that $(r^2u)\ du/dr$ is a spatial constant. Integrating, we find that the wind speed varies as

$$u(r) = V_\infty \sqrt{1 - \frac{R_*}{r}}\ , \tag{15.36}$$

in the supersonic regime.

15.3.2 Optical and Radio Emission

Observationally, the presence of winds near massive stars is revealed by optical and ultraviolet emission lines. Analysis of the line profiles gives not only the terminal velocity V_∞, but also the mass transport rate $\dot{M}_w$. Derived values of these quantities for typical O or early B-type stars are 2000 km s^{-1} and $10^{-6}\ M_\odot$ yr^{-1}, respectively, with substantial variation according to spectral type. The ultraviolet emission lines stem from multiply ionized species, confirming that the wind temperature is high, roughly equal to that of the stellar photosphere, or about 10^5 K.

Such observations concern relatively mature stars that are not cloaked in high column densities of dusty gas. The same objects sometimes display excess continuum radiation at infrared and radio wavelengths. In more embedded stars, this emission would stem from the associated HII region. Nevertheless, it is instructive to consider the longer-wavelength contribution from exposed stars, to see how it signals the existence of an ionized wind.

Figure 15.11 shows the radio spectrum of ζ Puppis, an O4 star with a luminosity of $8 \times 10^5\ L_\odot$ and an estimated mass of 70 $M_\odot$. This object is located in the Gum Nebula, about 450 pc distant. Note again that the radio fluxes shown in the figure represent emission above that from the normal photospheric values. The luminosity and effective temperature of

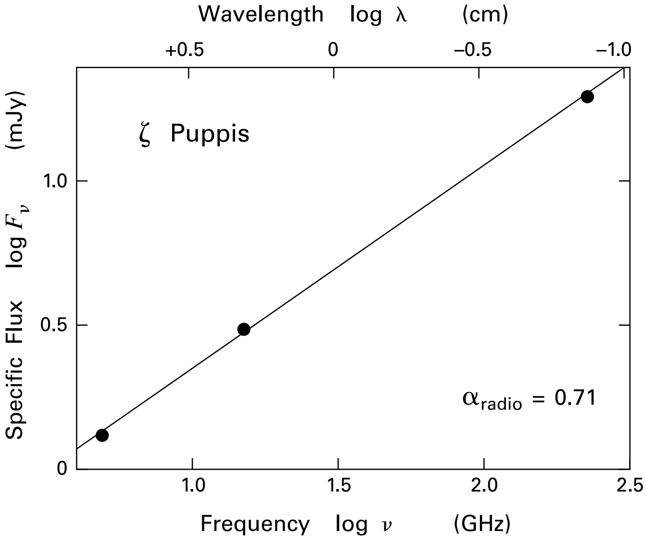


Figure 15.11 Excess radio emission from the supergiant ζ Pup. The straight line with the indicated slope is a best fit to the three data points.

ζ Pup place it close to, but not precisely on, the zero-age main sequence. The star is actually a more evolved supergiant. Such objects undergo considerably greater mass loss than their main-sequence counterparts, facilitating detection of the free-free emission. In this case, the relatively few available measurements indicate a slope of $\alpha_{\rm radio} \equiv d\log F_\nu / d\log\nu = 0.71$. This value differs from that expected from an HII region, either in the optically thick ($\alpha_{\rm radio} = 2$) or optically thin ($\alpha_{\rm radio} \lesssim 0$) regimes.

How does an ionized wind explain the altered slope? The key difference between a wind and an HII region in terms of their emission properties is the steep density falloff in outflowing gas. Suppose that the radio flux comes from a portion of the flow where the gas has already attained its terminal velocity. Then, in steady state, the density ρ declines as r^{-2}. If we further assume that we are observing the optically thick (*i. e.*, long-wavelength) portion of the spectrum, then the rapid density variation implies that different frequencies emanate from substantially different locations within the flow.

To quantify this idea, let R_ν be the effective radius producing photons of frequency ν. This radius is located where the optical depth, integrated from the outside, reaches a value of order unity. The luminosity at the same frequency, L_ν, is then given by

$$L_\nu = 4\pi R_\nu^2 F_\nu \, . \tag{15.37}$$

Here, the specific flux F_ν is $\pi B_\nu(T)$ evaluated at the radiating surface. (See equation (2.34) for the analogous relation involving F_λ.) Since the temperature gradient is small, and since B_ν is proportional to ν^2 in the Rayleigh-Jeans approximation, we see that L_ν varies as $R_\nu^2\,\nu^2$.

It remains to determine how R_ν changes with frequency. We have already seen, arguing from Kirkhoff's law, that α_ν for free-free emission varies as ν^{-2} and is also proportional to the product of electron and ion densities, *i. e.*, to ρ^2 in a fully ionized medium. Given the spatial variation of density, we find that $\alpha_\nu \propto \nu^{-2}\, r^{-4}$. The integrated optical depth thus varies as $\nu^{-2}\, r^{-3}$. We conclude that the value of R_ν decreases with rising frequency, as $\nu^{-2/3}$. Using this result in equation (15.37), we see that $L_\nu \propto \nu^{-4/3}\, \nu^2 = \nu^{2/3}$. A slope of 0.67 for the spectral energy distribution is in reasonable agreement with the empirical findings, such as those in Figure 15.11.

15.3.3 X-Ray Production

Satellite observations of the 1970s found that the brightest optically visible stars also emit in X-rays. The total luminosity in this component is relatively small, typically 10^{-7} of the $L_{\rm bol}$-value. Nevertheless, the inferred temperature of the emitting plasma is so high, of order 10^7 K, that some very energetic process is called for. It is natural to invoke the stellar wind. The temperature of the wind gas is much too low, but its kinetic energy content is interestingly high, provided it can be tapped through shock dissipation. According to equation (8.50), a fully ionized wind ($\mu = 0.61$) traveling at 2000 km s^{-1} produces, upon shocking, a temperature of 5×10^7 K.

The top panel of Figure 15.12 shows the broadband X-ray spectrum of ζ Pup, as observed by ROSAT. While a range of energies is present, the distribution peaks near 1 keV, *i. e.*, in the soft X-ray regime. The solid curve in the figure is a theoretical result, obtained by assuming that a dilute component of temperature 6×10^6 K is mixed into the relatively cold wind gas. If shocks indeed create the X-ray emission, then the observations show they must vary in time, and in a periodic manner. The bottom panel of the figure displays this variation in detail. Here the X-ray intensity within the indicated energy range fluctuates with 6 percent amplitude over a period of 16 h. The emission in Hα also varies, with identical period but a different phase.

What creates the shocks? Part of the answer is that radiatively accelerated winds are intrinsically unstable. That is, any small perturbation in the flow tends to amplify. Once a disturbance becomes sufficiently large, it shocks and produces the high temperatures and X-ray emission.

To understand the nature of the instability, let us idealize the wind as being driven by a single line which is optically thick. Assuming $\tau_1 \gg 1$ in equation (15.35a), we see that the radiative force is proportional to the velocity gradient du/dr. Figure 15.13 illustrates why the wind is unstable in this circumstance. In the lefthand panel, we have superposed a small perturbation on the underlying, steady-state velocity field, which itself is represented as the dashed curve. Material at point A has an abnormally high velocity and velocity gradient. After a time interval Δt, it thus travels farther *and* has an even greater velocity relative to the corresponding fluid element in steady state. Conversely, point B starts out with a smaller velocity and du/dr-value. It therefore travels less distance in Δt and its velocity deficit increases. Finally, the element at point C is accelerating faster, but begins at a lower speed. This element also travels less distance, but ends with a velocity closer to the steady-state value. The net result at time $t + \Delta t$, as sketched in the righthand panel, is that the velocity perturbation grows.

Wind instability by itself, however, cannot be the whole story. If it were, then even random initial perturbations would amplify and shock. We have seen, however, that both the Hα and X-ray fluxes of ζ Pup vary periodically. Similar results apply to θ^1 Ori C, the brightest star of the Trapezium. Optical, ultraviolet, and X-ray fluxes all vary, but this time with a 15-day period. Circular polarization of the optical light indicates the presence of a magnetic field. This

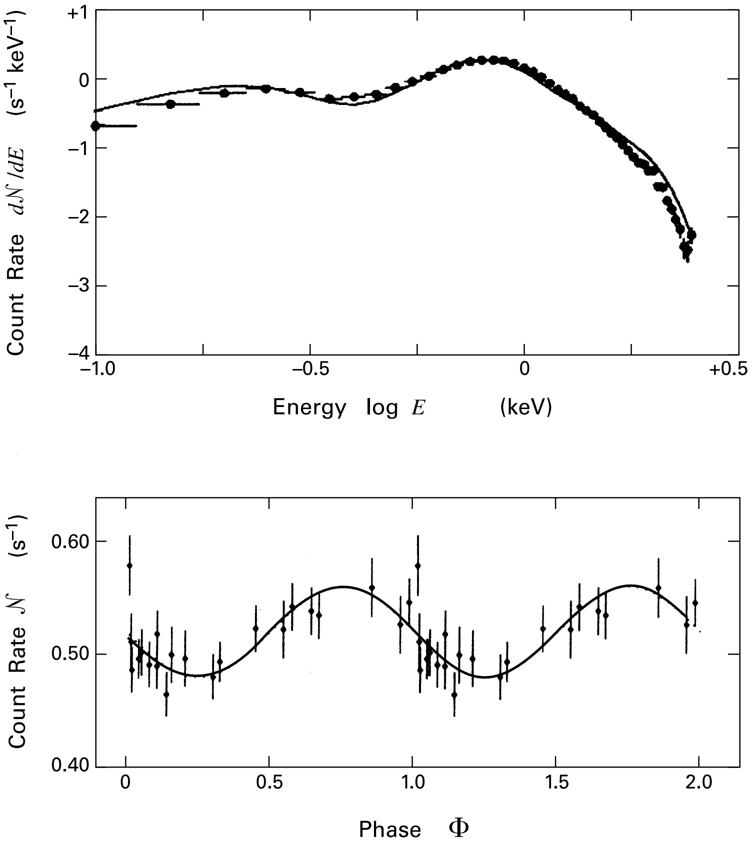


Figure 15.12 *Upper panel:* Observed X-ray emission from ζ Pup. Shown is the detector count rate per unit energy as a function of that energy. The solid curve shows the theoretical result for a gas at a temperature of 6×10^6 K. *Lower panel:* Temporal variation of the X-ray emission from 0.9 to 2.0 keV. A phase interval of unity corresponds to a time of 16 h.

field may be channeling the wind into some anisotropic pattern and generating shocks through the collision of gas streams. Radiating shocks would then sweep past the observer with each rotation of the star.

15.3.4 Massive Outflows

Let us now return to the younger stars of more direct interest. There is every reason to believe that these emit the same powerful winds as their more exposed counterparts. The reason is, first, that one does occasionally find both H_2O masers and Herbig-Haro objects associated with the star, indicating directly a shocked and collimated wind. Second, we observe cloud gas

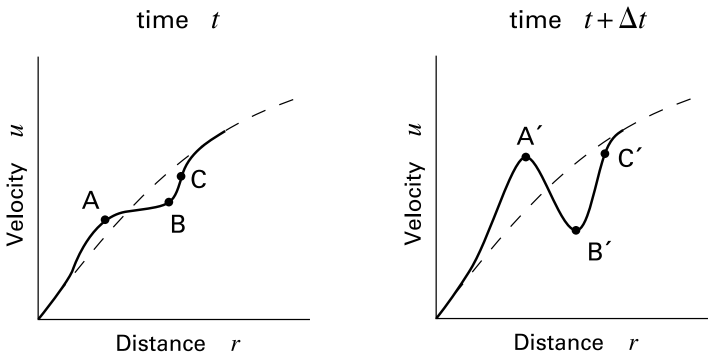


Figure 15.13 Instability of a radiatively driven wind. The dashed curve in both panels represents the unperturbed, steady-state velocity profile. At some initial time t (*left panel*), we apply a perturbation to this profile. By the later time $t + \Delta t$ (*right panel*), the velocity profile is more distorted, with the fluid element at A moving to A', *etc.*

surrounding the embedded object that has been stirred up and dragged outward as a molecular outflow.

We already saw, in the case of IRAS 20126+4104 (Figure 15.9) how a luminous star may excite molecular-line emission over a substantial distance. In this case, the $H^{13}C^{16}O^{+}$ $(J = 1 \rightarrow 0)$ flux displays the bipolar pattern familiar from low-mass examples. The alignment of the outflow axis with centrally located H_2O masers reinforces the notion that the cloud gas is entrained by an interior jet.

Given the large distances and heavy obscuration of young, massive stars, it is not surprising that optical manifestations of such jets are hard to find, at least close to the star. On the other hand, there a few examples of shock emission extending well beyond the molecular-line contours. Figure 15.14 is a composite image of the G192.16-3.82 system. The optical photograph at the top, taken in [S II], shows distended, irregular Herbig-Haro objects lying on either side of the central nebulosity. From end to end, the complex of emission knots spans over 10 pc. The central region, magnified in the lower inset, contains a highly structured and bipolar CO outflow. Near its intense peak lies a dense core and an even smaller region of heated dust, seen here through 2.6 mm continuum flux.

The central star is an infrared source of total luminosity $3 \times 10^3\ L_\odot$. Assuming the object to be on the main sequence, this figure identifies it as having a spectral type between B2 and B3, and a mass of 9 $M_\odot$. Such a star cannot create widespread ionization, although there is a small, central region of radio continuum emission. The system of optical knots naturally reminds us of the giant Herbig-Haro flows driven by low-mass stars (§ 13.1). In both cases, the underlying wind penetrates well beyond the parent molecular cloud gas into relatively empty space. A difference is that the shocked regions projecting from the B star are less collimated. Thus, the jet itself is either intrinsically broader, or else creates a greater lateral spray of shocked gas.

This observation is systematic of a larger trend. Not only do B stars drive wider jets, but objects of even greater mass have winds that seem to resist collimation altogether. No source with luminosity $10^5\ L_\odot$ or greater (*i. e.*, with $M_* \gtrsim 20\ M_\odot$) is known to drive a true optical jet. Of course, one must be cautious in drawing conclusions from the relatively small number of

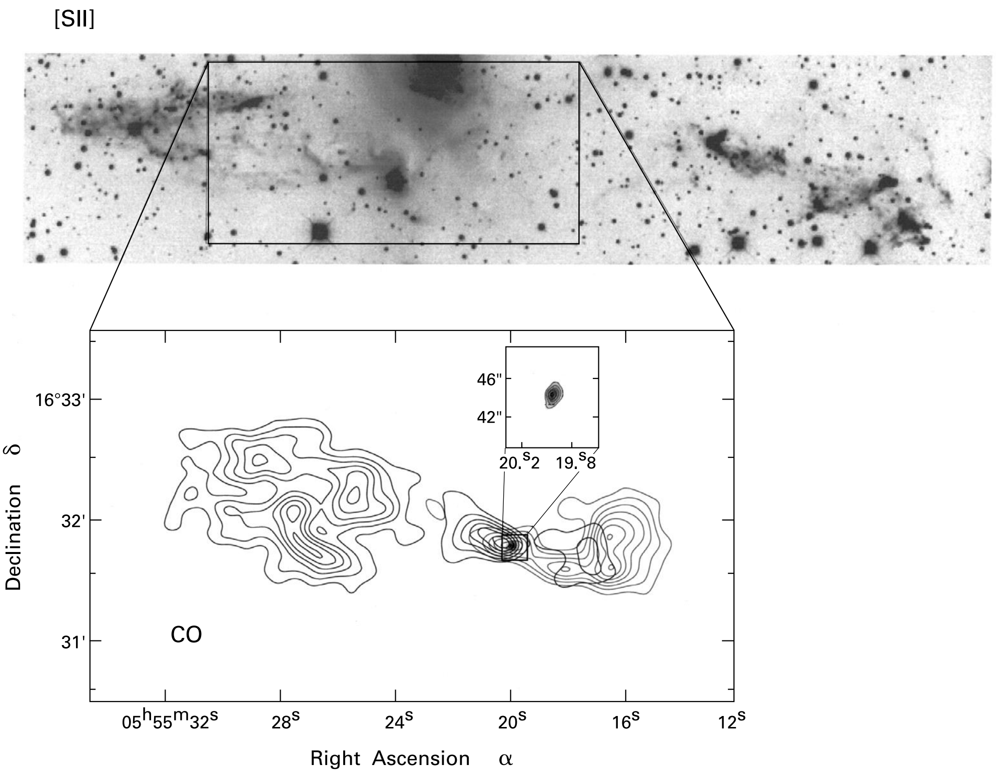


Figure 15.14 *Upper panel:* Optical photograph in [S II] of the massive outflow G192.16-3.82. Notice the Herbig-Haro knots on either side of the box. *Lower panel:* Molecular outflow within the central region. The inset shows a 2.6 mm continuum image of a dense core lying near the peak CO intensity.

embedded O stars. The OMC-1 outflow, nevertheless, continues to provide an example of what may happen generally in this mass regime. Optical radiation, along with shock emission from [Fe II] and NH_3, occurs at the tips of the many elongated "fingers" protruding from the vicinity of IRc2 (recall Plate 13). While the wind is uncollimated in a global sense, it is assuredly not isotropic and homogeneous.

Although optical jets disappear, molecular outflows persist even to the highest stellar masses. Here, too, there is a tendency toward lower collimation. Again we recall OMC-1, where the blue- and redshifted lobes in CO almost completely overlap. That this result is not a simple projection effect is attested by other examples. The aspect ratio of the outflows, including the lowest-velocity CO, rarely exceeds 2:1. In particular, we do not find analogues of the very narrow outflows driven by Class 0 sources.

The outflows from embedded, luminous sources have systematically more mass than those driven by solar-type stars. Thus, the lobes of G192.16-3.82 broadly resemble those of L1551/IRS 5 (Figure 1.13), but contain about 100 $M_\odot$, a figure larger by two orders of magnitude. The DR 21 outflow in Cygnus has a mass exceeding $10^3\ M_\odot$. For all systems, if we divide

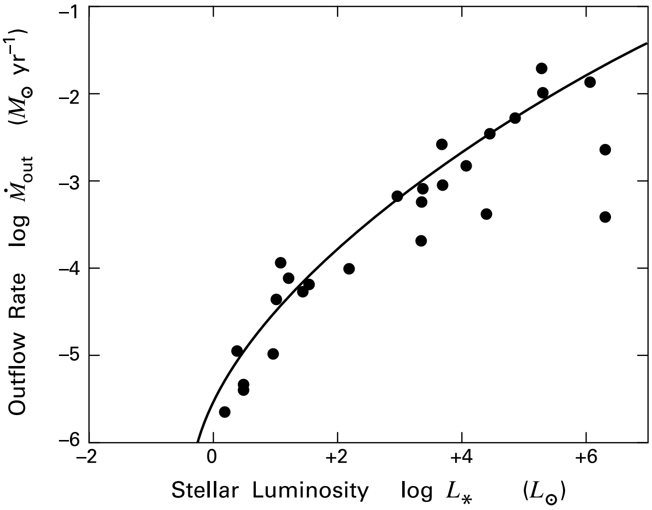


Figure 15.15 Mass transport rate in molecular outflows, as a function of the luminosity of the driving star. The smooth curve is an approximate fit to the data.

the estimated mass by the dynamical time $t_{\rm dyn}$, we obtain a measure of $\dot{M}_{\rm out}$, the net outward transport rate of mass. Figure 15.15 shows $\dot{M}_{\rm out}$, plotted as a function of the stellar luminosity. While the transport rates are indeed higher for the brighter (and hence more massive) central stars, the values continue a smooth rise that was already evident in the lower-mass systems.

In part, the increases in both the outflow mass and $\dot{M}_{\rm out}$ may be due to the wider dispersion of the underlying wind. Another factor must be the higher speeds of these winds, as seen in the proper motions of associated Herbig-Haro objects, which can exceed 1000 km s^{-1}. Further, indirect, evidence comes from the greater V_∞-values measured for optically exposed O and B stars. Within a given outflow, the distribution of mass with speed obeys the falloff documented in Figure 13.13, but exhibits a much steeper (negative) slope at higher velocities. That is, the exponent γ in equation (13.7) rises above 2. All these developments suggest that the basic interaction between wind and cloud may be qualitatively different.

There is at least one hint from theory why such a change occurs. Underlying the picture of wind collimation through crossing shocks (Figure 13.22) is the assumption that postshock cooling is relatively efficient. Thus, much of the wind's *energy* is radiated away, while its *momentum* in the jet direction is preserved. The faster winds from massive stars, however, give rise to higher postshock temperatures. Here, cooling becomes more difficult. The radiation no longer stems from forbidden transitions, such as [O II]$\lambda\lambda\,3726, 3729$, but from resonance lines within higher ionization states. The Einstein A-values are now much greater, and it is harder to maintain an excited population. For shock speeds exceeding 2000 km s^{-1}, the only cooling is from the relatively weak bremsstrahlung associated with free electrons.

The net result of diminished cooling is that massive outflows tend to be energy-conserving rather than just momentum-conserving. That is, shocking of the wind powers a hot, postshock region that itself expands outward by thermal pressure. This expansion should create a second, outer shock and compressed outer layer, just as in an overpressured HII region (Figure 15.2). However, massive outflows, like their low-mass analogues, are filled with turbulent cloud gas, and the majority are still bipolar. Even in the case of OMC-1, the complex morphology shows that a spherical picture is not applicable in any detail. Nevertheless, the basic idea that sufficiently fast winds create bubbles of hot gas should still hold. Perhaps the fingers of the OMC-1 flow result from such material erupting at numerous locations into the surrounding, ionized medium.

15.4 Photoevaporation of Gas

Ionization of the parent cloud surrounding a newborn massive star creates added pressure that tends to drive away the gas. A similar expansive effect also occurs for more compact material near exposed, luminous sources. Thus, the lifetime of a circumstellar disk around a massive star is limited by its eventual photoevaporation. The same holds for disks associated with low-mass stars that happen to be near one producing ionizing radiation. Indeed, any discrete clump, whether it contains its own star or not, will begin to dissipate once engulfed by an HII region.

15.4.1 Disk Irradiation

The destruction of a disk around an O or B star is a rather hypothetical situation, as there is still little evidence for such entities. When introducing embedded clusters in Chapter 4, we cited the case of S106, a bipolar HII region within a massive cloud. Here, two ionized lobes are separated by a ridge of extinction (Plate 2). While it is tempting to identify this feature as a disk, the associated thickness would be enormous, some 10^{16} cm. A much smaller system with a similar morphology is MWC 349 in Cygnus. The illuminating, B-type star has a heavily reddened optical spectrum that is rich with emission lines. A radio continuum map at 2 cm (Figure 15.16) shows two distinct lobes and a pinched waistband, reminiscent of S106. Indeed, the latter also has an hourglass shape in this wavelength regime.

The appearance of MWC 349 again seems to call for a planar configuration of dusty gas. Note, however, that radio emission can penetrate dust without suffering extinction. The morphology of Figure 15.16 indicates rather a lack of ionizing gas in the equatorial region. A large column of dust would indeed absorb stellar photons and thus halt the advance of ionization. The enhanced absorption could arise from a slab or disk, a view that has been bolstered by the detection of near-infrared radiation from this central region. Note also that an ionized wind is present in the extended lobes, since the integrated radio flux has a spectral index $\alpha_{\rm radio}$ of 0.6, close to that predicted by theory. The wind velocity, gauged from the width of emission lines, is surprisingly low, only 50 km s^{-1}.

These examples refer to exceptional objects whose interpretation is not yet clear. Pending additional study, let us accept provisionally that disks may indeed be present around massive stars, perhaps as a result of their formation process. How would such a structure respond to the onslaught of ultraviolet photons?

The top panel of Figure 15.17 shows schematically the expected result. At small radii, the upward, expansive force due to stellar heating is effectively opposed by the downward com-

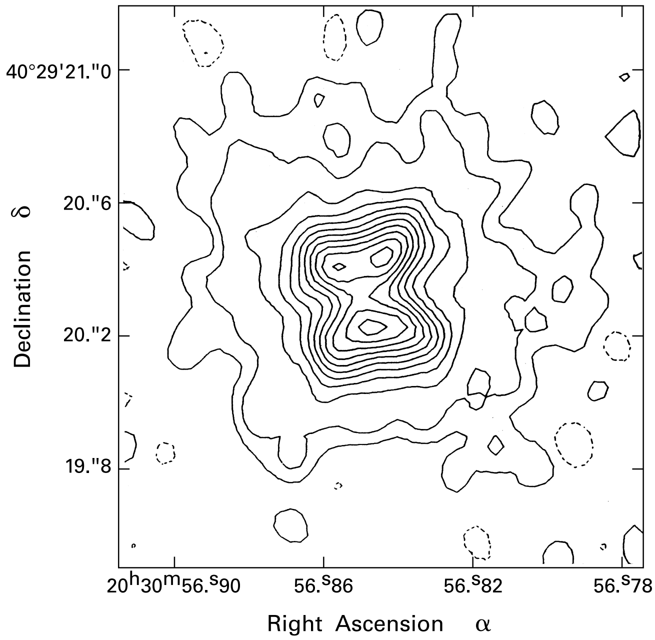


Figure 15.16 Radio emission from the embedded star MWC 349. The contours represent flux in 2 cm continuum radiation. Note the hourglass morphology.

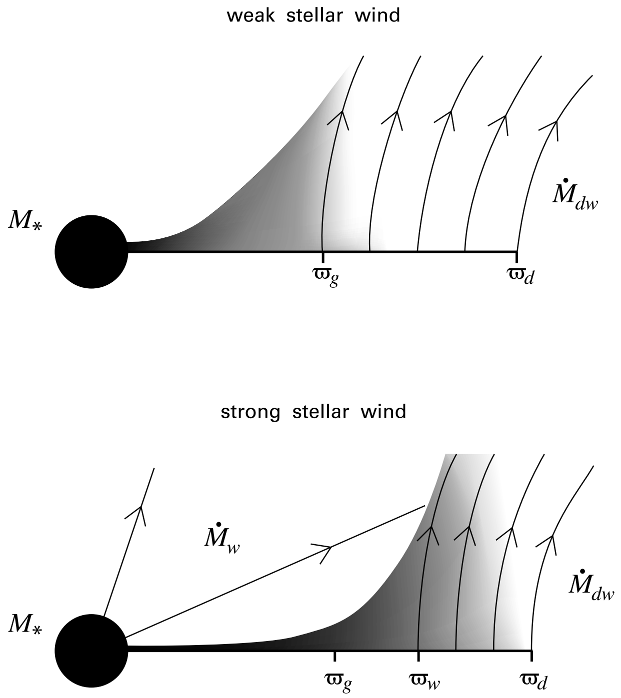
ponent of gravity. Thus, evaporated material forms a static atmosphere above and below the disk. Farther out, the vertical gravity becomes so weak that ionized gas flows off as a thermally driven wind. The transition occurs at the *gravitational radius*, ϖ_g. To find it, we first note that the atmosphere's scale height at any radius ϖ is approximately

$$\Delta z \approx \left(\frac{a_1}{V_{\rm Kep}}\right) \varpi \ , \tag{15.38}$$

according to equation (11.34). Here, a_1 is the isothermal sound speed within the atmosphere, where the temperature is expected to be nearly uniform and similar to that in an ordinary HII region. Since $V_{\rm Kep}$ is proportional to $\varpi^{-1/2}$, we see that the scale height increases radially as $\varpi^{3/2}$. The wind begins at the point where Δz grows equal to ϖ itself. Thus, a convenient definition is

$$\begin{aligned} \varpi_g &\equiv \frac{G\,M_*}{a_1^2} \\ &= 5 \times 10^{15}\ {\rm cm}\left(\frac{M_*}{40\,M_\odot}\right)\left(\frac{a_1}{10\ {\rm km\ s^{-1}}}\right)^{-2} \ . \end{aligned} \tag{15.39}$$

Note the similarity in the definitions of ϖ_g and the sonic point in a spherical, isothermal wind (equation (13.14)).


weak stellar wind

strong stellar wind

Figure 15.17 Photoevaporation of a disk around a massive star. If the stellar wind is weak (*upper panel*), gas flows out from the gravitational radius ϖ_g to the disk edge at ϖ_d. Inside ϖ_g is a static atmosphere. For a strong wind (*lower panel*), the atmosphere is suppressed, and the flow begins at the larger radius ϖ_w. The total mass flow rate from the disk is $\dot{M}_{\rm dw}$.

Material in the disk atmosphere is optically thick to the star's radiation. That is, while stellar ultraviolet photons bathe the atmosphere's upper layers, it is the internal, diffuse field that provides the true heating. This radiation arises from hydrogen recombinations to the ground state. The situation, once more, is analogous to a spherical HII region, where heating is also provided by internal recombinations. In both cases, cooling stems mainly from metallic forbidden lines. Assuming the disk atmosphere to be strictly isothermal, the reader may demonstrate that the density falls off vertically as a Gaussian. (Recall the discussion preceding equation (11.34).) For the full distribution of density, we need to know its value at the base of the atmosphere, *i. e.*, close to the disk surface. We shall return to this question momentarily.

The intensity of the diffuse radiation field within the atmosphere falls nearly to zero at the disk surface. If it did not, more gas would be lifted off the disk, and the local column would be too massive to be supported against gravity. An appropriate fraction of the gas would then settle back down. The actual diminution of the radiative intensity occurs when photons are absorbed by the small fraction of neutral hydrogen atoms.

In contrast, a finite ultraviolet intensity impinges on the disk for $\varpi \gtrsim \varpi_g$. Since direct starlight is largely absorbed by passage through the atmosphere, the ambient radiation is again

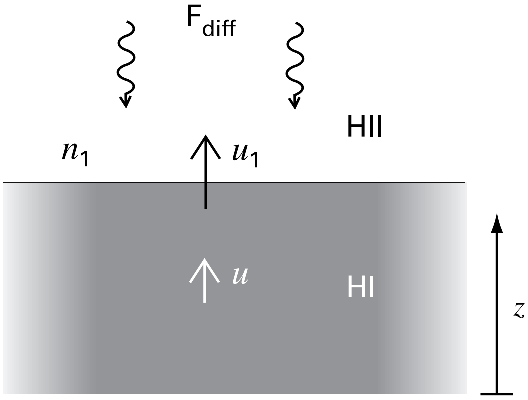


Figure 15.18 Inside the ionization front of a disk surrounding a massive star. Diffuse radiation from the wind impinges on the neutral disk. In response, gas flows upward and is ionized. The flow speed and density of this matter as it leaves the disk are u_1 and n_1, respectively.

the diffuse component, both from the upper reaches of the wind and from the atmosphere near $\varpi = \varpi_g$. Note that the wind interior is still nearly isothermal at a temperature similar to that of an HII region. The disk surface coincides with an ionization front, which slowly advances toward the midplane. As neutral disk gas is engulfed by the front and ionized, it flows away rapidly. The post-front velocity, we now demonstrate, is about equal to the sound speed in the ionized gas, or roughly 10 km s^{-1}.

15.4.2 Efflux of Disk Gas

Consider, as in Figure 15.18, the situation in the reference frame of the ionization front. Within the front itself, the equations of mass and momentum conservation, together with the heat equation, read

$$\rho\, u = \mathcal{C}_1 \tag{15.40a}$$

$$P + \rho\, u^2 = \mathcal{C}_2 \tag{15.40b}$$

$$\rho T \frac{Ds}{Dt} = \Gamma - \Lambda \,. \tag{15.40c}$$

Here, all variables are functions of the vertical coordinate z. The quantities $\mathcal{C}_1$ and $\mathcal{C}_2$ are spatial constants, while Γ and Λ are the volumetric heating and cooling rates, respectively. To evaluate the specific entropy s, we apply the second law of thermodynamics in the form

$$T\,\frac{Ds}{Dt} = \frac{D}{Dt}\left(\frac{3\,P}{2\,\rho}\right) - \frac{P}{\rho^2}\,\frac{D\rho}{Dt} \,, \tag{15.41}$$

where we have used the fact that the internal energy per unit mass in a monatomic gas is $(3/2)P/\rho$. That is, the ultraviolet radiation can excite no internal degrees of freedom, but only translational motion, which is reflected in the pressure.

In a steady-state flow, the convective derivative D/Dt is just $u\, d/dz$. After differentiating (15.40b) with respect to z and using equations (15.40a) and (15.41), the heat equation becomes

$$\rho\, u\,\frac{d}{dz}\left(\frac{1}{2}\,u^2 + \frac{5\,P}{2\,\rho}\right) = \Gamma - \Lambda \,. \tag{15.42}$$

The quantity in parentheses is the enthalpy per unit mass, which we denote by h. Just inside the ionization front, h rises sharply, since the ionization heating dominates cooling from metallic forbidden lines or hydrogen recombinations. Further downstream, Γ and Λ nearly balance, and h settles to a more constant value. The peak h-value is close to the net thermal energy gained from ionization, again measured per unit mass. This quantity, in turn, is about $k_B\,T_{\rm eff}/m_H$, where $T_{\rm eff}$ is the stellar effective temperature. (Recall the discussion preceding equation (15.18).)

On the other hand, mass and momentum conservation also imply a maximum possible value for h. Dividing equation (15.40b) by (15.40a) yields

$$\frac{a_T^2}{u} + u = \mathcal{C}_3 \ , \tag{15.43}$$

where $a_T^2 \equiv P/\rho$ is the isothermal sound speed, and $\mathcal{C}_3$ is another constant. We may thus write the enthalpy as

$$h = \frac{5}{2}\,\mathcal{C}_3\,u - 2\,u^2 \ . \tag{15.44}$$

The maximum value of h is $25\,\mathcal{C}_3^2/32$, attained when $u = 5\,\mathcal{C}_3/8$. Our previous argument then tells us that $\mathcal{C}_3$ is given by

$$\mathcal{C}_3 = \left(\frac{32\,k_B\,T_{\rm eff}}{25\,m_H}\right)^{1/2} \ . \tag{15.45}$$

Equation (15.43) implies finally that

$$\frac{a_T}{u} + \frac{u}{a_T} = \left(\frac{32\,k_B\,T_{\rm eff}}{25\,m_H\,a_T^2}\right)^{1/2} \ . \tag{15.46}$$

Using $a_T \equiv a_1 \approx 10\ {\rm km\ s^{-1}}$ in the post-front region and our representative $T_{\rm eff}$ of 3.8×10^4 K, we find a numerical value close to 2 for the nondimensional quantity on the righthand side. The post-front Mach number, u_1/a_T, is then near unity, as claimed.

The total rate at which mass leaves the disk as a wind is given by

$$\dot{M}_{\rm dw} = 4\,\pi\,m_H\,u_1 \int_{\varpi_g}^{\varpi_d} n_1(\varpi)\,\varpi\,d\varpi \ . \tag{15.47}$$

Here, ϖ_d is the outer disk radius, while $n_1(\varpi)$ is the number density of electrons (or protons) just outside the disk surfaces. At any radius smaller than ϖ_g, the ionizing flux falls to zero at the surface. To determine $n_1(\varpi)$, then, we may utilize equation (15.3) for the Strömgren radius, replacing R_s by ϖ and solving for the density. We find

$$n_1(\varpi) = \left(\frac{3\,\mathcal{N}_*}{4\,\pi\,\alpha'_{\rm rec}\,\varpi^3}\right)^{1/2} \ . \qquad \varpi \lesssim \varpi_g \tag{15.48}$$

Beyond ϖ_g, the density falls more rapidly. The ionizing flux no longer vanishes at the disk, but liberates electron-proton pairs that flow off as the wind. Here, the radiation originates mainly from recombining atoms located near ϖ_g. A numerical calculation finds that the associated flux declines roughly as ϖ^{-4}. The first factor of ϖ^{-2} is the usual one from a point source. One extra factor of ϖ^{-1} accounts for the increasing obliquity of the disk to radially outward photons. Finally, the flux attenuates even faster because of absorption by the small amount of hydrogen in the wind itself.

It is actually the divergence of the flux that provides the photons needed to ionize the disk. Taking the divergence introduces yet another factor of ϖ^{-1}, so that the local ionization rate declines as ϖ^{-5}. We equate this to the recombination rate, which is proportional to n_1^2. Thus, we write the density as

$$n_1(\varpi) = n_g \left(\frac{\varpi}{\varpi_g}\right)^{-5/2} , \qquad \varpi \gtrsim \varpi_g \tag{15.49}$$

where n_g is found by setting $\varpi = \varpi_g$ in equation (15.48). We next substitute equation (15.49) into (15.47) and do the integral. In the limit $\varpi_d \gg \varpi_g$, we find

$$\begin{aligned} \dot{M}_{\rm dw} &= 8\,\pi\, m_H\, u_1\, n_g\, \varpi_g^2 \\ &= 1 \times 10^{-4}\, M_\odot\ {\rm yr}^{-1} \left(\frac{\mathcal{N}_*}{10^{49}\ {\rm s}^{-1}}\right)^{1/2} \left(\frac{M_*}{40\ M_\odot}\right)^{1/2} . \end{aligned} \tag{15.50}$$

To obtain the second form of this equation, we have assumed precise equality of u_1 and a_1, which then cancel in the final expression.

The calculation previously mentioned, which includes a solution of the radiative transfer in both the disk atmosphere and wind, confirms the dependence of $\dot{M}_{\rm dw}$ on $\mathcal{N}_*$ and M_* given by equation (15.50), but yields a numerical coefficient lower by a factor of 3. In any case, we now see why extended disks, *i. e.*, those with $\varpi \gtrsim \varpi_g$, should *not* be present around observed, massive stars. A large circumstellar structure of 1 $M_\odot$ surrounding a 40 $M_\odot$ star survives only for a time of order 10^4 yr, or about 1 percent of the central object's main-sequence lifetime. The erosion of disk material inside ϖ_g has not been considered in detail, but the stellar wind must play a major role. If this wind carries sufficient momentum, as it does for a star of 40 $M_\odot$, it sweeps away the disk atmosphere. (See the lower panel of Figure 15.17.) Launching of the disk wind now starts at a radius $\varpi_w > \varpi_g$. At this point, the outward ram pressure of the stellar wind falls to a value comparable to the thermal pressure of evaporating material. We will not analyze this case further, but simply note that $\dot{M}_{\rm dw}$ is increased, since ionizing photons from the star itself are no longer attenuated by an atmosphere and can reach the outer disk directly.

15.4.3 Proplyds in Orion

Besides dispersing its own circumstellar disk, a massive star can also affect those around other stars. This circumstance arises because of the tendency for high-mass objects to be born near the centers of crowded stellar groups. Not surprisingly, the best-studied case is the Orion Nebula cluster associated with the Trapezium. The structures being photoevaporated by the Trapezium stars are called *proplyds*, a contraction of "protoplanetary disks."

Evidence for the Orion proplyds accumulated gradually. Optical imaging during the 1970s revealed a number of small patches clustered near θ^1 Ori C. These regions emitted strongly in such lines as Hα and [O III]λ 5007. Thus, they appeared to represent ionized gas, but were weaker in the low-excitation lines characterizing Herbig-Haro objects. Subsequent radio continuum observations with the VLA interferometer resolved the original group of emission objects and revealed dozens of others. Near-infrared surveys then found that almost all these structures contained their own low-mass stars. Since the regions could not be ionized locally, the obvious culprit was θ^1 Ori C and, to a lesser extent, the other Trapezium members. But what reservoir supplied the neutral matter to be ionized? The very detectability of the low-mass, central objects indicated that this matter was *not* distributed isotropically around each such star. A geometrically thin disk, on the other hand, could both provide ample material for ionization and yet leave its central star visible in the optical or near-infrared.

With the deployment of the Hubble Space Telescope in the 1990s, optical images with a resolution better than $0.''05$ became available. These spectacularly confirmed the disk hypothesis. At the same time, the images revealed a variety of morphological types that had been largely unsuspected. Figure 15.19 shows two outstanding examples, selected from the more than 100 objects now documented. The upper photograph, taken in Hα, illustrates the teardrop-shaped emission associated with many proplyds. Here, the elongated tail points directly away from θ^1 Ori C, which lies at a projected distance of 4×10^{17} cm. The disk itself is the dark slab located inside the teardrop. In this case, the disk does obscure the central star, at least in Hα. The bright crescent at the top is the main ionization front, arising where ultraviolet photons from θ^1 Ori C encounter neutral matter flowing off the disk. Other photons from the ambient HII region heat the disk from the rear and create a flow away from θ^1 Ori C. This neutral gas is hit nearly broadside by Lyman continuum radiation. The resulting ionization front assumes a tapered configuration.

Our description highlights an important difference between a disk around a massive star and one that is photoevaporated externally. In the first case, the ionization front coincides with the disk surface. In the second, the reduced flux of Lyman continuum photons is generally unable to penetrate to that depth. Instead, far-ultraviolet radiation with photon energies between 6 and 13.6 eV impinges on the disk, heating its surface layers to about 10^3 K. The heated gas moves out as a wind, until it meets the Lyman continuum radiation and forms the bright ionization rim (see Figure 15.20). For the proplyd shown in the top panel of Figure 15.19, this rim stands off from the disk with a separation of 6×10^{15} cm, or about twice the disk diameter. The same heating process occurs at the backside of the disk, but initiated by diffuse, far-ultraviolet radiation. On both sides, the neutral wind can only flow at disk radii exceeding ϖ_g in equation (15.39). Here, M_* refers to the low-mass, central star, while $a_1 \approx 3\ \mathrm{km\ s^{-1}}$ is the sound speed in the heated, neutral matter. Table 15.1 lists, for convenient reference, the far-ultraviolet output of massive stars.

For proplyds situated much closer to θ^1 Ori C, the gap between the leading ionization rim and the disk becomes narrower, while the tapered backside grows into a long and slender tail. Conversely, a proplyd lying farther from the massive star may have an ionization front so weak that it is dimmer than the background HII region. The lower panel of Figure 15.19 shows such a case. In this remarkable image, taken through a continuum filter centered on 5470 Å, the disk is seen clearly in silhouette. The bright patch above the structure and the much fainter one below are reflected light from the central, low-mass star.

It is interesting that the diameter of this quiescent disk, 1.5×10^{16} cm, exceeds that of the one undergoing heavy photodissociation. This trend is general and indicates that the evaporation

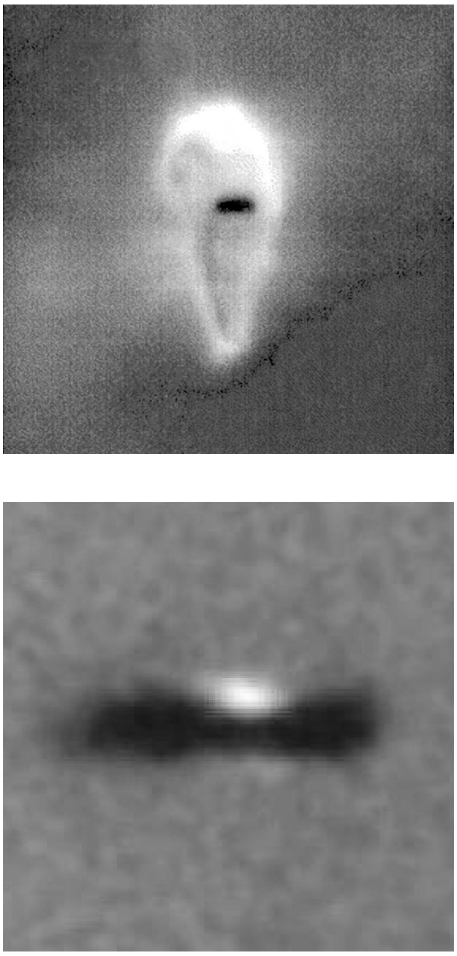


Figure 15.19 Two examples of proplyds in Orion. The top Hα image is of the proplyd 182-413. Note the dark interior disk and the bright ionization rim. The bottom image, taken through a filter centered at 5470 Å, is of the 114-426 proplyd. Here we see both the disk silhouette and scattered light from the star.

process first destroys the outermost region of lowest surface density. To estimate the total rate of mass loss, we again invoke equation (15.47). However, we no longer expect the base density $n_1(\varpi)$ to be a steeply declining function of radius. The neutral region interior to the ionization front is heated by the broad flux of far-ultraviolet radiation from the massive star. Taking n_1 to be a spatial constant in equation (15.47), we integrate to find

$$\dot{M}_{\rm dw} = 2\,\pi\,m_H\,u_1\,n_1\,\varpi_d^2 \ , \qquad (15.51)$$

where we have again used $\varpi_d \gg \varpi_g$.

The thickness of the neutral region is set by the condition that the far-ultraviolet flux diminish significantly close to the disk surface. The attenuation here is due to interstellar grains. Of course, the flux cannot vanish entirely, since photoevaporation would then cease. In summary, the neutral *column density* at each disk radius must be such that the optical depth to far-ultraviolet photons is a bit larger than unity. But this column density, N_1, is roughly $n_1\varpi_d$,

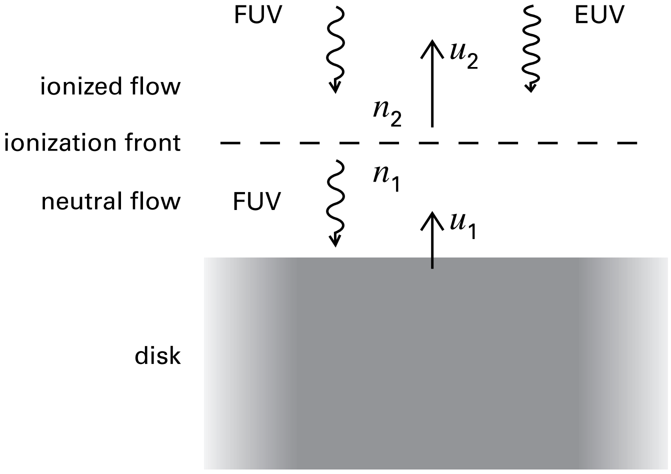


Figure 15.20 Photoevaporation of gas from a proplyd. Far-ultraviolet flux penetrates to the disk surface, where it creates an outward, neutral flow. Matter leaves the disk surface at speed u_1 and density n_1, then encounters the ionization front caused by extreme-ultraviolet photons. As gas crosses the front, its velocity increases to u_2 and its density falls to n_2.

where n_1 again refers to the base of the expanding flow (Figure 15.20). Note finally that the optical depth condition forces N_1 to be of order 10^{21} cm^{-2}. (Recall equations (2.43), (2.44), and (2.47), with $Q_\lambda \approx 1$.) We thus have

$$\dot{M}_{\rm dw} \;=\; 5 \times 10^{-8}\, M_\odot\, {\rm yr}^{-1} \left(\frac{N_1}{10^{21}\ {\rm cm}^{-2}}\right) \left(\frac{\varpi_d}{10^{15}\ {\rm cm}}\right) \ , \qquad (15.52)$$

where we have employed a u_1-value of 3 km s^{-1}. Molecular-line observations show that the disks themselves have masses of order $0.01\ M_\odot$. According to equation (15.52), the proplyd evaporation time is then 10^5 yr. Bearing in mind our rather uncertain estimate of $\dot{M}_{\rm dw}$, the Orion proplyds could not have survived for the Nebula age of 2×10^6 yr (Chapter 12). The implication is that θ^1 Ori C and the other Trapezium stars were relative latecomers, as we suggested earlier.

15.4.4 Destruction of Globules

The asymmetric appearance of the proplyds closer to θ^1 Ori C, with their brightest ionization rims facing the massive star, attests to a corresponding anisotropy in the radiation field. Although the objects in question are deep within the HII region, most of the ionizing photons stream directly from θ^1 Ori C and not from recombinations in the immediate, surrounding volume. In other HII regions, there are larger amounts of interior gas than in the Orion Nebula, often concentrated in clumpy, molecular form. We saw previously the dramatic elephant trunks protruding inward toward the O stars of M16. The bright rims that top each structure, conspicuous in Figure 15.5, clearly have a similar origin as the ionization fronts bounding the Orion proplyds. Individual globules engulfed by an HII region also glow in this manner. A prime ex-

Figure 15.21 Optical photograph in Hα of a cometary globule (designated Globule 1) in the Rosette Nebula. Notice the bright ionization rim on the right and the bifurcated tail to the left.

ample is the structure surrounding the HH 46/47 jet and outflow (Figure 13.2). Here, radiation from the supergiant ζ Pup has eroded enough of the cloud to bring the low-mass, embedded star driving HH 46/47 close to the surface.

Still other self-gravitating clouds are more strongly distorted by the radiation field. A striking example is shown in the Hα image of Figure 15.21. This *cometary globule* lies at the edge of the Rosette Nebula HII region, at a distance of 1.5 kpc. The structure consists of a rather flattened head and a bifurcated tail, which can be traced for $2'$ (1 pc). The bright ionization rim at the head faces the five O stars that power the HII region. The entire structure is reminiscent of some proplyds, but, like other globules, consists of dense molecular gas. We may picture the object being sculpted out of the inhomogeneous medium surrounding the newly formed O stars. The first material to be dispersed is the relatively sparse gas between the clumps. Matter that lies in the shadow behind these clumps, *i. e.*, shielded from the direct stellar flux, will survive longer. This selective evaporation may account for the elephant trunks and could also fashion cometary globules out of more isolated structures.

The bright rim of a globule advances slowly into the gas, eventually leading to its complete dispersal. How fast does the process occur? In principle, one could use equation (15.5) to determine the rate of advance of the ionization front. The difficulty is that the flux incident on the globule is no longer that emitted by the massive star. Evaporated material streaming off the cloud quickly recombines, becoming optically thick to the ultraviolet photons. Consequently, only a small fraction of the stellar flux F_* actually reaches the ionization front at any time.

We may use this very fact, however, to obtain a simple estimate for the evaporation rate. If an idealized, spherical globule of radius R emits a wind over half its surface, the associated mass loss is

$$\dot{M}_{\rm gw} = \pi\, m_H\, u_1\, n_1\, R^2 \ . \tag{15.53}$$

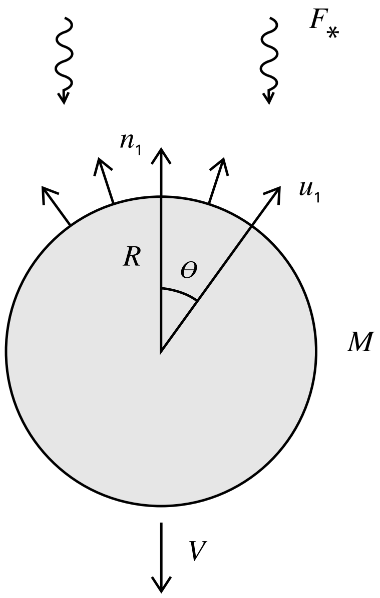


Figure 15.22 Photoevaporation of a spherical globule. The stellar flux F_* causes the cloud, of mass M and radius R, to emit a wind over half its surface. This wind has a uniform velocity u_1 and a base density that varies with the polar angle θ. Recoiling from its own wind, the globule moves away from the star with velocity V.

As usual, the subscript denotes values just downstream from the ionization front, in this case, at the base of the wind where it leaves the polar radius ($\theta = 0$ in Figure 15.22). We assume that the velocity is uniform over the hemisphere, but that the base density varies as $\cos\,\theta$, *i. e.*, as the projected area of the globule exposed to the stellar flux. The numerical coefficient in equation (15.53) then follows by averaging $\cos\,\theta$ over the hemisphere.

Let us now suppose that the stellar flux incident at the globule's polar axis is entirely absorbed. If d is the distance to the star, this condition is

$$\frac{\mathcal{N}_*}{4\pi d^2} = \int_R^\infty n_{\rm gw}^2\, \alpha'_{\rm rec}\, dr \ , \tag{15.54}$$

where $n_{\rm gw}$ is the density in the wind along $\theta = 0$, and where the upper limit of the integral reflects the fact that $d \gg R$. We assume for simplicity that the velocity maintains the value u_1 throughout the wind. Then the requirement of mass continuity dictates that $n_{\rm gw}$ fall with distance from the globule center as

$$n_{\rm gw} = n_1 \left(\frac{R}{r}\right)^2 \ , \tag{15.55}$$

so that

$$\frac{\mathcal{N}_*}{4\pi d^2} = \frac{(n_1)^2}{3}\, \alpha'_{\rm rec}\, R \ . \tag{15.56}$$

We solve this last relation for n_1 and substitute into equation (15.53) to find

$$\begin{aligned} \dot{M}_{\rm gw} &= \frac{\sqrt{3\pi}}{2}\, m_H\, u_1 \left(\frac{\mathcal{N}_*\, R^3}{\alpha'_{\rm rec}\, d^2}\right)^{1/2} \\ &= 2\times 10^{-5}\, M_\odot\ {\rm yr}^{-1} \left(\frac{\mathcal{N}_*}{10^{49}\ {\rm s}^{-1}}\right)^{1/2} \left(\frac{R}{0.1\ {\rm pc}}\right)^{3/2} \left(\frac{d}{1.0\ {\rm pc}}\right)^{-1} . \end{aligned} \tag{15.57}$$

Here we have assumed $u_1 \;=\; 10\ {\rm km\ s}^{-1}$.

Equation (15.57) indicates that a globule of several solar masses disappears relatively quickly once engulfed by an HII region. If the object had already begun internal, protostellar collapse or were on the verge of doing so, it might disperse before completing the process. On the other hand, most cometary globules seem to have masses larger by an order of magnitude, according to CO measurements. These should last longer, and indeed many have infrared stars embedded within their heads.

15.4.5 The Rocket Effect

There is another factor that tends to prolong the life of the clouds. We have described how the ionization of a globule creates a wind that, somewhat paradoxically, flows back toward the luminous star. In order to conserve momentum along this direction, the globule must suffer a recoil (Figure 15.22). The net result is that the wind-emitting cloud accelerates *away* from the star. This *rocket effect* can, in principle, drive structures out of the HII region entirely, so that they become protected from further photoevaporation.

To quantify this idea, we first note from Figure 15.22 that the wind velocity in the stellar direction is $u_1\,\cos\,\theta$. At the same latitude, the local mass loss rate is also proportional to cos θ, and the product of this rate and $u_1\,\cos\,\theta$ is the differential contribution to the thrust. After integrating over the hemisphere, we find that the *total* thrust on the globule is $(2/3)\,\dot{M}_{\rm gw}\,u_1$. If M and V denote, respectively, the globule mass and velocity, we then have

$$\begin{aligned} M\,\frac{dV}{dt} &= \frac{2}{3}\,\dot{M}_{\rm gw}\,u_1 \\ &= -\frac{2}{3}\,\frac{dM}{dt}\,u_1 \ , \end{aligned} \tag{15.58}$$

which immediately integrates to

$$M \;=\; M_0 \exp\left(-\frac{3\,V}{2\,u_1}\right) \ . \tag{15.59}$$

Here, M_0 is the initial value of the mass. The globule's starting velocity is taken to be zero.

The next step is to rewrite equation (15.57) in terms of the initial values of $\dot{M}_{\rm gw}$, d, and M:

$$\frac{dM}{dt} \;=\; -\dot{M}_0\,\frac{d_0}{d}\left(\frac{M}{M_0}\right)^{1/2} \ . \tag{15.60}$$

We have assumed here that the globule radius varies as $M^{1/3}$, which is true for isothermal spheres of modest density contrast. (Recall equation (9.15) and the accompanying discussion.) Combination of equations (15.58), (15.59), and (15.60) yields

$$\frac{dV}{dt} = \frac{2\,u_1}{3}\,\frac{d_0}{d}\,\frac{\dot{M}}{M_0}\,\exp\left(\frac{3\,V}{4\,u_1}\right) \ . \tag{15.61}$$

We now recast the lefthand side as the change of velocity with respect to distance. If we define $w \equiv V/u_1$ and $z \equiv d/d_0$, then the nondimensional form of (15.61) is

$$w\,\frac{dw}{dz} = \frac{\beta}{z}\,\exp\left(\frac{3\,w}{4}\right) \ , \tag{15.62}$$

where

$$\beta \equiv \frac{2\,d_0\,\dot{M}_0}{3\,u_1\,M_0} \ . \tag{15.63}$$

Notice, from equation (15.57) for $\dot{M}_{\rm gw}$, that β depends on $\mathcal{N}_*$, as well as the globule's initial mass and radius, but not on the displacement d_0.

Integration of equation (15.62) gives an implicit solution for the velocity:

$$\frac{4}{3}\left(w + \frac{4}{3}\right)\,\exp\left(-\frac{3w}{4}\right) = \frac{16}{9} - \beta\,\ln z \ . \tag{15.64}$$

Figure 15.23 plots, as a function of z, both w and the dimensionless mass $m \equiv M/M_0 = \exp(-3w/2)$. Here we have selected representative β-values of 1.0 and 0.5. With $\beta = 1.0$, the cloud accelerates and its mass falls by a large amount before the structure can move appreciably from its initial location. Changing β to 0.5 (equivalent, for example, to quadrupling the starting mass at fixed radius and $\mathcal{N}_*$), we find that the cloud reaches three times its initial distance before the mass drops by a factor of ten. At this point, the velocity is $1.5\,u_1$, or about 15 km s^{-1}. The rocket effect can therefore account for a relatively low-mass globule outside an HII region, but only if that object exhibits a substantial velocity.

15.5 Induced Star Formation

We mentioned earlier that the heads of cometary globules frequently contain one or more embedded stars. In some cases, there is no doubt that these objects are indeed young, as several drive molecular outflows and Herbig-Haro jets. An interesting question is whether the stars were born *prior* to the arrival of the HII region, or whether their formation was somehow *induced* by ionizing radiation impinging on the neutral globule. The analogous question arises regarding the supernova explosions that end the life of massive stars. Could the high pressure in the accompanying blast wave similarly trigger ambient clouds to form new stars?

Certainly there is no indication that all, or even most, lower-mass objects in the vicinity of massive stars form either way. We have already seen that the hundreds of pre-main-sequence stars in the Orion Nebula Cluster are probably *older* on average than the Trapezium itself. Nevertheless, are there instances where clouds engulfed by an HII region or supernova remnant indeed produce stars as a result?

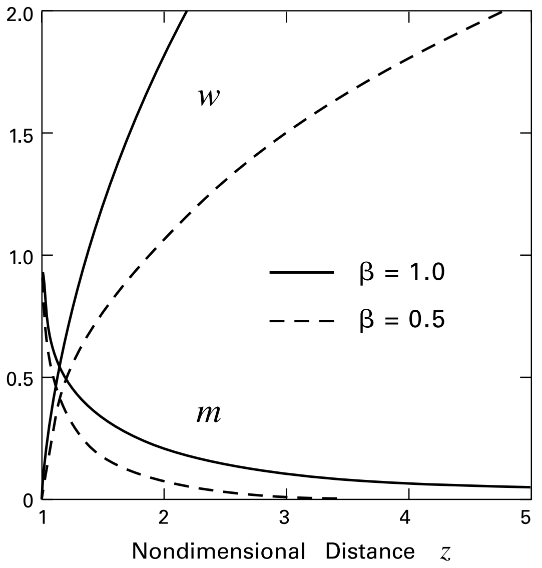


Figure 15.23 Acceleration and mass loss of a globule due to the rocket effect. Shown as a function of nondimensional distance z are the dimensionless mass m and velocity w, as defined in the text. Curves are displayed for indicated values of the parameter β.

15.5.1 Radiative Effects

Consider first the evidence for radiation-driven formation. Figure 15.24 is an optical image of a cometary globule in the HII region S199, at a distance of 1.9 kpc. The large cross represents a far-infrared source of total luminosity 1300 $L_\odot$, while the smaller crosses are near-infrared stars of lesser brightness. Two features are noteworthy. First, nearly all the embedded stars are situated close to, but behind, the prominent ionization rim. Second, the cluster members tend to lie between the far-infrared source and the rim, but *not* on the far side of the source.

A number of other cometary globules exhibit the same pattern. The presence of lower-mass objects close to an intermediate-mass one is familiar enough from our discussion of clusters and their production of higher-mass stars (Chapter 12). It is the *asymmetry* of the cluster distribution that stands out, and suggests a causal role for the external radiation field. After all, it is this same radiation that created the distinctive morphology of the globule in the first place. We are reminded of the swept-up appearance of some dark clouds within the ρ Ophiuchi complex (recall Figures 3.17 and 4.15). It is just these regions, lying adjacent to the Upper Scorpius OB association, that account for the most concentrated stellar production in the molecular complex. The important lesson is that massive, luminous stars can promote further star formation by increasing the density in nearby gas. Within individual globules, most of this pileup occurs in the swept-up material just inside the ionization rim, although there may be a secondary peak at the location of any intermediate-mass star.

It is clear that ultraviolet radiation does more than just erode a cloud through photoevaporation. It also exerts a mechanical *pressure* on that object, one that may radically alter its structure. The basis of this effect is surface heating by Lyman continuum photons. If we imagine the massive star suddenly turning on near a pre-existing globule, this heating causes a compressional wave to move inward toward the cloud's center. Meanwhile, the object is both losing

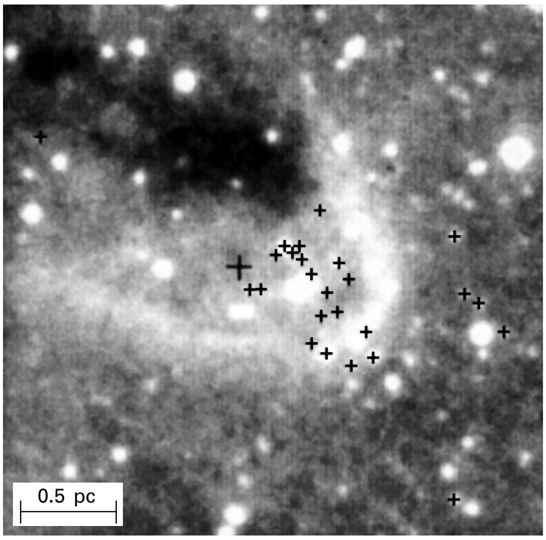


Figure 15.24 Optical image of Globule 13 in the HII region S199. The large cross near the center represents a bright, far-infrared source, and the smaller crosses are dimmer, near-infrared stars. These objects appear preferentially in the compressed region between the far-infrared source and the bright ionization rim.

mass through evaporation and accelerating away from the massive star. The readjustment of the cloud to dynamical equilibrium results in a new morphology that, as we have seen, is often markedly asymmetrical, even if the initial object were more spherical. Additional influence is exerted by the massive star's wind, whose ram pressure acts similarly to the radiation field, but is generally less important.

Numerical simulations of globule irradiation begin with a cloud already immersed in a hot, tenuous medium. The object is initially not self-gravitating, but confined by the background pressure. Once the radiation is switched on, an ionization front covers much of the surface. The compressional wave inside the front steepens into a shock and forces material toward the symmetry axis. A calculation illustrating this effect is displayed as Figure 15.25. The short-lived implosion phase is followed by a partial rebound, after which the compressed central object settles into its equilibrium, cometary structure. How, then, does an interior cluster form? The conventional view has been that star formation happens when the globule's mass and added pressure are so large that the object cannot find an equilibrium, but instead goes into dynamical collapse.

This view, however, runs contrary to our ideas on both cloud stability and cluster formation. Suppose the initial globule had significant support from an internal magnetic field. Then, from the discussion in § 9.4, *no* increase of the external pressure is enough to render it gravitationally unstable. On the other hand, the magnetic field may be negligible, and the globule may initially be in pressure balance with the background. Once the object is subjected to a high enough external pressure to cause collapse, the situation is essentially equivalent to starting with a cloud mass that exceeds the Jeans limit. In either case, the structure is out of force balance and undergoes dynamical fragmentation. It splits into a number of lower-mass clouds that are themselves gravitationally stable, and therefore incapable of producing stars.

Any induced star formation must therefore arise from a static or, at best, quasi-statically evolving entity. In the present case, this entity is a cometary globule molded by the radiation

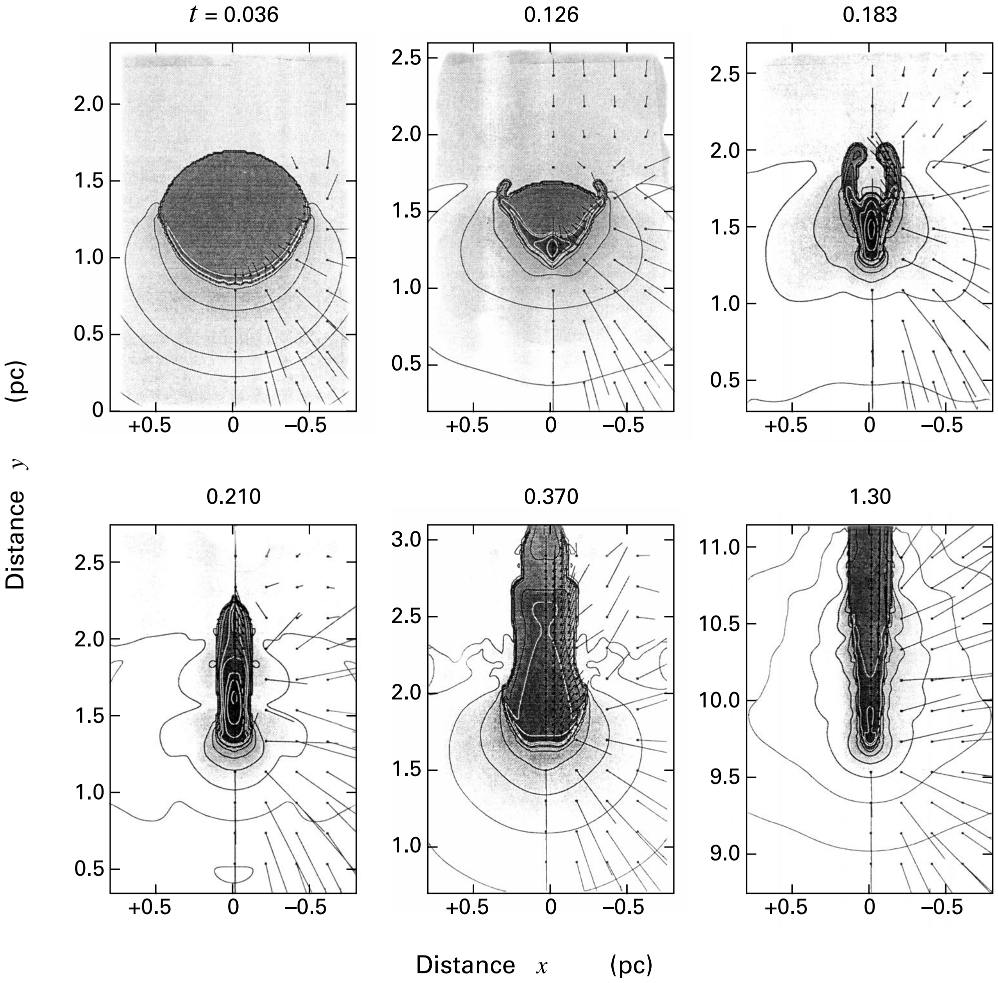


Figure 15.25 Numerical simulation of a globule subject to ionizing radiation. The times at the top of each panel are in units of 1×10^6 yr. Line segments are velocity vectors and represent the distance traveled by fluid in 2×10^4 yr. Exposed to radiation from below, the initially spherical cloud of mass 20 $M_\odot$ first collapses toward its symmetry axis, then partially reexpands, before reaching a final, elongated configuration at a much later time. The radiative flux corresponds to that from an O5 star at 10 pc.

field within an HII region. The structure may exist for a substantial period before it spawns interior dense cores, which in turn lose magnetic flux through ambipolar diffusion and collapse to stars. Precisely how the swept-up morphology stimulates core growth remains unknown. There is no reason to suspect, however, that the mechanism differs fundamentally from that in any other cluster-forming region. In any case, future observations that place a globule's stars within the HR diagram should help elucidate the chain of events.

15.5.2 Dynamics of Supernova Remnants

There is another way by which massive stars exert a dynamical influence on nearby clouds. This is the pressure increase associated with a supernova. As discussed in Chapter 1, all stars heavier than about 8 $M_\odot$, *i. e.*, those of spectral type B3 and earlier, undergo this violent explosion as their central region collapses to form a neutron star. Lower-mass objects that become white dwarfs also detonate if they gain sufficient mass from a binary companion. The latter event is termed a Type Ia supernova, as distinguished from the Type Ib, Ic, and Type II supernovae, all of which originate from high-mass stars. (This classification scheme is based on the spectral appearance of the optical light.) Spiral galaxies like the Milky Way produce several supernovae per century, with Type Ia events occurring less frequently than the others. Within our own Galaxy, the complex of glowing gas in Taurus known as the Crab Nebula is the most conspicuous example of a *supernova remnant*, the swept-up gas racing away from the explosion site. As we shall see presently, there are observational signatures of such remnants interacting with giant molecular clouds. Some of the ejected gas creates huge shells and bubbles, and eventually escapes the Galaxy entirely.

Most of our understanding of supernova remnant dynamics is based on consideration of an idealized, uniform background, just as in the corresponding analysis of HII regions. Following the detonation near the star's center, a shock wave propagates outward and ejects a few percent of the stellar mass. The initial speed of this gas is enormous, typically 1×10^4 km s^{-1}, or 3 percent the velocity of light. Long before this event, the cloud material in the stellar neighborhood had already inflated to some degree because of the stellar radiation field. In any case, the sparse ambient gas offers little resistance to the massive and dense envelope. Through its expansion, however, the remnant continually sweeps up external matter. Once its total mass has risen appreciably, the envelope begins to decelerate. For a representative initial envelope mass of 0.2 $M_\odot$, the initial, constant-velocity phase lasts about 100 yr, assuming a background number density of $n_{\rm tot}$ = 1 cm^{-3}. The outer edge of the remnant now lies a parsec or so from the explosion site.

This advancing boundary is a shock front, through which ambient gas flows.[4] Although the front slows down, its velocity is at first so high that the shocked gas cannot efficiently cool. (Recall, for comparison, the discussion concerning massive outflows at the end of § 15.3.) The expansion at this time, the so-called *Sedov-Taylor phase*, is therefore an adiabatic one. In fact, since the moving gas does little mechanical work on its surroundings, the total *energy* of the system maintains its initial value, *i. e.*, that imparted by the explosion itself. Both kinetic and thermal contributions are significant, arising from the bulk and random motions, respectively, of the mass-gaining remnant. Note that it is the pressure of shocked, interior gas that drives the expansion.

To assess the actual rate of deceleration, a simple dimensional argument suffices. Let $R_{\rm shock}$ denote the radius of the leading shock front, which travels at the speed $V_{\rm shock}$ into the static environment. Since the total energy is fixed, the volumetric energy *density* of the remnant falls

---

[4] The expansion speed in a young supernova remnant is so high, and the background density sufficiently low, that interparticle collisions are too infrequent to mediate the shock transition. Instead, the ambient magnetic field takes up this role, resulting in a *collisionless shock*. The tangled field within the front is believed to be responsible for cosmic ray acceleration.

as $R_{\rm shock}^{-3}$. But the average interior pressure of any nonrelativistic gas is just 2/3 of its energy density. Let us further suppose that the immediate postshock pressure is itself proportional to this average value. (A detailed analysis shows it is about twice the average.) The postshock pressure varies as $V_{\rm shock}^2$, as we showed, for example, in equation (8.50). Combining these results, we conclude that $V_{\rm shock}$ falls as $R_{\rm shock}^{-3/2}$. Since $V_{\rm shock} = dR_{\rm shock}/dt$, it also follows that $R_{\rm shock}$ is proportional to $t^{2/5}$, as opposed to the linear increase with time during the initial, free expansion phase.

Suppose that the remnant, initially traveling at 1×10^4 km s^{-1}, swells to a radius of 1 pc in 100 yr, at which point the velocity begins to fall. The postshock temperature also declines. When it reaches about 5×10^5 K, radiative cooling by metallic ions comes into play. Equation (8.50), with μ set equal to 0.61, implies that the required shock velocity is 200 km s^{-1}. This velocity is attained when the radius increases by a factor of $(5000/200)^{2/3} = 9$. Here we have crudely accounted for prior deceleration by setting the initial velocity at the Sedov-Taylor phase to 5×10^3 km s^{-1}. The elapsed time during this adiabatic period is longer by $(5000/200)^{5/3}$, yielding 2×10^4 yr. The accumulated mass, which is proportional to the swept-up volume, is $100\,M_\odot$.

Rapid postshock cooling now leads to increased compression behind the moving shock front. As a result, the remnant becomes essentially a thin shell, which accumulates mass as it continues to inflate. During this final, *snowplow* phase, the expansion is momentum-driven. By our previous reasoning, the total mass is proportional to $R_{\rm shock}^3$. But conservation of momentum requires that the mass is also inversely proportional to $V_{\rm shock}$. It follows that $V_{\rm shock}$ varies as $R_{\rm shock}^{-3}$ and that $R_{\rm shock}$ itself increases only as $t^{1/4}$. Once the velocity becomes comparable to the random speed characterizing the interstellar medium, the shell merges into the background. The total elapsed time is about 1×10^6 yr for our example, and the final radius is 23 pc. Here, we have assumed an ambient velocity dispersion of 10 km s^{-1}, appropriate for the warm, neutral gas that occupies most of the interstellar volume.

A shell-like structure for the remnant is often apparent at radio wavelengths, which have provided the means for detecting several hundred such objects across the Galaxy. The spectral energy distribution in the radio regime differs qualitatively from that of HII regions, in that the intensity declines rather steeply with frequency. It is clear that the underlying mechanism is *not* bremsstrahlung, but a *nonthermal* process. The emission, which is often highly polarized, represents *synchrotron radiation*. It arises from relativistic electrons spiraling about an internal magnetic field that expands along with the gas. These electrons, originating in the ambient medium, are accelerated by the advancing shock wave. Far more important in cooling the remnant are X-rays, which indeed stem from bremsstrahlung in the shocked gas. Images taken by X-ray satellites, starting in the 1970s, have provided much of our knowledge concerning physical conditions in this material. Finally, many remnants, ranging from the 300 yr old Cassiopeia A to the Cygnus Loop with an age of 4×10^4 yr, exhibit bright optical and ultraviolet emission in a system of extended, clumpy filaments. Both Hα and such familiar lines as [O III] $\lambda\,5007$ and the [S II] $\lambda\lambda\,6716, 6731$ doublet attest to the shock origin of this component. In contrast, continuum emission in the optical is again strongly polarized and stems from the synchrotron process.

15.5.3 Blast Waves and Clouds

We are concerned primarily with how the blast wave affects ambient cloud material. If the supernova progenitor is an O star, it has already driven away molecular gas from a large volume, at least 15 pc in radius, during its main-sequence lifetime. This dispersal occurs via thermal expansion of the relatively diffuse, interclump component, together with photoevaporation and the rocket effect acting on the clumps themselves. After the explosion, the shock associated with the supernova remnant is thus considerably weakened before it runs into molecular material. In contrast, stars of spectral type B3 to B0, *i. e.*, from 8 to 12 $M_\odot$, directly impact the original cloud complex once they detonate.

Observational examples of supernova-cloud interactions are actually quite rare. Of those remnants that have been studied in molecular lines, only a handful exhibit highly perturbed, dense gas. The best case is IC 443, a 10^4 yr old remnant in the Gem OB1 association, at a distance of 1.5 kpc. The left panel of Figure 15.26 shows, in the negative, an optical photograph taken through a red filter. Such an image is dominated by Hα emission from postshock gas. Here we see two partial shells whose radii of curvature are manifestly different. The implication is that cloud material to the upper left has a relatively high density, so that the associated shock is advancing more slowly. Notice the filamentary arcs protruding from the righthand shell, whose radius is 13 pc. A symbol in both panels marks the position of a strong, point-like X-ray source that is presumed to be the actual neutron star.

The lefthand panel also displays heavier contours of 2.12 µm H_2 emission, which is very high in this source. Morphologically, the arc of shocked molecular hydrogen neatly continues the lefthand shell, presumably into cloud regions that are too dense to be visible in Hα or other optical lines. This supposition is confirmed by observations in CO, which show a broad swath of clumpy, molecular gas cutting the remnant diagonally, between the optical arcs. Numerous other emission lines, from such species as OH, HCO^+, and SiO, are detected in this region. All the lines are relatively broad, with radial velocity widths up to 90 km s^{-1}. Additionally, the transitions are from both low-lying levels and more elevated ones. Clusters of maser spots are seen in the 1720 MHz line of OH. All of these molecules are being excited and set into motion by a shock traveling about 40 km s^{-1} into gas that was already clumpy. This material still partially surrounds the middle portion of the expanding remnant.

In the right panel of Figure 15.26, we display, again as an overlay on the optical image, contours of diffuse X-ray emission. The morphology is clearly different from that in the optical or near infrared. There is no longer any indication of a rim-like structure, but a more uniform distribution of hot gas throughout the interior volume. The original Sedov-Taylor solution predicts that both the density of the gas and its emitted X-ray radiation should indeed peak toward the shock boundary. That they do *not* suggests that significant amounts of material have been injected within the volume and heated to somewhat above 10^6 K. A natural assumption is that this matter represents initially cold, molecular clumps that were overrun by the blast wave. The smaller of these clumps were initially squeezed by the greatly enhanced ambient pressure. After partially re-expanding and establishing dynamical equilibrium, the clumps then evaporated and merged into the background gas.

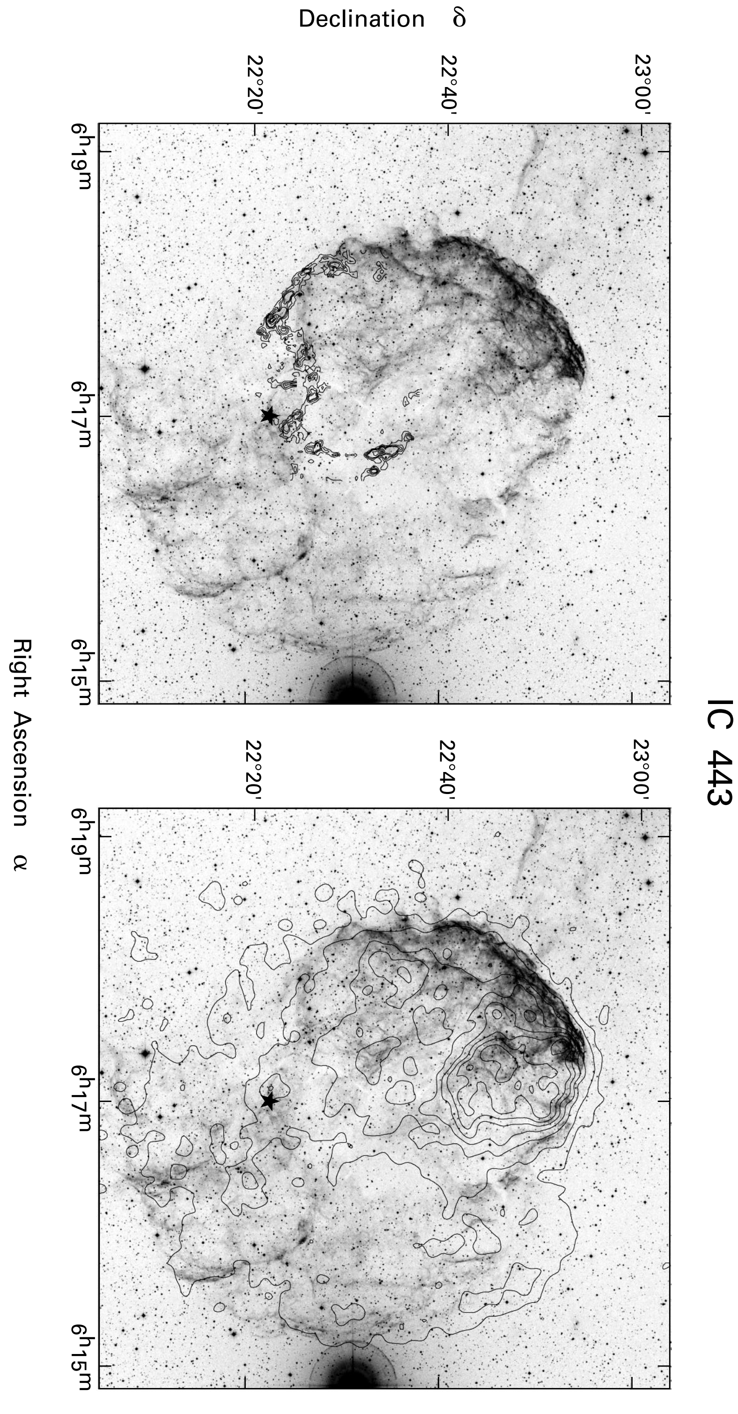


Figure 15.26 *Left panel:* Optical image of the supernova remnant IC 443, shown with contours of 2 μm emission from H_2. *Right panel:* The same optical image, but now combined with contours of X-ray emission in the range from 0.2 to 4.0 keV. The position of the neutron star is indicated in both panels.

15.5.4 Bubbles and Shells

An individual supernova remnant typically spans a few tens of parsecs and cools significantly after 10^6 yr or so. On the other hand, there are much larger regions in our Galaxy, with diameters ranging from 10^2 to 10^3 pc, that display a similar morphology. That is, they consist of a hot, tenuous interior surrounded by a thick surface layer of relatively cold and dense gas. Judging from the measured velocity of the outer component, expansion has been proceeding for some 10^7 yr. The process must be driven by *multiple* supernova explosions occurring in a quasi-continuous manner over the entire epoch. A single giant molecular cloud survives for the required length of time and could develop numerous sites of massive star formation over this period.

Our own solar system is immersed in just such a hot, expanding region. When discussing the interstellar radiation field in Chapter 7, we noted the existence of a component in the soft X-ray regime. Photons here originate from a 10^6 K plasma with an estimated diameter of 100 pc. This *Local Bubble* is also noteworthy for its low HI density compared to surrounding regions. Indeed, 21 cm observations reveal the presence of an exterior shell of atomic hydrogen, commonly referred to as *Lindblad's Ring*. Radio data are consistent with an elliptically shaped band of gas expanding within the Galactic plane (see Figure 15.27). Although the original shell may have been launched isotropically, the outward motion within the plane becomes sheared by the same gravitational field that produces the differential rotation of the stars. Figure 15.27 depicts how placement of the solar system just inside the Ring's border yields a pattern of radial velocities that matches the 21 cm observations.

If we compare Figure 15.27 with Figure 4.14, we see that the oval contour of Lindblad's Ring strikingly resembles Gould's Belt. This extensive system of bright stars, we recall, contains a number of prominent OB associations, including those of Orion and Scorpius-Centaurus. A precise comparison between the two structures is inappropriate, since the underlying ring calculation ignores such factors as deceleration of the HI shell from its encounter with external mass. Nevertheless, the similarity in both the spatial configuration and the expansion velocities is surely no coincidence. The generally accepted picture is that supernovae began detonating about 6×10^7 yr ago in the now largely dispersed Cassiopeia-Taurus association. Shock heating from multiple expanding remnants created the Local Bubble and also swept up gas into both Lindblad's Ring and the molecular complexes of Scorpius-Centaurus, Orion, Perseus, and ρ Ophiuchi. More recent OB associations, such as Ori OB1 and Per OB2, later arose within this gas, and now comprise Gould's Belt. We emphasize that the causal link from the original Cas-Tau supernovae to the present-day O stars is indirect, with the former helping to shape the environment within which the usual star formation processes have been operating.

Numerous other shells throughout the Galaxy are discernible from spatial maps at 21 cm. The atomic hydrogen away from the Galactic plane tends to be concentrated in filamentary arcs; many of these are likely to be parts of closed, nearly spherical structures. Within the plane, one needs to examine maps constructed in narrow slices of radial velocity, in order to minimize contamination by foreground and background gas. Such images display both arcs and complete shells. In some cases, the systematic variation of the shell diameter with radial velocity demonstrates convincingly and quantitatively an expansive motion. Utilizing the Galactic rotation curve, one may assign distances to the structures and convert angular to physical radii. The

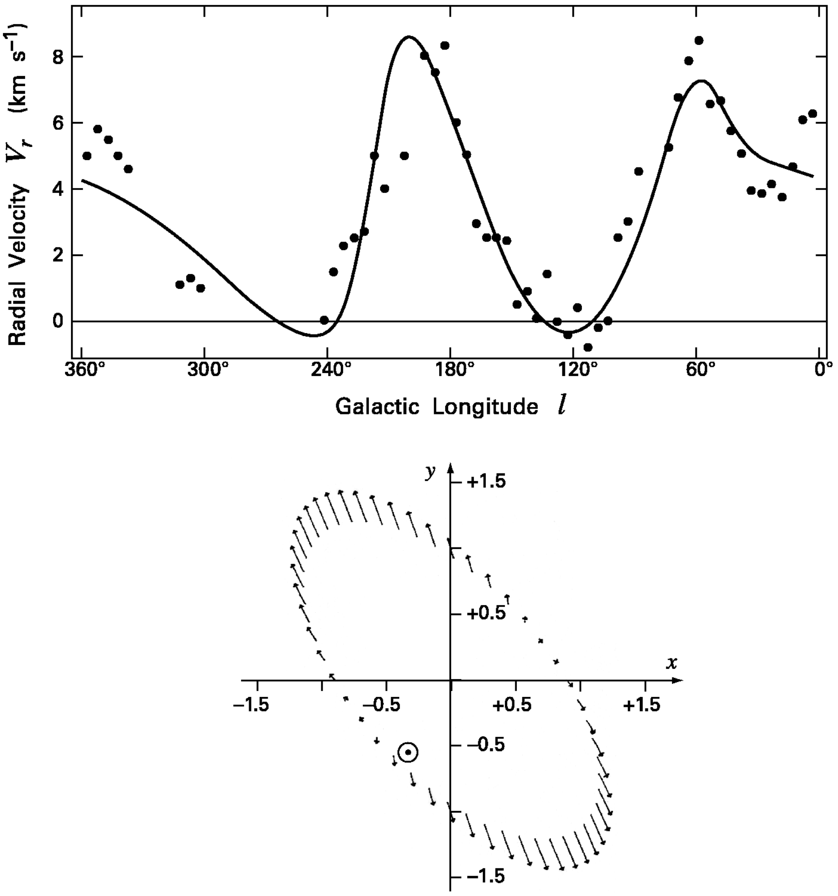


Figure 15.27 Expansion of Lindblad's ring. Data points in the upper panel are observed radial velocities in HI as a function of Galactic longitude. The smooth curve shows the expected distribution of radial velocity if the HI gas were expanding as shown below. The elliptical shape represents an initially isotropic expansion into the Galactic disk, after 6×10^7 yr of evolution. Calculated velocities refer to the indicated solar system position inside the ring.

latter range in magnitude from the resolution limit to values even larger than that for Lindblad's Ring, about 250 pc. At the top end of the scale are shells whose estimated mass (over $10^6\ M_\odot$) and size would seem to require an explosion energy above 10^{53} ergs. The latter exceeds the energy from a single Type I supernova by at least two orders of magnitude. However, the underlying assumption in this estimate is that injection occurs instantaneously. Less total energy would be needed for the shell to attain a given radius if that energy were instead added in a continuous fashion. Thus, the observed *supershells* can plausibly be formed through the combined action of winds and supernovae from large OB associations.

Within a supershell is a *superbubble*, and these too are seen, through X-ray observations. A well-studied example is the Cygnus Superbubble, located at a distance of 1.3 kpc. Here, a large region of X-ray emitting gas, some 13° (or 300 pc) in diameter, is bifurcated by a thick ridge of heavy obscuration. Within the latter is the highly luminous Cygnus OB2 association, thought to be energizing the entire structure. (See Figure 4.14 for the location of this stellar group.) The area surrounding the X-ray region contains shock-heated filaments glowing in Hα and coincides with an HI supershell. Four discrete OB associations are also found along this border. At one extremity lies Cygnus OB7, which the reader may also find in Figure 4.14. The Hipparcos satellite has provided proper motions for numerous stars in this peripheral region, showing them to be in a state of global expansion.

Bubbles and their surrounding shells constitute a large fraction by volume of the interstellar medium. In particular, the diameter of a supershell is comparable to the scale height of HI gas. That portion of the shell normal to the Galactic plane feels an additional retardation from the gravity of the local disk. The shell contour thus begins to flatten, an effect that is sometimes apparent in the 21 cm maps. With continued expansion, the interior bubble bursts from the disk entirely. Hot gas spews outward as a *Galactic fountain*. Once this matter cools, it may return to the disk at a location well removed from its injection point. Observed clouds of atomic hydrogen that are approaching the plane with speeds from about 30 to 100 km s^{-1} may represent this reentry component.

15.5.5 Supernova-Induced Collapse

Supernovae are thus primary agents for both heating and mechanically stirring gas on the Galactic scale. Assessing their collective effects is essential for understanding the interstellar medium. Our present concern, however, is to explore the means by which new stars may form as a result of these energetic events. Aside from the indirect route, *i. e.*, the compression of ambient gas into relatively dense molecular clouds, is it possible for an individual supernova to provoke collapse in a nearby globule?

From a theoretical perspective, the answer is clear and affirmative. Figure 15.28 displays snapshots from a three-dimensional numerical calculation simulating the impact of a planar shock front on an isolated globule. The $1\ M_\odot$ object was initially spherical and nearly in hydrostatic equilibrium. As the shock enters from the right, the cloud flattens and the front distorts into a broad, concave structure. The associated pileup eventually drives some of the cloud gas into the center, raising the density there substantially. The local velocity increases toward this location, a sign that collapse is imminent.

The particular simulation of Figure 15.28 employed a shock speed of 25 km s^{-1}, not far below the 40 km s^{-1} inferred for the supernova remnant IC 443. Other experiments find that a substantially higher shock speed does *not* lead to collapse, but to violent cloud shredding and dispersal. Conversely, a shock that is too weak results in temporary distortion and compression of the cloud, followed by rebound to a spherical equilibrium. It thus appears that over some range of shock speeds, including those appropriate for some supernova remnants within molecular environments, pre-existing globules are indeed set into collapse.

Even if the cloud collapses, however, it does not follow that new stars are always, or even usually, formed as a result. The reasoning here is the same as we applied to cometary globules inside HII regions. Indeed, much of the compression in the supernova case is not from ram pressure *per se*, but from the enhanced temperature behind the shock front, which has a similar

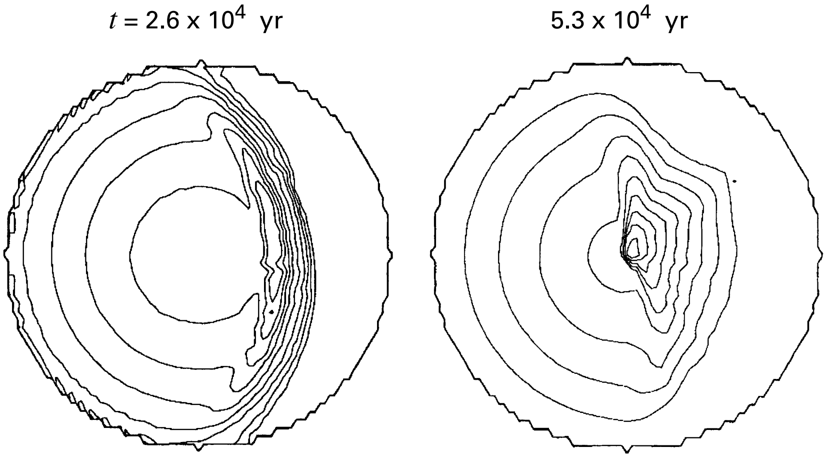


Figure 15.28 Numerical simulation of the shock-induced collapse of a cloud. Contours represent mass density and are spaced by factors of two in that variable. The cloud is initially spherical, with a mass and radius of 1 $M_\odot$ and 10^4 AU, respectively. At $t = 0$, it is struck by a planar shock front approaching from the right at 25 km s^{-1}.

effect as that created at the globule surface by ultraviolet heating. The resultant collapse, in either case, is unlikely to yield *stellar* densities unless the applied pressure marginally exceeds the maximum allowed for gravitational stability. Barring such a match, which could occur only through a fortuitous combination of supernova energy, distance, and the density of the medium surrounding the cloud, the outcome of induced collapse is dynamical fragmentation.

15.5.6 Extinct Radionuclides

The formation of low-mass stars through the induced collapse of molecular clouds would thus seem to be a rare event. Even so, a number of planetary scientists have maintained that just such a picture applies to our own solar system. Their evidence is a pattern of anomalous isotopic abundances found within certain meteorites. This pattern is a fossil record of short-lived, radioactive species that may have been injected just at the time of solar system formation. Since supernova remnants carry such species, it is argued that the blast itself triggered collapse and the subsequent creation of the Sun.

Meteorites are fragments of asteroids, the subplanetary masses orbiting within the solar system. These bodies, in turn, are relics of the solar nebula, *i. e.*, the protostellar disk within which the planets themselves formed. Most rocky meteorites (as opposed to those with a high metallic content) are classified as *chondrites*, and it is these which most faithfully represent the original nebular conditions.[5] Chondrites are not homogeneous rocks, but aggregates of smaller objects, termed *chondrules* and *CAIs*. The former are millimeter-size droplets, rich in iron,

---

[5] Abundance analysis of long-lived radioactive species within chondrites has given us the presently accepted age of the solar system, 4.6×10^9 yr.

magnesium and silicates. The term CAI is an acronym for “calcium-aluminum-rich inclusion.” These are larger, rounded structures, up to several centimeters in size. They are depleted in just those volatile elements that make up the chondrules and contain an array of minerals, each with its own crystalline structure.

For the most part, the relative elemental abundances within chondritic meteorites are close to those observed in the solar atmosphere. This basic fact supports the view that the original material was a homogeneous gas that later differentiated into the central star and its disk. Depletion in the more volatile elements, such as that occurring in CAIs, presumably reflects high temperatures in the solar nebula, at the sites where these objects condensed. There are also variations within the isotopic abundances of a given element. Some of these variations are attributable to fractionation processes, akin to the selective chemical reactions that alter the [^{12}C/^{13}C] ratio throughout the Galaxy (Chapter 6). Other isotopic shifts, however, do not fit this pattern and appear to be the result of prior radioactive decay.

Let us consider one case, the enrichment of ^{26}Mg in CAIs. This isotope may be produced by the decay of ^{26}Al. What is the evidence that such a decay actually occurred in the solar nebula? Figure 15.29 shows data from a single CAI of the Allende meteorite, recovered in Mexico in 1969. This particular specimen contains four distinct crystalline phases. Each has its own characteristic abundance ratio of the two principal isotopes, ^{27}Al and ^{24}Mg; this ratio is plotted on the horizontal axis. On the vertical axis is displayed the corresponding ratio of ^{26}Mg to ^{24}Mg. This latter quantity varies by some 10 percent throughout the CAI, with a minimum of 0.140, just the value found in terrestrial and lunar rocks.

The most significant feature of Figure 15.29 is that the local enrichment of ^{26}Mg increases *linearly* with the abundance of ^{27}Al. A plausible explanation is that a certain fraction of aluminum was in the isotopic form ^{26}Al at the time the inclusion solidified. Different parts of the body ended up with different amounts of aluminum, depending on their crystalline phase. The original fraction of the radioactive isotope, [^{26}Al/^{27}Al]$_0$, may be read from the slope of the line as 5×10^{-5}. Note that the half-life of ^{26}Al is only 1.1×10^6 yr. Assuming that this isotope was produced during a supernova, there was very little time from the explosion to incorporation of the hot gas into the forming solar nebula. Hence the hypothesis of triggered collapse.

Other elements tell a similar tale, with comparable fractions of extinct, radioactive isotopes. At present, there is no quantitative model that produces all these contaminants in their correct proportions. Nor is their a firm understanding of how the radioactive material penetrated the collapsing dense core to the relatively tiny disk in its deep interior. Some theorists have suggested that the trigger was not a supernova at all, but the wind shed by a lower-mass star during its asymptotic giant phase. Others have pointed out, in opposition to the triggering hypothesis, that energetic protons from flares in the young Sun could have reacted with ^{25}Mg in the disk to create ^{26}Al. (This same reaction produces the isotope in supernovae and giants.) Overshadowing all these competing ideas is our ignorance of how chondrules and CAIs form at all. The morphology of both shows that their constituent matter was heated to some 1500 K and that their solidification occurred very quickly after this heating. The nature of this thermal event is at least as mysterious as the isotopic anomalies.

Returning finally to the general issue of star formation, we have stressed that any triggering scenario relies on a finely tuned relation between the dense core and the external agent. Such

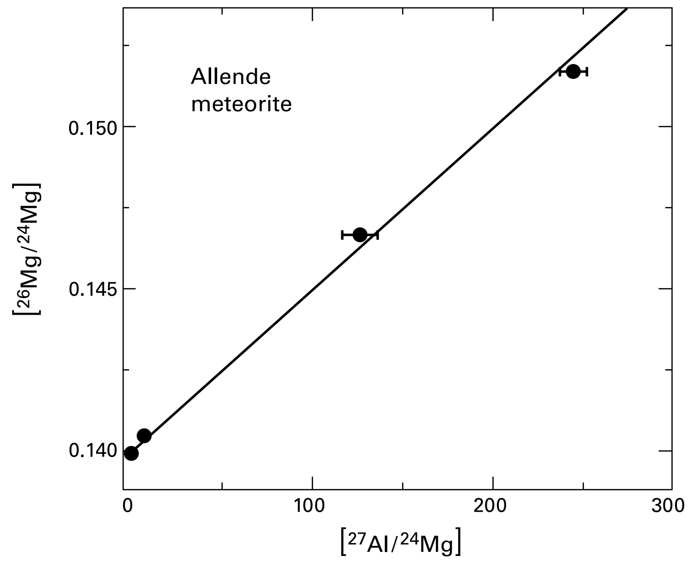


Figure 15.29 Observation of excess ^{26}Mg in the Allende meteorite. The indicated isotopic ratios are plotted for four different crystalline phases within a single CAI. The straight line is a best fit to the data.

conditions are unlikely to hold throughout an extensive cluster or association. On the other hand, the orchestrated pattern of star formation within such groups *does* suggest an environmental influence (Chapter 12). Part of this influence could well be the prior compression of the larger cloud by nearby, massive objects. If we are to reject any direct form of triggering, then the birth of each low-mass star must still proceed through the growth, to the point of instability, of its parent dense core.

Chapter Summary

The ultraviolet radiation from a high-mass star ionizes ambient gas. If the latter is a typical molecular cloud, the ionization front initially races out at high speed. The front decelerates as photons are depleted, then begins a more prolongued expansion at subsonic velocity. Eventually, the highly pressurized HII region bursts out of the cloud. The resulting champagne flow may be traced through its optical and radio emission.

Observations show that massive stars are actually born in environments far denser than ordinary clouds. Such hot cores emit copiously in the infrared and through radio molecular lines, but exhibit no internal HII region. At a later phase, the ionization front is temporarily stalled by the ambient density at a radius of order 0.1 pc. These ultracompact HII regions are especially luminous in far-infrared continuum radiation from heated dust.

O and B stars also disrupt clouds through their high-velocity winds. Here, the ionized gas is accelerated outward by radiation pressure. Wind-emitting stars have excess radio emission with a characteristic value of the spectral index. They also produce fluctuating X-rays that stem from internal wind shocks. Younger, embedded objects drive bipolar molecular outflows that are both broader and more massive than those from lower-luminosity stars.

Any disk within an HII region suffers photoevaporation. If the source of ultraviolet flux is the disk's own central star, the inner part of the structure is kept intact by stellar gravity, while the outer part flows off as a pressure-driven wind. Low-mass stars in Orion near θ^1 Ori C have evaporating disks known as proplyds that glow optically. Parsec-scale cloud globules within other HII regions are similarly destroyed. Radiation squeezes a globule, and the enhanced density stimulates production of new, low-mass stars. Finally, the smallest globules are pushed away from an ionizing star by the recoil from their evaporative flows.

Cloud material not dissipated during a star's main-sequence life by the wind or radiation is impacted by the terminal supernova event. A dense shell traveling from the exploded star creates a shock in the cloud, observed through broad emission lines. The interior of this shell glows in X-rays. Analogous structures, but on a much larger scale, are created by multiple, contiguous supernovae. These supershells strongly perturb gas throughout the Galaxy. An individual supernova is generally disruptive, but can drive a relatively small cloud into collapse. Such rapid implosion is more likely to result in cloud breakup than production of a new star.

Suggested Reading

Section 15.1 The basic properties of ionization fronts and HII regions are presented in

Dyson, J. & Williams, D. A. 1997, *Physics of the Interstellar Medium* (Bristol: Institute of Physics), Chapter 7

Spitzer, L. 1978, *Physical Processes in the Interstellar Medium* (New York: Wiley), Chapter 5.

For a more detailed treatment of atomic processes within HII regions and a discussion of their emitted radiation, see

Osterbrock, D. E. 1989, *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei* (Mill Valley: University Science Books).

Section 15.2 The first comprehensive study of ultracompact HII regions was that of

Wood, D. O. S. & Churchwell, E. 1989, ApJSS, 69, 831.

The lifetime problem had earlier been pointed out by

Dreher, J. W., Johnston, K. J., Welch, W. J., & Walker, R. C. 1984, ApJ, 283, 632.

For the stifling of ionization due to accretion, see

Yorke, H. W. 1983, in *Birth and Infancy of Stars*, eds. R. Lucas, A. Omont, and R. Stora (Amsterdam: North Holland), p. 645.

The properties of hot cores have been reviewed by

Kurtz, S., Cesaroni, R., Churchwell, E., Hofner, P., & Walmsley, C. M. 2000, in *Protostars and Planets IV*, eds. V. Mannings, A. P. Boss, and S. S. Russell (Tucson: U. of Arizona Press), p. 299.

Section 15.3 The radiative acceleration of winds is treated in

Lamers, H. J. G. L. M. & Cassinelli, J. P. 1999, *Introduction to Stellar Winds* (New York: Cambridge U. Press), Chapter 8.

For the theory of radio emission from massive winds, see

Felli, M. & Panagia, N. 1981, AA, 102, 424.

A review of molecular outflows from high-mass stars is

Beuther, H., Schilke, P., Menten, K. M., Walmsley, C. M., & Sridharan, T. K. 2002, in *The Earliest Stages of Massive Star Birth*, ed. P. A. Crowther, (San Francisco: ASP), p. 341.

Section 15.4 Our treatment of the erosion of disks around massive stars follows the studies by

Kahn, F. D. 1969, Physica, 41, 172

Hollenbach, D., Johnstone, D., Lizano, S., & Shu, F. 1994, ApJ, 428, 654.

For observations of the Orion proplyds and theoretical interpretation, see

McCaughrean, M. J. & O'Dell, C. R. 1996, AJ, 111, 1977

Johnstone, D., Hollenbach, D., & Bally, J. 1998, ApJ, 499, 758.

The rocket effect was discovered by

Oort, J. H. & Spitzer, L. 1955, ApJ, 121, 6.

Section 15.5 The compression of interstellar clouds by a stellar radiation field is the subject of

Dyson, J. E. 1973, AA, 27, 459

Bertoldi, F. 1989, ApJ, 346, 735.

For the physics of supernova remnants, see

Woltjer, L. 1972, ARAA, 10, 129.

A multi-wavelength study of the interacting remnant IC 443 was conducted by

Mufson, S. L., McCollough, M. L., Dickel, J. R., Petre, R., White, R., & Chevalier, R. 1986, AJ, 92, 1349.

Supershells in HI were discovered by

Heiles, C. 1979, ApJ, 229, 533.

The astrophysical significance of extinct radionuclides in meteorites is explored in

Zinner, E. 1998, AREPS, 26, 147.

Busso, M., Gallino, G. J., & Wasserburg, G. J. 1999 ARAA, 37, 239.

The last reference presents the case for triggered formation of the solar system by the wind from a nearby giant star.

Part V

Pre-Main-Sequence Stars

The Formation of Stars
Steven W. Stahler and Francesco Palla

16 Quasi-Static Contraction

The pre-main-sequence phase of stellar evolution marks both an end and a beginning. On the one hand, it represents the last period of a star's youth, before the object enters a protracted epoch of hydrogen fusion. It is also true that the star, as it begins pre-main-sequence contraction, is no longer buried within an opaque dust cloud. For the first time, radiation can emanate freely from the surface layers. The star thus behaves in some respects like a mature, largely stable object and is relatively easy to observe by traditional means. A number of phenomena, however, attest to the fact that youthful activity has not ceased altogether. It is these peculiarities that have kept pre-main-sequence stars a major focus of astronomical research for over half a century.

We begin our study, in the present chapter, by exploring this evolutionary phase from a theoretical perspective. After examining the physical origin of the stellar birthline, we address the central issue of how contraction persists over a relatively long period. We also describe the thermal and nuclear changes that occur along the way. The ignition of hydrogen ends contraction for normal stars. In brown dwarfs, which we also discuss, the central temperature is generally high enough to burn deuterium but not ordinary hydrogen, and contraction must eventually be halted by electron degeneracy pressure. In both cases, the depletion of another light element, lithium, provides a useful chronometer.

Another issue of prime concern is stellar rotation. All stars are born with some finite angular momentum, and the question of how their spin rate changes during contraction is not yet fully understood. Hence, we will rely heavily on the available observations to guide us through this developing topic. In the remaining chapters of Part V, we will summarize our current understanding of both T Tauri and Herbig Ae/Be stars.

16.1 The Stellar Birthline

Throughout this book, we have frequently utilized the birthline as a theoretical construct within the HR diagram. We presented the curve as the locus from which pre-main-sequence tracks descend, as well as the fiducial zero-point for stellar ages. But what is the motivation, from an observational viewpoint, for introducing this concept? And once the idea is accepted, with what precision can we pinpoint the curve's location, from either observation or theory? We saw in previous chapters that the birthline is not relevant for the most massive stars, which first appear on the main sequence. What is the stellar physics at the root of this distinction? One task in this chapter must be to address these questions.

16.1.1 Empirical Evidence

The idea of the birthline stems not from the observations of individual stars, but from surveys of stellar groups. Consider first T associations. These, as we described in Chapter 4, still contain

The Formation of Stars. Steven W. Stahler and Francesco Palla

ISBN: 3-527-40559-3

a substantial quantity of molecular gas, although the visual extinction toward the regions is low enough for most members to be placed in the HR diagram. A glance at Figure 4.9 shows immediately that there is a pattern to the distribution. Within each association, virtually every star is found above the main sequence. This fact, along with the presence of molecular gas, assures us that the systems are indeed young. Equally striking is the clear *upper* boundary in each distribution. Such a boundary appears in every T association with a sufficient population of well-observed members. It also appears in even richer groups containing massive stars. A good example is the Orion Nebula Cluster, whose HR diagram we displayed as Figure 12.29. The net impression is that the upper reaches of the HR diagram are somehow forbidden, *i. e.*, that pre-main-sequence stars, by their very nature, are restricted to relatively modest luminosities.

This interpretation, however, is immediately subject to a telling criticism. The contraction of pre-main-sequence stars decelerates with time. We also know that low-mass stars grow dimmer as they contract. Might not the paucity of very luminous objects be a statistical effect? Might there be no sharp upper boundary at all, but only a diminishing likelihood of detecting stars high in the diagram, simply because they spend so little time there?

To quantify this argument, we recall the characteristic time scale for pre-main-sequence contraction. This is the Kelvin-Helmholtz value:

$$t_{\rm KH} \,=\, \frac{G\,M_*^2}{R_*\,L_*} \ , \tag{16.1}$$

introduced originally as equation (1.6). The quantity $t_{\rm KH}$, we recall, measures the period required for the stellar radius to shrink appreciably. That is, the star's contraction obeys

$$\frac{dR_*}{dt} \,=\, -n_1\,\frac{R_*}{t_{\rm KH}} \ , \tag{16.2}$$

where n_1 is a dimensionless, positive constant of order unity. Later in this chapter, we will derive this equation more formally and supply the numerical coefficient. We shall also indicate the origin of another important fact concerning pre-main-sequence evolution. This is the near constancy of the star's effective temperature, especially for subsolar masses. Now the radius, effective temperature, and luminosity are related through equation (1.5):

$$L_* \,=\, 4\,\pi\,R_*^2\,\sigma_B\,T_{\rm eff}^4 \ . \tag{16.3}$$

We thus have

$$\frac{dL_*}{dt} \,=\, -n_2\,\frac{L_*}{t_{\rm KH}} \ , \tag{16.4}$$

where $n_2 \equiv 2\,n_1$. It also follows that $t_{\rm KH}$ itself varies as $L_*^{-3/2}$ for a star of fixed mass. If the number of stars in a portion of the HR diagram were indeed proportional to the length of time spent in that portion, then equation (16.4) implies that the number per logarithmic interval of luminosity would vary as $t_{\rm KH}$, *i. e.*, as $L_*^{-3/2}$.

It is a straightforward matter to test this prediction against observation. Figure 16.1 shows results for the Orion Nebula Cluster. Here, we have selected a vertical strip in the diagram of Figure 12.29. This strip is centered on $\log T_{\rm eff} \,=\, 3.52$, corresponding to a stellar mass of

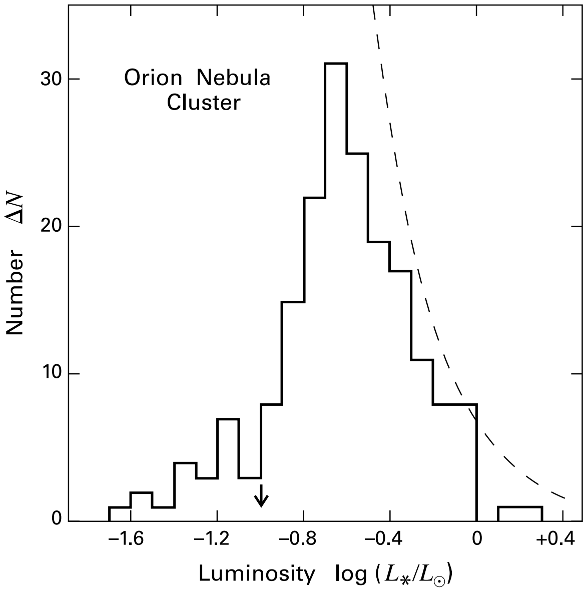


Figure 16.1 Stellar distribution by luminosity in the Orion Nebula Cluster. Plotted is the number of stars in each luminosity interval $\Delta\log\,(L_*/L_\odot) = 0.1$. The stars have $\log T_{\text{eff}} = 3.52$, within a width $\Delta\log T_{\text{eff}} = 0.03$. The survey's nominal flux limit is $0.1\ L_\odot$, as marked by the vertical arrow on the horizontal axis. The dashed curve is the prediction of a steady-state formation model.

$0.20\ M_\odot$. The histogram shows the observed number of stars, in fixed intervals of $\log L_*$. For comparison, the dashed curve is the $L_*^{-3/2}$ prediction. The two are manifestly different over the entire range of luminosities, especially at small L_*-values. What are we to make of this discrepancy?

A tacit assumption underlying our prediction is that pre-main-sequence stars appear at a steady rate. We know from Chapter 12, however, that real groups do not form stars for longer than about 10^7 yr. Thus, the eventual decline in the histogram toward *lower* luminosities, *i. e.*, older stars, is to be expected. Note that the luminosities displayed extend below the sensitivity limit for these observations, which is about $0.1\ L_\odot$. However, the decline is already well established by that point.

Of more immediate interest is the trend toward *higher* luminosities. Figure 16.1 shows a steep falloff in this direction and a sudden drop at $\log L_* \approx 0$. The production rate of many groups, including the Orion Nebula Cluster, accelerates with time through the present epoch (§ 12.4). This increased rate partially offsets the diminishing value of t_{KH} at higher luminosity. Indeed, it is clear from the figure that the decline in population is initially *less* steep than in steady state, just as expected. At a certain point, however, the number falls much more dramatically. Even under the steady-state hypothesis, about 17 additional stars would be present with $\log L_* > 0$. The observed number is only 2.

A similar, precipitous truncation in the stellar population is seen for other vertical strips in the Orion HR diagram. One possible explanation is that brighter objects would still be forming if not for the current O stars, which have stripped away the molecular gas within the recent past. However, essentially the same sharp boundary is seen for other groups, such as Lupus and Taurus-Auriga, where there are no massive stars at present. We conclude that brighter stars of low mass are not just undersampled in all these environments, but never appear at all. That is, a forbidden region in the HR diagram indeed exists.

The physical reality of the boundary is confirmed by the fact that empirical indicators of stellar youth show a marked enhancement in its vicinity. Some of these indicators, such as excess infrared continuum flux and optical emission lines, will be considered in detail in the next two chapters. Yet another sign, outflow activity, was the subject of Chapter 13. There we saw that most stars driving either jets or molecular outflows are themselves too embedded to be placed in the HR diagram. A few sources, however, exhibit optical photospheres. Their measured temperatures and luminosities invariably place them close to the birthline.

To illustrate the point, we present in Figure 16.2 a composite HR diagram of several dozen Herbig Ae and Be stars, taken from a number of different locations. As expected, the aggregate group lies almost wholly between the birthline and the ZAMS, both of which are included in the figure. The filled symbols represent objects, covering a range of masses, that have associated molecular outflows, as detected in CO. In some cases, the optically visible star itself may be powering its own flow. In others, an unresolved, and perhaps more embedded, binary companion could be the driver. We know from Chapter 12 that both components within such a binary are likely to be coeval, so the outflow would still serve as a sign of youth for the visible star. It is evident, in any case, that stars associated with outflows occur relatively high within the distribution, and that most fall close to the upper boundary.

16.1.2 Theoretical Derivation

Within the context of stellar evolution theory, the existence of the birthline is readily understood. The pre-main-sequence objects that populate the HR diagram are the optically visible descendents of embedded protostars. The most luminous, and therefore youngest, visible stars have characteristics that reflect their origin as accreting objects within collapsing dense cores. Both the surface luminosity and temperature of a protostar are set by infall dynamics. (Recall equations (11.5) and (11.8).) These quantities are expected to change significantly once the main accretion phase has ended. On the other hand, the *radius* of the object is determined by its internal structure, *i. e.*, by the balance of self-gravity and thermal pressure. This property is insensitive to surface infall and is the same for a protostar and a very young pre-main-sequence star of identical mass. The birthline is simply the locus in the HR diagram of pre-main-sequence stars with protostellar radii.

Once this essential point is appreciated, it is easy to see why visible stars of very high luminosity are not present, *i. e.*, why there is a forbidden region. The theoretical calculations described in Chapter 11 find that protostellar radii are greater than main-sequence values, but never by more than a factor of 10. (Recall Figure 11.6.) The luminosity of a pre-main-sequence star obeys the photospheric relation of equation (16.3). As we shall detail shortly, the surface properties of such objects result in effective temperatures that, for subsolar masses, are slightly

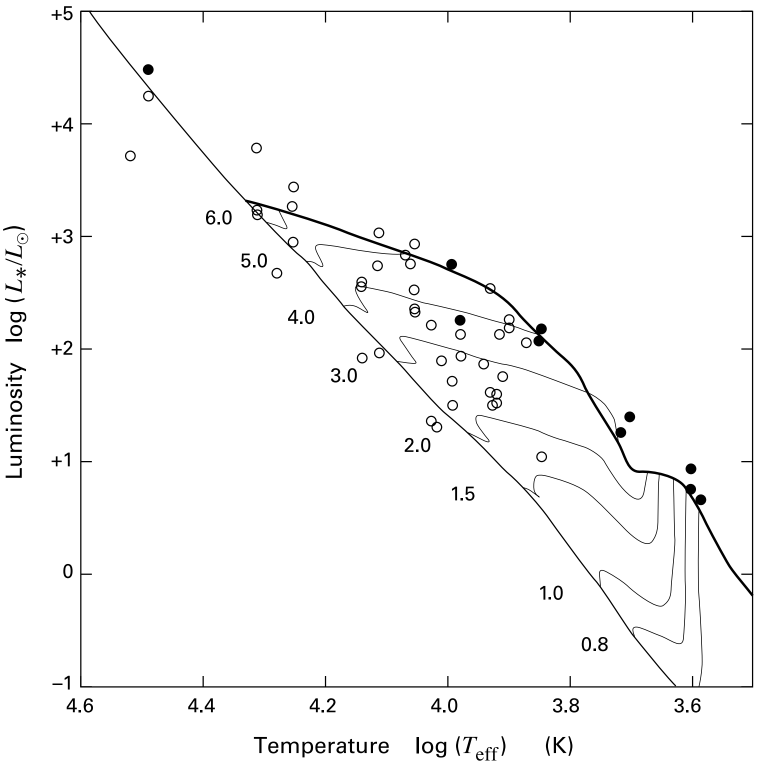


Figure 16.2 The distribution in the HR diagram of visible stars driving molecular outflows. The open circles represent Herbig Ae/Be stars from a variety of star-forming regions. Filled circles are that subset associated with CO emission. Also shown are the theoretical birthline, the zero-age main sequence, and pre-main-sequence evolutionary tracks. Masses for the latter, in solar units, are indicated.

less than main-sequence. It follows that the initial luminosities of pre-main-sequence stars cannot exceed those along the ZAMS by more than two orders of magnitude. The L_*-values are modest compared to those achieved in later stellar evolution (recall Figure 11.5).

Theoretical construction of the birthline begins by considering a sequence of stellar models, each taken to represent the youngest pre-main-sequence star of its respective mass. One solves, for a given M_*-value, the four stellar structure equations. The radius enters when imposing the appropriate boundary conditions. (See § 16.2 below.) Thus, one surface condition is equation (16.3), where $R_*(M_*)$ is adopted from protostar theory. Solution of the structure equations gives $T_{\rm eff}$, and therefore also L_*, at each mass. The birthline we have displayed so far uses the mass-radius relation from a spherical calculation of protostellar infall, in which the accretion rate $\dot{M}$ was set equal to the constant value $1 \times 10^{-5}\,M_\odot$ yr^{-1}. Table 16.1 lists the essential birthline properties under this assumption. The fifth column gives the period of active deuterium burning in the subsequent, pre-main-sequence phase (see § 16.3 below). Finally, we have listed the time $t_{\rm ZAMS}$ required for each mass to contract all the way to the main sequence.

Table 16.1 The Theoretical Birthline

Mass ($M_\odot$)	Radius ($R_\odot$)	log L_* ($L_\odot$)	log $T_{\rm eff}$ (K)	Δt_D (yr)	$t_{\rm ZAMS}$ (yr)
0.1	2.49	-0.28	3.49	1.5×10^6	3.7×10^8
0.2	2.52	-0.01	3.52	8.5×10^5	2.4×10^8
0.4	2.70	$+0.27$	3.56	3.0×10^5	1.1×10^8
0.8	4.32	$+0.78$	3.61	2.7×10^4	5.2×10^7
1.0	4.92	$+0.85$	3.63	6.9×10^3	3.2×10^7
1.5	5.09	$+0.89$	3.65	0	1.2×10^7
2.0	4.94	$+0.90$	3.67	0	8.4×10^6
3.0	5.66	$+0.94$	3.70	0	2.0×10^6
4.0	10.2	$+2.09$	3.84	1.4×10^4	8.2×10^5
5.0	8.20	$+2.83$	4.05	8.3×10^3	2.3×10^5
6.0	4.62	$+3.24$	4.27	1.1×10^3	2.9×10^4
7.0	3.28	$+3.40$	4.32	7.0×10^1	8.5×10^3
8.0	3.11	$+3.55$	4.36	0	0

The predicted birthline location clearly depends on the results of *both* protostar and pre-main-sequence theory. Deferring consideration of the latter, how sensitive is this location to the infall model? Given the considerable uncertainty regarding the infall phase, one must hope that the precise value of $\dot{M}$, its temporal behavior, or even the geometric pattern of infall, do not strongly influence the curve's location. Fortunately, this appears to be the case.

Consider the value of $\dot{M}$, assumed for the moment to be temporally constant. We saw in Chapter 10 how rates of 10^{-6} to $10^{-5}\ M_\odot\ {\rm yr}^{-1}$ are expected from the dynamics of inside-out collapse, in conjunction with the measured temperatures of dense cores. (The central relation connecting $\dot{M}$ and this temperature is equation (10.31).) But how does the birthline itself respond to a shift in the infall rate? The left panel of Figure 11.8 shows that the protostellar radius, at a given mass, diminishes for lower $\dot{M}$. Slower buildup of its mass gives the protostar more time to radiate internal energy and therefore shrink. Nevertheless, over the indicated range in $\dot{M}$, the radius change is severely restricted at subsolar masses. The reason, we recall, is the thermostatic effect of deuterium fusion. Although present in small quantities, this fuel effectively constrains the radius by altering the nuclear energy output during the main accretion phase. In calculations where deuterium fusion is artificially suppressed, the mass-radius curve undergoes a larger shift when $\dot{M}$ is altered. (See the right panel of Figure 11.8.)

Figure 16.3 shows the effect on the birthline of altering the infall rate, for the same $\dot{M}$-range as in Figure 11.8. As expected, inclusion of deuterium assures a relatively narrow spread in the predicted birthline. In contrast, the maximum luminosities for pre-main-sequence stars would have a larger dispersion in the hypothetical situation of zero deuterium abundance. To summarize, we now have one clue why the observed stellar groups show a rather sharp boundary in the HR diagram. This reasoning implies, incidentally, that any rise in the estimated deuterium abundance relative to hydrogen in the interstellar medium, a matter of ongoing investigation and here taken to be $[{\rm D/H}] = 2.5 \times 10^{-5}$, would tighten the theoretical boundary even more.

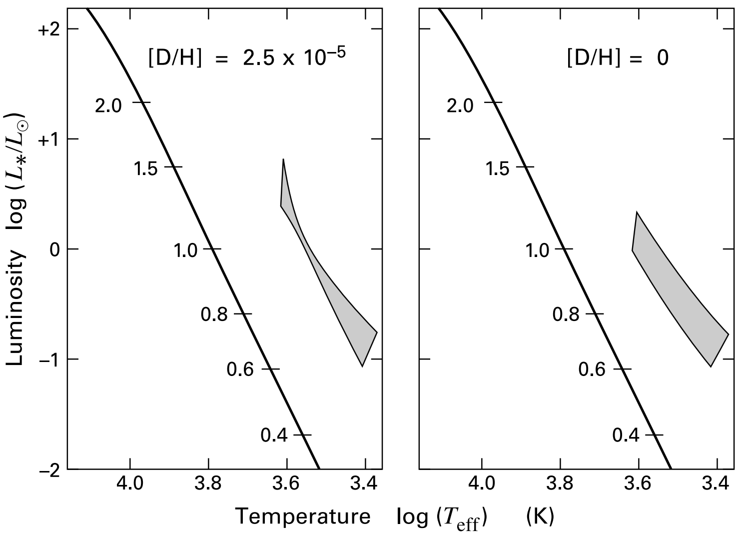


Figure 16.3 Effect on the theoretical birthline of changes in the mass accretion rate $\dot{M}$. The upper curve in each case is the birthline constructed with $\dot{M} = 1 \times 10^{-5}\, M_\odot\, {\rm yr}^{-1}$, while the lower curve corresponds to $2 \times 10^{-6}\, M_\odot\, {\rm yr}^{-1}$. In the left panel, the calculations assume the full interstellar deuterium abundance, while this abundance has been artificially set to zero in the right panel.

The other issues we raised — the temporal behavior of $\dot{M}(t)$ and the spatial pattern of infall — have not yet been addressed through detailed calculations. A few general remarks, however, are in order. The birthline concept rests on the supposition that a pre-main-sequence star's initial radius is that inherited from the infall epoch. Thus, the main accretion phase is assumed to end quickly compared to the characteristic time $t_{\rm KH}$. If this were *not* the case, the star would begin to contract even while gaining mass at a slowly diminishing rate. Theorists are still far from understanding, even in a qualitative fashion, the mechanism by which infall ends. However, the observations themselves offer valuable insight. In Chapter 17, we shall examine the evidence for continued mass addition onto visible T Tauri stars. While such late infall is almost certainly present, the estimated rate, even for stars relatively close to the birthline, is well under the protostellar value of equation (10.31). Hence, the hypothesis of a rapid decrease of infall is at least consistent with observations.

By the spatial pattern of infalling gas, we refer primarily to the effects of angular momentum. It is not difficult to determine the density and streaming velocity within a rotating, collapsing envelope, as we did in Chapter 11. The real problem, still outstanding, is the fate of this material after it lands in the disk and how it is ultimately transported to the protostellar surface. We argued in § 11.4 that the deposition of specific entropy on the surface should be insensitive to the details of this transport, but the calculation has yet to be done. As an extreme limit, one may replace the thermal boundary condition appropriate for spherical infall, equation (11.25), by the photospheric one in equation (16.3). At fixed $\dot{M}$, the effect of even this change on protostellar evolution should be minor, at least in the subsolar mass regime. The reason is that central deuterium fusion has more influence on the star's internal entropy than the surface condition.

16.1.3 Intersection with the ZAMS

If the protostar continues to increase its mass, other effects come into play. We described in Chapter 11 how deuterium begins to fuse within an interior shell, and how this burning dramatically swells the protostar's radius (Figure 11.21). At even higher mass, gravitational contraction becomes strong, and the protostar rapidly shrinks. These events occur once the star's internal luminosity, given by $\langle L_{\rm int} \rangle$ in equation (11.44), becomes so high that it dominates $L_{\rm acc}$.[1] Both processes clearly operate independently of the star's surface properties. Hence, the precise thermal boundary condition also has little impact on the radii of higher-mass protostars, and thus on the birthline in this regime.

The form of the birthline naturally reflects the evolutionary developments just described. Referring again to the curve in Figure 16.2, we note how, with increasing $T_{\rm eff}$, the boundary veers upward in the neighborhood of $\log T_{\rm eff} = 3.70$, corresponding to $M_* = 3\, M_\odot$. This change, which allows the birthline to encompass observed Herbig Ae and Be stars, results from the swelling induced by deuterium shell burning. For even higher $T_{\rm eff}$, the upward rise sags, as the protostar's radius begins to fall. Finally, at $\log T_{\rm eff} = 4.33$, corresponding to $8\, M_\odot$, the curve intersects the ZAMS.

Consider the physical significance of this intersection point. A star that has attained the mass in question is joining the main sequence just as infall ends. That is, the object begins to fuse hydrogen in its center at about the same time that it becomes optically revealed. Objects of lower mass are still too cold to fuse hydrogen following their main accretion phase. These optically visible stars thus enter a period of quasi-static contraction — the standard pre-main-sequence phase. Conversely, stars more massive than the intersection value have no pre-main-sequence phase at all. They are either embedded protostars or, following the end of infall, main-sequence objects.

Observations of the Orion Nebula Cluster (Figure 12.29) and other populous groups support this theoretical prediction. In the Orion HR diagram, we see a string of luminous objects above the intersection; these fall, at least approximately, along the ZAMS. It is important to bear in mind that the evolution of such objects *prior* to their main-sequence appearance must be very different from that of low- and intermediate-mass protostars. As we argued in Chapter 12, the time scale for their accumulation has to be greatly reduced to avoid the destructive effects of winds and radiation pressure. The actual interval over which mass buildup occurs has little bearing on their interior structure, in contrast to birthline objects. Instead, the properties of the newly visible star, following the dissipation of transients, are set entirely by the need to maintain hydrogen fusion at a steady rate.

---

[1] Equivalently, the Kelvin-Helmholtz time associated with $\langle L_{\rm int} \rangle$, which governs the rate of internal contraction, becomes less than the evolutionary time scale $M_*/\dot{M}$.

16.2 The Contraction Process

Although the main accretion phase involves dynamical motion, the protostar itself is always in a state of hydrostatic equilibrium. Its central region is also hot enough to power nuclear fusion, at least for part of the time. What distinguishes this entity from more mature objects is the rapid, continual addition of fresh material from the surrounding cloud. With the end of infall, this critical difference vanishes. The further evolution of the star, now of fixed mass, is driven almost entirely by radiation from its surface layers. Let us now examine more closely the course of events.

16.2.1 Tracking the Evolution

Radiation drains away internal energy. Since the object is gravitationally bound, its total energy, $E_{\rm tot}$, is negative and becomes more so with the passage of time. The virial theorem tells us, however, that the thermal contribution, U, is $-E_{\rm tot}$. Thus, U actually *increases*, as does the internal temperature. A pre-main-sequence star is therefore an object with negative heat capacity, one whose temperature rises as a result of heat loss. This behavior, odd by terrestrial standards, occurs because of the increased gravitational binding.

The energy loss from radiation causes the star to shrink with the characteristic time scale $t_{\rm KH}$. For the evolution to be quasi-static, the stellar interior must continually readjust to maintain force balance. This readjustment is achieved through pressure disturbances and takes place over the sound travel time, $t_s \equiv R_*/a_s$, where a_s is an appropriately averaged sound speed. The justification for treating the star at each instant as a hydrostatic object is that $t_{\rm KH} \gg t_s$ throughout the pre-main-sequence phase.

It is worthwhile to check this inequality through a numerical example. A $1\,M_\odot$ star near the birthline has, according to Table 16.1, a radius of $4.92\ R_\odot$. From equation (11.2a), the volume-averaged sound speed is

$$\begin{aligned} a_s &= \sqrt{\frac{\gamma\,\mathcal{R}\,T}{\mu}} \\ &= \sqrt{\frac{\gamma\,G\,M_*}{3\,R_*}}\ , \end{aligned} \tag{16.5}$$

which is 150 km s^{-1} for $\gamma = 5/3$. Thus, t_s is only 6.5 *hours*. Since $L_* = 7.1\ L_\odot$, the Kelvin-Helmholtz time is 8.7×10^5 *years*. Moreover, the ratio $t_{\rm KH}/t_s$ varies as $R_*^{-5/2}$ and so increases even more as contraction proceeds.

Numerical calculation of pre-main-sequence evolution utilizes the four stellar structure equations, along with the equation of state for an ideal gas, equation (11.16). As developed in Chapter 11, the first three structure equations are

$$\frac{\partial r}{\partial M_r} = \frac{1}{4\,\pi\,r^2\,\rho} \tag{16.6a}$$

$$\frac{\partial P}{\partial M_r} = -\frac{G\,M_r}{4\,\pi\,r^4} \tag{16.6b}$$

$$\frac{\partial L_{\rm int}}{\partial M_r} = \epsilon - T\,\frac{\partial s}{\partial t}\ , \tag{16.6c}$$

where we have suppressed all subscripts in the partial derivatives. In radiatively stable regions of the star, the fourth equation is

$$T^3\,\frac{\partial T}{\partial M_r} = -\frac{3\,\kappa\,L_{\rm int}}{256\,\pi^2\,\sigma_B\,r^4}\ . \tag{16.7}$$

This is replaced by

$$\frac{\partial s}{\partial M_r} = 0 \tag{16.8}$$

in convection zones within the deep interior. Convection in the *surface* layers of a pre-main-sequence star is inefficient, *i. e.*, $s(M_r)$ falls with a non-negligible slope. Here, the standard procedure is to use mixing-length theory to supply the relation between $\partial s/\partial M_r$ and $L_{\rm int}$.

Of the four boundary conditions needed to solve the four, first-order differential equations, three are already familiar:

$$r(0) = 0 \tag{16.9a}$$

$$L_{\rm int}(0) = 0 \tag{16.9b}$$

$$L_{\rm int}(M_*) = 4\,\pi\,R_*^2\,\sigma_B\,T_{\rm eff}^4\ . \tag{16.9c}$$

Note that equation (16.9c) does *not* simply serve as the definition of $T_{\rm eff}$. The equation relates three dependent variables: $L_{\rm int}$, r, and T. The latter is equal to $T_{\rm eff}$ only at the stellar photosphere, which serves as the outer boundary of the calculation.

The fourth boundary condition, which replaces equation (11.23) for protostars, concerns the pressure at the photosphere. Dividing equation (16.6b) by (16.6a), we obtain an expression for $\partial P/\partial r$ (see equation (11.14)). Spatial integration from the photosphere outward then yields

$$P_{\rm phot} = \frac{G\,M_*}{R_*^2}\int_{R_*}^{\infty}\rho\,dr\ . \tag{16.10}$$

We have assumed in our derivation that M_r and r have reached their final values at the photosphere, but that $\rho(r)$ continues to fall beyond that point. If the temperature decline is rather shallow, we may rewrite equation (16.10) approximately as

$$P_{\rm phot} \approx \frac{G\,M_*}{R_*^2\,\kappa_{\rm phot}}\int_{R_*}^{\infty}\rho\,\kappa\,dr \tag{16.11a}$$

$$= \frac{G\,M_*\,\Delta\tau}{R_*^2\,\kappa_{\rm phot}}\ . \tag{16.11b}$$

Here, $\kappa_{\rm phot}$ is the Rosseland mean opacity evaluated at the photosphere itself. The optical thickness $\Delta\tau$ is that seen by a photon propagating radially outward from that surface. This quantity is 2/3, according to the Eddington approximation (Appendix G). Thus, our final boundary condition is

$$P(M_*) = \frac{2\,G\,M_*}{3\,R_*^2\,\kappa_{\rm phot}} \ . \tag{16.12}$$

In stars of the lowest mass, the opacity in the outer stellar layers is especially complex because of the presence of molecules. The most accurate evolutionary calculations supplement the structure equations by integrating the full, frequency-dependent radiative transfer equation within a thin, plane-parallel atmosphere. The calculated temperature and pressure at the base of that atmosphere then supply the two outer boundary conditions in place of equations (16.9c) and (16.12). Since each atmosphere calculation requires knowledge of the radiative flux $F_{\rm rad}$ and the surface gravity $g \equiv GM_*/R_*^2$, the solution is iterative.

Following the evolution numerically for any mass requires that we specify the initial conditions of that object. These include the starting radius, which we obtain from protostar theory. In addition, we must prescribe the star's thermal structure, *i. e.*, the distribution of specific entropy, $s(M_r)$. This function is also obtained from the appropriate protostellar model. Over any subsequent time interval, one advances the entropy through the heat equation (16.6c) and then establishes the density and temperature distribution through spatial integration of the remaining three equations.

16.2.2 Fully Convective Phase

Table 16.2 gives numerical results for three representative stellar masses. Here we limit ourselves to low-mass objects. The special issues encountered for heavier stars will be covered in Chapter 18. The table displays the temporal evolution of a number of physical quantities. We have denoted by $M_{\rm con}$ the mass within any convection zone that is present. The luminosity $L_{\rm nuc}$ is the contribution from nuclear reactions:

$$L_{\rm nuc} \equiv \int_0^{M_*} \epsilon \, dM_r \ . \tag{16.13}$$

Early in the contraction, deuterium fusion releases energy, while hydrogen burning becomes significant later on. At the initial time, conditions are identical to those on the birthline, as given in Table 16.1. The final time listed for each mass is close to $t_{\rm ZAMS}$.

These numerical solutions exhibit two striking characteristics. First, the surface temperature is very nearly constant, at least until late in the contraction. The corresponding portions of the evolutionary tracks in the HR diagram are therefore essentially vertical. Second, all three masses, except for 0.2 $M_\odot$ during a relatively brief, initial phase, begin as *fully* convective objects ($M_{\rm con} = M_*$) and remain so for substantial times. What is the underlying physical explanation for both of these features?

To understand the behavior of the photospheric temperature, we first recognize that the opacity in the star's outer layers is largely provided by the H$^-$ ion. As discussed in Appendix G, $\kappa(\rho, T)$ therefore falls substantially with even a slight decrease in temperature, in the regime below 10^4 K. But the optical depth to the photosphere must always be close to unity. This

Table 16.2 Evolution of Low-Mass, Pre–Main-Sequence Stars

$M_* = 0.2\ M_\odot$

Time (yr)	R_* ($R_\odot$)	log T_{eff} (K)	log L_* ($L_\odot$)	log T_c (K)	L_{nuc}/L_*	M_{con} ($M_\odot$)
0	2.55	3.51	+0.02	5.81	0	0.06
1×10^5	2.29	3.51	−0.02	5.92	0.83	0.20
1×10^6	1.89	3.51	−0.57	6.08	0.12	0.20
3×10^6	1.30	3.52	−1.11	6.30	0	0.20
1×10^7	0.61	3.52	−1.40	6.48	0	0.20
3×10^7	0.41	3.52	−1.78	6.64	0	0.20
1×10^8	0.28	3.52	−2.10	6.75	0.40	0.20
2×10^8	0.26	3.52	−2.15	6.78	0.90	0.20

$M_* = 0.6\ M_\odot$

Time (yr)	R_* ($R_\odot$)	log T_{eff} (K)	log L_* ($L_\odot$)	log T_c (K)	L_{nuc}/L_*	M_{con} ($M_\odot$)
0	4.01	3.59	+0.54	6.04	0.85	0.60
1×10^5	4.00	3.59	+0.53	6.08	0.45	0.60
3×10^5	3.62	3.59	+0.43	6.17	0.09	0.60
1×10^6	2.13	3.59	−0.02	6.35	0	0.60
3×10^6	1.51	3.59	−0.35	6.51	0	0.60
1×10^7	0.90	3.58	−0.78	6.68	0	0.51
3×10^7	0.72	3.59	−0.98	6.77	0.06	0.30
9×10^7	0.58	3.61	−1.08	6.97	0.90	0.10

$M_* = 1.0\ M_\odot$

Time (yr)	R_* ($R_\odot$)	log T_{eff} (K)	log L_* ($L_\odot$)	log T_c (K)	L_{nuc}/L_*	M_{con} ($M_\odot$)
0	4.80	3.64	+0.85	6.20	0	1.00
1×10^5	4.25	3.63	+0.75	6.22	0	1.00
3×10^5	3.77	3.63	+0.62	6.26	0	1.00
1×10^6	2.59	3.63	+0.28	6.48	0	1.00
3×10^6	1.80	3.63	+0.00	6.64	0	0.78
1×10^7	1.22	3.64	−0.28	6.78	0.01	0.38
3×10^7	1.01	3.75	−0.02	7.09	0.90	0.03

requirement tightly constrains the surface temperature. Specifically, a star of fixed mass and radius has a *minimum* allowed value of T_{eff}. We shall call this value the *Hayashi temperature*, after the theorist who first demonstrated its existence.

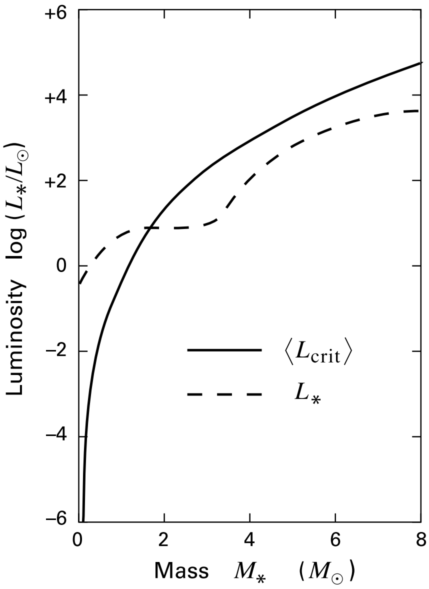


Figure 16.4 The critical luminosity for convective instability, $\langle L_\text{crit} \rangle$, displayed as a function of stellar mass. The corresponding radii are those along the theoretical birthline of Table 16.1. Also shown is the surface luminosity L_*.

If T_eff is bounded from below, then equation (16.9c) implies that the luminosity emanating from the photosphere also has a minimum. For a sufficiently large stellar radius, even this minimum exceeds that which can be transported by radiation. The latter, critical luminosity, evaluated locally at an internal mass shell, is given by equation (11.29). We are more interested in a global average of L_crit. From equation (11.44), this average may be written

$$\langle L_\text{crit} \rangle \approx 1 \, L_\odot \left(\frac{M_*}{1 \, M_\odot} \right)^{11/2} \left(\frac{R_*}{1 \, R_\odot} \right)^{-1/2} . \tag{16.14}$$

Figure 16.4 plots $\langle L_\text{crit} \rangle$ as a function of M_*, for stars along the theoretical birthline in Table 16.1. Also plotted is L_*, the actual surface luminosity from the table. For sufficiently low stellar mass, we see that L_* greatly exceeds $\langle L_\text{crit} \rangle$. Whatever the detailed, local variation of its critical luminosity, a star in this regime must be fully convective, at least following the decay of transients. If the object is initially radiative, as in the case of $0.2 \, M_\odot$, a convection zone quickly sweeps in from the surface. Figure 16.4 also indicates that a higher-mass object, with $M_* \gtrsim 2 \, M_\odot$, is radiatively stable.

We still have not explained what sets the star's true effective temperature, either along the birthline or below it. For stars that are fully convective, the surface temperature is just the minimum, Hayashi value. To see why, consider first the specific entropy, whose functional dependence on temperature and density was given in equation (11.21). The interior temperature scales as $M_* \, R_*^{-1}$. (See, *e. g.*, equation (11.2a).) Similarly, the average density is approximately $M_* \, R_*^{-3}$. We conclude that the *mean value* of the entropy is essentially fixed once we specify the star's mass and radius.

Figure 16.5 illustrates schematically the run of specific entropy, $s(M_r)$, for stars of identical M_* and R_*, but differing internal, thermal states. Within a fully convective object (*curve 1*),

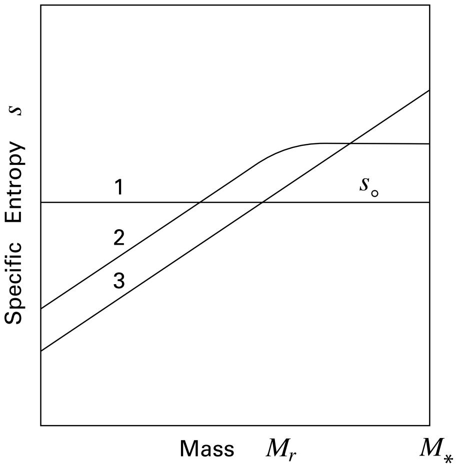


Figure 16.5 Three possible interior entropy distributions in a star of fixed mass and radius (*schematic*). Curve 1 represents a fully convective star with spatially uniform entropy $s_\circ$, curve 2 a star with radiative interior and outer convection zone, and curve 3 a fully radiative object.

$s(M_r)$ is nearly a constant value, which we denote as $s_\circ$. The entropy in a star with an outer convection zone (*curve 2*) first rises, then flattens at some interior mass. To preserve the mean value, the entropy must be less than $s_\circ$ in the deep interior, but greater at the surface. Finally, $s(M_r)$ in a fully radiative star (*curve 3*) monotonically increases and therefore attains its highest value at the outer edge. Note that the entropy increase in radiative portions is actually not linear, as shown here for simplicity.

The lesson is that any fully convective object, including a low-mass, pre-main-sequence star of sufficiently large radius, has the lowest possible surface entropy consistent with its mass and size. From equation (11.21), the combination $T^{3/2}\rho^{-1}$ must be a minimum at the boundary, and therefore $T^{5/2}P^{-1}$ must be, as well. But equation (16.12) says that P varies inversely with κ, so we need to minimize $T^{5/2}\kappa$. Since the opacity drops very steeply as the temperature falls, the configuration with the minimum surface entropy also has the lowest effective temperature.[2]

As the star contracts, its interior, nearly-constant entropy $s_\circ$ falls. From equation (11.21) and the scaling relations for temperature and density, this change is only logarithmic with the stellar radius. Thus, the photospheric entropy also decreases slowly, and the effective temperature is nearly constant. The vertical portion of the star's evolutionary curve in the HR diagram is frequently called the *Hayashi track*. If we idealize $T_{\rm eff}$ as being *precisely* fixed, then both the star's structure and evolution during this epoch may be described in a succinct manner.

---

[2] Even a "fully" convective, pre-main-sequence star is radiatively stable in its outermost layers. Moving inward, the specific entropy first drops, then climbs as hydrogen is ionized, and finally asymptotes to some interior value. The climb signifies a super-adiabatic region (recall § 11.2); the calculated thermal structure here is sensitive to the adopted model of convection. All these changes occur in a layer of negligible mass, and the photospheric entropy shifts in tandem with the asymptotic value just inside. Hence, our argument concerning the surface temperature still holds.

16.2.3 Polytropic Analysis

Spatial constancy of the specific entropy implies, from equation (11.21), that the local pressure and density obey the polytropic relation

$$P \,=\, K\,\rho^{5/3}\ . \tag{16.15}$$

Here, K, is a time-dependent coefficient related to $s_\circ$. If we divide equation (16.6b) by (16.6a) to obtain $\partial P/\partial r$, then a further differentiation yields

$$\frac{1}{r^2}\,\frac{\partial}{\partial r}\left(\frac{r^2}{\rho}\,\frac{\partial P}{\partial r}\right) \,=\, -4\,\pi\,G\,\rho\ . \tag{16.16}$$

We now substitute equation (16.15) for the pressure. The resulting equation is simplified if we define a nondimensional density and radius through

$$\rho \,=\, \rho_c\,\theta^{3/2} \tag{16.17a}$$

$$r \,=\, a\,\xi\ . \tag{16.17b}$$

Here, ρ_c is the central density. Suppose we let the coefficient a in equation (16.17b) be

$$a \,\equiv\, \left(\frac{5\,K}{8\,\pi\,G\,\rho_c^{1/3}}\right)^{1/2}\ . \tag{16.18}$$

Then the reader may verify that equation (16.16) transforms to

$$\frac{1}{\xi^2}\,\frac{d}{d\xi}\left(\xi^2\,\frac{d\theta}{d\xi}\right) \,=\, -\theta^{3/2}\ . \tag{16.19}$$

Equation (16.19) is the Lane-Emden equation, with index $n = 3/2$.[3] The equation is to be integrated subject to the initial conditions:

$$\theta(0) \,=\, 1 \tag{16.20a}$$

$$\theta'(0) \,=\, 0\ . \tag{16.20b}$$

The first condition follows from the definition of ρ_c, and the second by considering the behavior of $\partial P/\partial r$ in the limit of small radius. Figure 16.6 displays $\theta(\xi)$, as obtained by numerical integration of equation (16.19).

The function $\theta(\xi)$ contains all the information concerning the stellar interior, at least within our somewhat idealized model. The star's radius is given by

$$R_* \,=\, a\,\xi_\circ\ , \tag{16.21}$$

---

[3] In Chapter 9, we utilized the *isothermal* Lane-Emden equation, which effectively has infinite n; compare equation (9.7).

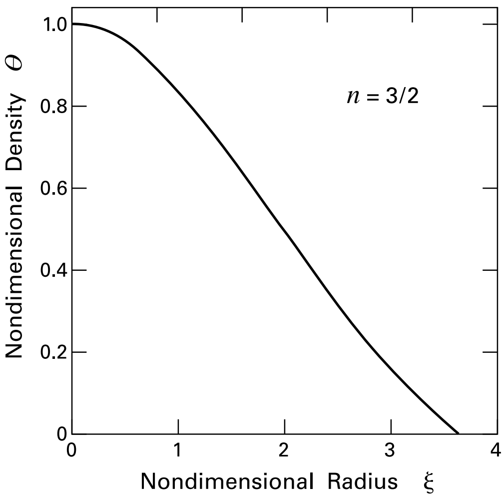


Figure 16.6 The function $\theta(\xi)$ for an $n = 3/2$ polytrope.

where $\xi_\circ$ is the point where $\theta(\xi)$ falls to zero. As shown in Figure 16.6, $\xi_\circ$ has the numerical value of 3.65. The total mass is

$$\begin{aligned} M_* &= \int_0^{R_*} 4\,\pi\, r^2 \rho\, dr \\ &= 4\,\pi\, a^3 \rho_c \int_0^{\xi_\circ} \xi^2\, \theta^{3/2}\, d\xi\ . \end{aligned}$$

If we use equation (16.19) for the integrand, we find

$$\begin{aligned} M_* &= -4\,\pi\, a^3 \rho_c \int_0^{\xi_\circ} \frac{\partial}{\partial \xi}\left(\xi^2 \frac{\partial \theta}{\partial \xi}\right) d\xi \\ &= -4\,\pi\, a^3 \rho_c \left(\xi^2 \frac{\partial \theta}{\partial \xi}\right)_\circ\ . \end{aligned} \tag{16.22}$$

The term in parentheses, which is evaluated at $\xi_\circ$, has the value –2.71.

16.2.4 Entropy Decrease

Our immediate goal is to analyze the star's contraction. To this end, we integrate the heat equation (16.6c) over mass shells. Neglecting the nuclear energy contribution ϵ as well as any spatial variation of the specific entropy, we find that

$$L_* = -\dot{s}_\circ \int_0^{M_*} T\, dM_r\ , \tag{16.23}$$

where the dot signifies a temporal derivative. Using equations (11.21) and (16.15) along with the equation of state, we find that

$$\dot{s}_\circ = \frac{3\,\mathcal{R}}{2\,\mu}\,\frac{\dot{K}}{K}\ .$$

Equation (16.18) relates K and two other time-dependent quantities: a and ρ_c. Since the latter is a numerical factor times $M_* R_*^{-3}$, we have

$$\begin{aligned}
\frac{\dot{R}_*}{R_*} &= \frac{\dot{a}}{a} \\
&= \frac{1}{2}\frac{\dot{K}}{K} - \frac{1}{6}\frac{\dot{\rho}_c}{\rho_c} \\
&= \frac{1}{2}\frac{\dot{K}}{K} + \frac{1}{2}\frac{\dot{R}_*}{R_*} \ .
\end{aligned}$$

Thus, the entropy decrease and the contraction rate are related through

$$\dot{s}_\circ = \frac{3\,\mathcal{R}}{2\,\mu}\frac{\dot{R}_*}{R_*} \ . \qquad (16.24)$$

It remains to evaluate the integral in equation (16.23). Using (16.21) and (16.22), this is

$$\int_0^{M_*} T\, dM_r = \frac{2}{5}\frac{\mu}{\mathcal{R}}\frac{G\,M_*^2}{R_*}\frac{\mathcal{I}}{\xi_\circ^3\,(\partial\theta/\partial\xi)_\circ^2} \ . \qquad (16.25)$$

Here, we have used the notation

$$\mathcal{I} \equiv \int_0^{\xi_\circ} \xi^2\,\theta^{5/2}\,d\xi \ . \qquad (16.26)$$

This nondimensional quantity may be evaluated through repeated integration by parts and use of equation (16.19). As we show in Appendix I, the result is

$$\mathcal{I} = \frac{5}{7}\xi_\circ^3\left(\frac{\partial\theta}{\partial\xi}\right)_\circ^2 \ . \qquad (16.27)$$

If we substitute this evaluation of $\mathcal{I}$ into equation (16.25), then equations (16.23) and (16.24) imply that

$$L_* = -\frac{3}{7}\frac{G\,M_*^2\dot{R}_*}{R_*^2} \ . \qquad (16.28)$$

As we know, the luminosity is a sink for the stellar energy. Thus equation (16.28) tells us that the total energy at any time is

$$E_{\rm tot} = -\frac{3}{7}\frac{G\,M_*^2}{R_*} \ . \qquad (16.29)$$

The virial theorem then implies that the gravitational and thermal energies are $2\,E_{\rm tot}$ and $-E_{\rm tot}$, respectively.

Equation (16.28) gives us the rate of stellar contraction, which we now write as

$$\begin{aligned}
\frac{dR_*}{dt} &= -\frac{7}{3}\frac{L_*\,R_*^2}{G\,M_*^2} \\
&= -\frac{7}{3}\frac{R_*}{t_{\rm KH}} \ .
\end{aligned} \qquad (16.30)$$

Thus, the coefficient n_1 in equation (16.2) is 7/3. If we evaluate L_* through equation (16.3) and idealize $T_{\rm eff}$ as a constant, we may find R_* as an explicit function of time. The quantity $t_{\rm KH}$ varies as R_*^{-3}. Thus, we may recast equation (16.30) in nondimensional form:

$$\frac{d\lambda}{d\tau} = -\frac{7}{3}\lambda^4 \, . \tag{16.31}$$

Here, we have defined

$$\lambda \equiv \frac{R_*}{R_0} \tag{16.32a}$$

$$\tau \equiv \frac{t}{t_0} \, , \tag{16.32b}$$

where R_0 and t_0 are initial (*i. e.*, birthline) values of R_* and $t_{\rm KH}$, respectively.

It is now a simple matter to integrate equation (16.31). Applying the initial condition $\lambda(0) = 1$, we find

$$\lambda = (1 + 7\tau)^{-1/3} \, . \tag{16.33}$$

Figure 16.7 compares this formula with the result of a numerical integration of the full stellar structure equations for a $0.6\ M_\odot$ star. Here, the values of R_0 and t_0 are $4.0\ R_\odot$ and 8×10^5 yr, respectively. While the agreement is generally good, the true contraction is briefly stalled at first, so that the radius stays slightly above the predicted value for some time. The initial delay is a result of internal deuterium fusion, as we will discuss shortly. At $\tau = 14$, corresponding to a dimensional time of 1×10^7 yr, the two curves touch. From that point on, the numerical curve veers away with a flatter slope. A glance at Table 16.2 shows that the star is no longer fully convective by this time. In addition, the effective temperature has begun to increase from its Hayashi value. This temperature increase is more pronounced in stars of higher mass, which also develop a radiative core earlier. The two changes are physically linked, and we should understand why they occur.

Before doing so, we stress that contraction along Hayashi tracks is of a particularly simple nature. All physical variables are functions of $\theta(\xi)$, multiplied by time-dependent, dimensional coefficients. Thus, the star appears essentially the same at each instant, except for an overall change of scale. Such *homologous* contraction owes its origin to the spatial uniformity of the specific entropy. In Chapter 12, we studied the homologous *collapse* of interstellar clouds with zero internal pressure. This highly idealized situation is useful conceptually, but never occurs in practice, both because of finite pressure effects and because a zero-pressure cloud is unstable to perturbations as it collapses. In contrast, uniform entropy within a star is enforced physically by the circulation of turbulent eddies.

16.2.5 Growth of Radiative Core

As the fully convective star shrinks, its luminosity decreases in proportion to the surface area, *i. e.* to R_*^2. The interior opacity also falls, since it is inversely related to the temperature. Thus, radiation is able to carry an increasing fraction of L_*. From equation (16.14), $\langle L_{\rm crit}\rangle$ is both proportional to $M_*^{11/2}$ and increases as $R_*^{-1/2}$. Even if $L_* > \langle L_{\rm crit}\rangle$ at the birthline, the two

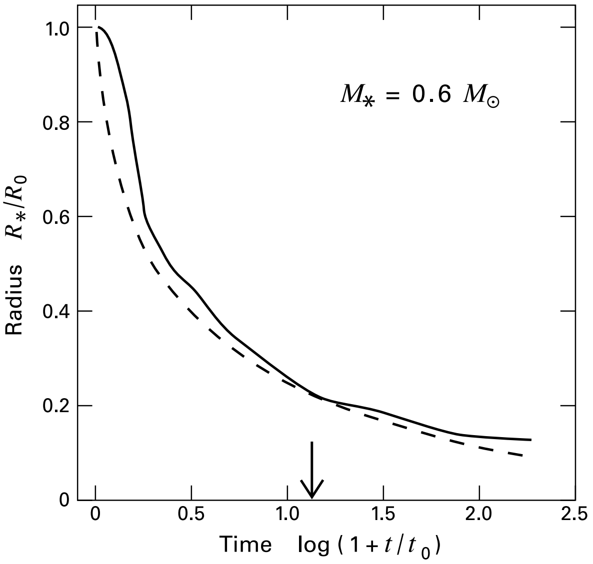


Figure 16.7 Contraction of a $0.6\ M_\odot$ pre-main-sequence star. The solid curve shows the result of a numerical integration of the full stellar structure equations, while the dashed curve is from equation (16.33) in the text. The vertical arrow marks the birth of the star's radiative core.

luminosities approach equality and eventually cross, *provided* the stellar mass is sufficiently high. If the mass is too low, the contraction ends before this crossover occurs, and the star remains fully convective throughout its pre-main-sequence life. The transitional M_*-value, for which the object starts to become radiative just as it reaches the main sequence, is $0.4\ M_\odot$.

The onset of radiative stability reminds us of the situation governing intermediate-mass protostars (§ 11.4). However, there are two important differences. For a protostar, it is the increase of M_* that causes $\langle L_{\rm crit} \rangle$ to grow; here, M_* is fixed. Additionally, the accreting object first becomes radiatively stable at some internal radius. (Recall Figure 11.18.) The analogous event occurs at the *center* of a pre-main-sequence star. In the absence of any nuclear fuel, the local value of $L_{\rm int}$ rises gently with enclosed mass M_r. Thus, when $L_{\rm crit}(M_r)$ eventually overtakes $L_{\rm int}(M_r)$, it does so even at $M_r = 0$. Convection gradually retreats to the surface layers as contraction proceeds.

During this retreat, the distribution of specific entropy resembles, at least schematically, curve 2 in Figure 16.4. The average entropy continues to fall, so the temporal *decrease* of $s(M_r)$ near the center is only partially compensated by an entropy *increase* within the flat, outer portion. But any increase of $s(M_r)$ in the surface layers pulls up the photospheric value. Consequently, $T_{\rm eff}$ rises above its previous, Hayashi level.

Figure 16.8 shows in more detail the pre-main-sequence tracks of the same masses used in Table 16.2. We have marked by open circles the points where the stars' central regions first become radiatively stable. For the $0.2\ M_\odot$ object, this event never takes place, while it occurs at $t = 6 \times 10^6$ yr and 1×10^6 yr, respectively, for $0.6\ M_\odot$ and $1.0\ M_\odot$. As convection

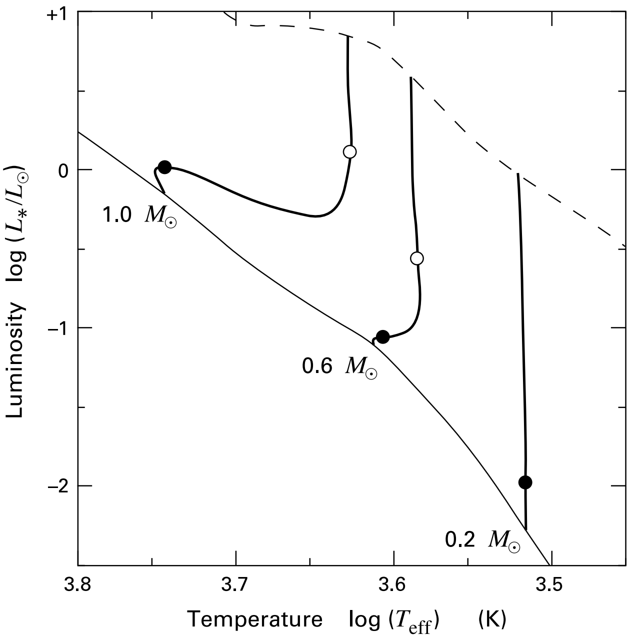


Figure 16.8 Evolutionary tracks of 0.2, 0.6, and 1.0 $M_\odot$ pre-main-sequence stars. Open and filled circles mark the onset of radiative stability and central hydrogen burning, respectively. Also shown are the birthline (*dashed curve*) and the ZAMS.

retreats in these more massive objects, the stellar luminosity becomes better approximated by equation (16.14) for $\langle L_{\rm crit} \rangle$. That is, L_* stops falling and begins to increase slowly, as $R_*^{-1/2}$. This change is evident in the figure. The more horizontal portion of the evolutionary curve is traditionally known as the *Henyey track*.

The Henyey track covers the final interval of a star's pre-main-sequence life, for masses of $0.4\ M_\odot$ and above. During this phase of the contraction, the object becomes more centrally concentrated in density than it was on the Hayashi track. Not only does the luminosity increase, but so does the star's surface temperature. From equations (16.3) and (16.14), $T_{\rm eff}$ climbs as $R_*^{-5/8}$. The *central* temperature, on the other hand, continues to increase approximately as R_*^{-1}, with consequences that we shall explore next.

Note, however, one additional property of the evolutionary curves. The ones we have displayed thus far were constructed assuming a standard solar composition for the gas. In star formation regions where the metallicity is lower, we expect the curves to change. To gauge the sense of this change, recall that the Hayashi temperature is set by the requirement that the photospheric optical depth be of order unity. With a lower abundance of heavy elements, the opacity at a given temperature falls. Material at the previous Hayashi temperature is now optically thin, and the true photosphere moves inward, to higher temperature. The full evolutionary curve, including the Henyey track, thus shifts leftward in the HR diagram. The temperature change is small, *e. g.*, about 150 K when the metallicity is halved in a 0.2 $M_\odot$ star. Nevertheless, such an alteration can be significant for quantitative work.

16.3 Nuclear Reactions

We have seen how self-gravity inexorably drives the interior temperature to higher values during pre-main-sequence contraction. The hotter interior facilitates a series of nuclear reactions, culminating in the hydrogen fusion that halts the contraction itself. Along the way, other light elements are consumed. One of these, lithium, has become an observational signature of stellar youth.

16.3.1 Exhaustion of Deuterium

Let us quantify the rise of the interior temperature. Consider first the fully convective phase. Here, the polytropic formalism allows us to derive the instantaneous distribution $T(r)$. From the equation of state, along with equations (16.15) and (16.17a), we find

$$\begin{aligned} T &= \frac{\mu}{\mathcal{R}} \frac{P}{\rho} \\ &= \frac{\mu}{\mathcal{R}} K \rho_c^{2/3} \theta \ . \end{aligned} \tag{16.34}$$

Thus, the spatial distribution of temperature is simply that of $\theta(\xi)$, as displayed in Figure 16.6. Note that the nondimensional radius ξ in that plot is $\xi_\circ$ times the ratio r/R_*.

The central value of the temperature is evidently

$$T_c = \frac{\mu}{\mathcal{R}} K \rho_c^{2/3} \ .$$

Using equations (16.21) and (16.22) for the stellar radius and mass, respectively, we find

$$T_c = n_3 \frac{\mu}{\mathcal{R}} \frac{G M_*}{R_*} \ , \tag{16.35}$$

where the coefficient is

$$n_3 \equiv -\frac{2}{5} \left(\xi \frac{\partial \theta}{\partial \xi} \right)_\circ^{-1} \ . \tag{16.36}$$

After evaluating n_3 numerically, we obtain the useful expression

$$T_c = 7.5 \times 10^6 \ \mathrm{K} \left(\frac{M_*}{1 \, M_\odot} \right) \left(\frac{R_*}{1 \, R_\odot} \right)^{-1} \ . \tag{16.37}$$

Knowledge of the contraction rate then yields the temporal evolution of T_c down the Hayashi track.

It is instructive to apply equation (16.37) to the ensemble of stars that constitute the birthline. Proceeding upward in mass, we find that the central temperature climbs, reaching 1×10^6 K by $0.4 \ M_\odot$. The temperature then hovers within a factor of two of this figure until $M_* = 1.5 \ M_\odot$. It is no coincidence that the T_c-value, over a significant mass range, is close to the ignition temperature for deuterium fusion:

$$^2\mathrm{H} + {}^1\mathrm{H} \rightarrow {}^3\mathrm{He} + \gamma \ . \tag{16.38}$$

During infall, the deuterium thermostat continually readjusts the stellar radius so that the ignition temperature is maintained at the center. Meanwhile, the deuterium abundance steadily declines below its full, interstellar value. The central region eventually becomes radiatively stable, after which it quickly consumes whatever fuel remains inside the radiative barrier. A reduced supply of deuterium continues to burn in a thick mantle that steadily retreats toward the surface (Figure 11.20).

Pre-main-sequence stars of sufficient mass thus inherit a deuterium abundance that has been reduced by prior burning in the protostellar phase. For protostars built up with $\dot{M} = 1 \times 10^{-5}\, M_\odot\, \mathrm{yr}^{-1}$, calculations find that their deuterium abundance is half the interstellar value at a mass of 0.6 $M_\odot$, while it falls to a tenth of interstellar by 0.9 $M_\odot$. The corresponding masses are smaller if $\dot{M}$ itself is lower, since the protostar then has more time to burn deuterium. *Any* amount of the fuel, however, serves to retard pre-main-sequence contraction to some degree. Integration of the heat equation (16.6c) over mass shells tells us that the entropy derivative $\dot{s}_\circ$, and therefore $\dot{R}_*/R_*$, is proportional to the reduced luminosity $L_* - L_D$, but of opposite sign.

If a protostar has appreciable deuterium-generated luminosity, then the pre-main-sequence star of the same mass has L_D close to L_* at the beginning of its contraction. Lower-mass objects initially have $L_D = 0$, but subsequent contraction raises T_c until L_D climbs to L_*. In both cases, contraction is held up (*i. e.*, $\dot{R}_* = 0$) for the period it takes to consume the deuterium. Assuming the star to be fully convective at the time of ignition, this interval is

$$\begin{aligned} \Delta t_D &= \frac{f_D\,[\mathrm{D/H}]\,X\,M_*\,\Delta E_D}{m_\mathrm{H}\,L_*} \\ &= 1.5 \times 10^6\ \mathrm{yr}\ \ f_D \left(\frac{M_*}{1\,M_\odot}\right)\left(\frac{L_*}{1\,L_\odot}\right)^{-1} . \end{aligned} \tag{16.39}$$

Here, f_D is the deuterium abundance relative to that in the interstellar medium, as found in the star just before it ignites the isotope. Our numerical version of equation (16.39) has assumed $[\mathrm{D/H}] = 2.5 \times 10^{-5}$. Recall from Chapter 11 that ΔE_D represents the amount of energy released in each fusion reaction. Note finally that the luminosity L_* is somewhat less than the birthline value if the star must contract before deuterium ignites.

Table 16.1 displays Δt_D over a range of masses. This quantity vanishes near 1 $M_\odot$, where the protostellar f_D has fallen to zero, but then resumes a finite value above 4 $M_\odot$. Pre-main-sequence stars of intermediate mass are partially radiative and burn the fuel only in their outer layers (Chapter 18). Thus, equation (16.39) can no longer be used to obtain Δt_D. It is clear, in any case, that the interval always represents a small fraction of the object's total pre-main-sequence lifetime. For the lowest-mass stars, Δt_D is about 10^6 yr and is therefore comparable to the ages of the youngest observed clusters. Some cluster members should therefore be undergoing active deuterium burning.

An historical note is of interest here. Before the advent of protostar theory, it was assumed that all pre-main-sequence stars begin contraction with their full complement of deuterium, and with such large radii that fusion is inactive. The locus in the HR diagram where ignition subsequently occurs (*i. e.*, where $T_c \approx 1 \times 10^6$ K) was termed the "deuterium main sequence." We now recognize that the fusion process can begin in the protostar phase, and that fusion helps set the initial radius for contraction. Thus, the modern birthline essentially coincides with the old deuterium main sequence over a range of subsolar masses.

16.3.2 Lithium Burning

The deuterium within young stars was produced cosmologically, with an abundance roughly twice the current interstellar value. Other light elements were created through this primordial nucleosynthesis, including helium, lithium, beryllium, and boron. By far the most useful astrophysically has been lithium. Inside stars, the dominant isotope, $^7\mathrm{Li}$, fuses with protons:

$$^7\mathrm{Li} + {}^1\mathrm{H} \rightarrow {}^4\mathrm{He} + {}^4\mathrm{He} \, . \tag{16.40}$$

In this case, the ignition temperature is about 3×10^6 K, so that the fusion event follows that of deuterium.

Relative to hydrogen, the number abundance of lithium in the interstellar medium is measured to be $[\mathrm{Li/H}] = 2 \times 10^{-9}$. This tiny fraction is itself higher than the primordial value by an order of magnitude. The increase has come largely from cosmic-ray α particles impacting the helium within interstellar clouds. Thus, the reaction that has produced most of the present-day [Li/H] is just the inverse of that in equation (16.40).

In spite of its small abundance, lithium is readily observed in stellar atmospheres. The neutral species Li I strongly absorbs in an electronic transition at 6708 Å. Figure 16.9 shows a portion of the spectrum of a T Tauri star, BP Tau, which contains the full interstellar abundance. Here we see the deep absorption trough. Note that the line is actually a doublet, $\lambda\lambda\, 6707.78, 6707.93$, but this structure is never resolved in practice. The figure also displays, for comparison, a main-sequence star of the same spectral type, in which lithium has been destroyed. Both objects show absorption in the Ca I line at 6718 Å, a prominent feature of late-type photospheres.

Low-mass protostars do not have high enough internal temperatures to burn lithium, so the corresponding pre-main-sequence stars inherit the full, interstellar supply. In protostars of higher mass, T_c indeed reaches 3×10^6 K, at which point the *central* lithium is promptly destroyed. This event occurs, however, when the protostar is already radiative in its deep interior. The temperature at the base of the convection zone is below the ignition value and subsequently declines as the zone retreats outward. (Recall § 11.4.) Thus, pre-main-sequence stars of *all* mass begin contraction with their surface lithium abundance equal to the interstellar value. The subsequent burning has no influence on contraction, but serves as an interesting test of the evolutionary tracks, as well as a stellar clock.

Figure 16.10 shows the pattern of subsequent depletion at the surface of pre-main-sequence stars, as predicted by theory. Objects within the white area, between the birthline and the ZAMS, retain their original lithium abundance. Light gray indicates that [Li/H] at the surface has declined by at most a factor of 10. Finally, stars within the darker gray zone have undergone depletion of *at least* that amount. Consider, for example, an object of $0.6\ M_\odot$. This star begins fully convective and ignites lithium at its center once its radius has shrunk by a factor of 1.7. Convection then starts to retreat toward the surface, but the base temperature remains high enough that the element continues to burn. During this period, rapid circulation of turbulent eddies ensures that the lithium abundance declines uniformly throughout the entire convection

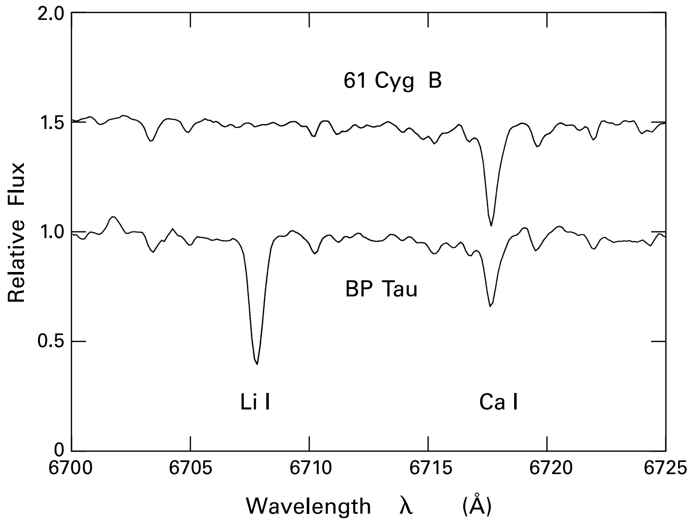


Figure 16.9 Lithium absorption in a pre-main-sequence star. Shown is a portion of the optical spectrum of BP Tau, a T Tauri star of spectral type K7, corresponding to an effective temperature of 4000 K. Also shown, for comparison, is a main-sequence star of the same spectral type, 61 Cyg B. Only in the first star do we see the Li I absorption line at 6708 Å. Both objects also have a strong line due to neutral calcium.

zone. Eventually, the base temperature falls below 3×10^6 K, but not before lithium has been fully destroyed. Such complete depletion is avoided in any star more massive than 0.9 $M_\odot$, for which convection retreats earlier. Stars with $M_* > 1.2\ M_\odot$ have such shallow convection that their surface lithium does not decline at all during the contraction phase.

One interesting aspect of Figure 16.10 is how the shaded regions cut across stellar isochrones. Consider, for example, the $t = 3 \times 10^6$ yr isochrone, shown here as a dashed curve. Within a group of pre-main-sequence stars of precisely that age, those with masses between 0.5 and 1.2 $M_\odot$ should be partially depleted in lithium, while objects outside that range should contain the full supply. The mass interval for depletion evidently widens with increasing age. Of course, any real stellar group consists of objects with a variety of ages. A cluster formed 3×10^6 yr ago still contains some stars even closer to the birthline; these will have the original, interstellar [Li/H]. Nonetheless, the prediction that all members of young groups with sufficiently low mass should suffer no depletion is a useful corroboration of theory that has yet to be directly checked through observations.

More attention has been paid to somewhat older groups with members close to the ZAMS. Here, the basic pattern of Figure 16.10 is broadly consistent with the available data. However, the observations reveal additional features whose origin is unclear. As an example, even cluster members of very similar luminosity and effective temperature may exhibit a substantial spread in their surface lithium abundance. The amount of depletion in such a star clearly depends on more than just its mass and age.

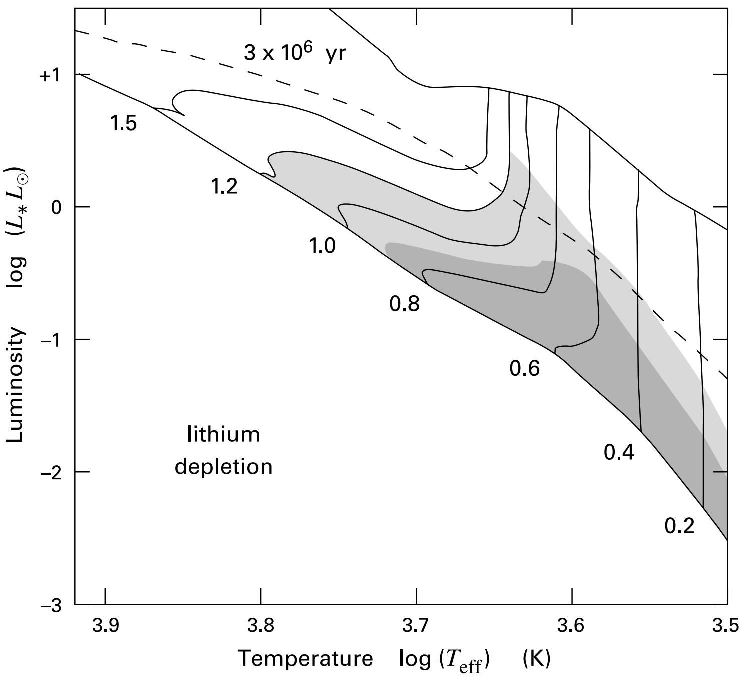


Figure 16.10 Theoretical prediction of pre-main-sequence lithium depletion. Within the white area between the birthline and the ZAMS, the surface [Li/H] is equal to its interstellar value of 2×10^{-9}. Stars in the lightly shaded region have depleted the element down to 0.1 times the interstellar value. The darker shading indicates depletion by at least this amount. Note also the masses on the ZAMS, in solar units, and the indicated isochrone.

Figure 16.11 displays [Li/H] as a function of effective temperature in the main-sequence stars of the Pleiades (*filled circles*) and the Hyades (*open circles*). The lithium abundances in both generally fall with decreasing $T_{\rm eff}$, as expected from Figure 16.10. The dispersion just mentioned is most evident in the Pleiades, which has an age of 1×10^8 yr. By the age of the Hyades (6×10^8 yr), this spread has greatly diminished. In addition, the temperature at which [Li/H] falls precipitously is now higher. Apparently, there is a process that continues transporting lithium to the hot, central region, even in main-sequence stars with radiatively stable interiors. This process has created an even lower abundance in the Sun, also shown in the figure. Finally, we note, for the Hyades, a sharp plunge in [Li/H] at a $T_{\rm eff}$-value near 6600 K. A similar *lithium dip* appears in other clusters of comparable age and in the field population.

16.3.3 Joining the Main Sequence

Whether or not a pre-main-sequence star exhausts its surface lithium, further contraction inevitably leads, once the central temperature reaches about 10^7 K, to the fusion of hydrogen at its center. The filled circles in Figure 16.8 mark the beginning of this process for the three

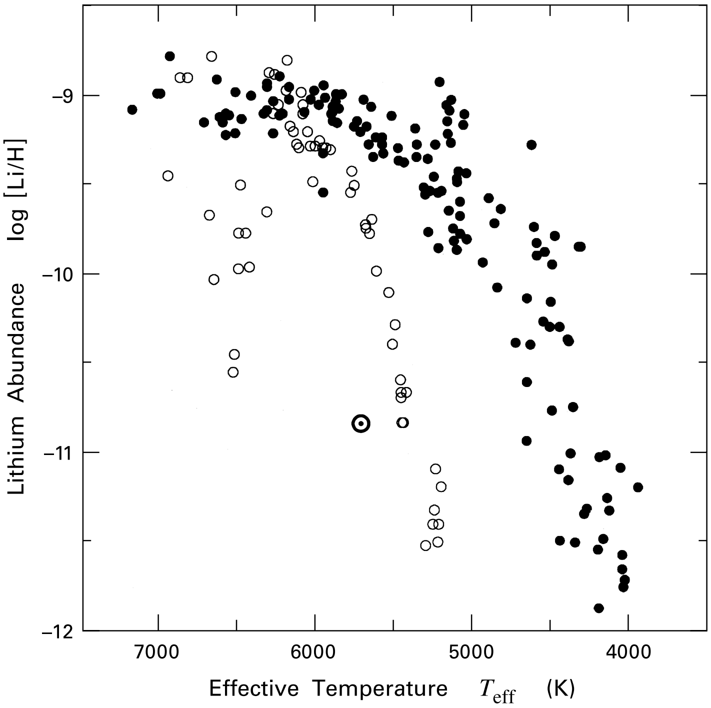


Figure 16.11 Observed values of [Li/H] as a function of effective temperature. Results are shown for the Pleiades (*filled circles*), Hyades (*open circles*), and the Sun, indicated by the usual symbol.

displayed masses. As the burning increases, the nuclear energy released can eventually balance the surface loss through radiation, and the object's radius stabilizes to its ZAMS value. In the heat equation (16.6c), the temporal derivative on the righthand side becomes negligible, and the internal luminosity is spatially constant outside the small region of active burning. (Recall Figure 1.16.) The approach to this state of thermal equilibrium is asymptotic, since increased hydrogen burning slows the very contraction process that drives up the central temperature. Thus, there is some arbitrariness in defining the total time involved. The quantity $t_{\rm ZAMS}$ in Table 16.1 represents the interval from the birthline to the point where the nuclear luminosity constitutes 90 percent of L_*.

The basic hydrogen fusion reaction consists of the production of helium:

$$4\,{}^1{\rm H} \rightarrow {}^4{\rm He} + 2\,{\rm e}^+ + 2\,\nu \ . \qquad (16.41)$$

The energy release from this process is 26.7 MeV, of which a few percent is carried off by the neutrinos and is thus unavailable for internal heating. The four protons actually never combine directly, but through a sequence of intermediate steps. We described one such reaction network, the CN cycle, in Chapter 11. Here, successive protons fuse with heavy nuclei, until the product eventually releases an α particle (see Figure 11.22). This sequence is part of the larger CNO bi-cycle, which is the dominant source of nuclear energy in main-sequence stars heavier than 1.2 $M_\odot$, *i. e.*, of spectral type F7 and earlier.

Less massive stars burn hydrogen primarily through the PP chains, also introduced in Chapter 11. As the name implies, this reaction network begins with the fusion of two protons:

$$^1\mathrm{H} + {}^1\mathrm{H} \rightarrow {}^2\mathrm{H} + \mathrm{e}^+ + \nu \ , \tag{16.42}$$

promptly followed by

$$^2\mathrm{H} + {}^1\mathrm{H} \rightarrow {}^3\mathrm{He} + \gamma \ . \tag{16.43}$$

In the dominant PPI chain, the creation of $^4\mathrm{He}$ occurs through

$$^3\mathrm{He} + {}^3\mathrm{He} \rightarrow {}^4\mathrm{He} + {}^1\mathrm{H} + {}^1\mathrm{H} \ . \tag{16.44}$$

Alternatively, the $^3\mathrm{He}$ in equation (16.43) fuses with ambient $^4\mathrm{He}$:

$$^3\mathrm{He} + {}^4\mathrm{He} \rightarrow {}^7\mathrm{Be} + \gamma \ . \tag{16.45}$$

The subsequent fate of the beryllium nucleus then determines whether the network is completed through the PPII chain

$$\begin{aligned} {}^7\mathrm{Be} + \mathrm{e}^- &\rightarrow {}^7\mathrm{Li} + \nu \\ {}^7\mathrm{Li} + {}^1\mathrm{H} &\rightarrow {}^4\mathrm{He} + {}^4\mathrm{He} \ , \end{aligned} \tag{16.46}$$

or through PPIII:

$$\begin{aligned} {}^7\mathrm{Be} + {}^1\mathrm{H} &\rightarrow {}^8\mathrm{B} + \gamma \\ {}^8\mathrm{B} &\rightarrow {}^8\mathrm{Be} + \mathrm{e}^+ + \nu \\ {}^8\mathrm{Be} &\rightarrow {}^4\mathrm{He} + {}^4\mathrm{He} \ . \end{aligned} \tag{16.47}$$

The PPII and PPIII chains together contribute at most 10 percent of the stellar luminosity for masses below 1.2 $M_\odot$.

As we have seen, all stars exceeding 0.4 $M_\odot$ approach the main sequence along the more horizontal, Henyey tracks. It is evident from Figure 16.8 that the luminosity does not increase monotonically during the final approach, but peaks and then declines shortly before the end. At the start of hydrogen burning, the star releases too much nuclear energy for thermal equilibrium. The excess heats the stellar interior and creates a central convection zone. Increasing the central entropy allows $L_r(M_r)$ to fall, so that L_* eventually drops back to its proper equilibrium value. In a 1 $M_\odot$ star, the central convection disappears at this point. However, in those stars powered by the CNO bi-cycle, this zone remains. Concurrently, any residual *outer* convection inherited from the contraction phase vanishes. This latter trend is to be expected from the decline in opacity at higher interior temperatures. In summary, more massive stars on the main sequence have a convectively unstable center and a radiatively stable envelope, while those in the range 0.4 $M_\odot < M_* <$ 1.2 $M_\odot$ have just the reverse.

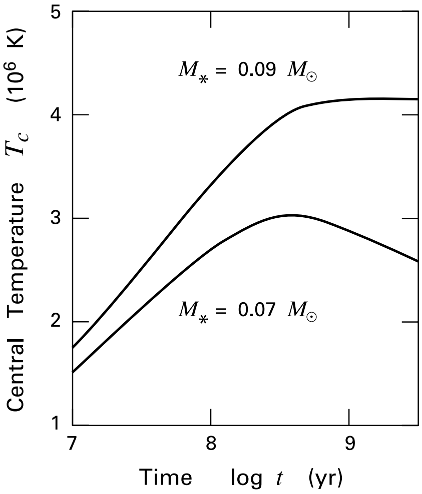


Figure 16.12 Central temperature as a function of time in contracting objects of 0.09 and 0.07 $M_\odot$.

16.4 Brown Dwarfs

Let us study in more detail objects of the lowest mass. Here we cross the border between true stars, capable of hydrogen fusion, and brown dwarfs, which are not. The latter have now been observed in abundance, an exciting development that is complementing our knowledge of stellar birth.

16.4.1 Electron Degeneracy Pressure

Lower-mass, young stars take longer to reach the main sequence as a consequence of their smaller surface areas for heat leakage. They also have higher densities at that final stage. To see why, first note that the hydrogen ignition temperature decreases, but only slowly, along with stellar mass. Thus, R_* is roughly proportional to M_* for objects along the ZAMS, at least for a limited mass range. Over this same range, the central density increases as $M_* R_*^{-3} \propto M_*^{-2}$ as we move down the main sequence to objects of lower mass and effective temperature.

Return now to the contraction phase. For an object of sufficiently low mass, the increased density means that the central temperature may *never* increase to the point of hydrogen ignition. Figure 16.12 shows what happens in more detail, as revealed through numerical calculation. Here we see that T_c in a 0.09 $M_\odot$ star increases steadily with time, until it levels off to the ZAMS value of 4.1×10^6 K. In an object of 0.07 $M_\odot$, however, T_c first increases, peaks at a lower value that creates only modest burning, then actually *declines* while contraction proceeds. The critical mass of 0.075 $M_\odot$ separates ordinary stars, which fuse hydrogen in a protracted main-sequence phase, from brown dwarfs, which may burn the element in a transient manner, but ultimately cool to zero luminosity and effective temperature.

If the central temperature can fall while the object is still shrinking, then equation (16.35) must no longer be valid. We derived this relation by applying hydrostatic balance, along with Poisson's equation for the gravitational field and the ideal equation of state. The first two continue to hold, of course, but the latter does not. At the relatively high densities of brown dwarfs, free electrons experience an effective, mutual repulsion that is independent of the ambient kinetic temperature. As the object contracts, this *electron degeneracy pressure* gradually supplants ordinary gas pressure. The internal temperature, unlike in ordinary stars, can fall as the object radiates energy into space. Eventually, degeneracy pressure fully counteracts gravity, and the radius stabilizes to a finite value.

It is not difficult to understand the quantum mechanical basis of this new effect and to derive a quantitative expression for $P_{\rm deg}$, the electron degeneracy pressure. Consider, for simplicity, a zero-temperature gas consisting *only* of electrons. The latter we may describe through plane-wave solutions of the Schrödinger equation. These solutions contain three independent quantum numbers, corresponding to the three spatial components of momentum. Additionally, one must specify the electron spin. Since the particles are fermions, the Pauli exclusion principle dictates that no two can share all four quantum numbers. At zero temperature, therefore, the particles simply fill all possible states. An elementary counting argument then shows that the number per unit volume with momentum between p and $p \, + \, dp$ is

$$dn_e \,=\, \frac{8\,\pi\,p^2\,dp}{h^3} \ . \tag{16.48}$$

In obtaining this result, we have accounted for the spin through an extra factor of 2 in the numerator on the righthand side.

Integration of equation (16.48) over all momenta yields the total electron density:

$$n_e \,=\, \frac{8\,\pi\,p_F^3}{3\,h^3} \ . \tag{16.49}$$

Here, p_F is the *Fermi momentum*, the largest value attainable for the given n_e at zero temperature. Degeneracy pressure arises because p_F can be appreciable even when the gas has no thermal energy in the classical sense. In fact, the true energy density associated with the differential population dn_e is

$$\begin{aligned} du_e \,&=\, \frac{p^2}{2\,m_e}\,dn_e \\ &=\, \frac{4\,\pi\,p^4\,dp}{m_e\,h^3} \ . \end{aligned} \tag{16.50}$$

Thus, the *total* energy density is

$$u_e \,=\, \frac{4\,\pi\,p_F^5}{5\,m_e\,h^3} \ . \tag{16.51}$$

The pressure in a nonrelativistic gas is 2/3 of this quantity, so that

$$\begin{aligned} P_{\rm deg} \,&=\, \frac{8\,\pi\,p_F^5}{15\,m_e\,h^3} \\ &=\, \frac{h^2}{20\,m_e}\left(\frac{3}{\pi}\right)^{2/3} n_e^{5/3} \ , \end{aligned} \tag{16.52}$$

where we have also utilized equation (16.49).

It is convenient to recast the electron degeneracy pressure in terms of the usual density ρ. To this end, we introduce μ_e, the mass per free electron, expressed in units of m_H. That is, μ_e obeys

$$\rho \,=\, n_e \, \mu_e \, m_H \ , \tag{16.53}$$

to which the reader may compare the analogous equation (2.2) for the mean molecular weight μ. Reasoning similar to that following equation (2.3) shows that μ_e in a fully ionized gas is nearly equal to $2/(1+X)$, or 1.2 for solar composition. Insertion of equation (16.53) into (16.52) now yields our final expression for the electron degeneracy pressure:

$$\begin{aligned} P_{\rm deg} \,&=\, \frac{h^2}{20\, m_e} \left(\frac{3}{\pi}\right)^{2/3} \frac{1}{m_H^{5/3}} \left(\frac{\rho}{\mu_e}\right)^{5/3} \\ &\equiv\, K_{\rm deg}\, \rho^{5/3} \ . \end{aligned} \tag{16.54}$$

Here $K_{\rm deg}$ has the numerical value 7.7×10^{12} in cgs units, again assuming solar composition.

Reference to equation (16.15) shows that our degenerate brown dwarf has the pressure-density relationship of an $n = 3/2$ polytrope, *i. e.*, the same as in a fully convective, ordinary star. The underlying physics is entirely different in the two cases. The star has a uniform specific entropy maintained by convection, while the degenerate object has *zero* entropy. Nevertheless, we may still utilize our previous mathematical development, simply replacing K by $K_{\rm deg}$. For example, we may combine equations (16.18), (16.21), and (16.22) to obtain a mass-radius relation for the zero-temperature configurations:

$$\begin{aligned} R_* \,&=\, \frac{5\, K_{\rm deg}}{4\,(2\pi^2)^{1/3}\, G} \left[-\left(\xi^5 \frac{\partial\theta}{\partial\xi}\right)_{\circ}\right]^{1/3} M_*^{-1/3} \\ &=\, 0.084\, R_\odot \left(\frac{M_*}{0.05\, M_\odot}\right)^{-1/3} \ . \end{aligned} \tag{16.55}$$

A more detailed structure calculation also accounts for the electrostatic attraction of the electrons for ambient ions. This *Coulomb pressure* is negative and causes the radius to deviate from equation (16.55). As we approach the realm of giant planets, *i. e.*, masses of 0.01 $M_\odot$ and below, R_* flattens and begins to decline.

We stress that equation (16.55) refers only to the *endstate* of brown dwarfs. For any finite interior temperature, the pressure is the sum of $P_{\rm deg}$ and the ideal gas expression arising from both ions and electrons. Even $P_{\rm deg}$ itself only obeys equation (16.54) at zero temperature, since otherwise the electrons are only partially degenerate. That is, they populate energy levels higher than the maximum ones allowed at zero temperature and can have momenta in excess of p_F. In time, however, the brown dwarf cools sufficiently that its radius settles to a value near our derived result.

16.4.2 Evolutionary Tracks

Figure 16.13 shows the evolution in the HR diagram of very low-mass stars and brown dwarfs, along with representative isochrones. Here, we have arbitrarily begun all tracks at

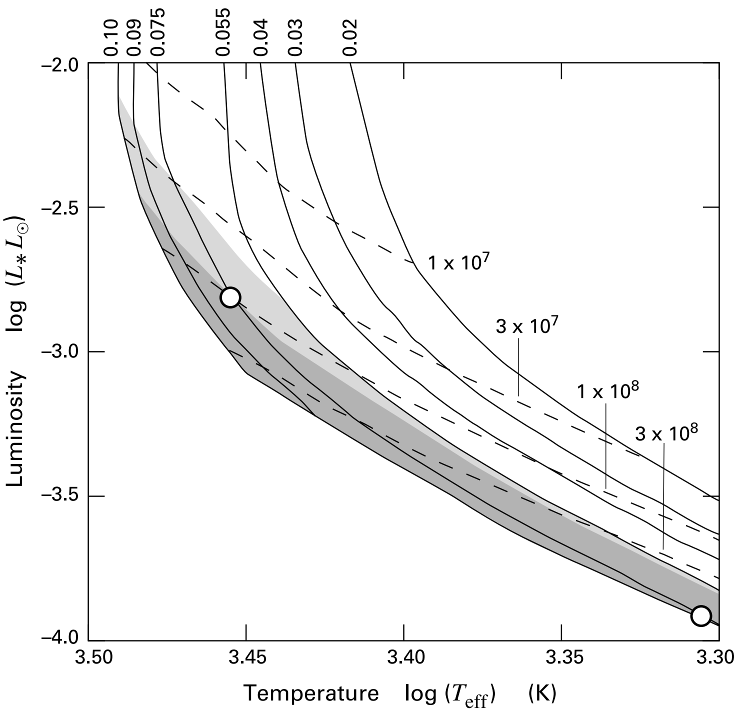


Figure 16.13 Evolutionary tracks of very low-mass stars and brown dwarfs. The corresponding masses, in solar units, and isochrones are also indicated. The light and dark shading signify the predicted depletion of surface lithium, as in Figure 16.10. Also shown is the point where a 0.075 $M_\odot$ star depletes 90 percent of its lithium (*upper open circle*), and where the object ignites hydrogen (*lower open circle*).

$\log L_* = -2$, as there are no detailed calculations for the proper starting conditions in this mass range. We see that the tracks are nearly vertical at relatively early times, just as in heavier stars. Similarly, the interiors of the present structures are fully convective. As electron degeneracy pressure begins to manifest itself, $T_{\rm eff}$ declines while L_* falls. Each track veers toward a diagonal line corresponding to fixed radius ($L_* \propto T_{\rm eff}^4$). It is also apparent that these asymptotic radii are larger for smaller masses, in accordance with equation (16.55).

Of particular interest is the curve defining the ZAMS. A star of $0.1\ M_\odot$ reaches the curve and halts contraction when $\log L_* = -3.07$.[4] For 0.09 $M_\odot$, the track approaches the ZAMS more gradually and does not reach it until $\log L_*$ has dropped to -3.25. An object of 0.075 $M_\odot$ halts contraction at $\log L_* = -3.90$ and $\log T_{\rm eff} = 3.31$. This critical point is indicated by the open circle near the bottom of the figure. All lower mass-values represent brown dwarfs. Their tracks never end, but eventually follow a constant-radius diagonal.

---

[4] The $0.1\ M_\odot$ track in Figure 16.13 differs from that in Figure 1.18, which reaches the ZAMS at higher luminosity, while the track is still nearly vertical. The calculations in Figure 16.13 include a detailed treatment of the atmosphere, which is essential at the very lowest masses.

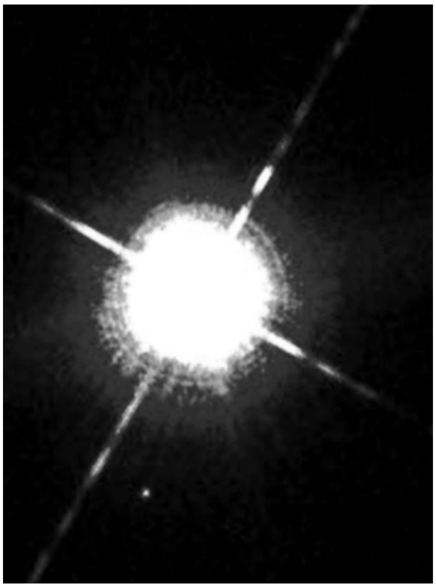


Figure 16.14 Optical image of the brown dwarf Gl 229B. Here, south is down and east is to the left. The brown dwarf is the isolated spot below the large sphere, representing the M star Gl 229A. The latter image is not the star itself, but an artifact of detector saturation. Note the prominent diffraction spikes, which also arise in the detector.

Figure 16.13 also displays the regions of light and heavy lithium depletion, using the same scheme as Figure 16.10. The upper boundary of the darker region is known as the *lithium edge*. This curve is parametrized by either the stellar mass or age, where the latter may be read off by the intersection of the boundary with various isochrones. Note that brown dwarfs of $0.055\ M_\odot$ and below never achieve a central temperature sufficient to burn lithium. Correspondingly, the upper boundary of the lightly shaded region, *i. e.*, the lithium ignition curve, touches the $0.055\ M_\odot$ track but does not cross it. Brown dwarfs of yet lower mass are still fully convective, but are eventually unable to ignite even deuterium. This boundary value of $0.013\ M_\odot$, or 14 times the mass of Jupiter, properly signifies the upper end of the giant planet regime.

16.4.3 Observed Properties

The main impediment to finding brown dwarfs is their intrinsically low luminosities. One reasonable strategy is to look for dim companions to nearby field stars. After many years of failed attempts, this approach ultimately proved successful in 1995, with the discovery of Gliese 229B. The object was found $7\overset{\prime\prime}{.}6$ from the M star now designated Gl 229A (see Figure 16.14). The distance to the latter, established through parallax, is 5.7 pc. Thus, the integrated flux of the companion translates into a bolometric luminosity. The latter is only $6 \times 10^{-6}\ L_\odot$, demonstrating unambiguously that this faint emission source is indeed substellar.

An alternative strategy that has also been fruitful is to examine the low-mass end of stellar groups. If the group is sufficiently young, the luminosities of member brown dwarfs are significantly greater than any in the field. Additionally, the distance to the group is usually known already through main-sequence fitting of the brighter members. This distance should allow one to obtain the luminosity through application of a bolometric correction. In practice, however, uncertain extinction renders L_* more problematic than $T_{\rm eff}$ in star-forming regions.

Figure 16.13 then shows that the inferred mass of the object depends critically on the cluster *age*, which may not be known with sufficient accuracy.

Observation of the lithium absorption line is of considerable value in this regard and has been critical in establishing the brown-dwarf nature of cluster members. Consider, for example, a star at the pivotal mass 0.075 $M_\odot$. This object has depleted 90 percent of its lithium when $T_{\rm eff}$ falls to 2800 K. The corresponding point in the HR diagram is marked by the upper open circle in Figure 16.13. It is clear from the figure that any cooler object (*i. e.*, one of spectral type M7 or later) must be a brown dwarf, *provided it exhibits strong lithium absorption*. More generally, the observation of lithium in an object of any surface temperature provides upper bounds to both its mass and age.

The survey of low-mass stars that revealed the first brown dwarf examined about 200 others with negative result. Thus, binaries consisting of this combination appear to be rare. Single brown dwarfs within clusters are far more common. Most of these objects have been discovered through near-infrared surveys. Such observations are helping to delineate the initial mass function in the regime below 0.1 $M_\odot$. (Recall § 4.5.) Deep surveys of the field are also finding numerous objects with the low luminosities and very cool surface temperatures expected of older brown dwarfs.

Spectroscopic examination of brown dwarfs shows peculiarities that further distinguish these objects from even the lowest-mass stars. Brown dwarfs have such low surface temperatures that dust grains, completely absent in pre-main-sequence stars, are able to grow within their atmospheres. This additional component alters the spectrum in several ways. Ordinary M stars display numerous absorption lines in the near-infrared and optical arising from rare molecular species such as TiO (titanium oxide). These molecular bands vanish at cooler temperatures, presumably because the relevant molecules are adsorbed onto dust grains. Concurrently, near-infrared lines of H_2 and CO become more prominent. Neutral atoms still in the gas phase, such as Na and K, exhibit lines that are greatly broadened by the high atmospheric pressure. To cover this ensemble of characteristics, a spectral type L has been added below the traditional M. The new type spans effective temperatures from 2200 K down to 1300 K.

In even cooler objects, including Gl 229B itself, the near-infrared is dominated by absorption lines of CH_4, which becomes the main repository of carbon. Such objects are designated spectral type T. The presence of gas-phase CH_4, along with other molecular species, would be a puzzle if dust were still present. Apparently, the low ambient temperature promotes such rapid growth of the grains that they settle out from the most tenuous region of the atmosphere. The enlarged dust continues to sink until the higher ambient temperature causes it to partially sublimate and find an equilibrium level. But grains residing in such optically thick, interior gas no longer affect the emitted spectrum. This hypothesis also helps explain the curious fact that T dwarfs exhibit bluer colors in the near-infrared than L dwarfs, despite their lower surface temperatures. In a color-color plot, of which Figure 16.15 is an example, these objects occupy a distinct region that is *not* the extension of that for M- and L-type objects.

It is important to realize that the effective temperature, and hence the spectral type, of any brown dwarf is intrinsically time-dependent. Figure 16.13 shows that all such objects should be of spectral type M ($3.35 < \log T_{\rm eff} < 3.58$) sufficiently early in their lives. As cooling proceeds, the surface inevitably enters the regime of type L, and eventually becomes type T before

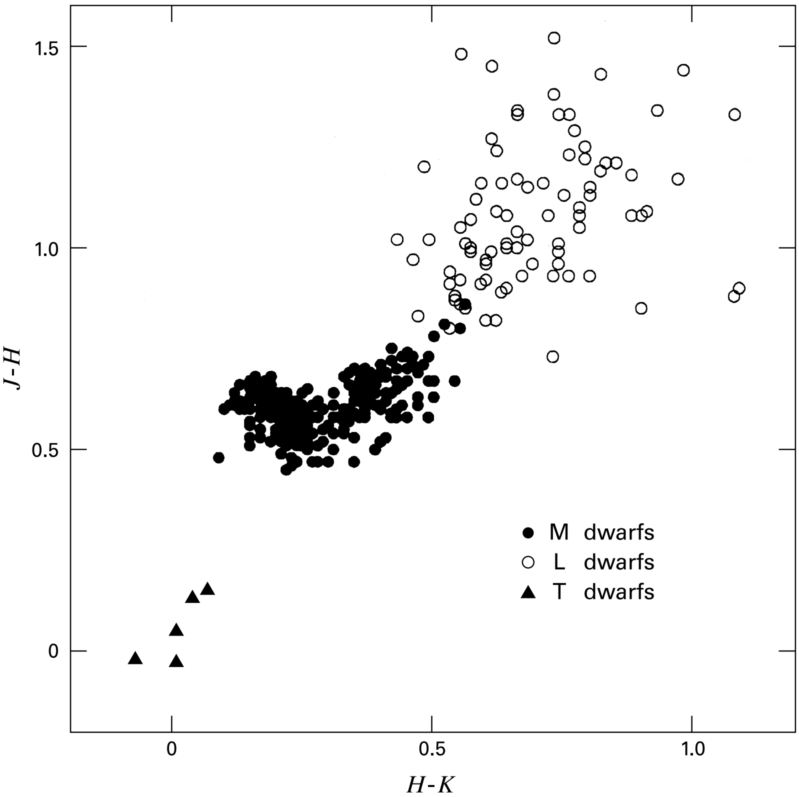


Figure 16.15 Near-infrared color-color plot of M dwarfs (*filled circles*), L dwarfs (*open circles*), and T dwarfs (*filled triangles*). The objects are from a variety of regions. Note that the typical measurement errors for the L- and T-dwarfs are quite large, about 0.13 mag.

the object fades to invisibility. This sequence of events holds regardless of the mass, provided it is below the critical $0.075\ M_\odot$.

The total number of brown dwarfs produced in a typical cluster or association may rival that of ordinary stars, although the net contribution in mass is small.[5] What is the origin of these objects? In addressing this issue, it is important to remember that the nature of the initial condensation or collapse process is unrelated to whether or not the final object fuses hydrogen. Thus, there is no reason in principle why the higher-mass brown dwarfs cannot form like low-mass stars, from the collapse of dense cloud cores. The discovery of PPl 15, a binary brown dwarf system in the Pleiades, lends credence to this view. On the other hand, it is difficult to envision the very lowest-mass brown dwarfs forming this way. Unless their parent dense cores have considerably more mass than the endproduct, the diffuse structure would have a thermal

---

[5] If the initial mass function were completely flat below $0.1\ M_\odot$, then equation (4.6) implies that 31 percent of all objects would have masses under this value.

pressure that overwhelms self-gravity. (That is, the cloud mass would be less than M_J; recall equation (9.24).) In this regime, brown dwarfs may form like planets, by condensation from a dusty disk surrounding a true star. The object Gl 229B, slowly orbiting its parent M star, would seem to be a case in point. The rarity of such systems may reflect a physical limit to the masses of objects that are able to coalesce from circumstellar disks.

16.5 Spinup and Spindown

After this digression into the substellar regime, we are ready to explore another important aspect of pre-main-sequence evolution. This is the question of how quasi-static contraction influences the star's rotation rate. Regardless of the details of protostellar collapse, the object must possess non-zero angular momentum at the birthline. If the contracting star is also fully convective, then rapid circulation of turbulent eddies may establish an interior with nearly solid-body rotation, as we discuss below. The moment of inertia of a polytropic sphere is some fixed, nondimensional coefficient times $M_* R_*^2$. It follows from angular momentum conservation that the rotation period should vary as R_*^2 and thus decline monotonically along a Hayashi track. In fact, observations reveal that spinup, when it occurs at all, progresses in a very different manner from this theoretical expectation. Moreover, the final approach to the main sequence is characterized by a rotational *deceleration*, at least for solar-type objects.

16.5.1 Measuring Rotation

Throughout this section, we shall concentrate on individual stars and ignore the issue of whether an object is actually single or part of a binary. We know, of course, that most stars actually fall into the second category. How does the presence of an orbiting companion alter an object's rotational history? We saw in Chapter 12 that the components within binaries exert mutual torques that both lower the system's eccentricity and induce synchronous rotation in the stars themselves. However, the tidal force responsible for these effects declines so rapidly with distance that it is negligible except in the very tightest systems. In such binaries, both circularization and synchronization occur as the components undergo quasi-static contraction. Numerical calculations of solar-mass pairs starting from the birthline find that those with periods less than about 8 days have circular orbits upon arrival on the main sequence, while wider pairs retain any initial eccentricity. This figure is consistent with the cutoff periods cited in Chapter 12. Since tidal gravity is most effective early in the contraction, when the stars are largest, the agreement provides additional support for the correctness of the starting radii and therefore the location of the birthline.

Returning to single stars and their rotation, the observational picture is not only complex, but still only partially filled in. We shall survey in due course the present landscape. First, however, we should review the tools we have at our disposal for assessing rotation in any given star, regardless of its age. One approach is spectroscopic and relies on a characteristic alteration in the profiles of absorption lines. A second, photometric technique is to monitor temporal variation of the object's broadband flux. Let us now examine in more detail both the conceptual basis and the implementation of each method.

Consider first the spectroscopic approach. Photons emanating from a rotating star appear redshifted if they originate from a surface patch that is receding from the observer. Approaching patches give blueshifted emission. Since the stellar surface is never resolved in practice, the net effect of rotation is to *broaden* any spectral line. Quantitatively, the original line profile $\phi(\lambda - \lambda_\circ)$ associated with a central wavelength $\lambda_\circ$ becomes smeared out to the new profile $\bar{\phi}(\lambda - \lambda_\circ)$, where

$$\bar{\phi}(\lambda - \lambda_\circ) = \int_{-1}^{+1} d\xi \; \mathcal{B}(\xi) \; \phi \left(\lambda - \lambda_\circ - \Delta\lambda_{\rm max}\, \xi\right) \; . \tag{16.56}$$

Here, $\Delta\lambda_{\rm max}$ is defined as

$$\Delta\lambda_{\rm max} \equiv \lambda_\circ \, \frac{V_{\rm eq}}{c} \, \sin i \; , \tag{16.57}$$

and represents the maximum Doppler shift from any point on the surface. The quantity $V_{\rm eq}$ is the velocity of a fluid element on the star's equator, and i is the inclination angle between the rotational axis and the line of sight. (Thus, the circular orbit of the equatorial element is seen edge-on when $i = 90^\circ$; compare Figure 12.6.) Finally, $\mathcal{B}(\xi)$ is the *broadening function*, which is obtained theoretically by integrating the Doppler-shifted radiation from the entire surface. We derive both equation (16.56) and a simple, illustrative form of $\mathcal{B}(\xi)$ in Appendix E.

To determine the rotational velocity of a star, one first requires the spectrum of a nonrotating star of identical spectral type. One uses equation (16.56) to artificially broaden this comparison spectrum to various degrees, *i. e.*, over a range of assumed values for $V_{\rm eq} \sin i$. The correct value of $V_{\rm eq} \sin i$ is obtained once the artificially broadened spectrum matches the one of interest.

Figure 16.16 shows, as a typical finding, the rotational broadening in the Taurus object DF Tau. The top spectrum is that of the T Tauri star itself, while the bottom is for a comparison star, DE Tau, that has negligible rotation. Both objects are of spectral type M1. It is evident that the spectra are also nearly identical. However, the lines in DF Tau are not quite as deep, and are slightly broader. Quantitative study gives a value for $V_{\rm eq} \sin i$ of 20 km $\rm s^{-1}$. Note that spinup alone does not convert the DE Tau spectrum into that of DF Tau. The lines must also be partially filled in by a broadband flux of energy. This additional luminosity is an important feature of many T Tauri stars and will be further discussed in Chapter 17.

The photometric method for obtaining rotational velocities utilizes the fact that many stars exhibit steady, periodic fluctuations in their broadband emission. In T Tauri stars, these changes are of modest amplitude, typically 0.1 mag at visual wavelengths. The periodic nature of the modulation, which may occur along with other, irregular variability, indicates that it is tied to stellar rotation. Indeed, associated changes in color, which we shall revisit in the next chapter, show that the dimming or brightening results from *starspots*. Such regions, like the better observed but intrinsically far weaker sunspots, represent intense local concentrations of magnetic flux. The surface layers of convective pre-main-sequence stars are permeated by strong magnetic fields and so are especially prone to starspots. The observed time between successive minima or maxima gives an accurate measure of the rotation period. Knowledge of the luminosity and effective temperature then yields the star's radius and hence the value of $V_{\rm eq}$ itself, with no correction for inclination.

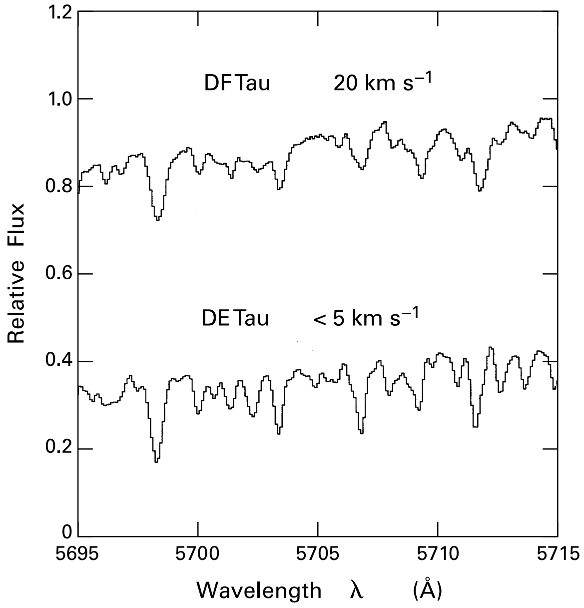


Figure 16.16 Rotational broadening of a T Tauri star. The top spectrum is that of DF Tau, a T Tauri star rotating with the indicated velocity, while below is DE Tau, a nonrotating object of the same spectral type.

16.5.2 Main-Sequence Trends

Any results concerning the rotation of young stars should be viewed in context. Specifically, it helps to understand first the behavior of mature objects in this regard. Figure 16.17 displays the variation of equatorial speed with mass, or equivalently, spectral type, for single main-sequence stars in the field. Each filled circle represents an average of $V_{\rm eq}\,\sin\,i$ over many equal-mass stars, as obtained spectroscopically. Note that the presence of the unknown inclination angle has little qualitative impact, since the average of $\sin\,i$ for an ensemble of randomly pointed axes is $\pi/4\,=\,0.79$.

Scanning Figure 16.17 from right to left, one sees that the surface velocity is constant or perhaps increases gently, reaching a broad maximum in late-B stars. Even this maximum is well below the critical speed for centrifugal breakup, $V_{\rm cen}\,\equiv\,\sqrt{G\,M_*/R_*}$. The latter is also shown here as a function of stellar mass. For spectral types of A or later, the average rotational speed falls, slowly at first, but then precipitously for spectral types later than early F, *i. e.*, for $M_*\,\lesssim\,1.5\,M_\odot$. This transition mass is not far from that where outer convection first appears. Theory predicts that stars below this point emit magnetized winds, whose transport of angular momentum we discussed in Chapter 13. Since pre-main-sequence stars also have significant outer convection, the falloff in $V_{\rm eq}$ seen in Figure 16.17 already hints that such young, contracting objects do not spin up monotonically as they descend from the birthline.

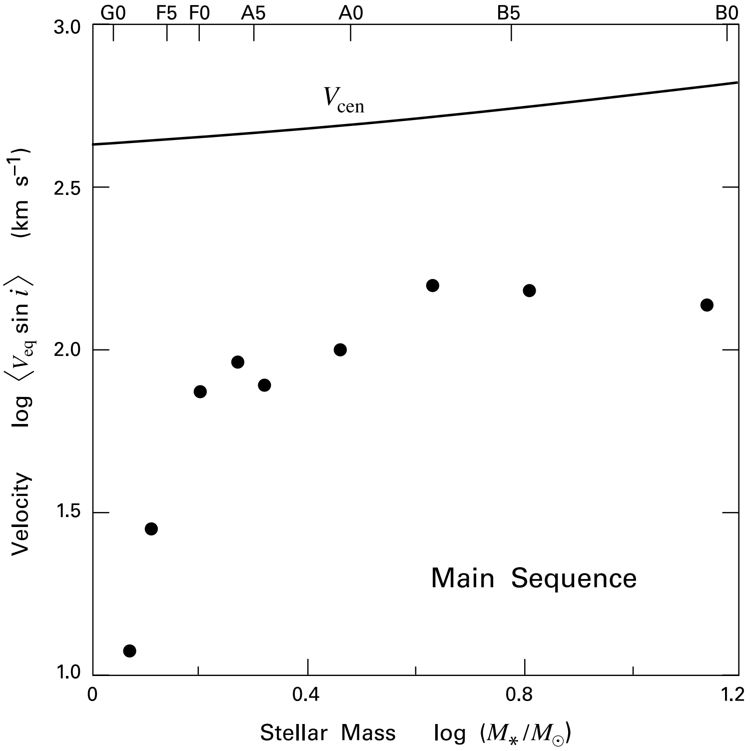


Figure 16.17 Rotational velocity of main-sequence field stars. Displayed is the equatorial velocity as a function of stellar mass. Also shown are the corresponding spectral types. The smooth curve at the top represents the critical speed for centrifugal breakup at each mass.

The temporal change in rotation is best documented in solar-type stars that have already ceased contraction and are slowly consuming their hydrogen on the main sequence. Figure 16.18 shows measured values of $V_{\rm eq}$ sin i as a function of age for four open clusters, along with the Sun. As in the previous figure, each point represents an average over a large number of individual objects, which in this case are confined to the spectral range from F8 to G8. The observations reveal a consistent pattern and show that the relatively slow rotation of low-mass stars in the field is achieved through a process of gradual deceleration. Ignoring the inclination factor, we find that

$$V_{\rm eq} \approx V_\circ \left(t/10^9\,{\rm yr}\right)^{-\alpha} \ . \tag{16.58}$$

The dashed line in Figure 16.18, a least-squares fit to the data, corresponds to $V_\circ = 4.5$ km s^{-1} and $\alpha = 0.51$. Thus, $V_{\rm eq}$ declines approximately as $t^{-1/2}$, a result known traditionally as the *Skumanich law*. This spindown must reflect the drain of angular momentum by the stellar wind.

We emphasize that the Skumanich law refers only to *average* values of the rotational speed. In fact, each group displays a significant dispersion in the observed values of $V_{\rm eq}$ sin i, even within the relatively narrow mass range used in Figure 16.18. The youngest cluster shown

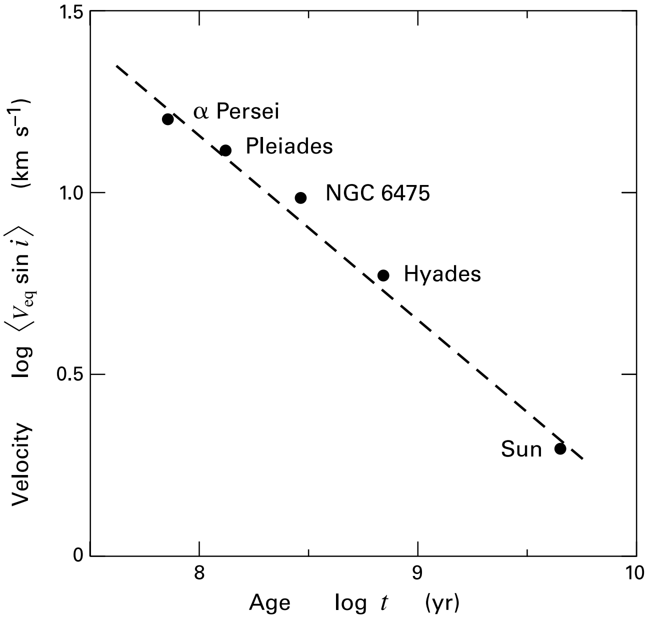


Figure 16.18 Average values of $V_{\rm eq}$ sin i for G-type stars in selected clusters and the Sun.

in the figure, α Persei, contains a few G-type stars with $V_{\rm eq}$ sin i in excess of 200 km s^{-1}. Significantly, the measured dispersion of G-star velocities shrinks with cluster age and is quite small in the Hyades. Stars that are rapidly rotating at the age of α Persei (7×10^7 yr) must initially decelerate at a *faster* rate than indicated by equation (16.58), while slower ones spin down at a more leisurely pace.

16.5.3 Orion and the Pleiades

Within any young cluster, a significant spread in rotational speeds exists at all stellar masses, not just those close to solar. How the distribution of speeds varies with mass will ultimately reveal much about both magnetic braking and the internal redistribution of angular momentum. The pattern is especially well documented in the Pleiades. Figure 16.19 summarizes graphically the results of several hundred spectroscopic velocity measurements. We see immediately that most stars have relatively low surface velocities, $V_{\rm eq} \ll V_{\rm cen}$. Within this majority population, many values of $V_{\rm eq}$ sin i are only upper limits, as indicated in the figure. The average speed for the G-type stars, given by the Skumanich law, is heavily influenced by these low values.

The most interesting fact revealed by Figure 16.19 is that numerous stars are spinning much more rapidly, with $V_{\rm eq}$ sin i approaching, and occasionally exceeding, 100 km s^{-1}. The latter value is still only about 1/3 of $V_{\rm cen}$, as shown in the figure. For stellar masses between about 0.5 and 1.0 $M_\odot$, there is a pronounced hole in the density of points. That is, stars in this range are rather neatly divided into two subsets – the bulk of slow rotators and a distinct minority of rapid rotators.

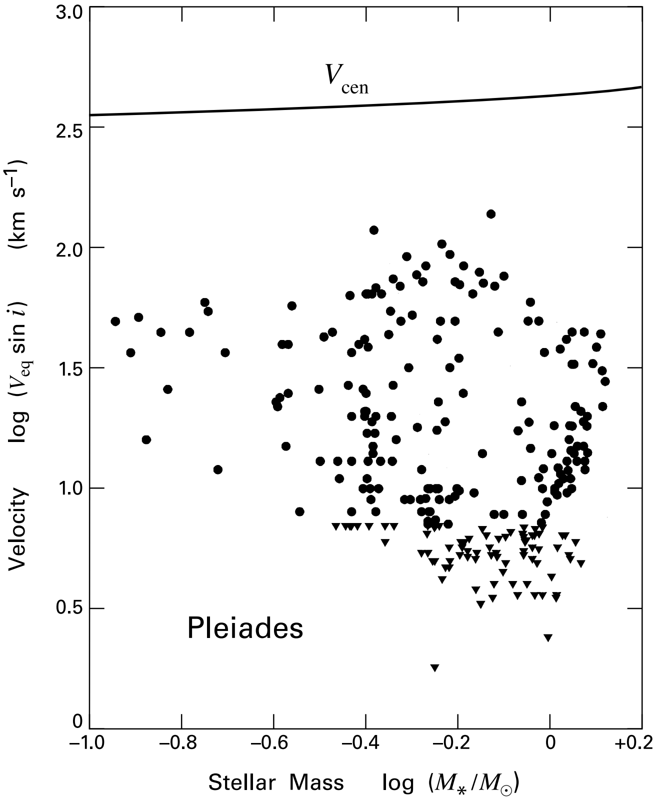


Figure 16.19 Spectroscopically measured values of $V_{\rm eq} \sin i$ for 266 stars in the Pleiades. Inverted triangles represent upper limits. Also shown is the centrifugal breakup speed $V_{\rm cen}$, calculated by using the main-sequence radius at each stellar mass.

For $M_* \gtrsim 1.0\,M_\odot$, the spread in $V_{\rm eq} \sin i$ narrows. There is also a narrowing at lower masses, *i. e.*, for $M_* \lesssim 0.5\,M_\odot$. We should view both results with caution, as there are undoubtedly more stars with velocities below the detection limit. Nevertheless, it is worth observing that the hole at intermediate speeds disappears for lower mass, and that many such stars exhibit relatively high $V_{\rm eq}$. Extrapolating Figure 16.17 to this mass regime, we see that most of the Pleiades objects must slow down considerably as they age. Note, however, that rapid rotation still persists even in some main-sequence field stars of spectral type M.

The breadth of Pleiades rotation rates at all masses is, of course, consistent with the α Persei data, but is nonetheless striking. The total spread in $V_{\rm eq} \sin i$, *e.g.* for $M_* \approx 0.8\,M_\odot$, is at least a factor of 15. This is much too great to be attributable solely to age differences among member stars, which are within 10^7 yr (§ 4.4). More plausibly, the dispersion reflects variations in rotation at a much earlier epoch. Apparently, the braking action of stellar winds has been insufficient, even at an age of 1×10^8 yr, to erase all memory of these primordial conditions.

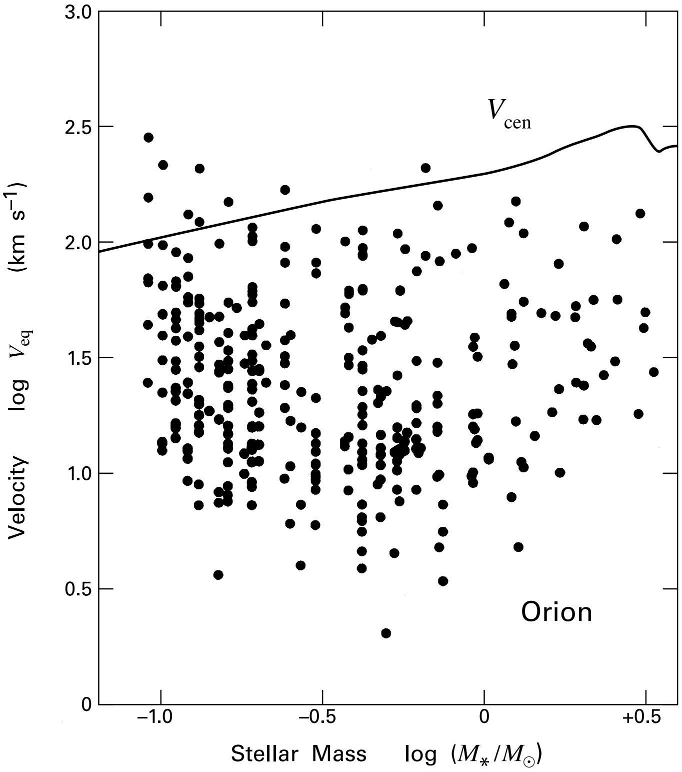


Figure 16.20 Values of the equatorial rotational speed $V_{\rm eq}$ for 330 stars in the Orion Nebula Cluster. These speeds were obtained photometrically. Also shown is the critical breakup speed $V_{\rm cen}$. The latter employs the birthline radius from Table 16.1 at each mass. Note the kink near $M_* = 3\,M_\odot$, associated with the swelling from deuterium shell burning.

The next logical step is to explore systematically the rotation rates within even younger clusters and associations. Such studies have employed both the spectroscopic and photometric approaches. By far the richest source of results has been the Orion Nebula Cluster. In Figure 16.20, we display rotational speeds for over 300 stars, ranging in mass from about 0.1 to 3 $M_\odot$. The $V_{\rm eq}$-values were obtained from photometric periods, after using the stars' known luminosities and effective temperatures to calculate stellar radii. Placement of each object in the HR diagram also yielded the proper mass.

It is clear that stars in this system, at an age of about 2×10^6 yr, again have a variety of rotational speeds. Indeed, the dispersion appears even broader than that in the Pleiades, especially when we consider that the Orion sample is again biased against very slow rotators. Note that the bimodal distribution of $V_{\rm eq}$ in the Pleiades, which occurs for masses between 0.5 and 1.0 $M_\odot$, is no longer evident.

Another striking feature in Orion is the large population of low-mass objects with rotational velocities exceeding 100 km s^{-1}. The highest speeds found here are closer to the critical, breakup value than in the Pleiades. Figure 16.20 even shows many objects ostensibly exceeding $V_{\rm cen}$. However, the latter was calculated here using the *birthline* radius at each mass. The contraction of a star raises the associated value of $V_{\rm cen}$, so that $V_{\rm eq}$ remains below that limiting value. Some of the most rapidly rotating objects could be members of tight binary systems. The tidal gravitational force of an undetected companion would then force the visible star into faster rotation than it would maintain on its own.

The lowering of the average equatorial velocity from Orion to the Pleiades indicates that a gradual deceleration, like that characterizing main-sequence G stars, holds both further back in time and over a broader range in mass. On the other hand, the existence of the Pleiades fast rotators shows that many stars rotating close to breakup at an age of 2×10^6 yr are still spinning quite rapidly at 1×10^8 yr. The deceleration time should generally increase toward lower stellar mass, to account for the fact that some M stars remain fast rotators even on the main sequence. Most of the rotational data is consistent with this broad picture. A possible exception is Taurus-Auriga. Since this system is older than Orion but younger than α Persei or the Pleiades, one would expect to find a substantial fraction of very rapid rotators. Spectroscopic study, however, has thus far revealed only a handful of members with $V_{\rm eq}\,\sin\,i$ exceeding even 50 km s^{-1}, and none over 100 km s^{-1}.

16.5.4 Embedded Stars

Probing the rotation of stars even younger than those in the Orion Nebula is technologically challenging. Such objects tend to be embedded in both circumstellar and cloud material, and may be weak or undetectable in the optical. In principle, one could apply the photometric method by dedicated monitoring at near-infrared wavelengths. More progress has been made through high-resolution spectroscopy in the regime of 2 µm. Class I objects are so heavily veiled that they exhibit no absorption lines at all. However, extreme Class II sources, *i. e.*, those with nearly flat spectral energy distributions between 2.2 and 10 µm, occasionally show weak features in CO. The few available measurements yield values of $V_{\rm eq}\,\sin\,i$ that usually exceed 20 km s^{-1}.

What are theoretical expectations in this regard? Suppose we first ignore the effect of magnetic braking during the protostellar accretion phase. Under this admittedly drastic simplification, most newly visible stars would have acquired so much angular momentum from their parent dense cores that they would be rotating at breakup speed. Indeed, this condition would be met even before another early event associated with cloud rotation, the appearance of a circumstellar disk.

It is instructive to follow the argument here in more detail. Recall first that protostars become convectively unstable from deuterium burning at a relatively low mass. Once this transition occurs, the turbulent stresses associated with convection are effective at eliminating internal shear. (See, however, the discussion of the solar case below.) In the idealized, and perhaps unrealistic, picture of zero differential rotation, we may characterize the fully convective protostar as a rigidly rotating $n = 3/2$ polytrope. If this object is indeed rotating close to breakup, then its angular velocity, $\Omega_{\rm max}$, is

$$\Omega_{\rm max} = n_4 \left(\frac{G\,M_*}{R_*^3}\right)^{1/2} . \qquad (16.59)$$

Here, n_4 is a constant near unity, which is derivable from the numerical study of rotating polytropes. The star's moment of inertia is

$$I_* = n_5 \, M_* \, R_*^2 \; , \tag{16.60}$$

where n_5 is another, known constant. The angular momentum contained in this maximally rotating structure is therefore

$$\begin{aligned} J_{\text{max}} &= I_* \, \Omega_{\text{max}} \\ &= n_4 \, n_5 \left(G \, M_*^3 \, R_* \right)^{1/2} \; . \end{aligned} \tag{16.61}$$

Our goal is to see when, during the infall process, such a state is achieved. We model the parent dense core as a singular isothermal sphere, rotating at angular speed $\Omega_\circ$. Utilization of the density distribution in equation (9.8) shows that the amount of *cloud* angular momentum eventually incorporated into the star is

$$J_{\text{cloud}} = \frac{2}{9} \, M_* \, \Omega_\circ \, r_{\text{cloud}}^2 \; . \tag{16.62}$$

The radius r_{cloud} is the outer boundary of the collapsing volume. From equation (9.8), this quantity is

$$r_{\text{cloud}} = \frac{G \, M_*}{2 \, a_T^2} \; . \tag{16.63}$$

After equating J_{max} and J_{cloud}, we solve for M_* to find

$$M_* = (18 \, n_4 \, n_5)^{2/3} \left(\frac{a_T^8 \, R_*}{G^3 \, \Omega_\circ^2} \right)^{1/3} \; . \tag{16.64}$$

The protostar is built up at the rate $\dot{M}$, given by equation (10.31). Thus, the time required to accumulate the mass M_* is

$$\begin{aligned} t_{\text{break}} &= M_* / \dot{M} \\ &= (18 \, n_4 \, n_5)^{2/3} \left(\frac{R_*}{m_\circ^3 \, \Omega_\circ^2 \, a_T} \right)^{1/3} \; . \end{aligned} \tag{16.65}$$

The numerical studies alluded to earlier give $n_4 = 1.02$ and $n_5 = 0.132$. Hence, the factor containing these constants in equation (16.65) has the value 1.80. Note that our expression for t_{break} has precisely the same algebraic form as that for t_0, the time when the rotating infall first creates a circumstellar disk. From equation (11.32), the corresponding numerical coefficient in t_0 is $16^{1/3} = 2.52$. We conclude that t_{break} is less than t_0, *i. e.*, the star achieves breakup speed prior to disk formation.

For times between t_{break} and t_0, the star must find some way to rid itself of the excess angular momentum it is gaining through direct infall. Spinup of the surface layers may lead to shedding of material, *i. e.*, to disk formation even prior to t_0. However, it is plausible that magnetized winds, already known for their torquing action in older stars, also operate at this early epoch. Sometime after t_0, the disk begins feeding additional mass and angular momentum to the star (§ 11.3). The twisting of magnetic field lines in an outflowing wind might then maintain the star's equatorial speed near, but slightly below, the breakup value. Alternatively, the wind could slow the object to a lower rotation rate. Future observations of $V_{\rm eq}$ in deeply embedded sources will tell us which path is actually followed.

16.5.5 Insights from Wind Theory

Returning to the pre-main-sequence phase, the empirical data from young groups demonstrates that substantial angular momentum loss occurs during contraction. It is difficult to be more quantitative, but a spindown of the Skumanich form may well hold over a broad mass range, at least for part of this epoch. Suppose that the torquing action is due solely to a magnetized wind. Then a simple theoretical argument, supplemented with input from observations, gives an expression for the deceleration rate.[6]

We begin with equation (13.32), which tells us that the specific angular momentum in any flux tube is

$$\mathcal{J} = \alpha\, \varpi_A^2 \ . \tag{16.66}$$

Recall that ϖ_A is the (cylindrical) Alfvén radius, while α is a constant along the field line in question. If $\boldsymbol{B}$ has only a small toroidal component near the stellar surface, then equation (13.19) implies that α is close to the star's rotational speed Ω_*. Suppose now that the full Alfvén surface is a sphere of radius r_A. Then the average specific angular momentum in this sphere is

$$\mathcal{J}_{\rm av} = \frac{2}{3}\,\Omega_*\, r_A^2 \ . \tag{16.67}$$

To obtain the total efflux of angular momentum, we multiply by the mass transport rate in the wind:

$$\begin{aligned} \dot{J} &= -\frac{2}{3}\,\dot{M}_w\,\Omega_*\, r_A^2 \\ &= -\frac{8\pi}{3}\,\rho_A\, u_A\,\Omega_*\, r_A^4 \ . \end{aligned} \tag{16.68}$$

Here, ρ_A and u_A are the spherically averaged wind density and radial velocity, respectively, at r_A.

When the ejected gas crosses r_A, the radial velocity matches the local Alfvén speed, evaluated using the poloidal component of the field:

$$u_A^2 = \frac{B_p^2(r_A)}{4\,\pi\,\rho_A} \ . \tag{16.69}$$

---

[6] Some investigators favor the idea that the star brakes itself by transferring angular momentum to its disk, via a linked magnetic field. As just demonstrated, angular momentum loss must occur *before* the appearance of a disk. In any case, we shall not explore this picture here. Recall § 13.4 for an introduction to the complexities of the star-disk interaction.

We may further relate B_p to the stellar surface field B_* via flux conservation:

$$B_p(r_A)\, r_A^2 \;=\; B_*\, R_*^2 \; . \tag{16.70}$$

Thus, equation (16.68) becomes

$$\begin{aligned} \dot{J} \;&=\; -\frac{2}{3}\,\frac{B_p^2\,(r_A)}{u_A}\, r_A^4\, \Omega_* \\ &=\; -\frac{2}{3}\,\frac{B_*^2}{u_A}\, R_*^4\, \Omega_* \; . \end{aligned} \tag{16.71}$$

The Alfvén radius is far enough from the stellar surface that the radial velocity has nearly reached the terminal value V_∞. Empirically, the latter is close to the escape speed. Hence, the value of u_A is likely to be insensitive to B_* and Ω_*. But what is the relation of these latter two quantities? The field is generated through the dynamo effect, *i. e.*, the interaction of differential rotation and convection. We therefore expect some dependence of B_* on Ω_*. Unfortunately, dynamo theory itself is not yet able to provide the desired relation, and we must turn to a direct, observational comparison.

Detection of stellar magnetic fields relies on the Zeeman splitting of optical spectral lines. In practice, one compares the profiles of selected lines with others formed at a similar depth that are less magnetically sensitive. In this manner, one recovers both the line-of-sight field $B_{||}$ and f_B, the fraction of the star's surface covered by the field. One interesting result, obtained for G and K dwarfs, is that $B_{||}$ depends on spectral type but *not* on the observed Ω_*. On the other hand, the product $f_B B_{||}$ *is* at least roughly proportional to Ω_*. Thus, faster rotation increases the number of flux tubes penetrating the surface, but without altering the field strength in each one. In any case, the larger-scale field affecting magnetic braking arises from the merging of individual flux tubes some distance above the photosphere. The combination $B_* R_*^2$ entering equation (16.71), a quantity proportional to the total magnetic flux emanating from a stellar hemisphere, thus varies as Ω_*.

Our conclusion is that $\dot{J}$ is proportional to Ω_*^3. For stars just arriving on the main sequence, the moment of inertia may be considered to be fixed. We then have

$$\frac{d\Omega_*}{dt} \;=\; -k\,\Omega_*^3 \; , \tag{16.72}$$

where k is a positive constant, dependent only on spectral type. The solution to this equation is

$$\Omega_*^{-2} \,-\, \Omega_\circ^{-2} \;=\; 2\,k\,t \; , \tag{16.73}$$

for any initial speed $\Omega_\circ$. After initial transient behavior, the equatorial velocity is indeed proportional to $t^{-1/2}$.

It is a simple matter to generalize equation (16.72) so as to include the shrinking of the star's radius. However, this broader Skumanich-type relation is insufficient to describe the entire pre-main-sequence phase. If $\dot{J}$ were actually proportional to Ω_*^3 within the Orion Nebula Cluster, all of its members would be slow rotators by the age of the Pleiades. Consequently, the magnetic torque needs to *saturate* at high rotational speed. That is, the braking must be relatively inefficient at early epochs. A likely reason is that the star is fully covered by intense magnetic flux above some minimum rotation speed. In support of this picture is a related observation. One signpost of surface activity, including the emergence of magnetic fields, is X-ray emission. (See also Chapter 17.) Among stars of a given spectral type, the X-ray luminosity generally correlates with $V_{\rm eq}\,\sin\,i$, but then levels off for sufficiently rapid rotation.

Another complication faced by theory is the case when the star is only partially convective. Indeed, this situation arises in the very stars that provided the original data for the Skumanich law. While it may be reasonable to assume that the outer convection zone spins down as a solid body, how does the interior respond? Does angular momentum diffuse outward as the result of some effective viscous stress between shearing layers? Do turbulent eddies in the convection zone generate waves that travel inward and alter the rotation? Or do the central regions maintain a state of high spin inherited from earlier contraction?

The most detailed empirical information concerning these questions has come from the field of helioseismology. By combining observations of the Sun's surface oscillations with models of its bulk, normal modes, one may derive the internal distribution of the angular velocity. It has long been known that the equatorial value of the *surface* angular speed exceeds the polar value by about 40 percent. Helioseismology has demonstrated that this difference persists to the base of the convection zone, with Ω showing little radial variation. Within the radiative interior, the rotation rate appears to be uniform. The transition from differential to solid-body rotation occurs in a remarkably thin layer known as the *tachocline*. How these results extend to younger, more fully convective objects is not yet clear. There is no doubt, however, that continued analysis of the solar case will aid our understanding of the evolutionary issues raised here.

Chapter Summary

The birthline manifests itself in the HR diagrams of stellar groups as a cutoff in the population at high luminosity. This upper boundary is present because pre-main-sequence stars first appear with the radii they attained as accreting protostars. The protostellar radius, in turn, is tightly constrained by deuterium fusion for subsolar masses and by gravitational contraction in intermediate-mass objects.

A star that has ceased protostellar accretion slowly contracts by radiating energy into space. The star's emitted luminosity is initially high enough to drive interior convection. Mixing associated with turbulent convection enforces a spatially constant specific entropy, and the star contracts in a simple, homologous fashion. Since its effective temperature remains nearly fixed, the object descends vertically in the HR diagram. Eventually, the interior opacity diminishes to the point where radiation can provide energy transport. The effective temperature then increases, and the evolutionary track becomes more horizontal.

The rising central temperature in a pre-main-sequence star ignites a sequence of nuclear reactions. First, the object consumes whatever deuterium remains from the protostar phase. This reaction can stall contraction for up to 10^6 yr in the lowest-mass stars. The subsequent fusion of lithium is energetically trivial. Since, however, the element is readily detected in stellar photosheres, its presence is an important observational signature of youth. Ordinary hydrogen eventually burns in all stars more massive than 0.075 $M_\odot$. Here, the released energy completely halts stellar contraction.

Objects of even smaller mass have a central temperature that peaks too low for hydrogen ignition and then declines further as the radius settles to a finite value. In these brown dwarfs, electron degeneracy pressure counteracts self-gravity. The presence of surface lithium is an important observational tool for identification. Spectra at these very low effective temperatures exhibit peculiarities due to the presence of dust grains and molecules such as CH_4. Numerous single brown dwarfs are present in stellar groups.

One important, if poorly understood, property of young stars is their rotation, which can be measured both spectroscopically and photometrically. Members of the Orion Nebula, at an age of 2×10^6 yr, have a large spread in rotation speed that must be inherited from the protostar phase. This spread diminishes by the age of the Pleiades (1×10^8 yr), but there remains a population of rapid rotators. The well-documented spindown of main-sequence stars, which is consistent with magnetic wind theory, cannot apply to the pre-main-sequence phase. Wind braking is present at this epoch, but is apparently less efficient.

Suggested Reading

Section 16.1 The stellar birthline was first described in

Stahler, S. W. 1983, ApJ, 274, 822

Palla, F. & Stahler, S. W. 1990, ApJ, 360, L47.

The first paper discusses the birthline for low-mass stars, while the second extends the concept to those of intermediate mass.

Section 16.2 Our present conception of pre-main-sequence contraction stems from

Hayashi, C. 1961, PASJ, 13, 450.

This was followed by more detailed studies that remain valuable references:

Hayashi, C., Hoshi, R., & Sugimoto, D. 1962, Progr. Theor. Phys. Supp., No. 22

Ezer, D. & Cameron, A. G. W. 1965, Can. J. Phys., 43, 1497

Iben, I. 1965, ApJ, 141, 933.

A summary of the historical development of the subject is

Stahler, S. W. 1988, PASP, 100, 1474.

Section 16.3 The classic numerical studies already cited included hydrogen fusion. A standard reference for the main-sequence reactions is

Rolfs, C. E. & Rodney, W. S. 1988, *Cauldrons in the Cosmos* (Chicago: U. of Chicago).

The idea of a deuterium main sequence comes from

Grossman, A. N. & Graboske, H. C. 1971, ApJ, 164, 475.

For a modern, theoretical treatment of lithium fusion, see

Ushomirsky, G., Matzner, C. D., Brown, E. F., Bildsten, L., Hilliard, V. G., & Schroeder, P. C. 1998, ApJ, 497, 253.

The observational situation has been summarized by

Martín, E. L. 1997, Mem. Soc. Astron. It., 68, 905.

Section 16.4 Brown dwarfs were discovered theoretically by

Kumar, S. 1963, ApJ, 137, 1121

Hayashi, C. & Nakano, T. 1963, Prog. Theoret. Phys., 30, 460.

The first confirmed sighting was

Nakajima, T., Oppenheimer, B. R., Kulkarni, S. R., Golimowski, D. A., Matthews, K., & Durrance, S. T. 1995, Nature, 378, 463.

For a lucid summary of the observational situation, see

Oppenheimer, B. R., Kulkarni, S. R., & Stauffer, J. R. 2000, in *Protostars and Planets IV*, ed. V. Mannings, A. P. Boss, & S. S. Russell, (Tucson: U. of Arizona), p. 1313.

Current research activities are described in

Martín, E. L. (ed.) 2003, *Brown Dwarfs* (San Francisco: ASP).

An overview of theoretical issues is

Chabrier, G. & Baraffe, I. 2000, ARAA, 38, 337.

Section 16.5 The spindown of G dwarfs was quantified by

Skumanich, A. 1972, ApJ, 171, 565.

There have been numerous papers investigating rotation within young clusters. Two examples focusing on Orion and the Pleiades, respectively, are

Stassun, K. G., Mathieu, R. D., Mazeh, T., & Vrba, F. 1999, AJ, 117, 2941

Terndrup, D. M., Stauffer, J. R., Pinsonneault, M. H., Sills, A., Yuan, Y., Jones, B. F., Fischer, D., & Krishnamurthi, A. 2000, AJ, 119, 1303.

One approach to the theory has been to adopt a parametrized law for angular momentum loss and investigate its consequences:

Krishnamurthi, A., Pinsonneault, M. H., Barnes, S., & Sofia, S. 1997, ApJ, 480, 303.

The physical issues that will need to be tackled are summarized by

Charbonneau, P., Schrijver, C. J., & MacGregor, K. B. 1997, in *Cosmic Winds and the Heliosphere*, ed. J. R. Jokipii, C. P. Sonett, & M. S. Giampapa, (Tucson: U. of Arizona), p. 677.

Determination of the Sun's internal rotation from helioseismology is described in

Thompson, M. J. *et al.* 1996, Science, 272, 1300.

The Formation of Stars
Steven W. Stahler and Francesco Palla

17 T Tauri Stars

Observations of star-forming regions have utilized, to great advantage, detectors spanning a broad range of wavelengths. It remains true, however, that our knowledge is most complete for young, but optically visible stars. Those of about 2 $M_\odot$ and below fall within the T Tauri class, and have been intensively studied over many decades. Throughout this book, we have had occasion to refer to various properties of these objects, but have never presented the empirical data or associated theory in a systematic fashion. It is the purpose of this chapter to fill that gap.

Apart from their proximity to molecular clouds, the youth of T Tauri stars is primarily manifested in their excess emission, *i. e.*, photospheric radiation beyond that found in main-sequence stars of the same effective temperature. We accordingly begin by describing the characteristics of this emission, which occurs in both spectral lines and the continuum. Here we cover not only optical radiation, but also that in the ultraviolet, infrared, X-ray, and radio. We then turn to the finer details of optical lines, to show how these lines reveal the presence of both outflowing and infalling matter (§ 17.2). In the next section, we see how excess continuum radiation furnishes evidence for the existence of circumstellar disks. We also summarize theoretical understanding of the disks' thermal and dynamical structure.

Another fundamental characteristic of T Tauri stars is their temporal variability (§ 17.4). Line and continuum fluctuations span a range of amplitudes, and can be either random or periodic. Discussion of the latter naturally leads us to the issue of T Tauri magnetic fields and their observational detection. We also devote significant attention to the FU Orionis phenomenon, an especially dramatic flareup whose origin remains a puzzle.

In the final section, we describe low-mass objects that, while still contracting, have diminished surface activity and excess emission. This class of post-T Tauri stars provides the observational link to the main sequence. Objects that are still more evolved, *i. e.*, relatively young dwarf stars, exhibit tantalizing clues of ongoing planet formation. We close the chapter by summarizing our present understanding of this complex phenomenon.

17.1 Line and Continuum Emission

Although T Tauri stars individually display a bewildering variation in their properties, they nevertheless share a set of common features. These intrinsic characteristics are more robust, from an observational perspective, than the star's placement in the proper region of the HR diagram. For the latter purpose, one must assign an effective temperature, usually based on the pattern of optical absorption lines. These are sometimes barely discernible, being so heavily filled in that the underlying spectral type is hard to identify. The contamination here is due to an extra component of continuum radiation not found in main-sequence stars. Whether or not this *veiling* is present, some spectral lines that would normally be in absorption appear in emission.

The Formation of Stars. Steven W. Stahler and Francesco Palla

ISBN: 3-527-40559-3

In fact, the presence of prominent emission lines is one basic, defining property of the T Tauri class. These spikes project above a background spectrum that usually, but not always, contains well-defined absorption dips. Let us begin, then, by describing both the emission and absorption features.

17.1.1 Optical Spectra

The optical emission lines of T Tauri stars arise from hydrogen, along with neutral and singly ionized metals. In the latter category are the prominent Ca II H and K lines at 3968 and 3934 Å, respectively. Additionally, all T Tauri stars have significant flux in the X-ray regime. Both optical emission lines and X-rays are also present in late-type, main-sequence stars. Here they arise from the dissipation of mechanical and magnetic energy in an extended chromosphere and corona lying above the stellar surface proper (recall Chapter 13). This activity, in turn, is a manifestation of subsurface convection. Since pre-main-sequence stars of low mass are also expected to be convectively unstable, the minimal T Tauri properties we have listed come as no surprise. We shall see, however, that the *level* of surface activity is often far greater than in main-sequence stars. Moreover, many T Tauri stars display additional features that are difficult to understand solely within this framework.

The hydrogen lines are those of the Balmer series, principally Hα at 6563 Å. Accordingly, the vast majority of T Tauri stars have been discovered through wide-field surveys that look toward visually obscured regions for sources of either Hα or X-ray flux. Additional studies are then necessary to confirm the pre-main-sequence nature of any candidate object. Detection of strong lithium absorption at 6708 Å is a frequently used diagnostic. Through such means, observers have identified and catalogued hundreds of objects.

Within this large dataset, those optical spectra displaying clear absorption features have associated spectral types. These range from late G to mid M. The correspondence of the spectral type with effective temperature is similar to that for main-sequence stars (Table 1.1), but differs slightly, as a result of the lower surface gravity. It is important to realize that stars with identical $T_{\rm eff}$-values may still have spectra that appear quite different. This difference reflects the amount of excess emission in both the continuum and lines.

To illustrate the situation, Figure 17.1 displays spectra of three T Tauri stars, spanning wavelengths from the ultraviolet to the red portion of the optical. All the stars have spectral types from late K to early M. Starting at the bottom, the objects show an increasing degree of emission beyond photospheric. Thus, V830 Tau, a weak-lined T Tauri star, has pronounced absorption dips. The lines in BP Tau are shallower, the result of continuum veiling. Finally, the extreme T Tauri star DR Tau shows very few absorption lines, but a much larger number in emission. Its spectral type is accordingly rather uncertain.

Other trends are noteworthy. The bluest region of each spectrum is dominated by the Balmer continuum, *i. e.*, by photons released through capture of free electrons into the $n = 2$ level of hydrogen. Classical T Tauri stars show an upward jump in flux below the ionization threshold of 3647 Å. In main-sequence objects and weak-lined T Tauri stars like V830 Tau, this portion of the spectrum is depressed *below* that at longer wavelengths. Note, in both BP Tau and DR Tau, the thicket of emission lines just redward of the jump. These are higher members of the Balmer series: Hγ ($n = 5 \rightarrow 2$), Hδ ($n = 6 \rightarrow 2$), *etc.* Also within this region are the Ca II H and K

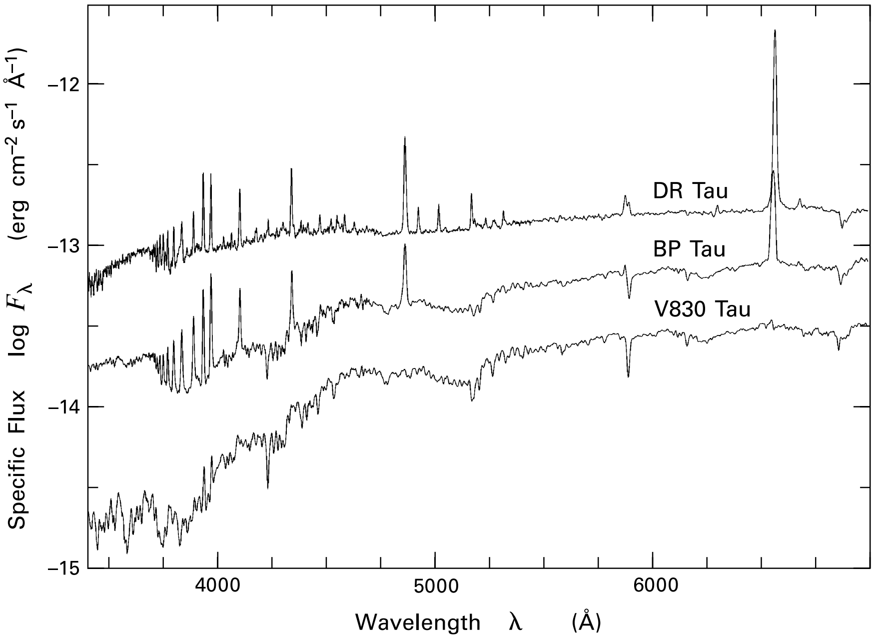


Figure 17.1 Medium-resolution spectra of three T Tauri stars. The received flux is correct only for BP Tau. The other two spectra have been shifted vertically for convenience.

lines. In DR Tau, there is a group of lines in Fe II, just to the red of Hβ at 4861 Å. The Na I D doublet at 5893 Å is in emission in this extreme T Tauri star, while it appears as an absorption feature in the other spectrograms.

Table 17.1 lists the principal emission lines found in classical T Tauri stars. Here we cover infrared, optical, and ultraviolet wavelengths. The second column gives, in spectroscopic notation, the electronic transitions, while the Einstein A-values are tabulated last. Among the optical lines, we note that both [O I] 6300 Å and [S II] 4076 Å have spontaneous emission rates smaller than the others by orders of magnitude. These are forbidden transitions that originate from material at relatively low densities. We will see later that these lines are valuable for probing T Tauri winds, just as they were for jets from more embedded stars. For the moment, however, we focus on the other, permitted, transitions.

17.1.2 Emission in Hα

Returning to Figure 17.1, the strongest line in both BP Tau and DR Tau is Hα. In contrast, the spectrum of V830 Tau is nearly flat in the same region. The presence of Hα in emission is, in fact, one of the defining spectral characteristics that separates classical from weak-lined T Tauri stars. It is apparent from our examples that increasing strength in this particular line is not an isolated phenomenon, but occurs in tandem with additional signposts of activity – a

Table 17.1 Main Emission Lines in Classical T Tauri Stars

Line	Transition	Wavelength (Å)	A_{ul} (s^{-1})
Infrared			
Br γ	$n = 7 \rightarrow 4$	21661	3.0×10^5
Pa β	$n = 5 \rightarrow 3$	12822	2.2×10^6
Ca II	$^2P_{1/2} \rightarrow {}^2D_{3/2}$	8662	2.8×10^5
Ca II	$^2P_{3/2} \rightarrow {}^2D_{5/2}$	8542	1.2×10^6
Ca II	$^2P_{3/2} \rightarrow {}^2D_{3/2}$	8498	6.3×10^5
Optical			
[S II]	$^2D_{3/2} \rightarrow {}^4S_{3/2}$	6731	8.8×10^{-4}
[S II]	$^2D_{5/2} \rightarrow {}^4S_{3/2}$	6716	2.6×10^{-4}
Hα	$n = 3 \rightarrow 2$	6563	1.0×10^8
[O I]	$^1D_2 \rightarrow {}^3P_2$	6300	$6.3{\times}10^{-3}$
Na I D$_1$	$^2P_{1/2} \rightarrow {}^2S_{1/2}$	5896	6.2×10^7
Na I D$_2$	$^2P_{3/2} \rightarrow {}^2S_{1/2}$	5890	6.2×10^7
He I	$^3D_3 \rightarrow {}^3P_2$	5876	7.1×10^7
Fe II	$^6P_{3/2} \rightarrow {}^6S_{5/2}$	4924	3.3×10^6
Hβ	$n = 4 \rightarrow 2$	4861	3.8×10^7
Hγ	$n = 5 \rightarrow 2$	4340	1.6×10^7
Fe I	$^3F_3 \rightarrow {}^3F_2$	4132	1.2×10^7
[S II]	$^2P_{1/2} \rightarrow {}^4S_{3/2}$	4076	9.1×10^{-2}
Ca II H	$^2P_{1/2} \rightarrow {}^2S_{1/2}$	3969	1.4×10^8
Ca II K	$^2P_{3/2} \rightarrow {}^2S_{1/2}$	3934	1.5×10^8
Ultraviolet			
Mg II h	$^2P_{1/2} \rightarrow {}^2S_{1/2}$	2803	2.6×10^8
Mg II k	$^2P_{3/2} \rightarrow {}^2S_{1/2}$	2796	2.6×10^8
C IV	$^2P_{3/2} \rightarrow {}^2S_{1/2}$	1548	2.7×10^8
Si IV	$^2P_{1/2} \rightarrow {}^2S_{1/2}$	1403	7.6×10^8
O I	$^3S_1 \rightarrow {}^3P_1$	1305	2.0×10^8
S I	$^3P_1 \rightarrow {}^3P_2$	1296	4.9×10^8
Lyα	$2p \rightarrow 1s$	1216	6.3×10^8

rise in metallic emission lines and others from the Balmer series, more pronounced veiling, and elevation of the Balmer continuum.

For detailed analyses of spectra, it is convenient to have a quantitative measure of the strength of any line, whether it is in absorption or emission. This measure is the *equivalent width*. Consider first an absorption line. As usual, we let F_λ denote the specific flux received

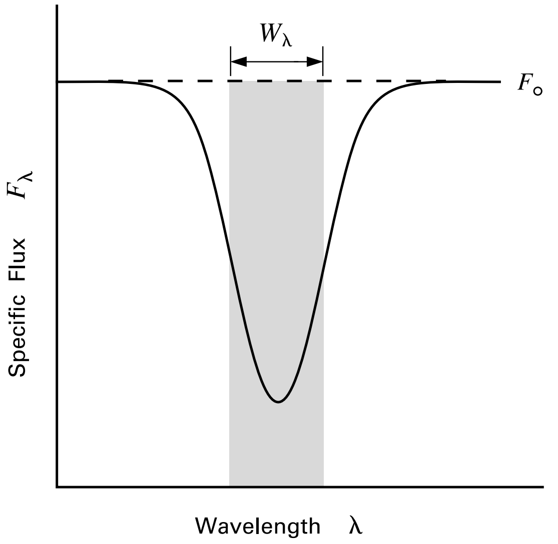


Figure 17.2 Definition of equivalent width for an absorption line.

from the star. Then we define the equivalent width by an integral over the line in question:

$$W_\lambda \equiv \int (1 - F_\lambda/F_\circ)\, d\lambda \ . \tag{17.1}$$

Here, $F_\circ$ is the continuum flux on either side of the absorption dip. Figure 17.2 shows that W_λ is the width of a hypothetical line with rectangular profile that represents the same integrated flux deficit from the continuum as the true one. In addition, the rectangular profile is chosen to be just deep enough to reduce the total flux at line center to zero. We also may apply the definition in equation (17.1) to emission lines. Here, W_λ is formally negative, and the absolute value is used in practice.

Return now to Hα in T Tauri stars and to the distinction between weak-lined and classical objects. The Hα equivalent width varies greatly from one source to another and may reach 150 Å or more in extreme cases. Conversely, the width becomes vanishingly small in an object such as V830 Tau. A T Tauri star of mid-K spectral type is considered to be weak-lined if $W_{\mathrm{H}\alpha}$ is less than 10 Å.[1] In practice, such an object also has little or no continuum veiling. It still exhibits surface activity through Ca II H and K emission, as well as X-ray flux. In addition, the Li I absorption line at 6708 Å is deeper than in main-sequence photospheres.

Table 17.2 summarizes the properties of classical and weak-lined stars. We stress that both types of objects are not only found near molecular gas, but often close to each other in the HR diagram. Thus, weak-lined stars are bona fide pre-main-sequence objects that lack some critical ingredient (or ingredients) in their circumstellar environments. We shall want to explore this issue in some depth. With reference to Hα, we note generally that the emission lines already present in weak-lined stars and very strong in classical ones are just those observed in the Sun's

---

[1] The defining criterion for weak-lined stars varies with spectral type. Consider G- and K-type T Tauri stars, both with $W_{\mathrm{H}\alpha} = 10\,$Å. The former has a greater flux in Hα, since the level of the background continuum is higher. The G star is therefore classical. Equivalently, earlier spectral types have a lower threshold in $W_{\mathrm{H}\alpha}$.

Table 17.2 Classical vs. Weak-lined T Tauri Stars

	Classical	Weak-lined
Hα emission	$\mathrm{W}_{H\alpha} \gtrsim 10$ Å	$\mathrm{W}_{H\alpha} \lesssim 10$ Å
other permitted emission lines	very strong and broad	moderately strong and narrow
forbidden-line emission	yes	no
Li absorption	yes	yes
optical veiling	yes	no
infrared excess	yes	no
X-ray emission	yes	yes
radio emission	free-free extended	nonthermal compact
photometric variability	periodic/aperiodic	periodic

chromosphere and in active, lower-mass dwarfs. Whatever physical mechanism powers main-sequence chromospheres should therefore also be operating at this earlier epoch. On the other hand, both the range and maximal extent of Hα emission from the full ensemble of T Tauri stars are so great that a purely chromospheric origin is unlikely for this line, as well as for others in the Balmer series.

We may better appreciate this fact by considering first a main-sequence G star. Here, Hα is an *absorption* line, as it is in almost all main-sequence objects.[2] If we next examine a somewhat younger object, say a G dwarf in the Hyades, Hα is still in absorption, but the line is shallower. The difference is an emission core that is superposed on an ordinary, photospheric absorption line. Suppose we plot the Hα surface flux, *i. e.*, the energy per unit time per unit area leaving the star itself, as a function of stellar age. As seen in Figure 17.3, this quantity declines with age, and in a manner at least very crudely consistent with a Skumanich-type ($t^{-1/2}$) law. Whatever its true form, the falloff only applies for $t \gtrsim 2 \times 10^8$ yr. A similar decline is seen in the strength of Ca II H and K, and of course in the surface rotational velocity (equation (16.57)). These parallel trends reinforce the belief in a chromospheric origin for *main-sequence* Hα emission.

If we now extrapolate back to solar-mass T Tauri stars a few million years old, the predicted Hα flux falls in the midst of a much broader, observed range (Figure 17.3). The most active

---

[2] The exceptions are anomalous stars of spectral type B and M, known as Be and dMe stars, respectively. Here, the "e" refers to the presence of emission lines.

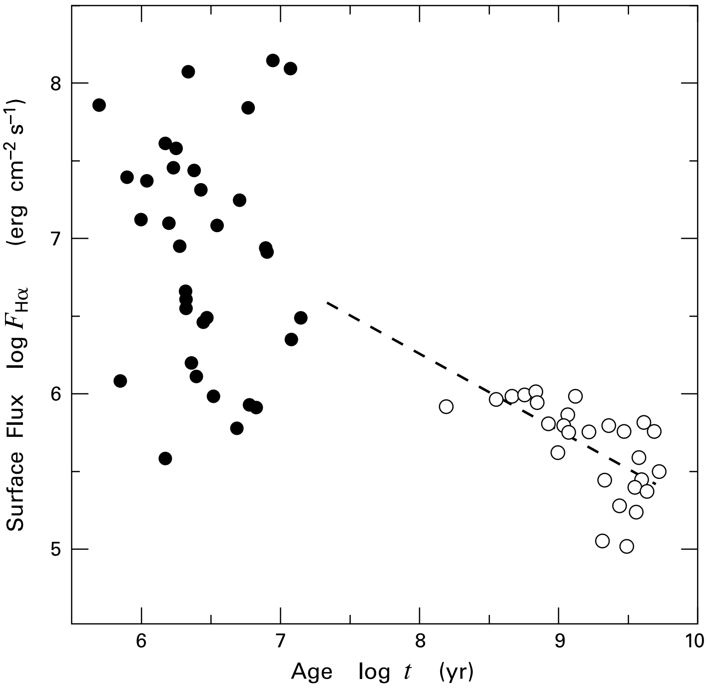


Figure 17.3 Evolution of Hα emission. Plotted is the surface flux for solar-mass stars, as defined in the text. Filled circles are T Tauri stars in Taurus-Auriga, and open ones are main-sequence objects in the field. Notice the gap between the two groups, a manifestation of the post-T Tauri problem. The dashed line represents a $t^{-1/2}$ decay.

classical T Tauri stars have fluxes larger than predicted by at least an order of magnitude. Thus, the Hα in these objects is unlikely to originate primarily in a chromosphere, although such a component is almost certainly present. It is also interesting that the *least* active T Tauri stars, *i. e.*, those designated weak-lined, have fluxes *below* the extrapolated value, but comparable to those seen at an age near 10^8 yr. Thus, any simple power law fails at early times, even for the minimal flux level. A contracting object that begins as a weak-lined T Tauri star may generate Hα from its chromosphere at a roughly steady level for as long as 10^8 yr, after which this activity gradually fades. In contrast, a classical T Tauri star starts out producing far more Hα, but the strong emission dies off quickly, within 10^7 yr.

17.1.3 Continuum Veiling

The question of what creates non-chromospheric Hα is a difficult one that is far from resolved. Before delving into the matter, let us first complete our description of the emission characteristics of T Tauri stars generally. Lines in the ultraviolet region of the spectrum also attest to elevated activity. This radiation was first extensively studied using the International Ultraviolet

Explorer (IUE) satellite, launched in 1978. The spectrum of classical T Tauri stars from 1100 to 3100 Å exhibits, at modest resolution, a smooth continuum with superposed emission lines. Transitions are partially from neutral and singly ionized species, as in the optical. The strongest flux is from the Mg II h and k doublet near 2800 Å. However, there are also lines from such ions as C IV and Si IV. These latter transitions require ambient temperatures near 10^5 K.

Both the roster of ultraviolet emission lines and even their flux ratios are similar to those observed in active dwarfs, including the Sun. This correspondence again implicates a chromosphere and corona in the pre-main-sequence objects. On the other hand, individual flux levels can be vastly greater than solar, three orders of magnitude or even more. Our reasoning concerning Hα would then suggest a non-chromospheric origin for the ultraviolet lines. Complicating this puzzle is the fact that the richness of the spectrum at the shortest wavelengths bears little relation to emission strength in the optical regime. To cite but one example, the extreme T Tauri star RW Aurigae displays few of the far-ultraviolet lines that originate from very high-temperature gas.

We stress that this apparently contradictory behavior applies only to spectral lines. The ultraviolet *continuum* flux in classical T Tauri stars is always greater than that of a weak-lined or main-sequence star, and to a degree that correlates well with optical emission activity. That is, the excess emission at the shortest observable wavelengths appears to be a smooth extension of that seen in the optical.

What creates the additional continuum radiation? We cannot hope to address this question without first establishing the wavelength dependence of the emission. That is, we must separate out the spectral energy distribution of the continuum excess from that of the underlying photosphere. This separation is currently feasible only in the optical regime, where the spectral resolution is highest, and even there requires some care. Suppose one has the spectrum of a classical T Tauri star and that of a weak-lined object with nearly the same effective temperature but negligible veiling. Then one cannot simply difference the two spectra and obtain the excess. All observed fluxes must first be dereddened to eliminate the effect of intervening dust. Dereddening, in turn, requires comparison of an object's apparent colors with those that would be seen by an observer close to the star. This correction is straightforward for the weak-lined star. In a classical T Tauri star, however, the observed colors are influenced both by reddening *and* by the presence of the continuum excess.

We may disentangle the two effects through detailed comparison of absorption line strengths in the classical and weak-lined stars. Recall that lines in the former are shallower as a result of the veiling. Furnished with high-resolution spectrograms of both objects, we may find the *relative* excess emission within small wavelength intervals.

Suppose, as in the top panel of Figure 17.4, that we have portions of a weak-lined and classical spectrum, both with a continuum flux level of F_0. Suppose further that the observed depths of a particular absorption line in the weak-lined and classical stars are δF_0 and δF_1, respectively. If we add a broadband, veiling flux ΔF_λ to the weak-lined spectrum, we preserve the depth of the line, but raise the background continuum flux from F_0 to $F_0 + \Delta F_\lambda$. We then rescale the augmented, weak-lined spectrum until the continuum level is again F_0. If we have added the proper veiling, then the new line depth matches δF_1, the one in the classical star:

$$\delta F_0 \left(\frac{F_0}{F_0 + \Delta F_\lambda} \right) = \delta F_1 \ .$$

We define the *veiling index*, r_λ, for the classical star as $\Delta F_\lambda / F_0$. The reader may verify that, in a wavelength interval where $r_\lambda = 1$, we would have $\delta F_0 = 2\,\delta F_1$. That is, the lines in the classical star would be half as deep as in the weak-lined object with no veiling.

The final step in the procedure is to take a dereddened, low-resolution spectrum of the weak-lined star and multiply it at each wavelength by r_λ. One thus obtains the properly dereddened ΔF_λ for the continuum excess. Figure 17.4 shows the calculated result for BP Tau. This is the moderately active T Tauri star whose full optical spectrum we displayed earlier. Here, we show the specific flux relative to that in a weak-lined, template star at 5500 Å. It is clear that the optical excess represents a significant fraction of the star's radiative output in this wavelength range. Discounting observational noise, the flux distribution appears featureless, and declines gently from 4000 to 6800 Å. The simplest model consistent with this result is a geometrically thin, gaseous slab with a temperature near 10,000 K. Since the absorption lines of BP Tau are only partially filled in, *i. e.*, since r_λ never greatly exceeds unity, the putative slab must cover only a small fraction (a few percent in this case) of the stellar surface.

Other classical T Tauri stars are amenable to such modeling, although the parameters of the slab vary widely. It is not difficult to find a more physical basis for this result. As we shall see in § 17.2, there is ample evidence for infalling gas surrounding many T Tauri stars. There are also, curiously enough, signs of outflow, often within the same object. The hot gas crudely represented as homogeneous slabs must actually arise in shocks located above or on the star's surface. The sparse areal coverage of a typical slab reflects a corresponding sparsity of the shock loci, where either outflowing gas impacts external material, or infalling streams collide with the surface layers.

17.1.4 Infrared Excess

The continuum excess produced by T Tauri stars spans a much broader range in wavelength than just the ultraviolet and optical. Figure 17.5 displays the full spectral energy distributions for the same three stars that were shown in Figure 17.1. The weak-lined star V830 Tau peaks in emission near 1 μm and differs little from a main-sequence star of the same spectral type. In contrast, the broadband spectrum of BP Tau has a far shallower decline toward the infrared. Finally, DR Tau has so much continuum emission at all wavelengths that the peak of its spectral energy distribution is shifted redward by a large amount.

We have encountered the infrared excess before, of course. According to the classification scheme introduced in Chapter 4, weak-lined stars such as V830 Tau fall into Class III, since they exhibit no excess in the near-infrared. The moderately active BP Tau is a Class II source. So is the extreme (or "continuum") star DR Tau, since $\lambda\, F_\lambda$ still declines between 2.2 and 10 μm. A true Class I object, we recall, has *rising* flux in this interval and a spectral energy distribution that may peak at wavelengths as large as 100 μm. We argued in Chapter 11 that these sources are also likely to be pre-main-sequence stars, but ones that are buried within especially massive and dense gaseous envelopes.

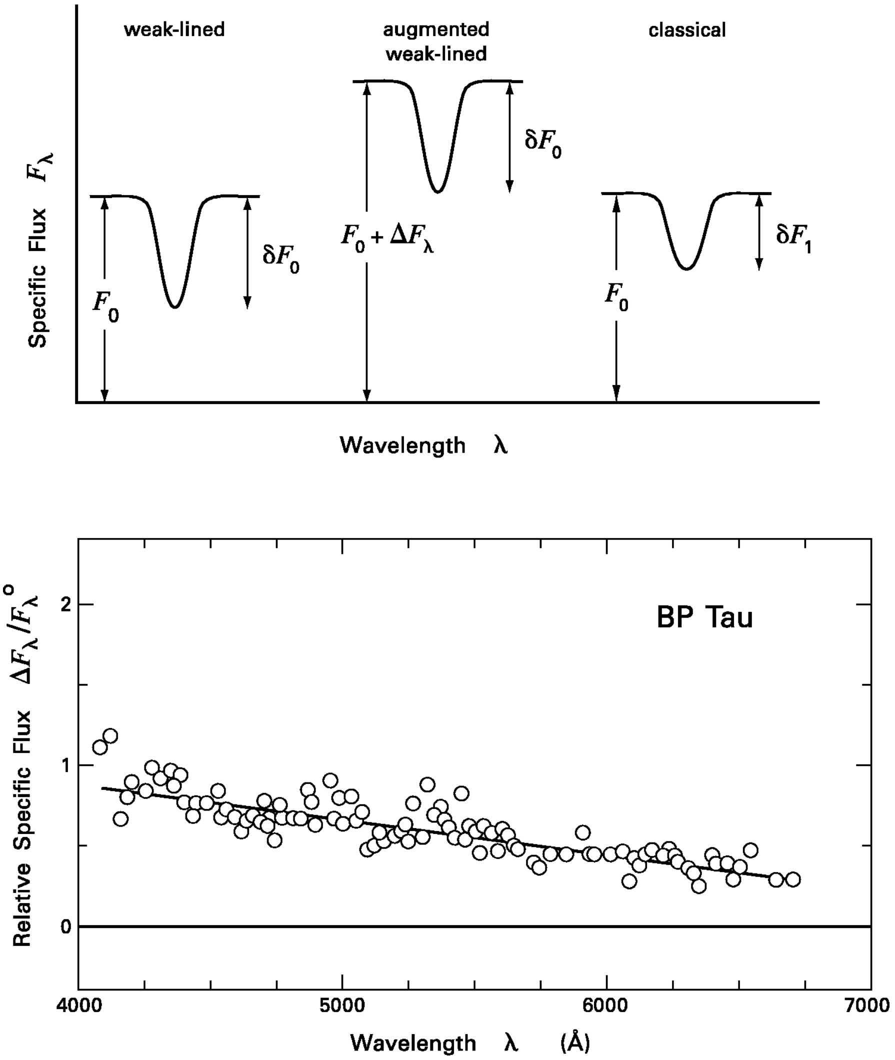


Figure 17.4 *Top panel:* Determining the veiling flux in a classical T Tauri star. *Bottom panel:* Spectral energy distribution of veiling emission in the star BP Tau. The specific flux is normalized to F_λ°, the value in a weak-lined, template star at 5500 Å. The curve is a linear fit to the data.

Returning to Figure 17.5, the star BP Tau is the only one for which we may reliably obtain an infrared excess. Weak-lined stars have essentially no excess, while the photospheric emission cannot be accurately gauged for DR Tau. Such extreme objects comprise only about 10 percent of all classical T Tauri stars. Thus, there exists a large number of documented infrared excesses. As we shall discuss in § 17.3, the most plausible source for this emission is a circumstellar disk. Much of the "excess" radiation simply represents incident stellar photons reprocessed by dust

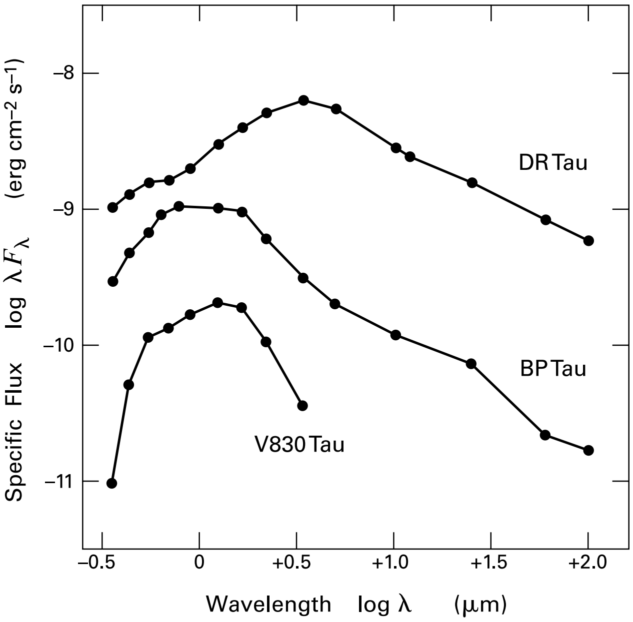


Figure 17.5 Broadband spectral energy distributions for the three stars in Figure 17.1. The spectra for DR Tau and V830 Tau have been displaced up and down, respectively, by $\Delta \log F_\lambda = 0.5$. Note that the fluxes for BP Tau and V830 Tau have been dereddened for interstellar extinction.

into the infrared. In some cases, the disk itself may be supplying additional energy. If so, the underlying physics is still unclear.

Having an infrared excess changes a star's intrinsic colors, as measured in the standard bandpasses. We discussed in Chapter 4 how near-infrared color-color plots are useful for picking out the classical T Tauri stars within populous groups. Figure 4.2 shows, in particular, that the observed $H - K$ values are too large relative to $J - H$ for the objects of interest, even after accounting for interstellar reddening. Of course, a reliable estimate for A_V allows one to deredden any single color index. Comparison with a main-sequence color then gives a quantitative measure of the amount of excess.

A dereddened color, such as $(H - K)_\circ$ or $(K - L)_\circ$, is thus useful and far more convenient than a direct assessment of the total, additional luminosity at longer wavelengths. Figure 17.6 compares one such index, $(K - L)_\circ$, with another, ostensibly unrelated one, the equivalent width of Hα. Here, the sample population consists of stars in Taurus-Auriga. Despite the large scatter, there is evidently a relation between the two numbers. Weak-lined stars, *i. e.*, those with $W_{\mathrm{H}\alpha} \lesssim 10\,\text{Å}$, have $(K - L)_\circ$ ranging from 0.1 to about $0.3\ \mathrm{mag}$. This is also the expected range for main-sequence stars of the appropriate spectral types. Classical T Tauri stars with higher Hα fluxes show correspondingly stronger infrared excesses. This trend illustrates the fact that the permitted line emission and the continuum infrared radiation both depend on the presence of circumstellar matter.

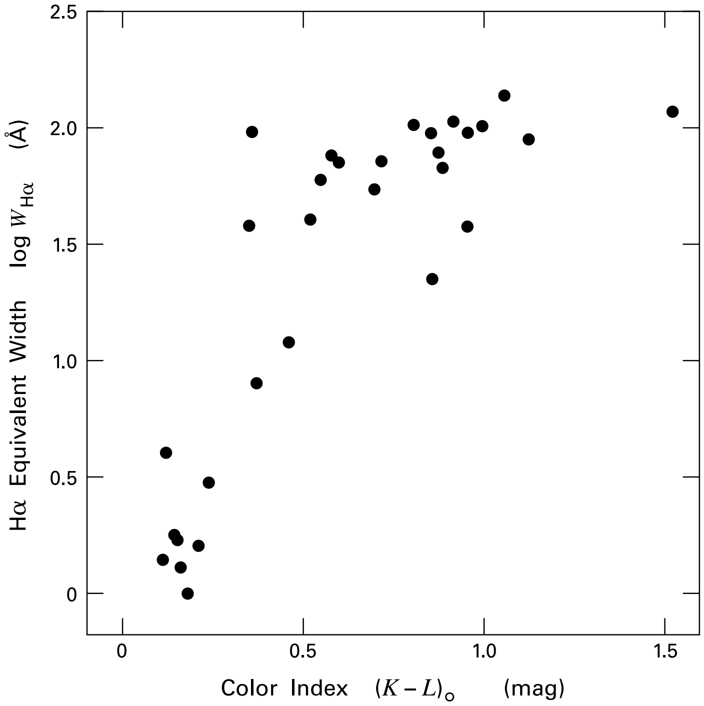


Figure 17.6 The Hα equivalent width of T Tauri stars plotted against their dereddened, $K - L$ color index. All stars are located in the Taurus-Auriga association.

17.1.5 X-Ray and Radio Flux

Let us finally examine two types of continuum radiation that are closely related to the stellar magnetic field. We have noted that X-ray activity is universal in T Tauri stars and has been the principal means of identifying weak-lined objects. We previously discussed, in Chapter 7, how this radiation is an important heating mechanism for molecular clouds. The flux from any individual star is mostly steady in time, with modest fluctuations, but is punctuated occasionally by strong flares. These represent a significant portion of the longterm energy output. The steady-state luminosity, L_X, typically constitutes about 10^{-4} of L_{bol}. However, the true fraction varies considerably from source to source. Few have L_X/L_{bol} exceeding 10^{-3}, which is also the observed upper limit for low-mass, main-sequence stars.

Figure 17.7 displays the spectral energy distribution of steady-state emission from one of the strongest known pre-main-sequence X-ray sources, V773 Tau. This weak-lined star is also the primary in a spectroscopic binary. The X-ray spectrum peaks near 1 keV and exhibits the characteristics of thermal bremsstrahlung. A plasma with temperature $T_X = 1.5 \times 10^7$ K and column density $N_H = 3 \times 10^{21}\ \mathrm{cm}^{-2}$ provides a good fit, as shown. Both of these values are normal for T Tauri stars generally. On the other hand, L_X is estimated at $1 \times 10^{31}\ \mathrm{erg\,s}^{-1}$. Since $L_{\mathrm{bol}} = 3\ L_\odot$, the ratio L_X/L_{bol} is near the maximal value of 10^{-3}.

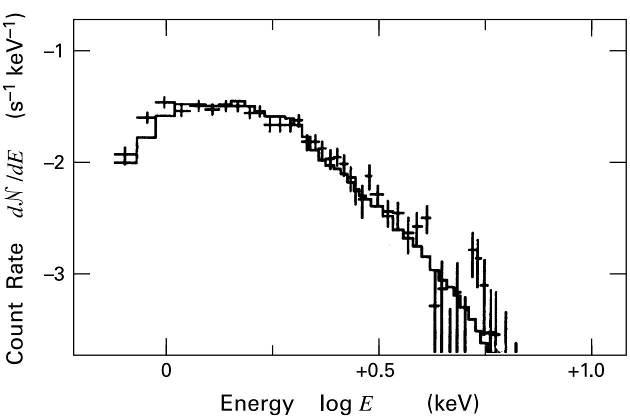


Figure 17.7 Spectral energy distribution of the steady-state X-ray emission from V773 Tau. The continuous, step-like curve represents the calculated emission from a uniform plasma.

Flares from this source are also much stronger than average. Rising within minutes, the flare luminosity can reach 10 percent of the stellar value, before fading in a matter of hours. At the peak of emission, the spectral energy distribution is characterized by $T_X \gtrsim 10^8$ K. Such hardening of the spectrum, while extreme in this case, is also a general feature. Within any populous group of T Tauri stars, about 5 percent exhibit flaring at any time.

When considering the origin of the X-ray emission, it is helpful to compare with the Sun, for which there is a wealth of detailed information. Here we also find steady-state and flaring components, both at a much reduced level. The steady-state portion arises from the Sun's outflowing corona and is characterized by a T_X of order 10^6 K. The temperature associated with a flare is again higher. Individual flares originate within closed magnetic field loops. The loop endpoints, embedded in the photosphere, move about as a result of subsurface convection. Stress within the loop eventually leads to reconnection and a sudden release of energy.

How much of this picture applies to T Tauri stars? One key empirical fact is that a large L_X-value does *not* imply strong Hα emission. (Remember that our previous example, V773 Tau, is a *weak-lined* star.) Nor does L_X correlate with the ultraviolet, optical, or infrared continuum excess. Thus, the observed X-rays do not depend on the presence of circumstellar matter. The radiation must originate, as in the Sun, from surface activity involving a dynamo-generated magnetic field. On the other hand, the spectral temperature T_X associated with the steady-state component of emission is too high to be coronal in origin. It appears instead that even this "quiescent" output is actually the sum of numerous, unresolved flares. These may again originate from reconnection in closed loops, but are only noticeable as discrete events when anomalously strong.

Magnetic fields are implicated in yet another emission feature, that in the radio continuum. We have seen at other points in this book how thermal bremsstrahlung (free-free radiation) at centimeter wavelengths is a valuable tracer of ionized winds, both from deeply embedded stars creating shocks in the surrounding medium (Chapter 13) and from exposed, massive objects that radiatively accelerate hot gas from their surfaces (Chapter 15). Only a few T Tauri stars in the nearest associations have detectable free-free emission. Observations using the Very Large Array (VLA) interferometer have established that this flux originates from a region of dimension 10^{14} cm, and is associated with optical jets.

Interestingly, it is only *classical* T Tauri stars, or rather a small subset of these, that exhibit such extended, ionized flow. A higher portion of weak-lined stars produces centimeter flux, but emanating from a much smaller volume. Moreover, the output here is more variable in time and occasionally exhibits flaring. One dramatic example is DoAr 21, a relatively luminous, weak-lined star in ρ Ophiuchi. Here the quiescent flux at 6 cm varies between 2 and 5 mJy. Over the course of several hours in 1983, the level was caught rising from 34 to 48 mJy.

What is the nature of the compact radio emission? Consider again V773 Tau. This star not only produces copious X-rays, but is also the brightest radio source in Taurus-Auriga, with numerous, documented flares. Interferometric measurements with subarcsecond resolution at 6 cm reveal the quiescent, *i. e.*, moderately variable, emission to originate within 15 stellar radii. As a function of frequency, the spectrum is essentially flat in this regime. We thus immediately discount optically thick radiation, either in a static atmosphere or an ionized wind. (Recall the discussion of the spectral index in Chapter 15.) Moreover, the high radio flux cannot represent the long-wavelength portion of an optically thin spectrum produced by bremsstrahlung, as the implied X-ray intensity would then be far greater than what is observed. One is forced to conclude that the emission is nonthermal. Note that the radio continuum from V773 Tau is circularly polarized, a telltale sign of magnetic influence. A likely emission process is gyrosynchrotron emission from electrons spiraling about stellar field lines.

It is tempting, of course, to seek a common origin for the activity in radio and X-ray wavelengths. The temporal pattern seen in both – longterm, quiescent emission interspersed with flares – is encouraging in this regard. So too is the theoretical prediction that magnetic reconnection can release a shower of energetic electrons. On the other hand, radio flaring is much less frequent than that in X-rays. It is also rather surprising that the magnetic field lines apparently responsible for the steady-state radio emission extend as far as 15 stellar radii. This is much greater than the loop sizes inferred for X-ray production, so that the two types of emission must arise in different locations.

17.2 Outflow and Infall

Early in the investigation of T Tauri stars, it was observed that many optical emission lines have profiles indicating the presence of outflowing gas. Even from the crude estimates available, it was clear that the typical rate of mass efflux was vastly greater than solar. The idea thus gained acceptance that powerful winds are a key aspect of pre-main-sequence evolution. Corroborating this picture was the later finding that younger, more embedded objects drive even more energetic flows. Other spectroscopic evidence, however, showed indisputably that gas is sometimes approaching the parent star. More curious yet was the discovery that the same object can exhibit both types of motion. The circumstellar environment is therefore both geometrically and dynamically complex.

In this section and in much of the remaining chapter, we shall present the main lines of evidence that have been used to develop a picture of the matter exterior to T Tauri stars. We stress at the outset that the process is ongoing, as no fully self-consistent physical description yet exists. Where appropriate, we will summarize quantitative models that have proved useful. At other points, we offer qualitative ideas that may serve as guidelines for future research.

17.2.1 Hα Profiles

As we have indicated, the study of line profiles has been the chief source of information. Let us accordingly begin with the most prominent emission line of all, that of Hα. Indeed, the very high Hα flux seen in classical T Tauri stars is one of the primary issues for theory to address. The essential requirement to obtain this emission is a region outside the photosphere with a temperature from 5,000 to 10,000 K, high enough to populate significantly the $n = 3$ level of hydrogen, but too low for complete ionization. One might postulate a layer analogous to an ordinary chromosphere,but with a higher temperature minimum and greater geometrical thickness. Such a structure, however, does not yield the requisite emission in the line. As one proceeds outward into lower densities, the $n = 3$ level population cannot be maintained through collisions, even if the local kinetic temperature continues to rise. The increasing $n = 2$ population causes absorption of Hα photons. The line profile develops a deep, central trough, and the total flux is diminished.

Usually, observed Hα profiles are not so deeply self-reversed, but their structure is of interest nonetheless. In fact, a range in line shapes is seen, and that of a given star may change in a matter of days. Figure 17.8 shows two representative profiles, whose shapes together account for the majority of the observations. About a fourth of the profiles are essentially symmetric, as illustrated by FM Tau in the top panel. The greatest number, roughly a third of the total, are asymmetric and resemble the CI Tau profile in the lower panel. Here, the main peak is still close to line center. A secondary maximum of relatively small amplitude is located at a shorter wavelength. This double-peaked structure is created by an absorption dip superposed on an underlying symmetric emission profile with broad wings. The dip, in turn, must arise from a partially opaque layer of gas moving *away* from an interior source of emission. An observer within such a layer sees incoming radiation shifted to the red because of the Doppler effect. Absorption therefore occurs for bluer photons whose wavelength is shifted back to line center in the comoving frame. To summarize, a profile like that of CI Tau in Figure 17.8 reveals the presence of a stellar wind.

High velocities are also evident from the overall width of the Hα line. For the symmetric emission from FM Tau, this width (FWHM) is 200 km s^{-1}. Thermal broadening alone would require a temperature of 9×10^5 K, great enough to ionize all hydrogen, and far above our previously estimated range. As we discussed in § 17.1, even higher temperatures *are* necessary to produce the X-ray emission. This component must originate in a different region from the Hα. Note finally that the broadening of the latter cannot be solely turbulent in origin, as internal shocks would quickly damp the eddies. In summary, the large Hα line width must arise from bulk motion within the emitting gas.

What is the nature of this bulk motion? First, it is important to appreciate that the offset from line center of the blueshifted dip measures directly the speed of absorbing wind material.

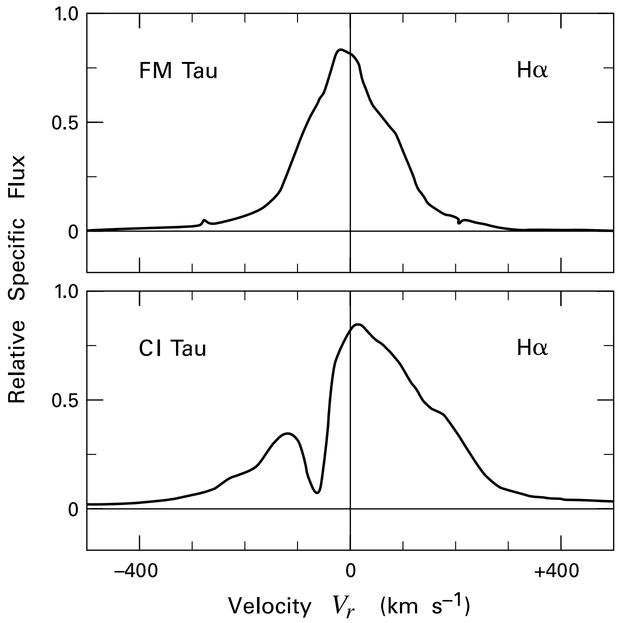


Figure 17.8 Two representative Hα profiles. Note that only *relative* specific flux is shown. Line center is indicated by a vertical line.

In the case of CI Tau, this displacement is 70 km s^{-1}. The velocity obtained in this manner is consistent with that inferred through another diagnostic, optical forbidden lines (see below). On the other hand, the full Hα profile of CI Tau extends to velocities, both blue- and redshifted, that are substantially higher. The *emission* of Hα, therefore, stems from some type of non-turbulent motion that may bear little resemblance to the smooth stellar wind relatively far from the star. The latter only provides partial *absorption* of outgoing radiation.

For both components, a key question is what heats the gas. We faced the same issue in Chapter 13, when discussing the optical jets from more deeply embedded stars. There we found that broadened line radiation, including Hα, arises from oblique shocks that partially interrupt the flow. Oblique shocks may also energize Hα emission in classical T Tauri stars, whether or not these shocks arise in jets. The flow close to the star is constrained to follow a magnetic field that is probably quite complex, consisting of both tight loops and more open portions. Recall from § 17.1 that those stars with the most intense Hα emission also have a large supply of circumstellar material, as revealed by their near-infrared excess. The system of shocks arising from impact of the outflowing gas with external matter could yield a broad, symmetric line, provided the shocks have a large range of inclinations. Short-term fluctuations in the field topology, such as those observed on the Sun, might then account for the temporal changes.

Consider once more the blueshifted dip, marking the wind component. In the well-studied winds of main-sequence stars and giants, the analogous absorption feature is often much stronger, so that the flux in the blue region goes below the adjacent continuum. Such a *P Cygni profile* is rarely seen in the Hα from T Tauri stars. There must be coherently outflowing matter in such objects as CI Tau, but the shallower absorption means that this exterior gas has a relatively low column density. At the same time, its temperature must still be high enough to populate significantly the $n = 2$ level of hydrogen. Our picture, then, is that Hα emission arises

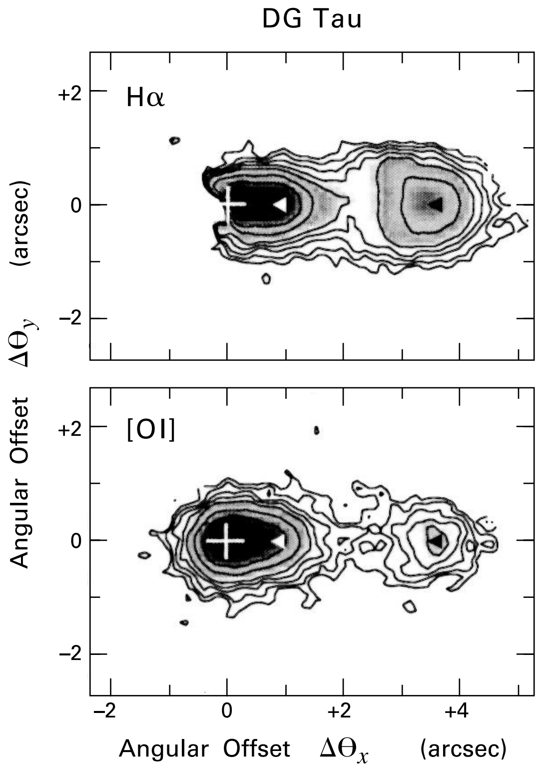


Figure 17.9 Microjet from DG Tau, as seen in blueshifted Hα (*top panel*) and [O I] 6300 Å (*bottom panel*). The white cross marks the position of the star. Filled triangles represent the positions of maximum velocity and the bowshock apex. The angular offsets are relative to the star.

primarily from numerous shocks within a central location relatively close to the star. There is also, at least in many sources, shock-heated gas in a more distended and well-directed wind. This latter material accounts for the absorption feature in the profile.

It is important to note in this context that there are some classical T Tauri stars, although a distinct minority, that drive more extended flows, of the kind described in Chapter 13. We may recall the example of AS 353A, where Hα is seen to emanate from several distinct bowshocks, including the bright HH 32A. These Herbig-Haro objects attest to wind collimation over a distance of about 0.1 pc. Most T Tauri stars lack the environment necessary to produce such narrow, elongated flows.

Powerful imaging techniques, such as adaptive optics, have revealed smaller-scale features in the Hα emission, even for stars without an extended jet. The top panel of Figure 17.9 shows a map of the blueshifted Hα from the classical T Tauri star DG Tau. Here the velocity interval covered is from 100 to 250 km s^{-1} relative to the star. We note, from the greyscale display, that the strongest flux emanates from an unresolved central source. There is also a local maximum to the right, as well as a connecting bridge to the inner region. The whole projecting structure is known as a *microjet*. Near the terminus of the flow, the contours resemble the bowshocks driven by embedded objects. (Recall, *e. g.*, Figure 13.5.) In the present case, the transverse arc is located only $4''$, or 560 AU, from the star.

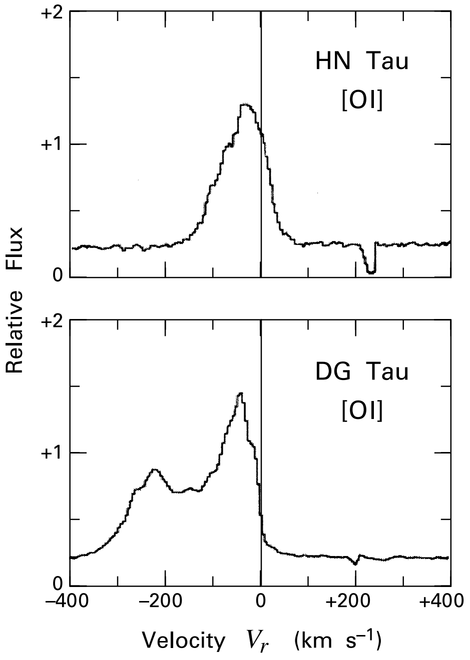


Figure 17.10 Two representative line profiles in [O I] 6300 Å. The thin vertical line demarcates the stellar rest velocity.

17.2.2 Forbidden-Line Emission

Equally detailed, complementary information on T Tauri environments is available from another source: forbidden-line emission. As already noted, this radiation emanates from relatively low densities, so that it specifically probes the wind region. Table 17.1 includes several of the lines most frequently used – the [O I] transition at 6300 Å and the [S II]$\lambda\lambda\,6716, 6731$ doublet. At the characteristic line formation temperature of 10^4 K, the associated critical densities are of order $n_e = 10^6$ and $10^4\ \mathrm{cm}^{-3}$, respectively. Note that these densities refer to *electrons*, the species most effective at collisional deexcitation.

The strongest emission is in [O I] 6300 Å. About 20 percent of T Tauri stars have appreciable flux in the line, *i. e.*, equivalent widths exceeding 1 Å. (The largest equivalent widths found are about 10 Å, an order of magnitude below the Hα maximum.) Figure 17.10 displays two of the most commonly observed profile types. The upper example has a single peak, but the line centroid is shifted to the blue. The lower profile, that of DG Tau, has a pronounced, secondary peak. Here again, almost all the emission is blueward of the stellar photosphere.

The blueshift of forbidden-line emission is a general characteristic that holds for other sources and other observed transitions. To appreciate its significance, we first note that the low densities of their emitting regions mean that forbidden lines are optically thin. Thus, the radiation stems from the entire volume of excited gas. The absence of redshifted emission means that parcels of retreating wind material are blocked from view by some extended, opaque object. This object is likely to be the circumstellar disk.

Aside from the presence of a disk, what do forbidden-line profiles tell us about the wind itself? In both examples shown, the full line width is several hundred km s^{-1}, similar to what we found using Hα. The more detailed wind morphology, including the distribution of gas density and velocity, is difficult to ascertain from the profile alone. Nevertheless, it has long been accepted that the presence of a secondary peak, such as that of DG Tau, represents a directed portion of the flow, with a relatively small spread in velocity. Additional evidence comes from long-slit spectroscopy, which effectively yields many contiguous line profiles along a fixed direction. This technique indicates the presence of additional material closer to the star that is also emitting in the forbidden lines. The gas here has a broader range in speeds but is, on average, moving more slowly.

These inferences from spectroscopy have been supplemented by high-resolution spatial mapping. Interestingly, the exterior wind component appears to coincide with that in Hα. The lower panel of Figure 17.9 is an image of the DG Tau region in [O I] 6300 Å. Both the spatial resolution and the selected velocity interval are the same as in the top panel. We see again that the emission stems from a jet-like structure that terminates in an apparent bowshock. To confirm the latter identification, the reader should compare Figure 17.10 with Figure 13.3, the line profile for HH 32A in Hα. The two profiles are essentially identical, apart from reversal of the velocities. Images of DG Tau in [O I] taken several years apart reveal that the terminal knot is moving outward at 200 km s^{-1}. Thus, the shock itself is a relatively weak one that arises from velocity fluctuations in a wind yet further inside. The lack of higher-excitation lines, such as [O III] emission, supports this view.

The discovery of microjets, first through spectroscopy and then through direct imaging, is an important step in our understanding. All of the optical emission lines associated with T Tauri stars require temperatures close to 10^4 K. Images such as those of Figure 17.9 tell us that at least some of the excitation occurs in shocks created by variable, jet-like winds. It is not yet clear what fraction of T Tauri stars possess flows with this structure. Nor do we understand the heating mechanism for the inner, unresolved portion of line emission. If, as we have argued in the case of Hα, shocks are again responsible, their morphology is unknown. We should also recognize that Hα and other permitted lines may well be optically thick in the central region. This circumstance adds to the difficulty of theoretical modeling.

Within the sparser, more extended portion of the flow, we may utilize the forbidden-line emission to quantify the mass transport along the microjet. The method here is the same as that described in § 13.5 for jets from embedded stars. That is, we first obtain the ambient electron density from the ratio of [S II] 6716 and 6731 Å fluxes. This density, in conjunction with the observed luminosity in [O I] 6300 Å, yields the mass of emitting gas (equation (13.40)). Finally, proper motion of the bowshock and the projected length of the microjet give the crossing time. The result of this exercise for the DG Tau microjet is that $\dot{M}_j$ is $8 \times 10^{-9}\ M_\odot\ \mathrm{yr}^{-1}$. Within a factor of two, this figure must represent the total wind loss rate, $\dot{M}_w$. The rate is comparable for other T Tauri stars, but with substantial variation from source to source.

Not all T Tauri stars have detectable emission in optical forbidden lines. To see the trend, Figure 17.11 plots the equivalent width in [O I] 6300 Å against our surrogate for thermal disk emission, the $(K - L)_\circ$ color index. The sample of stars displayed here is the same as in Figure 17.6. Apart from a systematic reduction in equivalent widths from those in Hα, the pattern is similar. There is a clear correlation between the line emission strength and the near-infrared excess. Furthermore, stars with low or undetectable [OI] flux are again those with $(K - L)_\circ \lesssim 0.3\ \mathrm{mag}$, *i. e.*, weak-lined objects that appear to lack dusty, circumstellar disks.

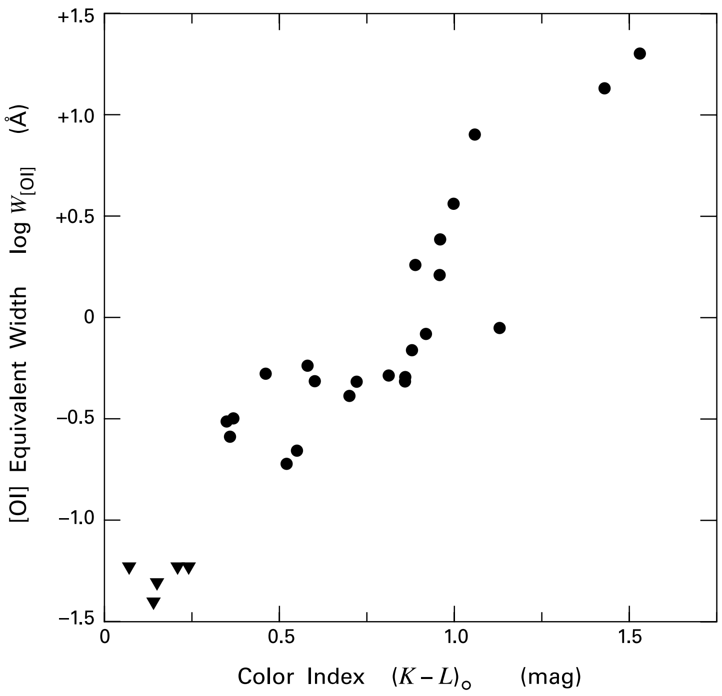


Figure 17.11 The [O I] 6300 Å equivalent width of T Tauri stars plotted against their dereddened, $K - L$ color index. The stars are the same as in Figure 17.6. Inverted triangles represent upper limits.

How should we interpet these facts? Since forbidden lines trace winds, many have concluded that circumstellar disks actually create these winds. Thus, stars without the near-infrared excess signifying disks would also have no outflowing matter and hence no forbidden-line emission. The exact causal link from disks to winds is unknown. As we described in Chapter 13, it may be that the disk themselves fling off material centrifugally, as this matter spirals inward on its way to the star.

Whatever the appeal of this idea, we should recognize that it does *not* follow inexorably from the data at hand. What we actually observe is that classical T Tauri stars, many of which certainly have disks, also emit in forbidden lines. Suppose, for the sake of argument, that the associated winds do not originate from disks, but from the stars themselves. Then weak-lined objects with age and rotation period similar to classical ones may have comparable winds. However, these flows would also be more isotropic than the observed microjets, since the parent stars lack the circumstellar matter that apparently provides collimation. Such winds (for which, we stress, there is no evidence to date) might not have detectable levels of forbidden-line emission. A search for cooler, outflowing gas from weak-lined stars would therefore be of potential interest.

Even the winds of classical T Tauri stars are almost certainly composed of largely neutral gas. It is only when this material is locally heated, through shocks, that it becomes partially ionized and produces the optical lines by which the wind itself is revealed. Collisional ionization also yields radio free-free emission, seen in such objects as DG Tau. Incidentally, the same process may also explain why classical T Tauri stars lack *nonthermal* radio luminosity. A modest amount of plasma can quench, through free-free absorption, radio photons from the deeper interior. It may be, then, that most T Tauri stars actually produce synchrotron emission, but that this component only emerges in the relatively clear environments of weak-lined objects.

17.2.3 YY Orionis Stars

Throughout this discussion, we have distinguished two spatial domains surrounding T Tauri stars: one in which there is a well-developed wind, and another where the geometry of the flow is poorly understood. We have also stressed that the bulk of permitted-line emission, including Hα, originates in the second, interior volume. The complexity of this region is underscored by the fact that many stars exhibit *redshifted* absorption in their optical lines. From our previous discussion, a dip in the profile at a wavelength greater than line center signifies motion toward the star, *i. e.*, gravitational infall.

For this spectral feature to reveal itself, the star must have strong and broad emission lines. Thus, the phenomenon is essentially limited to classical T Tauri stars. Objects with the spectroscopic signature of infall constitute the *YY Orionis* subclass. By definition, the profile distortion must occur in one of the Balmer lines of hydrogen. Members of this group invariably show infall in additional lines, such as Ca II H and K or the Na D doublet.

Figure 17.12 displays the pattern of emission in a typical YY Orionis star, S CrA. Like others in its class, this object has especially strong continuum veiling, with few absorption lines in the blue region of its spectrum. Here we show the lowest four emission lines in the Balmer series. The profile of Hα does *not* have a redshifted absorption dip, but a modestly blueshifted one, similar to that of CI Tau in Figure 17.8. Redshifted absorption first appears in Hβ, along with a dip closer to line center. The infall signature deepens in the higher members of the series. Notice that the velocity associated with this feature is nearly the same from line to line, about 250 km s^{-1}. Notice too that the redshifted absorption dip, when present, goes below the continuum flux level. The complex profiles of Hβ, Hγ, and Hδ are thus said to exhibit both P Cygni and *inverse* P Cygni character.

As we seek to understand the trends here, one important consideration is optical depth. A glance at Table 17.1 shows that the Einstein A-value declines from Hα to Hβ; this falloff continues for the higher lines. Since A is directly proportional to optical depth (see equation (C.15)), the medium is most opaque to Hα. The radiation in this line stems from more exterior gas. That the corresponding profile exhibits only a wind signature reaffirms that infall must occur relatively close to the star. Here, gas flowing along closed, magnetic loops could be returning at high speed to the surface.

Although the Hα profiles of YY Orionis stars do change with time, they are relatively stable compared to the higher transitions. In S CrA, the redshifted absorption in Hδ disappears and reemerges in a matter of hours. It seems that the infall is not a steady phenomenon, but that it fluctuates, probably along with outflow occurring on a similar scale. Complex and variable Balmer profiles also occur in DR Tau, our prototype of an extreme T Tauri star.

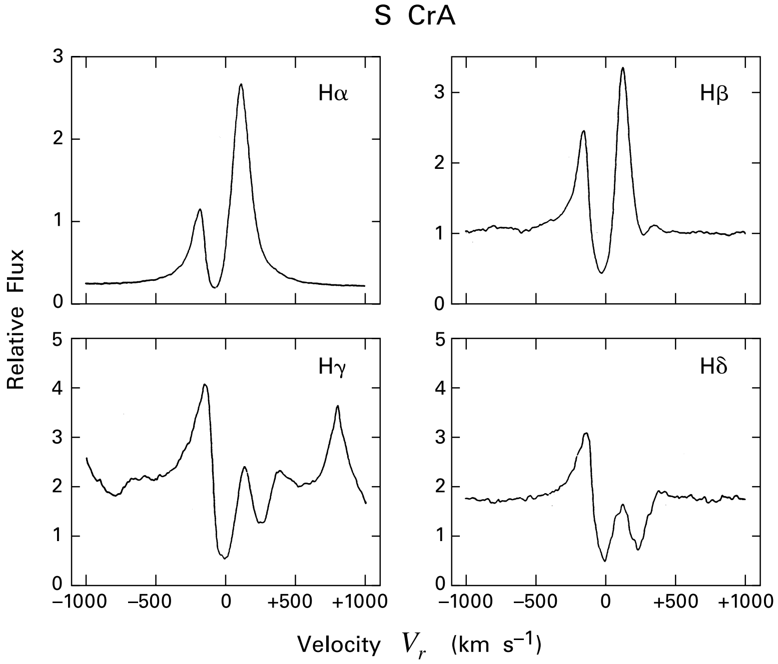


Figure 17.12 Balmer line profiles of the YY Orionis star S CrA. The outlying absorption and emission lines in the Hγ panel are due to Fe I and Fe II, respectively.

The YY Orionis phenomenon is not rare. As many as half of the classical T Tauri stars belong to this group, as traditionally defined. In other stars, inverse P Cygni profiles are seen in longer-wavelength hydrogen lines, such as the near-infrared members of the Paschen series. Infalling gas must be a common trait, provided one probes close enough to the central object. Those stars exhibiting redshifted absorption in more optically thick lines have, in some sense, more extensive infall. It is surely no coincidence that these same objects are among the most variable spectroscopically and have higher levels of continuum veiling.

17.2.4 Origin of Continuum Veiling

We noted earlier that the continuum output of a classical T Tauri star can be modeled crudely as arising from a homogeneous slab that partially covers the stellar surface. The spectroscopic indications of rapid motion, both inward and outward, lend credence to the idea that this flux originates in radiating shocks. Theorists have progressed beyond slab models and determined numerically the emission from such shocks, but only for the pure infall case. At the base of each magnetic flux tube, the collision of matter with the stellar surface is reminiscent of that occurring in an embedded protostar. The speed attained just before entering the shock front is again close to free-fall, as long as the gas starts from a distance of a few stellar radii or greater. Thus, the immediate postshock temperature is also similar, of order 10^6 K. This is high enough for the gas to emit in the soft X-ray regime. On the other hand, the total infall rate from all the flux tubes, $\dot{M}$, has to be much less than the protostellar value, so that the accretion luminosity does not overwhelm the stellar component.

The accretion calculations help to explain the origin of the 10,000 K temperatures needed within the homogeneous slab models. Each shock actually creates *two* distinct sources of outgoing, continuum flux. Most of this energy originates from the stellar surface. The gas here radiates more than usual because it is locally heated by the X-rays produced within each postshock region. (Recall Figure 11.5 and the discussion of § 11.2.) Shock-generated photons also stream outward, where they encounter gas falling inside each flux tube. As the X-rays are absorbed, they ionize and heat the material to peak temperatures of 10,000 to 20,000 K. Optical radiation from this preshock gas constitutes another component of the total excess emission.

Matching the observed optical continuum from any T Tauri star requires that $\dot{M}$ be of typical magnitude $10^{-8}\ M_\odot$ yr^{-1}. It is encouraging that this rate is consistent with that for the winds, as derived from observations of microjets. The infall models also find the flux tubes to cover a percent or so of the stellar surface. A more detailed treatment of the magnetized surface layers will be essential for improving these shock calculations. We need to know, for example, the temperature at the base of a flux tube containing *no* infalling matter, to gauge accurately the heating effect from accretion. Here, T_{eff} must be depressed *below* that of the surrounding gas, as is the case on the Sun. Observations of stellar variability make it clear that many starspots are indeed cooler, while others are elevated in temperature by several thousand degrees above the unperturbed photosphere. (See § 17.4 below.) The basic idea of multiple accretion shocks is thus attractive. A similar calculation of wind shocks will eventually be necessary to account for both the profiles and overall strengths of permitted emission lines.[3]

17.3 Circumstellar Disks

While shocks may explain the optical continuum in T Tauri spectra, they have little relevance at infrared and longer wavelengths. It is true that shocks produce radio free-free photons, but the total luminosity of this component is minor. Figure 17.5 shows, on the other hand, that the infrared excess typically comprises a large portion of the energetic output. Moreover, the amounts of emission at short and long wavelengths are evidently correlated. This empirical relationship naturally reminds us of those between optical *line* flux and the near-infrared continuum (Figures 17.6 and 17.11). All these trends reflect the importance of circumstellar matter, either for supplying material that falls onto the star, or for partially impeding outflowing gas.

At least part of this external matter lies within a orbiting disk. Images of the Orion proplyds, like those in Figure 15.19, show flattened configurations surrounding low-mass, pre-main-sequence objects. We will see later in this section that interferometric observations in the millimeter also reveal such structures.

---

[3] Within a pure accretion scenario, the preshock gas also emits in spectral lines, including Hα. The heated region, however, is far too small to yield the large observed equivalent widths.

17.3.1 Dust Emission: Optically Thick

The primary, and historically the first, source of evidence for disks is the infrared excess. To follow the chain of reasoning here, it helps to entertain a contrary hypothesis. Suppose that no disk at all were present around a classical T Tauri star like BP Tau. Could the long-wavelength spectral energy distribution of this source arise instead from a *spherical* distribution of matter? The matter of interest is dust grains, which both absorb starlight and reradiate this energy into the infrared. Ignoring the dynamics of this hypothetical dust envelope, could it explain the long-wavelength excess?

At this point, it helps to review the discussion of § 11.5 concerning more embedded stars. There we showed that the spectral energy distribution of Class I sources can, at least in some cases, be reproduced using a dust shell model (Figure 11.24b). The critical difference is that a star such as BP Tau, despite copious infrared emission, is nonetheless optically visible. Indeed, the observed optical spectrum is well fit as the sum of an ordinary photosphere and the continuum excess of Figure 17.4, together with a modest degree of stellar extinction ($A_V = 1\ \mathrm{mag}$). If we were to include *additional* dust, enough to supply the full mid- and far-infrared luminosity, the resulting column density would be much too high. That is, the A_V-value associated with this envelope would vastly exceed that derived from optical observations.

The real conclusion to be drawn is that the dust grains, which assuredly do supply the long-wavelength emission, are distributed anisotropically. Moreover, the presence of near-infrared emission shows that some of the dust is quite close to the star, within tens of stellar radii. A natural assumption is that this matter is undergoing orbital motion.

Once we accept that a disk is present and responsible for the bulk of infrared emission, we may also establish the run of dust temperature in that structure. We first separate out the contributions from the photosphere and optical continuum excess from the full spectral energy distribution. If we attribute the remaining emission solely to the disk, then the received flux is given by

$$F_\lambda \,=\, \int I_\lambda \, \mu \, d\Omega \ . \tag{17.2}$$

The integration is over the solid angle subtended by the disk. This solid angle, in turn, depends on the disk's projected surface area. Referring to Figure 17.13, we have

$$F_\lambda \,=\, \frac{2\pi}{D^2} \int_{\varpi_\circ}^{\varpi_D} I_\lambda \, \cos\theta \, \varpi \, d\varpi \ , \tag{17.3}$$

where $\varpi_\circ$ and ϖ_D are the inner and outer disk radii, respectively, and D is the distance to the system. The angle θ measures the inclination to the line of sight.

We next evaluate the specific intensity. The solution to the radiative transfer equation (2.20), applying Kirkhoff's law to the dust emissivity and opacity, is

$$I_\lambda \,=\, B_\lambda(T)\, \left[1 \,-\, \exp\left(-\Delta\tau_\lambda\right)\right] \ , \tag{17.4}$$

where $B_\lambda(T)$ is the Planck function and $\Delta\tau_\lambda$ the disk's optical thickness at each radius (see Figure 17.13). In practice, $\Delta\tau_\lambda \,\gg\, 1$ for plausible disk surface densities, as long as we restrict our attention to the infrared regime, *i. e.*, for $\lambda \,\lesssim\, 100\ \mu\mathrm{m}$. We may then neglect the second

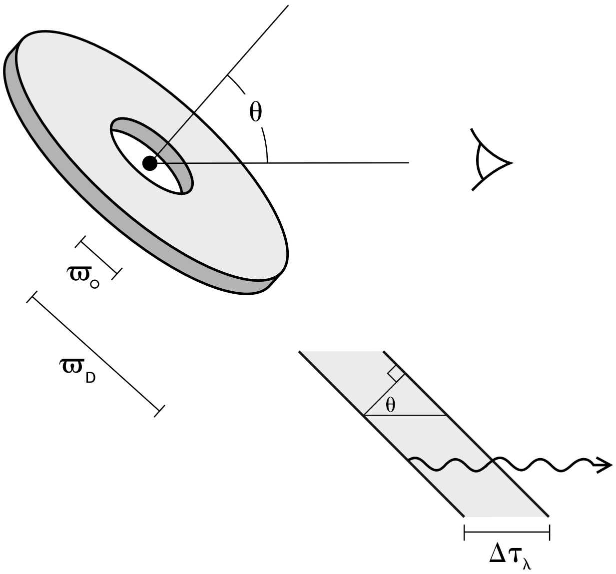


Figure 17.13 Dust emission from a circumstellar disk. Here the inner and outer radii are designated $\varpi_\circ$ and ϖ_D, respectively. The lower figure shows how the optical thickness depends on the disk's inclination to the line of sight.

factor in equation (17.4) and approximate I_λ by B_λ itself. Utilizing equation (2.29) for the latter, we find

$$F_\lambda \,=\, \frac{4\,\pi\,\cos\,\theta}{D^2}\,\frac{h\,c^2}{\lambda^5}\int_{\varpi_\circ}^{\varpi_D} d\varpi\,\frac{\varpi}{\exp\,(hc/\lambda k_B T)\,-\,1}\;. \tag{17.5}$$

As a first guess for the radial temperature distribution, we employ a power law:

$$T(\varpi)\,=\,T_\circ(\varpi/\varpi_\circ)^{-q}\;, \tag{17.6}$$

where $T_\circ$ is the inner-edge temperature and q is some positive constant. We define a nondimensional variable x as

$$x\,\equiv\,\left(\frac{h\,c}{\lambda\,k_B\,T_\circ}\right)^{1/q}\frac{\varpi}{\varpi_\circ}\;, \tag{17.7}$$

so that equation (17.5) becomes

$$F_\lambda \,=\, \frac{4\,\pi\,\varpi_\circ^2\,\cos\,\theta}{D^2}\,\frac{h\,c^2}{\lambda^5}\left(\frac{h\,c}{\lambda\,k_B\,T_\circ}\right)^{-2/q}\int_{x_\circ}^{x_D} dx\,\frac{x}{\exp\,(x^q)\,-\,1}\;. \tag{17.8}$$

In the nondimensional integral, the upper limit x_D can be set to infinity, because of the steep falloff in the integrand for large x. The corresponding falloff toward the inner edge allows us to take $x_\circ$ as zero, provided the disk extends inward all the way to the star. With these simplifications, the integral depends only on the index q.

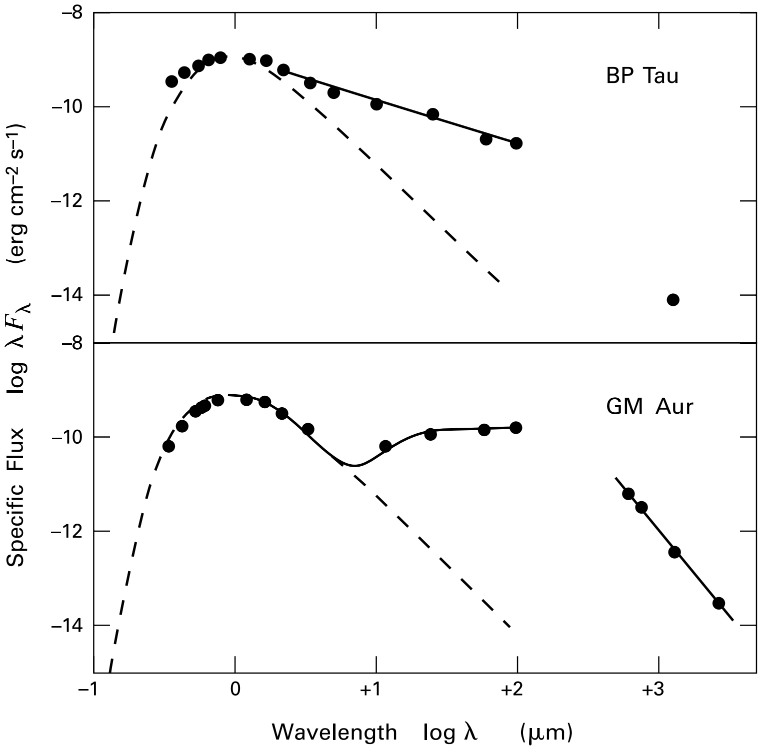


Figure 17.14 Spectral energy distributions of a star with evidence for a dusty disk (*upper panel*) and for a disk with a central hole (*lower panel*). Solid curves are derived from the theory presented in the text. Dashed curves represent the photospheric contributions of both stars.

The spectral energy distribution is generally displayed as λF_λ. From equation (17.8), the wavelength dependence of this product is

$$(\lambda F_\lambda)_{\rm disk} \propto \lambda^{(2-4q)/q} \, , \qquad \textit{optically thick} \qquad (17.9)$$

where we have emphasized that this portion of the flux arises from the disk alone. How does this predicted relation compare to observations? In the upper panel of Figure 17.14, we return to our prototypical example, BP Tau. The data are identical to those in Figure 17.5, but we have now added a point at 1.3 mm. We have also indicated, by a dashed curve, the photospheric contribution, here idealized as a blackbody at the stellar effective temperature of 4060 K. Apparently, the infrared flux does fall with wavelength in nearly power-law fashion; our straight line represents the best fit between 2 and 100 μm. From the slope of this line, we may use equation (17.9) to derive a q-value of 0.65.

For some stars, the infrared data clearly indicate a departure from our simple prescription for the temperature. The anomaly most often encountered is a depression in the flux at near- and mid-infrared wavelengths. In the lower panel of Figure 17.14, we use as an example the star GM Aur. It is again apparent that the observed flux greatly exceeds the photospheric contribution at longer wavelengths. However, the actual wavelength dependence of λF_λ is only reasonably fit by a power law beyond the infrared, for $\lambda > 100$ µm, where the radiation is optically thin.

What, then, creates the broad dip in the spectral energy distribution? One intriguing possibility is that this feature is telling us something about the disk geometry. Specifically, the disk in this case may be truncated before it gets too warm, *i. e.*, at an inner radius $\varpi_\circ$ that is further removed from the stellar surface. Referring again to equation (17.8), we may still set x_D to infinity, but we no longer neglect $x_\circ$. Equation (17.7) tells us, after specializing ϖ to $\varpi_\circ$, that $x_\circ$ varies inversely with wavelength. As λ decreases, the rise in $x_\circ$ causes the nondimensional integral in (17.8) to diminish in value. Accordingly, λF_λ from the disk is reduced below the prediction of equation (17.9).

Before concerning ourselves with how such an inner hole arises in practice, let us see how the idea works in the case of GM Aur. Note first that the observed spectral energy distribution flattens as λ approaches 100 µm from below. If we assume that the interior disk truncation has negligible effect by this point, then the underlying temperature distribution must have an associated q-value close to 0.5 (see equation (17.9)). After fixing q, equation (17.7), evaluated at $\varpi = \varpi_\circ$, and equation (17.8) together give us λF_λ from the disk, once we further specify θ, $\varpi_\circ$, and $T_\circ$. For simplicity, we take $\theta = 0$, corresponding to a face-on disk. We then vary $\varpi_\circ$ and $T_\circ$ until two conditions are simultaneously met. First, we require that the calculated disk flux equal the observed value at 100 µm. Second, we demand that the *total* spectral energy distribution, from both the disk and photosphere, match the empirical function for $\lambda < 100$ µm as closely as possible. The solid curve in Figure 17.14 shows the best result, obtained for $\varpi_\circ = 0.3$ AU and $T_\circ = 270$ K.

17.3.2 Dust Emission: Optically Thin

We have noted that the submillimeter portion of the spectral energy distribution for GM Aur *does* exhibit power-law behavior. It is not difficult to trace the origin of this trend, which has also been observed in numerous other sources. Since the radiation is now optically thin, equation (17.5) no longer applies. Instead, we return to equation (17.4). The smallness of $\Delta\tau_\lambda$ implies that we may now approximate I_λ as $B_\lambda\,\Delta\tau_\lambda$. For the Planck function, we utilize the Rayleigh-Jeans form, appropriate at these long wavelengths:

$$B_\lambda \approx \frac{2\,c\,k_B\,T}{\lambda^4}\ . \tag{17.10}$$

The reader may verify, again referring to Figure 17.13, that the optical thickness is

$$\Delta\tau_\lambda = \frac{\kappa_\lambda\,\Sigma}{\cos\theta}\ . \tag{17.11}$$

Here, κ_λ is the opacity due to dust grains and Σ the disk surface density in this component. Substitution of these results into equation (17.3) yields

$$F_\lambda = \frac{4\,\pi\,c\,k_B\,\kappa_\lambda}{D^2\,\lambda^4}\int_{\varpi_\circ}^{\varpi_D} T(\varpi)\,\Sigma(\varpi)\,\varpi\,d\varpi\ . \tag{17.12}$$

The wavelength dependence of F_λ is contained in the factors preceding the integral within equation (17.12). We mentioned in Chapter 2 that observations of the interstellar medium are consistent with the dust opacity declining as a fixed power of wavelength in the submillimeter regime. Denoting the relevant exponent as β, equation (17.12) predicts that the spectral energy distribution should obey

$$(\lambda\, F_\lambda)_{\rm disk} \,\propto\, \lambda^{-3-\beta}\ . \qquad\qquad \textit{optically thin} \qquad\qquad (17.13)$$

Of course, the disk contribution is the *only* significant one in this regime, as the stellar flux has fallen to comparatively negligible levels. The empirical decline in the GM Aur spectral energy distribution is consistent with equation (17.13), *i. e.*, with a nearly constant value of β. The best-fit slope, again displayed in Figure 17.14, yields $\beta \approx 0.7$. This value is less than those observed for dust in the interstellar medium, where β generally lies between 1 and 2. A shallower slope for κ_λ indicates that the grains are larger, perhaps because of coagulation that occurred within the disk. Other T Tauri stars observed in the submillimeter yield similar results.

A key difference between equation (17.12) and the analogous equation (17.5) is that only the former contains the disk's surface density. Physically, the detected submillimeter photons originate from every point along the line of sight, while optically thick radiation of shorter wavelength emanates only from a thin layer near the disk surface. Observations in the submillimeter are thus useful for probing the disk mass, M_d. What one must do in effect is invert equation (17.12), solving for the appropriate integral of $\Sigma(\varpi)$ as a function of the empirically known flux at some wavelength λ. The quantity M_d then follows by assuming the standard, interstellar ratio for the dust and gaseous mass densities.

There are several practical limitations to the implementation of this scheme. First, the actual value of the opacity κ_λ is imprecisely known, even when its wavelength dependence is well constrained from observations of a particular source (recall § 2.4). Second, it is not possible in general to pull out the temperature from the integral in equation (17.12). The calculations of M_d done so far have assumed *both* that $T(\varpi)$ obeys equation (17.6) *and* that $\Sigma(\varpi)$ itself declines as a fixed (and different) power of ϖ. While the first assumption is often well grounded empirically, as we have demonstrated, the second is not. The net result is that estimates of M_d are only reliable to within an order of magnitude. Quoted values are nonetheless of interest, since they consistently fall well below the associated *stellar* masses, usually at the level of a few percent.

17.3.3 Imaging Studies

We stress once more that the observations discussed so far are sensitive only to the *dust* component of the disk. The bulk of M_d, obtained here by extrapolation from the dust mass, consists of gas. To observe the latter, the relatively cold temperatures (varying within a given system from roughly 10 to 10^3 K) suggest that we utilize the molecular transitions that have proved so fruitful in exploring interstellar clouds. This expectation has been fully justified, with CO again playing a dominant role. High-resolution spectroscopic studies have been conducted, utilizing both near-infrared and millimeter lines. These investigations have probed, respectively, gas motion within the heated, inner portion of the disk and in the colder material comprising most of its surface area.

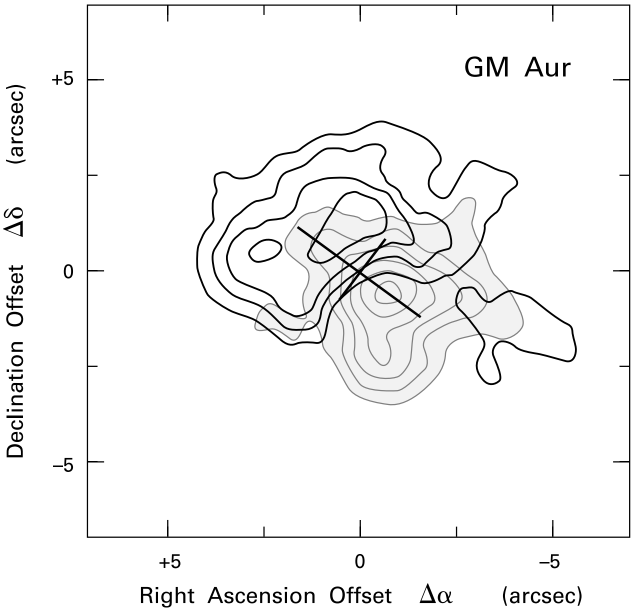


Figure 17.15 Contours of $^{12}C^{16}O\ (J = 2 \rightarrow 1)$ emission from GM Aur. Solid contours represent emission that is redshifted from the star by 0.5 km s^{-1}, while dashed contours, enclosing the shaded region, are for emission that is blueshifted by the same amount. The crosshairs indicate the disk's major and minor axes.

Let us focus, however, on the use of CO for direct imaging of the disks. As the need here is for very high *spatial* resolution, observers have employed interferometers operating in the millimeter regime. Figure 17.15 shows contours of the $^{12}C^{16}O\ (J = 2 \rightarrow 1)$ emission at 1.3 mm associated with GM Aur. The solid contours represent gas whose radial velocity is redshifted by 0.5 km s^{-1} with respect to the star, while the dashed ones are blushifted by the same amount. Evidently, most of the gas on one side of the star is moving away from us, and most on the opposite side is approaching. If we assume that the geometric aspect ratio of the lobes results from the tilt of an associated thin disk, then the inclination angle is 56°. Our earlier estimate of $\varpi_\circ$ should then be increased by $1/\cos\,\theta$, or about a factor of two.

The skeptical reader may well ask how we *know* that these red- and blueshifted contours represent a disk and not another molecular outflow. One reason is the modest extinction to the star itself ($A_V = 0.1$ mag), which would be unusually low for an outflow source. Far more persuasive, however, is the detailed pattern of the radial velocity. Suppose we draw, as in the figure, a straight line through the lobes' common major axis. At each position along this line, we may obtain the V_r-value corresponding to the maximum CO intensity. These numbers are positive in the redshifted lobe and negative in the blueshifted one. Away from the star, $|V_r|$ declines approximately as $\varpi^{-1/2}$, where ϖ is the distance. But this is just the falloff we expect for material in Keplerian rotation about a star of mass M_*. Here, we would have

$$V_r = \sqrt{\frac{G M_*}{\varpi}} \sin\theta \, , \tag{17.14}$$

where the last factor arises from projection onto the plane of the sky.

Since we previously inferred θ and also know $V_r(\varpi)$ from observations, we may use equation (17.14) to deduce the stellar mass. The answer is that $M_* = 0.8\ M_\odot$. On the other hand, we may also find M_* by placing the star in the HR diagram. That is, we may use the object's known values of L_* and $T_{\rm eff}$ to locate it on the appropriate pre-main-sequence track. Here again, the derived mass is close to $0.8\ M_\odot$. Similar agreement holds for most of the dozen or so T Tauri stars examined in this manner.

Returning once more to the case of GM Aur, we may also ask how the previously discussed disk truncation meshes with the CO data. If the putative inner hole is also present in the gas, it spans only about 1 AU, well below current spatial resolution. For comparison, the outermost CO lobes in Figure 17.15 have radii of some 500 AU. If we were also to show the detected *continuum* emission at 1.3 mm from heated dust, the corresponding radius would be about 200 AU. There is no reason to think that dust does not exist at greater distance, but the combination of falling surface density and temperature renders it invisible. On the other hand, the ^{12}C^{16}O transition used to probe the outer region is optically thick and so does not reliably trace the local column density. Our present uncertainty concerning the true sizes and gas distributions within disks further complicates the issue of their masses.

Having noted this basic limitation, we also point out that there are a few cases where an inner hole need not be surmised indirectly from broadband fluxes, but is directly observable. The objects in question are all *circumbinary* disks, *i. e.*, flattened configurations of gas and dust surrounding both members of a stellar pair. We previously described, in § 12.3, how such structures arise naturally during the cloud collapse that produces the stars themselves. Numerical simulations, like that displayed as Figure 12.21, find that the two protostars rapidly clear a gap in any gas residing within the orbital plane. The observations show that this gap survives past the epoch of cloud infall, well into the pre-main-sequence contraction phase.

Figure 17.16 displays the most striking example, the large disk associated with the binary GG Tau. We see here both contours and a greyscale image of 1.4 mm continuum radiation. The figure also indicates the positions of both stars, which are evidently at the center of an elliptically shaped, ring-like structure. Accounting for the finite inclination angle, this emission from warm dust extends from 180 to 260 AU, where distance is measured relative to the binary center of mass. For comparison, the stars themselves are separated by 44 AU. The reality of an interior gap is confirmed by near-infrared images. These show light from the central objects being scattered from an inner edge whose radius matches the figure derived from 1.4 mm data.

While the dimensions of the inner hole are vastly larger than in the GM Aur disk, the relative disposition of heated dust and cold gas is similar. Thus, observations here of ^{13}C^{16}O ($J = 2 \rightarrow 1$), also at 1.4 mm, show that molecular gas in the disk extends to at least 800 AU. The falloff in CO radial velocity from the binary again follows equation (17.14) to good accuracy. In this case, the inferred central mass of $1.28\ M_\odot$ represents the *sum* of the two binary components. One may also place both of these stars in the HR diagram and independently derive

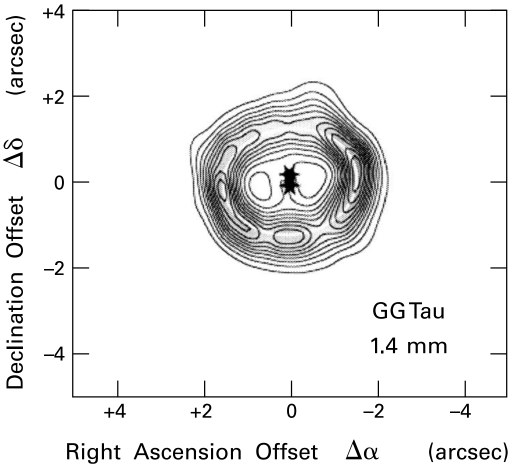


Figure 17.16 Map of GG Tau in 1.4 mm continuum radiation. The positions of the stars in the central binary are indicated. Angular offset is measured relative to the binary centroid. What appears as a narrow, interior gap, about 1.″5 from the stars, is actually a region of strong, but clumpy emission.

their individual masses to be 0.78 and $0.54\,M_\odot$. The sum is $1.32\,M_\odot$, in excellent agreement with the dynamical measurement.

A similar configuration of gas and dust is seen around UY Aur. This binary, discussed in Chapter 12, consists of an optically visible star and an infrared companion (see Figure 12.15). An evacuated ring of scattered light appears in the near infrared, as well as a much more extensive disk in CO. Since the visible object is a classical T Tauri star, complete with infrared excess, it probably has its own circumstellar disk. A similar structure, but oriented edge-on, could account for the high extinction toward the companion.

It appears, then, that young binary systems can simultaneously possess two types of disks – circumbinary and circumstellar. This hierarchy of diffuse structures is simply a reflection of orbital dynamics. Far from the binary's center of mass, fluid elements orbit in paths nearly indistinguishable from those in a central, point-mass potential. Close to either star, the paths again degenerate into ellipses and circles. It is in the intermediate region where the complex potential allows few, if any, stable orbits. One also sees, incidentally, how binary stars that are tightly bound might have severely truncated circumstellar disks. Empirically, the submillimeter continuum radiation from stellar pairs starts to decline once the binary separation becomes less than about 100 AU.

17.3.4 Disk Frequency

Let us return, however, to objects without such relatively close, disruptive companions and explore the basic issue of how often one finds associated disks. We first remark that all classical T Tauri stars exhibit *both* excess infrared flux *and* submillimeter continuum radiation. These two observational features probe the same circumstellar disk, albeit on different spatial scales. Any star with an infrared excess, which arises from dust orbiting inside roughly 10 AU, also has submillimeter emission, attesting to the presence of grains at a distance of order 100 AU. The converse situation, however, does *not* hold. That is, a star detectable in the submillimeter may have no infrared excess at all. The objects in question constitute a subset, perhaps 10 percent by number, of weak-lined T Tauri stars. Thus, they also lack the strong optical lines that signal circumstellar matter even closer to the central object. The extended disks they do possess must be relatively sparse. The highest 1.3 mm flux from a weak-lined star is an order of magnitude below the minimum seen in classical objects.

Since weak-lined and classical T Tauri stars overlap in age, these facts bring into even sharper focus an essential mystery. For reasons that remain obscure, a large portion of newly visible objects either have no disks, or are surrounded by material with very low surface density. Moreover, even the gross properties of the more substantial disks, including their outer radii, cannot yet be predicted through our present theory concerning the prior collapse phase and its aftermath. We saw in Chapter 11 that disks building up through rotating infall rapidly gain mass while they still have relatively small sizes. Although it is believed that steady, internal torquing can plausibly distend these structures both during and after infall, the idea has yet to be realized in an evolutionary calculation.

On the observational front, there is scant but growing evidence concerning the disks around protostars and their immediate descendents. Any optical or near-infrared disk signatures are naturally absent in embedded sources of Class 0 or I. Longer-wavelength emission generally lacks an unambiguous interpretation because of contamination by other matter. Thus, submillimeter radiation is indeed present in these sources, but mainly stems from much larger cloud cores, such as that shown in Figure 3.16. Additional, compact sources of submillimeter flux are known to be present in some cases. These plausibly represent circumstellar disks. However, no interferometric images are available, such as those toward visible T Tauri stars.

17.3.5 Surface Irradiation and Flaring

The infrared data from T Tauri disks suggests that our original, admittedly ad hoc assumption of a power-law temperature profile may actually hold in many cases. What underlying physics accounts for such behavior? We now show that a simple consideration of thermal balance for disk matter indeed yields a power-law profile. Other conceptual ingredients, however, are needed to explain the relative shallowness of the observed temperature gradients.

Let us picture, as in Figure 17.17, a disk patch being irradiated by a central star. The latter has a radius R_* and its center lies at the distance ϖ. We have singled out, through shading, a narrow annulus on the surface whose angular size, seen from the point of interest, is θ. As indicated by the shading, only half of this annulus irradiates the disk's upper surface. For a finite annular width $\Delta\theta$, corresponding to a small solid angle $\Delta\Omega$, we may write the incremental flux striking the patch as

$$\Delta F_{\rm rad} \,=\, f_\theta \,\Delta\Omega \int d\lambda \, B_\lambda \ . \tag{17.15}$$

Here, f_θ is a projection factor introduced to account for the grazing angle photons make with the patch normal. We have also integrated the specific intensity, assumed to be of blackbody form, over all wavelengths.

The integration of B_λ gives $\sigma_B \, T_{\rm eff}^4/\pi$, according to equations (2.34) and (2.35). Straightforward calculation also yields $\Delta\Omega \,=\, \pi \,\sin\,\theta\,\Delta\theta$. To derive f_θ, one must average the projection factor associated with each small element within the half-annulus. We forgo the necessary

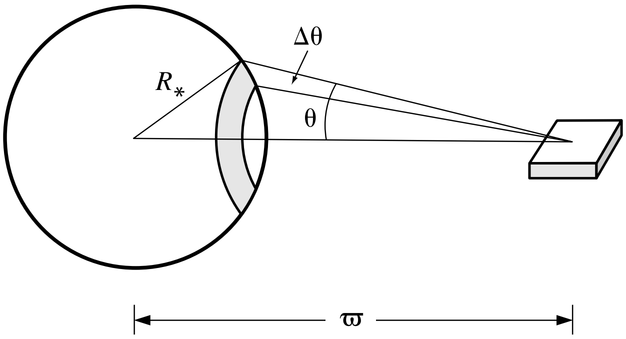


Figure 17.17 Irradiation of a disk patch by a spherical star.

geometric manipulations and simply quote the result:

$$f_\theta = \frac{2}{\pi} \sin\theta \; . \tag{17.16}$$

It follows that

$$\Delta F_{\rm rad} = \frac{2}{\pi} \sigma_B T_{\rm eff}^4 \sin^2\theta \, \Delta\theta \; . \tag{17.17}$$

We obtain the full flux by integrating $\Delta F_{\rm rad}$ over θ, where the latter ranges from 0 to $\theta_{\rm max} \equiv \sin^{-1}(R_*/\varpi)$. We then have

$$\begin{aligned} F_{\rm rad} &= \frac{2}{\pi} \sigma_B T_{\rm eff}^4 \int_0^{\theta_{\rm max}} \sin^2\theta \, d\theta \\ &= \frac{1}{\pi} \sigma_B T_{\rm eff}^4 \left[\sin^{-1}\left(\frac{R_*}{\varpi}\right) - \frac{R_*}{\varpi} \sqrt{1 - \left(\frac{R_*}{\varpi}\right)^2} \right] \\ &\approx \frac{2}{3\pi} \sigma_B T_{\rm eff}^4 \left(\frac{R_*}{\varpi}\right)^3 \; . \end{aligned} \tag{17.18}$$

In the last step, we have expanded in terms of the ratio R_*/ϖ, assumed to be small.

For the disk to maintain a steady temperature, the total heat input per unit area, given by $F_{\rm rad}$, must equal the corresponding efflux of radiant energy. Let us further assume that the disk is optically thick to its own cooling radiation. Then the outgoing flux from the top surface is

$$F_{\rm disk} = \sigma_B T^4 \; . \tag{17.19}$$

Equating $F_{\rm rad}$ and $F_{\rm disk}$, we find T as a function of ϖ:

$$T = \left(\frac{2}{3\pi}\right)^{1/4} \left(\frac{R_*}{\varpi}\right)^{3/4} T_{\rm eff} \; . \tag{17.20}$$

This expression only holds for ϖ well in excess of R_*, so that $T(\varpi)$ is much lower than $T_{\rm eff}$.

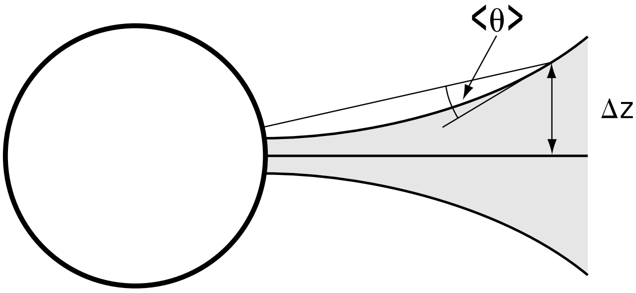


Figure 17.18 Irradiation of a flared disk. Depicted is a typical photon traveling from the stellar surface.

Equation (17.20) indeed states that the disk temperature declines with radius as a power law. The associated q of 0.75, however, is greater than we previously inferred for BP Tau (0.65) or estimated more crudely for GM Aur (0.50). In general, few T Tauri disks seem to have temperatures that fall as rapidly as our derivation predicts. To see the difficulty from another perspective, we may integrate $2\,F_{\rm disk}$ over all radii to find L_D, the total luminosity emanating from both disk surfaces. The answer, based on equation (17.20), is

$$L_D \,=\, \frac{2}{3\,\pi}\left(\frac{R_*}{\varpi_\circ}\right) L_* \ , \tag{17.21}$$

where L_* is the stellar luminosity, and where we have assumed $\varpi_D \gg \varpi_\circ$. Since the smallest possible value of $\varpi_\circ$ is R_*, equation (17.21) implies that L_D/L_* has an upper bound of about 0.2. The integrated spectral energy distributions of most classical T Tauri stars give a fractional disk luminosity higher than this limit.

Our derivation of the temperature assumed a razor-thin disk, *i. e.*, we neglected the scale height Δz compared to the typical radius ϖ. However, as illustrated schematically in Figure 17.18, any increase of Δz with ϖ allows the disk to intercept more photons from the star. The structure will thus be warmer than predicted earlier, and $T(\varpi)$ will decline more slowly. Such flaring of the surface is indeed expected and may be quantified through scaling arguments.

Let us suppose that the disk is vertically isothermal at each radius. Then, according to the dimensional analysis presented as equation (11.34), the scale height is related to the local, isothermal sound speed a_T by

$$\Delta z \,\approx\, \left(\frac{a_T}{V_{\rm Kep}}\right) \varpi \ .$$

Since $a_T \propto T^{1/2}$, and since the Keplerian orbital speed varies as $\varpi^{-1/2}$, we evidently have

$$\Delta z \,\propto\, T^{1/2}\,\varpi^{3/2} \ .$$

The grazing angle seen by the typical stellar photon, designated $\langle\theta\rangle$ in Figure 17.18, equals $\varpi\, d(\Delta z/\varpi)/d\varpi$, at least for $\varpi \gg R_*$. This angle therefore scales as $\Delta z/\varpi$, and we have

$$\langle\theta\rangle \,\propto\, T^{1/2}\,\varpi^{1/2} \ . \tag{17.22}$$

We complete the argument by repeating, in generalized form, our earlier derivation of $T(\varpi)$. The incoming flux $F_{\rm rad}$ is proportional, for small grazing angle, to $\langle\theta\rangle\,\varpi^{-2}$. On the other hand, the outgoing flux still obeys equation (17.19) and thus scales as T^4. It follows that the disk temperature now obeys

$$T \propto \langle\theta\rangle^{1/4}\,\varpi^{-1/2} \ . \tag{17.23}$$

Equations (17.22) and (17.23) together yield two interesting relationships:

$$\langle\theta\rangle \propto \varpi^{2/7} \tag{17.24a}$$

$$T \propto \varpi^{-3/7} \ . \tag{17.24b}$$

The first equation verifies that flaring occurs, since $\langle\theta\rangle$ increases with radial distance. Equation (17.24b) gives a new q-value of 0.43. Thus, flaring does produce a shallower falloff of temperature with distance.

It is also clear that our argument would need further development to reproduce observed spectral energy distributions. According to equation (17.9), any q-value less than 0.5 leads to $\lambda\,F_\lambda$ from the disk that *increases* with λ in the optically thick regime. Such a trend is rarely seen. A more careful derivation takes account of the fact that dust grains near the surface are actually warmer than those in the midplane. The former are exposed directly to stellar radiation, but cool rather inefficiently via infrared photons (for which Q_λ falls with λ; recall § 2.4). At sufficiently large radius, our implicit assumption that the bulk of the disk radiates as a blackbody also starts to break down. Finally, one must consider the effect on the observed flux of the disk's inclination with respect to the line of sight.

17.3.6 Magnetic Effects

These technical and physical issues are both nontrivial and have a significant impact on the final spectral energy distribution. Thus, a major question of principle is still unanswered. Can all the data be accommodated by models of *passive* disks, *i. e.*, those which merely reprocess impinging stellar radiation? Or does the non-photospheric contribution in some cases demand an *active* disk, *i. e.*, one generating part of its own luminosity?

The idea of active disks receives at least partial support from the observations of T Tauri emission lines. We have seen that many stars, including all those in the YY Orionis subclass, experience infall from circumstellar gas. To attain the observed speeds, this material need only start from a distance of a few stellar radii, much less than that associated with the infrared excess. Nonetheless, it is worth exploring whether some mechanism could both disrupt the interior portions of disks, allowing matter to fall onto the star, and also cause orbiting gas much farther out to migrate inward. The gravitational potential energy released during such migration could then supply some of the disk's luminosity.

Magnetic fields, which already must play a role in the observed winds, are also capable of disrupting the inner disk. Very close to a magnetized star, the field is strong enough to control the motion of *both* outgoing and incoming matter. Farther out, it is the matter that drags the field in the course of the fluid motion. The critical boundary is found by equating the local magnetic and kinetic energy densities. The calculation is straightforward, at least in principle, for a centrifugal wind with a known outflow rate $\dot{M}_w$ (§ 13.4). A disk is more problematic, since the material kinetic energy, and indeed the entire disk structure, depends on whatever torquing mechanism produces mass transport farther out. Nevertheless, there is certainly some region within which disk gas can no longer maintain planar orbits. If it is to join onto the star, this gas must flow along essentially rigid field lines.

What we are describing here is the magnetic dead zone, introduced in Chapter 13 during our discussion of wind theory. There we noted that any accretion onto the star occurs within the disk's corotation radius, *i. e.*, that point where the angular rotation speed $\Omega(\varpi)$ matches the stellar value Ω_*. If a steady state is reached, such that the magnetic field linking the disk to the star does not change over an orbital time scale, then both this field linkage and the inflow must occur precisely at corotation. To see why, let us denote this radius as ϖ_c. Suppose first that we have a disk where $\varpi_\circ < \varpi_c < \varpi_D$. Then all fluid elements between the inner edge $\varpi_\circ$ and ϖ_c would, in the absence of a magnetic field, be rotating faster than the star. If magnetically linked, this annulus drags forward the field lines. Conversely, the condition $\varpi_c < \varpi_\circ$ would mean that all points in the disk rotate more slowly than the star. Any stellar field lines penetrating the disk would be pulled back, and the field again wrapped up.

It is not evident, on general grounds, that a steady state *does* hold. The field connecting the disk to the star might indeed wrap up continually, only to periodically release energy through reconnection. (Recall Figure 13.19 and the accompanying discussion.) Even inflow that starts from the disk with a rotation speed of precisely Ω_* will spin up the star, unless the field has just the right azimuthal component.

Despite these caveats, some theorists argue that the star and disk must rapidly adjust themselves until a steady state is reached. That is, they *assume* that $\varpi_\circ = \varpi_c$ holds at all times, and that inflow to the star occurs only at that radius. The equality can, at least in principle, be checked observationally. We readily calculate ϖ_c to be

$$\begin{aligned} \varpi_c &= \left(\frac{G\,M_*}{\Omega_*^2}\right)^{1/3} \\ &= 12\ R_\odot \left(\frac{M_*}{0.5\ M_\odot}\right)^{1/3} \left(\frac{\Omega_*}{1\times 10^{-5}\ \mathrm{s}^{-1}}\right)^{-2/3} \quad . \end{aligned} \tag{17.25}$$

Here we have inserted, within our numerical evaluation, a representative Ω_*-value deduced from surface rotation speeds (§ 16.5). We see that ϖ_c is reasonable as a disk inner edge, in the sense that it comfortably exceeds the stellar radius. On the other hand, return to our example of GM Aur. Here the observed Ω_* is $1.3 \times 10^{-5}\ \mathrm{s}^{-1}$, while M_* is $0.8\ M_\odot$. The ϖ_c-value of $12\ R_\odot$ is much less than the 0.6 AU we inferred using the spectral energy distribution, after correcting for the disk inclination. Of course, any determination of ϖ_c from broadband fluxes is necessarily imprecise. But if our value for $\varpi_\circ$ is correct, then the stellar and disk speeds should *not* match. At a radius of 0.6 AU, both the disk temperature and its internal ionization are quite low. We thus expect any mechanical coupling between the field and the predominantly neutral matter to be weak.

This last example demonstrates that there are limits to the dynamical influence of the stellar magnetic field and therefore to the regime where inflow can occur. Let us focus, however, on a hypothetical disk that does become warm enough, and therefore sufficiently ionized, near its inner edge for coupling to be strong. That is, the field is effectively frozen to the matter. Then

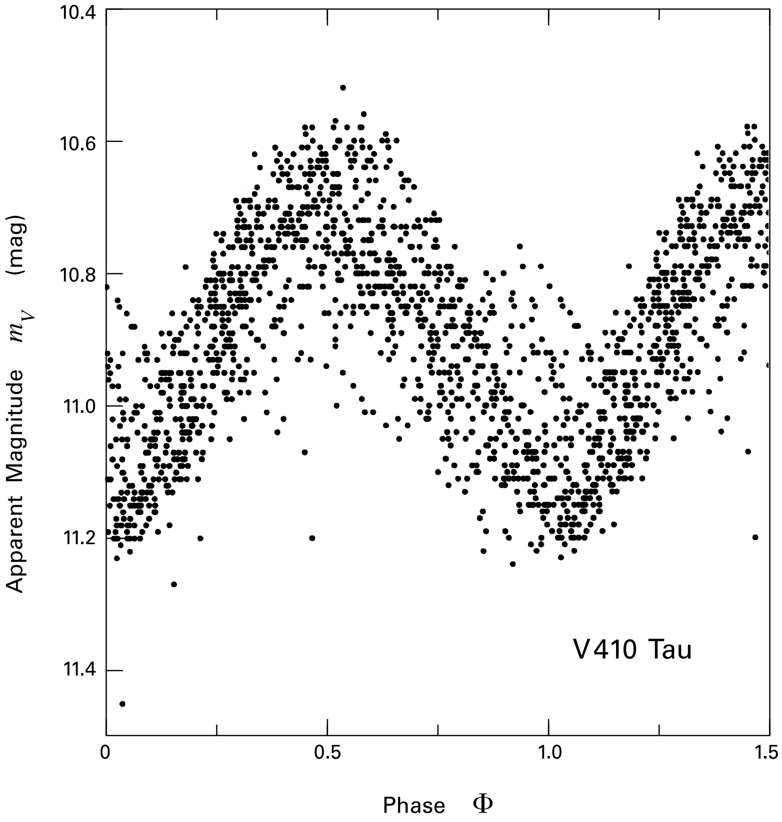


Figure 17.19 Phase-wrapped light curve in the V-band for the weak-lined T Tauri star V410 Tau.

there is another process in such a region that may be of importance for T Tauri stars. This is an *instability* associated with field lines penetrating any differentially rotating structure.

Consider the idealized situation of a straight magnetic field passing normally through a disk. Suppose further that each field line extends infinitely far in the vertical direction and that no external matter (or, more precisely, current) is present. The dynamical state of this disk is unperturbed by the field lines, which are simply transported around by the rotating fluid elements. Now imagine displacing an element radially outward, still within the disk plane. Then the attached field lines bend in such a way as to resist shear (Figure 10.12). That is, Ω remains nearly constant during the small excursion. However, the *background* Ω declines outward in all realistic disks, *i. e.*, $d\Omega/d\varpi < 0$. The centrifugal force per unit mass of the perturbed element, $\Omega^2\varpi$, is thus too great, and impels the element even farther out.

It is instructive to compare this *magnetorotational* instability with the purely *rotational* stability discussed in § 9.2. There we found a disk to be stable provided its specific angular momentum $j \equiv \Omega^2\varpi$ increases outward. Even if this criterion is met, however, the structure is *still* dynamically unstable if $\Omega(\varpi)$ declines. One need only invoke the presence of a weak magnetic field. Indeed, the disk is restabilized for too strong a field, *i. e.*, one for which the associated Alfvén speed V_A exceeds the local sound speed a_T. This latter criterion is eventually met as one approaches the star and the disk's inner edge.

Numerical calculations have explored the nonlinear evolution of the instability. These studies have largely focused on the dynamics within a single, isolated disk patch. Any small, initial perturbation in such a patch quickly grows and leads to large, time-varying disturbances of both the magnetic field structure and fluid orbits. Whether such turbulence can persist for many orbital periods throughout a large portion of the disk remains an open question. If turbulence does last, it could provide an effective viscosity between disk annuli that promotes mass transport onto the star. Any calculation purporting to demonstrate such a long-term, global effect must also address the origin and maintenance of the penetrating magnetic field. Does it arise wholly from the star, or is it partially regenerated by dynamo activity within the disk itself? Additionally, one would need to connect behavior in the turbulent region with that near the inner edge and farther out, in the cooler and dynamically stable exterior.

17.4 Temporal Variability

Historically, the recognition that T Tauri stars are intrinsically variable came early and indeed was one of their original, defining characteristics. In optical and ultraviolet light, these objects exhibit erratic fluctuations. The observed changes are usually relatively minor, but the amplitude in broadband flux can reach several magnitudes in some objects. The variables first identified by Joy in the 1940s were exclusively classical T Tauri stars. With the later addition of weak-lined objects, and with dedicated, photometric monitoring, other patterns of variability came to light. The ensemble of T Tauri stars exhibits fluctuations on every accessible time scale, from hours to decades. What further clues does this information provide about both the objects themselves and their environments?

17.4.1 Cool Spots

Breaking with the historical sequence of discoveries, we begin by examining weak-lined stars. Here the variability pattern is clearer, as is its physical interpretation. The changes in flux for any one object are generally smaller than in classical stars. For example, the fluctuation amplitude in the V-band is less than 1 mag. Another, and more striking, difference is that these changes tend to be *periodic*. The observed periods, typically measured in days or weeks, can be stable over observations spaced many years apart.

One possible cause for such alternate dimming and brightening of a star is pulsation of its surface layers. This type of oscillatory motion is well documented in the Sun and other dwarfs, and has been observed in a number of pre-main-sequence objects (Chapter 18). However, the associated periods are much shorter than the ones we are now considering. Pulsations also create changes in a star's radial velocity that are not detected. The observations of variability are better interpreted under the alternate assumption that the underlying period is that of the star's *rotation*.

Figure 17.19 is the optical light curve for the weak-lined star V410 Tau. In this plot, comprising several decades of observations, the independent variable is not time, but the elapsed fraction of a single period, *i. e.*, the rotational phase. The period itself was determined empirically by minimizing the scatter in the plotted points. The derived result in this case is 1.87 days. This figure is consistent with the independently known values of the stellar radius R_* and $V_{\rm eq}\,\sin\,i\,=\,77\,{\rm km\,s^{-1}}$, where the latter is determined by rotational broadening of absorption lines. Since $R_*\,=\,3.46\,R_\odot$, we deduce a value for the inclination angle i of $54°$.

In this and numerous other weak-lined stars, the variations seen in the V-band flux are essentially repeated when the star is observed through other filters, such as R and I. More specifically, the *period* of the fluctuations is identical, but the *amplitude* is not. The latter diminishes as one considers successively longer wavelengths. This falloff is gradual, suggesting that the modulations stem from rotation into the field of view of relatively cool surface regions.

To see the reasoning here, it helps to consider a simplified model. Suppose the star has photospheric temperature $T_{\rm star}$ and contains a single spot of surface area $\mathcal{A}_{\rm spot}$ and temperature $T_{\rm spot}$. Suppose further that rotation brings this feature to the precise center of the projected stellar surface, which has area $\mathcal{A}_{\rm star}$. At this phase, the total received flux is

$$F_1\,=\,c\,\left[\mathcal{A}_{\rm spot}\,B_\lambda(T_{\rm spot})\,+\,(\mathcal{A}_{\rm star}-\mathcal{A}_{\rm spot})\;B_\lambda(T_{\rm star})\right]\ , \tag{17.26}$$

where the coefficient c includes the stellar distance. Here we are idealizing both the unperturbed photosphere and the spot as blackbodies. Half a rotation period later, the spot has disappeared from view. The flux is then

$$F_2\,=\,c\,\mathcal{A}_{\rm star}\,B_\lambda(T_{\rm star})\ . \tag{17.27}$$

The ratio of these two fluxes, expressed in magnitudes, is the full range of variation seen in the light curve. This range is

$$\begin{aligned}\Delta m_\lambda\,&=\,-2.5\,\log\,(F_1/F_2)\\&=\,-2.5\,\log\left\{1\,-\,f_B\left[1-\frac{B_\lambda(T_{\rm spot})}{B_\lambda(T_{\rm star})}\right]\right\}\ .\end{aligned} \tag{17.28}$$

Here, the term f_B is the fractional coverage of the star's projected surface:

$$f_B\,\equiv\,\mathcal{A}_{\rm spot}/\mathcal{A}_{\rm star}\ . \tag{17.29}$$

Figure 17.20 plots Δm_λ as a function of λ for the case of $T_{\rm star}\,=\,4000$ K and $f_B\,=\,0.2$. For a cool spot, $T_{\rm spot}\,=\,3000$ K, Δm_λ indeed falls gently with λ (*solid curve*). The reader may verify, using equation (2.29) for $B_\lambda(T)$, that the argument of the logarithm in equation (17.28) *increases* from its zero-wavelength value of $1-f_B$ toward unity, and therefore that Δm_λ itself *decreases*. Qualitatively, this is the behavior seen in weak-lined stars. One may utilize, for any particular object, the actual variation of Δm_λ to estimate both the depression in temperature and the fraction of the stellar surface covered by spots. The temperature deficit is generally about 1000 K, while the areal coverage can be significantly greater than the 20 percent we have used. Even this last figure is enormous compared with analogous spotting on the Sun.

Thus far, we have discussed only results obtained through photometry. Multiple observations of spectral lines, when available, yield considerably more information. Through the

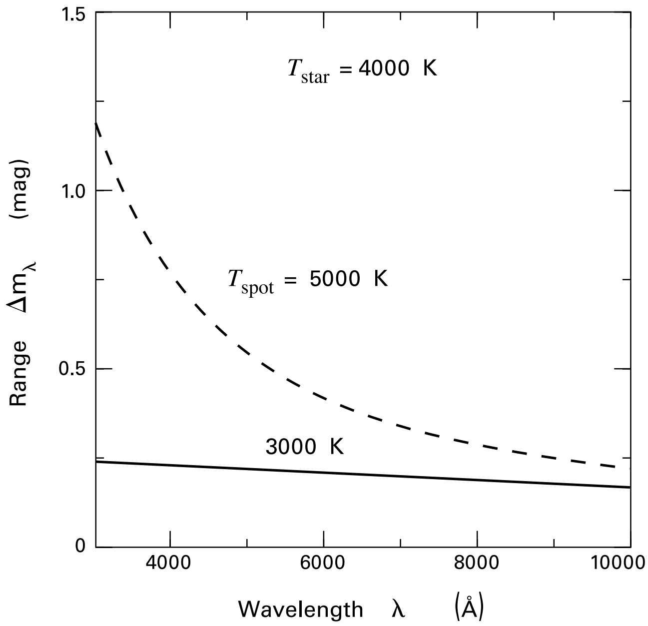


Figure 17.20 Theoretical range of variation in the flux from a star with a cool spot (*solid curve*) and a hot spot (*dashed curve*). The range is expressed in magnitudes and displayed as a function of wavelength.

technique known as *Doppler imaging*, one may reconstruct in detail the distribution of $T_{\rm eff}$ over the rotating stellar surface. One begins with a model stellar surface divided into many small zones. In each patch, an assumed $T_{\rm eff}$-value, together with the known surface gravity, allows one to solve the radiative transfer equation for the profile of any line of interest. Note that the result depends also on the inclination of the patch to the line of sight and hence on the rotational phase. Adding up all the patches gives the total, predicted profile from the star at one time. Finally, one varies the $T_{\rm eff}$-distribution to obtain the closest match to the observed temporal sequence of profiles.

For this technique to be feasible, the rotational broadening of the line profiles must significantly exceed the thermal contribution. Only a handful of T Tauri stars meet this requirement, including V410 Tau. Figure 17.21 shows graphically the results of Doppler imaging in this case. Here, temporal sequences of profiles from six absorption lines near 6400 Å were utilized. In these four images, equally spaced in rotational phase, the darkened patches represent locations with $T_{\rm eff}$-values more than 500 K below the unperturbed, photospheric value. We see that starspot activity concentrates heavily at high latitudes, *i. e.*, close to the stellar pole. Comparison with earlier observations indicates that the temperature distribution persisted for at least seven years.

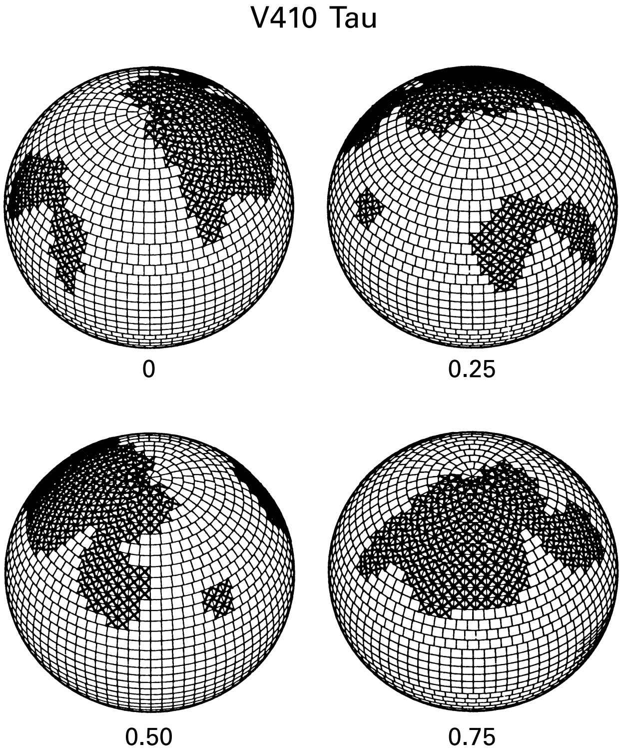


Figure 17.21 Doppler images of V410 Tau, shown at the indicated phases of rotation. The dark patches represent regions cooler than the surrounding photosphere by at least 500 K.

17.4.2 T Tauri Magnetic Fields

The presence of cool regions suggests that, as in the Sun, bundles of concentrated magnetic field are penetrating the stellar surface. Such fields locally inhibit the process of convection, which transports thermal energy from below. This is not our first indication, of course, that T Tauri stars are magnetized. We noted early in this chapter that the roster of optical emission lines is similar to that in active regions of the Sun. Continuum radiation in the X-ray and radio further indicates not only the presence of a static field, but also attests to violent recombination within migrating flux tubes. The reasoning that links such observations to magnetic fields is rather indirect, and is guided by analogy with the solar case. More convincing evidence, along with an assessment of the surface field strength, would be provided by Zeeman splitting of spectral lines. Let us digress momentarily to consider the results of such investigations.

Recall the essential logic of the technique (§ 6.3). Unpaired electrons within atoms give rise to circularly polarized radiation. In the presence of a magnetic field, two lines of differing polarization are shifted in opposite directions. The total observed separation of each such line pair yields the magnitude of the longitudinal field $B_{\|}$, while the sense of the shift (*i. e.*, which polarization component increases in wavelength) gives the field direction. As summarized in

Table 9.1, the method has proved successful in a number of environments and would seem especially promising in T Tauri stars, where the fields should be the highest of all. Polarimetric observations, however, find no detectable shift in optical absorption lines. The reason, once again, must be that the surface fields of T Tauri stars are geometrically complex. Adjacent flux tubes leaving and entering the photosphere combine to give zero net shift for any individual polarized line.

On the other hand, consider observations of a spectral line taken *without* an intervening polarimeter. The intensity one detects in this case represents the sum of polarized pairs of lines. Since a magnetic field causes each component within a pair to shift in wavelength, the total line profile broadens. The important point is that the amount of broadening does not depend on the direction of the field, but only on its magnitude. A potential difficulty, however, is that other physical effects, such as stellar rotation, also broaden lines and could be confused.

The trick is to observe a large number of lines with differing magnetic sensitivity (*e. g.*, different values of the Landé g-factor; recall § 6.3). Any systematic rise in the line equivalent width with increasing sensitivity reveals the presence of a field and also yields its magnitude. Especially useful here are lines of longer wavelength, including the near-infrared. From equation (6.33), the Zeeman splitting in *frequency*, $\Delta\nu$, is equal to $B_{\|}$ times a coefficient that is independent of the line frequency ν. Thus, the shift in *wavelength* $\Delta\lambda = -\lambda^2\Delta\nu/c$, is larger for transitions with greater λ.

Observations of Zeeman broadening have so far provided field strengths for a few weak-lined and classical T Tauri stars, including the primary of T Tauri itself. Measured values are all of order 10^3 G. This figure actually represents $f_B B_{\|}$ the *product* of the mean field strength and the fractional coverage of that field on the stellar surface. As we have indicated, f_B is at least 20 percent in weak-lined stars, with the fraction being even higher in many cases. It is interesting that the field strength is the same order of magnitude in the Sun, whose flux tubes cover only a few percent of the surface. The $B_{\|}$-values are also comparable in younger main-sequence stars, where the areal coverage is intermediate between weak-lined objects and the Sun (§ 16.5).

17.4.3 Hot Spots

Returning to the main topic of this section, let us now examine how the variability of classical T Tauri stars compares with their weak-lined counterparts. One general difference is that the amplitude of fluctuations tends to be greater, especially as one considers shorter wavelengths. Despite the historical identification of these objects as irregular variables, a sizable fraction do change over well-defined intervals, where the latter range from several days to a week or more. Even in such cases, however, the modulations display this character only for a limited duration, usually no longer than a few months. A new period, differing as much as 20 percent from the old, may then be established, after an epoch of seemingly random fluctuations. Analysis of many stars reveals no period at all, even over relatively brief epochs.

Consider now a classical T Tauri star that does display, at least temporarily, regular variations in flux. As in weak-lined objects, the same period emerges from observations over a range of wavelengths. The range Δm_λ once again declines as λ increases. This time, however, the falloff is much steeper. Such a trend is best explained by the rotation of zones that are *hotter* than the surrounding photosphere.

Our previous, simplified model again helps elucidate the situation. The dashed curve in Figure 17.20 shows Δm_λ as a function of λ when the putative single spot has a temperature of 5000 K, *i. e.*, 1000 K above the stellar photosphere. Since F_1 is now greater than F_2 in equation (17.28), Δm_λ is formally negative; we have nonetheless plotted the range as a positive quantity in the figure. The curve's steep slope is readily understood if we consider the regime of short wavelengths. In this limit, we have

$$\frac{B_\lambda(T_{\rm spot})}{B_\lambda(T_{\rm star})} \approx \exp\left[\frac{h\,c}{\lambda\,k_B}\left(\frac{1}{T_{\rm star}} - \frac{1}{T_{\rm spot}}\right)\right] \; .$$

The exponential, which becomes vanishingly small in the case of a cool spot, now diverges and dominates the argument of the logarithm in equation (17.28). We thus find

$$\Delta m_\lambda \approx -2.5\,(\log e)\,\frac{h\,c}{\lambda\,k_B}\left(\frac{1}{T_{\rm star}} - \frac{1}{T_{\rm spot}}\right) \; . \tag{17.30}$$

We now see both why U-band observations of classical T Tauri stars tend to exhibit large-amplitude fluctuations, and also why this amplitude dies away rapidly in the visible and near-infrared.[4]

The assumption that variability is again tied to spotting and rotation immediately raises several questions. Under what circumstances do hot, as opposed to cool, spots arise on the surface? How can a star's apparent period of rotation shift over the course of weeks or months? Finally, why do many classical T Tauri stars exhibit no regularity at all in their flux variations?

To address the first issue, consider the actual values of spot temperature obtained by more careful modeling of Δm_λ. The contrast with the unperturbed photosphere is now greater than for weak-lined objects; a typical $T_{\rm spot}$ is 7000 K. Conversely, the fraction of the total surface area covered by spots is far less, of order 1 percent. Both figures are intriguingly similar to those obtained in slab models of continuum veiling (§ 17.1). This fact, in turn, suggests a common origin for the two phenomena. Hot spots plausibly represent areas where external matter falls onto the stellar surface, creating elevated temperatures through shock heating.

As we have already stressed, the precise trajectory of this matter, and indeed its very origin, are still murky. Gas flowing along magnetic loops remains an attractive picture. The detailed modeling of photometric changes finds that a combination of hot and cool spots may best fit the data. In these stars, some loops apparently contain flowing gas while others do not. Both types of magnetic structures should partake in the local surface rotation, as they do on the Sun. Pursuing the analogy, we might then attribute a shift in periods to migration of the largest spots with latitude. Here, we further assume that the surface, like that of the Sun, is in a state of differential rotation.

---

[4] Our model also predicts that the photosphere should appear normal when the spot is hidden from view, *i. e.*, at minimum light. Indeed, the photometric colors of a classical star are most consistent with the object's $T_{\rm eff}$-value during this phase. Colors are anomalously warm at maximum light.

17.4.4 Aperiodic Changes

Especially puzzling are the seemingly random fluctuations. The arrangement of magnetic loops, or at least their associated mass flow, presumably changes rapidly during such epochs. Alterations must occur on a time scale comparable to or less than the star's rotation period. To repeat a familiar theme, the shocks here may arise not only from surface infall, but also from the interaction of outflowing and circumstellar gas. In any case, there is little understanding of why this more chaotic situation appears to dominate in many objects.

Return now to hot spots that do arise from localized infall. Surface rotation should not only move these features in and out of view, but should also lead to concurrent shifts in the measured level of veiling. It is encouraging that the continuum excess (as measured by r_λ) indeed fluctuates in a number of objects where it has been carefully tracked, and that these shifts are sometimes periodic. For other stars, the veiling changes erratically. Obtaining simultaneous photometry and high-resolution spectroscopy is difficult, so it is generally not possible to say in these latter cases that the two changes are actually in phase. Exceptions are the extreme objects such as DR Tau. Here the photospheres are dominated by an especially strong continuum flux. Hence it is certain that any brightness modulations are coincident with those in veiling, even when both are aperiodic.

The strong and broad emission lines of classical T Tauri stars also vary with time. Previously, we mentioned the unstable nature of Hα profiles. We shall not explore this topic in any detail. Rather, we simply note that the profile changes do *not* exhibit the periodicity sometimes found in the photometric data. Although it is thus difficult to link these two kinds of variations quantitatively in any one object, an interesting pattern does exist across the ensemble of classical stars. Suppose one measures the *range* of variation in the Hα equivalent width within a group of such stars. This quantity correlates well with the photometric range in any fixed waveband. In other words, stars that fluctuate greatly in their Hα-production also do so photometrically.

All the changes we have described thus far occur on time scales ranging from days to months. T Tauri stars also display interesting variations over both shorter and longer intervals. Documenting such activity naturally presents challenges, but there is nonetheless a significant body of data. In the regime of short times, weak-lined stars generally exhibit more erratic behavior in the U- and B-bands than at longer wavelengths. These fluctuations, superposed on the usual periodic light curve, stem from flaring activity. The actual rise in brightness, fortuitously seen in a few objects including V410 Tau, happens within a matter of hours, while the decay is more protracted.

Such events remind one, at least superficially, of those occurring in the X-ray domain (§ 17.1). The detailed character of the flares, including both their prominence at shorter wavelengths and temporal behavior, resembles that seen in the low-mass dwarfs known as dMe stars. Here, the activity stems from energy release above magnetized surface regions. Corroborating this interpretation for weak-lined stars is the fact that photometric flares appear most often near minimum light, *i. e.*, when the cool spots are most visible. Similar episodes must occur in classical T Tauri stars, at least those with an admixture of cool spots. Observationally, any rapid brightness changes in these objects tend to be more symmetric in time. There are also occasional surges in the strengths of emission lines. The relation of these spectroscopic occurrences to the photometric ones is unclear. A quick rise followed by an equally rapid decay may signify change of the infall rate onto a bright hot spot, rather than a coronal event.

17.4.5 FU Orionis Outbursts

We end this section by describing temporal changes of an entirely different character. These occur over decades, by far the longest intervals we have considered. The shorter-term phenomena described earlier do exhibit secular drifts over such epochs. Weak-lined stars, for example, may gradually alter both the amplitude and temporal form of their variation within each cycle. We are interested, however, in a much more drastic alteration of the star's appearance, one in which the luminosity rises by two orders of magnitude in the course of a year or more, and then fades again relatively slowly. Although few of these *FU Orionis outbursts* have actually been seen, the phenomenon represents an important aspect of the pre-main-sequence phase.

The first object discovered, FU Ori itself, was, for at least a decade, an inconspicuous source within the dark cloud B35. This cloud lies just beyond the spherical rim of the extensive λ Ori HII region. In 1936, FU Ori was caught brightening. Within a few months, the luminosity peaked and has been falling smoothly ever since. Figure 17.22 displays the B-band light curve, both for FU Ori and for the other two well-documented cases. Notice that m_B increased between 4 and 6 mag, with the change being relatively slow in V1515 Cyg. All three sources are located within star-forming regions and have strong lithium absorption. These facts, together with the objects' known irregular variability prior to outburst, indicate that they are low-mass, pre-main-sequence stars. Only V1057 Cyg was observed spectroscopically, albeit at low resolution, during its earlier epoch. The spectrum exhibits standard T Tauri emission lines.

Other events uncovered since the FU Ori flareup may also represent the same phenomenon. For example, comparison with archival images shows that a few embedded objects brightened sometime in the past. We might also count as "FUors" all stars currently exhibiting spectroscopic properties similar to the original three in their post-outburst state (see below). Even this expanded list comprises no more than about a dozen sources, and the effect we are describing would seem to be of little significance. Consider, however, the *rate of appearance* of new outbursts. Over the 60 yr time span of observations, this rate (based on the expanded list) is 0.2 yr^{-1}. Suppose that sources can be detected out to $d = 1\,\mathrm{kpc}$. The local stellar birthrate in the Galaxy, according to Chapter 1, is $\dot{m}_* = 5 \times 10^{-9}\,M_\odot\,\mathrm{yr}^{-1}\,\mathrm{pc}^{-2}$. Within a cylinder of radius d, therefore, new stars appear at the rate $\pi\,\dot{m}_*\,d^2$. This figure is 0.02 yr^{-1} for the quoted numbers. On average, therefore, every star undergoes an outburst ten times during its pre-main-sequence lifetime. The number is higher if only a fraction of young stars is involved. In summary, outbursts seem to be a *recurring feature* of early stellar life.

From our examination thus far of pre-main-sequence evolution, there is no apparent reason why stars should undergo large surges in brightness. In fact, T Tauri stars are in the convective phase of quasi-static contraction, for which theory predicts a steady *decrease* in luminosity. We shall later describe how objects of intermediate mass do brighten near the start of their contraction phase. However, the change is a permanent one, rather than a periodic flareup, and the relevant fraction of young stars is small in any case. If FU Orionis outbursts are as common as our statistical argument indicates, then either the contraction theory for low-mass

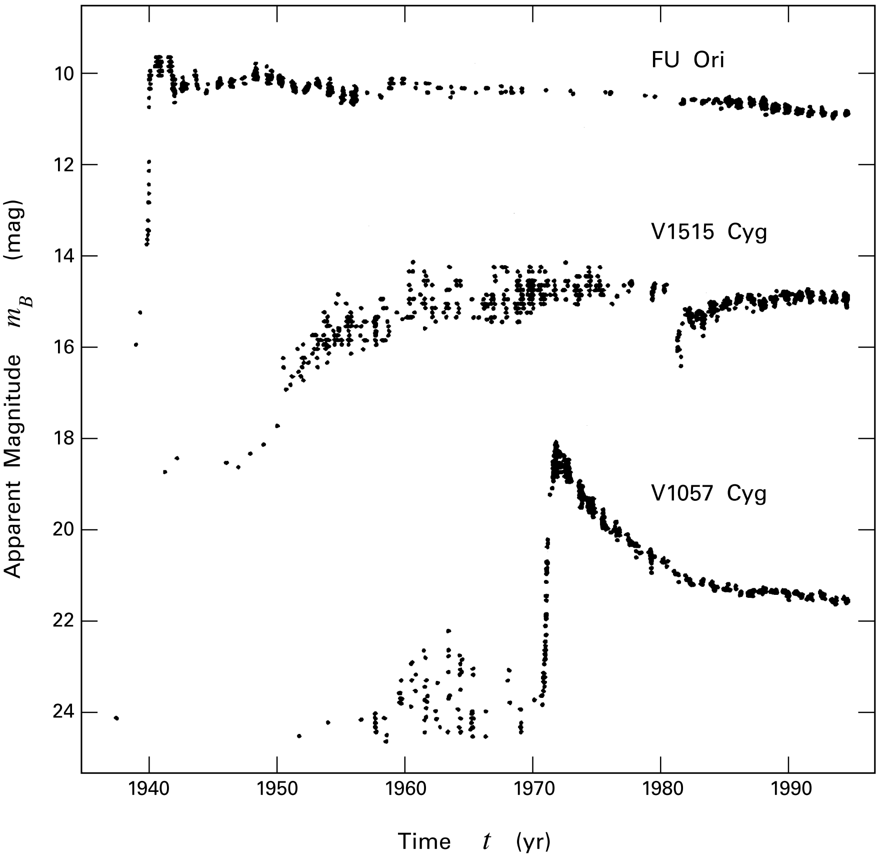


Figure 17.22 Historical light curves for the three visually erupting FU Orionis stars. The vertical scale of V1515 Cyg has been displaced downward by 1 mag, that of V1057 Cyg by 8 mag.

stars needs refinement or the effect originates externally, through some interaction of the star and its immediate environment.

Could each flareup occur simply from the removal or displacement of opaque material blocking the visible starlight? The answer is no, for several reasons. First, the two more recent, well-documented examples were not infrared objects prior to outburst. Second, the star's *effective temperature* apparently changes during the event. In the case of V1057 Cyg, both the emission lines mentioned previously and the broadband fluxes in the pre-outburst spectrum are consistent with those of a T Tauri star. Just afterward, the object lacked these emission lines and had colors appropriate to an A-type star. Similarly, FU Ori first appeared in its luminous state as a G star. Such dramatic shifts in temperature reflect either a change in the underlying stellar photosphere or else heating of circumstellar material.

Under the first hypothesis, the object becomes much too luminous for its spectral type, even on the birthline, and therefore acquires a large radius. Analysis of absorption lines yields, in fact, a very low surface gravity in all three cases. Both the swelling and attendant spectral features change as the star fades. The apparent spectral type of V1057 Cyg went from A to F within 2 yr after the peak luminosity. In this star and the others, Hα exhibits a P Cygni profile, with a deep, blueshifted absorption dip. That is, the present-day object emits a strong wind.

Other features of the spectra are more problematic. The optical absorption lines are quite broad, ostensibly signifying high rotational velocity. This property must be reconciled with the enlarged radius. Consider again V1057 Cyg, for which the line broadening yields $V_{\rm eq}\,\sin\,i \approx 50\,{\rm km\,s^{-1}}$. From the estimated surface gravity, the immediate post-outburst radius was about $12\,R_\odot$. Here we have assumed a mass of $0.5\,M_\odot$, appropriate for a T Tauri star. The pre-outburst radius for this same object would have been at most $3\,R_\odot$ (Table 16.1). If angular momentum were conserved during the swelling, the initial rotation speed would exceed the breakup value of 180 km $\rm s^{-1}$.

Adding to the mystery is the curious fact that the line broadening itself changes in different wavelength intervals. The trend is for the apparent rotational speed to decrease as λ increases. Thus, profiles of near-infrared lines on V1057 Cyg yield a $V_{\rm eq}\,\sin\,i$ of only about 30 km $\rm s^{-1}$. The pattern of absorption lines in this regime, and especially the deep CO feature at 2.2 μm, is characteristic of a late-type giant atmosphere, again in disagreement with the inferences from optical data.

The spectral energy distributions for the three sources in Figure 17.22 all display large infrared excesses. It is therefore plausible that each object has its own circumstellar disk. Could a very hot and luminous disk, rather than the star itself, be the source of the outburst? Some of the observational problems would then be neatly resolved. Consider the combination of low surface gravity and high rotational speed. The vertical gravitational acceleration in an extended disk is much smaller than at the stellar surface (Figure 11.10). Absorption lines from an annulus rotating at Keplerian speed would also be significantly broadened. One thus avoids the troubling notion of a rapidly spinning giant star. Within a disk, near-infrared lines must originate in relatively cool material, while optical features arise in hotter gas. Thus, if the disk temperature falls with distance from the star, one would naturally expect the line broadening to diminish at longer wavelengths. The deduced spectral type would also be later in this regime.

While the disk picture offers insight concerning puzzling aspects of FU Orionis spectra, it is less helpful on the most fundamental issue of all, the origin of the outbursts themselves. A disk can store very little thermal energy compared to a star. Thus, it can only generate luminosity through orbital decay, *i. e.*, through accretion onto the central object. Rapid mass transfer could in principle be stimulated by close passage of a companion star. Unfortunately, the known FUors are not spectroscopic binaries. We argued in Chapter 11 that continued infall onto a disk also leads to mass transfer via gravitational torquing. Here, the infall arises from the collapse of the parent molecular cloud. None of the original three stars, however, is situated within a dense core. Finally, many theorists have investigated disk instabilities and stellar accretion that might arise spontaneously. The models rely on a specified internal viscosity to carry matter onto the star. Since the origin of the putative viscosity is unknown, it must be admitted that repetitive outbursts in disks are no more comprehensible than those originating within stars.

We have noted that the well-documented FUors show evidence, through their Hα profiles, for strong winds. The presence of this signature in a few other stars has led to their classification as post-outburst objects. Whether or not this identification is correct, it is tempting to speculate

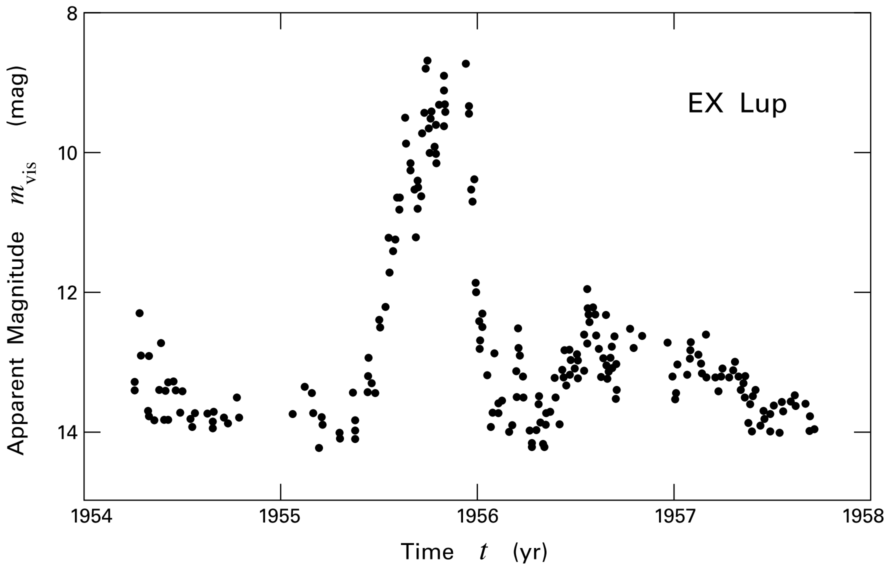


Figure 17.23 Eruption of EX Lup during the 1950s. Note that the apparent magnitudes are visual estimates.

that each flareup is accompanied by enhanced ejection of gas from the star or disk. Such a wind burst would create a Herbig-Haro shock in nearby cloud matter. This does not mean, of course, that all the observed Herbig-Haro objects are necessarily driven by such impulsive events. Figure 13.20 demonstrates how traveling shocks arise also from *continuous* variation of an underlying wind. More generally, there is still no evidence that the deeply embedded objects driving jets are, as a group, subject to the flareups seen primarily in optically visible stars.

A small fraction of T Tauri stars have fluctuations that, in several respects, are intermediate between FU Orionis events and the small-scale irregularities of the main population. Members of this subclass are sometimes called "EXors." The name derives both from the analogous FUors and from the prototypical example, EX Lupi. This object exhibits outbursts of several magnitudes every month or so. Larger increases occur about once a decade, with the bright phase lasting roughly a year. An especially strong flareup, rivaling FU Orionis outbursts in magnitude, is shown as Figure 17.23.

Even at maximum light, the spectrum of EX Lup continues to display the same T Tauri emission lines as before. Moreover, the strongest ones have clear *inverse* P Cygni profiles. The previously existing absorption spectrum becomes filled in by continuum veiling in the course of brightening. As in YY Orionis stars (§ 17.2), these two observations indicate that each event stems from episodic infall, presumably from material not far beyond the stellar surface. The differences with FU Orionis observations are noteworthy. In the latter, the spectrum changes so radically that the original, strong emission lines are no longer present. Nor is there any ostensible infall signature, either through inverse P Cygni line profiles or elevated continuum flux. A viable theory of the larger outbursts must naturally accommodate these facts.

17.5 Post-T Tauri Stars and Beyond

All of the characteristics peculiar to T Tauri stars naturally disappear in the course of pre-main-sequence contraction. Some of these properties, we have seen, are created by shocks associated with inflow and winds, others arise from orbiting circumstellar disks, while still others, such as X-ray emission, are related to the presence of strong magnetic fields. It is by no means obvious that the *rate* of decline is the same for all properties, even at fixed stellar mass. Indeed, empirical documentation of the various rates might help elucidate the T Tauri environment.

17.5.1 Fading of T Tauri Properties

A glance back at Figure 17.3 shows us immediately the essential complication in this enterprise. While it is clear from the figure that the Hα surface flux generally decays with time, there is considerable variation among subsamples of essentially identical age. This variation itself gradually diminishes, but does not entirely vanish even as the main sequence is reached. The only feasible way to quantify evolutionary trends is to compare stellar groups with disparate ages. Any result obtained is an average, both with respect to the star-to-star variation and over some finite spread in mass.

Figure 17.24 uses this approach with the near-infrared emission. Here we plot, within each indicated group, the fraction of stars having excess emission, as judged through near-infrared color-color diagrams. (Figure 4.2 showed such a plot using the $J-H$ and $H-K$ color indices.) The "age" of each system in Figure 17.24 is a mean of the individual values determined from the HR diagram. We see that the fraction of stars with excess flux smoothly falls in time and disappears entirely by 6×10^6 yr. This last figure depends heavily on observations of a single cluster and should not be given too much weight. In any case, the emission in question originates in dust grains heated to roughly 10^3 K. Thus the falloff seen here reflects the sublimation or orbital decay of grains located within a few stellar radii.

The bulk of circumstellar matter, including any extended disk, is too cold to be seen at near-infrared wavelengths. As described in § 17.3, cooler dust at greater distance from the star still emits in the submillimeter continuum. We recall that every classical T Tauri star radiates in this regime, while most weak-lined sources are undetectable. Observations thus far do not show any clear evolutionary trend, beyond the fact that the emission vanishes by a stellar age of roughly 10^7 yr. Gaseous tracers of disks, such as the ^{12}C^{16}O line emission shown in Figure 17.15, give a similar result.

T Tauri stars also display activity related to surface magnetic fields, convection, and winds. While some degree of activity remains in low-mass dwarfs, the erratic optical flaring characteristic of younger stars eventually disappears in all but a small subset. To record the decline, one must first identify the flaring objects in any given group. Repeated spectroscopic or photometric observations are necessary for this purpose. Such a study of Taurus-Auriga reveals that 80 percent of weak-lined stars undergo flaring, compared to 56 percent of classical objects. The figure drops to just 16 percent for zero-age main-sequence stars in the α Persei cluster, with an age of 7×10^7 yr.

We also discussed how yet another manifestation of surface activity, X-ray emission, has both flaring and steady contributions. Satellite observations have given us the distribution of the

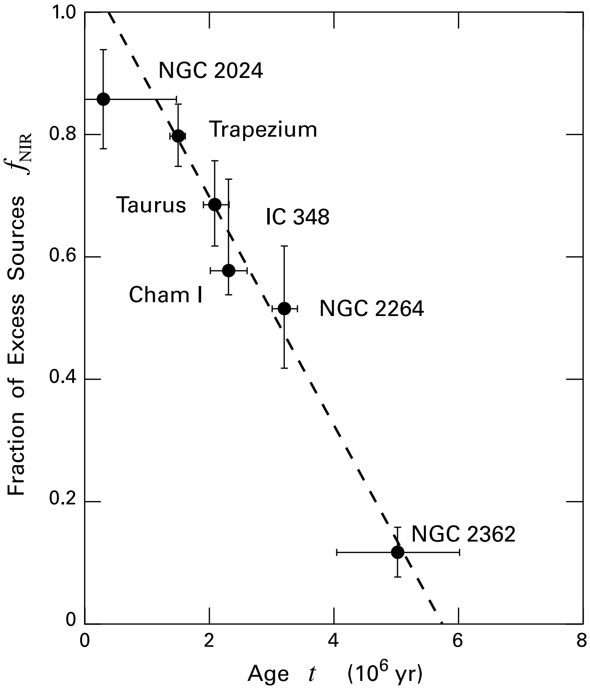


Figure 17.24 Decline of near-infrared excess with time. The quantity $f_{\rm NIR}$ is the fraction of sources displaying an excess, as gauged through color-color plots employing J, H, K, and L-band fluxes. Cluster ages are average values within each group. Error bars indicate the statistical uncertainty arising from the finite sample sizes. The straight line is a least-squares fit.

steady X-ray luminosity, L_X, within populous groups. Figure 17.25 shows representative findings, displayed here as the *cumulative luminosity function*, Φ_X. This is the fraction of sources within a group that have L_X exceeding any given value. The figure shows results for K stars in Taurus-Auriga, the Pleiades, and the Hyades. The first sample is missing numerous, relatively faint sources; their inclusion may significantly alter the histogram. With this caveat, there is evidently a steady fall in the mean L_X, although the upper limit does not change appreciably from the Pleiades age (1×10^8 yr) to that of the Hyades (6×10^8 yr).

All these observations indicate that alterations occurring in the course of pre-main-sequence contraction and beyond are continuous. At what point, then, does a given low-mass object cease being a T Tauri star? Any precise demarcation is clearly a matter of convention. However, the question is meaningful because some phenomena die away before the onset of hydrogen fusion. Thus, there is a realm of *post*-T Tauri stars.

One might be tempted to define such objects as those lacking any signature of disks, either at infrared or submillimeter wavelengths. Such a definition, however, would encompass most weak-lined T Tauri stars, including those close to the birthline. A more relevant property is the X-ray luminosity, which is high in both classical and weak-lined objects. Unfortunately, there is a wide spread in L_X even at fixed mass and age, partially due to the influence of stellar

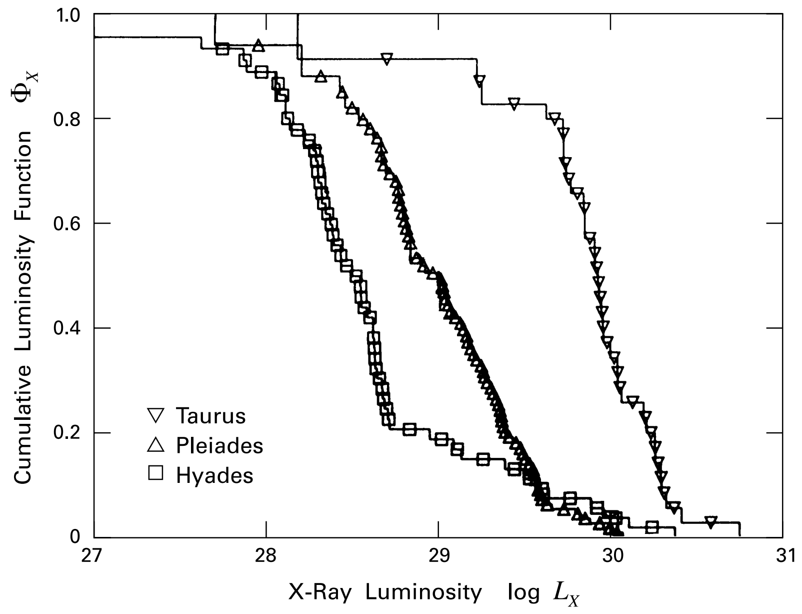


Figure 17.25 Cumulative X-ray luminosity functions in Taurus, the Pleiades, and the Hyades.

rotation. Yet another choice is the surface lithium abundance. This is always enhanced above the interstellar value in sufficiently young objects. While the abundance still varies within stars of identical L_* and T_{eff}, it declines substantially before the zero-age main-sequence.

Figure 17.26 displays, as open circles, the observed equivalent widths in Li I $\lambda 6708$ for a number of Taurus-Auriga members. Here, the data are shown as a function of the stellar effective temperature. It is apparent that nearly all points fall above the locus corresponding to 50 percent depletion (*dashed curve*). The filled circles in the figure represent measurements in the Pleiades.[5]

Conspicuous in this plot is the widening gap between the dashed curve and the upper envelope to the Pleiades, shown here as another solid line. We may conveniently define post-T Tauri stars as being objects that fall within this wedge-shaped region. Note that warmer stars to the left of the intersection ($T_{\text{eff}} \gtrsim 4800\ K$) suffer less than 50 percent depletion by the age of the Pleiades and may even have their full, interstellar supply upon arrival at the zero-age main sequence (recall Figure 16.10). Such objects, according to our definition, have no post-T Tauri epoch prior to hydrogen ignition.

The sparseness of points in the region of interest (see also Figure 17.3) is a graphic illustration of the venerable "post-T Tauri problem," introduced in Chapter 12. For many years, the dearth of such objects in Taurus-Auriga and other populous associations has been a puzzle.

---

[5] The Pleiades data in Figure 17.26 are equivalent to those already displayed within Figure 16.11. There we showed, instead of the directly observed equivalent width, the inferred surface lithium abundance. The latter is obtained theoretically via the *curve of growth*. One solves the equation of transfer for the spectral line to obtain its equivalent width as a function of the lithium column density within a model stellar atmosphere.

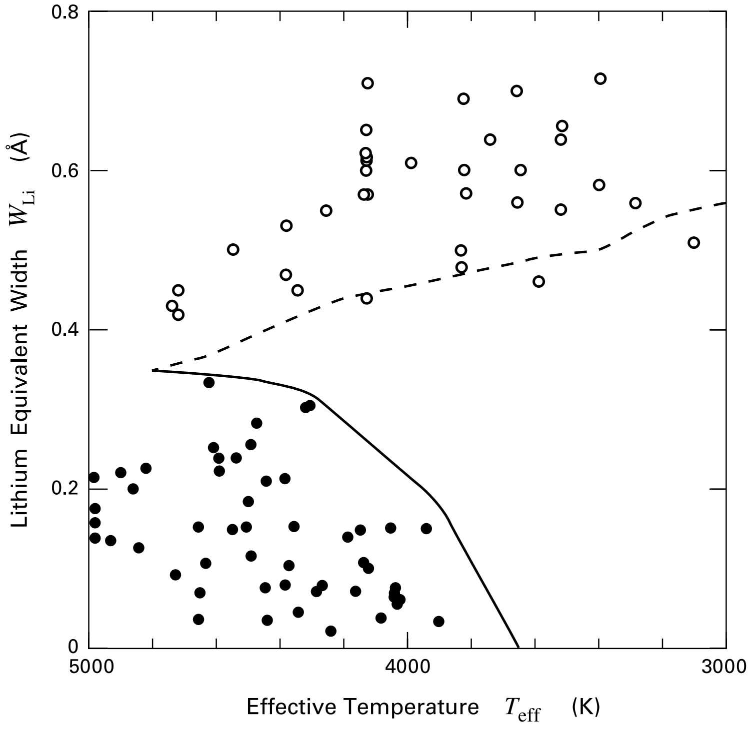


Figure 17.26 Equivalent widths of Li I 6708 Å absorption in Taurus-Auriga (*open circles*) and the Pleiades (*closed circles*). The equivalent widths are plotted as a function of stellar effective temperature.

Since a star's pre-main-sequence contraction rate slows with time, the later stages should be especially well represented, *provided* associations remain intact for the requisite time. Apparently, they do not. How the parent cloud disappears after a few million years, leaving the member stars to drift into space, is poorly understood (§ 12.4).

17.5.2 Older Associations

Large-scale surveys have uncovered groups of stars that individually display the characteristics of pre-main-sequence objects, but are not situated within any molecular cloud. These are evidently T associations, or subgroups of such associations, that have been caught in the act of dispersing. Ironically, the absence of dusty gas is a disadvantage for identification of these systems, since the members are more easily confused with field stars. The persistence of T Tauri signatures, however, allows one to pick out the objects of interest.

The best-studied example is the TW Hydrae Association. This group of two dozen stars lies at an average distance of only 55 pc and subtends a full 20° on the sky. We previously depicted this association, the closest known, in Figure 4.7. The star TW Hya itself is a classical T Tauri object, with both strong Hα emission and a near-infrared excess. Placement in the HR diagram gives it an age of 1×10^7 yr. The other members have comparable ages. While none have the Hα emission found in TW Hya, a few display remnant circumstellar disks through mid-infrared and longer-wavelength fluxes. An elliptical ring around one object, the A star HR 4796A, has been directly imaged in the optical and infrared. This structure appears to be a disk containing a substantial inner hole (see also Chapter 18).

It was excess infrared fluxes that provided the first clues to the existence of the extended TW Hya group. Additional objects were then revealed by X-ray observations and confirmed spectroscopically through their lithium absorption. Distances here are small enough that they can be established individually through parallax measurements. The objects also share a common proper motion that further serves to identify members. Both the large physical diameter spanned by the stars, some 30 pc, and their observed streaming motion are at least broadly consistent with a picture in which this small group originated as part of the Sco-Cen OB association. Over the course of 10^7 yr, the stars spread apart, while retaining the bulk motion inherited from their parent cloud.

Within the Chamaeleon region, X-ray study has revealed another grouping of similar age. This observational tool already found many stars displaced far from the molecular clouds (Figure 4.11). The B-type star η Chamaeleontis, with a luminosity of about 100 $L_\odot$, is tightly surrounded by at least a dozen other objects lying at the same distance of 100 pc. The angular extent of this group, also dubbed η Cha, is only $0.^\circ5$, corresponding to a physical size of 0.8 pc. Thus, the stars may represent a bound cluster. Almost all the remaining members are late-type objects, exhibiting optical variability, infrared excess, and modest Hα emission.

Lithium observations in both the η Cha and TW Hya groups give equivalent widths similar to those in Taurus-Auriga. That is, most objects lie above the dashed curve in Figure 17.26. Coherent associations of true post-T Tauri stars may not exist, or at least are difficult to trace observationally. By the time lithium has been reduced to less than half its initial level, the objects have apparently dispersed so widely that they are no longer distinguishable from the field population.

17.5.3 Debris Disks

As a star ends contraction and begins its main-sequence life, the excess continuum emission signifying the presence of circumstellar matter continues to decline. Sensitive array detectors nevertheless detect residual, long-wavelength flux in a significant fraction of dwarf stars. The level of emission is far below that predicted by equation (17.8). Hence the warmed dust must be optically thin to the photons it emits. Moreover, searches in millimeter lines, principally from CO, detect little or no accompanying gas. The circumstellar structures being detected at this relatively late epoch consist essentially of orbiting grains and are known as *debris disks.*

Main-sequence objects with the photometric signature of debris disks are often called *Vega-like*. Vega itself is an A star, as are many other members of this class, including the previously mentioned HR 4796A. This curious preponderance of more massive (and thus intrinsically rarer) stars is a selection effect. Such objects more easily provide the luminosity required to warm the dust to detectable levels. In any case, we shall revisit the structures around intermediate-mass stars in Chapter 18, when we describe the aftermath of the Herbig Ae/Be phase.

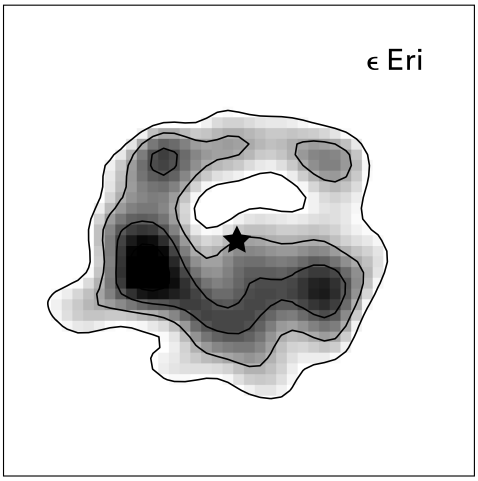


Figure 17.27 The debris disk around ϵ Eridani. Displayed, in both greyscale and contours, is the 850 μm continuum flux.

With regard to solar-type objects, the evidence for debris disks comes from surveys carried out in the far-infrared and submillimeter continuum. When detected, the total excess luminosity is at most 10^{-3} of the stellar value. This fraction declines steadily with the age of the parent star. As usual, the age may best be determined if the surveyed object is within a cluster that can be placed in the HR diagram.

On occasion, a star with an attendant disk may be so nearby that the structure can be spatially resolved. Such a fortuitous case is presented by ϵ Eridani, a K2 star only 3.2 pc from the Sun. Observations at 850 μm reveal a broad, face-on ring, with lumpy substructure (Figure 17.27). The ring peaks in flux at a radius of about 60 AU. A fit to the spectral energy distribution over a number of wavebands, along with the assumption of optically thin emission, yields a total dust mass of 4×10^{-3} *Earth* masses, or $1 \times 10^{-8}\, M_\odot$.

The age of ϵ Eri, as inferred from its Ca II H and K emission, is about 1×10^9 yr. If the progenitor was a classical T Tauri star, then the tiny amount of circumstellar material now observed is all that remains of a far more massive disk, composed of both gas and dust, that largely vanished within the first 10^7 yr of the star's life. The subsequent evolutionary path of the material is far from clear. Especially puzzling is the large central hole. This seems to violate our expectation that the mass density within any disk rises toward the central star. One intriguing possibility is that the grains well inside of 60 AU have been incorporated into larger bodies. The fractional surface area occupied by an ensemble of such objects would be reduced from that associated with small grains. Hence, a swarm of much larger objects would be undetectable, even if the surface *mass* density of the debris disk continues to rise toward the star.

The grains we do observe are quite different from those associated with interstellar gas. Multi-wavelength observations allow us to estimate their average size as 30 μm, larger than what one finds generally. Moreover, a simple theoretical argument shows that the present-day grains did *not* coagulate directly from interstellar ones, but are being continuously replenished from some other source. The reason is that any given grain experiences collisions with its neighbors on a time scale of 10^6 yr. These impacts arise because the debris disk has a finite thickness. Orbits are therefore inclined, at least to a modest degree, with respect to the plane. Collisions between intersecting orbits are sufficiently energetic that they shatter both grains, instead of leading to coalescence. The even smaller fragments are then susceptible to radiation pressure from the star, which either drives them outward or causes them to spiral into the star, again over a time brief compared to the stellar age. We shall reconsider the radiative effects in Chapter 18.

In summary, a number of processes would clear the debris disk unless material were being resupplied at a compensating rate. It is again attractive to invoke larger bodies. In this case, their mutual collisions could provide a spray of debris, including perhaps the presently observed, 30 μ grains. The original colliding objects may be asteroids, planets, or the smaller predecessors of planets known as planetesimals.

17.5.4 Planet Formation

These arguments in favor of large, orbiting bodies are of course indirect and by no means ironclad. They attain additional weight, however, from the incontrovertible fact that some main-sequence stars, including many akin to ϵ Eri, do have one or more giant planets, *i. e.*, objects with mass from about that of Jupiter ($1 \times 10^{-3}\ M_\odot$) to the deuterium fusion limit ten times greater (Chapter 16). In 1995, M. Mayor and D. Queloz announced the discovery of a planet orbiting close to 51 Pegasi, a G star at a distance of 14 pc. The actual detection was not of the planet itself, but of a tiny, periodic shift in the star's radial velocity. This shift was interpreted as the wobble induced by an orbiting mass. Because the orbital inclination was unknown, the observers could place only a lower bound on the mass value, which was 0.5 times the Jovian value. A large number of similar detections quickly followed.

The discovery of extrasolar planets confirms that formation of these objects is a common occurrence in solar-type stars. It is fitting that we close this chapter by posing, and at least attempting to answer, the following question. Why does the evolution of a young star commonly lead to attendant planets? To establish the context, let us first briefly describe the physical steps in the formation process (see Figure 17.28). This theoretical framework, admittedly incomplete, is still largely based on the vast amount of data accumulated from our own solar system.

Planets, or at least their solid cores, form through the aggregation of grains within a circumstellar disk. Many detailed investigations directed toward our own system employ as a specific disk model the *minimum-mass solar nebula*. This structure, it is assumed, had the same chemical composition as the Sun. It is observed that planets today are greatly enriched in heavy elements relative to that standard. If these elements were inherited from the primordial epoch, then one may readily determine the amount of hydrogen and helium needed to restore the proper balance in each body. The result of this exercise is that the disk within 30 AU comprised some 0.01 $M_\odot$, *i. e.*, about 10 times the total mass contained in the present-day planets. This figure represents a lower bound, since other material that never formed planets could have been lost from the disk. By considering the orbital spacing of these objects, one further concludes that the surface density within the minimum-mass nebula fell with heliocentric radius ϖ roughly as $\varpi^{-3/2}$.

The first step toward forming a planetary system is physical separation of the dust and gas components of the disk. Even if the dust were initially well mixed with the gas, it could not

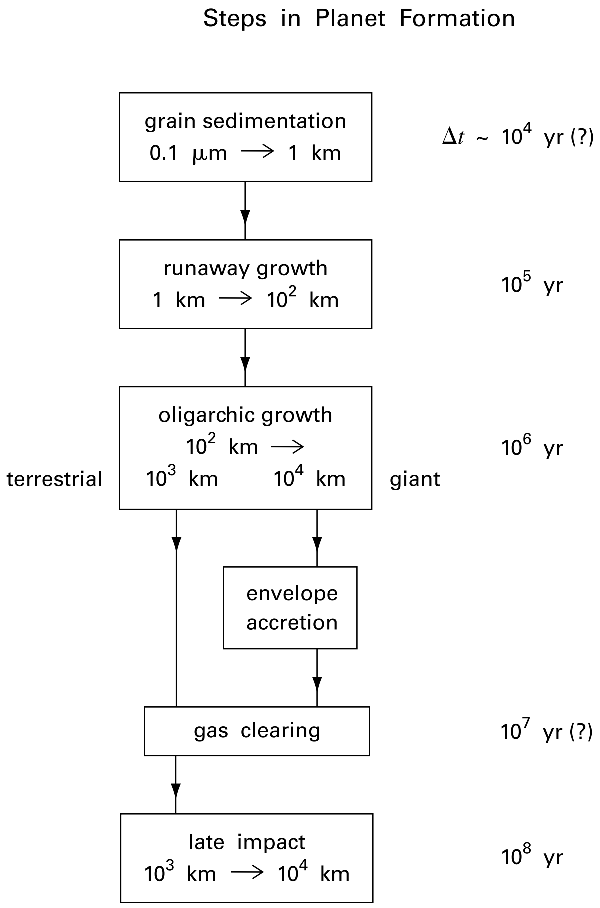


Figure 17.28 Essential elements in planet building, according to theory. The required time intervals are indicated. That for grain sedimentation is a lower limit, based on the simplest, quiescent disk model.

remain suspended forever, since individual grains have a tendency to settle under the force of gravity. The drift is toward the midplane, and the relevant gravitational force is the vertical component of that from the central star (recall Figure 11.10). As a grain falls, with the slow terminal velocity set by aerodynamic drag, it encounters and sticks to neighboring particles. The disk midplane thus becomes steadily more enriched in solid particles, which themselves are enhanced in size and mass above interstellar values.

Two circumstances complicate this picture. First, gas near the disk surface rotates about the star with an orbital speed that is sub-Keplerian. Each fluid element experiences, in addition to the radially inward component of stellar gravity, an outward force induced by the thermal pressure gradient (equation (9.25a)). While the reduction in orbital speed is relatively minor, it is even less closer to the midplane. Here, gas is enriched by dust, which has inertia but no associated pressure. Each fluid element, consisting of dust and entrained gas, therefore rotates even closer to the Keplerian speed. The net result is a vertical velocity shear at each radius. With continued settling, the fluid eventually becomes susceptible to the Kelvin-Helmholtz instability, a phenomenon we previously encountered in § 13.5 when discussing propagating jets. The ensuing turbulence in the disk must inhibit any gentle, downward drift of grains.

A second complication is that the same drag force acting in the vertical direction also removes energy from the particles' orbital motion. This loss is negligible for the smallest grains, which are well coupled to the local gas and rotate with it. Conversely, the inertia of the largest solid masses overwhelms the drag, which again has little effect. Orbital decay is strongest at some intermediate scale. In a minimum-mass solar nebula, at a heliocentric distance of 1 AU, the critical size is about 1 m. A solid object of this dimension would spiral into the Sun extremely quickly, within about 100 yr.

To date, there are no theoretical models of dust settling and coalescense that successfully address both of these concerns. The orbital decay problem implies that the growth process, however it occurs in detail, must build up objects very rapidly through the meter regime. Ironically, turbulence might actually promote such fast growth, by increasing the frequency of particle encounters. On the other hand, even a modest rise in relative velocity leads to shattering, as we have seen in the case of debris disks. It is hardly surprising that estimates of the total time needed to reach any average size for solid material vary widely. If turbulence is neglected and the sticking probability during encounters remains high, then even growth to the kilometer scale requires less than 10^4 yr at 1 AU. In this quiescent picture, the final size increase occurs through gravitational clumping of enlarged grains already residing near the midplane.

A diameter of roughly 1 km conventionally defines the lower bound for planetesimals. The evolution of a disk composed of such objects, still embedded within the original gas, is governed mainly by the pairwise gravitational attraction between the solid bodies. This attraction perturbs their orbits about the central star. Physical collisions between planetesimals lead either to merger or to breakup into smaller pieces.

The path of an individual planetesimal changes impulsively as a result of the gravitational tug from a close neighbor. In a statistical sense, the effect of such mutual perturbations is to introduce a random motion into the swarm of objects, *i. e.*, a velocity component in addition to that associated with circular, coplanar orbits. A large number of perturbations result in equipartition of this random kinetic energy among the population. Thus, more massive objects tend to have lower random speeds. (Compare the discussion of dynamical relaxation in § 4.4.) This velocity reduction, together with the stronger gravitational pull of these bodies, increases their collision cross section with other planetesimals. In addition, any fragments generated during a collision tend to fall back toward the main body rather than escape. The result is *runaway growth*. The most massive planetesimal in a given region of the disk gains even more mass by devouring its neighbors.

As runaway growth proceeds, the initial high population of kilometer-size planetesimals is partially transformed into a reduced number of larger objects spaced at regular intervals in semimajor axis. During the phase of *oligarchic growth*, this ensemble evolves in unison, as the background particles continue to be absorbed through collisions. The process ultimately yields several hundred *protoplanets*, with sizes ranging from 10^3 km to an order of magnitude greater. Terrestrial planets like the Earth build up slowly, over 10^8 yr, through merger of the smaller protoplanets. These perturb each others' orbits until large radial excursions lead eventually to orbit crossing and high-impact collisions.

Giant planets form by another route. Here, a more massive protoplanet is able to draw in gas from the surrounding disk. This accretion onto the rocky planetary core is predicted to be very rapid. On the other hand, the process must also end abruptly, as evidenced by the composition of our own giant planets. Relative to its heavy elements, Jupiter contains only a tenth the hydrogen and helium of a gas with solar composition. One attractive idea for the termination is that the growing object eventually opens a gap in the disk through the action of gravitational torques. Prior to gap formation, these same torques also cause the planets to migrate toward the central star.[6]

Return finally to these central objects and to the observational data at the heart of this chapter. The theory just outlined places little constraint on when planet formation actually begins in the life of a star. Dust sedimentation could proceed within any disk, including one being created during protostellar infall. On the other hand, the millimeter fluxes from classical T Tauri stars originate from small dust grains, as does the infrared luminosity much further inside. Sedimentation has therefore *not* consumed all the dust at this time, up to 3×10^6 yr or so after protostellar collapse. Apparently, the quiescent disk environment envisioned in the simplest version of sedimentation theory is not maintained.

At a more advanced age, classical T Tauri observations tell us that both signatures of dust vanish. Have planets formed by this point? We do not know, but the coalescence of most grains into planetesimals would indeed curtail both millimeter and infrared fluxes. Significantly, the stellar observations also show no indication, either through cold CO or warm H_2, of a substantial gaseous component to the disk. The gas must have either accreted onto the star or else been driven into space, perhaps by photoevaporation or action of the stellar wind. Our ignorance of this important event constitutes a major deficit in both stellar and planetary theory. In any case, the formation of giant planets must occur prior to the dispersal. The relatively long buildup of terrestrial planets takes place throughout the post-T Tauri phase and may well be continuing in the young dwarfs that now exhibit debris disks.

Chapter Summary

Spectra of classical T Tauri stars include a large number of broad optical and ultraviolet emission lines, of which Hα predominates. Weak-lined objects, by definition, are lacking in this regard. Also present in classical objects is an excess continuum flux in the blue and ultraviolet, as well as one at infrared wavelengths. The former arises from shocks associated with winds and infall, the latter from heated, circumstellar dust. All T Tauri stars have enhanced X-ray emission, which appears to be the superposition of numerous, unresolved flares. Some also exhibit flaring

---

[6] Theorists have also suggested, on the basis of simulations, that giant planets may form intact, through prompt, dynamical fragmentation of the underlying disk. As in the analogous hypothesis for stellar cluster formation in molecular clouds (§ 12.1), it is difficult to see how the parent object evolved to a configuration susceptible to such breakup. The rapidly growing data on extrasolar planets should eventually test both this and the gas accretion picture.

in the radio continuum. Both X-ray and radio flares arise from reconnecting magnetic field lines outside the stellar surface.

The bulk of Hα emission originates in numerous shocks located relatively close to the star, although there is frequently also shock-heated gas in a more directed wind. These microjets are less than 10^3 AU in extent, much smaller than the jets from more embedded stars. Microjets also appear in optical forbidden-line emission, such as that from O I. Analysis of the forbidden lines reveals that the average mass loss rate in T Tauri winds is about $10^{-8}\ M_\odot$ yr^{-1}. A subclass dubbed YY Orionis stars additionally shows, through permitted optical line profiles, evidence for infall. The mass infall rate appears comparable to that in the wind, and both flows change rapidly with time.

The spectral energy distribution of classical T Tauri stars in the infrared gives the run of dust temperature with radius in their circumstellar disks. The shallowness of observed temperature falloffs indicates that the disks are substantially flared. Observed dips in the spectral energy distribution apparently signify a central hole in the surface density. In any case, the total mass in dust may be crudely estimated from continuum radiation in the submillimeter regime. The gas component reveals itself through millimeter lines, whose Doppler shift reflects rotational motion about the central object. The total emission from a disk includes, at a minimum, reradiated light from this star. There may also be energy release from internal mass torquing, including that associated with the tangling and reconnection of magnetic field lines.

T Tauri stars exhibit both regular and erratic variability in their emission. The first type, seen in weak-lined stars, arises from the rotation of cool spots, where strong magnetic fields emerge from the stellar surface. Classical T Tauri stars show more aperiodic fluctuations and often have both cool and hot spots. Flaring at optical wavelengths is also a hallmark of young, pre-main-sequence stars. The class known as EXors display large outbursts on decade time scales. FU Orionis stars exhibit the most dramatic changes. Here, the luminosity increases by two orders of magnitude in the course of a year or so. These outbursts, of unknown origin, occur repeatedly during pre-main-sequence contraction.

Even the youngest weak-lined objects have scant observational evidence of circumstellar disks, while the signatures in classical stars fade within 10^7 yr. Both types continue to exhibit X-ray emission and surface lithium into the ensuing, post-T Tauri phase. Observed associations contain so few of these older objects that they must disperse relatively quickly. Some of the recently dispersed members are still visible as discrete groups. Still older, low-mass stars can possess sparse remnant disks composed of dust grains. These grains must be continually produced through the mutual collision of solid, orbiting masses. The discovery of extrasolar giant planets shows that such bodies are commonly formed around T Tauri stars, before the disappearance of their gaseous disks.

Suggested Reading

Section 17.1 The general properties of T Tauri stars are summarized in

Bertout, C. 1989, ARAA, 27, 351.

For their X-ray and radio emission, see

Feigelson, E. D. & Montmerle, T. 1999, ARAA, 37, 363

Güdel, M. 2002, ARAA, 40, 217.

Section 17.2 Permitted optical emission lines, including Hα, have been extensively studied observationally. See, for example,

Alencar, S. H. P. & Basri, G. 2000, ApJ, 119, 1881.

An infall model for Balmer emission is presented in

Calvet, N. & Hartmann, L. 1992, ApJ, 386, 239.

Forbidden-line emission and its utility as a wind probe are discussed by

Edwards, S., Ray, T. P., & Mundt, R. 1993, in *Protostars and Planets III*, ed. E. H. Levy & J. I. Lunine (Tucson: U. of Arizona), p. 567

Hirth, G. A., Mundt, R., & Solf, J. 1997, AAS, 126, 437.

The shock interpretation of optical continuum radiation is detailed in

Lamzin, S. A. 1998, Astron. Rep., 42, 322.

Section 17.3 For a description of how observed spectral energy distributions yield disk properties, see

Beckwith, S. V. W. 1999, in *The Origin of Stars and Planetary Systems*, ed. C. J. Lada & N. D. Kylafis (Dordrecht: Reidel), p. 579.

An influential study of disk flaring and surface heating is

Chiang, E. I. & Goldreich, P. 1997, ApJ, 490, 368.

Imaging studies in the millimeter were preceded by optical polarization maps on larger scales. Here, the geometrical pattern of polarization vectors suggests flattened structures. See

Bastien, P. 1996, in *Polarimetry in the Interstellar Medium*, ed. W. G. Roberge & D. C. B. Whittet (San Francisco: ASP), p. 297.

The basic magnetorotational instability has long been known and was rediscovered in the disk context by

Balbus, S. A. & Hawley, J. F. 1991, ApJ, 376, 214.

Section 17.4 Detailed accounts of T Tauri variability are in

Herbst, W., Herbst, D. K., Grossman, E. J., & Weinstein, D. 1994, AJ, 108, 1906

Menard, F. & Bertout, C. 1999, in *The Origin of Stars and Planetary Systems*, ed. C. J. Lada & N. D. Kylafis (Dordrecht: Reidel), p. 341.

For stellar magnetic field measurements, see

Johns-Krull, C. M., Valenti, J. A., Piskunov, N. E., Saar, S. H., & Hatzes, A. P. 2001, in *Magnetic Fields Across the HR Diagram*, ed. G. Mathys, S. K. Solankik, & D. T. Wickramasinghe (San Francisco: ASP), p. 527.

Observational knowledge of FU Orionis outbursts is summarized by

Herbig, G. H. 1977, ApJ, 217, 693.

There is also much useful information in

Hartmann, L. 1998, *Accretion Processes in Star Formation*, (New York: Cambridge U. Press), Chapter 7.

This monograph analyzes many features of young stars through phenomenological disk models.

Section 17.5 A statistical analysis that documents the evolution of T Tauri flaring is

Guenther, E. W. & Ball, M. 1999, AA, 347, 508.

Various authors describe the ongoing search for older associations in

Jayawardhana, R. & Greene, T. P. (ed.) 2002, *Young Stars Near Earth: Progress and Prospects*, (San Franscisco: ASP).

The discovery of the first extrasolar giant planet around a main-sequence star was due to

Mayor, M. & Queloz, D. 1995, Nature, 378, 355.

For a review of this developing field, see

Marcy, G. W., Cochran, W. D., & Mayor, M. 2000, in *Protostars and Planets IV*, ed. V. Mannings, A. P. Boss, & S. S. Russell, (Tuscon: U. of Arizona), p. 1285.

Planet formation theory is summarized in

Lissauer, J. J. 1993, ARAA, 31, 129

Ruden, S. P. 1999, in *The Origin of Stars and Planetary Systems*, eds. C. J. Lada & N. D. Kylafis (Dordrecht: Reidel), p. 643.

The Formation of Stars
Steven W. Stahler and Francesco Palla

18 Herbig Ae/Be Stars

In this last chapter on the pre-main-sequence phase, we describe objects of intermediate mass. These Herbig Ae/Be stars share certain key features with the T Tauri class, and others with even more massive and luminous O and B stars. The latter first become optically visible after they have already begun to fuse hydrogen. Thus, the stars we now discuss are the only remaining class of young objects that evolve purely through gravitational contraction.

The study of Ae/Be stars is less advanced than that of their low-mass counterparts. The main impediment here is the relative scarcity of the objects. According to equation (4.6) for the initial mass function, the number of stars produced in the range of interest, from 2 to 10 $M_\odot$ is only 3 percent of those from 0.1 to 2 $M_\odot$. Exacerbating the problem is the shorter contraction time at larger masses. On the other hand, the higher luminosity does facilitate discovery; the number of known sources now exceeds 100.

We begin this chapter by summarizing the general properties of this class. We examine not only the observed characteristics of the stars themselves, but also their spatial relationship to neighboring objects. Section 18.2 turns from these empirical matters to theory. We demonstrate that quasi-static contraction in this mass range differs qualitatively from that in T Tauri stars. This difference reflects the interior mode of energy transport. Theory also predicts that the more massive objects should undergo a transient period of bulk oscillations, which has now been observed.

In the final two sections, we shift our attention to the stellar environments. We show how infrared and millimeter studies, while providing evidence for dusty, circumstellar material, have not yet clarified its structure or dynamics. We also cover the winds from Ae/Be stars, which dissipate remnant cloud gas. These outflows, whose origin is poorly understood, are apparently responsible for the observed X-ray emission. Energetic radiation, including the ultraviolet component of the stellar spectrum, ionizes and heats surrounding matter. Finally, § 18.4 focuses on circumstellar disks and on the gradual depletion of their dust components.

18.1 Basic Properties

Within the HR diagram, Herbig Ae/Be stars fall close to, but below, the intersection of the birthline and the zero-age main sequence. A glance at the diagram in Figure 16.2, which includes several dozen of these objects, shows that they have luminosities ranging from about 10 to $10^3\ L_\odot$, and $T_{\rm eff}$-values from 8,000 to 20,000 K. One natural question is where to find pre-main-sequence objects with these characteristics. Truly massive stars, as we know, are born in the most crowded regions of giant cloud complexes. Low-mass T Tauri stars are also found in such locales, but appear in sparser associations and more isolated dark clouds as well. What type of environment spawns objects of intermediate mass?

The Formation of Stars. Steven W. Stahler and Francesco Palla

ISBN: 3-527-40559-3

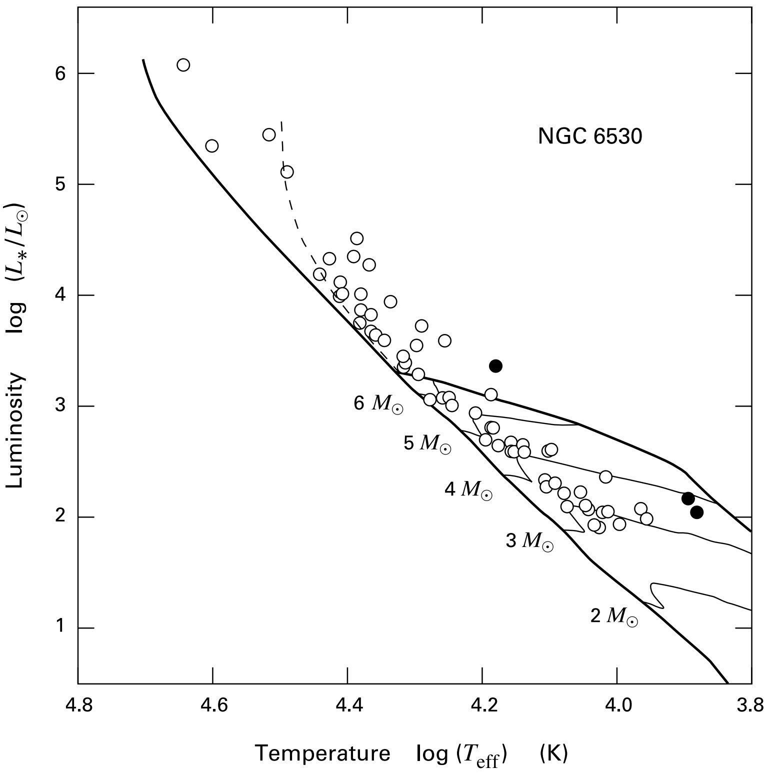


Figure 18.1 HR diagram for the cluster NGC 6530 in Sagittarius. Filled circles represent member stars that have optical emission lines. The birthline and ZAMS are shown, as well as selected pre-main-sequence tracks (*solid curves*) and the 5×10^6 yr post-main-sequence isochrone (*dashed curve*).

18.1.1 Birth Environments

Toward the Galactic center in the constellation Sagittarius is a conspicuous HII region, the Lagoon Nebula (M8). This patch of ionized gas, although at a distance of 1.8 kpc, is nevertheless bright enough to be seen with the naked eye. It is situated in front of a massive molecular cloud and engulfs a cluster known as NGC 6530. Other clusters nearby were also born from the background molecular gas. All are part of the extended Sgr OB1 association.

Many stars ostensibly within the area encompassed by NGC 6530 are actually foreground objects. Establishing cluster membership in this case is no simple task. Additionally, the large distance limits observations to the more luminous objects in the system. Figure 18.1 shows the HR diagram for a few dozen stars that have been studied both photometrically and spectroscopically. In most cases, members were selected by their anomalous ratios of color excesses, an indication that their reddening arises from circumstellar and intracluster material, and not simply from interstellar grains situated along the line of sight.

In the diagram of Figure 18.1, most stars are quite close to the main sequence, but there is increasing dispersion above this curve toward both higher and lower masses. If the high-mass effect reflects true post-main-sequence evolution, then the relevant stars were born at most 5×10^6 yr ago. We show the isochrone corresponding to this age in the figure. The dispersion among intermediate-mass members arises from pre-main-sequence contraction, although some of the scatter across the diagram represents observational errors. Despite the generous sampling of objects from 3 to 10 $M_\odot$, few meet all the criteria for true Herbig Ae/Be stars. By definition, the source must not only have a spectral type of A or B and be located near cloud gas, but must also display optical *emission* lines, as indicated by the nomenclature.[1] The filled circles in Figure 18.1 represent the handful of stars with this last property.

The NGC 6530 group is not unusual in this regard. Another cluster within the Sgr OB1 association, NGC 6611, also has a substantial, intermediate-mass population, of which perhaps 4 objects exhibit line emission. We presented as Figures 4.9d and 12.29, respectively, the HR diagrams for NGC 2264 in Monoceros and the especially crowded Orion Nebula Cluster. In both diagrams, there appear to be many objects in the intermediate range. Yet a search for Hα emission, in particular, turns up only a few sources. Note the technical issue one faces in making this assessment, a problem that has occasionally misled observers. Since there are even more massive, ionizing stars in all these regions, the *ambient* Hα emission arising from diffuse gas must first be subtracted at each location.

The paucity of emission-line objects within young clusters is a striking departure from the T Tauri situation. If one were to posit that Herbig Ae/Be stars are essentially similar to classical T Tauri objects, then the weak-lined analogues would comprise the vast majority of the intermediate-mass population. However, this basic identification is not secure. As we shall describe, the Ae/Be stars exhibit other key differences from low-mass objects generally.

Large clouds are not the only structures producing intermediate-mass stars. While giant complexes do spawn the largest contiguous populations, individual sources are also seen within more isolated dark clouds. On the other hand, Ae/Be stars are seldom found in the sparser molecular gas that gives rise to T associations. These findings are simply manifestations of the statistical rarity of intermediate-mass objects, together with the unremarkable fact that higher-mass clouds tend to form more stars in total.

Of greater interest is the proximate cloud environment, *e. g.*, the material within roughly 0.1 pc of the star itself. In the original definition of the class, Herbig Ae/Be stars were required to illuminate optical reflection nebulae. This stipulation was added to ensure that the star was physically associated with cloud material, and not a foreground or background object. Millimeter observations in spectral lines or continuum radiation accomplish the same goal. Additionally, this radio mapping has provided more detail concerning the state of nearby gas.

While a variety of results are found, Figure 18.2 illustrates the extremes. Here we see, through ^{13}C^{16}O ($J = 1 \rightarrow 0$) emission at 2.7 mm, a portion of cloud material near the edge of the Cepheus Bubble, a large shell created by massive stars in the Cep OB2 association (Figure 4.14). The region shown coincides optically with a reflection nebula known as

---

[1] It is common practice to include within the Herbig Ae/Be class F-type pre-main-sequence stars. The reason is simply that T Tauri stars are conventionally restricted to having a spectral type later than G0. One thus avoids establishing yet a third class for the F stars alone. We have no objection to the expanded definition, but will focus in this chapter on A- and B-type stars.

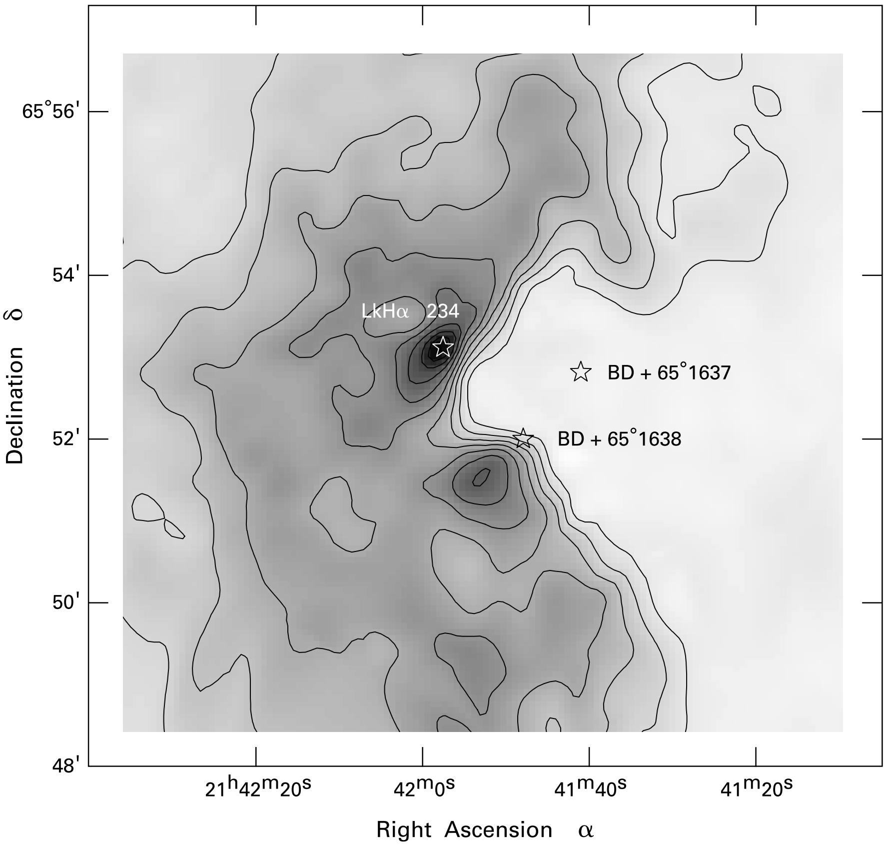


Figure 18.2 Three Herbig Ae/Be stars in the NGC 7129 region. Both the contours and greyscale shading represent emission from $^{13}C^{16}O$ ($J = 1 \rightarrow 0$). At a distance of 1250 pc, the physical size of the map is about 3×3 pc.

NGC 7129 and contains three visible Herbig Ae/Be stars. That on the upper left, a B6 object called LkHα 234, is still embedded within cloud material. The star coincides with a local peak in gas, similar to a dense core that might produce a low-mass object. In contrast, the star on its right, BD+65°1637, lies within relatively sparse HI gas and has apparently evacuated its own, local cavity, further compressing adjacent cloud material. An intermediate case is presented by the star below these two, BD+65°1638, which sits just at the edge of molecular gas.

Judging from several dozen of these observations, there are two factors promoting the gas dispersal we are seeing from BD+65°1637. The first is stellar age. Not surprisingly, the very youngest Ae/Be stars are also the most obscured and are apt to be situated near peaks of millimeter emission. The second factor is effective temperature. Earlier-type Be objects are more able to clear extensive volumes of cloud gas. The star BD+65°1637 has a spectral type of B3, placing it close to the intersection of the birthline and zero-age main sequence. The physical agents that are available for dispelling the gas are stellar winds and ultraviolet radiation. We shall explore both in § 18.3 below.

18.1.2 Low-Mass Companions

Return to the clump of molecular gas surrounding LkHα 234 in Figure 18.2. If this entity is truly analogous to a dense core, then it recently created the interior star through inside-out collapse. Are all Herbig Ae/Be stars formed in a manner analogous to low-mass objects? We have argued throughout this book that the most luminous and massive stars probably do *not* form through the process of dense core buildup to the point of gravitational collapse, but through the coalescense of previously formed stars. The main line of evidence bolstering this view is the finding that the youngest massive stars are invariably located at the crowded centers of stellar clusters. It is gratifying that intermediate-mass objects show evidence for *both* formation modes.

As we discussed in § 12.5, there is again a division according to spectral type. While Herbig Ae stars are relatively isolated entities, hotter stars are frequently, though not always, surrounded by a retinue of lower-mass companions, *i. e.*, T Tauri stars. We showed an example in Plate 5, an infrared image of the Be star BD+40°4124 and its attendant objects, both visible and embedded. Of the latter, 40 objects have an infrared excess and therefore cannot be background sources. The star BD+65°1637 in Figure 18.2 is surrounded by many low-mass stars, not shown in the map. These objects are revealed by virtue of the fact that the central object has aided in dispersing nearby cloud gas. More young, low-mass stars undoubtedly lie within the thick shell. We emphasize that such groups, while they may be visually impressive, are still less dense than those around newly formed O stars, such as the Orion Nebula Cluster or NGC 3603 (Plate 7). Nevertheless, the existence of a Be star at the center of an aggregate does suggest that the more massive, central object formed through some coalescense process.

Additional support for this hypothesis comes from examination of orbiting companions to Ae and Be stars. High-resolution surveys for visual systems find that the differential binary frequency is at least 0.5 for $\log P$ between 5 and 7. Here, P is the orbital period in days, as estimated using the total system mass and the projection-corrected separation. Figures 12.7 and 12.13 show that this frequency is more than double that of G dwarfs in the same period interval and also exceeds the figure for T Tauri primaries. The suggestion is that the formation process for intermediate-mass stars, occurring in a crowded environment, often results in the capture of neighboring stars. In contrast, capture is not a plausible origin for even the widest T Tauri binaries.

Whether this overabundance of binaries holds also for shorter periods is difficult to say. As with O stars, both the high flux ratio and the greater distances to these systems make the imaging of tighter pairs more problematic. It is also possible to misinterpret the members of a low-mass cluster surrounding a given star as gravitationally bound companions. For periods less than 100 days, one may detect, through spectroscopy, radial velocity variations of the primary. While the overall fraction of such close binaries is small, the numbers are comparable to those for T Tauri stars, which in turn mimic G-dwarf primaries (Figure 12.13). Chapter 12 discussed in detail the system TY CrA, consisting of a 3.2 $M_\odot$ primary and a T Tauri secondary. Here the primary has already reached the main sequence (Figure 12.16). In other observed systems, even the more massive component is still undergoing pre-main-sequence contraction.

18.1.3 Rotation

We stress that the current census of tight binaries with intermediate-mass primaries is undoubtedly incomplete. Yet another difficulty is presented by the fact that Herbig Ae/Be stars tend to rotate quite rapidly. We saw in Chapter 16 how surface rotation broadens all photospheric absorption lines. It thus becomes harder to discern the periodic shift in the line center stemming from gravitational tug of an orbiting, companion object.

Of course, the faster rotation in Herbig Ae/Be stars is itself of interest. Figure 18.3 summarizes observations of the projected equatorial velocity, $V_{\rm eq}\,\sin\,i$, here shown as a function of stellar mass. These speeds were obtained by comparing a portion of the optical spectrum with that of nonrotating stars of identical spectral type. Thus, the same broadening that hinders detection of binaries provides key information about the star itself, just as for T Tauri stars. The typical equatorial speed for Ae/Be stars is about 150 km s^{-1}. Here we may recall Figure 16.20, which displayed $V_{\rm eq}$ in the Orion Nebula Cluster, for $M_* < 3\,M_\odot$. While some Orion stars have speeds exceeding 100 km s^{-1}, the vast majority of $V_{\rm eq}$-values are much lower. On the other hand, intermediate-mass objects are still rotating well under their respective breakup speeds $V_{\rm cen}$, displayed in Figure 18.3 for stars along the birthline.

Recall from our previous discussion that solar-type young stars rotate more slowly than those on the zero-age main sequence. The situation for intermediate-mass stars is not as clearcut. The dashed curve in Figure 18.3 shows the average observed values of $V_{\rm eq}\,\sin\,i$ for main-sequence objects. Herbig Ae/Be stars have rotation velocities that are roughly consistent with this curve, but there is considerable scatter, both above and below. Those objects with the very lowest equatorial speeds are also the most displaced above the main sequence in the HR diagram. Their future contraction, through roughly a factor of two in radius, will spin them up close to main-sequence velocities, assuming strict conservation of angular momentum.

18.1.4 Spectroscopic Features

It is only the *absorption* lines of a Herbig Ae or Be star whose broadening yields rotational speeds. The additional emission lines superposed on the spectrum have more complex profiles that cannot arise simply from Doppler shifts on a rotating surface. As in T Tauri stars, the emission must stem from circumstellar shocks. The physical origin and disposition of the infall and wind activity powering these shocks is at least as problematic in the more massive objects.

The roster of absorption lines in a Herbig star is essentially the same as in a main-sequence object of similar mass. There is also little continuum veiling, so that the relative line strengths are true photospheric values. The assignment of spectral types thus presents no special difficulty, even for the youngest, optically revealed sources. Figure 18.4 shows the spectrum of a representative Herbig Be star, LkHα 220. This $4\,M_\odot$ object, with a luminosity of $170\,L_\odot$ and $T_{\rm eff}$ of 1.4×10^4 K, is located in an R association within Canis Majoris, at a distance of 1 kpc. The star formation region is part of a ring-like nebula within a large molecular cloud.

Near its blue end, the spectrum is marked by a sharp rise in the continuum toward longer wavelengths. This rise occurs just past the Balmer jump at 3647 Å. Recall from Figure 17.1 that classical T Tauri stars also exhibit a discontinuity here, but in the opposite sense, *i. e.*, the continuum is depressed at longer wavelengths. T Tauri stars also display a cluster of emission

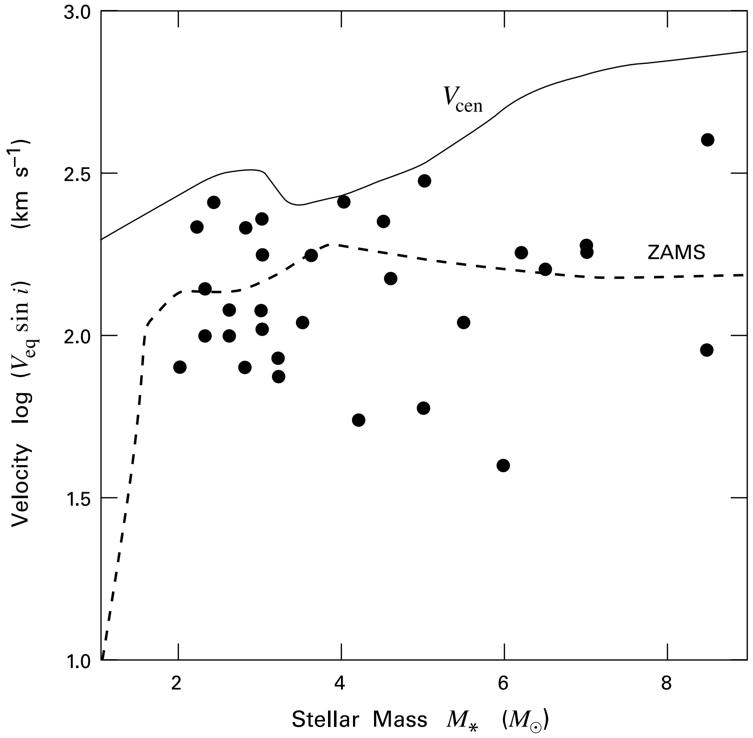


Figure 18.3 Spectroscopically measured rotation speeds of Herbig Ae/Be stars. The solid curve represents breakup speed for each stellar mass along the birthline. Main-sequence rotational speeds are shown by the dashed curve.

lines atop this lowered continuum. These lines include both Hβ at 4861 Å and higher members of the series, which terminates at the Balmer jump. In the Be star, the same transitions appear as deep absorption troughs, although Hβ has an emission core. Also in absorption here are the He I line at 5876 Å and the two Na I D lines near 5900 Å.

Figure 18.4 includes, for comparison, a main-sequence star of nearly the same spectral type. It is evident that both the shape of the continuum and the prominent absorption lines are reproduced. The one noticeable difference is the presence of emission lines in the Herbig star. These include, most prominently, Hα at 6563 Å. Others are the core of Hβ, a triplet of Fe II lines near 5000 Å, and a weak [O I] line at 6300 Å. Reference to Table 17.1 shows that all these features appear as well in classical T Tauri stars.

Their strong emission in Hα is a valuable tool for locating Herbig Ae/Be stars in observational surveys. Dedicated monitoring of this line shows that the profile varies on time scales at least as short as days. While the same is true of classical T Tauri stars, there are differences. In the low-mass case, we recall, most profiles either have a blueshifted absorption dip or are symmetric (Figure 17.8). Absorption dips are also common in Herbig Ae/Be stars, and we shall

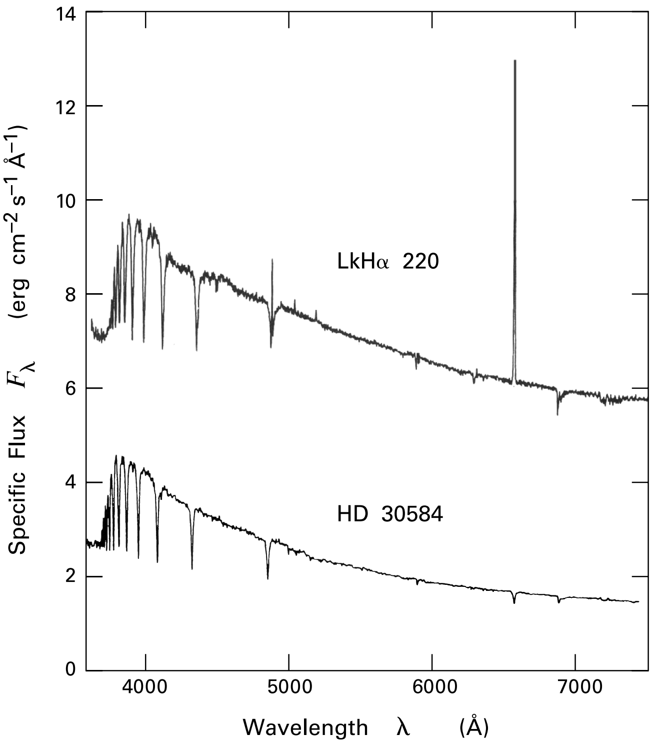


Figure 18.4 Optical spectrum of a Herbig Be star (*top*), compared to a main-sequence star of type B6 (*bottom*), which nearly matches in effective temperature. Note the prominent Hα emission in the Herbig star.

discuss their dynamical significance in § 18.3 below. Symmetric Hα profiles, on the other hand, are only occasionally seen. This curiosity may eventually provide a significant clue in future modeling of the circumstellar environment.

It is important to distinguish Herbig Be stars from another group of older objects that spans the same range in spectral type and also exhibits Hα emission. These *classical Be stars*, which lie on the main sequence or have evolved past it, are actually quite numerous, constituting some 15 percent of the entire B-star population. The optical emission here is also highly variable and can occur in other Balmer lines of hydrogen. Their rapid rotation is one of the most striking features of these stars. The average $V_{\rm eq} \sin i$ is 250 km s^{-1}, greater than in a typical Herbig star, although still below breakup. Several lines of indirect evidence, as well as interferometric imaging in a few cases, show that the emitting gas is configured in a flattened disk, where it is being irradiated by the central star. Some objects also show, through narrow absorption lines, the presence of nonrotating circumstellar matter. This "shell," which appears intermittently in a given object, represents absorbing disk material seen edge on. The disk itself arises from the star shedding material in the equatorial plane. There is also a higher-velocity wind emanating mainly from the poles. Note finally that classical Be stars exhibit a modest infrared excess, attributable to free-free emission from hot, circumstellar gas. Most Herbig Ae/Be stars, as we shall see again in § 18.4, have much larger excesses arising from thermal dust emission.

18.1.5 UX Ori Stars

Variability of the line profiles in Ae/Be stars is a ubiquitous phenomenon that holds for more than Hα. For example, the well-studied Ae star AB Aur exhibits emission in the Mg II h and k doublet near 2800 Å. The profiles have blueshifted absorption dips that change periodically in strength. Other lines, including pure absorption features, vary either erratically or in a periodic fashion. Both types of changes are also found in the broadband flux. Some stars exhibit longterm fading or brightening over decades of photometric observation. Many have relatively low-amplitude fluctuations over much shorter time spans. However, strictly periodic alterations in broadband flux, such as those observed in weak-lined T Tauri stars, are rare.

The most striking photometric changes are those exhibited by *UX Ori stars*. These objects undergo, in a stochastic manner, sudden drops in optical light. The change can be as large as 3 mag in V (a factor of 16 in flux). After a few days near minimum, there ensues a relatively slow recovery to normal brightness, typically lasting several weeks. Interestingly, all the sources for which such behavior has been documented are of spectral type B8 and later. The UX Ori phenomenon occurs in about a fourth of all Herbig Ae stars and also has been seen in a few F- and G-type objects.

Other effects that accompany the sudden dimming help us to understand its origin. The light from Herbig Ae/Be stars is often linearly polarized, although the degree of polarization is usually low and erratically variable. During a UX Ori-type fading, the polarization level increases markedly, reaching up to 8 percent at minimum light. The variation of the level with wavelength makes it clear that we are witnessing the effect of scattering by circumstellar dust. Additional grains, interposed directly along the line of sight, are presumably responsible for the dimming itself.

If the latter statement is true, then we expect the starlight to redden as its fades. The actual behavior is more interesting and complex. Figure 18.5 is a 3 yr record of photometric magnitudes and colors for the prototypical object UX Ori. Plotted here is the apparent V-magnitude versus the $B-V$ color index. Recall that a numerical increase in $B-V$ signifies reddening. Progressive reddening indeed occurs during the earlier part of the fadeout, when $V \lesssim 11.5\ \mathrm{mag}$. As the V-band flux continues to decline, $B - V$ reaches a maximum, then starts to decrease. This *blueing* phenomenon is a common feature of all UX Ori stars.

Figure 18.6 sketches a sequence of events that neatly accounts for the observations. Suppose that an optically thick dust cloud passes in front of the star. Suppose further that the star and cloud have comparable sizes. At first, the cloud both diminishes and reddens the received flux, as the absorbed photons are reemitted by grains at longer wavelengths. A fraction of the incoming radiation also consists of photons scattering off dust that is not directly in the line of sight. Eventually, the extinction by the cloud becomes so great that *most* of the observed radiation arises from scattering. This process is most efficient at shorter wavelengths and also creates linearly polarized radiation for any asymmetric distribution of grains. We may thus understand both the blueing effect and the enhanced polarization level near minimum light.

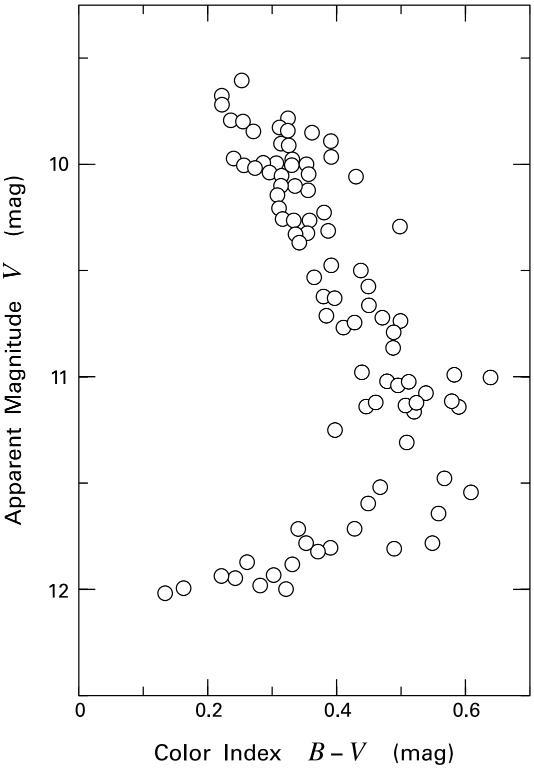


Figure 18.5 Record of apparent V-magnitude as a function of $B-V$ color in the star UX Ori. As the magnitude dims, the star first reddens, then becomes more blue.

The origin of the dust clouds themselves is less clear. These may be clumps in orbit around the star, perhaps residing in a disk seen nearly edge on. If so, then the aperiodic character of the dimming implies that we are witnessing several clumps at different radii. The same stars also exhibit transient absorption lines in hydrogen and heavy elements. These dips are redshifted, as in YY Orionis objects (§ 17.2). Hence, a complete dynamical picture must also include infalling motion of circumstellar gas. Perhaps a flattened distribution of circulating and infalling clumps exists around most Herbig Be stars, but only those in the UX Ori category have this material properly oriented with respect to our line of sight.

18.2 Nonhomologous Evolution

Let us now turn from our empirical description of Herbig Ae/Be stars to considerations of a more theoretical nature. We note at the outset that theory has not yet succeeded in explaining, at least in any detail, numerous features of these objects, such as their emission-line spectra or the X-rays we shall later discuss. On the other hand, Figure 16.2 indicates that the most basic property of all, the distribution of the population within the HR diagram, is essentially understood. Calculations have also shown how both the structure and evolution of these stars differ from low-mass objects. Our purpose in this section is to delineate these differences.

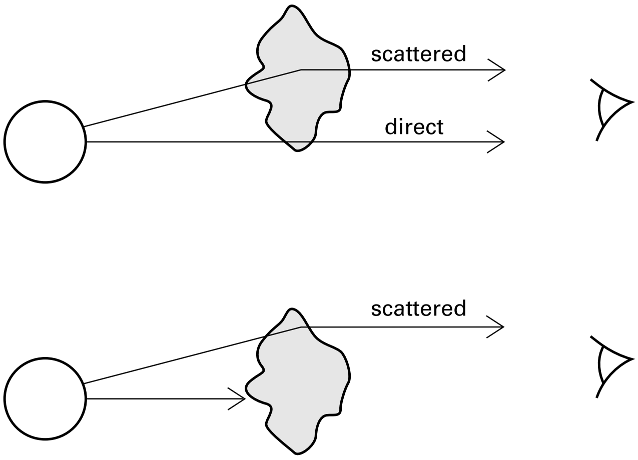


Figure 18.6 Theoretical explanation of the UX Ori phenomenon. A dust cloud partially obscuring the star transmits both direct, reddened light and scattered, blue photons. Eventually, obscuration is so great that only the scattered component appears in the substantially diminished flux.

18.2.1 Thermal Relaxation

Herbig Ae/Be stars evolve from intermediate-mass protostars. We previously saw, in § 11.4, how a protostar becomes stable against convection once its mass exceeds about 2 $M_\odot$. After the end of infall, the newly visible pre-main-sequence object is also radiatively stable, in contrast to a fully convective star of lower mass. Thus, the mode of interior energy transport has changed, with consequences for the process of contraction itself.

At this point, we should recall the physical basis for this transition to radiative stability within protostars. Consider first those of relatively low mass. Here the internal luminosity, which is largely supplied by central deuterium fusion, exceeds the maximum value that can be transported through radiation. The latter, expressed as an average throughout the stellar interior, is given by $\langle L_{\rm crit}\rangle$ in equation (16.14). According to this equation, $\langle L_{\rm crit}\rangle$ rises steeply as the mass M_* increases. This rise, in turn, reflects the declining opacity throughout the protostar. Eventually, $\langle L_{\rm crit}\rangle$ exceeds the deuterium-generated luminosity, which is therefore able to traverse the interior through radiative diffusion.

Another key development in protostars of intermediate mass is the shifting locus of deuterium burning. Once the central portion of the star becomes radiatively stable deuterium is rapidly consumed, and persists only in an outer mantle. Fusion occurs at the base of this mantle and leads to a dramatic swelling of the stellar radius (Figure 11.21). This swelling endows young, pre-main-sequence stars in the proper mass range with greater luminosity. The consequent lifting of the birthline in the HR diagram is just what is needed to encompass the known positions of Herbig Ae/Be stars (Figure 16.2). With even further mass addition, the radius of the protostar shrinks, and gravitational contraction takes over as the main source of luminosity.

A star of intermediate mass that has just finished accreting finds itself in a peculiar situation. According to Figure 16.4, the luminosity $\langle L_{\rm crit}\rangle$ is higher than L_*, the value emitted by the surface layers. But $\langle L_{\rm crit} \rangle$ is not just the *maximum* luminosity that can be carried by radiation. It is also, to the same degree of approximation, the *actual* interior luminosity within any radiatively stable star. Thus, the pre-main-sequence object of interest is generating, at least initially, more luminosity than it can radiate into space. This imbalance cannot last indefinitely. Both the stellar interior and surface layers must change until $\langle L_{\rm crit} \rangle$ and L_* match. In other words, the star must undergo *thermal relaxation*.

The interior structure of any star is determined by its run of specific entropy, $s(M_r)$. We recall from Chapter 11 that the specific entropy added, layer by layer, to a protostar tends to increase with time, as a result of the accretion process itself. If deuterium fusion drives global convection, mixing by turbulent eddies enforces a spatially uniform $s(M_r)$, regardless of the star's accretion history. In a radiatively stable object, however, the rising entropy accurately reflects this buildup. There is no reason to expect that the resulting $s(M_r)$ should correspond, in a young pre-main-sequence object, to a distribution in *luminosity*, $L_{\rm int}(M_r)$, that joins smoothly onto the new surface value, L_*. Thermal relaxation alters $s(M_r)$ until the interior and surface luminosities are indeed consistent.

A specific example helps to illustrate the process. Figure 18.7 is taken from a numerical calculation, *i. e.*, a solution of the stellar structure equations with photospheric boundary conditions, as presented in § 16.2. The figure shows the evolution of both the specific entropy and luminosity within a pre-main-sequence star of $3.5\ M_\odot$. At $t = 0$, *i. e.*, just after the cessation of protostellar infall, $s(M_r)$ rises gently to $M_r = 2.8\ M_\odot$, after which it is constant. The exterior, flat portion of the profile occurs within a convective mantle created by the burning of residual deuterium. In the righthand panel, we show the corresponding luminosity profile. The function $L_{\rm int}(M_r)$ rises to a maximum near $60\ L_\odot$ and then falls steeply. The outer portion of this star is abnormally cold relative to a thermally relaxed object. Correspondingly, the opacity is too high, and the energy outflow blocked.

We can see, at least qualitatively, how this system must evolve by referring to the heat equation (16.6c). Within the region of rising interior luminosity, $\partial L_{\rm int}/\partial M_r > 0$. Thus, $\partial s/\partial t$ is negative, *i. e.*, the specific entropy falls with time. Conversely, the entropy must *increase* with time in the region beyond the luminosity peak. Heat is therefore transfered from the deep interior to more exterior layers. This outward transfer facilitates contraction of the interior. It also leads, however, to temporary *expansion* of the surface region. Clearly, the star's evolution is more complex than in a fully convective, low-mass object. There, the spatially uniform entropy declines globally, and the star undergoes homologous contraction (§ 16.2). The present evolution, driven both by internal transfer and surface loss of thermal energy, is distinctly *non-homologous*.

Returning to the lefthand panel of Figure 18.7, we see that the convective portion of the star eventually vanishes, while the entropy profile develops a very steep, exterior spike. Concurrently, the luminosity falloff in the outer layers becomes more shallow. For the last time shown, $L_{\rm int}(M_r)$ is a monotonically increasing function. At this point, $\partial s/\partial t < 0$ everywhere, and the star may be considered thermally relaxed.

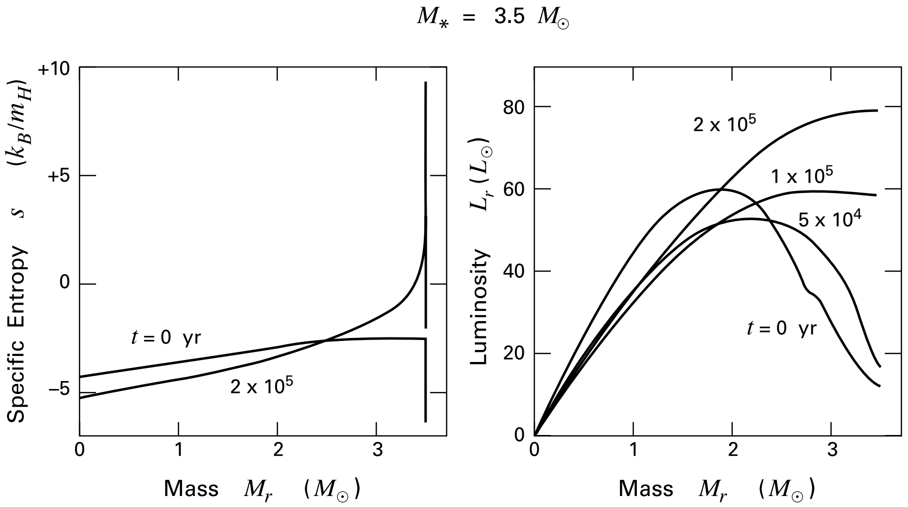


Figure 18.7 Thermal relaxation in a $3.5\ M_\odot$ star. The left panel shows the evolution of the specific entropy, while the right follows the internal luminosity. Both plots cover the first 2×10^5 yr of quasi-static contraction. The zero point of specific entropy is arbitrarily set at $T = 2.05 \times 10^5$ K, $\rho = 5.16\ \mathrm{g\,cm^{-3}}$.

18.2.2 Luminosity Jump

During thermal relaxation, the star's surface luminosity increases from the relatively low value imposed by protostellar initial conditions to the higher one consistent with energy transport in the deep interior. The net increase, as we see from the figure, is nearly an order of magnitude in the present example. We may describe this change as a *luminosity jump*. As the term indicates, the rise occurs rather quickly. The time scale for any bulk thermal change is the Kelvin-Helmholtz value, given in equation (16.1). However, t_{KH} overestimates the time interval if one substitutes for L_* the low, initial luminosity. The change actually occurs over the shorter Kelvin-Helmholtz time obtained by using $\langle L_{\mathrm{crit}} \rangle$. In our example, this latter interval is of order 10^5 yr.

Figure 18.8 shows the evolutionary path of the $3.5\ M_\odot$ star in the HR diagram. Because of the early luminosity jump and the concurrent swelling of the radius, the pre-main-sequence track looks very different from that of a solar-type object. The track first climbs steeply during thermal relaxation, and even temporarily crosses to the high-luminosity side of the birthline. The path then becomes shallower and more horizontal as the stellar radius begins to shrink. This final locus corresponds to the Henyey track in low-mass stars.

We also display in the same figure the pre-main-sequence tracks for other masses. A $2.0\ M_\odot$ star begins as a fully convective object. Hence, it starts to contract homologously, following a nearly vertical path in the HR diagram. Within 2×10^6 yr, however, the central opacity falls so low that a radiatively stable core develops. This core spreads with time, driving convection to the surface. Meanwhile, some of the heat expelled from the deep interior is absorbed by the retreating mantle, and the radius of the star swells. This swelling, and an accompanying

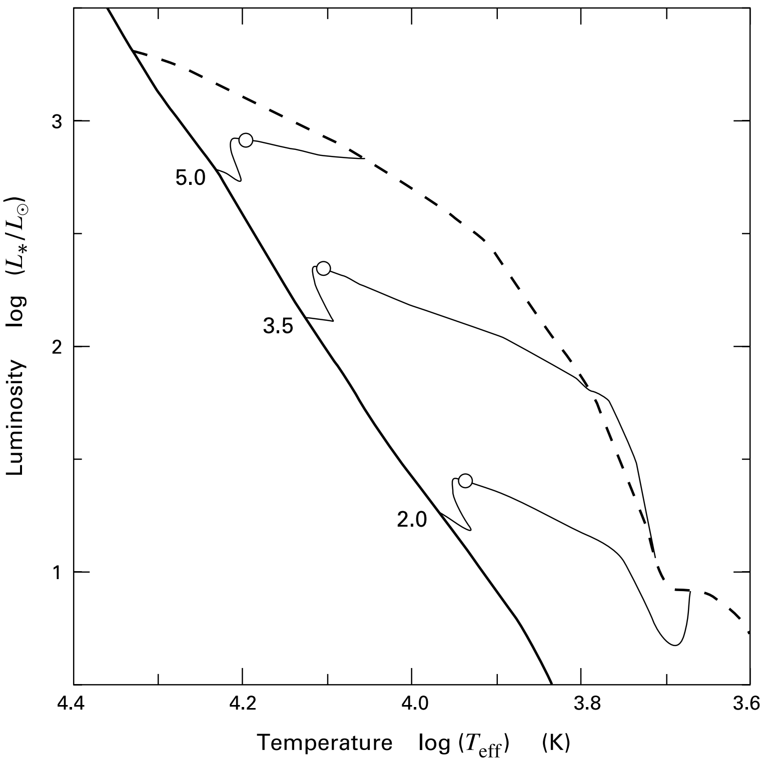


Figure 18.8 Evolution in the HR diagram of three intermediate-mass stars. The masses of these objects are given in solar units. Open circles mark the onset of central convection due to hydrogen fusion. Also shown are the stellar birthline and ZAMS.

luminosity rise, are more modest than in the $3.5\ M_\odot$ star. Once it is fully radiative, the object again follows a shallower track in the diagram.

A star of $5.0\ M_\odot$, also included in Figure 18.8, begins with no convection at all. During the protostar phase, an object of this mass derives its luminosity from rapid gravitational contraction. The same is true for the pre-main-sequence star shown here. The object is thermally relaxed, so that the specific entropy at every mass shell falls with time.[2] Correspondingly, the luminosity profile, $L_{\rm int}(M_r)$, is monotonically increasing at any epoch. Since the star is fully radiative, equation (16.14) tells us that the surface luminosity L_* rises gently during contraction, as $R_*^{-1/2}$.

Table 18.1 presents more detailed results for these same three masses. As in the analogous Table 16.2 for low-mass objects, the initial entry represents conditions at the birthline, while the final one corresponds in each case to the zero-age main sequence. We note immediately the

---

[2] Because the object in question was already contracting as a protostar, its initial entropy profile is *not* the one established through infall. This prior evolution explains how the pre-main-sequence star can be radiatively stable and yet avoid nonhomologous evolution.

much shorter evolutionary time scale for intermediate-mass stars. The main reason, of course, is that $t_{\rm KH}$ itself is less at these elevated luminosities. A second, and related, cause is that the initial, protostellar radii are even closer to main-sequence values. A star of 5 $M_\odot$ has barely any pre-main-sequence life, since a protostar of only slightly more mass would be burning hydrogen at its center while still accreting.

Table 18.1 Evolution of Intermediate-Mass, Pre–Main-Sequence Stars

$M_* = 2.0\ M_\odot$

Time (yr)	R_* ($R_\odot$)	log $T_{\rm eff}$ (K)	log L_* ($L_\odot$)	log T_c (K)	$L_{\rm nuc}/L_*$	$M_{\rm con}$ ($M_\odot$)
0	4.94	3.67	+0.90	6.54	0	1.60
1×10^5	4.14	3.67	+0.87	6.56	0	1.54
3×10^5	3.83	3.67	+0.81	6.58	0	1.35
1×10^6	3.18	3.68	+0.68	6.67	0	0.81
3×10^6	3.21	3.73	+0.86	6.85	0	0.05
8×10^6	1.77	3.94	+1.19	7.26	0.90	0.28

$M_* = 3.5\ M_\odot$

Time (yr)	R_* ($R_\odot$)	log $T_{\rm eff}$ (K)	log L_* ($L_\odot$)	log T_c (K)	$L_{\rm nuc}/L_*$	$M_{\rm con}$ ($M_\odot$)
0	4.23	3.71	+1.04	6.81	0	0.72
5×10^4	4.89	3.72	+1.21	6.81	0	0.17
1×10^5	7.15	3.77	+1.73	6.83	0	0
3×10^5	5.94	3.88	+2.03	6.96	0	0
5×10^5	4.44	3.99	+2.20	7.10	0	0
1×10^6	2.52	4.09	+2.11	7.32	0.90	0.72

$M_* = 5.0\ M_\odot$

Time (yr)	R_* ($R_\odot$)	log $T_{\rm eff}$ (K)	log L_* ($L_\odot$)	log T_c (K)	$L_{\rm nuc}/L_*$	$M_{\rm con}$ ($M_\odot$)
0	6.73	4.07	+2.89	7.08	0	0
1×10^4	5.93	4.08	+2.81	7.12	0	0
3×10^4	5.33	4.10	+2.80	7.16	0	0
5×10^4	4.88	4.13	+2.84	7.20	0	0
1×10^5	3.89	4.20	+2.91	7.32	0.03	0.04
2×10^5	3.09	4.18	+2.66	7.35	0.90	1.04

18.2.3 Nuclear Burning

Young stars of intermediate mass contain light elements other than hydrogen that are susceptible, at least in principle, to fusion during pre-main-sequence evolution. As we discussed in § 16.3, all stars start at the birthline with their full, interstellar supply of lithium. However, the convection zone in even a star of $1.2\ M_\odot$ is too thin for its base temperature ever to exceed the ignition value of 3×10^6 K. Herbig Ae/Be stars thus retain their initial lithium abundance as they evolve toward the main sequence.

The situation regarding deuterium is more involved. Returning momentarily to protostars, only their outer mantle contains the fuel at the time the radiative barrier is established, while the interior is soon thoroughly depleted (§ 11.4). Deuterium within the mantle burns, but is also resupplied continually through accretion. In fact, this convection zone retreats so quickly toward the surface that the net deuterium abundance first rises with increasing protostellar mass, then falls gradually to zero.

A pre-main-sequence star within a range of intermediate masses inherits the enriched mantle. Table 16.1 listed Δt_D, the subsequent time required to consume deuterium. Lower-mass Ae stars begin with such a small abundance that this time is negligible. However, the enhanced deuterium in the thin mantle of a $4\ M_\odot$ star yields a finite burning time. During this relatively brief interval, convection driven by the fusion reaction vanishes, as we have already seen. The luminosity yield from burning in both the radiative and convective outer layers is never comparable to L_*, and the process does not affect the star's mechanical evolution.

The contraction of intermediate-mass stars eventually initiates the fusion of ordinary hydrogen. As in lower-mass objects, the rate of nuclear energy generation increases rapidly with the central temperature T_c. For some T_c-value in excess of 10^7 K, this energy production becomes great enough to supply the full, interior luminosity, and contraction stops. The fusion of hydrogen to helium proceeds through the CN cycle, detailed in equation (11.46). Note again that these reactions are unimportant in solar-type stars, whose central temperatures are too low to overcome the strong Coulomb repulsion between protons and neavy nuclei such as ^{13}C.

Referring again to equation (11.46), we see that one of the intermediate products in the CN cycle is ^{15}N. Protons fusing with this nucleus usually create ^{12}C and an α particle, as indicated. Occasionally, however, ^{16}O is produced instead. The oxygen nucleus then further reacts with protons, initiating another cycle that eventually regenerates ^{15}N:

$$\begin{aligned}
{}^{15}\mathrm{N} + {}^{1}\mathrm{H} &\rightarrow {}^{16}\mathrm{O} + \gamma \\
{}^{16}\mathrm{O} + {}^{1}\mathrm{H} &\rightarrow {}^{17}\mathrm{F} + \gamma \\
{}^{17}\mathrm{F} &\rightarrow {}^{17}\mathrm{O} + \mathrm{e}^+ + \nu \\
{}^{17}\mathrm{O} + {}^{1}\mathrm{H} &\rightarrow {}^{14}\mathrm{N} + \alpha \\
{}^{14}\mathrm{N} + {}^{1}\mathrm{H} &\rightarrow {}^{15}\mathrm{O} + \gamma \\
{}^{15}\mathrm{O} &\rightarrow {}^{15}\mathrm{N} + \mathrm{e}^+ + \nu\ .
\end{aligned} \tag{18.1}$$

The net effect of the above cycle, like that of equation (11.46), is to consume four protons and thereby create ^{4}He. Note also that the last two reactions here are also contained within the other set. Thus, the two cycles are interlocked and constitute the CNO bi-cycle, as illustrated

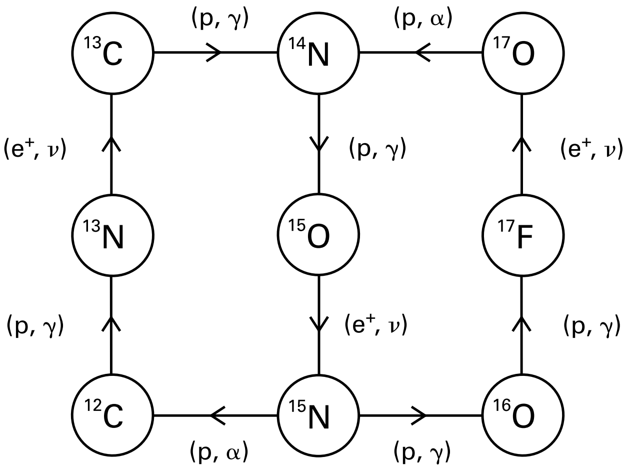


Figure 18.9 Coupled reactions of the CNO bi-cycle. Compare Figure 11.22, which shows only the lefthand cycle.

in Figure 18.9. Note that the probability of forming ^{16}O from ^{15}N is only about 10^{-3} that of forming ^{12}C. Hence the reactions of equation (18.1) contribute little to the star's total energy production. On the other hand, this side cycle *is* important for nucleosynthesis, being the chief source for such nuclei as ^{17}O.

Intermediate-mass objects on the main sequence lack the surface convection found in those of solar mass and below. Conversely, all stars with mass greater than about $1.5\ M_\odot$ have a *central* convection zone immediately surrounding the nuclear-burning core. The reason is that the CNO reactions are highly temperature sensitive. Thus, all the luminosity is generated very close to the center. According to equation (16.7), the temperature gradient $\partial T/\partial M_r$ rises steeply with diminishing radius r. Eventually, the specific entropy also begins to rise, the sign of convective instability. The birth of a central convection zone during pre-main-sequence contraction is marked by open circles along the three evolutionary paths of Figure 18.8. This event can also be seen in Table 18.1 as the final rise of the convective mass $M_{\rm con}$ in the models of 3.5 and $5.0\ M_\odot$. For the $2\ M_\odot$ star, the initial, surface convection does not vanish entirely until 3×10^6 yr, shortly before the appearance of the new, central zone.

18.2.4 Stellar Pulsations

Return once more to the earlier contraction of a Herbig Ae star, such as our $2\ M_\odot$ example. One of the subtler evolutionary effects concerns the mechanical stability of the object's outer layers. Over some fraction of the star's pre-main-sequence life, this region is subject to alternating expansion and contraction. Such *stellar pulsation* is well understood theoretically, as the same basic effect occurs in numerous other objects throughout the HR diagram.

Pulsation is manifested observationally through small, temporal variations of the stellar brightness. Figure 18.10 is a brief record of the U-band flux from the Herbig Ae star V351 Ori. This object, with an observed luminosity of $14\ L_\odot$ and effective temperature of 7400 K, has a theoretically derived mass of $1.8\ M_\odot$, at the low end of the Herbig Ae range. Over the interval covered in Figure 18.10, the U-magnitude increases and decreases periodically, but with an amplitude that also varies. The data are well reproduced by a combination of 5 sinusoidal

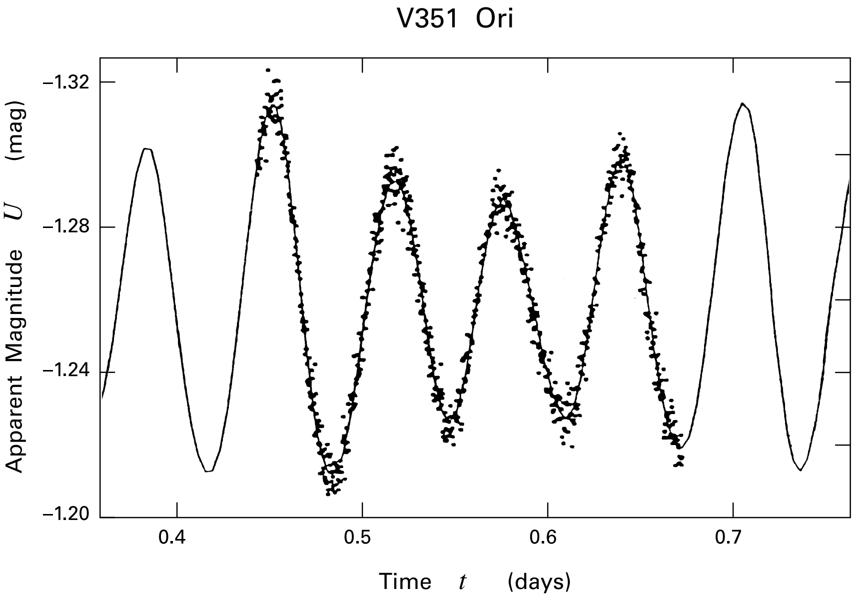


Figure 18.10 Oscillations of the Herbig Ae star V351 Ori. Shown as a function of time is the U-magnitude, relative to an arbitrary value. The data was recorded on a single night in December, 2000. The smooth curve is a best-fit combination of five frequencies.

oscillations. That with the largest amplitude has a period of 0.0841 days, or a little over 2 hours. It is gratifying that theory predicts the same period for the oscillation in a model of the appropriate mass and luminosity. Indeed, this agreement is strong confirmation of the underlying pre-main-sequence calculation.

How do pulsations arise in the first place? The key fact here is that, because of the star's T_{eff}-value, hydrogen and helium become ionized inside, but not *too* far inside, the photosphere. That is, ionization occurs in a region with sufficiently low density that it can undergo motion of appreciable amplitude. Imagine, then, that an outer shell of the star contracts slightly inward because of a pressure fluctuation. Surrounding layers squeeze the shell. This external work creates further ionization of hydrogen and helium without increasing the kinetic temperature of the gas. Since the density increases during contraction, so does the local opacity. As a result, more of the star's outgoing luminosity is absorbed, adding thermal energy to the compressed layer. This shell then expands more vigorously, until its density falls below the starting value. At this point, the opacity drops. The shell loses thermal energy as it collapses back toward its starting location, under the influence of stellar gravity. The region is thus executing oscillatory motion with increasing amplitude, a condition dubbed *overstability*.[3]

---

[3] In a linear, perturbational analysis, all physical variables are assumed to vary with time as $\exp(i\omega t)$. For overstability, ω is complex, so that the temporal function has both oscillatory and exponentially growing factors. Compare the simpler case of isothermal cloud dynamics in § 9.1.

This means of producing self-excited, radial pulsations through opacity variation is known as the *κ-mechanism.* Calculations show that the specific layer most responsible for driving the motion is that where helium becomes doubly ionized. The transition from He II to He III occurs at a temperature of 4×10^4 K. A glance at Figure G.2 in Appendix G shows that κ decreases with T in this regime more strongly than it increases with the density ρ. Slightly interior to the helium ionization zone, the temperature rises during a shell's contraction. The local opacity *falls*, and motion tends to be damped in this region. Indeed, such damping is what eventually limits the growing amplitude of radial excursions.

Thus far, our description applies to a single, normal mode of oscillation. If one such pattern can be excited in a given star, then others can, as well. Each normal mode has its own characteristic period. In the case of V351 Ori, the figure of 0.0841 days refers to the fundamental, *i. e.*, the longest period allowed. The observations show that additional modes are also present, but with lower amplitude. These are overtones and have periods shorter than the fundamental. Our particular star provides an especially clean example, where the full variation in flux can be reproduced simply by a combination of discrete modes. In other objects, the fluctuations are not as well understood. More complex, nonradial oscillations may be present, or the flux changes may arise in part from variations in circumstellar dust extinction.

18.2.5 The Instability Strip

We have emphasized that the self-excitation of a normal mode requires fortuitous conditions pertaining to ionization. In a star with the same mass as V351 Ori but a younger age, the surface temperature is lower. Thus, helium is ionized too deeply for oscillatory motion to occur. Conversely, ionization occurs at such a shallow depth in an older and hotter star that the outgoing luminosity cannot be absorbed during a shell's contraction. A star of fixed mass, therefore, is only susceptible to pulsation over a limited range of age and effective temperature. Within the HR diagram, the object must fall within the *pre-main-sequence instability strip.* Figure 18.11 displays this region, as ascertained through numerical calculations. Here, stellar models of various masses and ages were subjected to perturbations corresponding to the three lowest radial modes, and the oscillations were followed over many cycles to check for growth or decay. Note that the luminosity and temperature of V351 Ori place it squarely in the middle of the region, as also shown in the figure.

The lefthand edge of the instability strip touches the birthline at a $T_{\rm eff}$-value of 7100 K, corresponding to an early-F spectral type. Because the strip veers to higher $T_{\rm eff}$ with decreasing luminosity, most pulsating objects should be Ae stars, a prediction in accord with the dozen or so examples currently known. Depending on its mass, a given star only spends from 5 to 10 percent of its pre-main-sequence lifetime within the strip. Thus, the potential number of candidate sources is rather limited. Many more pulsating stars, however, are also observed in just this region of the HR diagram. These are older objects, also of intermediate mass, that are either still in, or have recently evolved past, their main-sequence phase. Both the photometric oscillation amplitudes and periods of these *δ Scuti variables* are similar to those in the relevant subclass of Herbig stars. We may thus consider the latter as another instance of the well-documented, δ Scuti-type variation.

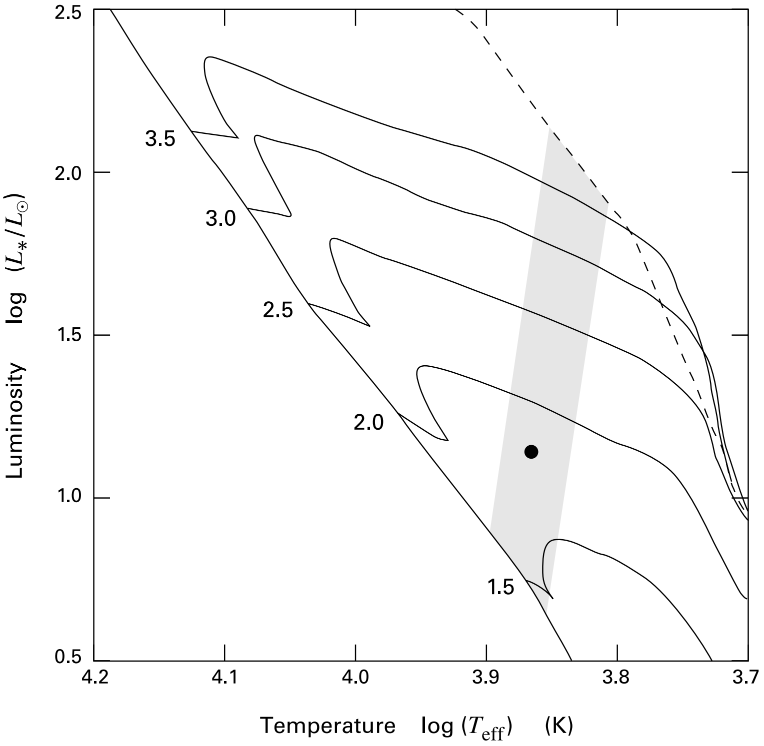


Figure 18.11 The pre-main-sequence instability strip in the HR diagram. Stars within this shaded region are subject to self-excited, radial pulsations. Also shown are the birthline, ZAMS, and evolutionary tracks for the indicated masses, in solar units. The filled circle represents V351 Ori.

A point made earlier bears some elaboration. Despite the agreement between the predicted fundamental period for V351 Ori and the observed one, the current photometric data on this and other young pulsators are rather sparse, making it generally difficult to identify modes with confidence. The situation will improve dramatically when satellites begin monitoring both photometric and spectroscopic variations. More precise and extensive observations will not only pinpoint the masses of stars that may not be components of binary pairs, but will also act as probes of the objects' internal structure, just as helioseismology currently does for the Sun.

As we leave the topic of stellar pulsation, let us broaden our perspective to note that the shaded area demarcated in Figure 18.11 is actually the lowest portion of a much larger instability strip. Stars with effective temperatures similar to those of interest here, but with luminosities from 500 to $3 \times 10^4\ L_\odot$, constitute the *Cepheid variables*. These are evolved supergiants that are either burning hydrogen in a thin shell or helium in their central cores. A given star can pass through the instability strip many times as its post-main-sequence evolutionary path loops around in the HR diagram. The period for radial pulsations, which are again driven by the κ-mechanism, is greater by an order of magnitude than for δ Scuti variables. Moreover, this period is systematically longer for stars of higher average luminosity. Historically, this last fact is the most important of all, since it enabled observers to gauge the absolute magnitude of a pulsating star, and thereby measure its distance. It is through this means that both the structure of our own Galaxy and the location of analogous, neighboring systems were first determined.

18.3 Thermal and Mechanical Effects

All Herbig Ae/Be stars exhibit, to varying degrees, both optical reddening and excess radiation at longer wavelengths, above that emitted by a bare photosphere. The dust grains responsible for these two effects are located in the sparse medium between the source and Earth, as well as in relatively dense configurations surrounding the stars themselves. To analyze the situation, we proceed in the same manner as for T Tauri stars. That is, we first obtain the spectral energy distribution for each source, properly augmenting the flux in each waveband to compensate for interstellar dust extinction. As we now see, the results point to key differences between Ae/Be stars and their low-mass counterparts.

18.3.1 Infrared Excess

The variety of spectral energy distributions actually observed divides neatly into three categories. These are displayed, through representative examples, in Figure 18.12. The figure also shows, as dashed curves, the expected flux at each wavelength from the stellar photospheres. Construction of these curves requires, of course, a prior determination of each star's spectral type. This information was also needed when comparing observed and intrinsic colors, *i. e.*, when dereddening the full spectral energy distribution. It is an important feature of Ae/Be stars, already mentioned in § 18.1, that their visible absorption lines are *not* partially filled in by continuum emission. Thus, the spectral type and the star's own flux contribution are relatively secure.

In AB Aur, the A0 star shown in the top panel, the flux climbs above stellar values for wavelengths exceeding 1 µm. On the other hand, λF_λ still declines with λ beyond the near-infrared. This so-called "Group I" spectral energy distribution characterizes the majority of Herbig Ae/Be stars. Perhaps one fourth of observed sources fall into Group II, exemplified by LkHα 198 in the middle panel. Here, λF_λ is flat or rising, at least between 3 and 10 µm. In the remaining, Group III sources, λF_λ not only declines sharply in this range, but the full, dereddened spectral energy distribution is barely distinguishable from that of the star alone. Our example here is the B3 star BD+65°1637, previously shown in Figure 18.2. One sees only a hint of excess emission at the longest wavelengths.

It is tempting to link Group III stars with the Class III subset of lower-luminosity sources. We already know that BD+65°1637, in particular, has cleared out much of its immediate environment. The minor infrared excess that remains plausibly stems from free-free emission by hot, circumstellar gas, as in classical Be stars. This lack of a dusty environment is indeed reminiscent of the Class III situation. On the other hand, BD+65°1637, like all Herbig Ae/Be stars, produces Hα in emission, while Class III objects are either weak-lined or post-T Tauri stars, without significant flux in the line. Whatever the origin of Hα emission in our B3 star, it is assuredly *not* tied to the presence of a disk, for which there is no evidence.

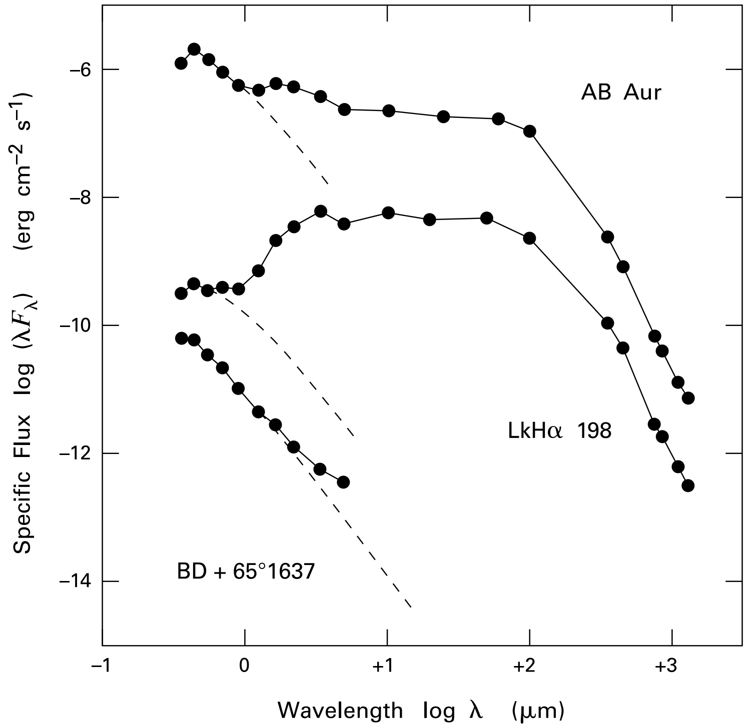


Figure 18.12 Spectral energy distributions for three representative Herbig Ae/Be stars. The plot for AB Aur has been shifted upward by $\Delta \log(\lambda F_\lambda) = +1.5$, while that for BD+65°1637 has been lowered by $\Delta \log(\lambda F_\lambda) = -2.5$. Dashed curves are blackbodies at the corresponding effective temperatures.

At the other extreme, Group II objects, since they exhibit the largest infrared excess, should also be the most deeply embedded. This expectation is confirmed through molecular line observations. The dense tracers CS and HCN are detected in a region spanning 10^4 AU around our sample star LkHα 198. Heated dust within this envelope, and perhaps also from an interior disk, must supply the copious infrared emission. The closest analog in the low-luminosity regime is Class I sources. We stress, however, that Group II stars are still optically visible, despite rather high values of A_V.

Most detailed studies have focused on the relatively populous Group I sources. Referring again to Figure 18.12, the declining λF_λ in AB Aur in the mid- and far-infrared recalls Class II T Tauri spectra, for which we ascribed disks with power-law temperature profiles (§ 17.3). The near-infrared dip between 1 and 3 μm may similarly be explicable by a central hole in the disk, as in the T Tauri star GM Aur. A closer look, however, reveals problems with this interpretation. In all Group I sources, the infrared excess is so large that it rivals the stellar luminosity (For AB Aur, $L_{\rm IR}/L_* = 0.48$.) A flat, passive disk is thus ruled out. A flaring structure would intercept more stellar radiation and hence give a larger ratio $L_{\rm IR}/L_*$. However, the resulting spectral energy distribution peaks at a longer wavelength than observed.

One might argue that the disk is self-luminous, *i. e.*, that it somehow releases sufficient energy to allow massive accretion onto the central star. We have seen, however, that Ae/Be stars lack the optical continuum excess created by shock-generating accretion flows. More generally, we recall that disks are invoked for T Tauri spectra because a more isotropic distribution of dust would obscure the central object at visible wavelengths. However, if the luminosity from disk accretion were to approach that from the star, we would have

$$\frac{G\,M_*\,\dot{M}_d}{R_*} \approx L_* \ , \tag{18.2}$$

where $\dot{M}_d$ represents the rate of mass transport from the disk to the stellar surface. Here we have utilized the fact that most of the energy release occurs close to the star, because of its deepening potential well. We also know that M_* exceeds the disk mass M_d, so we find

$$\frac{M_d}{\dot{M}_d} < \frac{G\,M_*^2}{R_*\,L_*} \ . \tag{18.3}$$

The expression on the right is the Kelvin-Helmholtz time for the star. Evidently, our self-luminous disk would drain itself in a time short compared to that over which the star evolves. The disk must continually be replenished, presumably by matter from a more extended, infalling envelope. We are thus led back to the very structure we sought initially to avoid. In summary, neither active nor passive disks seem able to explain the Group I infrared excess.

18.3.2 Dusty Envelopes

The resolution of this paradox is twofold. First, the bulk of excess emission in Group I sources must not originate in a self-luminous disk, but in some more extended structure. Second, this dusty configuration must have sufficiently low column density to leave the central star optically visible. Here, we are aided by the larger intrinsic luminosities of Herbig Ae/Be stars. But it is clear that the circumstellar matter cannot resemble the infalling envelope we described for spherical protostars in Chapter 11; the resulting densities are simply too large. A sparser structure is called for; one with little or no associated infall.

Theorists have offered a number of detailed prescriptions for spherical envelopes around Herbig Ae/Be stars, all designed to match observed spectral energy distributions. In some models, the envelope surrounds a much smaller, interior disk. The latter is invoked to account for the flux at submillimeter wavelengths, which is difficult to produce in a dust shell. Others account for this long-wavelength component by diminishing β, the exponent characterizing the falloff in grain opacity with wavelength. There is evidence, from intercomparison of color excesses, that the grains surrounding Ae/Be stars are larger than in the general interstellar medium. Grain coagulation would indeed create a relatively shallow decline in κ_λ at large λ.

It should be clear from our brief description that there is considerable latitude in building models of this kind. Even within the arbitrarily restrictive framework of a perfectly spherical envelope, one may freely alter not only the spatial distribution of dust, but also the composition and size of the grains themselves. This freedom, of course, reflects our ignorance of the true circumstellar environment. One may at least hope to discern common features of those models that successfully reproduce observed spectra. We find, as anticipated, that the average density is lower than in canonical protostellar envelopes, but also higher than that of a cloud core forming a T Tauri star. A figure of $n_H \sim 10^5 \text{ cm}^{-3}$ is representative, with an associated linear scale of 10^3 AU. These numbers refer either to the inner, heated portions of an individual cloud fragment that spawned the star, or to an entity formed through coalescense of preexisting cores.

The inner hole required in disk models finds an analog within the envelope picture. As we saw when discussing protostellar infall, dust approaching too close to the central object becomes hot enough to sublimate. Thus, there is indeed a hole in the dust distribution, regardless of details of the model. On the other hand, it is not obvious that there should be a corresponding absence of gas. Even a relatively modest gas density could produce appreciable radiation when bathed by the star's copious ultraviolet photons. Earlier-type Herbig Be stars should create an interior HII region. Photons leaving this region have energies below 13.6 eV, but still ionize carbon and other elements farther out (see below). Free-free and bound-free transitions in these species contribute to the near-infrared spectral energy distribution.

Direct imaging of Herbig Ae/Be environments has been achieved in some cases. On a scale of 0.1 pc, optical reflection nebulae and the single-dish radio observations cited in § 18.1 reveal a clumpy structure. Mapping of CO lines also gives asymmetric contours, centered on the star but elongated. We shall return to these latter observations in the next section, when we summarize evidence for circumstellar disks. Within the same size range, from 10^2 to 10^3 AU, interferometry shows halos of scattered infrared light, particularly around more embedded objects like LkHα 198. These structures may persist quite close to the star itself, just as the theoretical dust shell models indicate. Our prototypical Group I object, AB Aur, scatters near-infrared light in a roughly symmetric region spanning 0.6 AU. Note that few T Tauri stars have shown similar, direct evidence for truly three-dimensional distributions of dust.

Circumstellar grains not only scatter continuum radiation, but also produce their own spectral lines. Aircraft and satellite observations of Herbig Ae/Be stars in the mid-infrared reveal sharp emission features. Lines such as those at 3.3 and 6.2 μm are signatures of PAHs, the very tiny grains of nearly molecular size. This emission is more prominent near Be stars, which supply more of the ultraviolet photons that excite the transitions (§ 2.4).

The best-studied spectral line is the broad feature near 10 μm, associated with the SiO bond in silicate grains. Recall that deeply embedded, low-mass stars have a strong absorption dip in this wavelength regime (see Figure 11.24). Optically revealed T Tauri stars usually exhibit emission at 10 μm, but a significant fraction show the feature in absorption. Those for which the absorption is especially deep also have the highest polarization in their optical light. A plausible explanation for this trend is that the silicates reside mainly in circumstellar disks. When these are viewed nearly edge-on, infrared continuum light from near the star is strongly absorbed, leading to a pronounced dip in the spectral energy distribution. Any optical light received is scattered off material above and below the disk surface, resulting in polarization.

Many Herbig Ae/Be stars also exhibit a spectral signature at 10 μm. Again, one sees the line both in emission and absorption. Those stars with absorption dips are often optically polarized. However, there is no longer a correlation between the degree of polarization and the depth of the dust feature. The inference is again that circumstellar grains are not wholly confined to a flattened configuration. On the other hand, the very fact of polarization demonstrates that the stellar environment, at least that part responsible for scattering optical photons, cannot be spherically symmetric. We shall return to this point, as well as to the 10 μm feature, in § 18.4.

18.3.3 Photodissociation Regions

The cloud matter surrounding some Herbig Be stars glows brightly in the far-infrared. This luminosity is much too great to represent scattered emission from the star itself. Instead, the radiation must stem from heated dust grains. Earlier, we noted that far-ultraviolet photons escaping an HII region can still ionize carbon and other elements, thereby contributing to the observed near-infrared component of the spectral energy distribution. We recall from Chapter 8 that these same photons also heat both the dust and gas, the latter through the grain photoelectric effect. In other words, Herbig Be stars create photodissociation regions, just like their main-sequence counterparts.

While most of the emitted radiation is in the far-infrared continuum, there is also substantial flux in atomic forbidden lines, especially [O I] 63 μm and [C II] 158 μm. Consider again the NGC 7129 region shown in Figure 18.2. Satellite studies find emission in both lines peaking toward the embedded B6 star LkHα 234. There is also more diffuse radiation surrounding BD$+65^\circ 1637$. It appears, in fact, that there are *two* photodissociation regions, one associated with each star. That created by LkHα 234 has a 63 μm flux higher by about a factor of two than in 158 μm. Accounting for self-absorption by cold O I along the line of sight, the intrinsic ratio $\Lambda_{\mathrm{OI}}/\Lambda_{\mathrm{CII}}$ is even greater. In § 8.3, we learned that this ratio becomes elevated in denser cloud material. In fact, the derived n_H from heated gas near LkHα 234 is of order 10^5 cm^{-3}, with a far-ultraviolet enhancement factor of $G_\circ \sim 100$. The photodissociation region created by BD$+65^\circ 1637$ lies in the molecular ridge just to the left of the evacuated cavity. Here n_H is lower by a factor of 100.

The denser and more compact photodissociation region also emits molecular lines in the far-infrared regime. Satellite observations have detected various rotational transitions of CO. Here, the upper J-values are very high, from 14 to 19. Energizing these levels requires a relatively nearby star's ultraviolet radiation bathing the cloud medium. The alternative mechanism, shock heating from a stellar wind, is ruled out observationally. In both C- and J-type shocks, a principal coolant is H_2O formed downstream in the postshock gas. Some H_2O masers are actually seen near LkHα 234, demonstrating that a wind is indeed present. Nevertheless, the total flux in H_2O lines is well below that predicted by shock models.

Far-ultraviolet photons entering a photodissociation region destroy all molecular hydrogen within a substantial volume. Those with energy between 11.2 and 13.6 eV excite the molecule to higher electronic states, from which it may fall apart into isolated atoms (§ 5.2). Viewing the matter globally, a fraction $\langle\beta_i\rangle \approx 0.15$ of all emitted photons in this energy range destroy a hydrogen molecule somewhere within the region. The other photons are reemitted at longer wavelengths and escape. In steady state, each dissociation is balanced locally by the creation of H_2 from a pair of hydrogen atoms. Thus, we may determine the radius R_{H_2} of dissociated H_2 in essentially the same way we obtained the Strömgren radius marking the boundary of ionized hydrogen.

We recall that H_2 forms through the recombination of atoms on grain surfaces. The volumetric formation rate is given by equation (5.10):

$$\mathcal{R}_{\mathrm{H}_2} = \frac{1}{2}\,\gamma_H\, n_d\, \sigma_d\, n_{\mathrm{HI}}\, V_{\mathrm{therm}} \; . \tag{18.4}$$

Here, n_d and σ_d are the grain number density and geometric cross section, respectively, while γ_H is the probability that each hydrogen atom, traveling at the thermal speed V_{therm}, sticks to the grain after collision. To pursue the analogy with HII regions, we first rewrite $n_d\,\sigma_d$ as $n_H\,\Sigma_d$, following equation (2.42). After further equating n_H to n_{HI}, we have

$$\mathcal{R}_{\mathrm{H}_2} = \alpha_{\mathrm{rec}}(\mathrm{HI})\, n_{\mathrm{HI}}^2 \; . \tag{18.5}$$

The recombination coefficient for HI gas is

$$\alpha_{\mathrm{rec}}(\mathrm{HI}) \equiv \frac{1}{2}\,\gamma_H\, \Sigma_d\, V_{\mathrm{therm}} \; , \tag{18.6}$$

and has the numerical value $3.3 \times 10^{-13}\ \mathrm{cm}^{-3}\,\mathrm{s}^{-1}$ at the representative temperature of 500 K for the photodissociation region.

Let us denote by $\mathcal{N}_{\mathrm{diss}}$ the rate at which the central star emits photons between 11.2 and 13.6 eV. Then the radius of the HI region is given by an expression analogous to equation (15.3):

$$\begin{aligned} R_{\mathrm{HI}} &= \left[\frac{3\,\langle\beta_i\rangle\,\mathcal{N}_{\mathrm{diss}}}{4\,\pi\,\alpha_{\mathrm{rec}}(\mathrm{HI})\, n_H^2}\right]^{1/3} \\ &= 0.01\ \mathrm{pc}\left(\frac{\mathcal{N}_{\mathrm{diss}}}{10^{45}\ \mathrm{s}^{-1}}\right)^{1/3}\left(\frac{n_{\mathrm{H}_2}}{10^5\ \mathrm{cm}^{-3}}\right)^{-2/3} \; . \end{aligned} \tag{18.7}$$

In the second form of this equation, we have used representative values of $\mathcal{N}_{\mathrm{diss}}$ and the surrounding density n_{H_2} that are appropriate for LkHα 234.

Although the typical size of the HI region would appear to be quite small, one should bear in mind that the corresponding Strömgren radius, for the same star immersed within the same cloud, is smaller yet. Within equation (18.7), the coefficient α_{rec}(HI) is four orders of magnitude less than α'_{rec} in equation (15.3), while $\mathcal{N}_{\mathrm{diss}}$ always exceeds the ionizing flux from the same star. For a B6 star like LkHα 234, this flux ratio is 2×10^4. A more accurate evaluation of R_{HI} accounts for partial absorption of outward photons by dust grains and lowers the numerical coefficient in equation (18.7) by a factor of 2 at the B6 spectral type.

The HI region formed around a Herbig Ae/Be star is initially overpressured with respect to its surroundings. The dissociated volume thus expands into the cloud medium. This expansion, in the case of LkHα 234, has presumably been halted by ambient pressure. In contrast, most of the cloud material around BD+65°1637 has been driven outward, leaving only a relatively thin ridge of heated and dissociated gas.

18.3.4 Winds

Herbig Ae/Be stars also disrupt their surroundings through the action of strong winds. As in their low-mass counterparts, the primary evidence for winds is in profiles of emission lines.

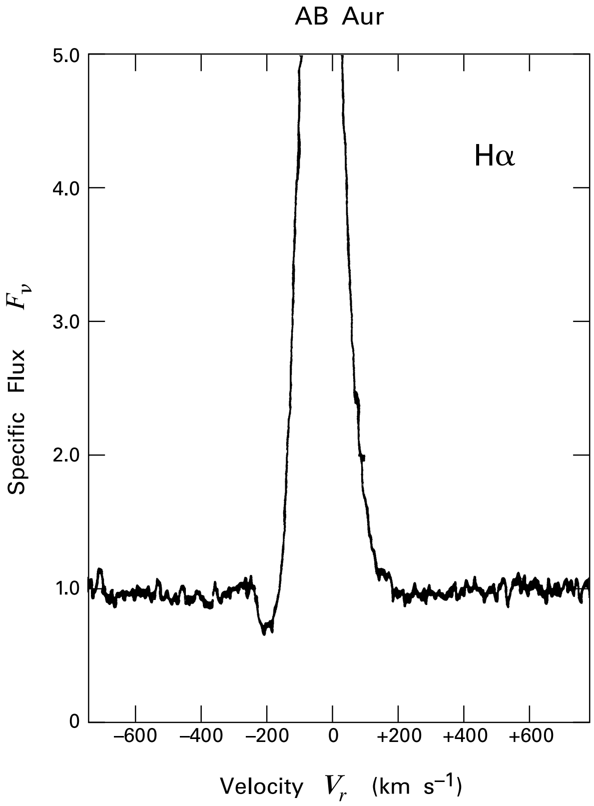


Figure 18.13 Observed Hα profile of the Herbig Ae star AB Aur. Specific flux is measured relative to the adjacent continuum level.

Consider the strongest line of all, that of Hα. The detailed profile of a given star can change erratically over a very short time. Nevertheless, these variations generally occur within a well-defined morphological type. Most Hα profiles are double-peaked. As in T Tauri stars, the central absorption feature is usually blueshifted by a modest velocity. (Recall the case of CI Tau, seen in Figure 17.8.) There are fewer symmetric examples, analogous to FM Tau. However, a fraction of stars exhibit true P Cygni profiles, in which the blueshifted absorption component dips below the continuum. In summary, most Herbig Ae/Be stars show evidence, through Hα, of material receding from the central object, *i. e.*, a wind.

Figure 18.13 displays the Hα emission from AB Aur. At this epoch, the profile is unambiguously of the P Cygni type. The minimum in the blueshifted absorption dip occurs at a radial velocity of $V_r = -200 \text{ km s}^{-1}$. This figure must represent the speed of relatively cold material moving toward us against a background of hotter gas emitted by shocks outside the star. As in T Tauri stars, the large overall breadth of the Hα profile indicates that the motion of the shocked gas is complex, with velocity vectors spanning a range of orientations.

Nevertheless, there are other indications that the wind eventually attains a well-ordered, anisotropic structure. Stars like AB Aur, with P Cygni profiles in Hα, also exhibit emission in other lines traditionally associated with chromospheric activity, including the Mg II doublet near 2800 Å. Figure 18.14 shows the flux in this spectral region, as detected via satellite, on two

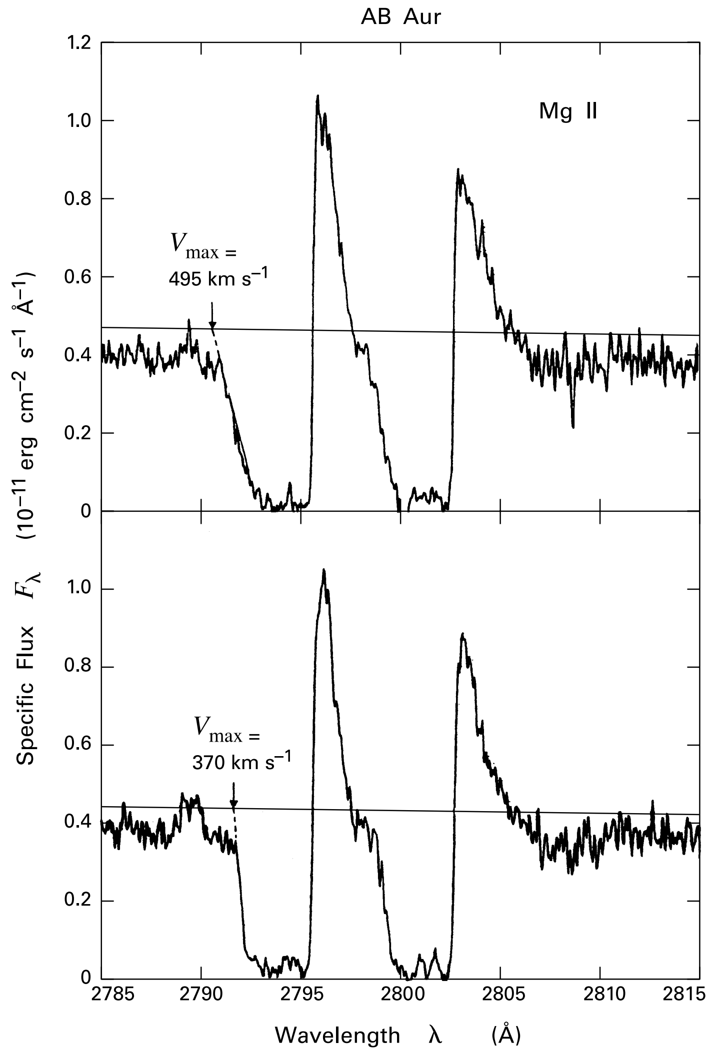


Figure 18.14 Profiles of the Mg II h and k lines in AB Aur, taken on two successive nights. The continuum level is shown in each panel. Also indicated are the values of $V_{\rm max}$, the edge velocity in the absorption dip of the Mg II k line.

successive days. Both the Mg II k line (centered at 2796 Å) and Mg II h (2803 Å) again have P Cygni profiles, with strong absorption components. It is also clear that the detailed profiles change with time, even over this brief interval. In particular, the blueward edge of the Mg II k absorption trough has shifted to a less negative velocity.

The absorption edge traces the highest-speed material present in the wind at any time. If we plot this velocity, $V_{\rm max}$, over a more extended time, we find that it varies *sinusoidally*, as

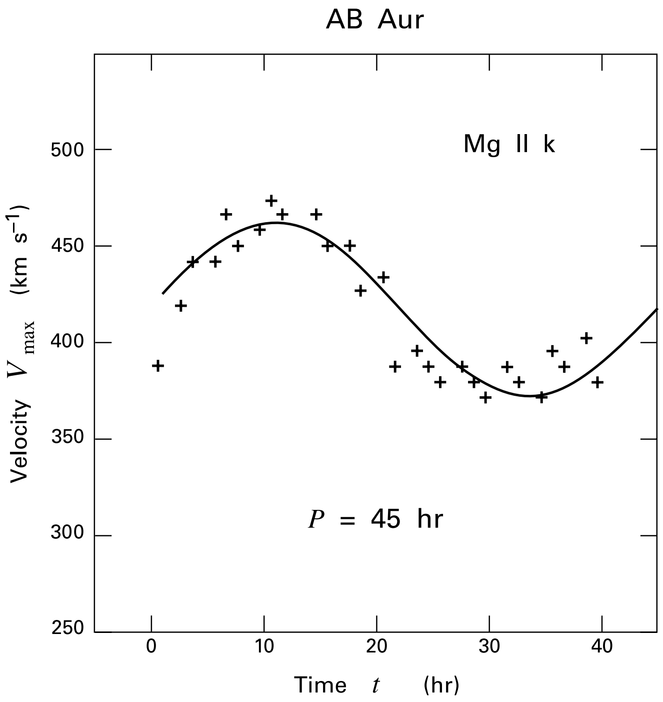


Figure 18.15 Temporal variation of $V_{\rm max}$, the edge velocity of the Mg II k absorption dip in AB Aur. The smooth curve is the best-fitting sinusoid, with indicated period.

seen in Figure 18.15. The derived period is 45 h. Note that the equatorial velocity of the star, as obtained from the broadening of optical absorption lines, is $V_{\rm eq}$ sin $i = 75\,{\rm km\,s^{-1}}$. The object's effective temperature and luminosity imply a radius of $3\ R_\odot$. Thus, the maximum rotation period, corresponding to an inclination $i = 90°$, is 49 h.

It therefore appears that the wind from AB Aur is being modulated in step with the star's rotation. The outflowing gas, at the point where Mg II absorption occurs, is clearly not spherically symmetric, but is somehow tracking the conditions with which it was launched at the stellar surface. One intriguing possibility is that the wind consists of high- and low-velocity streams, each emitted from a different portion of the surface. The velocity $V_{\rm max}$ then peaks whenever the high-speed stream is directed toward the observer.

A handful of Ae/Be stars exhibit a wind signature in the ultraviolet: redshifted or blueshifted emission in Lyα at 1216 Å. There is also indirect evidence. Some sources appear to be driving CO outflows. Figure 16.2, which displayed a heterogeneous collection of intermediate-mass stars in the HR diagram, singled out those with outflows. We noted that this subclass is situated especially high up in the diagram, just below the birthline. We also cautioned that, at least in some instances, the star actually powering the fast-moving CO may be an unresolved neighbor or binary companion.

Such potential confusion of sources is a general problem for Herbig Ae/Be stars, arising both from their location in crowded regions and from their greater distance. The same issue enters any discussion of jets. Several dozen Ae/Be stars have, at one time or another, been identified as the sources of either continuous jets or chains of discrete, Herbig-Haro knots. An additional difficulty here is that the intermediate-mass object may create a reflection nebula whose continuum emission overwhelms the relatively weak jet. In any case, higher-resolution imaging has sometimes led to retraction of the original claim. A good example is the Herbig Be star LkHα 234 (Figure 18.2), which appears, from optical images, to drive a jet down the central axis of a broad reflection nebula. Additional, millimeter and mid-infrared studies show that the source of this jet, and of an oppositely directed CO outflow lobe, is more likely to be an even more embedded object located nearby.

The situation regarding molecular outflows and jets is thus less settled than one might hope. Nevertheless, the spectroscopic evidence, both from Hα and other lines, leaves no doubt that Ae/Be stars do power winds. But this fact itself presents a major conundrum. Our theoretical account of pre-main-sequence winds in § 13.4 focused on the mechanism of centrifugal ejection from magnetized objects. This picture, in turn, is motivated by our knowledge that T Tauri stars are at least partially convective and hence generate magnetic fields through dynamo activity. The situation is entirely different for intermediate-mass objects. Apart from transient effects during their thermal relaxation, these stars are stable against convection. Any surface magnetic fields that exist must have a different origin.

There is, in fact, scant empirical evidence for magnetic fields associated with Herbig Ae/Be stars. Measurements of Zeeman broadening have thus far been inconclusive. It is true that a dozen or so objects have radio flux at centimeter wavelengths. The emission, however, is generally unpolarized and appears to be thermal, presumably arising from the ionized component of the wind. Again, the situation is unlike that for T Tauri stars such as V773 Tau, whose polarized, compact radio emission indicates the presence of an underlying field (§ 17.1).

What, then, powers Herbig Ae/Be winds? Our previous, general discussion of wind origins further sharpens the dilemma. The earliest-type Herbig Be stars still have too small a luminosity to drive winds through radiative acceleration. Furthermore, even the crudest estimates of mass loss from Ae/Be stars (typically, $\dot{M} \lesssim 10^{-7}\ M_\odot\ \mathrm{yr}^{-1}$) show that the wind cannot be driven by thermal pressure. The associated corona would be readily detectable through its powerful X-rays, contrary to observation. Speculations about the other driver traditionally considered, Alfvén-wave pressure, would seem premature in the absence of good evidence for surface fields. For the time being, the origin of Ae/Be winds remains a mystery.

18.3.5 X-Ray Emission

Although X-rays are too weak to help explain the wind origin, this emission component does exist, with characteristics that are gradually becoming clear. A succession of spaceborne instruments, beginning with the Einstein satellite, have found X-rays in about a dozen stars, roughly half of the observed sample. While such numbers make generalizations difficult, the data reveal a luminosity L_X that is a fraction 10^{-6} to 10^{-5} of the corresponding stellar value.

We may place this result into context by first recalling that L_X/L_* for T Tauri stars tends to be greater, of order 10^{-4}. On the other hand, O and early-B stars, which also display X-rays, typically have $L_X/L_* \sim 10^{-7}$. The physical origin of the emission in the two groups is also very different. In T Tauri stars, X-rays plausibly arise from reconnection of dynamo-generated magnetic loops near the stellar surface (§ 17.1). Massive stars, which lack outer convection zones, create this emission through shocks generated by instabilities within the stellar wind (§ 15.3).

Which of these two explanations, if either, applies to Herbig Ae/Be stars? Our question assumes, of course, that the X-rays do not stem from an unseen, lower-mass companion, a nontrivial assertion. Some Herbig Be stars have an associated L_X that would be too large for a T Tauri star. For the others, the attribution remains uncertain. In any event, a solar-type dynamo mechanism is inapplicable to radiatively stable objects. We are left to consider the wind hypothesis. Until the origin and configuration of the flow itself is better understood, we cannot gauge the likelihood of internal shocks, arising either from the wind itself or from impact with stationary matter. Note that the temperature required for the X-rays ($T \sim 10^7$ K) can indeed be produced, at least in principle, by a shock speed comparable to the highest observed wind speeds, about 500 km s^{-1}. As the star ages and its wind dies away, so does the X-ray flux. This picture is at least consistent with the fact that L_X falls to undetectable levels in main-sequence stars with spectral types from mid B to early F.

18.4 Gaseous and Debris Disks

We have seen the difficulty of invoking thermal disk emission to account for the infrared excess seen in most Herbig Ae/Be stars. Nevertheless, such entities must exist, if only because we witness their remnants – flattened configurations of orbiting dust – around older objects in the same mass range. Let us now review more systematically the observational study of disks around Ae/Be stars, including the evolutionary development of these structures.

18.4.1 Spectral Line Studies

The spectral energy plots for AB Aur and LkHα 198 in Figure 18.12 include significant components in the millimeter regime. Historically, strong continuum flux at $\lambda \gtrsim 1\ \mathrm{mm}$ was used to bolster the idea that a star is surrounded by a dusty disk. The basic argument is the familiar one that a spherical distribution of grains with this amount of long-wavelength emission would obscure the object completely in the visual. As already noted, however, the shell models themselves can accommodate a diversity of observations through appropriate changes in volume density and internal composition of the grains. Additionally, direct imaging fails to resolve the continuum emission from relatively unobscured objects like AB Aur. More embedded stars, such as LkHα 198 and LkHα 234, do have spatially extended emission, but this could arise from warmed dust in their clumpy environments.

Stronger evidence for disks is potentially available from spectral lines originating in the gas. Here, one hopes to image flattened structures that are manifestly rotating about their central stars, as seen through Doppler shifts of the line in question. This hope has been realized to some degree, although the case for disks remains weaker than in T Tauri stars. Figure 18.16 provides two examples of interferometric maps utilizing CO. We also show, for comparison, the telescope beam in each case. On the left is the gaseous structure around AB Aur, as seen in the ^{13}C^{16}O $(J = 1 \rightarrow 0)$ transition. While the configuration is clearly asymmetric, its true

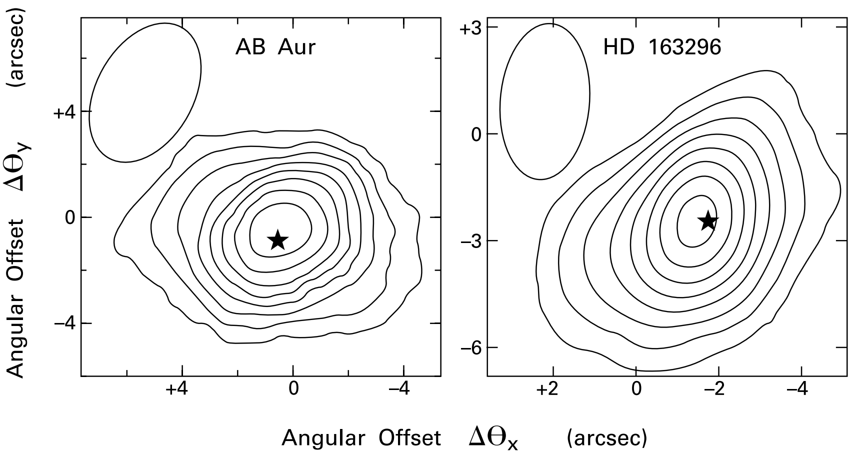


Figure 18.16 Interferometric images of ^{13}C^{16}O $(J = 1 \rightarrow 0)$ emission around the stars AB Aur (*left*) and HD 163296 (*right*). Angular offsets are relative to the phase center of each field, which does not coincide precisely with the stellar position, shown here symbolically. Also shown in each case is the beam size and orientation.

shape is hard to discern, as the minor axis is only marginally resolved. The mean, *i. e.*, intensity-weighted, radial velocity exhibits a smooth gradient along the major axis, with a total range of $2\ \mathrm{km\ s^{-1}}$. Whether or not this variation represents Keplerian motion is still unclear. In the right panel, we show gas around the star HD 163296, now mapped in the ^{12}C^{16}O $(J = 2 \rightarrow 1)$ line at 1.3 mm. Here, the full structure is better resolved, but it is also more irregular in shape. The mean radial velocity again changes along the major axis.

Both AB Aur and HD 163296 have spectral types of early A. It is significant that nearly all the objects with excess millimeter emission, whether in the continuum or lines, are either Ae or late-Be stars. Conversely, stars whose spectral type is earlier than about B5 rarely have significant flux in this regime. It is true that more massive stars tend to be farther away, but the higher millimeter luminosities already seen would be detectable. The implication is that the circumstellar gas and dust being sampled by these observations is absent in the higher-mass population.

We have encountered this situation before, in a slightly different context. After noting the paucity of evidence for disks around main-sequence O stars, we saw in Chapter 15 how their ultraviolet photons are able to destroy, through photoevaporation, any structures inherited from the formation epoch. However, this mechanism is inadequate for the objects of interest here. Equation (15.50) gives the theoretical rate of efflux from a disk outside its gravitational radius ϖ_g. Consider a Herbig star of spectral type B4, corresponding to $T_{\mathrm{eff}} = 1.7 \times 10^4$ K. If we substitute into equation (15.50) $M_* = 7\ M_\odot$ and use the appropriate ionizing photon output of $\mathcal{N}_* = 6 \times 10^{42}\ \mathrm{s^{-1}}$, we find a mass loss rate of $3 \times 10^{-8}\ M_\odot\ \mathrm{yr^{-1}}$. Thus, a disk with initial mass 10 percent of M_* would need 2×10^7 yr to vaporize, far longer than the object's pre-main-sequence lifetime. Either the disk never formed, or other factors, such as wind erosion, predominate.

Even among objects that do have millimeter flux, the additional disk signatures found in T Tauri stars are curiously weak or absent. For example, we recall that optical forbidden lines, most notably [OI] 6300 Å, tend to be blueshifted. The common interpretation of this fact is that the redshifted emission that would arise from receding gas in the stellar wind is being occulted by an opaque disk. Many Herbig Ae/Be stars also exhibit [O I] emission, presumably arising from shocks generated in their winds. Usually, however, the line profile is either symmetric or only modestly suppressed in the red wing. This finding may mean that the winds are relatively broad, as suggested also by the apparent lack of molecular outflows and jets (§ 18.3).

Permitted optical emission lines in Herbig Ae/Be stars, such as those of the Balmer series, do not commonly exhibit deep absorption dips in the red. Such inverse P Cygni profiles are the spectroscopic sign of infall for the T Tauri class. Of course, we do not know the true source of the infalling gas, even among these objects. But if even part of the material flows from the inner regions of disks, then the rarity of the effect in intermediate-mass stars is puzzling. We note in passing that the infall signature does appear more frequently in *ultraviolet* lines. Perhaps a third of Herbig Ae stars show the effect, which is unseen at earlier spectral types.

These negative or ambiguous results should not allow us to lose sight of the secure fact that matter just outside Herbig Ae/Be stars is distributed anisotropically. The difficulty has been in establishing the specific, disk-like geometry and kinematics. As further evidence of anisotropy, we need only recall that the optical emission is often linearly polarized. Photons are evidently being scattered by dust grains in the stellar environment. These grains cannot reside in a purely spherical envelope, which would yield no net polarization. The effect is strongest, and most variable, in the UX Ori stars discussed previously. Note that these sources all have spectral types of either A or late B. It appears, once again, that the more massive Herbig Be stars have efficiently cleared out their immediate surroundings.

18.4.2 Dust Grain Processing

Those stars that still retain dusty material often exhibit broad emission near 10 µm, as mentioned in § 18.3. The detailed profile of this spectral feature is sensitive to the grain structure. It is interesting, therefore, that the profile evolves with time. Figure 18.17 illustrates the point through two representative spectra. The top panel shows observations of the now-familiar AB Aur, an A0 star with an age of 2×10^6 yr. Below is HD 100546, an object similar in spectral type (B9) but closer to the main sequence, with an age of about 1×10^7 yr. The most obvious change in the profile is the shift in its peak, from 9.7 µm in AB Aur to 11.3 µm in the older object.

Dust radiating in this regime is being heated to a temperature of roughly 500 K. Laboratory studies, coupled with theoretical modeling, find that it is the smaller, and more amorphous, silicates whose emission peaks near 9.7 µm. (The same feature, often seen in absorption, is created by dust in the interstellar medium.) Conversely, larger and more crystalline material has a profile that peaks at a longer wavelength. The particular value of 11.3 µm seen in HD 100546 indicates the presence of the mineral forsterite (Mg_2SiO_4) and also coincides with a PAH emission feature.

The picture, then, is that the grains, while heated, are nonetheless in a quiescent environment that allows them both to agglomerate and to gradually assume a more ordered structure. This finding is in accord with the inference, based on optical colors, that grains near Herbig Ae/Be

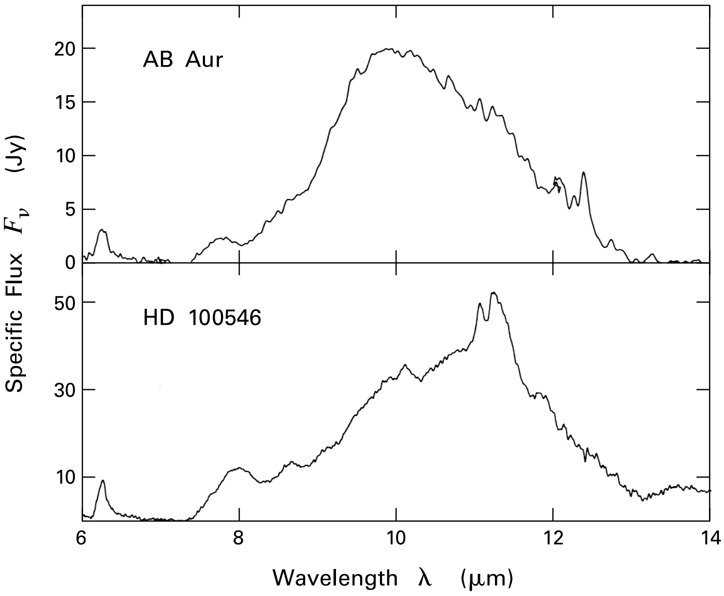


Figure 18.17 Spectra of mid-infrared emission from AB Aur (*top panel*) and the older star HD 100546 (*bottom panel*). Note the shift with age of the peak flux.

stars are larger, on average, than those in the interstellar medium. Finally, we noted earlier that spherical envelope models need the dust opacity to have a relatively weak dependence on wavelength in the millimeter regime to match observed flux values. Enhanced grain size naturally gives this property.

The top panel of Figure 18.18 shows the full spectral energy distribution for HD 100546. It is clear that the object still emits copiously in the infrared. The excess, long-wavelength luminosity is 0.51 times the stellar value of 32 $L_\odot$. Indeed, the rise in flux from the near- to mid-infrared is reminiscent of Group II Herbig Ae/Be stars (recall Figure 18.12). In the present case, however, the visual extinction is too low for the star to be embedded within a massive, opaque envelope. Again, a flattened structure is more appropriate. Note that the spectral energy distribution exhibits a strong dip just shortward of 10 μm. Such a depression, we saw in Chapter 17, suggests truncation of the disk at its inner edge.[4]

Absence of gas close to the star is not a surprise. Ultraviolet photons from this B9 object can dissociate ambient molecules and ionize the atomic component. The dip in the spectral energy distribution, however, attests to the lack of *dust* within roughly 10 AU from the stellar surface. There is copious material beyond that distance, as seen through the large infrared excess. It appears, then, that the system of orbiting grains has been somehow cleared, perhaps by infall onto the star. Infall motion is directly seen in the ultraviolet, where gaseous emission lines exhibit redshifted absorption. Curiously, the species displaying this effect are either neutral

---

[4] All three spectral energy distributions in Figure 18.18 also have strong dips shortward of 0.4 μm. This feature is intrinsic to the photosphere and represents absorption by partially ionized hydrogen in the star's outer layers.

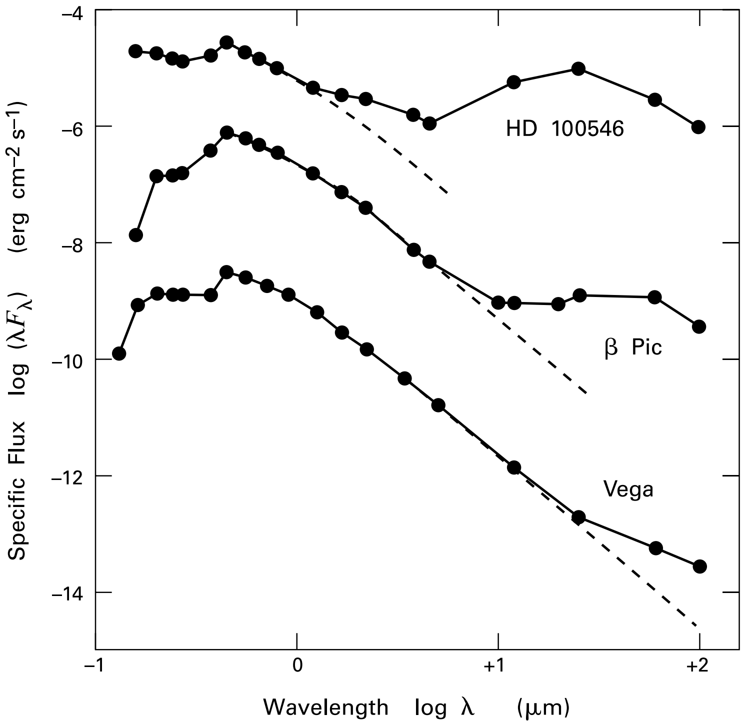


Figure 18.18 Spectral energy distributions of intermediate-mass stars close to, or on the main sequence. From top to bottom, the three objects have increasing age and decreasing infrared excess, as measured relative to blackbodies at the appropriate effective temperatures (*dashed curves*). Note the strong dips in photospheric flux near $\log \lambda = -0.4$. The top curve has been shifted upward by $\Delta \log (\lambda F_\lambda) = +2.5$, the bottom downward by $\Delta \log (\lambda F_\lambda) = -4.0$

(*e. g.*, C I) or lightly ionized (S II). Given our previous remarks, the presence of gas moving at high speed, and thus situated relatively close to the star, is unexpected. Perhaps this gas arises through evaporation of infalling, solid material.

The dust component of the disk around HD 100546 is directly observable through scattered optical and near-infrared light. Since the star is relatively luminous and close (100 pc), this structure has been probed in some detail. It can be traced out to 500 AU in radius and has a surface density that rises inward, at least to the (unresolved) truncation edge. Several dark lanes are seen that may be segments of larger, spiral structures. There is also a much lower-density envelope of grains above and below the disk plane, some 1000 AU in radius. This more diffuse material appears as a haze in scattered light.

Another well-studied disk is that associated with the A0 star HR 4796A. The star itself is the brightest member of the TW Hya Association, described earlier in § 17.5. The age of HR 4796A is thus known to be about 1×10^7 yr, close to that of HD 100546. Nevertheless, the excess, long-wavelength luminosity is a good deal less, only 5×10^{-3} of the stellar value. Spectral observations in CO have failed to detect any flux. It appears that most circumstellar gas has been driven off, leaving behind the collection of orbiting grains we have termed a debris disk.

The upper left panel in Figure 18.19 displays an image of this disk in scattered, near-infrared light. To reduce contrast, the star was masked with a coronagraph, as shown. We see, with remarkable clarity, that the circumstellar material in this case is actually confined to a narrow, highly inclined *ring*. The width of this ring is measured to be 17 AU. Interior to it, *i. e.*, within a radius of some 60 AU outside the star, all grains have been cleared. We are reminded of the interior gap in the debris disk of ϵ Eri, the K dwarf situated in the immediate solar vicinity (Figure 17.27). Another point of similarity is that the observed dust grains in both cases are larger than those generally found in the interstellar medium.

18.4.3 Main-Sequence Disks

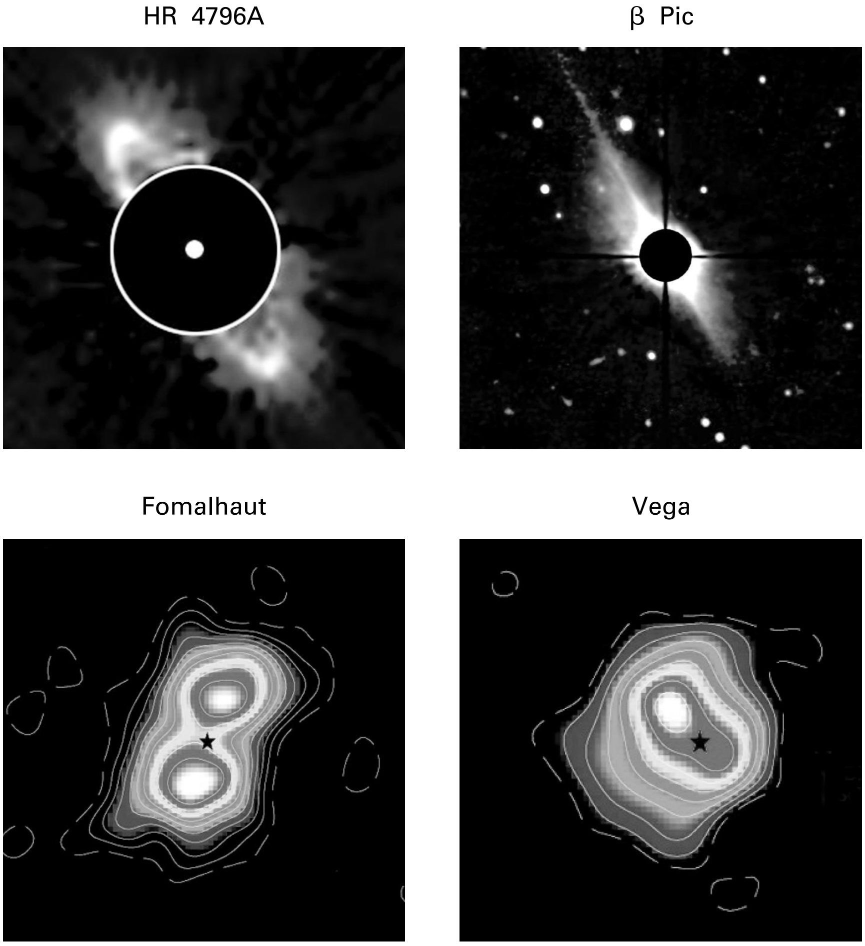
Scattered-light images of disks can be striking in their detail, but are available only for a handful of relatively nearby systems. Detection of unresolved, infrared flux from a star has always provided the initial evidence for circumstellar matter. This longer-wavelength radiation represents emission from heated dust grains. Because of their higher luminosities, A stars, such as the zero-age main-sequence object HR 4796A and its older counterparts, have constituted the bulk of the sample. All told, a substantial (but quantitatively uncertain) fraction of field stars are Vega-like, *i. e.*, display excess thermal emission, but have little sign of circumstellar gas.

The most intensively studied debris disk has been that around β Pictoris, a Southern hemisphere star lying at a distance of 19 pc. One may assess the age of this $1.8\ M_\odot$ object by utilizing the fact that it is moving through space with several, lower-mass companions that are still on pre-main-sequence tracks. Through this means, the age appears to be 1.5×10^7 yr, not much older than HR 4796A. The spectral energy distribution in Figure 18.18 shows that there is slightly less infrared excess luminosity, about 3×10^{-3} times the L_*-value of $8.5\ L_\odot$.

Coronagraphic optical imaging reveals a remarkably narrow structure (Figure 18.19; upper right panel). Apparently, we are viewing the debris disk almost exactly edge-on. Also striking is the manifest asymmetry. While the northeast portion of the disk extends out to 790 AU from the star, the opposite side can only be seen in this image to a radius of 650 AU. The gravitational influence of another body, either a planet or passing star, may have created the distortion.

A wide range of observing facilities, including both spaceborne and groundbased telescopes, have provided many more details. The general picture, in any case, follows the pattern we have already seen. From the thermal, infrared emission, we know that the orbiting grains are relatively large, typically 10 µm in size. There is a steep falloff in this material within some tens of AU from the central star, but smaller grains persist farther inside. A relatively small amount of gas exists, some of it lying in a ring within 1 AU of the star. All of this attention, we note, has been lavished on a structure whose total mass amounts to 10 percent that of the Earth. The interest lies mainly in the generation and erosive depletion of this material, and in the evidence from both processes for more massive, orbiting bodies.

As a star continues to age, its circumstellar dust eventually becomes undetectable through scattered light. However, thermal radiation at submillimeter wavelengths still yields images, as exemplified in the two lower panels of Figure 18.19. On the left is an 850 µm map of Fomalhaut (α Piscis Austrini), an A3 star 8 pc away. All main-sequence objects in the field, including



HR 4796A

β Pic

Fomalhaut

Vega

Figure 18.19 Debris disks around four main-sequence A stars. The top two images are in scattered starlight (optical and near-infrared), while the bottom two display submillimeter thermal emission from grains.

this one, have highly uncertain ages. Here, there happens to be a nearby K star with identical proper motion and radial velocity. Assuming the two objects were born together, we may use the fact that the lower-mass companion has a detectable lithium abundance to assign an age to both of 2×10^8 yr. The submillimeter image of Fomalhaut itself shows a ring-like debris disk seen edge-on, *i. e.*, essentially in cross section. A hollow central cavity extends 30 AU in radius. Within the ring are relatively large (10 μm) grains but no observed gas.

The prototype of all objects with excess, long-wavelength emission but little accompanying gas is Vega (α Lyrae), shown in the lower panel of Figure 18.19. The spectral energy distribution, plotted in Figure 18.18, reveals that the actual luminosity contributed by dust has declined to $2 \times 10^{-5}\, L_*$. Judging from its position in the HR diagram, the star has already left the zero-age main sequence and has an age of 4×10^8 yr. The 850 µm map in Figure 18.19 shows a disk viewed nearly face-on. In addition to a low-emission gap of radius 60 AU surrounding the star itself, there is also an interior, bright peak that is curiously off-center. Note, however, that this region is marginally resolved in the image.

Table 18.2 summarizes key properties of the debris disk systems we have discussed. Here we have included $L_{\rm IR}/L_*$, the fraction of the stellar luminosity reemitted at long wavelengths, as well as estimates or upper bounds for $\varpi_{\rm in}$, the interior radius of the disk itself. While the age uncertainty is large, there is clearly a tendency for the total mass in orbiting dust to diminish with time. Erosion apparently proceeds faster on the inside, as evidenced by the central holes. What accounts for these trends?

Table 18.2 Nearby Stars with Debris Disks

Name	Distance (pc)	Spectral Type	M_* $(M_\odot)$	L_* $(L_\odot)$	$L_{\rm IR}/L_*$	$\varpi_{\rm in}$ (AU)	Age (10^6 yr)
HD 100546	103	B9	2.4	32	0.51	<50	10
HR 4796A	67	A0	2.5	20	5×10^{-3}	55	10
β Pictoris	19	A5	1.8	8.5	3×10^{-3}	<15	15
Fomalhaut	7	A3	2.0	13	8×10^{-5}	30	200
Vega	8	A0	2.5	60	2×10^{-5}	60	400

18.4.4 Effects of Radiation

The first point to appreciate is that, regardless of their origin, all grains below a minimum size are rapidly expelled by stellar radiation pressure. Suppose, for simplicity, that the dust perfectly absorbs all photons and reradiates them isotropically. Recall that each photon of energy ϵ carries momentum ϵ/c. Then $f_{\rm rad}$, the radiative force on a single grain of cross section σ_d, is

$$f_{\rm rad} = \frac{\sigma_d\, L_*}{4\,\pi\, c\, \varpi^2} \ , \tag{18.8}$$

where ϖ is the distance from the star. Here, we have neglected scattering and assumed that the absorption efficiency is unity, so that the cross section is the geometric value. The opposing gravitational force is

$$f_{\rm grav} = \frac{G\, M_*\, m_d}{\varpi^2} \ , \tag{18.9}$$

for a grain with mass m_d.

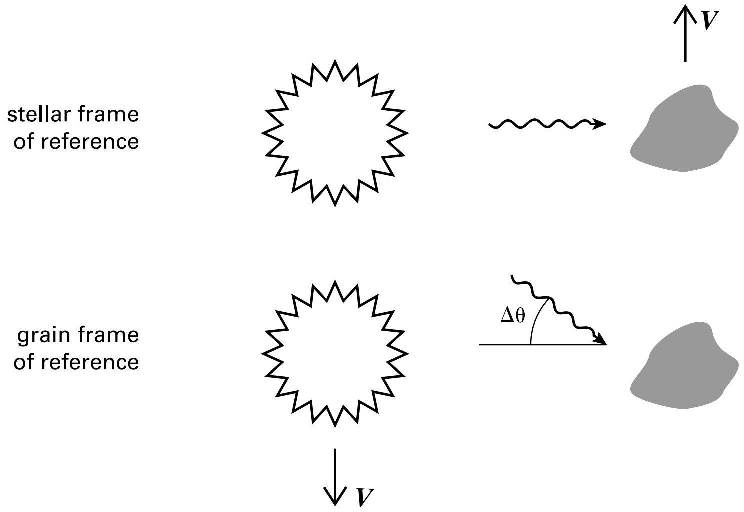


Figure 18.20 Mechanism of Poynting-Robertson drag. In the stellar frame of reference, the photon path is orthogonal to the velocity $\boldsymbol{V}$ of the dust grain. In the grain's reference frame, this path is tilted by the small angle $\Delta\theta$, creating a retarding force.

We now evaluate both σ_d and m_d by assuming that the grain is a sphere of radius a_d and internal density ρ_d. Equating $f_{\rm rad}$ and $f_{\rm grav}$ gives us the desired minimum radius:

$$\begin{aligned} a_{\rm min} &= \frac{3\,L_*}{16\,\pi\,c\,\rho_d\,G\,M_*} \\ &= 1\ \mu{\rm m}\,\left(\frac{L_*}{10\,L_\odot}\right)\left(\frac{M_*}{2\,M_\odot}\right)^{-1} \ . \end{aligned} \tag{18.10}$$

In making the numerical estimate, we have set ρ_d equal to 3 g cm^{-3}. It is no surprise that grains of ordinary interstellar size are *not* seen in debris disks around A stars. Radiation pressure long ago drove them away, within an interval at most equal to an orbital period.

Larger grains are also vulnerable. Curiously, radiation pressure causes these to spiral *inward*, toward the central star. Figure 18.20 explains why. In the star's reference frame, photons impact the grain broadside as it moves with orbital speed V, at a radial distance ϖ. The only effect of the transverse force, if it is less than gravity, is to increase ϖ slightly above the Keplerian value. In the rest frame of the grain, however, each incoming stellar photon acquires a small, retrograde velocity component. That is, the propagation direction is tilted by the angle $\Delta\theta$, where

$$\Delta\theta \approx V/c \ . \tag{18.11}$$

The tilt in photon direction means that the grain effectively experiences a retarding force from the impinging radiation field. This force is known as *Poynting-Robertson drag*.

The grain spirals inward because the drag exerts a torque, equal in magnitude to $\varpi\, f_{\rm rad}\,\Delta\theta$. This torque, in turn, drains orbital angular momentum, which has the instantaneous value

$\varpi\, m_d\, V$. Since the decay of the orbit is very gradual in practice, we may approximate V at any time as Keplerian. Setting the torque equal to the rate of change of angular momentum and utilizing equations (18.8) and (18.11), we find the evolution of the orbital radius:

$$\frac{d\varpi}{dt} \,=\, -\frac{\sigma_d\, L_*}{2\,\pi\, m_d\, c^2\, \varpi}\ . \tag{18.12}$$

Notice, from this last equation, that the radial decay is slowest at the beginning and then speeds up as the grain moves inward. Thus, the total time required to drift all the way to the star may be found by integration from the initial radius $\varpi_\circ$ to $\varpi \,=\, 0$. Denoting this *Poynting-Robertson decay time* as $t_{\rm PR}$, we find it to be

$$t_{\rm PR} \,=\, \frac{\pi\, m_d\, c^2\, \varpi_\circ^2}{\sigma_d\, L_*}\ . \tag{18.13}$$

We again idealize the grain as a uniform-density sphere of radius a_d. Then $t_{\rm PR}$ becomes

$$\begin{aligned} t_{\rm PR} \,&=\, \frac{4\,\pi\, a_d\, \rho_d\, c^2\, \varpi_\circ^2}{3\, L_*} \\ &=\, 2\times 10^7 \ {\rm yr}\ \left(\frac{a_d}{10\ \mu{\rm m}}\right)\left(\frac{\varpi_\circ}{100\ {\rm AU}}\right)^2\left(\frac{L_*}{10\, L_\odot}\right)^{-1}\ . \end{aligned} \tag{18.14}$$

A more careful derivation would properly account for the fact that longer-wavelength stellar photons are inefficiently absorbed and that *all* photons may instead be scattered. Nevertheless, our result is adequate for illustrating the main effect of the drag. It is apparent that even relatively large grains spiral inward over a period less than the stellar age. Moreover, the Poynting-Robertson decay time is shorter at smaller orbital radius. The inner disk is thus the first region to be cleared by this means.

18.4.5 Episodic Infall

We are led to the same conclusion we reached in Chapter 17, when discussing the ϵ Eri debris disk. The relatively short times for orbital decay and radiative expulsion imply that the presently observed grains are not primordial, but must be continually replenished. An attractive source of material is again provided by the mutual collision of larger solid bodies, *e. g.*, asteroids and planets. Such objects could well be lurking within the presently observed disks, but have so far escaped detection. Any fragments they produce are subject to further collisions and shattering. The current grain population must result from balancing this input source against radiative clearing.

The total amount of mass locked up in solid bodies is difficult to ascertain, and arguments to this effect are necessarily indirect. Besides acting as a potential source for grains, these objects may also supply whatever *gas* is observed. Recall that HD 100546 exhibits redshifted absorption troughs in ultraviolet lines. The transitions are in atoms that are either neutral or singly ionized. We speculated that this material represents evaporation by infalling bodies. The absorption features vary in time and can disappear entirely. Thus, the inward flux of the parent solid bodies appears to be episodic, rather than continuous.

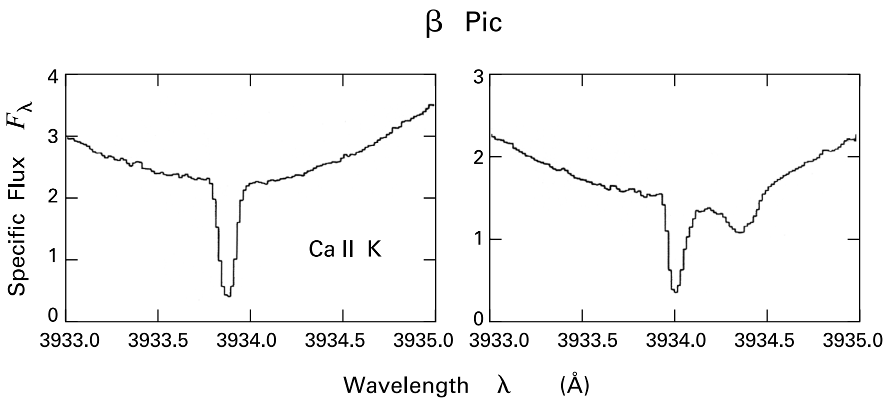


Figure 18.21 Evidence for episodic infall onto β Pic. The left and right spectra of the Ca II K line were taken in 1984 and 1986, respectively.

A similar phenomenon occurs in other debris disks and has been extensively monitored in β Pic. Figure 18.21 displays two spectra, taken several years apart, of the Ca II K line. What we are seeing is actually the bottom of the rotationally broadened photospheric absorption trough. Exactly at line center is an additional, narrower dip, which does not change with time. Redward of this feature, however, there appears a new dip at the later epoch. This depression signifies inward motion of the Ca II ion, with a radial velocity in this instance of 25 km s^{-1}.

While the Ca II features come and go at day-long intervals, ultraviolet lines, such as Mg II h and k, change within hours and show infall speeds as high as 400 km s^{-1}. The total number of events is on the order of several hundred per year. A model, which has been elaborated in considerable detail, posits that the objects ultimately responsible are comets. Originating somewhere in the debris disk, they plunge toward the star on highly eccentric orbits. Increased heating from the central object eventually drives off grains from the bodies' surfaces. These form a dust coma around the comet nucleus, as witnessed in solar system objects. Ultraviolet radiation then sublimates the grains, creating the atoms and ions that we detect through their absorption of stellar light.

According to this picture, the intermittency of the gaseous infall signature means that the flux of comets has a similar temporal pattern. The fact that the absorption is nearly always redshifted rather than blueshifted is curious; one would expect an equal number of bodies that are approaching us after a near miss of the star. Equally puzzling is the narrow absorption dip at line center. The atoms and ions creating this feature have little radial velocity. Yet they experience strong radiation pressure, which should quickly drive them away.

Comets in the solar system plunge inward from the outer disk because of gravitational encounters with giant planets. Pursuing the analogy, the β Pic comets may be so perturbed by an unseen planet that their elongated orbits lock to it. Numerical simulations find that their major axes can align with the planet's in a configuration that minimizes close approach to the larger body. Each passage of the planet across the star would then generate a flurry of absorption events in the spectrum. If the comets are completely evaporated before they round the star, then only redshifts would result. Of course, new comets then need to be continuously supplied, presumably by this or another planet. Such ideas still lie in the realm of speculation. Most welcome at this point would be direct evidence for the planets themselves, either around β Pic or other stars with debris disks. Such a finding would demonstrate unequivocally that formation of these bodies is not limited to solar-type objects.

Chapter Summary

Many young clusters, when displayed in the HR diagram, contain a large population of intermediate mass, pre-main-sequence stars. Yet only the small minority of these objects with optical line emission are classified as Herbig Ae/Be stars. Each is surrounded by dusty cloud material, seen either in optical reflection or in the millimeter continuum. Some create local bubbles, driving nearby gas away through radiative heating and winds. Those of earlier spectral type are surrounded by a retinue of T Tauri stars.

Rotation in Herbig Ae/Be stars is more rapid than in lower-mass objects, but still significantly under breakup. Few stars display significant periodic changes in broadband emission. Those in the UX Ori subclass undergo sudden dimming, accompanied by enhanced polarization and a blueward shift in color. Both developments indicate temporary obscuration by circumstellar dust clouds.

From a theoretical perspective, the evolution of intermediate-mass objects is influenced by the fact that the deep interior is radiatively stable. The star begins with a modest surface luminosity and then undergoes thermal relaxation. Here, the central region contracts, while transferring heat to an expanding, exterior mantle. After a relatively short time, the surface luminosity jumps upward. Subsequent contraction to the main sequence is more uniform. However, theory predicts that the star's outer layers are unstable to pulsation for a limited time. These pulsations have been observed. Arriving on the main sequence, the star develops a central convection zone in which hydrogen fusion occurs via the CNO bi-cycle.

The spectral energy distributions of all Herbig Ae/Be stars show some degree of infrared excess. The majority, Group I objects, have excesses that are too large to originate solely from disks. An additional, dusty envelope is needed. This is also the case for the even more infrared-luminous Group II stars. Some of these envelopes have been imaged directly through interferometry. The relatively small infrared excess in Group III stars plausibly stems from hot, circumstellar gas.

Ultraviolet radiation creates photodissociation regions in surrounding cloud material. These regions are observed in the infrared, through both continuum dust emission and through atomic and molecular transitions. Heating of the cloud gas disperses it, along with the ram pressure from stellar winds. While the winds are well-documented through P Cygni profiles, there origin is not understood. Equally mysterious is the X-ray luminosity, which is intermediate in strength between that in T Tauri stars and massive, main-sequence objects.

Direct observational evidence for disks is scant, and absent entirely for earlier-type Herbig Be stars. Nevertheless, we do see remnant debris disks around older objects. The dust surrounding an aging pre-main-sequence star gradually agglomerates into larger, more crystalline grains. Poynting-Robertson drag from stellar photons causes these grains to spiral inward.

Others must be resupplied, presumably through the mutual collisions of larger orbiting bodies. In any case, both optical and submillimeter observations show that debris disks gradually disappear during the main-sequence phase. Episodic infall is seen through gaseous absorption lines. This material may originate from the evaporation of star-grazing comets.

Suggested Reading

Section 18.1 The original identification of Herbig Ae/Be stars is due to

Herbig, G. H. 1960, ApJSS, 4, 337.

Two comparatively recent reviews are

Pérez, M. & Grady, C. A. 1997, Sp. Sci. Rev., 82, 407

Waters, L. B. F. M. & Waelkens, C. 1998, ARAA, 36, 233.

For the clustering of other stars around these objects, see

Testi, L., Palla, F., & Natta, A. 1999, AA, 342, 515.

A useful comparison of Herbig Ae/Be and classical Be stars is

T. Böhm & L. A. Balona 2000, in “The Be Phenomenon in Early-Type Stars," eds. M. A. Smith, H. F. Henricks, & J., Fabregat (San Francisco: ASP), p. 103.

Section 18.2 Thermal relaxation in the contraction of intermediate-mass stars was studied numerically by

Palla, F. & Stahler, S. W. 1993, ApJ, 418, 414.

For a detailed account of the CNO bi-cycle, see

Rolfs, C. E. & Rodney, W. S. 1988, *Cauldrons in the Cosmos* (Chicago: U. of Chicago), Chapter 6.

The pre-main-sequence instability strip was first described by

Marconi, M. & Palla, F. 1998, ApJ, 507, L141.

Section 18.3 The classification of Herbig Ae/Be stars according to their spectral energy distributions is due to

Hillenbrand, L. A., Strom, S. E., Vrba, F. J., & Keene, J. 1992, ApJ, 397, 613.

The disk interpretation offered in this paper for Group I objects has subsequently been challenged, in part as a result of interferometric imaging in the near-infrared. See, for example,

Millan-Gabet, R., Schloerb, F. P., & Traub, W. A. 2001, ApJ, 546, 358.

Photodissociation regions are analyzed theoretically in

Diaz-Miller, R. I., Franco, J., & Shore, S. N. 1998, ApJ, 501, 192.

The presence of winds was first documented in an influential study:

Finkenzeller, U. & Mundt, R. 1984, AAS, 55, 109.

Section 18.4 Observational evidence supporting the existence of disks around Herbig Ae/Be stars has been summarized by

> Natta, A., Grinin, V. P., & Mannings, V. 2000, in *Protostars and Planets IV*, eds. V. Mannings, A. P. Boss, & S. S. Russell (Tucson: U. of Arizona), p. 559.

Debris disks have been studied intensively over many years. Two reviews are

> Backman, D. E. & Paresce, F. 1993, in *Protostars and Planets III*, ed. E. H. Levy & J I. Lunine (Tucson: U. of Arizona), p. 1253
>
> Zuckerman, B. 2001, ARAA, 39, 549.

The first paper, which predates the submillimeter observations, still provides a lucid summary of physical clearing mechanisms. For the infalling comet model of β Pic, see

> Beust, H. 1994, in *Circumstellar Dust Disks and Planet Formation*, eds. R. Ferlet & A. Vidal-Madjar (Gif-sur-Yvette: Editions Frontières), p. 35.

Part VI

A Universe of Stars

The Formation of Stars
Steven W. Stahler and Francesco Palla

19 Star Formation on the Galactic Scale

Our exploration of stellar birth has led us through many different environments. These have ranged widely in physical extent, from the compact, nuclear-burning cores of young stars themselves to the largest coherent gas clouds spawning these objects. The underlying empirical data have been drawn almost entirely from our own Galaxy, indeed largely from the solar neighborhood. In the final two chapters comprising Part VI, we broaden our perspective in two senses. First, we describe what is known about star formation activity over the dimensions of entire galaxies. We see how the consumption of diffuse gas is affected by, and in turn may influence, the morphology of these systems. Our description here is necessarily more qualitative and less detailed than in our coverage of nearby regions. Chapter 20 will then attempt, in a summary fashion, to reconsider the physical problem of star formation as a whole. We assess both the insights we have gained and the areas where understanding remains poor.

We begin our galactic investigation with the Milky Way itself. After describing our own pattern of stellar birth, we turn in § 19.2 to that in other spiral galaxies, as well as the two other major classes, ellipticals and irregulars. In the first type, star formation ended long ago. We describe the clues that allow us to reconstruct the early history of activity. In gas-rich irregulars, including small galaxies neighboring ours, the production of new stars is ongoing.

This activity reaches an extreme in starburst systems, the subject of Section 19.3. Here, a region of the galaxy, usually close to the center, is undergoing a dramatic, and ultimately short-lived, flareup in formation activity. As we shall see, the influx of gas that fuels this explosive event is often stimulated by interaction or merger with a neighboring galaxy. Exotic phenomena also occurred in much older systems, including some types no longer seen, and others that appear to antedate present-day galaxies (§ 19.4). As we extend our investigation, in Section 19.5, to formation of the very first stars, we reach the limits, in both time and space, of current knowledge.

19.1 The Milky Way Revisited

In Chapter 1, we introduced two basic measures of star formation in our Galaxy. One, which we denoted as $\dot{m}_*$, is the *local* birthrate, expressed in units of $M_\odot$ yr^{-1} per square parsec of the Galactic disk.[1] This figure refers to activity only within the solar neighborhood. The second quantity, $\dot{M}_*$, is the *total* rate (in $M_\odot$ yr^{-1}), summed over the entire disk. How are these numbers obtained? Moreover, if ϖ denotes the Galactocentric radius, what is known about $\dot{m}_*(\varpi)$, the radial variation of the star formation rate? Our first goal should be to give

---

[1] We limit our discussion to the traditional "thin" disk, with scale height of order 100 pc. There is also evidence for a *thick disk*, consisting of older stars. The scale height of the latter is greater by an order of magnitude.

The Formation of Stars. Steven W. Stahler and Francesco Palla

ISBN: 3-527-40559-3

the empirical basis for these key measures of activity. We may then see how our own Galaxy compares with others.

19.1.1 Activity in the Disk

The fact that we are embedded within the Galactic plane makes it impossible to observe the full stellar content of young, distant groups, and hence their associated birthrates. Optical light is strongly absorbed by interstellar dust, so one must turn to infrared and radio wavelengths. Figure 1.19 demonstrates how near-infrared photons penetrate to the central bulge of the Galaxy. Much of this light, however, is contributed by red giants. The daunting task of separating out the youthful component over the entire disk remains unsolved. A more feasible approach is to survey far-infrared and radio-wavelength emission known to arise exclusively from *massive* stars. Since these objects have relatively short lifetimes, one may obtain what is essentially their instantaneous formation rate without needing to consider the prior history at the location of interest. Summing the birthrate over all masses then requires that we augment the massive population employing the initial mass function.

Let us focus on the radio emission, which is both thermal and nonthermal. Within the first category is a diffuse background of centimeter radiation that pervades interstellar space. While the received flux is weak compared to other interstellar sources, it is much too strong to originate from the photospheres of stars. Rather, it represents free-free emission from HII regions surrounding massive objects. We described in Chapter 15 how these volumes are ionized by Lyman continuum photons. It is straightforward, at least in principle, to convert the centimeter flux, observed as a function of Galactic longitude, to a corresponding emission rate of stellar ionizing photons, now expressed in terms of Galactocentric radius. The latter function, in turn, yields the surface density of O and early-B stars. Their birthrate follows by utilizing the appropriate main-sequence lifetimes (Table 1.1).

A very similar analysis may be applied to two sources of *nonthermal* radio flux, again most conveniently at centimeter wavelengths. One is synchrotron emission from supernova remnants (§ 15.5). It has been established empirically that the surface brightness, *i. e.*, the specific flux divided by the solid angle subtended by the source, declines steeply with the physical diameter of a remnant's surface. This relation allows one to assess the distance to individual remnants, provided they are spatially resolved. The observed space density of the objects then translates into a distribution of progenitor, high-mass stars.

Finally, extensive radio surveys have detected thousands of Galactic pulsars. Each, we recall, is a rapidly spinning neutron star. The distance in this case is established utilizing the phenomenon of *interstellar dispersion*. The source itself emits brief, repetitive bursts of radiation over a broad spectral range. As photons traverse the ionized component of interstellar gas, those of shorter wavelength have higher group velocity. The net result is that the original burst is spread out (dispersed) by an amount proportional to the distance traveled.[2] Again, the areal distribution of sources may be reinterpreted as the spatially varying birthrate of the associated, massive stars.

---

[2] The dispersion actually depends on the emission measure, *i. e.*, the square of the electron volume density integrated over distance (§ 15.1). Extracting the distances to numerous pulsars thus requires a global model of the electron density throughout the Galactic disk.

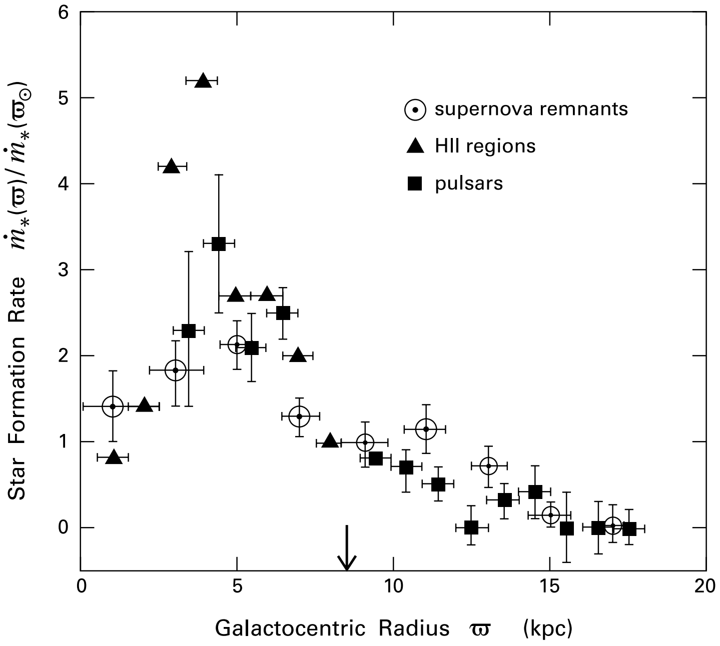


Figure 19.1 Star formation rate per unit area in the Milky Way disk. The rate, shown as a function of Galactocentric radius, is normalized to that of the solar system, whose radial position is indicated here by a vertical arrow.

We emphasize that the birthrates thus obtained represent only the small fraction pertaining to the high-mass tail of the stellar distribution. If we assume that the initial mass function is the same throughout the disk, we may plot up the full, empirical birthrates normalized to $\dot{m}_*(\varpi_\odot)$, the value in the solar neighborhood. Figure 19.1 shows the result of this exercise. Allowing for the considerable measurement uncertainties, it is significant and gratifying that all three methods exhibit a similar trend. The star formation rate increases steadily inward to $\varpi \approx 3$ kpc. There is indication of a steep decline interior to that radius.

Assigning absolute values to $\dot{m}_*(\varpi)$ now requires only a single number, the local birthrate. Here, we recall the method used to establish the initial mass function itself (§ 4.5). Equation (4.4) shows explicitly how $\dot{m}_*(\varpi_\odot)$ is needed to convert the present-day luminosity function $\Phi(M_V)$, to its initial counterpart $\Psi(M_V)$. The latter, in turn, yields the initial mass function $\xi(M_*)$ through equation (4.5). Since the luminosity function $\Phi(M_V)$ is known in physical units (see Figure 4.21), the conversion to $\Psi(M_V)$ yields a true value for $\dot{m}_*(\varpi_\odot)$. Under the additional assumption that the local birthrate has been *temporally* constant (see below), we find $\dot{m}_*(\varpi_\odot) \approx 3 \times 10^{-9}\ M_\odot\,\mathrm{yr}^{-1}\,\mathrm{pc}^{-2}$, with a 50 percent uncertainty in either direction. Spatial integration of a smoothed approximation to the plot in Figure 19.1 then yields the global formation rate throughout the disk. This is $\dot{M}_* \approx 4\ M_\odot\,\mathrm{yr}^{-1}$, as quoted in Chapter 1.

We know that stars are born within molecular clouds. How, then, does the radial variation of $\dot{m}_*(\varpi)$ compare with that of the H_2 surface density? A glance at Figure 2.3 reveals substantial, qualitative agreement. There is again a slow, inward rise in the gaseous surface density, with a peak at $\varpi \approx 4\ \mathrm{kpc}$. There is an apparent decline further inside, also consistent with the spatial pattern in the stellar birthrate.

19.1.2 The Galactic Center

An additional, and striking, feature of the H_2 distribution in Figure 2.3 is the high, central peak. Is there a corresponding rise in the star formation rate close to the Galactic center? If so, this rise is certainly not apparent in Figure 19.1. We must keep in mind, however, the inherent limitations of our empirical methods for measuring $\dot{m}_*(\varpi)$. For example, the dispersion of pulsars residing in the dense, central portion of the Galaxy is so great that discrete pulses are smeared out. The identification of individual supernova remnants also becomes problematic. Nevertheless, the suggestion provided by the molecular spike in Figure 2.3 is true. There is indeed a very high level of star formation activity close to the Galactic center.

Before delving into this topic, however, consider once again the spatially extended gas in the disk and its rate of conversion into stars. In Chapter 2, we cited a total mass in Galactic H_2 of $2 \times 10^9\ M_\odot$ (see Table 2.2). Using $\dot{M}_* = 4\ M_\odot\,\mathrm{yr}^{-1}$, all of the molecular gas should disappear in 5×10^8 yr, a small fraction of the Galactic age of 1×10^{10} yr. We must account, of course, for the fact that a portion of the molecular gas is recycled back to the interstellar medium by winds, novae, and supernovae. But this effective reduction of $\dot{M}_*$, fractionally perhaps 25 percent, does not alter the essential result. On the other hand, the mass in *atomic* hydrogen within the disk is larger, roughly $7 \times 10^9\ M_\odot$ (Table 2.2). The depletion time for the total gas reservoir is thus increased to 2×10^9 yr, a more palatable result. There must be a steady conversion of HI to H_2 throughout this epoch. Since the HI surface density is quite flat at large ϖ, there is evidently radial transport of the gas, as well.

Some of that gas makes its way to the Galactic center. In Chapter 5, we discussed the rich chemistry of Sgr B2, a giant molecular cloud displayed, in Figure 5.3, through its submillimeter emission. While Sgr B2 holds the record as the most massive cloud in the Galaxy, it is just one of a half dozen similar entities within the region. All have masses exceeding $10^6\ M_\odot$ and densities $n_H \gtrsim 10^4\ \mathrm{cm}^{-3}$, more characteristic of dense cores in the surrounding disk. The kinetic temperature in these clouds is also relatively high, typically 70 K, as are their internal velocity dispersions, which range from 15 to 50 km s^{-1}.

The giant clouds all appear to lie within the Galactic plane. Surrounding them is a massive ring, also composed of molecular hydrogen. This latter structure has a radius of some 200 pc and is tilted with respect to the plane. The ring and clouds within it together make up the *central molecular zone*. A magnetic field of milligauss strength permeates the gas. This field manifests itself through long filaments that glow in synchrotron, radio emission. Other high-energy phenomena include a hot ($T \gtrsim 10^7$ K) plasma seen in X-rays and extending 100 pc in radius. There is also evidence for a massive black hole lurking within Sgr A^*, the intense radio source lying even closer to the center (Figure 5.3).

But let us return to our main topic, star formation activity. Application of the radio-continuum method previously described, together with a similar analysis of far-infrared emission, yields a star formation rate in the central molecular zone of $0.3\ M_\odot\ \mathrm{yr}^{-1}$, nearly 10 percent of that in the entire Galaxy. Where, precisely, are the new stars being created? Consider first Sgr B2. When viewed in continuum, radio emission with the VLA interferometer, the complex breaks up into dozens of HII regions, distributed within a 9×11 pc area. Many are ultracompact and display cometary or shell-like morphologies (recall Figure 15.7). The continuum observations also reveal a more extended halo of emission surrounding the discrete HII regions. We are again witnessing the heating and ionizing effects of massive stars. These are undoubtedly accompanied by many more, lower-mass objects, but the latter are undetectable, at least through such means.

Some clouds within the central molecular zone appear to be sterile, while others are forming similar populations of massive stars, as revealed through radio observations in both the continuum and recombination lines. The latter indicate the velocity of outflowing, ionized material. The most intense star formation activity, however, occurs in three clusters that are *not* embedded within giant molecular clouds, presumably because the gas has already been driven off. One such cluster, about 1 pc in radius, surrounds the central black hole and radio source Sgr A*. Indeed, the mass of the black hole ($2 \times 10^6\ M_\odot$) has been obtained by tracking the proper motion of member stars. Just outside the stellar group is a gaseous disk or torus whose inner edge is being illuminated and ablated by stellar radiation and winds.

Two other isolated clusters, also seen only in the infrared, lie about 30 pc in projection from the center. One, dubbed the Arches, is named for a system of nearby radio filaments. After it was realized that these structures are emitting thermally, and therefore require an external source of heating, a deep near-infrared search uncovered the stars. The cluster is an impressive entity. While only a few tenths of a parsec in size, it contains over 150 O stars. Extrapolation gives a total cluster mass in excess of $10^4\ M_\odot$. A similar group is the Quintuplet Cluster. This is also found near a region of extended radio emission. Figure 19.2 shows some of the more prominent stars, as well as the radio source, which is aptly called The Pistol. The name of the cluster itself derives from the original group of five stars detected in near-infrared surveys. Thousands of other members were later observed by the Hubble Space Telescope. The color-magnitude diagram exhibits a main-sequence turnoff, giving a rough age of 4×10^6 yr. This figure is twice that for the Arches.

The Central, Arches, and Quintuplet clusters are the most massive concentrations of young stars in the Galaxy. For example, they collectively account for 10 percent of all the O stars. Their central densities exceed by two orders of magnitude that in the Orion Nebula Cluster. And yet these compact groups must be short-lived, as we find no analogous systems in the Galactic center with ages of 10^7 yr or longer. Furthermore, the total stellar production rate of the clusters, prodigious as it is, represents a minor fraction of that within the central molecular zone. We should view the clusters as spectacular, but transient and highly localized, flareups of activity within a larger arena of persistent star formation.

Broadening our perspective still further, the central production of stars within the Galaxy depends on a steady inflow of gas from the disk. It is not difficult to see why such a drift occurs. Recall from our discussion of protostellar disks (Chapter 11) how self-gravity within a differentially rotating, flattened structure establishes trailing spiral arms. The gravitational perturbation associated with these arms creates internal torquing that in turn leads to mass inflow. Protostellar disks are gaseous entities, while galaxies are composed mainly of discrete stars. Nevertheless, the same basic physics applies, as demonstrated by the spiral arms of our own and similar galaxies. We can thus understand, at least in general terms, how the Galactic cen-

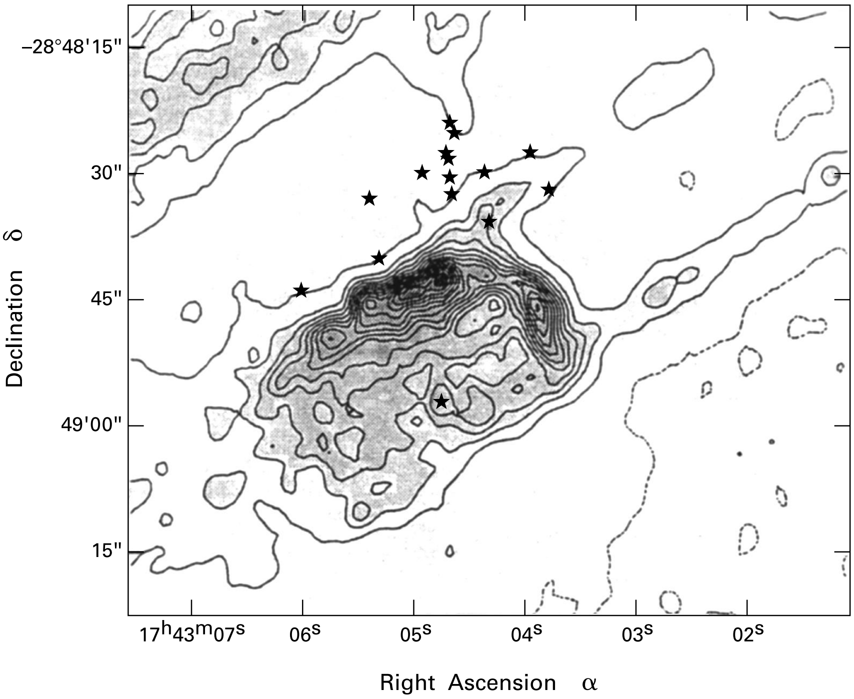


Figure 19.2 The Quintuplet Cluster near the Galactic Center. The grayscale shading and contours display 6 cm continuum emission. Also indicated are a few of the brightest, infrared stars.

ter obtains its reservoir of diffuse material for new stars. Note finally that the spike in central density creates a strong tidal force of gravity. Thus, material creeping inward is eventually disrupted. This process accounts for the brief lifetimes of even dense clusters like the Arches and Quintuplet. [3]

19.1.3 Schmidt's Law

We noted earlier that the H_2 surface density and $\dot{m}_*(\varpi)$, the azimuthally averaged star formation rate, both decline for $\varpi \lesssim 4 \text{ kpc}$. Similarly, the volumetric gas density and formation rate also vary together, attaining their highest values toward the Galactic center. It is hardly surprising that the production of new stars at any location depends on the availability of fuel, in the form of molecular gas. Whether a simple, functional relationship holds throughout the Galaxy is another matter, one that has been discussed for many years. In 1959, M. Schmidt originally proposed that the star formation rate in the solar neighborhood is proportional to the square of

---

[3] The tidal force vanishes at $\varpi = 0$ and thus may not be disrupting the Central Cluster. Unfortunately, it is also more difficult to ascertain the age of this group from its color-magnitude diagram, which is heavily contaminated by older stars from the Galaxy's central bulge.

the local gas density, where both quantities are measured per unit volume. He cited as evidence the scale height of young stars, which is about half that of the diffuse gas. Note that the latter referred to *atomic* hydrogen. However, it is reasonable to assume that the relevant quantity is the *total* gas density, atomic plus molecular. This assumption is consistent with Schmidt's original hypothesis, since HI dominates H_2 in the solar neighborhood (Figure 2.3).

Models of large-scale star formation and the recycling of disk gas employ a similar, but mathematically distinct, relationship. The most common form of *Schmidt's law* is

$$\dot{m}_*(\varpi) \,=\, A\,\left[\Sigma_{\rm gas}(\varpi)\right]^N \ . \tag{19.1}$$

Here, A and N are constants and $\Sigma_{\rm gas}$ is the total surface density of interstellar gas. A direct comparison of Figures 2.3 and 19.1 shows that equation (19.1) cannot be fully correct. On the one hand, the relation accords with the fact that both $\dot{m}_*$ and $\Sigma_{\rm gas}$ peak near $\varpi\,=\,4\ {\rm kpc}$. Considering only this interior region, one would conclude that $N\,\sim\,1$. On the other hand, $\Sigma_{\rm gas}$ is roughly constant for $8\,\lesssim\,\varpi\,\lesssim\,16\ {\rm kpc}$, while $\dot{m}_*$ falls to zero in this interval. It may be that star formation requires a *threshold* surface density, which is too high in the outer region. Indeed, we will see shortly that yet another, global form of the Schmidt law seems to apply to other spiral galaxies, as long as $\Sigma_{\rm gas}$ exceeds some minimum value.

Why the power-law relation in equation (19.1) should hold is not at all clear from a theoretical perspective. In principle, $\dot{m}_*(\varpi)$ must depend on (at least) the production rate of new molecular clouds and how rapidly these create stars before their demise. We have seen in Chapters 5 and 8 that the conversion of HI gas to H_2 is well understood as a *local* process. It is quite another matter to predict the bulk formation rate of cloud complexes out of atomic gas. Our understanding is equally limited concerning the stellar output of such clouds, once created. Schmidt's law must therefore be viewed as a strictly observational result. Its simplicity has made it a key ingredient in Galactic models, especially those of chemical evolution.

19.2 Other Galaxies

As we turn our attention to external galaxies, we are immediately struck by their wide variety in shapes and sizes. These morphological differences correlate, as we will show, with an equally broad dispersion in the pattern of star formation. The connection here is not obvious. Some giant systems of high luminosity are completely inactive with respect to star formation. Conversely, some intrinsically dim galaxies are producing stars at a much faster pace than the Milky Way.

19.2.1 The Hubble Sequence

We first establish the basic categories and nomenclature. Spiral galaxies like the Milky Way contain both a central bulge and an extended, flat disk. Elliptical galaxies, such as NGC 4472 in Figure 19.3, are fully three-dimensional entities, consisting essentially only of a bulge. They display a range of projected shapes, from spheres to cigars. At the other end of the spectrum are irregulars, illustrated in the figure by the Large Magellanic Cloud, an orbiting companion to the Milky Way. This galaxy has traces of spiral structure, not evident in this figure, but there is no central bulge. Between these two extremes are the spirals, classified according to the

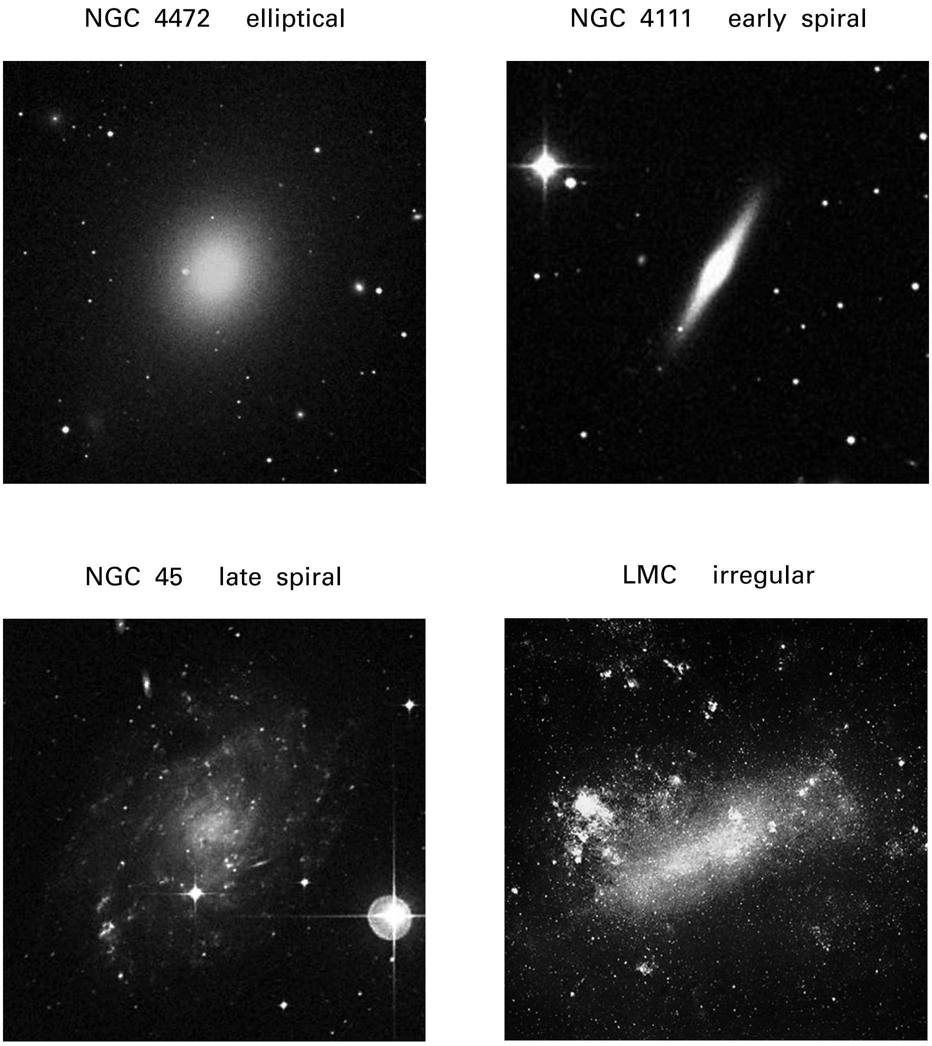


Figure 19.3 Four sample galaxies that span the Hubble sequence. All photographs are optical images.

Hubble scheme as Sa through Sd. The order here is one of decreasing bulge size and increasing openness of the disk's spiral pattern. Figure 19.3 gives examples of an early (Sa) and late-type (Sd) system. We may also recall Figure 1.20 showing M51, a system classified as Sc. Our own Galaxy is thought to fall midway between Sb and Sc in its morphology. Finally, we note the existence of a transitional type (S0) between the most elongated ellipticals and Sa, the earliest true spiral.

This sequence is based solely on the optical appearance of the galaxies. However, the ordering correlates well with other observed characteristics. One of these is the radial velocity of the stars themselves, as measured spectroscopically. Detailed study of these velocities, supplemented by theoretical modeling, shows that the stars within ellipticals are on plunging, nearly radial orbits, whose planes are randomly oriented. The disks of spiral galaxies, as one might expect, consist of objects moving on essentially coplanar, circular orbits. Central bulges again display the more random velocities.

We defer until § 19.4 and § 19.5 the fundamental, and still poorly understood, issue of how the various galaxy types arose cosmologically. At this point, it suffices to remark that the placement of a system within the Hubble sequence does *not* primarily reflect its age, but rather the detailed circumstances of its origin. That is, most relatively nearby galaxies have comparable ages, about 1×10^{10} yr. A given region of primordial gas develops into a certain Hubble type by virtue of its mass and velocity distribution, as well as the amount of its associated dark matter. The latter, we recall, is manifestly still present in the halos of present-day spiral galaxies and actually dominates the total system mass. A key fact, surely relevant for this discussion, is that most ellipticals are found at the centers of galaxy clusters. Spirals, on the other hand, may occur either as isolated objects or in crowded regions.

Although an elliptical and spiral galaxy may have the same age, their star formation histories are entirely different. An interesting way to see this fact is through the variation of the photometric *color* along the Hubble sequence. Figure 19.4 shows the average $B-V$ color index for hundreds of catalogued galaxies, displayed as a function of Hubble type. Ellipticals are the reddest, while the color of spirals becomes progressively bluer. The explanation is that much of the light from early-type galaxies is from relatively old, low-mass stars presently in their red giant phase. Later-type galaxies contain numerous main-sequence stars of high mass and effective temperature. The two panels of Figure 1.20 illustrated, for the case of M51, the predominance of spiral arms in blue light. Massive, hot stars are necessarily young. In summary, star formation activity ended early in the life of elliptical galaxies, but is ongoing in spirals and irregulars. This view of ellipticals is bolstered by the observation, from radio studies, that they are devoid of molecular and atomic gas.

19.2.2 Early and Late Spirals

Return now to spiral galaxies and their integrated, optical spectra. Two examples, representing averages over a number of systems, are given in Figure 19.5. Here we see how the redder color of earlier Hubble types is associated with a depression in the spectrum at shorter wavelengths. This depression is filled in for Sd systems. Another change that occurs along the Hubble sequence is the appearance of strong emission lines, notably [OII]$\lambda\,3727$, an [OIII] doublet near 5000 Å, and Hα at 6563 Å. These arise, not from stellar photospheres (which create the blue continuum), but from HII regions surrounding young, massive stars. The latter are produced with a much larger complement of low-mass objects, in the proportion dictated by the initial mass function. Thus a galaxy's Hα luminosity, in particular, effectively measures its *total* star formation rate and indeed has become the most widely used means of doing so. This technique cannot be used in our own Galaxy, because of the high optical extinction within the plane. Even for external galaxies, one must correct for extinction in more edge-on systems, such as

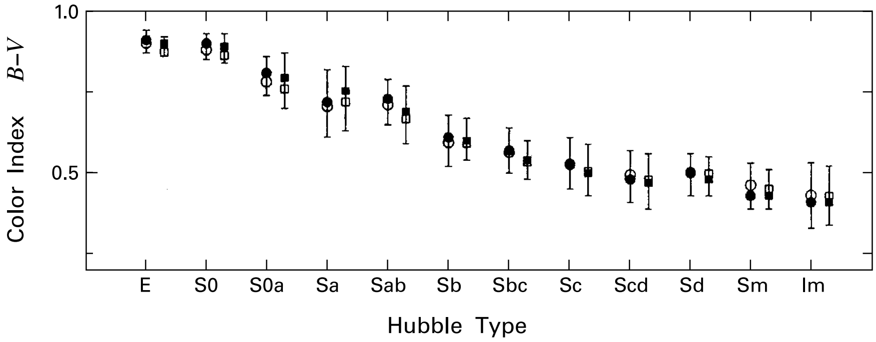


Figure 19.4 The color index $B - V$ of galaxies as a function of their Hubble type. Note that the latter contains intermediate spiral types (such as Sbc), as well as the Magellanic irregulars Sm and Im (see text). The two sets of data points were produced by averaging from two different galaxy catalogs. Error bars represent counting statistics.

NGC 4111 in Figure 19.3. Note finally that other diagnostics at ultraviolet, infrared and radio wavelengths are also used and may serve to complement the Hα data.

Figure 19.6 displays the Hα equivalent widths for many individual spirals, grouped according to Hubble type. Also included, for comparison, are a small number of ellipticals (here labeled E/S0) and irregulars (Sm/Im). As indicated in the plot, the observed fluxes include emission from nearby [NII] lines; this extra component must be subtracted to obtain star formation rates. Recall from equation (17.1) and the accompanying discussion that the equivalent width of a line measures its strength relative to the adjacent continuum. Thus, Figure 19.6 actually gives the intrinsic formation activity per unit galactic luminosity in the red, rather than an absolute rate. In any case, the trend is clear. Ellipticals, as expected, have no star formation at all. Later-type spirals display an increasing *range* in activity, with the maximum level systematically rising along the Hubble sequence. When the data are translated into true star formation rates, these rise, on average, by a factor of 20 from Sa to Sc. A typical $\dot{M}_*$-value for the latter is 10 $M_\odot$ yr^{-1}.

It is also of interest to compare the activity in spirals with their supply of interstellar matter. In the spirit of equation (19.1), Figure 19.7 plots $\langle\dot{m}_*\rangle$, the total star formation rate per unit area of the galactic disk, versus $\langle\Sigma_{\rm gas}\rangle$, the average surface density in diffuse gas. The latter is the total mass in molecular and atomic gas, as measured through CO and HI, respectively, divided by the disk area. Despite appreciable scatter, the points display a linear trend in this logarithmic plot. That is, spiral galaxies roughly obey a global form of the Schmidt law:

$$\langle\dot{m}_*\rangle \;=\; A_1\,\langle\Sigma_{\rm gas}\rangle^{N_1}\;. \tag{19.2}$$

The best-fit curve, also shown in the figure corresponds to $N_1 \;=\; 1.4$.

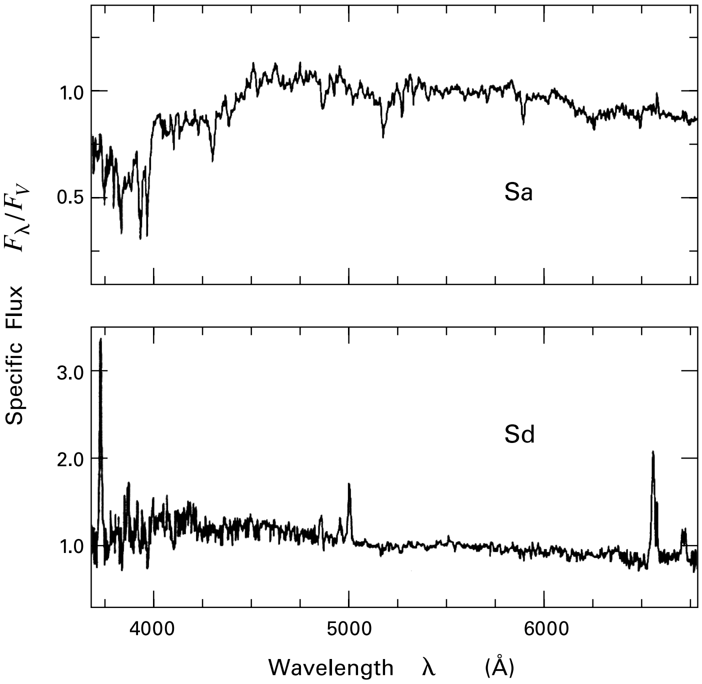


Figure 19.5 Galactic optical spectra along the Hubble sequence. The top panel is an average spectrum for 11 galaxies of type Sa. The bottom represents the average of 7 systems, classified Scd or Sd. In both panels, the specific flux is normalized to the value at 5500 Å.

19.2.3 Density Threshold

Earlier we introduced, when discussing $\dot{m}_*$ in the Milky Way, the idea of a threshold surface density for the gas, below which it is impossible to sustain vigorous star formation activity. This notion actually derives from observations of other, well-resolved spirals, where a spatial limit to the activity is more apparent. Thus, we often see the stellar disk of a galaxy extending beyond the region of Hα emission. The latter exhibits a sharp cutoff in radius. Beyond this break, one still finds individual HII regions, but their surface density is much lower.

Given this finding, we may now recast the issue of threshold density in a more precise manner. Near the boundary of strong Hα emission in any galaxy, we should examine $\Sigma_{\rm crit}$, the total gas surface density. In fact, there is a simple, theoretical prescription for this quantity that turns out to be surprisingly accurate. Suppose we think of the gas within a galaxy as constituting its own disk. Let us further suppose that molecular clouds, and therefore stars, only form at radii where this fluid is gravitationally unstable. The stability of a self-gravitating disk is determined by the Toomre Q-parameter, given in equation (11.41). Setting $Q = 1$ and rearranging terms gives our expression for $\Sigma_{\rm crit}$:

$$\Sigma_{\rm crit} = \frac{\kappa\, a_s}{\pi\, G} \, . \tag{19.3}$$

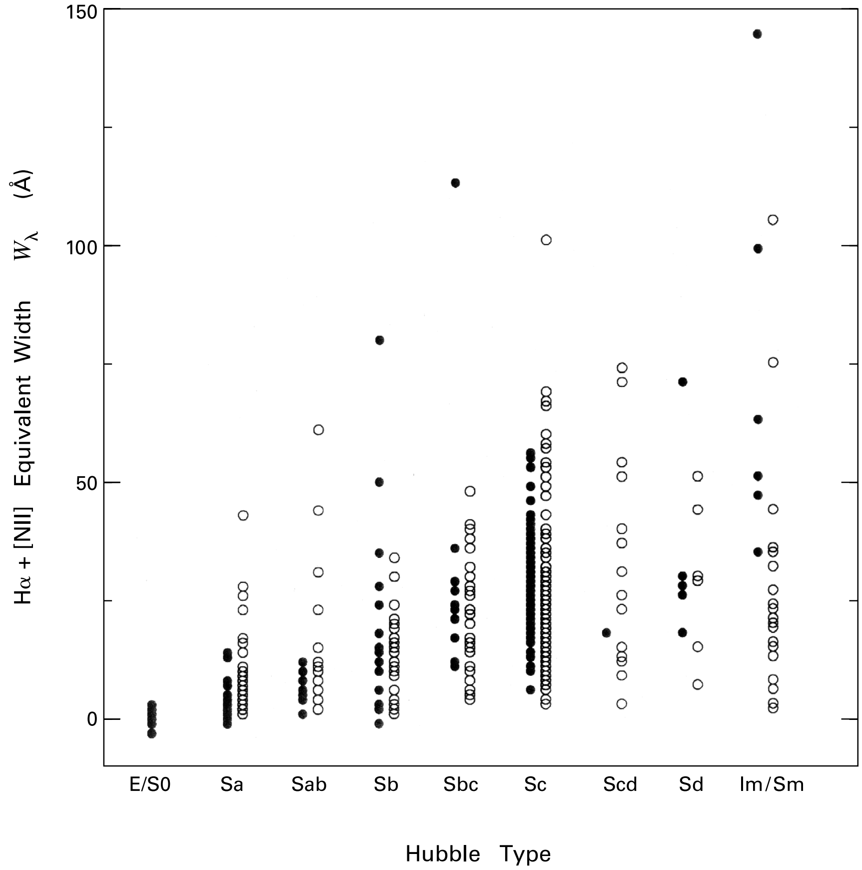


Figure 19.6 Combined equivalent widths of Hα and [NII] for field galaxies (*filled circles*) and those in the Virgo Cluster (*open circles*). The galaxies have been binned by Hubble type. Note again the intermediate types, such as Sab.

We calculate the epicyclic frequency κ from equation (11.42), using for $\Omega(\varpi)$ the observed distribution of rotation speeds in HI gas. Additionally, we reinterpret the sound speed a_s as a weighted mean velocity dispersion of the atomic and molecular material at each radius. This quantity is also supplied by observation.

How well does equation (19.3) account for the star formation threshold? Figure 19.8 shows the run of surface density in a large number of spiral galaxies. Here, $\Sigma_{\rm gas}$ has been normalized to $\Sigma_{\rm crit}$. The radius ϖ has similarly been scaled to $\varpi_{\rm HII}$, the position of the Hα cutoff. While $\Sigma_{\rm gas}$ itself can fall by two orders of magnitude from the center to $\varpi_{\rm HII}$ in a typical galaxy, $\Sigma_{\rm gas}/\Sigma_{\rm crit}$ rarely changes even by a factor of 3. Thus, the interior gas is never very far from marginal stability. Even more striking is the fact the $\Sigma_{\rm gas}/\Sigma_{\rm crit}$ reaches a nearly universal value at the point where $\varpi/\varpi_{\rm HII} = 1$. It appears, then, that equation (19.3) predicts the observed

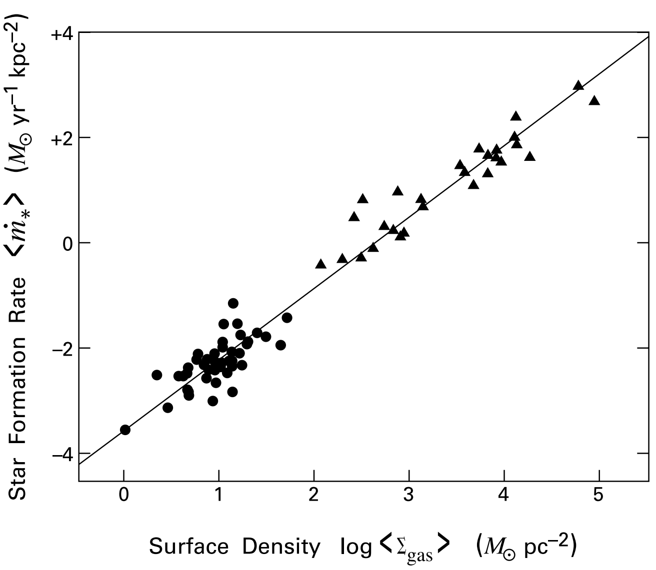


Figure 19.7 Star formation rate per unit disk area as a function of total gas surface density. The circles represent quiescent galaxies, the triangles starburst systems (see § 19.3). For the latter, the gas surface density is the circumnuclear value inferred from CO observations.

truncation rather well, especially if we supply a numerical coefficient (about 0.7) on the right-hand side.

This success notwithstanding, the underlying model leading to equation (19.3) is surely oversimplified and hard to justify in any detail. Diffuse gas represents a small fraction of the mass in a typical spiral. Thus, it is the stars that supply most of the gravitational force on a gaseous fluid element. This fact is clear from the observed distribution of rotation speeds. That is, $\Omega(\varpi)$ obtained at any radius from HI is identical to that inferred from Doppler shifts of the integrated stellar spectra. How can the gravitational stability of the gas then depend on $\Sigma_{\rm gas}$?

There is still no fully satisfactory explanation of why equation (19.3) works as well as it does, but we can at least indicate a key point on which there is consensus. Proper assessment of a galaxy's gravitational stability must indeed take into account both the gas and stars. For the purposes of such an analysis, we may also consider the stars to be a fluid with internal pressure. The latter stems from the stars' finite velocity dispersion, *i. e.*, their departure from perfectly circular motion. Moreover, the magnitude of this dispersion, as high as 35 km s^{-1} in our own Galaxy, is significantly greater than that associated with the gas. (This is about 8 km s^{-1} locally in the Milky Way.) But it is the pressure of each fluid component, along with its rotation, that opposes gravity. Hence, the stability of the combined star-gas system is sensitive to the gas, even if its fractional mass is low.

The view that star formation ultimately arises from the gravitational instability of a galaxy is an appealing one that has stimulated much research. Unfortunately, there remains a significant gap between global calculations in this vein and those pertaining to the evolution and collapse

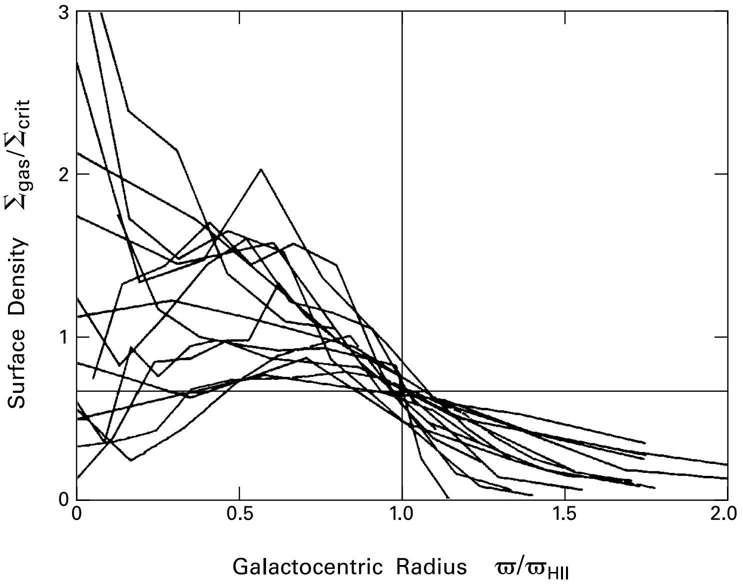


Figure 19.8 Total gas surface density as a function of radius in the disks of spiral galaxies. The surface density is normalized to the critical value, and the radius to the star formation threshold, as measured by Hα. The straight, horizontal line corresponds to a numerical coefficient of 0.67 on the lefthand side of equation (19.3).

of individual clouds. One important element in the picture, at least for the galaxies we are now considering, is that self-gravity drives the formation of spiral arms and would do so even if there were no gas. We know from observation that HI concentrates along the arms, presumably settling into the gravitational trough they provide. The resulting increase of column density promotes self-shielding and the creation of molecular gas. High-resolution studies of CO find that it too exhibits spiral structure (Figure 1.21). The azimuthally averaged CO intensity always first rises toward the center and then may decline, sometimes peaking at $\varpi = 0$. Our own molecular ring thus finds its analog in these other galaxies.

We have noted that Hα is produced, at least indirectly, by short-lived, massive stars. Any assessment of formation activity based on this technique refers only to present-day conditions. What do we know concerning the *history* of activity on the galactic scale? To address this issue, the most widely used method is *population synthesis modeling*. The idea here is to combine the light from a range of stellar photospheres to match a galaxy's observed spectrum, *i. e.*, both its distribution of broadband fluxes and discrete absorption lines. Accounting for galactic *emission* lines, on the other hand, is more difficult, as it requires a model of interstellar clouds at each epoch.

In any case, one effects the match by first obtaining, from theory, the fluxes from individual stars as a function of their mass and evolutionary age τ. Holding τ fixed, one adds together many stellar fluxes, weighted according to the initial mass function, and obtains $f_\lambda(\tau)$, the galactic spectral energy distribution at each isochrone. The true predicted flux at time t follows by utilizing the star formation rate $\dot{M}_*(t)$ and integrating over the single-isochrone results:

$$F_\lambda(t) \,=\, A \int_0^t \dot{M}_*(t-\tau)\, f_\lambda(\tau)\, d\tau \;. \tag{19.4}$$

Here, A is a normalization constant, with units $M_\odot^{-1}$. The integral at any time t has contributions from many epochs. Physically, the observed spectral energy distribution is sensitive to previous history, since old stars of sufficiently low mass are still present. The practice has been to fix the present galactic age and then vary the star formation rate, assuming a simple exponential time dependence. It is also common to adopt as a free parameter the *initial metallicity* of the gas, which is naturally well below solar. The metallicity is updated by assuming prompt mixing of stellar winds and supernova remnants with previously existing gas.

The best-fit parameters obtained through this modeling indicate significant variation of the histories even among galaxies of the same Hubble type. Nevertheless, there is a definite trend along the sequence. In earlier-type spirals, the present-day rate of star formation is as much as an order of magnitude *lower* than the rate averaged over the entire galactic lifetime. In Sc galaxies, the current and average rates are closer, implying a relatively slow change. The simplified assumption of a temporally constant Galactic star formation rate, used in deriving the initial mass function, is therefore not unreasonable. In any case, the pattern of histories among spirals is consistent with our previous remarks. Considering the extreme ends of the sequence, ellipticals have no present-day star formation, while this rate exceeds the average in gas-rich irregulars. However, the findings still present a puzzle. Every Sc galaxy must have been predominantly gaseous early in its history. Presumably, $\langle\Sigma_{\rm gas}\rangle$ declined with time. But then a Schmidt-type law at each epoch would be inconsistent with a steady consumption rate. A significant caveat is the possibility, to be discussed later, that the galaxy continued to accrete fresh gas while it was making stars.

19.2.4 Activity in Irregulars

Optical images of galaxies with relatively high, current values of $\dot{M}_*$ show a larger number of HII regions. Moreover, the average *size* of each region is greater. In a typical Sa galaxy, a single patch of Hα emission requires only a handful of O stars to provide the observed luminosity. This number rises to hundreds in an Sc-type system. As we proceed down the Hubble sequence into the realm of irregular galaxies, we find that giant HII regions predominate. Indeed, the mottled appearance of irregulars largely reflects the asymmetric distribution of massive star formation sites, within an underlying, disk-like geometry.

The prototypical example of an irregular galaxy is the Large Magellanic Cloud (Figure 19.3). Located relatively close to the Milky Way (50 kpc), it is one of the best studied systems in its class. The total mass within the brighter optical portion, as indicated by the speed of outlying HI gas, is $4 \times 10^9\ M_\odot$. This figure is a few percent that for a typical spiral. Irregulars generally are also among the dimmer galaxies. Nevertheless, these low-luminosity objects of unimposing size and mass constitute a large fraction, at least one third, of all catalogued systems. Their high degree of star formation activity is reflected not only in their blue colors (Figure 19.4), but more directly, as just indicated, by the presence of giant HII regions.

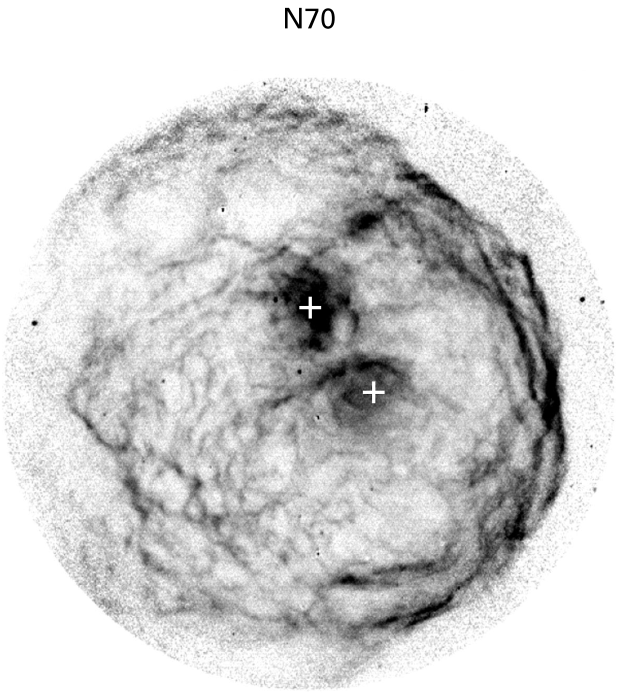


Figure 19.9 Image in Hα of the N70 shell, created by winds from an OB association in the Large Magellanic Cloud. The shell radius is 55 pc. Two O stars are marked by crosses.

Within the Large Magellanic Cloud, the most impressive such region is 30 Doradus. This is the bright patch situated to the upper left of the galaxy's main body in Figure 19.3. (Recall also Plate 8.) The entire complex is 1 kpc in diameter. It contains over 100 O stars, including an especially dense concentration in the area known as R136. Images in Hα show the presence of many shells, some enclosing cavities that are glowing in X-rays. Much of this hot gas has been energized by supernova explosions. Other expanding shells are created by winds from *Wolf-Rayet stars*. These are massive, post-main-sequence objects that have already shed much of their hydrogen-enriched mantles of gas. The exposed helium core emits a fast wind that shocks as it plows into the previously ejected mantle. Figure 19.9 displays an example, from outside 30 Doradus itself, of an especially symmetric, wind-driven shell. Here, gas streaming from an OB association impacts a surrounding, HI cloud.

Of course, fast-moving shells, hot gas, and other energetic phenomena are only the most spectacular manifestations of a process involving many more, lower-mass objects. The Large Magellanic Cloud is near enough that this larger population is accessible through direct imaging. Hundreds of classical T Tauri stars have now been resolved in Hα, and detailed spectroscopy of individual objects has also begun. From such studies and others, we infer a metallicity averaging only one fourth the solar value, yet another indication of low prior activity. The copious production of new stars accords with the quantity of interstellar gas, comprising some

15 percent of the total galactic mass. Curiously, nearly all the diffuse matter is HI; the CO luminosity indicates only a small molecular component.

The giant HII regions of the Large Magellanic Cloud, while striking in terms of their productivity, are in essence no different from many others observed in late-type spiral galaxies. It is apparent that, once a region attains a density sufficient for rapid production of massive stars, the nature of the larger environment fades in significance. A second valuable lesson is that vigorous star formation in a galaxy does *not* require well-delineated spiral arms. As mentioned earlier in this section, such structure is actually present in the Large Magellanic Cloud, but the arms are fragmentary and of relatively low contrast.[4]

Most other irregular galaxies resemble, in their broad characteristics, the two closest examples, the Large Magellanic Cloud and the even more HI-rich Small Magellanic Cloud. A minority, the so-called *amorphous* irregulars, present a much smoother optical appearance. One should not infer, however, an absence of HII regions. The integrated spectra still have conspicuous emission lines, and Hα luminosities, in particular, are relatively high. A safer conclusion is that star formation is proceeding strongly, as in the Magellanic-type irregulars, but is more diffuse spatially. Activity is also strong in *blue compact dwarfs*. Considered a third subclass of irregulars, these are small systems with both exceptionally blue photometric color and strong line emission.

19.2.5 The Local Group

The Magellanic Clouds are not only our galactic neighbors, but also the brightest systems of the three dozen that make up the *Local Group*. This gravitationally bound entity, of which the Milky Way is a prominent member, extends out to a radius of nearly 2 Mpc. The third brightest system is the Andromeda Galaxy (M31), a spiral of type Sb. Andromeda and the Milky Way together account for nearly all of the group's visible mass. Remaining members consist of a third spiral (M33) and a collection of smaller, low-luminosity (*i. e.*, dwarf) irregulars and ellipticals. For the most part, the minor galaxies are clustered about the two massive members. Those ellipticals without a strong, central peak in stellar density are also known as *dwarf spheroidals*. Some two dozen Local Group systems fall into this latter category. Table 19.1 lists essential properties of some nearby galaxies. Note that we have tabulated the luminosity in the B-band, a conventional measure of massive stellar content.

All galaxies in the Local Group are close enough that we can spatially resolve individual stars of high and intermediate mass, both on and off the main sequence. Color-magnitude diagrams may then be constructed. Within these, one may discern separate populations of differing age. Spectroscopy of the brightest objects also yields the metallicity of the subgroups. In other words, one may read off the galaxy's detailed history, both in terms of stellar production and chemical enrichment.

This basic strategy is clear enough, but the interpretation of color-magnitude data is not a simple exercise. One must accommodate inevitable errors in the photometry, patchy extinction across the galaxy, and a distance that is usually far from secure. A valuable approach is to

---

[4] Irregulars also lack the opaque dust lanes conspicuous in spiral galaxies. A paucity of grains is consistent with the reduced metallicity of these systems. Note that ambient ultraviolet radiation is more destructive in a low-dust medium. We can thus understand the small amount of molecular gas found in these systems.

Table 19.1 Properties of Some Nearby Galaxies

Name	Type	Distance (kpc)	L_B ($\mathrm{L}_\odot$)	M_HI ($\mathrm{M}_\odot$)	M_{H_2} ($\mathrm{M}_\odot$)	M_tot ($\mathrm{M}_\odot$)
Milky Way	Sbc		2×10^{10}	6×10^9	2×10^9	2×10^{11}
LMC	Im	50	3×10^9	7×10^8	6×10^7	1×10^{10}
SMC	Im	60	9×10^8	6×10^8	4×10^6	9×10^8
Leo I	dSph	250	1×10^6	$< 10^4$		2×10^7
NGC 185	dE	630	6×10^8	1×10^5	4×10^4	7×10^8
M31	Sb	750	6×10^{10}	5×10^9	5×10^8	3×10^{11}
M33	Scd	840	6×10^9	2×10^9	2×10^8	5×10^{10}

Note that the luminosity is in the B-band only (spectral width 1000 Å). The total masses for all systems except Leo I and the Small Magellanic Cloud are based on the HI rotational velocity, and thus include dark matter.

search for characteristic features in the diagram or special classes of stars. For example, a horizontal branch well above the main sequence indicates that the corresponding population is of low metallicity and older than 10^{10} yr. The slope and position of the red giant branch are also sensitive to metallicity. From the presence of numerous O and early-B stars, one may conclude that stars have formed recently, *i. e.*, within the last 2×10^7 yr. On the other hand, a concurrent *absence* of Wolf-Rayet stars means that this activity ceased more than 6×10^6 yr ago, which is the time needed for such objects to fade from view.

Figure 19.10 shows a color-magnitude diagram (using the V- and I-bands) for Leo I, a dwarf spheroidal companion to the Milky Way. We have included this system in Table 19.1, where it is conspicuous for its low total mass and luminosity. The diagram, constructed from 28,000 observed stars, displays both a main sequence and red giant branch. The extraordinary thickness of the main sequence largely reflects random errors in the photometry, which have significant impact at the lowest magnitudes. Nevertheless, there is still a well-defined turnoff; this translates into a minimum stellar age of 1×10^9 yr.

The red giant branch of Leo I consists of stars with intermediate ages, *i. e.*, from 1×10^9 to 1×10^{10} yr. Those undergoing central helium burning form the prominent clump to the left. Note also that the whole red giant branch is shifted toward the blue with respect to stars of the solar neighborhood. This change indicates a reduced metallicity. The single, best-fit Z-value, used to construct the star formation history in the righthand panel, is 4×10^{-4}, or 2 percent of the solar figure. As presented here, the history is actually a record of $\dot{M}_*(t)$, constructed from the stellar ages. Note the low level of star formation in the remote past. Activity during this epoch would have manifested itself as a horizontal branch in the color-magnitude diagram.

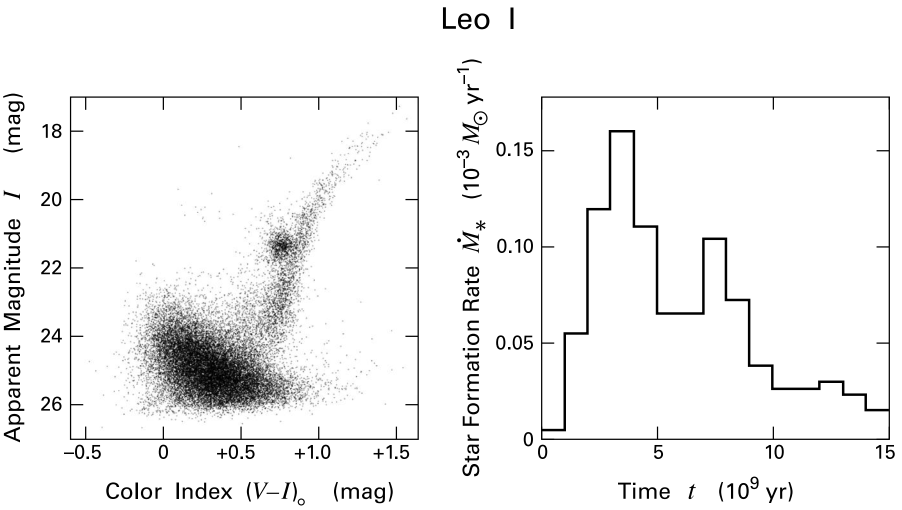


Figure 19.10 (*left panel*:) Color-magnitude diagram for 28,000 stars in the Leo I galaxy. Both the color and I-magnitude have been corrected using a spatially averaged reddening of $E(V - I) = 0.04$ mag. (*right panel*:) Derived global star formation rate as a function of time t into the past.

The fairly recent spike in star formation is curious, as there is no detected interstellar gas at the present time (see Table 19.1).

A very different history is displayed in Figure 19.11, representing data from another Local Group member, NGC 185. This dwarf elliptical satellite of Andromeda has a strong central concentration of stars. The color-magnitude diagram, based on 16,000 objects, includes both the central region and less crowded areas, for which the photometric accuracy is higher. Both the main sequence and horizontal branch are below the flux limit in this diagram, which consists essentially of a thickened red giant branch. Here reside low-mass objects with ages of 1×10^{10} yr and greater. Accordingly, the reconstructed star formation history shows a strong, initial burst of activity, followed by a steep decline. Observations show a small, but nonzero amount of atomic and molecular gas in the central region (Table 19.1). This material is not primordial, but was ejected by stars passing through their planetary nebula phase.

Similar analysis of the remaining Local Group systems reveals a broad variety in star formation histories. The diversity itself may be a consequence of the low galactic masses. As in the case of NGC 185, fresh gas regenerated from dying stars often constitutes the bulk of the interstellar medium at any epoch. This gas is vulnerable to dispersion by a few supernovae. There will then be a lag in star formation activity, which could end when more gas is supplied by the aging stellar population. Such intermittent behavior on a global scale would not occur in a galaxy with greater mass and deeper gravitational potential, which can retain the bulk of its gas over longer periods.

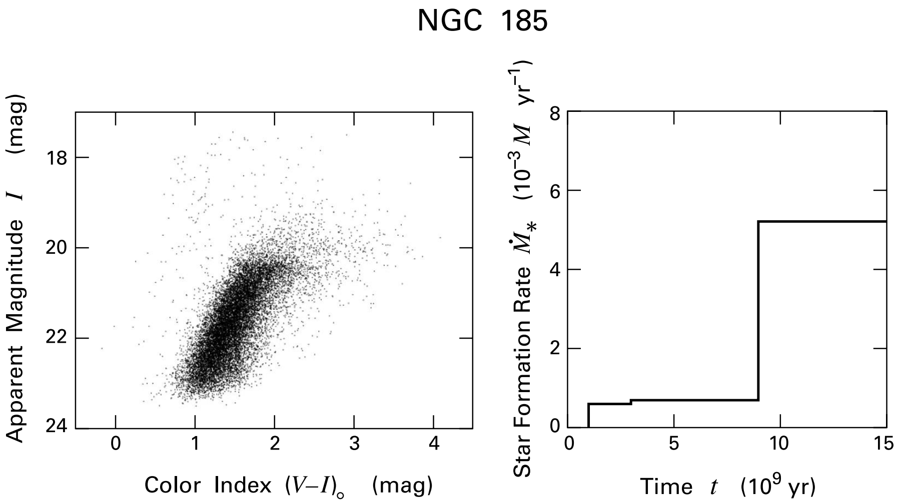


Figure 19.11 (*left panel*:) Color-magnitude diagram for 16,000 stars in NGC 185. As in Figure 19.10, both colors and magnitudes have been dereddened, here using $E(V - I) = 0.22$ mag. (*right panel*:) Star formation rate as a function of time into the past.

19.3 The Starburst Phenomenon

We have seen that the majority of galaxies within our own cluster are objects whose mass and luminosity are very low compared to the Milky Way or M31. This preponderance of dwarf systems appears to hold in more distant regions, as well. Morphologically, we again find dwarfs outside the Local Group to be either ellipticals or irregulars. What little interstellar gas resides in the first type is presumably generated by the aging stellar population. The dwarf irregulars, like those nearby, have copious gas, with the material largely being HI. This gas is of low metallicity, typically 10 percent of the solar value. Thus, there has not been a great deal of nucleosynthesis through successive generations of stars.

On the other hand, many dwarf irregulars do contain localized sites that are not only actively producing new stars, but at extraordinary rates. Such a *starburst* region, with a diameter of perhaps 300 pc, has an associated $\dot{M}_*$ exceeding that of the entire Milky Way by one to two orders of magnitude. Note that the host galaxy itself spans only a few kpc, but often contains several of these regions. Other, much more massive, galaxies may also harbor starbursts, as we shall describe. In all cases, the limited supply of gas tells us that this activity could not have been sustained for a long period. Population synthesis models give quantitative estimates of the duration, typically 10^8 yr.

19.3.1 Anomalous Spectra and Colors

Starbursts are most conspicuous in dwarf irregulars because of the relatively low background flux from old stars. Hence, we begin our survey with these gas-rich but metal-poor systems. Optical spectra of starbursts are characterized by strong emission lines from hydrogen and other ionized elements. These are the same lines displayed by late-type spirals (Figure 19.5), but now with higher equivalent widths because of the greater contrast. The lines are again created in the gaseous envelopes around O and early-B stars, of which there may be thousands in a single site. Indeed, the spectrum of a starburst closely resembles that of any nearby HII region, despite the difference in physical scale. Starburst irregulars where the background light is undetectably faint have accordingly been dubbed "HII galaxies" or "extragalactic HII regions."

In our description of the nearest true starburst region, 30 Doradus, we noted the presence of Wolf-Rayet stars. Here, we recall, the surrounding envelope has been supplied by the stellar wind itself. Ionized species within the envelope generate emission lines, as in HII regions, but there is a difference. Because of the high wind speed, the observed lines are significantly broadened. Such characteristic emission, generally in the 4686 Å line of He II, has been used to identify over 100 starburst irregular galaxies. In particular, these are systems where vigorous star formation began 3 to 6×10^6 yr in the past, corresponding to the predicted interval for the phenomenon in an individual, massive star.

Another point to recall about 30 Doradus is that it contains a bright core, designated R136. This is the compact region, only 5 pc in diameter, that includes the most massive stars, along with a large complement of dimmer objects. Analogous clusters, some of even greater density, are located in more distant galaxies. For example, the dwarf irregular NGC 1569 has two such luminous knots (see Figure 19.12). In both locales, we detect the broadened He II emission line from Wolf-Rayet stars. Other starburst regions probably have a number of these tight formations, not yet spatially resolved. If the groups remain bound after gas dispersal, they may slowly evolve into globular clusters like those orbiting within our own halo.

Starbursts in larger galaxies were first recognized, in the 1970s, by their effect on the integrated system *color*. The argument here is of more than just historical interest and bears repeating. Figure 19.4 showed that the $B - V$ colors of galaxies along the Hubble sequence, while exhibiting a scatter at each type, nevertheless change systematically. This variation reflects the different rates at which global star formation declines in each type. A similar plot holds for any other color index, such as $U - B$. Thus, there is a *correlation* between a galaxy's $U - B$ and $B - V$ colors. The empirical relationship is apparent in the left panel of Figure 19.13. Each point here represents a galaxy taken from a magnitude-limited survey.

We next consider galaxies that, while they have luminosities comparable to the first set, are morphologically peculiar in some way. The spirals display narrow, asymmetric features, variously labeled as tails, filaments, or bridges, in addition to a standard bulge and disk. We shall later see some examples. The important point for now is that the color-color diagram for this new group, shown in the right panel of Figure 19.13, is also peculiar. While some galaxies still lie close to the average curve representing the normal $(U - B, B - V)$ relation, others lie well above it. At a given $B - V$ value, these systems have $U - B$ indices that are low, *i. e.*, the galaxies are *anomalously* blue.

A blue galaxy quite generally is one with a significant portion of hot, massive objects. Thus, it is actively forming stars. In our peculiar group, however, this activity must be relatively short-lived, so that a growing, older population does not change the colors and bring the $U - B$ index back to its normal range. Population synthesis modeling, either of the $(U - B, B - V)$ values or of the optical spectrum, confirms that the surge in formation activity cannot last longer than

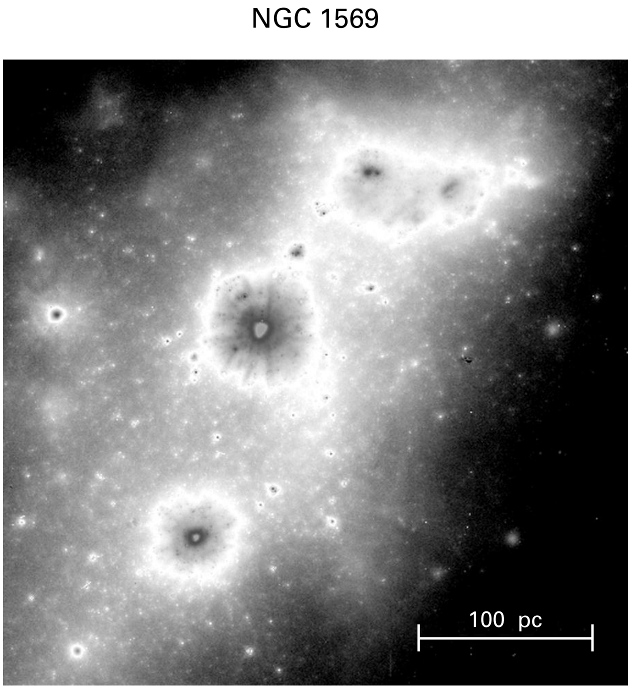


Figure 19.12 Central portion of the dwarf irregular NGC 1569, as seen in the V-band. The dark rings around each of the two lower clusters are an artifact of the false colors used in the original image. The more distended patch at the top is a complex of HII regions.

about 10^7 yr. Moreover, the event cannot have *ended* much more than 10^8 yr in the past. In other words, many of the peculiar galaxies have ongoing or comparatively recent starbursts.

Optical imaging reveals the activity to be occurring close to the galaxy's center. This circumnuclear region, with a radius of less than a kpc, is blue in color and gives the galaxy as a whole its unusual photometric flux ratios. Moreover, a spectrum of the area shows the same emission lines that we find in both HII regions and the starbursts of dwarf irregulars. We also detect nonthermal radio continuum flux at centimeter wavelengths, the signature of many supernova remnants ploughing through the local interstellar medium. It is clear that we are again witnessing copious production of stars in a restricted volume.

Such production requires, of course, a certain amount of fuel. Return to Figure 19.7, where the filled triangles now represent individual spiral galaxies harboring circumnuclear starbursts. The inferred star formation rate per disk area, which stems almost entirely from the central region, is evidently correlated with the total gas surface density. Interestingly, the points still lie close to the line representing the Schmidt law for ordinary spirals. This result, besides corroborating the general, underlying relationship, also begins to clarify the nature of starbursts themselves. The key question is how the system accumulates a large amount of interstellar matter in a localized area, and in a relatively brief time. Note that the Schmidt law exponent N_1

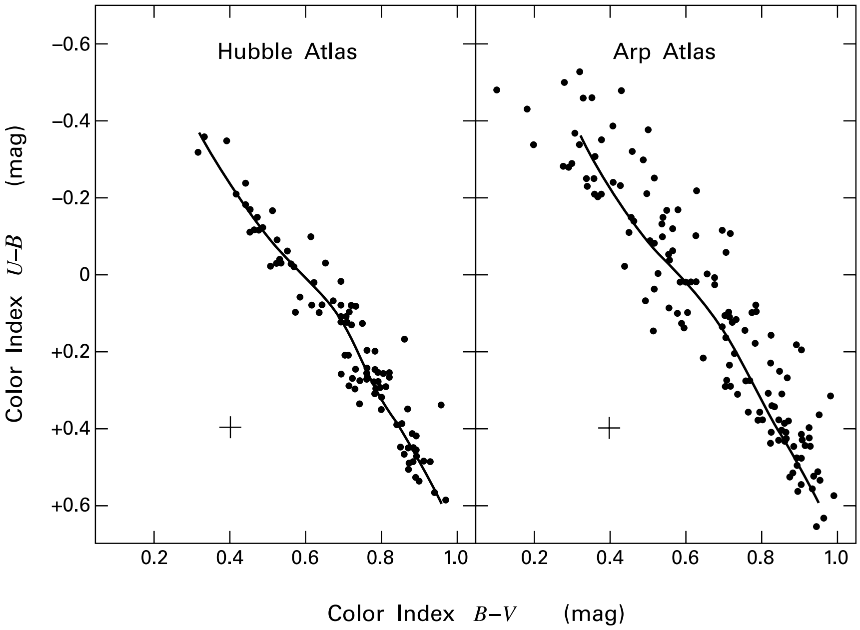


Figure 19.13 (*left panel*:) Color-color plot of galaxies from the Hubble Atlas; these systems are, for the most part, normal spirals. (*right panel*:) The same plot, but for galaxies from the Arp Atlas; these are all morphologically peculiar. The solid curve in both panels represents a smoothed average of the color-color relation for normal systems. Crosshairs indicate the mean error in both color indices.

is not far above unity. The gas, once in place, thus forms stars at a rate *per unit mass* that does not greatly exceed that in a normal spiral.

19.3.2 Infrared Galaxies

The actual values of $\dot{M}_*$ for the starburst systems in Figure 19.7 were not obtained from the Hα luminosity, as in the other sample. Dust extinction in the circumstellar region is sufficiently high that it becomes more convenient (and accurate) to utilize instead the far-infrared component of the flux. We assume that this longer-wavelength radiation stems from grains that are heated, principally by massive stars. Thus, the far-infrared luminosity serves, as did Hα, to identify the high-mass population, and thereby to set the total stellar birthrate.

Among galaxies exhibiting starburst activity, the proportion of the energy output at infrared wavelengths rises dramatically with the total luminosity. The most extreme systems were therefore not discovered optically, but through the infrared satellite IRAS, in the mid 1980s. Only later did optical study reveal these galaxies to have the emission-line circumnuclear regions characteristic of starbursts.

The general trend toward higher infrared brightness is illustrated by Figure 19.14. Here we reproduce the spectral energy distributions of three well-studied galaxies. In the spiral M51,

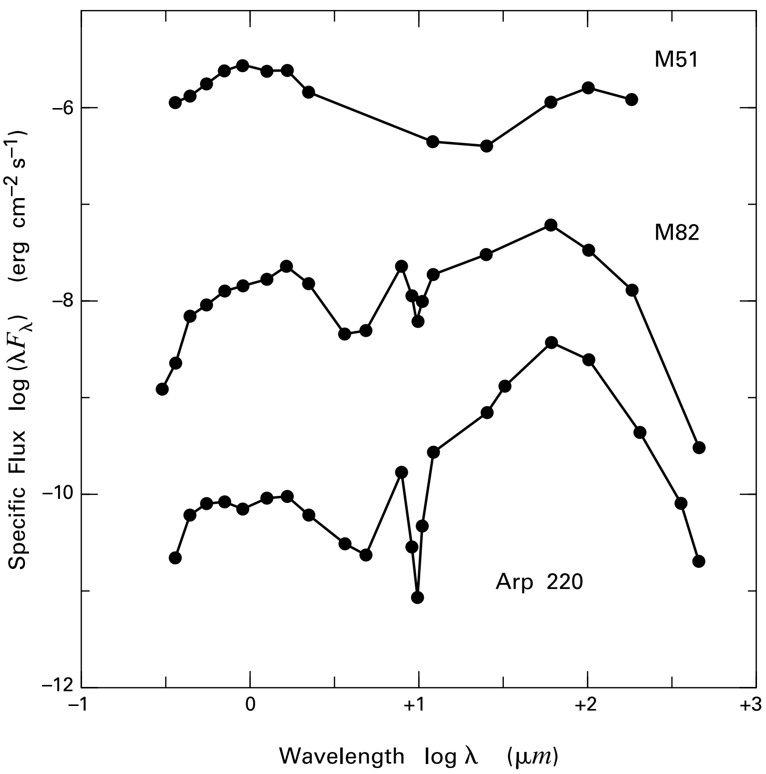


Figure 19.14 Spectral energy distributions of three galaxies. The plot for M51 has been shifted vertically upward by $\Delta\log\lambda\, F_\lambda\ =\ +2$, while Arp 220 has been shifted downward by 0.5. Note the prominent silicate absorption features near 10 μm in M82 and Arp 220.

we see that the flux has two major peaks, in the near- and far-infrared; these have comparable strength. The long-wavelength portion begins to dominate in M82, an edge-on irregular just outside the Local Group. This system has $L_{\rm IR}\ >\ 10^{11}\,L_\odot$ and exemplifies the class designated *luminous infrared galaxies*. Here, the quantity $L_{\rm IR}$ refers to the energy output between 8 and 1000 μm. Those rare galaxies, like Arp 220, with $L_{\rm IR}\ >\ 10^{12}\,L_\odot$ are classified as *ultraluminous*. The total star formation rate in this latter group approaches $10^3\ M_\odot$ yr^{-1}, while $\lambda\, F_\lambda$ at 100 μm exceeds that at 1 μm by two orders of magnitude.

Most star formation activity of infrared galaxies, and indeed most of their long-wavelength radiation, comes from the dusty, circumnuclear region. One would expect that the steep, central concentration of grains should be accompanied by plentiful gas, and this appears to be the case. Interferometric measurements of ^{12}C^{16}O $(J = 1 \rightarrow 0)$ emission reveal that the flux peaks strongly toward the galactic center. New stars are being formed within what is essentially an inner disk of gas and dust.

While this qualitative picture is well established, accurate assessment of the *mass* in central gas, and hence the star formation efficiency, has not been straightforward. Using the CO as a tracer for the hydrogen content, *i. e.*, application of the X-factor introduced in equation (6.14), leads to high column densities. In some cases, including M82, the integrated mass is clearly too large, as it exceeds the virial theorem estimate using the width of the CO lines themselves. The implication is that there must be less hydrogen for a given CO flux. In other words, the CO that is present radiates more strongly, probably because its excitation temperature is higher than in our own Galactic disk, where X is calibrated. The difficulty is how to account for this effect in a systematic way. It is still unclear how the correction alters various estimates of the circumnuclear surface density, including those employed for the starburst systems in Figure 19.7.

Nevertheless, other evidence attests to the elevated *volumetric* gas density near the centers of starburst galaxies. Observations of molecular tracers such as CS and HCN yield typical values for $n(\mathrm{H_2})$ between 10^5 and 10^7 cm^{-3}, while kinetic temperatures range from 60 to 90 K. Both figures are reminiscent of the molecular cloud Sgr B2 in our own Galaxy. The CO mapping, on the other hand, indicates that the filling factor of such entities is higher in the starburst systems, and that the clouds persist out to a larger galactocentric radius.

The proliferation of massive and luminous stars within such a dense environment is conducive to maser activity. Indeed, strong emission in the main lines of OH, at 1665 and 1667 MHz, was first seen toward Arp 220 in 1982. Since then, numerous other starburst galaxies have yielded positive detections. Radio interferometers have shown that much of the emission in Arp 220 comes from several, parsec-size spots. In all systems, this emission is very high, with a flux equivalent to an isotropic luminosity approaching $L_{\mathrm{OH}} = 10^4\, L_\odot$. Since this figure exceeds by a factor of 10^6 the luminosities of even the most powerful Galactic sources, infrared galaxies are said to power OH *megamasers*. The copious, far-infrared radiation from dust may act as the pump, enabling the OH molecules to amplify background, radio continuum flux. The latter plausibly stems from supernova remnants in the vicinity.

What accounts for the impressive, if transient, energetics displayed by starburst galaxies? At the very highest luminosities, the pattern of emission lines in the optical and near-infrared begins to resemble that in active galactic nuclei and the closely related quasars. Some of the power in these cases therefore stems from accretion onto a central black hole. The bulk of the luminosity from most infrared galaxies, however, is from massive, young stars. A basic problem, as mentioned earlier, is how the molecular fuel for these stars is transported to the circumnuclear region. Even with reasonable corrections for the CO flux conversion, the diffuse mass within this region is of order $10^{10}\, M_\odot$. This figure may well represent the entire interstellar medium of the system, gathered into the central kpc or so in radius.

19.3.3 Mergers and Tidal Interactions

To make headway on the transport issue, let us amplify our previous remark that starburst galaxies are morphologically “peculiar.” We begin with Arp 220. At a distance of 75 Mpc, this is one of the nearest ultraluminous systems. Optical emission, while dim compared to the far-infrared, is nonetheless revealing (see Figure 19.15). Most light comes from the more or less spherical area in the center. We also see two plumes of lower intensity to the right. Another, related feature becomes apparent once we probe the central region using higher-resolution, near-infrared imaging. Here, we are sampling a mix of stars, many of them red supergiants. As displayed in the figure inset, there is not just one peak in stellar density, but at least two. Radio continuum maps also show the two strongest concentrations, separated by a mere 360 pc. The intriguing suggestion is that Arp 220 is not a single galaxy at all, but a pair caught in the act of merging. The optical plumes are then a byproduct of the merger process.

This general picture is assuredly not established solely by the present example. However, it quickly gains credence as we look at additional starburst systems. The left panel of Figure 19.16 is an optical image of The Antennae (Arp 244). Two bright lobes are evident, as are a remarkable pair of faint and thin tails. The lobes, it is clear from spectroscopic examination, are the main bodies of two merging, disk galaxies. Their current separation is 15 kpc. The tails are also composed of stars, as well as HI gas.

The righthand panel of Figure 19.16 is a closer view, taken in the near-infrared, of the interaction region. Here we see much more clearly the two galactic nuclei, each with attendant spiral arms. We also see a stellar bridge connecting the two galaxies. Each clump within this bridge and beyond it is a massive cluster. Population synthesis models yield a range of cluster ages, from less than 10^7 yr, within the bridge itself, to a maximum of 5×10^8 yr. Interestingly, the youngest groups are larger by an order of magnitude than the older ones, which in turn resemble globular clusters. The region between the galactic nuclei also has a strong concentration of molecular gas.

Surveys of many tight galaxy pairs find that their total star formation rate, as measured by the far-infrared luminosity, increases dramatically with smaller galaxy separation. Conversely, the fraction of galaxies that are evidently interacting with a neighbor climbs with $L_{\rm FIR}$ and is almost unity for $L_{\rm FIR} > 10^{12}\,L_\odot$. We *cannot* draw the conclusion from these results that *all* starbursts are created by mergers. Recall the dwarf irregulars, many of which exhibit similar activity and yet appear relatively isolated. What the studies do show clearly is that dynamical interaction is an effective trigger of galactic starbursts, probably the most significant one for spirals. How, then, does the close passage of two galaxies stimulate formation of stars? What is the origin of the structural changes accompanying this event?

We begin with the second issue, which is more completely understood. A long series of numerical simulations, beginning in the 1970s, has demonstrated that bridges, tails, and related features are streams of stars and gas, torn away from each galaxy by the gravity of its companion. The key point to recognize is the *tidal* nature of the interaction. Each galaxy, besides pulling bodily on its neighbor, stretches it along the line joining the two centers of mass. Stars closest to the approaching system are drawn toward it, creating a bridge. Stars diametrically opposite are pushed *outward*, relative to the center of mass, and may fail to complete their orbits, instead flying off as tails. Both effects are strongest when the approaching galaxy is orbiting in the same sense as the target's internal rotation. In addition, the response peaks for those stars in the target galaxy whose angular speed closely matches the highest orbital angular speed of the neighbor. This resonance condition explains the striking narrowness of many observed streams.

All close interactions between galaxies do not create starbursts. The distended arm of M51 (Figure 1.20), reaching toward its smaller companion, is yet another tidal bridge. The main galaxy, however, is a normal Sc, without significantly enhanced star formation. A second general point is that any induced, narrow features are transient. The stars in distended tails escape both systems, while bridge material is either captured by the companion or falls back to the galaxy from which it was originally drawn. Finally, the numerical studies make it clear that

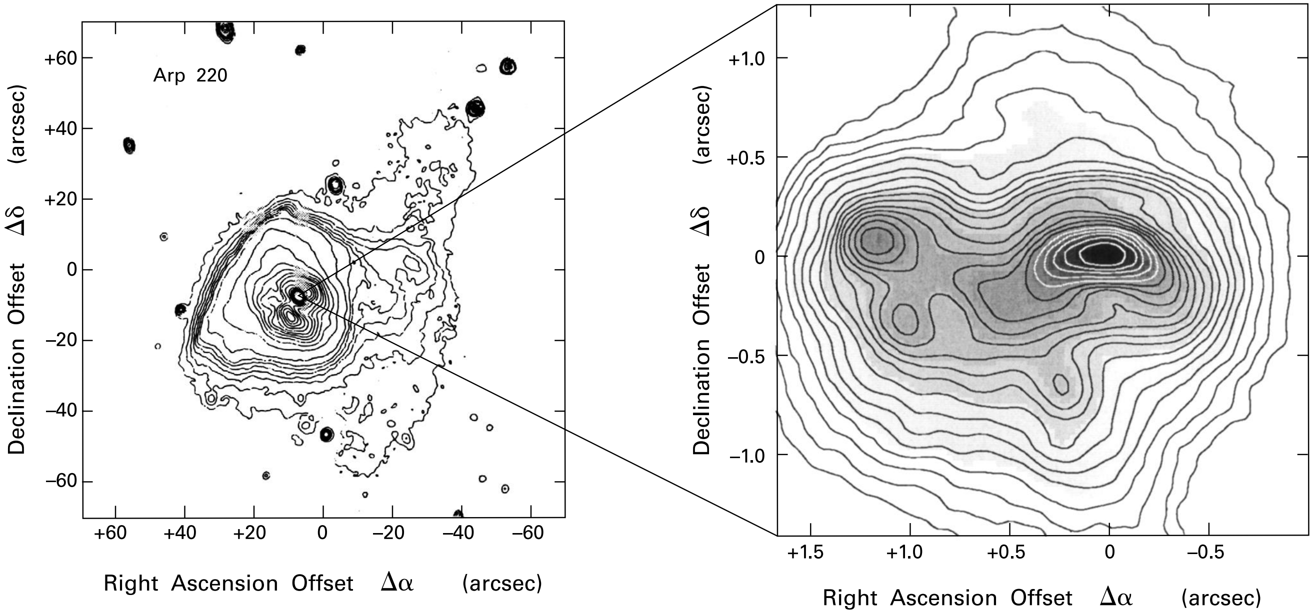
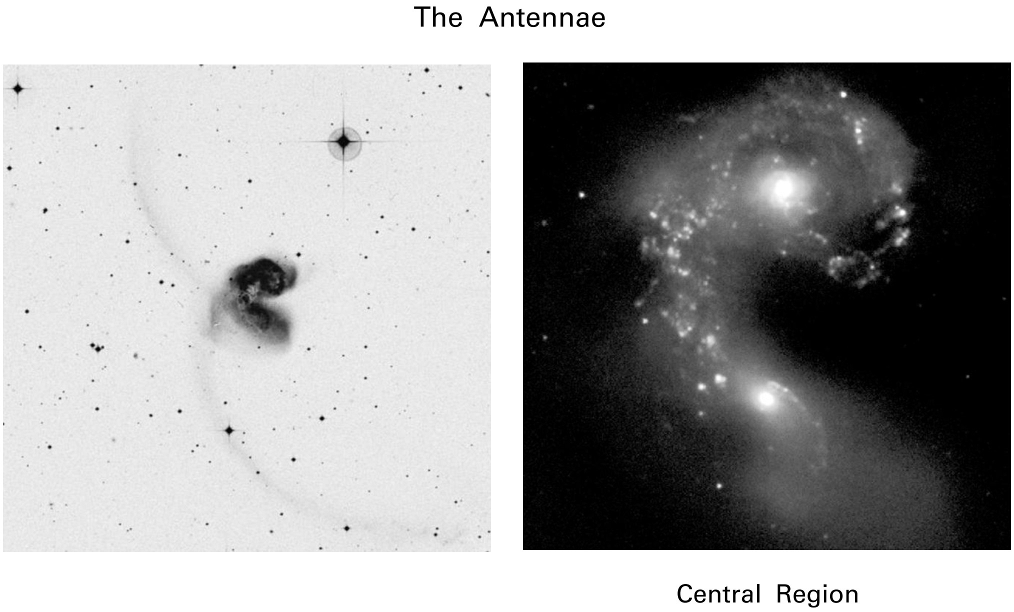


Figure 19.15 Optical image of the ultraluminous infrared galaxy Arp 220. The inset is a 2.2 μm image of the nuclear region only. Offsets in the main figure are relative to the nominal IRAS position of the galaxy. Offsets within the inset are relative to the peak emission at 2.2 μm.


The Antennae

Central Region

Figure 19.16 (*left panel:*) Optical image of The Antennae, spanning $18'$. (*right panel:*) Image in 2.12 μm of the innermost $2\overset{\prime}{.}5$, or 14 kpc at the galactic distance of 19 Mpc.

even two galaxies locked in a binary orbit cannot undergo many interactions. The ejection of stars during each encounter robs energy and angular momentum from the large-scale, orbital motion. After only a few close passages, a pair of gravitationally bound galaxies spirals together and truly merges, creating a system quite different from the original ones. The presence of dark halos accelerates this process, both by increasing the geometric cross section and through halo spinup, which further diminishes the orbital angular momentum of the visible component.

Each spiral galaxy of an interacting pair contains some 10^{11} stars. These are so thinly distributed, however, that there is no chance for direct collisions, even when the systems partially or wholly merge. The two stellar populations simply interpenetrate. During all subsequent motions, including violent ejection of stars, total energy is conserved. Similar statements cannot be made of the *gas*. This component, relatively small in mass, responds to the changing gravitational field created by the stars. Fluid streamlines may then cross, even when the galaxies themselves are still some distance apart. Streamline crossing creates shocks, which radiate energy.

This energy loss within the circulating gas of each galaxy allows that material to spiral rapidly inward, deeper into the potential well. If the gas shocks repeatedly, the spiraling occurs over just a few galactic orbital periods, or a total interval of order 10^8 yr. Such a time scale accords with that deduced empirically for the starburst phenomenon. We can thus understand, at least in broad terms, how circumnuclear gas accumulates. But how, in detail, do stars form within this new medium? And what is the nature of the perturbation leading to crossed streamlines and shocks?

The best-resolved examples of starburst regions show formation activity occurring in discrete clusters. The youngest of these do not widely differ within the system or from one galaxy to another. While their internal stellar densities may be a 100 times those in even the most vigorous Galactic sites, the inferred cluster formation times of order 10^7 yr are similar. The indication, then, is that localized regions are making stars through the usual process, albeit in an unusually dense and massive environment.

We have noted, in fact, that the molecular clouds producing individual clusters in starbursts are probably more akin to Sgr B2 than to anything in the solar neighborhood. There is a lingering uncertainty regarding the mass distribution of the stars being formed. The best-studied case is 30 Doradus. Here, the color-magnitude diagram yields a distribution that falls off at high M_* in a manner similar to the Galactic initial mass function. In the other direction, the flattening and subsequent turnover may occur at a larger M_*-value. If so, the region is forming a greater proportion of relatively massive stars. We must bear in mind, however, such technical issues as the proper dereddening procedure in an inhomogeneous, dusty environment. High extinction effectively removes objects from a magnitude-limited sample. The corrections are problematic but important, since they alter the low-mass end of the distribution.

19.3.4 Stellar Bars

Returning once more to the large-scale picture, we have noted that gas transport is facilitated by energy dissipation. The material also needs to lose angular momentum if it is to spiral inward. Any nonaxisymmetric perturbation to the stellar surface density will torque the gas gravitationally, and a trailing spiral wave induces mass flow in the proper direction. (Recall our discussion of the Galactic center.) To generate a very high mass flow, the perturbation itself must be exceptionally strong. Numerical simulations have demonstrated two fundamental results in this context. First, a trailing, spiral disturbance can be enhanced in a galaxy through tidal interaction with a neighbor. Second, a sufficiently large-amplitude disturbance becomes a linear *bar* in the central region.

Figure 19.17 illustrates both points through a representative calculation. This sequence shows the response of a face-on spiral galaxy to another, passing spiral. The latter, which is inclined 71° to our line of sight, only appears in the second frame, at its point of closest approach. Both systems are modeled as containing stars and gas, with the latter comprising 10 percent of the disk mass. There are also massive, dark halos, not shown. The stars (*upper panels*) are treated as point masses, interacting only through their mutual gravity. Gas elements (*lower panels*) have an extra, repulsive force to mimic pressure, following the technique of SPH (Chapter 12). Each unit of nondimensional time corresponds to 2.5×10^8 yr.

Before the companion galaxy appears, both the gas and stars exhibit a clear spiral pattern whose amplitude is modest, especially for the stellar component. The patterns are greatly disturbed in the second pair of frames, when the center of the neighboring galaxy reaches 8 kpc from the target nucleus. Following this close passage, the two-armed spiral is much more open and of greater amplitude. A central bar develops and is dominant in the last frames shown. The rotation sense of the face-on galaxy is clockwise. Hence, the gaseous bar slightly *leads* the stellar one. The stars exert a braking torque on the gas, draining its angular momentum.

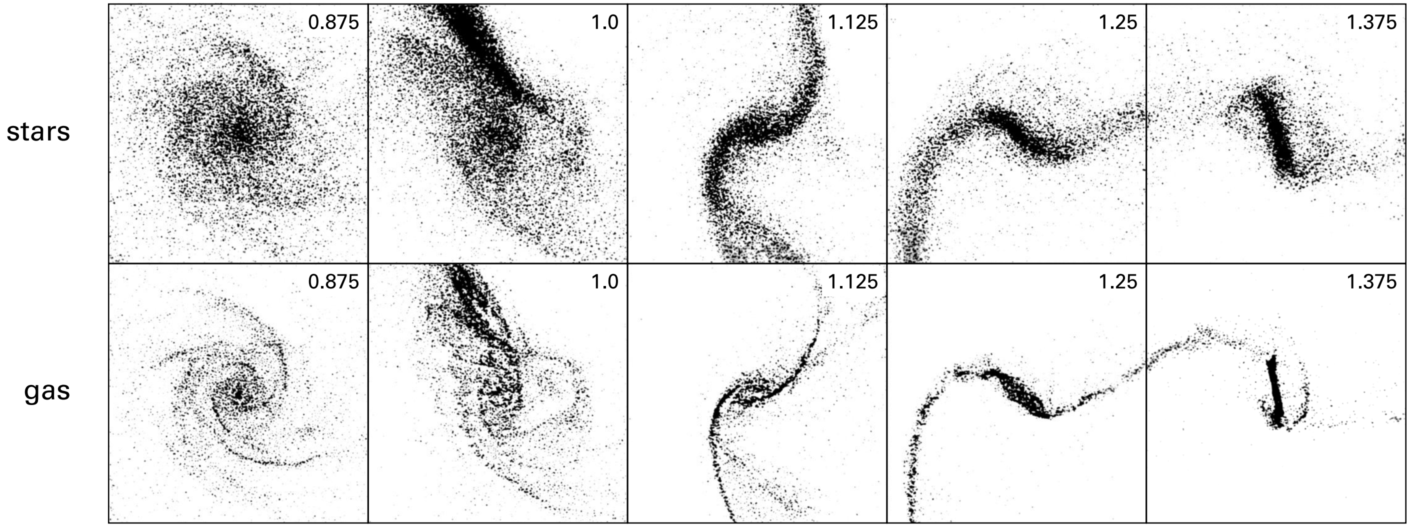


Figure 19.17 Numerical simulations of stars (*upper panels*) and gas (*lower panels*) in two interacting galaxies. One galaxy, tilted 71° to our line of sight, only appears in the second pair of panels. Each unit of the displayed, nondimensional time corresponds to 2.5×10^8 yr.

While shocks are not evident in Figure 19.17, they are nonetheless occurring within the simulated gas. Shocks, like gravitational torquing, arise from the nonaxisymmetric potential created by the bar itself. In the frame of reference rotating with the bar, there exist stable orbits both elongated in the bar direction and perpendicular to it. Many orbits have cusp-like segments or internal loops. An idealized test particle could trace such a region smoothly, but a real fluid undergoes a shock. After such a transition, the fluid temporarily follows another orbit further inside, until it shocks again. In this way, gas inexorably flows towards the center.

Most observed spiral galaxies are not interacting with a nearby companion. Hence, it must be possible for a spiral pattern to arise spontaneously within the system. The same is true for central bars. A portion of starburst galaxies, comprising the lower end of the luminosity scale, are isolated, barred spirals. These display circumnuclear formation activity and have a central structure that is strongly influenced by the bar. A disturbed pattern of this nature can also be found in galaxies with $\dot{M}_\star$-values closer to our own. The top two panels of Figure 19.18 provide an interesting example. On the left is a large-scale view of the galaxy NGC 4314, showing two spiral arms originating from a central bar. The inset to the right demonstrates that the nucleus is surrounded by a clumpy ring. Inside this ring, which has a radius of 320 pc, is a dark dust lane. Apparently, the bar has caused inward flow of gas from the body of the disk, but the material has become stalled, long enough for some of it to produce new stars.

One also sees bars at the centers of many otherwise normal spirals. These are classified, according to the Hubble scheme, as “SB” and fall along a sequence parallel to unbarred systems. The lower panel of Figure 19.18 shows an especially striking case, NGC 1300. This galaxy has, besides a well-defined bar of diameter 14 kpc, a prominent, interior bulge and relatively tight, outer arms. Hence it is assigned a Hubble type of SBbc. We expect that both the trailing arms and bar stimulate inflow of diffuse gas. However, this transport has not created a true starburst, either in the nucleus proper or in a circumnuclear ring.

A similar situation holds within the Milky Way. Near-infrared light from the central bulge (Figure 1.19; top panel) is slightly brighter to one side of the Galactic center than the other. This light emanates mainly from old stars and suffers only modest extinction by intervening dust. A plausible explanation for the asymmetry is that the stars are distributed in a bar-like pattern, oriented obliquely with respect to our line of sight. The semi-major axis of the bar must be an appreciable fraction of our Galactic radius for its nearer side to be noticeably brighter. In fact, this length is estimated to be from 3 to 4 kpc, so that the entire structure spans the molecular ring seen in CO. The associated mass is of order $10^{10}\ M_\odot$.

Our own bar, like those in other galaxies, sets the velocity pattern of the local interstellar medium. Indeed, it was the noncircular motion displayed by HI gas that provided the first hint of the bar’s existence. Numerous other studies have confirmed this finding and provided other details. The distinctive entities we have described in the Galactic center, such as the giant cloud Sgr B2, are orbiting within the nonaxisymmetric gravitational potential largely created by the bar. Such objects slowly but continually spiral inward, to be replaced by fresh material supplied from the outside. We are witnessing in this region a complex interplay of forces, where the distribution of older stars creates gas flow and the condensation of molecular clouds. The latter, in turn, are responsible for the addition of new members to the stellar component.

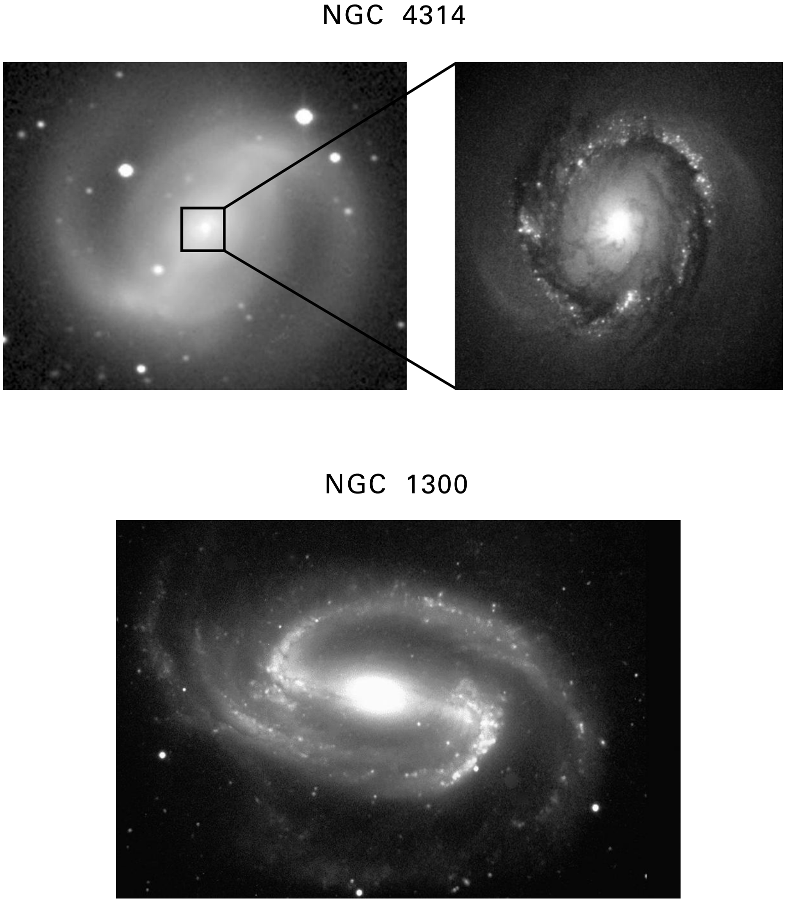


Figure 19.18 Optical images of barred galaxies. The upper left panel and inset show the bar and central ring of NGC 4314, while the lower image highlights the bar in the spiral NGC 1300.

19.4 Galaxies in Their Youth

Galaxies evolve with time. Beginning with the Milky Way, we know that the processes now occurring in various regions have changed qualitatively. At some point in the past, the Galaxy consisted entirely of gas, which transmuted to stars at different rates, depending on its location. This activity was most vigorous in what is now the stellar halo and the central bulge, since these components presently contain the highest proportion of old red giants. On the other hand, the transformation in the disk was so delayed that it is very much ongoing, at a roughly steady rate.

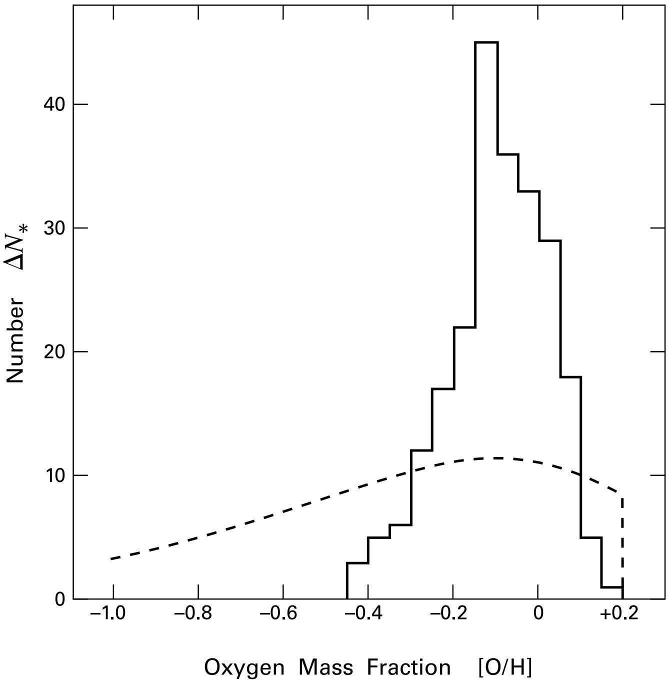


Figure 19.19 Distribution of oxygen abundance for 231 stars in the solar neighborhood. The independent variable is the logarithmic oxygen mass fraction relative to the solar photosphere. Shown are the number of stars in bins of $\Delta[\mathrm{O/H}] = 0.05$. The dashed curve is the theoretical prediction from equation (19.12).

19.4.1 The G-Dwarf Problem

How did the Galaxy arrive at its present-day structure? We are far from having a complete answer to this question. As just suggested, and demonstrated previously with regard to the Local Group, stellar ages inferred from the HR diagram provide the key input for empirical reconstruction of galactic history. Another valuable clue is the observed *metallicity* of stars. The mass of heavy elements within any region inexorably rises, as successive generations of objects die and return processed material to the interstellar medium. Elemental abundances are well documented in the solar neighborhood and have at least ruled out the simplest picture of how our Galactic patch could have evolved. This contradiction between the pattern of stellar metallicities and basic theoretical expectations is known traditionally as the *G-dwarf problem.*

The theoretical prediction is easy enough to grasp, at least qualitatively. Any stars initially formed within the Milky Way should have been entirely metal-free. While high-mass members of this group died quickly and enriched the gaseous environment, objects of sufficiently low mass should still be observable. According to Table 1.1, any star with $M_* < 1\,M_\odot$, corresponding to a spectral type later than G2, lives more than 1×10^{10} yr. It is true that low-mass stars born somewhat later incorporated gas already polluted with heavy elements, but the enrichment must have increased only gradually. Hence, there should still be an appreciable fraction of very metal-poor stars.

To quantify this argument, consider the evolution of $M_{\rm gas}(t)$ and $M_{\rm stars}(t)$, the mass in gas and long-lived stars, respectively, within the solar neighborhood. The second quantity increases at the expense of the first, so that their sum is a constant, equal to $M_{\rm gas}(0)$. We also follow the metallic enrichment of the gas, focusing specifically on $X_O \equiv M_O(t)/M_{\rm gas}(t)$, the mass fraction in oxygen. This element is created in Type II supernovae, from progenitor stars that live a short period, in terms of Galactic evolution. Such massive objects therefore do not enter the long-lived population of interest, but they do contribute a mass of oxygen proportional to Δt, the Galactic time elapsed. Since $\Delta M_{\rm stars}$ is also proportional to Δt, the oxygen mass produced by supernovae during this interval may be written as $y_O\,\Delta M_{\rm stars}$. Here, y_O is the *yield* in oxygen. This quantity, which we assume to be temporally constant, depends on the initial mass function. As usual, we also take the latter distribution to be invariant over Galactic history.

At the same time that supernovae are distributing fresh oxygen, the creation of new, low-mass stars has depleted some of the element from the gas. The net change is

$$\begin{aligned}\Delta M_O &= y_O\,\Delta M_{\rm stars} + X_O\,\Delta M_{\rm gas} \\ &= (X_O - y_O)\,\Delta M_{\rm gas}\ .\end{aligned} \tag{19.5}$$

where we have employed the fact that $\Delta M_{\rm stars} = -\Delta M_{\rm gas}$. Note that ΔM_O is positive and $\Delta M_{\rm gas}$ negative, so that $X_O < y_O$. Utilizing equation (19.5), we find the change in X_O over Δt to be

$$\begin{aligned}\Delta X_O &= \frac{\Delta M_O}{M_{\rm gas}} - \frac{M_O}{M_{\rm gas}^2}\,\Delta M_{\rm gas} \\ &= -y_O\,\frac{\Delta M_{\rm gas}}{M_{\rm gas}}\ .\end{aligned} \tag{19.6}$$

We may readily integrate the last equation. Since $X_O(0) = 0$, we have

$$M_{\rm gas}(t) = M_{\rm gas}(0)\,\exp\bigl[-X_O(t)/y_O\bigr]\ , \tag{19.7}$$

which is conveniently recast as

$$M_{\rm stars}(t) = M_{\rm gas}(0)\,\Bigl\{1 - \exp\bigl[-X_O(t)/y_O\bigr]\Bigr\}\ . \tag{19.8}$$

Suppose now that we observe the metallicities in a subset of long-lived stars. What is ΔN_*, the expected number of objects with oxygen mass fraction between X_O and $X_O + \Delta X_O$? Within the Galactic region, the cumulative mass of stars that were produced in this abundance range is

$$\Delta M_{\rm stars} = \frac{M_{\rm gas}(0)}{y_O}\,\exp\,(-X_O/y_O)\;\Delta X_O\ .$$

Here we have combined the second form of equation (19.6) with (19.7), and again utilized mass conservation. We are interested only in a certain fraction f of these stars, set by some range of spectral types. If $\langle M_* \rangle$ denotes the average stellar mass created at any time, then

$$\begin{aligned}\Delta N_* &= f\,\frac{\Delta M_{\rm stars}}{\langle M_* \rangle} \\ &= f\,\frac{M_{\rm gas}(0)}{\langle M_* \rangle}\,\exp\left(-X_O/y_O\right)\,\frac{\Delta X_O}{y_O}\ .\end{aligned} \tag{19.9}$$

At the present Galactic age $t_\circ$, the total number of stars in our sample, with all oxygen abundances, is $N_{\rm tot} = f\,M_{\rm stars}(t_\circ)/\langle M_* \rangle$. Let us denote by $\alpha_{\rm max}\,y_O$ the highest oxygen mass fraction within the sample, where $\alpha_{\rm max}$ is another nondimensional quantity. This maximal fraction should be above the solar one, which we denote $\alpha_\odot\,y_O$. Equation (19.8) then gives

$$N_{\rm tot} = f\,\frac{M_{\rm gas}(0)}{\langle M_* \rangle}\,\left[1 - \exp\left(-\alpha_{\rm max}\right)\right]\ . \tag{19.10}$$

Combining Equations (19.9) and (19.10), we find

$$\frac{\Delta N_*}{\Delta x_O} = \frac{\alpha_\odot\,N_{\rm tot}}{1 - \exp\left(-\alpha_{\rm max}\right)}\,\exp\left(-\alpha_\odot\,x_O\right)\ . \tag{19.11}$$

Here we have defined $x_O \equiv X_O/(\alpha_\odot y_O)$, the oxygen mass fraction relative to the solar value.

Our final transformation of equation (19.11) is to bin stars into equal increments of $\log x_O$, a quantity traditionally denoted [O/H]. We thus rewrite our equation as

$$\frac{\Delta N_*}{\Delta[{\rm O/H}]} = \frac{\alpha_\odot\,x_O\,N_{\rm tot}}{\log {\rm e}\left[1 - \exp\left(-\alpha_{\rm max}\right)\right]}\,\exp\left(-\alpha_\odot\,x_O\right)\ , \tag{19.12}$$

where $x_O = 10^{[{\rm O/H}]}$. The dashed curve in Figure 19.19 is this theoretical prediction. To set $\alpha_\odot$ and $\alpha_{\rm max}$, we first utilized nucleosynthesis calculations, which give $y_O = 0.016$. In combination with the observed oxygen abundance in the solar photosphere, we find $\alpha_\odot = 0.52$. The maximum abundance for local G dwarfs then gives $\alpha_{\rm max} = 0.82$. The histogram in Figure 19.19 summarizes data for 231 stars, with spectral types from G2 to G9, all located within a cylindrical radius of 25 pc from the solar Galactic position. There is evidently a steep falloff in stars with [O/H] both above and below about -0.2, in conflict with equation (19.12). The latter discrepancy with theory is the essence of the G-dwarf problem. Where are the truly low-metallicity stars?

One possible answer is that the gas in our region of the Galaxy was already enriched by the time its stars began to form. Perhaps the heavy elements were spewed out by the central bulge, which created stars at an earlier epoch. In any case, initial enrichment *per se* does not solve the problem. It is not difficult to repeat the foregoing derivation with a finite level of X_O at $t = 0$. Although this assumption sets a lower bound on the present-day oxygen abundance, the shallow rise of the distribution above that point remains.

An alternative suggestion is that the earliest, metal-poor stars were all massive objects that quickly exploded. The hypothesis, then, is that the initial mass function becomes tilted toward high M_* within a lower-metallicity cloud environment. There is, as yet, little empirical evidence supporting this notion. We have seen that 30 Doradus, where the mean metallicity is one third solar, has a falloff in the upper end of its stellar distribution that is similar to our own. The inferred initial mass function in globular clusters, whose metallicity is lower by another two orders of magnitude, is also close to the local result, at least for $M_* < 1\,M_\odot$. In § 19.5.3, we will describe how cloud collapse in a truly *zero*-metal environment may nonetheless have led to massive objects.

Yet a third solution is perhaps the most radical conceptually, but also one of the most appealing. The idea is to abandon the 'closed-box' nature of our theoretical model and allow inflow of gas. Suppose, for example, that the Galaxy, initially starless, had a much smaller mass than at present. As this structure began to produce stars, fresh gas was added continually from the outside. This material was presumably metal-free. Heavy elements created by the existing stars were mixed into a reservoir of less total mass, so that the enrichment quickly climbed. As more gas entered, it too engulfed the metals distributed by the growing stellar population. The extant members of this population that are very metal-poor would then be few in number.

There is still no firm evidence for steady gas accretion, either past or present, and therefore little basis for modifying our simple model in a quantitative manner.[5] Generally, however, the accretion picture is in accord with observations that galaxy mergers were more common in the past. We shall discuss these observations shortly. It is also worth noting that the ongoing search for zero-metallicity, or "Population III," stars, has still found no suitable objects in the disk, but interesting candidates in the Galactic halo. Here, individual stars can have metallicities far below that of the globular clusters. The 0.8 $M_\odot$ red giant HE 0107-5240, located at a distance of 11 kpc, has $[\mathrm{Fe/H}] = -5.3$. That is, the iron mass fraction is 5×10^{-6} times the solar value. Finding other objects of this nature will help fill in early Galactic history.

19.4.2 Trends with Redshift

The most direct approach to the broad problem of galaxy origins is to probe systems much less evolved than our own, *i. e.*, those which formed closer in time to the Big Bang. These younger entities may be identified by the *redshift* of all their spectral lines. If λ_{obs} and $\lambda_\circ$ are the observed and laboratory wavelengths, respectively, of a representative line, then we define the redshift as

$$z \equiv \frac{\lambda_{\mathrm{obs}}}{\lambda_\circ} - 1 \ . \tag{19.13}$$

This increase in wavelength, corresponding to a decrease in photon energy, arises from the background gravitational field. According to General Relativity, the Universe has expanded between the time of photon emission and detection. The galaxy is receding from us with a radial velocity V_r. This velocity, to a first approximation, is proportional to the galaxy's distance:

$$V_r = H_\circ\, D \ . \tag{19.14}$$

Here, the Hubble constant $H_\circ$ is inferred empirically to be $70\ \mathrm{km\ s^{-1}\ Mpc^{-1}}$, with an uncertainty gauged to be about 5 percent.

---

[5] Speculation has centered, for many years, on *high-velocity clouds* as material accreting onto the Milky Way. These patches of HI gas are seen in 21 cm surveys. If located outside the Milky Way but within the Local Group, their sizes and masses correspond to dwarf galaxies.

It is important to remember that we do *not* directly observe V_r, only the redshift. The usual practice is to assign the recessional velocity through the Doppler relation $V_r = z\,c$. However, the redshift is not simply a manifestation of the Doppler effect. Theory shows that the righthand side of equation (19.14) is actually the initial term from a series expansion in D. Higher-order coefficients are conventionally obtained within the context of homogeneous and isotropic cosmological models. These same models also give us both the age and precise distance of the galaxy from its redshift. For example, a redshift $z = 2$ corresponds to a time since the Big Bang of 3×10^9 yr and a distance of 1700 Mpc.[6] These figures pertain to a specific set of cosmological parameters, as we shall later describe. Other models within the currently accepted parameter range give similar ages and distances, with the difference between model results growing more pronounced at higher redshift.

How, then, do we find redshifted galaxies among the myriad of objects in the field? A traditional, and effective, method has been photometric surveys of relatively faint systems. Here, the choice of waveband is important. A $z = 1$ galaxy with high B-band luminosity in its own rest frame, indicative of ongoing star formation, will appear to us bright in I. Surveys in the latter wavelength regime, utilizing ground-based telescopes, have netted thousands of candidate galaxies. Spectroscopic examination has then culled those systems with appreciable redshift.

A key result of the surveys, already established by the 1980s, is a marked change in galaxy *color* with increasing redshift. Higher-z systems tend to be bluer, as measured, for example, by their rest-frame $U - V$ color index. From the discussion of § 19.2, it follows that star formation activity in this group was higher in the past. Another significant trend is a rapid rise in the number of low-luminosity galaxies. In summary, dwarf systems of intrinsically blue color become a major component of the population for $z \lesssim 1$.

Spatial imaging of the galaxies allows us to identify them morphologically and helps clarify the situation. The majority of faint, blue galaxies turn out, upon examination, to be irregulars undergoing starbursts. Recall that many present-day irregulars are also blue in color and vigorously producing stars. Their activity, however, began only recently in cosmological terms. To put the matter another way, the older irregulars found in deep surveys are too numerous to be the progenitors of the current group. The left panel of Figure 19.20 shows the number of galaxies as a function of apparent I-magnitude from one particular survey, in which the median redshift was $z \approx 0.5$. Here, galaxies classified as irregulars have been grouped together with peculiar systems suspected to represent mergers. It is apparent that the galaxy count rises more steeply toward fainter I-magnitudes than predicted by a simple extrapolation from the local population (*solid and dashed curves*).

The same statement does *not* apply to ellipticals or spirals. In the right panel of Figure 19.20, we see that the number of the latter type increases with magnitude in just the manner expected. In other words, the population has changed little from $z \approx 1$ to the present, at least in terms of its luminosity distribution. The intrinsic color of spirals, however, has on average become less blue, as star formation has slowed. This trend, while significant, is less dramatic than in the dwarf irregulars, most of which are no longer seen. Their disappearance may signify that

---

[6] We are observing the galaxy as it was in the past, when the separations between all such entities were less. The distance quoted here is a *proper* value. Specifically, it is the separation between the young Milky Way and the galaxy when the latter actually emitted its light. Also of interest is the *comoving* value of $1700 \times (1 + 2) = 5100$ Mpc. This is the distance the galaxy would have now if it had continued following the universal expansion.

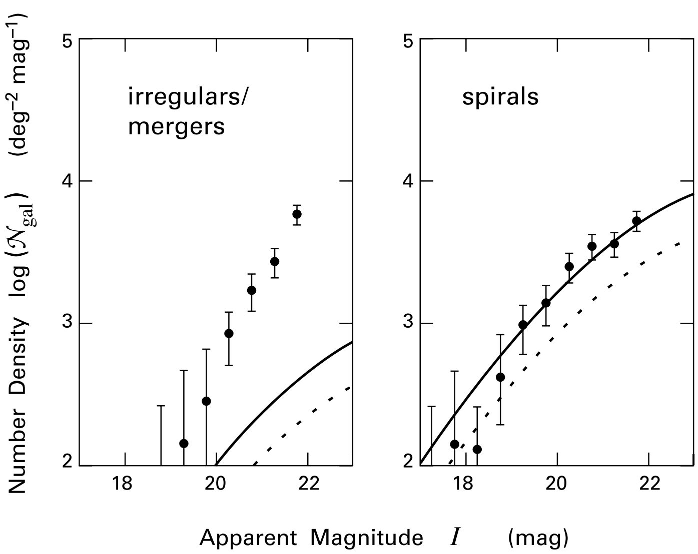


Figure 19.20 Density of galaxies as a function of I-band magnitude. The quantity $\mathcal{N}_{\rm gal}$ is the number of galaxies observed per square degree of sky, and per unit magnitude in I. The left panel shows a combination of irregulars and merger systems, while the right panel shows spirals. Solid and dashed smooth curves are the predictions based on the hypothesis that the galactic distributions have not evolved in time. The two curves are based on different estimates for the local galaxy density.

individual irregulars have remained intact, while fading in luminosity. It may also indicate something very different, that most of the small systems have vanished by merging with larger ones. This latter picture is supported by the general finding, already mentioned, that morphological signatures of mergers and close encounters increase with redshift.

By measuring the flux in various wavebands from many galaxies, we may reconstruct a quantity of key importance for the evolutionary issue. This is the *comoving luminosity density*, $\mathcal{L}(\lambda, z)$. To obtain it, we first gauge the total luminosity of a group of galaxies within a narrow redshift interval and waveband. We account for global expansion during the intervening epoch by using the comoving volume of the group in question. Reconstruction of this volume from the observed redshifts again requires adoption of a cosmological model.

The important point is that $\mathcal{L}(\lambda, z)$ is related to the total star formation rate $\dot{\mathcal{M}}_*$ of the group, where this rate is again measured per comoving volume. In fact, the energy output in the ultraviolet of an individual galaxy stems almost entirely from its massive stars. The ultraviolet luminosity, assessed in the galaxy's rest frame, thus correlates with its instantaneous star formation rate, independent of past history. As for nearby systems, we quantify this correlation through population synthesis models, after assuming an initial mass function and metallicity.

Figure 19.21 summarizes these analyses by displaying the volumetric star formation rate as a function of redshift. We see that there is a dramatic rise in $\dot{\mathcal{M}}_*$ from $z = 0$ to $z = 1$. Much of this increase reflects the appearance, at earlier time, of dwarf irregulars. At even greater

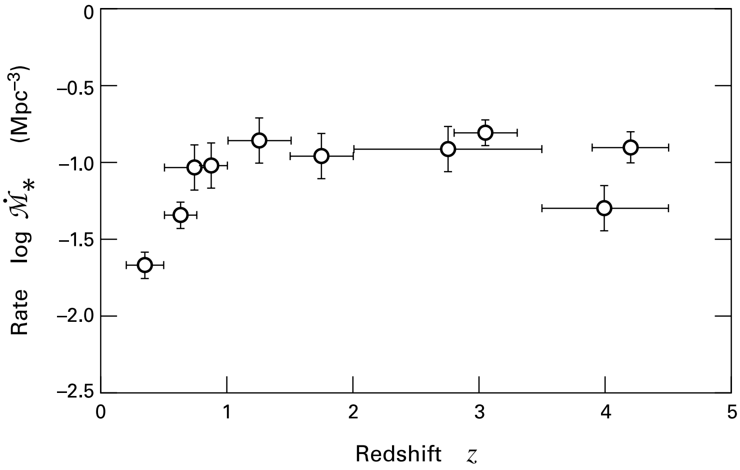


Figure 19.21 Global, volumetric star formation rate, displayed as a function of redshift.

redshifts, the total star formation rate appears to level off or even decline slightly. Needless to say, there is a great deal of information concealed in this deceptively simple figure. Why $\dot{\mathcal{M}}_*$ has this particular history is still far from clear.

19.4.3 Lyman-Break Galaxies

Before delving into the greater evolutionary issue, we need to explain how individual galaxies with redshifts significantly above unity are detected at all. The surveys we have described, conducted within a single waveband, no longer suffice. As we search for increasingly dim objects, their number rises so steeply that direct, spectroscopic determination of redshifts becomes infeasible. What we require is some characteristic spectral signature of a young galaxy. Searching for this signature within a large population would be far more efficient than analyzing spectra one by one.

Two such features have been exploited, both arising from atomic hydrogen. One is the Lyman α transition at 1216 Å. While this line has been detected in emission toward many systems, it has proved even more useful when seen in absorption. The strategy is to observe a distant quasar in the optical regime. Any HI gas lying between us and the quasar will partially absorb its radiation. The strongest dip occurs at 1216 Å in the rest frame of the HI mass. If this entity has redshift z, we detect the line at a wavelength increased by the factor $1 + z$, which shifts it to the optical regime for $z \gtrsim 2$.

Many quasars display, in fact, a large number of closely spaced absorption lines. This *Lyman α forest* arises, through the above mechanism, from intervening HI clouds spanning a broad range of distances and redshifts. (See the top panel of Figure 19.26 and later discussion.) In *damped Lyman α systems*, one sees additional, deeper absorption at a certain wavelength. The name comes from the fact that the line wings have the profile characteristic of natural broadening, which in turn mimics frictional damping (Appendix E). Column densities associated with the enhanced absorption are sufficiently high ($N_{\rm H} \gtrsim 2 \times 10^{20}\,{\rm cm}^{-2}$) that these systems may represent the gaseous component of old, galactic disks. Further spectroscopy reveals a wealth of information concerning this material, including both its heavy element enrichment and kinematics.

Let us turn, however, to other, complementary studies that directly observe radiation from the stars themselves. We now utilize a second spectral feature, the *Lyman continuum discontinuity*. Consider a typical O star in a young galaxy. The light from this object and its kin dominates the ultraviolet continuum of the star-forming system. Radiation being transported through the stellar atmosphere to the surface has a much higher optical depth if $\lambda < 912$ Å, where the threshold corresponds to the ionization energy of an $n = 1$ electron in hydrogen. The emergent flux is therefore diminished in this short-wavelength regime. As seen in the model spectral energy distribution of Figure 19.22 (*top panel*), the integrated flux from the entire galaxy also suffers a sharp discontinuity at this point. Indeed, the break is even stronger in the galactic flux, since interstellar HI gas absorbs another large portion of those short-wavelength photons emitted by the stars themselves.

The discontinuity shown in Figure 19.22 occurs in the far-ultraviolet. But this is only so in the rest frame of the galaxy. If the system were redshifted, say to $z = 3.0$, the feature would appear at a wavelength of $912 \times 4.0 = 3648$ Å. The increased wavelength happens to lie near the U-band, centered at 3600 Å. Thus, photometry of the galaxy in question should yield a relatively low flux at U and a pronounced upward jump at B. Here, we recall that the B-band is centered at 4400 Å, corresponding in the galaxy's rest frame to 1100 Å, a wavelength longward of the discontinuity. Many such "U-dropout" galaxies have, in fact, been seen. The bottom two panels of Figure 19.22 display one example. Here, a galaxy visible in B and V disappears in U.

What we have just described is known generally as the *Lyman break* technique for finding high-redshift galaxies. Among a population of faint sources, those exhibiting a jump in flux between two adjacent wavebands are good, initial candidates. We then follow up the photometric survey by examining each object spectroscopically, thus determining its precise redshift. The technique is not limited to the U- and B-bands. When comparing B- and V-fluxes to find B-dropouts, we are selecting candidate galaxies with $z \approx 4$. Extension to V and R is more problematic, as the background flux increases and sensitivity is lowered. Given that the U-band filter still represents the short-wavelength limit in both ground-based and satellite observations, Lyman-break galaxies currently span a redshift range $2.5 \lesssim z \lesssim 5$.

Spectroscopic study yields much more than just the numerical value of z. Remember that we are effectively probing these systems in the far-ultraviolet, a regime dominated by massive stars. The pattern and strength of emission lines, as well as the continuum level, make it clear that we are witnessing starbursts. In fact, the star formation rate per galaxy is comparable to that seen locally. A significant caveat here is that we must properly account for interstellar dust within each system. Dust both reduces the emergent ultraviolet flux and redistributes emission over a broad range of wavelengths, especially the far-infrared. This complication is a major source of uncertainty when assessing both the stellar production of individual galaxies and the overall, volumetric formation rate $\dot{\mathcal{M}}_*$ shown in Figure 19.21.

With the imaging capability of the Hubble Space Telescope, it has been possible to discern spatial structure in many Lyman-break galaxies. We find a broad range of morphologies, but these do not mimic the standard Hubble types. A fraction of the systems, it is true, have a smooth distribution of surface brightness, reminiscent of present-day ellipticals. Most, however,

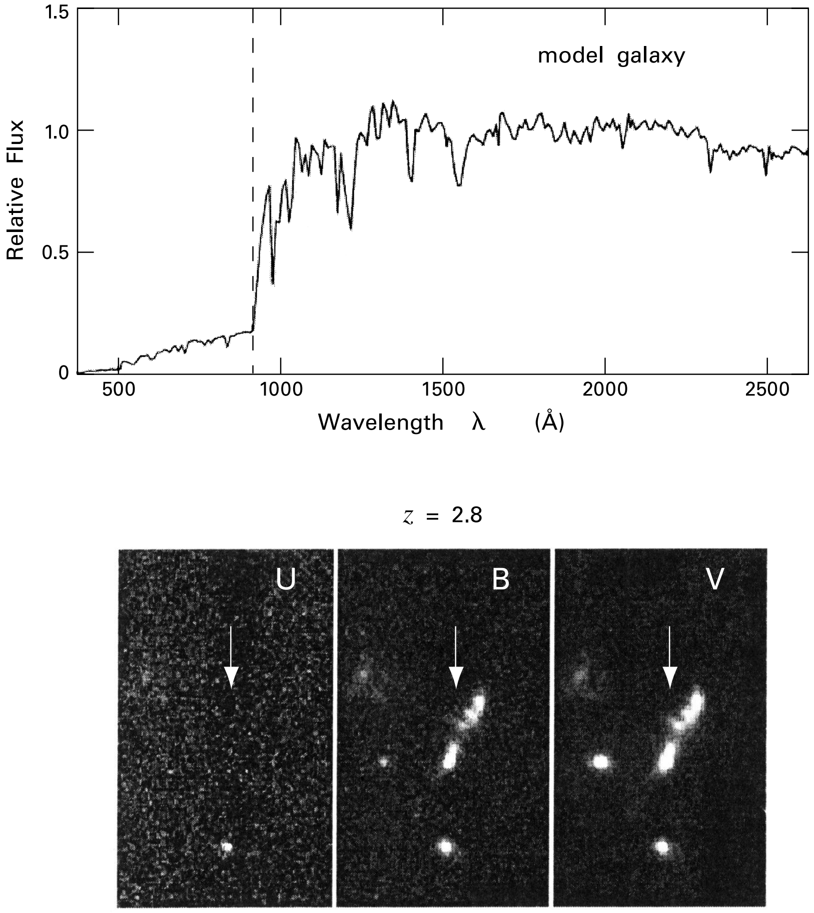


Figure 19.22 The Lyman break technique. *upper panel:* Model spectrum of a starburst galaxy, as seen in its rest frame. Note the severe drop in flux shortward of 912 Å. This Lyman limit is indicated by the vertical dashed line. *lower panel:* Hubble Space Telescope image of a $z = 2.8$ galaxy in the U, B, and V-filters.

display a pattern of bright knots embedded in a dimmer and more diffuse background (see Figure 19.22). Some appear clearly fragmented. Such images strengthen our earlier conclusion, based on spectral data, that the systems are analogous to local starbursts, which themselves arise from mergers and interactions. Lyman-break galaxies may be the antecedents to ellipticals and the central bulges of present-day spirals. Both systems completed formation in the remote past, perhaps in the burst-like fashion we seem to be witnessing.

The most powerful starbursts in the local Universe are the ultraluminous, infrared galaxies, such as Arp 220 (Figure 19.15). These radiate too weakly in the ultraviolet to be detected by the Lyman-break technique. There may thus exist a population of ancient starbursts even stronger than the ones we see. A dusty galaxy radiating most of its energy in the far-infrared should be bright in the submillimeter when seen at sufficiently high redshift. In fact, surveys have detected numerous sources with just this character. The system may be matched, in some cases, to a Lyman-break galaxy found in the optical. Generally, however, the relatively poor angular resolution in the submillimeter has hampered such identification. Both the nature of this population and its significance in early galactic evolution remain key problems for the future.

19.5 The First Stars

If accretion and mergers were common in the past, then the gas reservoir within any particular galaxy may have been relatively small when star formation began. Nevertheless, this matter must have evolved to the point of creating the analogues to today's dense cores, which thereafter collapsed. A key element in this picture is the ability of clouds to radiate energy efficiently while undergoing compression. This cooling ensures the ultimate onset of gravitational instability. The first difficulty we face is that the mechanisms described in Chapter 7, whether radiation in CO lines or infrared continuum from grains, would have been unavailable in the absence of metals.

19.5.1 Recombination and Molecule Synthesis

The general question is how cooling occurs in a gas composed of only hydrogen, helium, and the smattering of other light elements produced in the Big Bang. The answer depends on the relative population of ions, atoms, and molecules within the gas. Early on, the Universe was filled with ionized plasma taking part in the cosmic expansion. Free electrons and protons were immersed in a sea of blackbody radiation. Photons traveled only a short distance before scattering off electrons. As both the matter and radiation temperatures fell through adiabatic expansion, protons and electrons began recombining to form neutral hydrogen. This process accelerated once the number of Lyman continuum photons within any volume became less than the number of hydrogen atoms.

The recombination epoch occurred at a temperature of 4×10^3 K, about half the equivalent transition value in stellar interiors. Once hydrogen became neutral, its opacity fell drastically. Photons could now propagate freely, and some traversed the distance and time needed to reach our own location and epoch. Recombination, therefore, represents the earliest event in cosmological history that is presently observable. The arriving photons constitute the microwave background, with its precisely measured blackbody temperature of $T_\circ = 2.73$ K.

A simple argument places the recombination epoch in time, or equivalently, redshift. Consider the photons themselves as a relativistic gas, expanding adiabatically from recombination to the present. The first law of thermodynamics tells us that, as a representative volume V grows, its enclosed energy density and radiation pressure obey

$$0 = \Delta \left(u_{\rm rad} \, V \right) + P_{\rm rad} \, \Delta V \ . \qquad (19.15)$$

We now use Equations (2.37) and (11.48) to relate $u_{\rm rad}$ and $P_{\rm rad}$ to the radiation temperature $T_{\rm rad}$. We quickly find that

$$\frac{\Delta T_{\rm rad}}{T_{\rm rad}} = -\frac{1}{3}\frac{\Delta V}{V} \ . \tag{19.16}$$

Equation (19.16) tells us that $T_{\rm rad}$ declines as $V^{-1/3}$. If $V_\circ$ is the volume at the present time, it was $V_\circ\,(1\,+\,z)^{-3}$ at redshift z. It follows that the redshift associated with recombination is

$$\begin{aligned} z_{\rm rec} &= \left(\frac{V_{\rm rec}}{V_\circ}\right)^{-1/3} - 1 \\ &= \frac{T_{\rm rec}}{T_\circ} - 1 \\ &\approx 1500 \ , \end{aligned} \tag{19.17}$$

where we have again employed an approximate recombination temperature, $T_{\rm rec}$, of 4×10^3 K.

If the hydrogen gas were *completely* transformed to HI by the end of this phase, the subsequent formation of stars would be difficult to explain. Pure hydrogen at temperatures below 4000 K cools so inefficiently that any incipient contraction would quickly be reversed. The solution is that there is a small, but nonzero, population of free protons and electrons remaining for $z\,<\,z_{\rm rec}$.

Two factors inhibit recombination within the gas. One is the absorption of ultraviolet photons that arise when electrons cascade from higher levels or jump down from the continuum to the $n\,=\,1$ state. Such radiative excitation is ameliorated by the fact that hydrogen can sometimes make the transition from $n\,=\,2$ to $n\,=\,1$ by releasing a pair of lower-energy photons. This two-photon decay is intrinsically rare, but important for setting the overall rate of recombination. A second inhibiting factor is the global expansion itself. Recall that the volumetric recombination rate is proportional to the product of electron and proton densities (equation (15.1)). Both of these fall in time. Consequently, the declining electron fraction in the gas "freezes out" at a certain point. The residual fraction is of order 10^{-4} for cosmological parameters of interest.

Neither protons nor electrons by themselves provide significant cooling. However, they lead to the creation of molecular hydrogen, which eventually does. The first step is taken in the post-recombination era when protons react with the dominant, HI gas:

$$\begin{aligned} {\rm H} + {\rm H}^+ &\rightarrow {\rm H}_2^+ + h\nu \\ {\rm H}_2^+ + {\rm H} &\rightarrow {\rm H}_2 + \ {\rm H}^+ \ . \end{aligned} \tag{19.18}$$

This reaction sequence, presented earlier in Chapter 5, leads to a small fractional H_2 abundance, of order 10^{-7}, by a redshift of 300. By $z\,\approx\,100$, the abundance climbs another order of magnitude through the formation and subsequent reaction of H^-:

$$\begin{aligned} {\rm H} + {\rm e}^- &\rightarrow {\rm H}^- + h\nu \\ {\rm H}^- + {\rm H} &\rightarrow {\rm H}_2 + \ {\rm e}^- \ . \end{aligned} \tag{19.19}$$

The stepwise increase in the H_2 abundance, along with the prior freezeout of the electrons, can be seen in Figure 19.23. Note that H^- is rather easily ionized by the background radiation, thus explaining its delayed production relative to H_2^+.

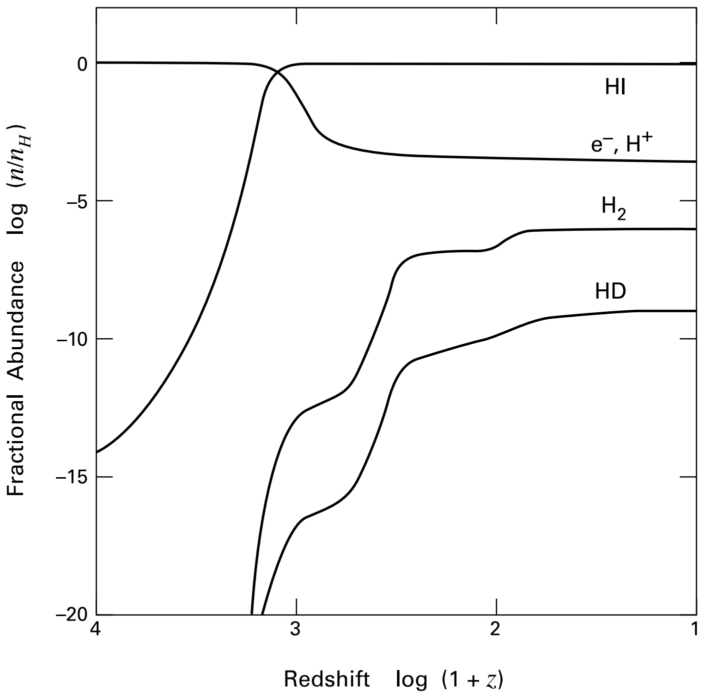


Figure 19.23 Growth of molecular abundances in the expanding Universe. Shown are the number densities of the indicated species relative to hydrogen as a function of redshift.

This last comment leads us to a more general observation. Tracking the evolution of the molecular abundance in the expanding gas requires that we fully account for the interaction of the matter and the ambient radiation. As discussed earlier, photons scatter off free electrons. This process heats the gas, provided the temperature of the radiation exceeds that of the matter. Radiation is also absorbed by the newly formed molecules. If an excited molecule collides with gas atoms before undergoing spontaneous decay, the gas temperature is again raised.

Figure 19.24 shows in more detail the decline of both the radiation and gas temperatures as a function of redshift. Prior to recombination, $T_{\rm rad}$ and $T_{\rm gas}$ are kept nearly equal by the frequent photon-electron scattering. The heat capacity of the photons greatly exceeds that of the gas. Thus, it is actually $T_{\rm gas}$ that is locked to $T_{\rm rad}$, which in turn declines as $(1 + z)^{-1}$. The small electron fraction left over from recombination keeps the two temperatures close until $z \approx 500$, after which $T_{\rm gas}$ falls more quickly.

Suppose the gas subsequently cooled adiabatically, with declining temperature $T_{\rm ad}$. Then the expression analogous to equation (19.15) would lead instead to

$$\frac{\Delta T_{\rm ad}}{T_{\rm ad}} = -\frac{2}{3}\frac{\Delta V}{V} \tag{19.20}$$

Here, we have employed $u_{\rm gas} = 3\rho k_B T_{\rm ad}/(2m_H)$ for a pure hydrogen gas, and $P_{\rm gas} = (2/3)\,u_{\rm gas}$. We have also used the fact that ρV is a constant during the expan-

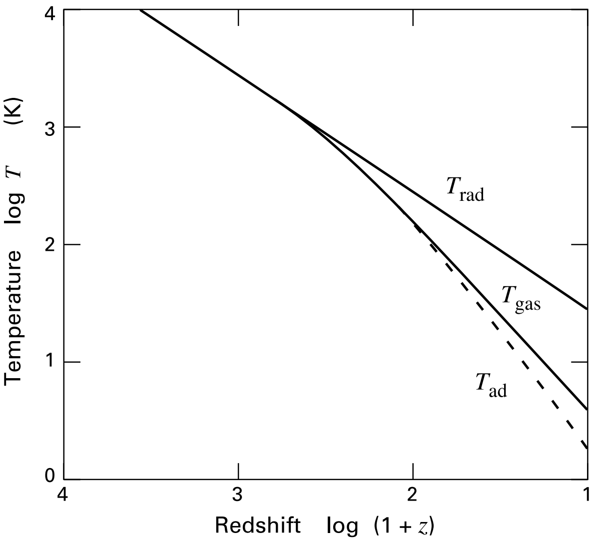


Figure 19.24 Decay of the radiation and matter temperatures during cosmological expansion. Also shown is the temperature in a hypothetical gas that expanded adiabatically.

sion. It follows from equation (19.20) that $T_{\rm ad}$ falls as $(1 + z)^{-2}$. Figure 19.24 demonstrates that the actual gas temperature declines slightly less rapidly. The difference reflects absorption of radiation by H_2, followed by collisional deexcitation.

The nucleosynthesis that created hydrogen soon after the Big Bang also produced helium, deuterium, and lithium. Most important is the second of these. Indeed, the deuterated form of molecular hydrogen, HD, appears along with H_2 itself (Figure 19.23). This species plays a role in subsequent cooling of the gas. It is produced by the reaction

$$D^+ + H_2 \rightarrow HD + H^+ \ . \tag{19.21}$$

As seen in the figure, the fractional abundance of HD reaches 10^{-9} during the global expansion. Lower concentrations of HeH^+, LiH and LiH^+ are also formed during this period.

19.5.2 The Role of Dark Matter

Our description thus far of the expanding Universe has included the ionized and neutral components of the gas, along with the background radiation field. We know, however, that a considerable amount of matter is present in some non-luminous form. Again, spiral galaxies reveal the presence of dark matter through their optical and radio rotation curves. Outlying stars and gas are orbiting too rapidly for the attracting mass to be only that seen directly. Analogous results hold for other galaxy types. Within the Local Group, one measures the central velocity dispersions of dwarf spheroidals or the rotation curves of dwarf irregulars. Comparison with the total system luminosity shows that the kinematics is dominated by a dark halo. The stellar velocity dispersion within ellipticals also reveals this extra component. On larger scales, the motion of galaxies within rich clusters is unaccountably fast unless one posits a rich supply of dark matter in the space *between* the orbiting systems.

These observations indicate that the dark matter density, averaged over suitably large volumes, exceeds that of ordinary ("baryonic") matter by an order of magnitude. Cosmological models therefore need to include this substance, even if we are completely ignorant of its nature. Indeed, its presence is critical for the growth of *structure* in the Universe.

Galaxies must have begun as small density perturbations within the uniform, expanding fluid. The self-gravity of such a lump retarded its expansion. Eventually, the object began to contract and formed the molecular clouds that in turn produced early stars. A theoretical stumbling block is that the initial growth of the density perturbation is *not* exponential in time, as in the Jeans analysis of a (fictitious) static medium (§ 9.1). The density, measured relative to the changing background, grows only algebraically in time, just as in the dynamical fragmentation of a uniform, collapsing sphere (§ 12.1). Quantitatively, $\Delta\rho/\rho_\circ$ must have been of order 10^{-3} at recombination to have reached unity by the present epoch.

This prediction is flatly contradicted by observation. The putative density perturbation would have imprinted itself on the background radiation field, giving rise to an *angular* fluctuation of $T_{\rm rad}$ that would be still be detectable. However, numerous experiments show that the anisotropy of the microwave background is much smaller, if measured in the proper reference frame. The actual fluctuations indicate that $\delta\rho/\rho_\circ$ at recombination was of order 10^{-5}. The accepted solution to this conundrum is that the requisite, initial perturbation occurred not in the baryonic matter, but in the dominant, dark component. Baryonic matter fell into the local potential wells created by these perturbations, and only then evolved to the point of gravitational collapse. It was in this later evolution that radiation from molecules played a crucial role.

Before describing this development, let us consider further the clumping of dark matter. The effectiveness of self-gravity depends on the velocity of the constituent particles. If this velocity is relativistic, *i. e.*, if the dark matter is "hot," then particles can readily stream out of a perturbed volume. In effect, this leakage stifles density growth, unless the perturbation has a very large physical extent. Numerical simulations confirm this fact and find that the dark matter condenses into a network of huge sheets, each of dimension 100 Mpc in comoving coordinates. Individual galaxies could only have separated out recently from these structures, in contradiction to observations of quasars at $z \gtrsim 5$.

One is forced instead to the *cold dark matter* picture. Here, the particles are moving at non-relativistic speeds. The initial uniform sea breaks up into a more randomly dispersed collection of smaller clumps. These coalesce in a hierarchical manner, gradually creating structure on larger scales. Meanwhile, baryonic matter is also condensing. In contrast to the previous picture, galaxies build up through the fusion of the central, gaseous portions of merging, dark matter clumps. This activity was well underway before even the oldest observed quasars were formed.

Numerical simulations of the merging process have led to an intriguing suggestion for the origin of various Hubble types. Any gas collecting in dark matter potential wells has finite angular momentum. But a rotating cloud will naturally flatten along its central axis, provided it can radiate away a substantial portion of its initial energy. The first disks formed in this way were far less massive than those in present-day spirals. We must therefore consider what happens when two or more small disks merge. One result is that the new, combined entity puffs up again into a more three-dimensional structure. It is also plausible that the merger triggers a starburst. Once all the gas is either consumed or driven off, we are left with a stellar system that can no longer dissipate energy. This basic picture, while incomplete in many aspects, may explain both ellipticals and the bulges of spirals. Perhaps spiral disks represent additional gas, accreted later, that did not participate in the original starburst.

Our sketch of the modern cosmological landscape requires one final element. The anisotropy measurements of the microwave background radiation also tell us how photon paths were curved gravitationally, from the recombination epoch until now. This curvature, in turn, depends on $\rho_{\rm tot}$, the present-day, total density of the Universe. According to General Relativity, the quantity of interest is actually $\Omega_{\rm tot} \equiv \rho_{\rm tot}/\rho_c$. Here, ρ_c is the *critical density*, defined as

$$\begin{aligned} \rho_c &\equiv \frac{3\,H_\circ^2}{8\,\pi\,G} \\ &= 9 \times 10^{-30}\ {\rm g\ cm^{-3}}\ . \end{aligned} \tag{19.22}$$

We have used for $H_\circ$ the value $70\ {\rm km\ s^{-1}\ Mpc^{-1}}$.

The anisotropy observations indicate that $\Omega_{\rm tot}$ is close to unity. On the other hand, calculations of primordial nucleosynthesis tell us, that in order to produce the present-day abundance of deuterium, the baryonic density Ω_B can be only a few percent. Observations of galaxies and clusters, together with the virial theorem, indicate that the dark matter density can exceed this by an order of magnitude but not more. The remaining density in the Universe is ascribed to another pervasive component known as *dark energy*. This exotic substance provides a repulsive form of gravity. It was unimportant in the past, during the initial growth of perturbations, but is now causing the cosmic expansion to accelerate. Its origin and nature are even less understood than cold dark matter. The popular "ΛCDM" models typically prescribe a baryonic density $\Omega_B = 0.05$, a cold dark matter component with $\Omega_{\rm DM} = 0.25$, and a dark energy with associated $\Omega_\Lambda = 0.70$.[7]

19.5.3 Protostellar Collapse

Return to the gravitational settling of gas within a dark matter clump. During this condensation, the matter first heats up adiabatically. Molecular hydrogen, however, continues to form by the processes already mentioned. A critical point is reached once the fractional abundance in H_2 reaches about 10^{-3}, when the gas density n_H is of order 10 cm^{-3}. This event may have occurred as early as $z \approx 20$. Emission in the rovibrational lines then creates a sharp drop in temperature, which stabilizes at a value near 200 K. With further accrual of gas, both the density and temperature rise again. If the calculation is run until n_H exceeds 10^8 cm^{-3}, then all of the hydrogen quickly becomes molecular via the three-body reactions:

$$\rm H + H + H \rightarrow H_2 + H \tag{19.23a}$$

$$\rm H + H + H_2 \rightarrow 2H_2\ . \tag{19.23b}$$

The temperature concurrently reaches a plateau of about 1500 K.

---

[7] The symbol Λ associated with dark energy represents the *cosmological constant*. This is a term added to the stress-energy tensor of ordinary matter in Einstein's field equations of General Relativity.

We do not know if primordial clouds much more massive than individual stars were able to reach such elevated densities. Our limited understanding of the dynamical balance in present-day dark clouds and complexes should be chastening in this regard. The compression at an earlier epoch could have been stymied, at least temporarily, by internal, turbulent motion. In simulations, the quiescent settling of gas tends to be disrupted by mergers of parent, dark matter clumps. Nonetheless, the joint action of molecular emission and slow condensation must have eventually led to true protostellar collapse at many locations.

The cloud fragment that ultimately became unstable was still much less dense than the star it produced. As in present-day star formation, the only known way to effect this transition is through inside-out collapse onto a central nucleus. Presumably, a disk acted as an intermediate conduit for infalling matter of higher angular momentum. If the temperatures we have quoted apply to this dense core analogue, then the protostellar infall rate, given by equation (10.31), was as much as three orders of magnitude above current ones. Thus, primordial protostars were extraordinarily luminous.

In § 11.4, we described a major impediment to forming massive protostars: radiation pressure acting on the infalling envelope. The absence of grains, however, means that the envelope opacity was far lower. Infalling matter close to the star was fully ionized, so that the appropriate κ is the electron scattering value (Appendix G). It may thus have been possible for the central protostar to attain much greater mass, even in the simplified picture of spherical collapse.

Figure 19.25 displays the mass-radius relation for a protostar accreting at $\dot{M} = 4 \times 10^{-3}\ M_\odot\ \mathrm{yr}^{-1}$. The initial rise in R_* is analogous to that in low-mass protostars forming at more modest rates. (See Figure 11.6, but note the different radius scale.) Also analogous is the first turnover (compare Figure 11.21). The latter is again due to the onset of gravitational contraction. This process continues until hydrogen ignites at a mass of 80 $M_\odot$. The resulting luminosity increase exerts additional pressure on the star's outer layers, as manifested by the second rise in R_*. Concurrently, L_{acc} falls and the star temporarily contracts. If the stellar mass reaches 300 $M_\odot$, then the sharply increasing nuclear luminosity creates very rapid surface expansion. The L_*-value at this point has reached $6 \times 10^6\ L_\odot$.

It is interesting that even such an extraordinarily luminous object does not create an HII region. Ultraviolet photons are completely absorbed in the massive accretion flow. Radiation diffuses outward through an optically thick precursor outside the accretion shock, ultimately escaping when the associated color temperature falls to about 6000 K. Stars of this nature are subject to internal pulsations, driven by their central nuclear reactions. These pulsations, in turn, could drive winds. Nevertheless, the foregoing calculation at least suggests that zero-metal objects of very high mass formed, along with the much lower-mass population represented by the halo star HE 0107-5240.

19.5.4 Reionization

The idea of massive, primordial objects is intriguing from another perspective. Following the accretion phase and the disappearance of the infalling envelope, each star shines copiously in the ultraviolet. Indeed, the effective temperature at a given mass is greater than that of its Population I counterpart, because of the reduced surface opacity (§ 16.2). The combined ultraviolet

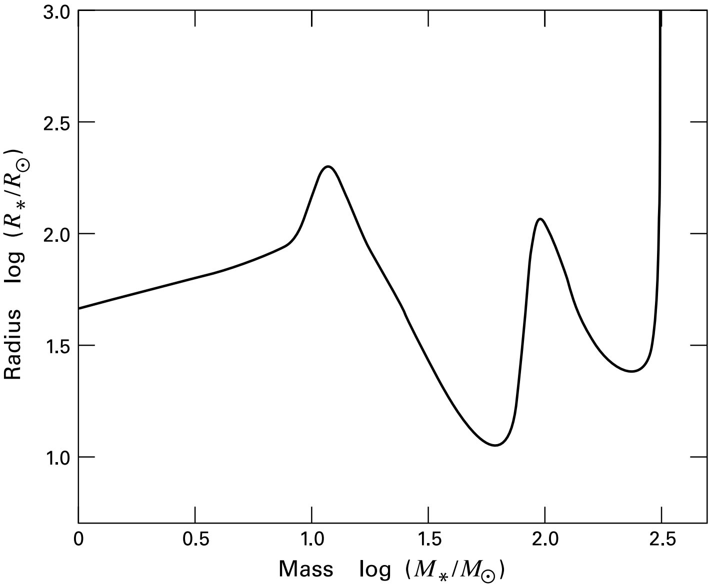


Figure 19.25 Mass-radius relation for primordial protostars accreting at a rate of $\dot{M} = 4 \times 10^{-3}\, M_\odot\, \mathrm{yr}^{-1}$. Note the extremely rapid swelling when M_* reaches $300\, M_\odot$.

luminosity from these massive stars could have ionized vast expanses of gas. Is there actually evidence for this effect?

The answer is that there is. We turn again to those distant beacons, quasars. As before, we utilize the fact that these objects emit continuum, ultraviolet light over a broad wavelength range. If hydrogen gas between us and the quasar were predominantly neutral, it would absorb this radiation strongly at the Lyman α transition. An ionized intergalactic medium would not.

If this argument sounds familiar, it is very similar to one we used to explain the Lyman α forest observed toward many quasars. The top panel of Figure 19.26 now shows an example. In this spectrum of a quasar at $z = 3.54$, we see the prominent Ly α peak, redshifted in wavelength to $1216 \times 4.54 = 5520$ Å. The panoply of lines blueward of the peak arises from many individual HI clouds. Each such entity absorbs in Ly α, but each is also less redshifted than the background quasar. Thus, the lines appear at shorter wavelength than the main emission feature.

Suppose, however, that there were additional, intervening gas not confined to discrete clouds. This material would span a *continuous* range of redshifts, though all still less than that of the quasar. We would now observe, superposed on the Lyman α forest, a broad depression of the spectrum. This *Gunn-Peterson effect* is *not* seen in the top panel of Figure 19.26, where the spectrum is basically symmetric about the Ly α peak. We conclude that, at the cosmological time corresponding to $z = 3.54$, the intergalactic medium was largely ionized. Since this medium was almost completely neutral following recombination, we may say that the Universe went through an epoch of *reionization*. This transition was created by the ultraviolet emission from massive, primordial stars.

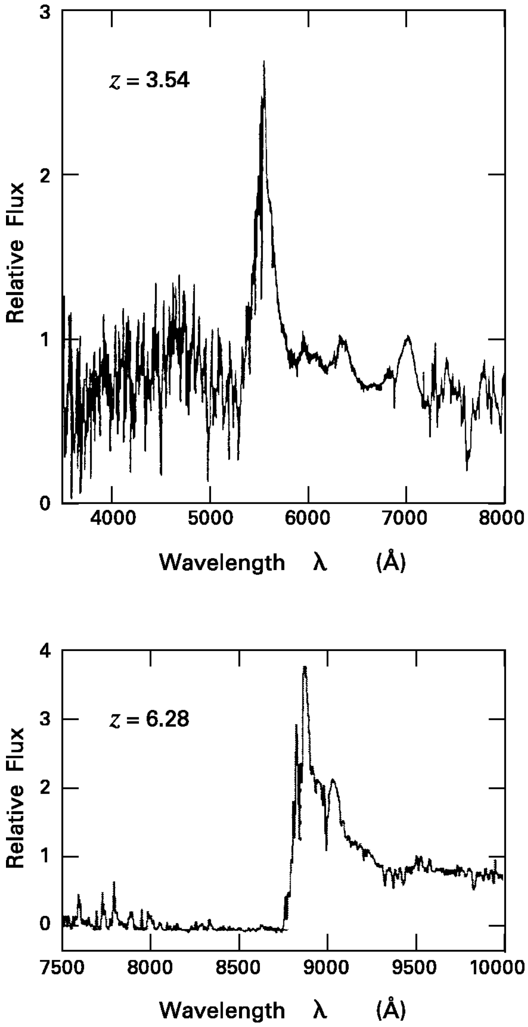


Figure 19.26 The Gunn-Peterson Effect. *upper panel*: Optical spectrum of a quasar at $z = 3.54$, showing symmetric emission about the redshifted Ly α line. Also apparent is the Lyman forest shortward of the line. *lower panel*: Spectrum of a quasar at $z = 6.28$, displaying a depression of the continuum shortward of Ly α.

The bottom panel of Figure 19.26 helps to date the epoch. Here, we see an even more distant quasar $(z = 6.28)$ that *does* show the Gunn-Peterson effect. The spectrum is strongly depressed blueward of the Ly α peak, which now occurs at 8850 Å. Intergalactic gas was therefore still neutral at $z = 6.28$, or 9×10^8 yr after the Big Bang, using the now standard cosmological parameters. Star formation had not yet begun. Other studies confirm that the critical redshift value was close to 6.

By now, the reader will not be surprised to learn that things are a bit more complicated than just indicated. First, it is not clear that the objects emitting the ultraviolet flux were actually stars, as opposed to other quasars, *i. e.*, accreting black holes. Second, the same hydrogen molecules that were instrumental in forming stars are readily dissociated by ultraviolet radiation. So one must be careful about the placement of massive stars within clouds. Finally, there was apparently a distinct, earlier epoch of reionization.

Satellite observations of the microwave background radiation find it to be linearly polarized. While the degree of polarization is tiny, of order 10^{-6}, it has important implications. The effect stems from scattering of photons by free electrons. From the angular scale of the observed polarization, one may conclude that the requisite ionization already existed by $z = 17$. Perhaps a very early generation of high-mass stars, born only 2×10^8 yr after the Big Bang, was responsible. If so, then full ionization of the intergalactic medium did not occur until another stellar population appeared, at $z = 6$. Did this second starbirth event mark the formation of the most primitive galaxies now observed? In what environment did the previous generation form? These are but a few of the questions whose answers we await.

Chapter Summary

The production rate of stars within the Milky Way varies with Galactocentric radius. We gauge this variation by observing the emission from young, massive stars and then extrapolating to obtain the full population. Both the formation rate and the supply of molecular gas rise inward, begin falling at a radius of about 4 kpc, then rise again at the very center. The latter region is extraordinarily active, containing both giant cloud complexes and the richest young clusters in the Galaxy.

Most relatively nearby galaxies have similar ages, regardless of morphological type. In ellipticals, there was a high level of stellar production early on, which has now abated. This activity is ongoing in the disks of spirals, as measured primarily through Hα emission. The spatial variation of star formation in many galaxies seems to correlate in a simple way with the gas supply (Schmidt's law), but only above a threshold surface density of the fuel. Theoretical modeling of a galaxy's optical spectrum yields the temporal history of activity. The galaxies with the largest current rise in production are the gas-rich and metal-poor irregulars, of which there are many within our Local Group. Other galaxies within this system, the dwarf spheroidals, are more akin to ellipticals.

Star formation activity reaches extraordinarily high levels in localized regions of some irregulars. Such starbursts are recognized spectroscopically, by the emission lines associated with a large population of massive stars. Larger, spiral galaxies may also contain starbursts, located near their centers. In these cases, it appears that the galaxy is being tidally disturbed or is merging with a close neighbor. Tidal gravity induces a rapid inflow of diffuse gas, often through the formation of a central, stellar bar. The most prodigiously active starburst galaxies are so heavily cloaked in dust that they are primarily seen in the infrared.

The paucity of extremely metal-poor, low-mass stars in the solar neighborhood demonstrates that our region of the Galaxy has not evolved in isolation. One possibility is that fresh gas has accreted from the outside. Surveys of older, redshifted galaxies find additional evidence for interactions. Many starbursting, dwarf irregulars existed in the past but have now disappeared, perhaps as a result of mergers. The Lyman break technique find even more primitive systems of higher luminosity that are also undergoing starburst activity.

Primordial gas must have both cooled and contracted substantially to create the first generation of stars. In modern cosmological models, the gas first coalesced at the centers of dark matter clumps, which had earlier separated out from a more uniform, expanding background. The cooling of the gas was effected by hydrogen molecules. These formed in abundance once the gas density reached values well in excess of present-day clouds. The first protostars therefore experienced rapid infall during their main accretion phase. Formation of these objects reionized the diffuse, intergalactic gas.

Suggested Reading

Section 19.1 For the radial distribution of star formation in the Milky Way, with references to the underlying observations, see

Prantzos, N. & Aubert, O. 1995, AA, 302, 69.

Research on the Galactic Center is summarized in

Morris, M. & Serabyn, E. 1996, ARAA, 34, 645.

The original paper on Schmidt's law is

Schmidt, M. 1959, ApJ, 129, 243.

Section 19.2 A general reference on galaxies, at the advanced undergraduate level, is

Sparke, L. S. & Gallagher, J. S. 2000, *Galaxies in the Universe: An Introduction* (New York: Cambridge U. Press).

Star formation throughout the Hubble Sequence is covered by

Kennicutt, R. C. 1998, ARAA, 36, 189.

The classic paper on galaxy stability via a two-fluid analysis is

Jog, C. & Solomon, P. M. 1984, ApJ, 276, 114.

For population synthesis modeling, see

Bruzual, G. & Charlot, S. 1993, ApJ, 405, 538.

Section 19.3 Both observational and theoretical aspects of the starburst phenomenon are described in

Kennicutt, R. C., Schweizer, F, & Barnes, J. E. (ed) 1998, *Galaxies: Interactions and Induced Star Formation*, (Berlin: Springer).

Observations of luminous infrared galaxies are the subject of

Sanders, D. B. & Mirabel, I. F. 1996, ARAA, 34, 749.

A summary of evidence for the Milky Way central bar is

Blitz, L., Binney, J., Lo, K. Y., Bally, J., & Ho, P. T. P. 1993, Nature, 361, 417.

Section 19.4 A full discussion of the G-dwarf problem is in

Pagel, B. E. J. 1997, *Nucleosynthesis and the Chemical Evolution of Galaxies*, (New York: Cambridge U. Press), Chapter 8.

A clear introduction to both the methods and key results of modern cosmology is

Longair, M. 1998, *Galaxy Formation* (New York: Springer-Verlag).

The first study displaying the cosmological star formation rate was

Madau, P., Ferguson, H. C., Dickinson, M. E., Giavalisco, M., Steidel, C. C., & Fruchter, A. 1996, MNRAS, 283, 1388.

For a detailed review of Lyman-break galaxies, see

Giavalisco, M. 2002, ARAA, 40, 579.

Section 19.5 Processes during the recombination epoch are described in

Peebles, P. J. E. 1993, *Principles of Physical Cosmology*, (Princeton: Princeton U. Press), Chapter 6

Galli, D. & Palla, F. 1998, AA, 335, 403.

The idea that gas settles into dark matter potential wells is due to

White, S. D. M. & Rees, M. J. 1978, MNRAS, 183, 341.

For a numerical simulation of this effect in the context of ΛCDM cosmology, see

Yoshida, N., Abel, T., Hernquist, L., & Sugiyama, N. 2003, ApJ, 592, 645.

Identification of the three-body reaction as a means to produce molecular hydrogen is from

Palla, F., Salpeter, E. E., & Stahler, S. W. 1983, ApJ, 271, 632.

Evidence for the Gunn-Peterson effect in a distant quasar was first presented in

Becker, R. H. et al. 2001, AJ, 122, 2850.

20 The Physical Problem: A Second Look

We now step back from the technical development of our subject and assess, in summary fashion, our grasp of its essential points. Of course, it is only through the course of this study that we have gained an appreciation of the salient issues. It is also true in general that the very questions we are capable of framing reflect the limitations of our understanding. Many current quandaries will fade in importance or even lose their meaning entirely as our knowledge base expands. Nevertheless, let us try and see how far we have advanced, and where future progress can be made.

20.1 Clouds

Half or more of all stars form within giant molecular cloud complexes. Where do these come from? We touched very lightly on this issue, in part because it is tangential to star formation proper, but also because it is incompletely understood. The idea that molecular clouds condense from HI gas has been prevalent for decades, and is strongly bolstered by the theory of hydrogen self-shielding (§ 8.1). How this transformation occurs on galactic scales is less clear, although it is a problem where numerical simulations could still make significant headway. To complicate the matter, a substantial, but uncertain fraction of the molecular gas is being converted back into atomic form, through the action of stellar radiation in photodissociation regions.

Roughly 10^7 yr following the birth of a giant cloud, it disappears from view, leaving behind multiple groups of stars that disperse into space. Combined molecular line maps of the clouds and optical imaging of the nascent clusters show the dispersal in action (§ 3.1). It is the massive stars themselves that evidently do the damage through their winds and ionizing radiation. A far more vexing issue is the disappearance of smaller, dark cloud complexes containing *no* massive stars. Stellar jets appear to be the culprit, at least by default, but the theory is woefully undeveloped (§ 12.4). Observations of HI gas surrounding aging T Tauri or post-T Tauri stars would be of great benefit in this regard.

Another central issue is the mechanical support of clouds against gravity. For dense cores, the smallest entities producing stars, the issue is resolved in principle. A combination of thermal pressure and tension from an embedded, largely static magnetic field seems adequate. How these two forces endow cores with their observed, elongated structure is less secure (§ 9.4). It is on the larger scale of dark clouds and giant complexes that things indeed become murky. Thermal pressure is manifestly unimportant, leaving the magnetic field as the dominant influence. The morphology of the clouds, as well as their large, internal line widths tell us that neither the gas nor the field is homogeneous and static. A field composed of planar Alfvén waves and one tangled up through turbulent motion are both unlikely to persist (§ 9.5). Some alternative

The Formation of Stars. Steven W. Stahler and Francesco Palla
Copyright © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
ISBN: 3-527-40559-3

picture is called for, in which magnetized streams of gas interact but somehow avoid creating shocks and thereby rapidly dissipating energy.

Given our ignorance of clouds' large-scale support, it is not surprising that the origin of substructure is also problematic. The molecular gas in giant complexes resides in clumps whose simple mass distribution may be a clue to the assembly process (§ 3.1). Only the most massive clumps are forming stars. Somehow, isolated regions in these entities lose their effective magnetic pressure, and settle into quiescent cores that subsequently undergo protostellar collapse. How this transition occurs is a matter of paramount importance, for it underlies the manifest *inefficiency* of star formation itself. Only a small fraction, perhaps a few percent, of the total mass of a giant cloud or dark cloud complex ends up in stars. Why should this be so? To put the matter another way, why is one cloud volume destined to become a star, while most neighboring volumes are not?

The empirical pattern of stellar births, as manifested in the histories of known associations and clusters, provides an interesting clue. That there is a coherent pattern at all, that the global star formation rate accelerates in regions with abundant molecular gas, is significant (§ 12.4). This increase demonstrates that the formation of a dense core cannot be a random event in the cloud medium, but is occurring *in response* to some global change in that medium. One such change would be large-scale, gravitational contraction of the parent cloud. This contraction presumably occurs through gradual loss of the mechanical support. Theory can say little more until the nature of that support is better established.

In this context, two traditional and influential ideas from the past can now be put to rest. One is *hierarchical fragmentation*, already mentioned briefly in § 12.4. The picture is that a large, relatively diffuse cloud transforms into small, dense stars in stages. Through the action of gravity, the original cloud both contracts and splits into several daughter fragments. Each of these fragments suffers an analogous fate, subdividing further into more compact entities. Eventually, we arrive at a collection of stars.

The very inefficiency of star formation tells us immediately that such a process cannot occur ubiquitously throughout a giant complex or dark cloud. But hierarchical fragmentation is not a plausible sequence of events even for a hypothetical region where a large mass fraction is destined to become stars. One way for a self-gravitating cloud to split apart is through dynamical fragmentation (§ 12.1). Here, the original object must be significantly out of force balance. It strains credibility to suppose that every observed giant molecular cloud is essentially in a state of free-fall collapse. But even if that were the case, breakup would continue only until the combination of thermal and magnetic pressure became significant in the latest fragment. At that point, the fragment would rebound as a static, equilibrium object.

Inside-out collapse is the other possibility. This is indeed the *only* known means by which a cloud can achieve the enormous density increase (a factor of 10^{20}) required to become a star. The requisite initial condition is that the cloud be in force balance, but that self-gravity be just over the maximum allowed for dynamical stability. It is unlikely that this delicate condition is attained through breakup of a larger parent. More plausibly, regions *evolve* to such a state. Dense cores presently observed to be starless may thus, in the future, gain mass from their surroundings until they reach the point of marginal stability. Gas will drift across magnetic field lines during this period (§ 10.1). In summary, stars do not arise from the "top-down" fragmentation of a larger entity. The dense cores producing them grow in a "bottom-up" fashion, out of a less static and homogeneous medium, which itself may be slowly contracting. The ensuing, inside-out collapse of cores involves no dynamical fragmentation.

The second traditional idea is *bimodal star formation*. Creating a massive star through inside-out collapse is problematic, because of the repulsion by stellar winds and radiation pressure (§ 11.4). Pending a solution to this difficulty, one may draw the conclusion that stars above some critical mass form in a very different way. We have said as much in this book, and have favored the view, still largely untested, that high-mass objects result from the coalescence of dense cores and previously formed stars within a crowded environment (§ 12.5).

Adherents of bimodal star formation, on the other hand, invoke the dynamical fragmentation of large clouds. Smaller clouds more nearly in force balance produce low-mass stars via inside-out collapse; larger clouds yield massive stars through prompt breakup. The objection here is the same as before. No such breakup can lead from cloud to stellar densities. Furthermore, observations show that every group of massive stars is accompanied by an even larger retinue of lower-mass objects. The true situation is not best described as bimodal, but as a *continuous* variation in the pattern of star formation in diverse environments. We shall return to this issue shortly.

20.2 Stars

Consider the region within a dense core that is undergoing inside-out collapse onto an embedded star and its disk. Note first that there are generally *two* such regions, since most stars are born as binaries (§ 12.1). The observed, geometric elongation of dense cores may promote the growth of two separate concentrations (§ 12.3). These gradually evolve, through accretion of their surroundings and further gravitational settling, to the point of true collapse. Since most binaries have separations much less than 0.1 pc, the stars must have transferred most of their angular momentum to the surrounding gas.

Let us concentrate, however, on the infall taking place onto a single star. A critical, unsolved problem is what *ends* the process. Equivalently, we are asking what circumstance limits the mass of stars in general. This latter question would be meaningless if the initial mass function were a simple power law. In fact, most stars form in the decade from 0.1 and 1 $M_\odot$ (§ 4.5). Having established the existence of a characteristic stellar mass scale, it is natural to seek its origin.

A basic realization from cloud studies is that the potential *reservoir* of material is very large. It is true that the estimated masses of dense cores do not greatly exceed stellar values (§ 3.3). But these figures simply represent the integrated signal from a molecular line or warmed dust, both of which essentially select high column density. The outermost contour from such a tracer does not mark a true discontinuity in the medium. To put the issue in more physical terms, we recall that the dense core is permeated by a magnetic field that pinches together in the vicinity of the star (as sketched in Figure 10.8). Once cloud collapse has begun, why doesn't cloud gas continue to slide along curved field lines, filling in the material draining onto the central object?

It is worth repeating the two observations that may be relevant in this regard. First, as we proceed to sufficiently large distances beyond the conventional boundary of the dense core, the gas is no longer quiescent. It has been difficult to characterize this transition quantitatively, as maps of the exterior, lower-density region are typically made in CO and have poorer spatial resolution than those of the interior. An improved theoretical treatment of dense core origins might clarify to what degree this change in the cloud inhibits accretion onto the star. The second significant fact, or rather inference from other data, is that the protostar itself may power a wind. This flow could be what reverses infall, even for a low-mass object.

Whether protostars in fact drive strong winds cannot presently be stated with any confidence, since we have no direct evidence that any particular embedded source is a true protostar, *i. e.*, that its luminosity stems mainly from the release of infall kinetic energy (§ 11.5). The fact is, however, that essentially *all* embedded stars have associated molecular outflows, and these tend to have higher mechanical luminosities for more deeply embedded objects (§ 13.2). Strong winds may arise in response to the buildup of stellar angular momentum during infall. Here, one should recall our demonstration in § 16.5 that a fully convective protostar accreting from a rotating cloud would attain breakup speed even before the appearance of any circumstellar disk. Thus, the eruption of the wind must occur very early, again as indicated by the outflow observations.

Molecular outflows are not themselves composed of wind gas, but of cloud matter that has been stirred up and dragged outward by an internal jet. This jet is the stellar wind channeled into a narrow, bipolar configuration (§ 13.1). Even as it leaves the star, the wind is already anisotropic because of rotational and magnetic forces. How the subsequent channeling over larger distances occurs is still not fully understood. It is ironic that jets first revealed themselves through shocks (Herbig-Haro objects) arising from relatively minor fluctuations in the wind. The spacing of the shocks yields, in principle, a record of these fluctuations. Shocks of the greatest intensity occur very far from the star and have associated intervals of order 10^3 yr. Weaker shocks occur within 0.1 pc and represent decade-long variations.

The growth of a disk represents another response of the protostar to increasing angular momentum. Infalling gas enters into orbit around the central object if that material is rotating too fast to reach the stellar surface. It is important to recognize the limitations of this statement. Imagine the extreme case, where essentially *all* the matter within a region collapses into a rotating slab. If such a slab can cool to a geometrically thin configuration, then it will break apart into numerous fragments, none of which attains stellar density. Forming a star requires an initial supply of low-angular momentum gas, presumably from the central column of a slowly rotating dense core (§ 10.4).

Since we observe disks around optically revealed stars (§ 17.3), something prevents the earlier, protostellar disk from building up mass to the point of fragmentation. It is attractive to think of the process as self-limiting, through the torquing action of spiral waves that both drive mass inward and also spread the configuration (§ 11.3). This idea is secure conceptually, but still needs development in an evolutionary setting. A more complex interaction occurs closer to the star, where the local magnetic field exerts dynamical influence.

Accretion from the disk onto the star appears to continue during the visible, pre-main-sequence phase. At least that is one relatively simple interpretation of the spectroscopic infall signature evident in the YY Orionis subclass of T Tauri stars (§ 17.2). Estimates of the mass transport rate are crude, but indicate that it is relatively low. Thus, numerical calculations of quasi-static contraction that ignore stellar mass variation are still trustworthy. Observations of microjets show that more than just infall is occurring. There are still no quantitative models for the full motion of gas in the vicinity of a pre-main-sequence object. Whatever its detailed nature, shocks resulting from this activity give rise to the strong emission lines that help define classical T Tauri stars (§ 17.1). A major puzzle is that weak-lined stars, which match their classical counterparts in both mass and age, lack strong emission lines in gas, as well as the heated dust indicative of nearby disks.

The bulk contraction of T Tauri stars is a relatively simple process that is well understood physically. Good corroboration of the basic theory is provided by the similar contraction ages for binary components of disparate masses (§ 12.3). This finding not only tells us something significant about binary origins, but would not hold unless the ages themselves were reasonably accurate. A key feature of quasi-static contraction is the steady decrease of stellar luminosity (§ 16.2). Pre-main-sequence stars are relatively bright, long before the fusion of hydrogen. Their source of energy is self-gravity. In no sense, therefore, do young stars "flare up" as a result of nuclear ignition.

Intermediate-mass stars are important conceptually as a bridge between solar-type objects, which originate from individual dense cores, and truly massive stars that do not. Above 2 $M_\odot$, the process of quasi-static contraction itself grows more complex, as the partially radiative star adjusts to thermal equilibrium (§ 18.2). A natural upper boundary for intermediate masses is about 10 $M_\odot$. Stars heavier than this have no pre-main-sequence phase at all, but begin fusing hydrogen even if they are still experiencing infall.

These findings are more secure than those concerning the stellar environment. We do not understand why emission-line activity is relatively rare at intermediate masses (§ 18.1). Presumably, the complex flows occurring outside lower-mass objects tend to be absent. Does this change reflect the detailed circumstances of formation? Certainly there is less evidence for disks, especially among Be stars. The volume immediately surrounding a massive O star may be exceptionally clean, as any residual matter would have been driven off by the intense radiation and winds.

The HII region generated by such an object sweeps through a large portion of the parent cloud (§ 15.1). Clumps overtaken by the ionization front undergo both compression and surface evaporation. The first process demonstrably induces low-mass star formation in the clump interior (§ 15.5), presumably through the stimulated production of dense cores and their subsequent collapse. But the net effect of the O star on its surroundings is a destructive one. If the cloud is not completely driven off before the star explodes as a supernova, then the resulting blast wave finishes the job.

A problem that involves star formation at all masses, along with cloud dynamics, is the origin of stellar groups. Why does one region form an OB association, another a bound cluster, while a third creates a T association? Here one can only speculate, as we did in Chapter 12. The star formation record, we have noted, suggests that the clouds undergo large-scale contraction, even while they are creating dense cores that in turn produce low-mass stars and binaries. Since the putative contraction is driven by self-gravity, it leads to the greatest density increase in the most massive parent clouds. The centers of these, therefore, are ripe environments for the coalescence leading to high-mass stars. Once formed, these objects rapidly disperse ambient gas.

Contraction also occurs in less massive clouds, but here it is reversed by the energy input and disruption from stellar outflows. The cloud therefore is unable to progress beyond the stage of forming a loose collection of low-mass objects, *i. e.*, a T association. In between are entities that contract and shed mass through stellar outflows on the same time scale. Within these environments, dense cores become sufficiently crowded that the stars they spawn remain as a bound cluster even after all background gas has been driven off. Both theoretical analysis and observations of very young, bound clusters should be brought to bear in exploring this picture.

20.3 Galaxies

The largest stellar groups of all are galaxies. In these, one encounters the full range of star formation activity. At one extreme are the moribund ellipticals, composed of aging main-sequence and red giant stars. At the other extreme are starburst regions in spirals and irregulars (§ 19.3). Many of these regions are deeply enshrouded in interstellar dust. We may thus be reassured that molecular gas, which provides the raw material for stars in the Galactic neighborhood, plays an analogous role even under these circumstances. But the clouds themselves are denser and warmer, the latter characteristic undoubtedly arising from the intense star formation itself. In the single well-resolved example, the 30 Doradus region in the Large Magellanic Cloud, we observe directly that massive stars are created together with a dominant population of solar-type objects. It would be even more comforting to witness the latter forming out of dense cores, but these will be hard to find in gas so thoroughly ravaged by its most luminous products.

Galaxies in general give us the perspective to witness large-scale processes that must also be occurring locally. For example, we know that gas in the solar neighborhood has not simply been consumed to produce the current population of stars. The essential lesson of the G-dwarf problem (§ 19.4) is that fresh material enters the region even as this production continues. Some of this accretion may represent the global transport of gas within the Galaxy, induced by the torquing action of spiral arms. The extraordinary activity at the Galactic center (§ 19.1) must ultimately arise from this effect. It remains to be seen if the pileup in gas manifested as the molecular ring reflects the same transport process.

Even apart from the issue of mass transport, spiral arms are central to molecular cloud buildup and therefore star formation. The arms are fundamentally perturbations of the *stellar* population. Since it is the stars that provide most of the gravitational force within galactic disks, the gas responds by collecting into the spiral pattern. We see the result with striking clarity in the CO images of M51 (Figure 1.21) and many other systems.

Equally striking in disk galaxies is the pattern in the azimuthally averaged star formation rate. Despite the complexity of spiral arms, the rate per surface area varies smoothly with total gas density. This behavior, the global form of the Schmidt law, still lacks a convincing theoretical basis. Within any galaxy, star formation also truncates abruptly below a well-defined, gas surface density (§ 19.2). There is much still to be learned from this simple fact.

Any complacency that galaxy structure is an old and settled issue must be shaken by the existence of dark matter. This unknown substance comprises most of the mass of spirals, and is at least as dominant in nearby dwarf irregulars. Needless to say, dark matter must enter any discussion of star formation on the galactic scale. Spiral arms, for example, are often stimulated by the presence of a companion galaxy. This mode of arm production must hold especially for systems with starburst nuclei, which are almost all members of interacting pairs. The interaction, however, is not between stellar disks, but between dark halos containing both stars and ordinary gas. Simulations demonstrate the complexity of the dynamics under these circumstances (§ 19.3).

Reaching back further in time, our knowledge of early galaxy evolution is advancing rapidly through direct observations. At modest redshifts ($z \lesssim 2$), the global star formation rate rises quickly, and is largely supplied by analogues to present-day dwarf irregulars. The behavior of the global rate at higher z is far less certain. What does seem clear, however, is that the primitive systems being revealed through spectroscopic and photometric surveys have little in common with those of today (§ 19.4).

In retrospect, this marked difference is perhaps not surprising. Certainly the interstellar medium in the first galaxies was exotic by our standards. The low abundance of metals implies far less efficient cooling in the ambient gas. What emission did occur was from molecules created initially during the cosmological expansion (§ 19.5). Self-gravity must still overwhelm thermal pressure to create stars. It follows that only clouds of extraordinary density, at least with respect to nearby examples, were susceptible to collapse. Detection of such entities would represent a key advance in understanding this distant epoch.

Suggested Reading

Section 20.1 For the unorthodox view that *most* Galactic HI is generated by photodissociation of molecular gas, see

> Allen, R. J. 2001, in *Gas and Galaxy Evolution*, eds. J. E. Hibbard, M. Rupen, & J. H. Van Gorkom, (San Francisco: ASP).

The original picture of hierarchical fragmentation is from

> Hoyle, F. 1953, ApJ, 118, 513.

In this model, breakup occurs in a quasi-static fashion. However, the role of the cloud's internal density structure was ignored. A later, influential implementation of Hoyle's idea was

> Low, C. & Lynden-Bell, D. 1976, MNRAS, 176, 367.

For one variant of the bimodal star formation hypothesis, see

> Larson, R. B. 2002, in *Modes of Star Formation and the Origin of the Field Population*, eds. E. K. Grebel & W. Brandner (San Francisco: ASP), p. 442.

Section 20.2 Some of the processes still marginally understood in young stars may be elucidated by analogous findings elsewhere. It is interesting, for example, that Herbig-Haro jets can be produced by evolved red giants:

> Cohen, M., Dopita, M. A., Schwartz, R. D., & Tielens, A. G. G. M. 1985, ApJ, 297, 702.

The coalescence that forms massive objects may resemble the merger process creating the anomalous main-sequence stars known as "blue stragglers." A review of the latter is in

> Lombardi, J. C. & Rasio, F. A. 2002, in *Stellar Collisions, Mergers, and their Consequences*, ed. M. M. Shara (San Francisco: ASP), p. 35.

Section 20.3 Starburst galaxies not only contain a large amount of diffuse gas and dust, but also drive off this material at extraordinary rates. For the detection of *superwinds* in these environments, see

> Lehnert, M. D. 1999, in *Wolf-Rayet Phenomena in Massive Stars and Starburst Galaxies*, eds. K. A. van der Hucht, G. Koenigsberger, & P. R. J. Eenens (San Francisco: ASP), p. 645.

We have not been able to describe in any detail how galactic disks settle within dark matter halos. An influential theoretical study has been

> Fall, S. M. & Efstathiou, G. 1980, MNRAS, 193, 189.

The ΛCDM cosmological models, while successfully reproducing structure at the largest scales, still face difficulties. See, for example,

> Moore, B., Ghigna, S., Governato, F., Lake, G., Quinn, T., Stadel, J., & Tozzi, P. 1999, ApJ, 524, L19.

These authors show that the simulated breakup of dark matter produces an excess of small, satellite bodies around each galaxy.

The Formation of Stars
Steven W. Stahler and Francesco Palla

A Astronomical Conventions

A.1 Units and Constants

Here we list some standard units of astronomy. We have also included other physical constants employed in the book.

Name	Symbol	Value
Angstrom	Å	10^{-8} cm
Astronomical Unit	AU	1.496×10^{13} cm
Bohr magneton	μ_B	5.788×10^{-3} eV μG^{-1}
Boltzmann constant	k_B	1.381×10^{-16} erg K^{-1}
electron charge	e	4.803×10^{-10} esu
electron mass	m_e	9.110×10^{-28} g
electron volt	eV	1.602×10^{-12} erg
gas constant	$\mathcal{R}$	8.314×10^{7} erg g^{-1} K^{-1}
gravitational constant	G	6.672×10^{-8} cm^3 g^{-1} s^{-2}
Hubble constant	$H_\circ$	70 km s^{-1} Mpc^{-1}
Jansky	Jy	10^{-23} erg s^{-1} cm^{-2} Hz^{-1}
nuclear magneton	μ_N	3.152×10^{-6} eV μG^{-1}
parsec	pc	3.086×10^{18} cm
Planck constant	h	6.626×10^{-27} erg s
	$\hbar$	1.055×10^{-27} erg s
proton mass	m_p	1.673×10^{-24} g
radiation density constant	a	7.565×10^{-15} erg cm^{-3} K^{-4}
solar luminosity	$L_\odot$	3.827×10^{33} erg s^{-1}
solar mass	$M_\odot$	1.989×10^{33} g
solar radius	$R_\odot$	6.960×10^{10} cm
speed of light	c	2.998×10^{10} cm s^{-1}
Stefan-Boltzmann constant	σ_B	5.670×10^{-5} erg s^{-1} cm^{-2} K^{-4}
Thomson cross section	σ_T	6.652×10^{-25} cm^2

The Formation of Stars. Steven W. Stahler and Francesco Palla

ISBN: 3-527-40559-3

A.2 Photometric Systems

The wavelength listed here is an effective one, weighted by the transmission function of the filter. We also list the full width half-maximum of the transmitted intensity.

System	Filter	Wavelength	Width
Johnson-Morgan	U	3600 Å	700 Å
	B	4400 Å	1000 Å
	V	5500 Å	900 Å
Infrared	R	7000 Å	2200 Å
	I	9000 Å	2400 Å
	J	1.22 µm	0.26 µm
	H	1.65 µm	0.29 µm
	K	2.18 µm	0.41 µm
	L	3.55 µm	0.57 µm
	M	4.77 µm	0.45 µm
	N	10.5 µm	5.2 µm
	Q	20.1 µm	7.8 µm
Strömgren	u	3500 Å	340 Å
	v	4100 Å	200 Å
	b	4700 Å	160 Å
	y	5500 Å	240 Å
	β	4860 Å	150 Å

A.3 Equatorial and Galactic Coordinates

From our perspective, astronomical objects appear to reside on a two-dimensional spherical surface centered on us. Because of diurnal rotation, this *celestial sphere* itself rotates from east to west, carrying planets and stars, including the Sun, along with it. The celestial equator is the outward projection of our own, and its north and south poles are the imaginary extensions of the Earth's rotational axis (see Figure A.1). Over a longer time scale, the Sun and planets migrate along the *ecliptic*, a band representing the intersection of the solar system's orbital plane and the celestial sphere. The ecliptic is tilted by 23° with respect to the celestial equatorial plane.

Several astronomical coordinate systems are used to locate objects within the celestial sphere, in such a way that the position is independent of diurnal motion. In the *equatorial* system, depicted in Figure A.1, we measure the *right ascension* (α) of the object as its angular displacement eastward along the celestial equator. Note that the angle is measured in hours, minutes, and seconds, where 1 hour is equivalent to $360/24 = 15^\circ$. The zero point for this measurement is the *vernal equinox*. This is the point where the Sun, in the northward phase of its migration along the ecliptic, crosses the celestial equator. The vernal equinox is located in the constellation Aries and is designated by an open circle in Figure A.1. The second coordinate in the equatorial system is the *declination* (δ). This is the angular displacement (in degrees, arcminutes ($'$), and arcseconds ($''$)), measured north or south of the celestial equator. Positive declinations are northern ones.

The Earth's rotational axis precesses slowly in time, with an associated period of 26,000 yr. Consequently, both the north and south poles of the equatorial system, as well as the celestial equator itself,

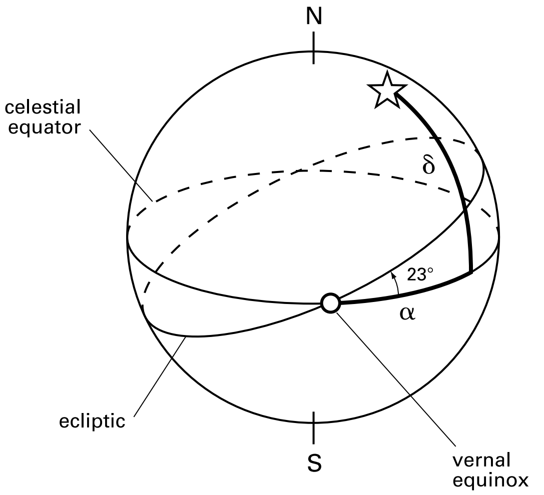


Figure A.1 Equatorial coordinate system. The celestial sphere, with the observer at the center, and the north and south celestial poles are indicated. The right ascension α is measured relative to the vernal equinox (*open circle*). This latter point is the intersection of the celestial equator and the northern leg of the ecliptic. The declination δ is measured from the celestial equator.

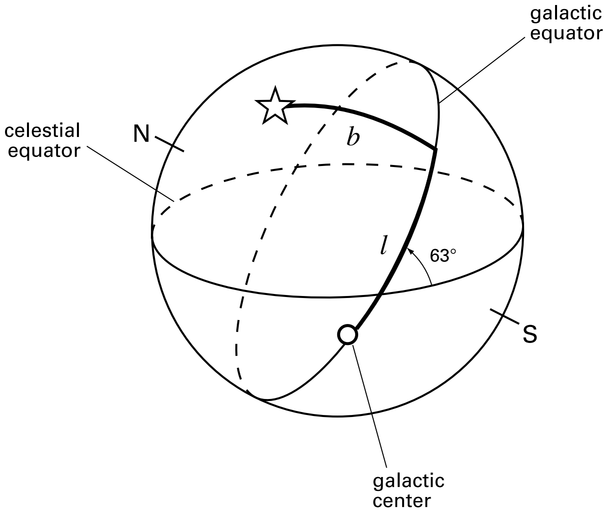


Figure A.2 Galactic coordinate system. The Galactic equator is the intersection of the Milky Way's plane with the celestial sphere. This equator is tilted relative to the celestial one; the Galactic north and south poles are indicated. The longitude l is measured relative to the Galactic center (*open circle*). The latitude b is is the angular distance above or below the Galactic equator.

drift relative to the background of distant stars. This phenomenon, known as *precession of the equinoxes*, continually alters the coordinate values for any object. In any particular observation, these values are referred to the appropriate *epoch*, the current one being the year 2000. Our maps in this book reproduce the coordinates as published in the literature. The reader who needs a precise location for any object should refer to the original article, as listed in the Sources section, to ascertain the epoch used. Standard programs can, if necessary, shift the coordinate values to the current ones.

In the *galactic* coordinate system, one places objects with respect to the plane of the Milky Way (see Figure A.2). This plane intersects the celestial sphere in the *Galactic equator*, a band that is tilted by 63° from the celestial one. The *longitude* (l) of the object is measured eastward along the Galactic equator. Here, the zero point is the Galactic center (open circle in the figure), which is located in the constellation Sagittarius, in the Southern hemisphere. For the *latitude* (b), we measure the angle north or south of the Galactic equator. Both the longitude and latitude are given in degrees, arcminutes, and arcseconds. A positive b-value indicates that the object is north of the equator. The origin ($l = 0$; $b = 0$) corresponds, in equatorial coordinates, to ($\alpha = 17^h 45^m 37^s$; $\delta = -28^\circ 56' 10''$).

B The Two-Level System

We consider a species of atom or molecule of number density n dispersed throughout a gas of total density $n_{\rm tot}$ and homogeneous composition. This species has only two energy levels, separated by ΔE (Figure B.1). In any real system, there are always other levels connected by possible physical transitions. Our two-level approximation is valid to the extent that these other transitions are slow compared to the one of interest. We include the possibility of degeneracy, *i. e.*, we suppose that there exist g_u and g_l sublevels of identical energy in the upper and lower levels, respectively. Our problem is to find the level populations n_u and n_l as a function of the ambient kinetic temperature $T_{\rm kin}$ and density $n_{\rm tot}$.

As illustrated in the figure, each atom in the lower level can be excited both collisionally and radiatively. The total rate of collisional excitations per unit time and per unit volume can be written $\gamma_{\rm lu}\, n_{\rm tot}\, n_l$, where the coefficient $\gamma_{\rm lu}$ depends on atomic properties of the species of interest and the background gas, as well as on their relative velocity distribution. The probability per unit time of a single atom being excited radiatively must be proportional to the ambient radiation intensity. Thus, we write this probability as $B_{\rm lu}\, \bar{J}\, n_l$. Here $B_{\rm lu}$ is the *Einstein coefficient for absorption*. The quantity $\bar{J}$ is related to the mean intensity J_ν by

$$\bar{J} \equiv \int_0^\infty J_\nu\, \phi(\nu)\, d\nu \tag{B.1}$$

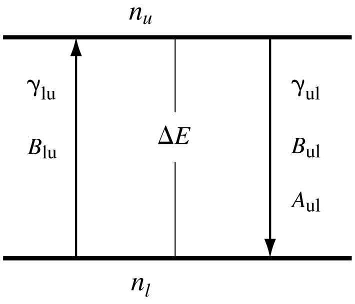


Figure B.1 Processes governing the populations in the two-level system. These include: collisions with other molecules, interaction with ambient radiation, and spontaneous emission.

where $\phi(\nu)d\nu$ is the relative efficiency with which a photon of frequency ν and $\nu + d\nu$ excites an upward transition. This absorption profile is normalized so that

$$\int_0^\infty \phi(\nu)\, d\nu = 1 \ . \tag{B.2}$$

Note that $\phi(\nu)$ is always strongly peaked at $\nu_\circ \equiv \Delta E/h$. The factors determining this function are discussed in Appendix E.

Turning to downward transitions, the volumetric collisional rate is $\gamma_{\rm ul}\, n_{\rm tot}\, n_u$, with $\gamma_{\rm ul}$ being the deexcitation coefficient. Two kinds of radiative transition can occur. First, an atom can spontaneously emit a photon of energy near $h\nu_\circ$ and drop to the lower level. The corresponding rate is $A_{\rm ul}\, n_u$, where $A_{\rm ul}$, the *Einstein coefficient for spontaneous emission*, is a function only of atomic properties. Secondly, an ambient photon of the right frequency can stimulate the excited atom to radiate a photon of identical energy and momentum, thereby again decaying to the lower level. Here, the associated volumetric rate of transitions is $B_{\rm ul}\, \bar{J}\, n_u$, where $B_{\rm ul}$ is the *Einstein coefficient for stimulated emission.*

In order for the populations of both levels to remain constant in time, we need

$$\gamma_{\rm lu}\, n_{\rm tot}\, n_l \ + \ B_{\rm lu}\, \bar{J}\, n_l = \gamma_{\rm ul}\, n_{\rm tot}\, n_u \ + \ B_{\rm ul}\, \bar{J}\, n_u \ + \ A_{\rm ul}\, n_u \ . \tag{B.3}$$

A relation between $\gamma_{\rm lu}$ and $\gamma_{\rm ul}$ can be obtained by considering a value of $n_{\rm tot}$ so high that radiative transitions are negligible compared to collisional ones. From equation (B.3), we then have $\gamma_{\rm lu}/\gamma_{\rm ul} = n_u/n_l$. In this high-density limit, the two-level system reaches *local thermodynamic equilibrium (LTE)* with the ambient temperature $T_{\rm kin}$, so that the level population obeys the Boltzmann relation. We therefore find

$$\frac{\gamma_{\rm lu}}{\gamma_{\rm ul}} = \frac{g_u}{g_l} \exp\left(-\frac{\Delta E}{k_B T_{\rm kin}}\right) \ , \tag{B.4}$$

an equation which holds generally at any density $n_{\rm tot}$.

We next seek relations between the Einstein coefficients. Accordingly, we consider the opposite limit, *i. e.*, a value of $n_{\rm tot}$ so low that radiative transitions completely dominate. Under these conditions, the two-level system is in thermodynamic equilibrium with the radiation field. The level populations again follow the Boltzmann relation, but with the exponential now evaluated at the radiation temperature $T_{\rm rad}$ that characterizes the photon energy distribution. Neglecting $n_{\rm tot}$ in equation (B.3) and solving for $\bar{J}$, we find

$$\begin{aligned} \bar{J} &= \frac{A_{\rm ul}\, n_u}{B_{\rm lu} n_l - B_{\rm ul} n_u} \\ &= \frac{A_{\rm ul}/B_{\rm ul}}{(g_l B_{\rm lu}/g_u B_{\rm ul}) \exp(\Delta E/k_B T_{\rm rad}) - 1} \ , \end{aligned} \tag{B.5}$$

where the second form utilizes the Boltzmann relation for the ratio n_u/n_l.

Under the assumed condition of thermodynamic equilibrium, the radiation intensity J_ν obeys Planck's law:

$$\begin{aligned} J_\nu &= B_\nu \\ &= \frac{2h\nu^3/c^2}{\exp(h\nu/k_B T_{\rm rad}) - 1} \ . \end{aligned} \tag{B.6}$$

This distribution is much broader than the profile function $\phi(\nu)$ in equation (B.1), so that substitution of (B.6) into (B.1) yields

$$\begin{aligned} \bar{J} &\approx B_{\nu_\circ} \\ &= \frac{2h\nu_\circ^3/c^2}{\exp(\Delta E/k_B T_{\rm rad}) - 1} \ . \end{aligned} \tag{B.7}$$

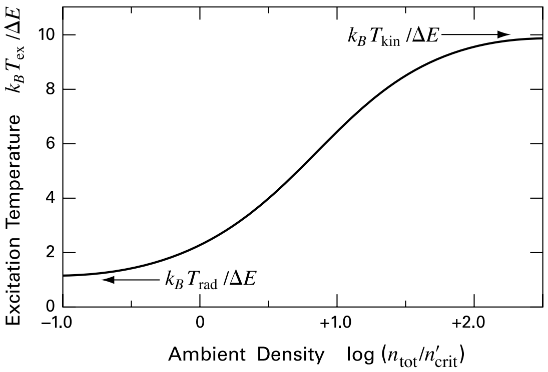


Figure B.2 Variation of the excitation temperature $T_{\rm ex}$ as a function of the ambient density $n_{\rm tot}$. In this numerical example, $T_{\rm rad} = \Delta E/k_B$, while $T_{\rm kin} = 10\,T_{\rm rad}$.

Comparison of (B.7) with (B.5) gives the desired relations between the Einstein coefficients:

$$A_{\rm ul} = \frac{2h\nu_\circ^3}{c^2}\,B_{\rm ul} \tag{B.8a}$$

$$g_l\,B_{\rm lu} = g_u\,B_{\rm ul} \tag{B.8b}$$

These relations, like that in equation (B.4), are valid also in environments far out of thermodynamic equilibrium.

We can now use equations (B.4) and (B.8) in (B.3) to derive an expression for n_u/n_l valid for any $n_{\rm tot}$ and $T_{\rm kin}$. We find

$$\frac{n_u}{n_l}\left[1 + \frac{n_{\rm crit}}{n_{\rm tot}}\left(1 + \frac{c^2\bar{J}}{2h\nu_\circ^3}\right)\right] = \frac{g_u}{g_l}\left[\exp\left(-\frac{\Delta E}{k_B T_{\rm kin}}\right) + \left(\frac{n_{\rm crit}}{n_{\rm tot}}\right)\frac{c^2\bar{J}}{2h\nu_\circ^3}\right] , \tag{B.9}$$

where $n_{\rm crit} \equiv A_{\rm ul}/\gamma_{\rm ul}$ is the critical density introduced in § 5.2. Equation (B.9) can be recast into a more physically transparent form by using $T_{\rm rad}$ instead of $\bar{J}$, through equation (B.7). Note that the actual radiation field need not be Planckian to make this substitution. If we further express n_u/n_l through the excitation temperature $T_{\rm ex}$ as in equation (5.14), we finally derive

$$\exp\left(-\frac{\Delta E}{k_B T_{\rm ex}}\right) = f_{\rm coll}\exp\left(-\frac{\Delta E}{k_B T_{\rm kin}}\right) + (1 - f_{\rm coll})\exp\left(-\frac{\Delta E}{k_B T_{\rm rad}}\right) . \tag{B.10}$$

Here the parameter $f_{\rm coll}$ represents the fraction of all downward transitions due to collisions:

$$f_{\rm coll} \equiv \frac{n_{\rm tot}}{n_{\rm tot} + n_{\rm crit}(1 + c^2\bar{J}/2h\nu_\circ^3)} . \tag{B.11}$$

For fixed values of $n_{\rm tot}$, $T_{\rm kin}$, and $T_{\rm rad}$, we may solve equation (B.10) for $T_{\rm ex}$, *i. e.*, for n_u/n_l. The individual level populations n_u and n_l follow by further imposing the condition $n_u + n_l = n$.

Equation (B.10) shows that $T_{\rm ex}$ always lies between $T_{\rm rad}$ and $T_{\rm kin}$. When collisions are relatively rare ($f_{\rm coll} \ll 1$), n_u and n_l equilibrate with the radiation field, and $T_{\rm ex} \approx T_{\rm rad}$. As $n_{\rm tot}$ increases and collisions become more important, the ratio n_u/n_l first rises proportionally to $n_{\rm tot}$. For $n_{\rm tot}$ sufficiently greater than $n_{\rm crit}$, equation (B.11) indicates that $f_{\rm coll} \rightarrow 1$, so that $T_{\rm ex} \rightarrow T_{\rm kin}$ and the system comes into LTE. Figure B.2 shows the behavior of $T_{\rm ex}$ as a function of $n_{\rm tot}/n'_{\rm crit}$, where $n'_{\rm crit} \equiv n_{\rm crit}(1 + c^2\bar{J}/2h\nu_\circ^3)$. The smooth increase of $T_{\rm ex}$, and therefore of n_u/n_l, implies that the emission from the $J = 1 \rightarrow 0$ transition will also rise smoothly and then level off. In particular, the local maximum in emissivity seen in Figure 5.4 is *not* present, reflecting the fact that the rotational ladder of CO is poorly approximated as a two-level system, even at the relatively low kinetic temperatures of quiescent molecular clouds.

C Transfer of Radiation in Spectral Lines

As we saw in Chapter 2 (equation (2.20)), the propagation of the specific intensity I_ν is governed by the fundamental equation

$$\frac{dI_\nu}{ds} = -\alpha_\nu\, I_\nu + j_\nu \ , \tag{C.1}$$

where α_ν was written as $\rho\kappa_\nu$. We suppose that a line of sight passing through the medium of interest (*e. g.*, a molecular cloud) covers a total path length $s = \Delta s$, where $s = 0$ at the backside of the medium (see Figure C.1). In general, the absorption coefficient α_ν and emission coefficient j_ν will be functions of s. We can write down the general solution of (C.1) by noting that

$$\frac{dI_\nu}{ds} + \alpha_\nu\, I_\nu = \exp\left[-\int_0^s \alpha_\nu(s')\, ds'\right] \frac{d}{ds}\left(I_\nu \exp\left[\int_0^s \alpha_\nu(s')\, ds'\right]\right) \ .$$

Equation (C.1) then becomes

$$\frac{d}{ds}\left(I_\nu \exp\left[\int_0^s \alpha_\nu(s')\, ds'\right]\right) = j_\nu \exp\left[\int_o^s \alpha_\nu(s')\, ds'\right] \ .$$

Integration from $s = 0$ to $s = \Delta s$ then yields the general solution:

$$\begin{aligned} I_\nu(\Delta s) = {} & I_\nu(0) \exp\left[-\int_0^{\Delta s} \alpha_\nu(s')\, ds'\right] \\ & + \int_0^{\Delta s} ds'\, j_\nu(s') \exp\left[-\int_{s'}^{\Delta s} \alpha_\nu(s'')\, ds''\right] \ . \end{aligned} \tag{C.2}$$

Here, $I_\nu(0)$ is the specific intensity of any background radiation field impinging at $s = 0$ (Figure C.1). If we now specialize to a *uniform* medium in which both α_ν and j_ν are spatial constants, equation (C.2) simplifies to

$$I_\nu(\Delta s) = I_\nu(0) \exp\left[-\alpha_\nu\, \Delta s\right] + \frac{j_\nu}{\alpha_\nu}\left(1 - \exp\left[-\alpha_\nu\, \Delta s\right]\right) \ . \tag{C.3}$$

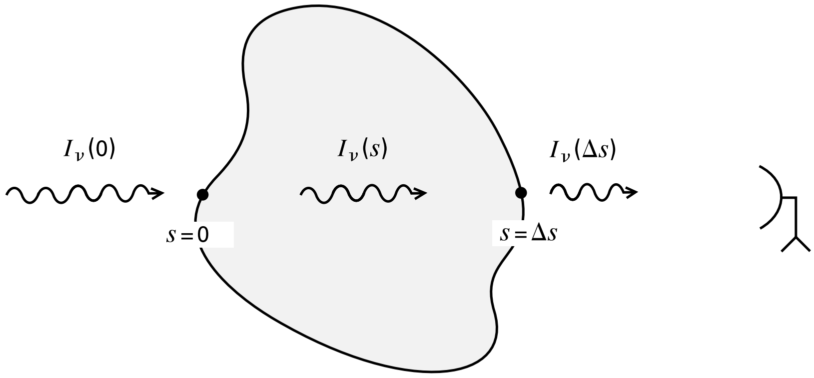


Figure C.1 Propagation of radiation through a gaseous medium. Both the absorption by the medium and its emissivity change the specific intensity I_ν over the path length Δs.

We consider the case in which both emission and absorption are due to transitions between two discrete levels in an atom or molecule. Employing the notation of Appendix B, the volumetric transition rate due to spontaneous emission is $A_{\rm ul} n_u$. We assume that this emission is isotropic, so that the coefficient j_ν, which is measured per unit solid angle, is given by

$$j_\nu = \frac{h\nu_\circ}{4\pi}\, n_u\, A_{\rm ul}\, \phi(\nu)\ . \tag{C.4}$$

Here, $\phi(\nu)$ is the relative probability of emission into a photon frequency ν near the line center frequency $\nu_\circ$. This profile function can generally be taken as identical to the absorption profile introduced under the same notation in Appendix B.

The volumetric absorption rate is $B_{\rm lu}\bar{J} n_l$. Hence, the energy absorbed per unit time per unit solid angle is $(h\nu/4\pi) B_{\rm lu}\bar{J} n_l$. The contribution to the absorption coefficient, which is measured per unit frequency, is

$$(\alpha_\nu)_{\rm true} = \frac{h\nu_\circ}{4\pi}\, B_{\rm lu}\, n_l\, \phi(\nu)\ . \tag{C.5}$$

The reader may check that this expression, upon integration over solid angle and frequency, yields the correct volumetric absorption rate in the case of an isotropic radiation field ($J_\nu = I_\nu$). However, equation (C.5) holds even when the radiation is anisotropic. The subscript indicates that this is a *true* absorption process. In contrast, stimulated emission can be viewed as negative absorption. Such an identification is natural because the stimulated emission rate is also proportional to the ambient radiation intensity. By analogy with equation (C.5), we can write down the associated coefficient as

$$(\alpha_\nu)_{\rm stim} = \frac{h\nu_\circ}{4\pi}\, B_{\rm ul}\, n_u\, \phi(\nu)\ . \tag{C.6}$$

The total absorption coefficient, corrected for stimulated emission, is then

$$\alpha_\nu = \frac{h\nu_\circ}{4\pi}\, (n_l\, B_{\rm lu} - n_u\, B_{\rm ul})\, \phi(\nu)\ . \tag{C.7}$$

From (C.4) and (C.7) we find that

$$\begin{aligned} \frac{j_\nu}{\alpha_\nu} &= \frac{n_u\, A_{\rm ul}}{n_l\, B_{\rm lu} - n_u\, B_{\rm ul}} \\ &= \frac{2h\nu_\circ^3/c^2}{\exp(h\nu_\circ/k_B T_{\rm ex}) - 1}\ , \end{aligned} \tag{C.8}$$

where we have used equation (B.8) and the definition of $T_{\rm ex}$ in equation (5.14). After specializing to $\nu = \nu_\circ$ in equation (C.3), substitution of (C.8) yields

$$I_{\nu_\circ}(\Delta\tau_\circ) = I_{\nu_\circ}(0)\, {\rm e}^{-\Delta\tau_\circ} + \frac{2h\nu_\circ^3/c^2}{\exp(h\nu_\circ/k_B T_{\rm ex}) - 1}\left(1 - {\rm e}^{-\Delta\tau_\circ}\right)\ . \tag{C.9}$$

Here we have set $\Delta\tau_\circ \equiv \Delta\tau_{\nu_\circ}$, where the medium's optical thickness $\Delta\tau_\nu$ is defined as $\alpha_\nu \Delta s$.

The quantity of interest to the observer is not I_ν itself, but rather the *difference* between I_ν and the background intensity. Accordingly, we define a *brightness temperature* T_B by

$$T_B \equiv \frac{c^2}{2\,\nu^2\, k_B}\, [I_\nu(\Delta\tau_\nu) - I_\nu(0)]\ . \tag{C.10}$$

To motivate this definition, recall the Planckian formula for the specific intensity at frequency ν from a blackbody source at temperature T:

$$B_\nu = \frac{2h\nu^3/c^2}{\exp(h\nu/k_B T) - 1} \tag{C.11}$$

In the Rayleigh-Jeans limit appropriate for low-energy transitions ($h\nu \ll k_B T$), we see that $T/B_\nu \rightarrow c^2/2\nu^2 k_B$. Note also that equations (C.8) and (C.11) together imply that j_ν/α_ν, a quantity known as the *source function*, is simply $B_{\nu_\circ}(T_{\rm ex})$ in the case of line radiation. Equation (C.9) further shows that $I_{\nu_\circ} \rightarrow B_{\nu_\circ}(T_{\rm ex})$ in the limit $\Delta\tau_\circ \gg 1$, as claimed in § 6.1.

The radio astronomer usually reports observations in terms of an *antenna temperature* T_A, defined as

$$T_A \equiv \eta \, \frac{\Omega_S}{\Omega_A} \, T_B \ . \tag{C.12}$$

Here, η is the *beam efficiency*, which accounts for such factors as the losses in the system's optics. The quantities Ω_S and Ω_A are, respectively, the solid angles subtended by the source and the telescope beam. The ratio of these two solid angles, known as the *beam dilution factor*, can be less than unity for observations of a clumpy molecular cloud, and is always very small for an interstellar maser.

If we make the final assumption that the background radiation field is Planckian with an associated temperature $T_{\rm bg}$, then the definition (C.10) applied to (C.9) yields the *detection equation* introduced in § 6.1:

$$T_{B_\circ} = T_\circ \left[f(T_{\rm ex}) - f(T_{\rm bg}) \right] \left[1 - {\rm e}^{-\Delta\tau_\circ} \right] \ , \tag{C.13}$$

where $T_{B_\circ} \equiv T_B(\nu = \nu_\circ)$, $T_\circ \equiv h\nu_\circ/k_B$, and

$$f(T) \equiv \left[\exp(T_\circ/T) - 1 \right]^{-1} \ . \tag{C.14}$$

Since the absorption coefficient $\alpha_{\nu_\circ}$ is proportional to the total volume density n of the species of interest, the optical thickness $\Delta\tau_\circ$ can also be written in terms of the column density $N \equiv n\Delta s$. Using equation (C.7) for $\alpha_{\nu_\circ}$, we have

$$\begin{aligned} \Delta\tau_\circ &= \frac{h\nu_\circ}{4\pi} \left(n_l \, B_{\rm lu} - n_u \, B_{\rm ul} \right) \phi(\nu_\circ) \, \Delta s \\ &= \frac{c^2 A_{\rm ul} (g_u/g_l)}{8\pi\nu_\circ^2} \left[1 - \exp\left(-\frac{\Delta E}{k_B T_{\rm ex}} \right) \right] \frac{n_l \Delta s}{\Delta\nu} \ . \end{aligned} \tag{C.15}$$

Here, we have again used equations (B.8) relating the Einstein coefficients and equation (5.14) for $T_{\rm ex}$. In addition, we have replaced $\phi(\nu_\circ)$ by the inverse of a characteristic frequency width $\Delta\nu$. This quantity equals the full-width half-maximum of the line in the case of a Doppler-broadened profile (Appendix E).

For the two-level system considered in Appendix B, $n_l \approx n$ and $n_l \Delta s \approx N$, so that equation (C.15) expresses the desired relation between $\Delta\tau_\circ$ and N. Such an approximation is not adequate, however, for a system in which the two levels are closely spaced rungs in a ladder. Consider, as in § 6.1, the lowest rotational transition of CO. Here the upper level corresponds to $J = 1$ and the lower to $J = 0$. Assuming that a single excitation temperature characterizes *all* the level populations, then repeated application of the definition for $T_{\rm ex}$ yields

$$\begin{aligned} n &= n_\circ \left[1 + \frac{g_1}{g_0} \exp\left(-\frac{\Delta E_{10}}{k_B T_{\rm ex}} \right) + \frac{g_2}{g_1} \exp\left(-\frac{\Delta E_{20}}{k_B T_{\rm ex}} \right) + \ldots \right] \\ &\equiv n_\circ \, Q \ , \end{aligned} \tag{C.16}$$

where the *partition function* Q is given in this case by

$$\begin{aligned} Q &= \sum_{j=0}^{\infty} (2J+1) \exp\left[-\frac{BhJ(J+1)}{k_B T_{\rm ex}} \right] \\ &= \sum_{j=0}^{\infty} (2J+1) \exp\left[-\frac{T_\circ J(J+1)}{2\,T_{\rm ex}} \right] \ . \end{aligned} \tag{C.17}$$

In writing equation (C.17), the formula (5.6) for the rotational energy has been employed. If $T_{\rm ex} \ll T_\circ$, we effectively have a two-level system, since $Q \rightarrow 1$ and $n_\circ \rightarrow n$. Since, however, $T_\circ$ for ^{12}C^{16}O is only 5.5 K, it is more useful to consider the opposite regime. Since the index J increases by unit intervals in (C.17), we may adopt the approximation

$$\begin{aligned} Q &\approx \int_0^\infty dJ\,(2J+1)\exp\left[-\frac{T_\circ J(J+1)}{2\,T_{\rm ex}}\right] \\ &= 2\,T_{\rm ex}/T_\circ\ . \end{aligned} \tag{C.18}$$

Substitution of this result into (C.15) yields, after rearrangement, equation (6.9) in the text.

D Derivation of the Virial Theorem

We form the scalar product of the position vector $\boldsymbol{r}$ with the equation of motion (3.6) and integrate over volume. For the manipulations that follow, it is most convenient to work in Cartesian coordinates, where $\boldsymbol{r}$ has the components x_i ($i = 1,\, 2,\, 3$). We adopt the notation

$$\partial_i \equiv \frac{\partial}{\partial x_i}\ ,$$

and follow the Einstein summation convention for repeated indices. The identities $\partial_i x_i = 3$ and $\partial_i x_j = \delta_{ij}$ will be used extensively.

Using the definition (3.4), the lefthand side of equation (3.6) becomes

$$\int \rho\, x_i \left(\frac{\partial v_i}{\partial t}\right)_{\boldsymbol{x}} d^3\boldsymbol{x} \,+\, \int \rho\, x_i\, v_j\, \partial_j v_i\ d^3\boldsymbol{x}\ .$$

The first integrand can be written

$$\rho\, x_i \left(\frac{\partial v_i}{\partial t}\right)_{\boldsymbol{x}} = \rho\left[\frac{\partial\,(x_i v_i)}{\partial t}\right]_{\boldsymbol{x}}\ , \tag{D.1}$$

while the second integrand is

$$\begin{aligned} \rho\, x_i\, v_j\, \partial_j v_i &= \rho\, v_j\, \partial_j\,(x_i\, v_i) \,-\, \rho\, v_j\, v_i\, \partial_j x_i \\ &= \partial_j\,(\rho\, v_j\, x_i\, v_i) \,-\, (x_i\, v_i)\ \partial_j\,(\rho\, v_j) \,-\, \rho\, v_j\, v_i\, \delta_{ij} \\ &= \partial_j\,(\rho\, v_j\, x_i\, v_i) \,+\, \left(\frac{\partial \rho}{\partial t}\right)_{\boldsymbol{x}} (x_i\, v_i) \,-\, \rho\, v_i\, v_i\ , \end{aligned} \tag{D.2}$$

where we have used the mass continuity equation (3.7) to replace $\partial_j\,(\rho v_j) \equiv \boldsymbol{\nabla}\cdot(\rho \boldsymbol{v})$.

The first righthand term in the final form of equation (D.2) is the divergence of a vector. After the volume integration, this term generates a surface integral which vanishes in the absence of any external mass flux. Adding equations (D.1) to (D.2) and integrating, we find

$$\int \rho\, x_i \left(\frac{\partial v_i}{\partial t}\right)_{\boldsymbol{x}} d^3\boldsymbol{x} \,+\, \int \rho\, x_i\, v_j\, \partial_j v_i\ d^3\boldsymbol{x} = \frac{\partial}{\partial t}\int \rho\, x_i\, v_i\ d^3\boldsymbol{x} \,-\, 2\,\mathcal{T}\ , \tag{D.3}$$

where the kinetic energy $\mathcal{T}$ is defined in equation (3.10). The integrand on the right is

$$\begin{aligned}\rho\, x_i\, v_i &= \frac{1}{2}\,\rho\, v_i\, \partial_i\,(x_j\, x_j)\\ &= \frac{1}{2}\,\partial_i\,(\rho\, v_i\, x_j\, x_j) - \frac{1}{2}\,x_j\, x_j\, \partial_i\,(\rho\, v_i)\\ &= \frac{1}{2}\,\partial_i\,(\rho\, v_i\, x_j\, x_j) + \frac{1}{2}\left(\frac{\partial\rho}{\partial t}\right)_{\boldsymbol{x}} x_j\, x_j\ .\end{aligned}$$

Invoking once more the condition of zero external mass flux, we have

$$\int \rho\, x_i\, v_i\ d^3\boldsymbol{x} = \frac{1}{2}\,\frac{\partial I}{\partial t}\ , \tag{D.4}$$

where I is a scalar quantity resembling the moment of inertia:

$$I \equiv \int \rho\, x_j\, x_j\ d^3\boldsymbol{x}\ . \tag{D.5}$$

With the help of (D.4), equation (D.3) now becomes

$$\int \rho\, x_i \left(\frac{\partial v_i}{\partial t}\right)_{\boldsymbol{x}} d^3\boldsymbol{x} + \int \rho\, x_i\, v_j\, \partial_j v_i\ d^3\boldsymbol{x} = \frac{1}{2}\,\frac{\partial^2 I}{\partial t^2} - 2\,\mathcal{T}\ . \tag{D.6}$$

We turn next to the righthand terms of equation (3.6). We first have

$$\begin{aligned}-\int x_i\, \partial_i P\ d^3\boldsymbol{x} &= -\int \partial_i\,(x_i P)\ d^3\boldsymbol{x} + \int (\partial_i x_i)\, P\ d^3\boldsymbol{x}\\ &= -\int P\, \boldsymbol{r}\cdot\boldsymbol{n}\, d^2\boldsymbol{x} + 3\int P\ d^3\boldsymbol{x}\ ,\end{aligned}$$

where $\boldsymbol{n}$ denotes an outward normal vector in the surface integral. For a nonrelativistic fluid, the internal pressure is two-thirds the energy density in thermal motion, so we find

$$-\int x_i\, \partial_i P\ d^3\boldsymbol{x} = -\int P\, \boldsymbol{r}\cdot\boldsymbol{n}\, d^2\boldsymbol{x} + 2\,U\ , \tag{D.7}$$

where U is the total thermal energy in equation (3.11).

To handle the gravitational term in equation (3.6), we first write the potential $\Phi_g(\boldsymbol{r})$ as an integral over all the mass elements in the volume:

$$\Phi_g(\boldsymbol{r}) = -G\int \frac{\rho(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\, d^3\boldsymbol{x}\ .$$

We thus find

$$-\int \rho\, x_i\, \partial_i \Phi_g\ d^3\boldsymbol{x} = -G\iint \rho(\boldsymbol{r})\, \rho(\boldsymbol{r}')\, \frac{x_i(x_i - x_i')}{|\boldsymbol{r}-\boldsymbol{r}'|^3}\, d^3\boldsymbol{x}'\, d^3\boldsymbol{x}\ , \tag{D.8}$$

where we have used the identity

$$\partial_i \frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|} = -\frac{(x_i - x_i')}{|\boldsymbol{r}-\boldsymbol{r}'|^3}\ .$$

Within the double integral in (D.8), we next switch the dummy variables $\boldsymbol{r}$ and $\boldsymbol{r}'$. After changing the order of integration, we can add the resulting expression to the original integral, thereby doubling its value. We conclude that

$$\begin{aligned} -\int \rho\, x_i\, \partial_i \Phi_g \; d^3\boldsymbol{x} \;&=\; -\frac{1}{2}\, G \iint \rho(\boldsymbol{r})\, \rho(\boldsymbol{r}')\, \frac{|\boldsymbol{r}-\boldsymbol{r}'|^2}{|\boldsymbol{r}-\boldsymbol{r}'|^3}\, d^3\boldsymbol{x}'\, d^3\boldsymbol{x} \\ &=\; \frac{1}{2}\int \rho(\boldsymbol{r})\, \Phi_g(\boldsymbol{r})\, d^3\boldsymbol{x} \\ &=\; \mathcal{W}\ , \end{aligned} \tag{D.9}$$

where $\mathcal{W}$ is the gravitational potential energy of equation (3.12).

Lastly, we consider the two magnetic terms of equation (3.6). Integration of the first leads to

$$\begin{aligned} \frac{1}{4\pi}\int x_i\, B_j\, \partial_j B_i\; d^3\boldsymbol{x} \;&=\; \frac{1}{4\pi}\int \partial_j \left(x_i\, B_j\, B_i\right)\, d^3\boldsymbol{x} \;-\; \frac{1}{4\pi}\int B_i\, \partial_j\left(x_i\, B_j\right)\, d^3\boldsymbol{x} \\ &=\; \frac{1}{4\pi}\int \partial_j \left(x_i\, B_j\, B_i\right)\, d^3\boldsymbol{x} \;-\; \frac{1}{4\pi}\int B_i\, B_i\, d^3\boldsymbol{x} \\ &=\; \frac{1}{4\pi}\int (\boldsymbol{r}\cdot\boldsymbol{B})\, \boldsymbol{B}\cdot\boldsymbol{n}\, d^2\boldsymbol{x} \;-\; 2\,\mathcal{M}\ , \end{aligned} \tag{D.10}$$

where $\mathcal{M}$ is the magnetic energy defined in equation (3.13). Here we have used Maxwell's equation $\partial_j B_j \equiv \boldsymbol{\nabla}\cdot\boldsymbol{B} = 0$ to eliminate one integral in the second line of (D.10). Integration of the last term of equation (3.6) gives

$$\begin{aligned} -\frac{1}{8\pi}\int x_i\, \partial_i\left(B_j\, B_j\right)\, d^3\boldsymbol{x} \;&=\; -\frac{1}{8\pi}\int \partial_i\left(x_i\, B_j\, B_j\right)\; d^3\boldsymbol{x} \\ &\quad +\frac{1}{8\pi}\int \left(\partial_i x_i\right)\; B_j\, B_j\; d^3\boldsymbol{x} \\ &=\; -\frac{1}{8\pi}\int B^2\, \boldsymbol{r}\cdot\boldsymbol{n}\, d^2\boldsymbol{x} \;+\; 3\,\mathcal{M}\ . \end{aligned} \tag{D.11}$$

Equations (D.6), (D.7), (D.9), (D.10), and (D.11) add together to yield the virial theorem:

$$\begin{aligned} \frac{1}{2}\,\frac{\partial^2 I}{\partial t^2} \;&=\; 2\,\mathcal{T} \;+\; 2\,U \;+\; \mathcal{W} \;+\; \mathcal{M} \\ &\quad -\int\left(P + \frac{B^2}{8\pi}\right) \boldsymbol{r}\cdot\boldsymbol{n}\, d^2\boldsymbol{x} \;+\; \frac{1}{4\pi}\int (\boldsymbol{r}\cdot\boldsymbol{B})\, \boldsymbol{B}\cdot\boldsymbol{n}\, d^2\boldsymbol{x}\ . \end{aligned} \tag{D.12}$$

E Spectral Line Broadening

E.1 Natural Width

We consider an isolated, stationary atom or molecule with only an upper (u) and lower (l) energy level, separated by ΔE. Let $\nu_0 \equiv \Delta E/h$ be the central frequency of the spectral line emitted during a downward transition. Our first goal is to determine the *natural line broadening, i.e.*, how the radiation intensity varies with frequency, in the vicinity of ν_0. Thus, we will obtain the emission profile function $\phi(\nu - \nu_0)$. Along the way, we will find a quantum-mechanical expression for the Einstein coefficient $A_{\rm ul}$.

Let $\Psi(t)$ be the wavefunction describing *both* the atom and its surrounding radiation field. This function obeys Schrödinger's equation:

$$H\,\Psi \,=\, i\,\hbar\,\frac{\partial\Psi}{\partial t}\ . \tag{E.1}$$

We decompose the Hamiltonian operator H into

$$H \,=\, H^\circ \,+\, V\ . \tag{E.2}$$

Here, V represents the interaction between the atom and the electromagnetic field, while H° governs the unperturbed system of separate atom and field. We assume that the disturbance is relatively weak and apply the methods of perturbation theory.

If V were strictly zero, there would still exist a set of wavefunctions obeying equation (E.1). Each of these solutions, denoted $\Psi_n(t)$, corresponds to an energy E_n of the combined system and has the time dependence

$$\Psi_n(t) \,=\, \Psi_n\,{\rm e}^{-i\omega_n t}\ , \tag{E.3}$$

where $\omega_n \,\equiv\, E_n/\hbar$. Moreover, each Ψ_n, a function of spatial coordinates alone, is a solution to the time-independent Schrödinger equation:

$$H^\circ\,\Psi_n \,=\, E_n\,\Psi_n\ . \tag{E.4}$$

Since we are now concerned with emission, we let the initial state Ψ_0 represent the excited atom in the upper energy level together with zero photons in the radiation field, while Ψ_1 is the atomic ground state and a single photon.

We now expand $\Psi(t)$ in terms of the unperturbed eigenfunctions:

$$\begin{aligned}\Psi_n(t) \,&=\, \sum_{n=0}^{1} c_n(t)\,\Psi_n(t)\\ &=\, \sum_{n=0}^{1} c_n(t)\,\Psi_n\,{\rm e}^{-i\omega_n t}\ .\end{aligned} \tag{E.5}$$

Each coefficient $c_n(t)$ is a complex number whose square $c_n^*(t)c_n(t) \equiv |c_n(t)|^2$ gives the probability that the system is in the corresponding eigenstate at time t. Physically, we expect that $|c_1(t)|^2$ will grow, while $|c_0(t)|^2$ diminishes. Substitution of (E.5) into (E.1) yields

$$\begin{aligned}i\hbar\sum_{n=0}^{1}\left[\frac{dc_n}{dt} \,-\, i\omega_n c_n\right]\,\Psi_n\,{\rm e}^{-i\omega_n t} \,&=\, \sum_{n=0}^{1} c_n[H^\circ\,\Psi_n \,+\, V\,\Psi_n]{\rm e}^{-i\omega_n t}\\ &=\, \sum_{n=0}^{1} c_n[E_n\,\Psi_n \,+\, V\,\Psi_n]{\rm e}^{-i\omega_n t}\ ,\end{aligned} \tag{E.6}$$

where we have also applied equation (E.4). We next utilize the orthonormal property of the unperturbed eigenfunctions:

$$\int\Psi_m^*\,\Psi_n\,d^3x \,=\, \delta_{mn}\ . \tag{E.7}$$

Multiplying (E.6) through by Ψ_0^* and integrating yields, after simplification,

$$i\,\hbar\,\frac{dc_0}{dt} \;=\; c_0\,V_{00} \;+\; c_1\,V_{01}\,\mathrm{e}^{i(\omega_0-\omega_1)t} \;, \tag{E.8}$$

where we have adopted the convenient notation

$$V_{mn} \;\equiv\; \int \Psi_m^*\,V\;\Psi_n\,d^3x \;.$$

Similarly, multiplication of (E.6) by Ψ_1^* gives

$$i\,\hbar\,\frac{dc_1}{dt} \;=\; c_0\;V_{10}\,\exp^{i(\omega_1-\omega_0)t} \;+\; c_1\,V_{11} \;. \tag{E.9}$$

The task now is to solve the coupled differential equations (E.8) and (E.9). These equations are subject to the initial conditions

$$c_0\,(t=0) \;=\; 1 \qquad\qquad c_1\,(t=0) \;=\; 0 \;. \tag{E.10}$$

For times short enough that $|c_1(t)|^2$ has not grown close to unity, it is a reasonable approximation to ignore c_1 in equation (E.9) and write instead

$$i\,\hbar\,\frac{dc_1}{dt} \;=\; c_0\,V_{10}\,\mathrm{e}^{i(\omega_1-\omega_0)t} \;, \tag{E.11}$$

which has the integral solution

$$c_1(t) \;=\; -\frac{i\,V_{10}}{\hbar}\int_0^t c_0(t')\,\mathrm{e}^{i(\omega_1-\omega_0)t'}\,dt' \;. \tag{E.12}$$

Substituting (E.12) into (E.8), we obtain a differential equation for $c_0(t)$ alone:

$$\frac{dc_0}{dt} \;=\; -\frac{|V_{10}|^2}{\hbar^2}\int_0^t c_0(t')\,\mathrm{e}^{i(\omega_1-\omega_0)(t'-t)}\,dt' \;-\; \frac{ic_0V_{00}}{\hbar} \;. \tag{E.13}$$

Equation (E.13) is best solved through Fourier transforms. Omitting the details, the result, for $t \gg (\omega_0-\omega_1)^{-1}$, is

$$c_0(t) \;=\; \exp\left\{-\frac{i}{\hbar}\left[V_{00} \;+\; \frac{|V_{00}|^2}{E_0-E_1} \;-\; i\,\pi\,|V_{10}|^2\,\delta(E_1-E_0)\right]\,t\right\} \;, \tag{E.14}$$

where $\delta(E_1-E_0)$ is the Dirac delta function.

This expression for $c_0(t)$ has a clear physical interpretation. Referring again to the second form of equation (E.5), we see that the two imaginary terms in the exponential above effectively constitute a shift of the initial state energy E_0. The shift is small because $V_{00} \ll H_{00}^\circ = E_0$. The remaining term leads to an exponential decay of the occupation probability $|c_0(t)|^2$. Such behavior indicates a transition rate from the initial to the final state that is constant in time. The presence of the delta function in (E.14) means that no transition occurs unless the final state has the same *total* energy, *i.e.*, unless the emitted photon has energy equal to the difference of the atomic levels. In fact, there are many final, composite states of the same energy, each distinguished by the direction of the emitted photon, *i.e.*, by the vector wavenumber $\boldsymbol{k}$, where $\boldsymbol{k} \equiv E/\hbar c$. Suppose there are $\rho_E dE$ such states in the energy interval dE. Then the *total* transition rate to these states is $\gamma \equiv |c_0(t)|^2/dt$, or

$$\begin{aligned}\gamma \;&=\; \frac{2\,\pi\,|V_{10}|^2}{\hbar}\\ &=\; A_{\rm ul} \;.\end{aligned} \tag{E.15}$$

As indicated, the calculated rate is also equal to the Einstein coefficient for spontaneous emission associated with the atomic transition.

Leaving aside the small energy change in the initial state, which contributes a phase shift to the oscillating wavefunction $\Psi_0(t)$, we may write $c_0(t)$ in the simplified form

$$c_0(t) \;=\; \mathrm{e}^{-\gamma t/2} \;. \tag{E.16}$$

We now substitute this expression into the right side of equation (E.12), in order to obtain $c_1(t)$. We find

$$\begin{aligned} c_1(t) \;&=\; -\frac{i\,V_{10}}{\hbar}\int_0^t \mathrm{e}^{i(\omega_1-\omega_0)t'-\gamma t'/2}\,dt' \\ &=\; \frac{V_{10}}{\hbar}\,\frac{1-\mathrm{e}^{i(\omega_1-\omega_0)t-\gamma t/2}}{(\omega_1-\omega_0)+i\gamma/2} \;. \end{aligned} \tag{E.17}$$

Hence, the probability that the atom has decayed to its lower level is

$$\begin{aligned} |c_1(t)|^2 \;&=\; \frac{|V_{10}|^2}{\hbar^2}\,\frac{1\,-\,2\,\mathrm{e}^{-\gamma t/2}\cos(\omega_1-\omega_0)t\,+\,\mathrm{e}^{-\gamma t}}{(\omega_1-\omega_0)^2+\gamma^2/4} \\ &\approx\; \frac{|V_{10}|^2}{(E_1-E_0)^2\,+\,\gamma^2/4} \;, \end{aligned} \tag{E.18}$$

where the limiting expression holds for times such that $\gamma t \gg 1$.[1]

Once we recall that $E_1 - E_0 = h\nu - \Delta E = h(\nu - \nu_\circ)$, we see that our final asymptotic form for $|c_1(t)|^2$ gives the probability for emission of a photon of frequency ν. This probability, normalized so that its integral is unity, is the *Lorentz profile function*:

$$\phi_L(\nu) \;=\; \frac{\gamma/4\pi^2}{(\nu-\nu_\circ)^2+(\gamma/4\pi)^2} \;, \tag{E.19}$$

which we display graphically in Figure E.1. It is interesting that the same formula describes, in classical electromagnetic theory, the spectrum of radiation from an oscillating charged particle subject to a damping force per unit mass equal to γ times its velocity. However, this analogy is no guide once we consider multiple energy levels. As the simplest example, suppose that our initial ($n = 0$) level can decay to *either* the $n = 1$ or an even lower level, designated $n = 2$. Suppose further that $n = 1$ cannot decay into $n = 2$, so that $V_{12} = 0$. Then our previous method of solution shows that $|c_0(t)|^2$ still declines exponentially, but with a time constant equal to

$$\gamma \;=\; \frac{2\,\pi\,|V_{10}|^2}{\hbar} + \frac{2\,\pi\,|V_{20}|^2}{\hbar} \;.$$

The profile for the 0⟶1 spectral line is again given by (E.19), with this larger value of γ. The most general case is an ensemble of many levels connected by radiative transitions. We define the "width" of a single level to be the sum of the Einstein A-coefficients to all lower levels. Then the appropriate γ for the spectral line connecting any two levels is the sum of their respective widths.

Equation (E.19), although derived for the case of emission, also describes the profile of the corresponding *absorption* line. To see why, consider a spatially uniform gas composed of radiating atoms in an enclosure at temperature T. If the system is in thermal equilibrium, the radiation spectrum is isotropic

---

[1] Strictly speaking, the expression for $|c_1(t)|^2$ in (E.18) violates conservation of probability, *i. e.*, $|c_0|^2 + |c_1|^2 \neq 1$. This error stems from our neglect of the phase factor in equation (E.16) and has a minor effect on the final result embodied in (E.19), effectively shifting the central frequency $\nu_\circ$ by a small amount.

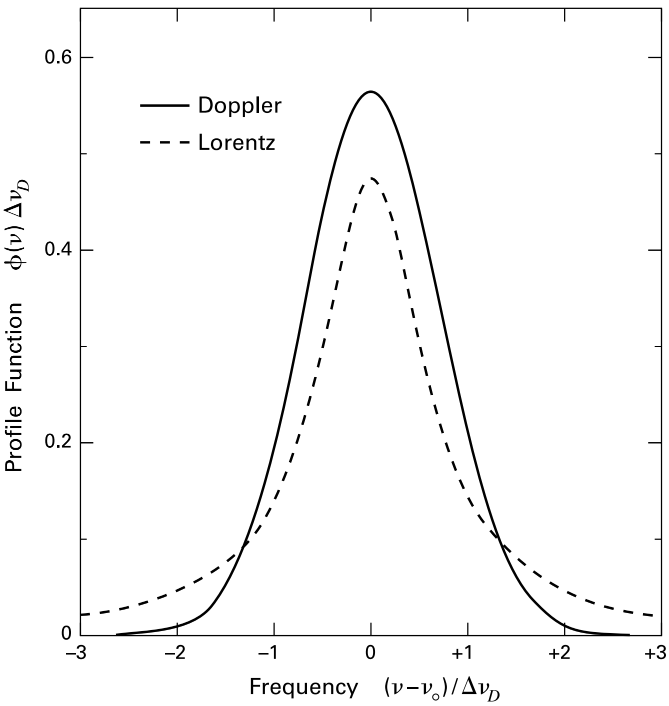


Figure E.1 Doppler (*solid*) and Lorentz (*dashed*) line profiles. The frequency is written as $(\nu - \nu_\circ)/\Delta\nu_D$, and the profile functions as $\phi(\nu)\,\Delta\nu_D$. We have chosen $\Delta\nu_D = 1.5 \times \gamma/4\pi$.

and Planckian. Referring to the radiative transfer equation (2.20), we have $I_\nu = B_\nu(T)$ and $dI_\nu/ds = 0$. Thus, the emission coefficient j_ν is related to the opacity κ_ν through

$$j_\nu \;=\; \rho\,\kappa_\nu\,B_\nu(T)\ , \tag{E.20}$$

where ρ is the gas density. Equation (E.20) is known as *Kirkhoff's law* and applies to any material whose emission depends only on its temperature and internal properties. The frequency dependence of j_ν in (E.20) is given by equation (E.19). Over the narrow interval $|\nu - \nu_\circ| \sim \gamma$, $B_\nu(T)$ changes negligibly. Hence, κ_ν and j_ν must share the same profile.

E.2 Thermal, Turbulent, and Collisional Broadening

Returning to emission lines, let us suppose that the atom has a velocity component V_r along the line of sight. Here, a positive V_r describes a velocity directed *away* from the observer. As a result of the Doppler effect, any photon emitted with frequency $\nu_\circ$ is actually detected at a different frequency ν, given by

$$\frac{\nu_\circ - \nu}{c} \;=\; \frac{V_r}{c}\ , \tag{E.21}$$

where we assume that $V_r \ll c$. If the emitting material is in thermal equilibrium at temperature T, the probability $p_{\rm therm}(V_r)dV_r$ that the velocity lies in the range dV_r is found from the Maxwell-Boltzmann relation:

$$p_{\rm therm}(V_r)\,dV_r \;=\; A\exp\left(-\frac{m\,V_r^2}{2\,k_B\,T}\right)\,dV_r\ , \tag{E.22}$$

where m is the atomic mass and A is a normalization constant. Since the velocity is related to the frequency by equation (E.21), this probability is effectively the profile function of the emitted radiation. After normalization, we obtain

$$\phi_D(\nu) \;=\; \frac{1}{\sqrt{\pi}\,\Delta\nu_D}\exp\left[-\frac{(\nu-\nu_\circ)^2}{\Delta\nu_D^2}\right]\ , \tag{E.23}$$

where $\Delta\nu_D$, the *Doppler width*, is

$$\Delta\nu_D \;\equiv\; \frac{\nu_\circ}{c}\left(\frac{2\,k_B\,T}{m}\right)^{1/2}\ . \tag{E.24}$$

Note that we obtained the coefficient in (E.23) by integrating $\nu-\nu_\circ$ from $-\infty$ to $+\infty$, *i. e.*, by making the reasonable assumption that $\Delta\nu_D \ll \nu_\circ$. The Doppler profile is also shown in Figure E.1.

The foregoing discussion employed a hypothetical emission line that was infinitely sharp in the atom's rest frame. Actually, the line will be subject to natural broadening. For each velocity V_r, the Lorentz emission profile is $\phi_L(\nu-\nu_\circ V_r/c)$. Hence, the net profile for the ensemble of atoms is

$$\phi(\nu) \;=\; \int_{-\infty}^{+\infty} dV_r\,p_{\rm therm}(V_r)\,\phi_L(\nu-\nu_\circ V_r/c)\ . \tag{E.25}$$

The explicit form of this function (known as the *Voigt profile*) is of less interest than its qualitative features. Under most astrophysical conditions, $\Delta\nu_D \gg \gamma/4\pi$, so that the *line core*, $|\nu-\nu_\circ| \lesssim \nu_D$, closely follows equation (E.23). However, the sharp falloff of the Gaussian implies that the Lorentz profile dominates in the *line wings*. Here, the intensity declines much more slowly, as $|\nu-\nu_\circ|^{-2}$.

In radio astronomy, observations of a spectral line are generally given in terms of the antenna temperature T_A, with the independent variable being the radial velocity V_r. From equation (E.22), the full-width half-maximum breadth of a purely thermally broadened line is

$$\begin{aligned}\Delta V_{\rm FWHM}({\rm therm}) \;&=\; \left(\frac{8\,\ln 2\,k_B\,T}{m}\right)^{1/2}\\ &=\; 2\,(\ln 2)^{1/2}\left(\frac{\Delta\nu_D}{\nu_\circ}\right)\,c\ ,\end{aligned} \tag{E.26}$$

where $(\Delta\nu_D/\nu_\circ)c$ is the one-dimensional velocity dispersion, *i.e.*, the root-mean-square value. If the radio beam includes many gas parcels in turbulent motion, this bulk, random velocity creates additional broadening. Suppose the distribution of turbulent velocities, $V_{\rm turb}$ along the line of sight is also Gaussian. Then the probability distribution $p_{\rm turb}(V_{\rm turb})$ analogous to equation (E.22) is

$$p_{\rm turb}(V_{\rm turb}) \;=\; A'\exp\left[-\frac{4\,\ln 2\,V_{\rm turb}^2}{\Delta V_{\rm FWHM}^2({\rm turb})}\right]\ , \tag{E.27}$$

where A' is another normalization constant and where $\Delta V_{\rm FWHM}({\rm turb})$ is the width of the distribution. This width is $2\,(\ln 2)^{1/2}$ times the corresponding one-dimensional velocity dispersion. For any fixed value of $V_{\rm turb}$, the observer detects a total velocity V_r only if the additional thermal component is $V_r - V_{\rm turb}$. Hence the net profile function is described by

$$\begin{aligned}p(V_r) \;&=\; \int_{-\infty}^{+\infty} dV_{\rm turb}\,p_{\rm turb}\,(V_{\rm turb})\,p_{\rm therm}(V_r - V_{\rm turb})\\ &=\; A''\int_{-\infty}^{+\infty} dV_{\rm turb}\exp\left[-\frac{4\,\ln 2\,V_{\rm turb}^2}{\Delta V_{\rm FWHM}^2({\rm turb})}\right]\exp\left[-\frac{4\,\ln 2\,(V_r - V_{\rm turb})^2}{\Delta V_{\rm FWHM}^2({\rm therm})}\right]\ .\end{aligned}$$

After factoring out the term involving V_r^2, the integral over $V_{\rm turb}$ can be evaluated by completing the square of the exponential argument. The result is another Gaussian:

$$p(V_r) \,=\, A''' \exp\left[-\frac{4\,\ln 2\, V_r^2}{\Delta V^2_{\rm FWHM}({\rm tot})}\right] \ , \tag{E.28}$$

where the *total* width is now given by

$$\Delta V^2_{\rm FWHM}({\rm tot}) \,=\, \Delta V^2_{\rm FWHM}({\rm therm}) \,+\, \Delta V^2_{\rm FWHM}({\rm turb}) \ . \tag{E.29}$$

At sufficiently high densities, the direct collision with neighboring atoms leads to additional broadening of any emission (or absorption) line. Since we will not have occasion to use this *collisional broadening* quantitatively, we limit ourselves to a heuristic account of its origin. Our analysis of the two-level atom shows that *any* process enhancing downward transitions from the excited level must also widen the corresponding spectral line. Indeed, if $\gamma_{\rm coll}$ is the mean rate of collision-induced transitions, we expect $c_\circ(t)$ to decay exponentially with a time constant of $2/(\gamma+\gamma_{\rm coll})$. In the absence of thermal motion, therefore, the profile function should again be a Lorentzian, but now with an associated width of $\gamma\,+\,\gamma_{\rm coll}$. A detailed derivation confirms the Lorentzian shape, but gives the true width as $\gamma\,+\,2\gamma_{\rm coll}$.

E.3 Rotational Broadening of Stellar Lines

A star that is rotating exhibits additional spectral broadening beyond any that can be attributed to thermal, turbulent, or collisional effects. The reason is that opposite limbs of the star have differing radial velocities. The corresponding frequency shifts are not seen directly, since one cannot resolve the stellar surface. Instead, the profile of the total emission across the projected surface is smeared out by an amount proportional to the rotation speed. Let us now obtain a quantitative expression for this broadening.

Figure E.2 depicts a spherical star of radius R_*, rotating with angular speed Ω_*. We erect a coordinate system with origin at the star's center, and with a z-axis pointing toward the observer. We further constrain the rotation axis to lie within the $y-z$ plane, and to be tilted by the inclination angle i with respect to the z-axis. Consider now a point lying on the surface. If $\boldsymbol{r} \,\equiv\, (x,y,z)$ is the vector displacement of this point from the origin, then its rotational velocity is

$$\boldsymbol{v} \,=\, \Omega_*\,\hat{\boldsymbol{n}} \,\times\, \boldsymbol{r} \ , \tag{E.30}$$

where $\hat{\boldsymbol{n}} \,\equiv\, (0,\,\sin\,i,\,\cos\,i)$ is a unit vector along the rotation axis. Radiation emanating from our point will be Doppler-shifted by an amount that depends on the velocity component v_z. Using equation (E.30), this is

$$\begin{aligned} v_z \,&=\, -x\,\Omega_*\,\sin\,i \\ &=\, -\frac{x}{R_*}\,V_{\rm eq}\,\sin\,i \ . \end{aligned} \tag{E.31}$$

Here, $V_{\rm eq} \,\equiv\, \Omega_* R_*$ is the star's equatorial velocity.

In our figure, a surface element with positive v_z approaches the observer. Thus, the central wavelength of any spectral line is changed from λ to λ', where

$$\begin{aligned} \lambda' \,-\, \lambda \,&=\, -\lambda_\circ\,\frac{v_z}{c} \\ &=\, \lambda_\circ\,\frac{x}{R_*}\,\frac{V_{\rm eq}}{c}\,\sin\,i \ . \end{aligned} \tag{E.32}$$

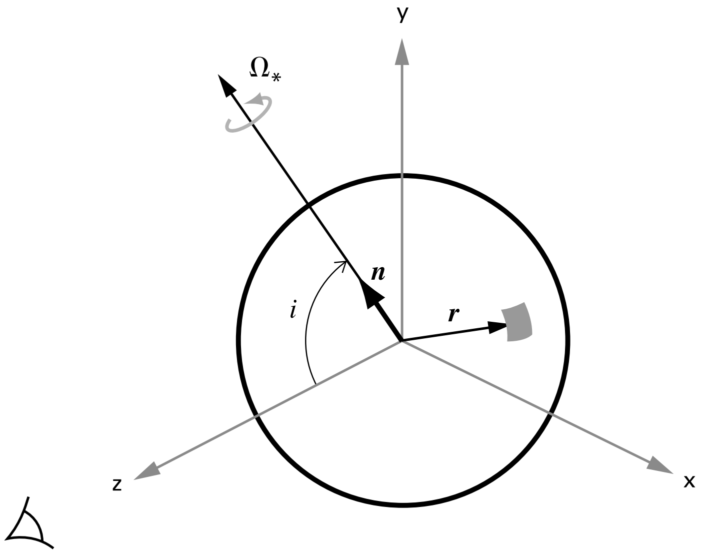


Figure E.2 Geometry of a rotating star. The rotation axis is in the $y - z$ plane, and tilted at an angle i with respect to the z-axis, where the latter coincides with the observer's line of sight. The unit vector $\boldsymbol{n}$ also lies along the rotation axis. Also shown is a typical patch on the surface and its vector displacement $\boldsymbol{r}$ from the origin.

Note that we have used λ, rather than ν, for our independent variable, as the former is more appropriate for optical stellar observations. The maximum possible wavelength shift occurs when $x = R_*$:

$$\Delta\lambda_{\rm max} = \lambda_\circ \frac{V_{\rm eq}}{c} \sin i \ . \tag{E.33}$$

To derive the line broadening, we first let $I_\lambda^{\rm NR}(x,y)$ be the specific intensity emanating from any point (x,y) within the projected surface of a *nonrotating* star with identical radius. We may write this intensity as

$$I_\lambda^{\rm NR}(x,y) = I_c \, \Delta\lambda_\circ \, \phi(\lambda - \lambda_\circ) \ . \tag{E.34}$$

Here, I_c is the continuum intensity, *i. e.*, that observed at wavelengths just outside the line of interest. We assume for simplicity that I_c does not vary over the stellar disk. The quantity $\Delta\lambda_\circ$ characterizes the intrinsic width of the line and depends on the profile $\phi(\lambda - \lambda_\circ)$. The latter is determined by the various physical effects already covered. Under the influence of rotation, the intensity becomes $I_\lambda(x,y)$. This function is identical to $I_\lambda^{\rm NR}(x,y)$, but with $\lambda_\circ$ replaced by $\lambda_\circ'$. Then from equation (E.32), we have

$$I_\lambda(x,y) = I_c \, \Delta\lambda_\circ \, \phi\left(\lambda - \lambda_\circ - \lambda_\circ \frac{x}{R_*} \frac{V_{\rm eq}}{c} \sin i\right) \ . \tag{E.35}$$

Each point on the disk contributes an intensity of the above form. The effective, smeared profile of the *total* received intensity is thus an average over the disk:

$$
\begin{aligned}
\bar{\phi}(\lambda - \lambda_\circ) &= \frac{\int\!\int dx\, dy\, I_\lambda(x,y)}{I_c\, \Delta\lambda_\circ\, \pi\, R_*^2} \\
&= \frac{1}{\pi\, R_*^2} \int_{-R_*}^{+R_*} dx \int_{-\sqrt{R_*^2 - x^2}}^{+\sqrt{R_*^2 - x^2}} dy\ \ \phi\!\left(\lambda - \lambda_\circ - \lambda_\circ \frac{x}{R_*} \frac{V_{\rm eq}}{c} \sin i\right) .
\end{aligned}
\tag{E.36}
$$

The y-integration is trivial, and we find

$$
\bar{\phi}(\lambda - \lambda_\circ) = \frac{2}{\pi\, R_*^2} \int_{-R_*}^{+R_*} dx \sqrt{R_*^2 - x^2}\ \ \phi\!\left(\lambda - \lambda_\circ - \lambda_\circ \frac{x}{R_*} \frac{V_{\rm eq}}{c} \sin i\right) \tag{E.37a}
$$

$$
= \frac{2}{\pi} \int_{-1}^{+1} d\xi\, \sqrt{1 - \xi^2}\ \ \phi\left(\lambda - \lambda_\circ - \Delta\lambda_{\rm max}\, \xi\right) . \tag{E.37b}
$$

To obtain the last expression, we have introduced the nondimensional variable $\xi \equiv x/R_*$, and have further utilized equation (E.33) for $\Delta\lambda_{\rm max}$.

It is instructive to recast equation (E.37b) as

$$
\bar{\phi}(\lambda - \lambda_\circ) = \int_{-1}^{+1} d\xi\ \mathcal{B}(\xi)\ \phi\left(\lambda\, -\, \lambda_\circ\, -\, \Delta\lambda_{\rm max}\, \xi\right) . \tag{E.38}
$$

That is, the new profile is the convolution of the original one and the broadening function:

$$
\mathcal{B}(\xi) = \frac{2}{\pi} \sqrt{1 - \xi^2} . \tag{E.39}
$$

The reader may verify that the broadening function is normalized, in the sense that $\int_{-1}^{+1} d\xi\, \mathcal{B}(\xi) = 1$.

Figure E.3 shows the rotational distortion of a purely thermal profile. Here, we have let $\Delta\lambda_{\rm max}$ be 2.0 and 0.5 times $\Delta\lambda_D \equiv (c/\nu_\circ^2)\Delta\nu_D$. To better understand the plotted results, consider that for a very rapidly rotating star, *i. e.*, for $\Delta\lambda_{\rm max} \gg \Delta\lambda_\circ$, the broadening function is much wider than the intrinsic profile. Approximating the latter as a delta function, we find that

$$
\bar{\phi}(\lambda - \lambda_\circ) \approx \mathcal{B}\left(\frac{\lambda - \lambda_\circ}{\Delta\lambda_{\rm max}}\right) . \qquad \textit{rapid rotation} \tag{E.40}
$$

Conversely, a slowly rotating star has such small $\Delta\lambda_{\rm max}$ that we may neglect this term in equation (E.38). We then recover the intrinsic profile:

$$
\bar{\phi}(\lambda - \lambda_\circ) \approx \phi(\lambda - \lambda_\circ) . \qquad \textit{slow rotation} \tag{E.41}
$$

A more detailed analysis of rotational broadening accounts for the spatial variation of the continuum intensity I_c across the face of the stellar disk. The main effect is *limb darkening*, *i. e.*, the fact that radiation originating near the edge has diminished intensity because it samples gas of lower temperature. This complication alters $\mathcal{B}(\xi)$ from our simple elliptical form.

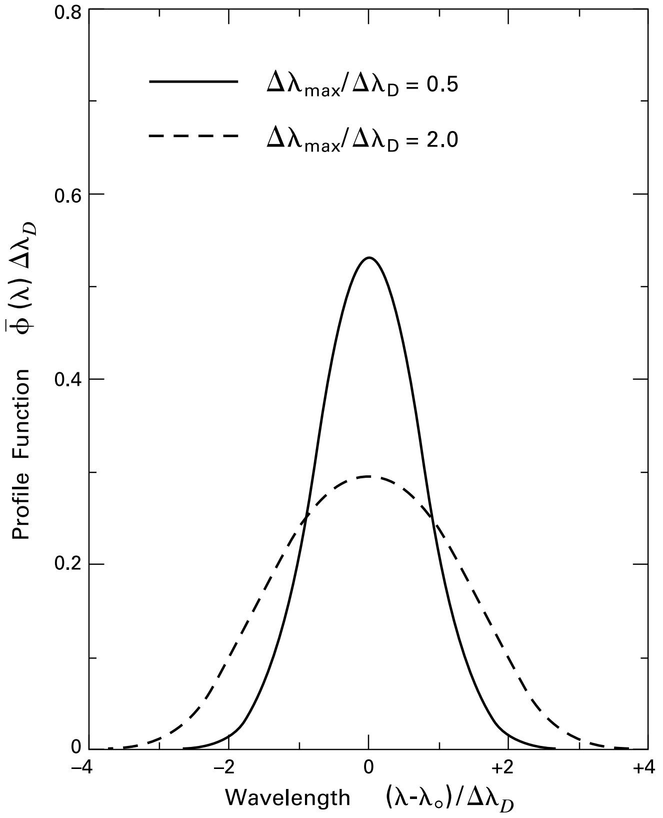


Figure E.3 Rotational broadening of a spectral line. The emission line from the nonrotating star has a thermal profile. Dashed and solid curves show the result of applying the simplified broadening function of equation (E.39) for the indicated ratios of $\Delta\lambda_{\rm max}$ to $\Delta\lambda_D$.

E.4 Two Examples: H_2 and CO in Clouds

What is the relative importance of the various broadening mechanisms we have discussed? Consider, for example, the passage of radiation through molecular clouds. What is the expected absorption line profile of molecular hydrogen for external stellar photons? We recall that H_2 is dissociated near cloud surfaces through electronic excitation followed by radiative decay (§ 8.1). Although many levels contribute to this process, the ultraviolet photon causing the initial excitation has the typical frequency $\nu_\circ = 3 \times 10^{15}$ s^{-1}. The core of the absorption line is broadened mainly by turbulent motions, where $\Delta V_{\rm FWHM}({\rm turb}) = 1.67$ km s^{-1} for a one-dimensional velocity dispersion of $\Delta V = 1$ km s^{-1}. In contrast, at the temperature of 37 K for the generic cloud boundary in § 8.1, the thermal width is only $\Delta V_{\rm FWHM}({\rm therm}) = 0.92$ km s^{-1}. From equation (E.29), the full line width is therefore $\Delta V_{\rm FWHM}({\rm tot}) = 1.91$ km s^{-1}, corresponding to a width in frequency of 8×10^9 s^{-1}. This width exceeds the typical natural broadening for such lines of $\gamma \approx 2 \times 10^9$ s^{-1}. Thus, the core indeed follows a (nonthermal) Doppler profile. The line wings, however, are naturally broadened, a fact we utilized in the detailed analysis of § 8.1. To see this, note from equation (B.4) that the rate of collisional excitation is multiplied by a thermal Boltzmann factor which is very small when, as here, $\Delta E \gg k_B T$. Thus, we have $\gamma_{\rm coll} \ll \gamma$ in this particular case.

Another important example in molecular clouds is the $J = 1 \rightarrow 0$ emission line of ^{12}C^{16}O. Once again, the low gas temperature implies that thermal broadening is generally small compared to that caused by turbulent motion. A significant exception is dense cores, where the inequality is reversed (see § 3.3). In any case, the line core again has the form of equation (E.25). The line wings have a Lorentzian shape, but the natural and collisional widths are now competitive. To see their relation, note that $\gamma_{\rm coll}$ is equal to $\gamma_{10} n_{\rm tot}$, in the notation of Appendix B. Thus, $\gamma_{\rm coll}$ equals the natural γ when $n_{\rm tot}$ equals the critical density $n_{\rm crit}$ introduced in § 5.3. For $n_{\rm tot}$ less than about 3×10^3 cm^{-3}, natural broadening dominates the line wings, while collisions are paramount at higher densities.

F Shock Jump Conditions

Consider a fluid with density ρ_1 and pressure P_1, moving with velocity u_1 in the x-direction (see Figure F.1). The flow occurs within a uniform, frictionless tube of rectangular cross section, whose sides lie in the x-y and y-z planes. At position $x = s(t)$, the flow encounters a shock front, which itself travels at velocity $u_0 = ds/dt$. Downstream from the shock, the fluid has a different density, pressure, and velocity, each distinguished here by the 2–subscript. At any instant, we consider two imaginary surfaces in the y-z plane, before and after the shock. These surfaces, located at $x_1(t)$ and $x_2(t)$, respectively, are comoving with the fluid, so that $dx_1/dt = u_1$ and $dx_2/dt = u_2$.

To derive the jump conditions, we first note that no mass leaves the volume enclosed by the comoving surfaces. Thus we have

$$\frac{d}{dt} \int_{x_1(t)}^{x_2(t)} \rho \, dx \, = \, 0 \ , \tag{F.1}$$

where we have divided through by the tube's cross-sectional area. The upstream pressure P_1 acting from the outside on the left surface increases the linear momentum of the fluid, while the momentum is lowered

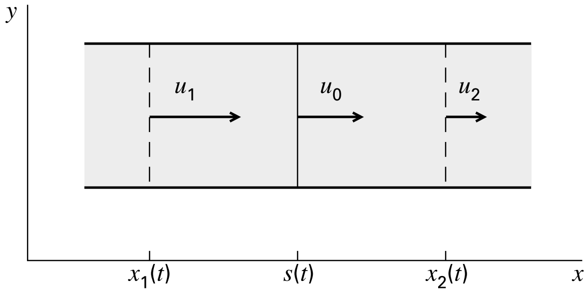


Figure F.1 Portion of a tube containing a shocked fluid. The tube has a rectangular cross section and extends in the z-direction, *i. e.*, out of the plane of the page. The two dashed lines represent surfaces that move with the fluid at the indicated speeds. The interior, solid line is the shock front.

by the pressure P_2 which the right surface exerts on the external medium. Hence

$$\frac{d}{dt}\int_{x_1(t)}^{x_2(t)} \rho\, u\, dx \;=\; P_1 \;-\; P_2 \;, \tag{F.2}$$

Finally, we note that the net mechanical power input per unit area, $P_1 u_1 - P_2 u_2$, serves to increase the energy of the fluid between the surfaces:

$$\frac{d}{dt}\int_{x_1(t)}^{x_2(t)} \rho\, \left(u^2/2 \,+\, \epsilon\right)\, dx \;=\; P_1\, u_1 \;-\; P_2\, u_2 \;, \tag{F.3}$$

where ϵ is the internal energy content per unit mass. Here we have omitted the gravitational potential energy, which does not change appreciably across the shock. Equation (F.3) is invalid if the fluid can lose energy through radiation; we consider this case below.

Each of the left-hand sides in equations (F.1)–(F.3) is of the form

$$\frac{dJ}{dt} \;\equiv\; \frac{d}{dt}\int_{x_1(t)}^{x_2(t)} \Psi(x,t)\, dx \;,$$

where $\Psi(x,t)$ is spatially discontinuous at $x \,=\, s(t)$. If we split the integral into up– and downstream contributions before differentiating, we obtain

$$\begin{aligned}\frac{dJ}{dt} \;&=\; \frac{d}{dt}\int_{x_1(t)}^{s(t)} \Psi\, dx \,+\, \frac{d}{dt}\int_{s(t)}^{x_2(t)} \Psi\, dx \\ &=\; \int_{x_1(t)}^{x_2(t)} \frac{\partial \Psi}{\partial t}\, dx \,+\, \Psi_1\, u_0 \,-\, \Psi_1\, u_1 \,+\, \Psi_2\, u_2 \,-\, \Psi_2\, u_0 \;,\end{aligned}$$

where Ψ_1 and Ψ_2 are, respectively, the upstream and downstream values of the generic integrand. Since the positions of the comoving surfaces are arbitrary, we consider the limit as both $x_1(t)$ and $x_2(t)$ approach $s(t)$. We are motivated to take this limit because the actual thickness of the shock front is almost always smaller than the dimensions of interest. The term $\partial\Psi/\partial t$ is finite, so the first righthand term above vanishes. We conclude that

$$\lim_{x_1, x_2 \rightarrow s} \frac{dJ}{dt} \;=\; \Psi_2\, v_2 \;-\; \Psi_1\, v_1 \;, \tag{F.4}$$

where $v_1 \equiv u_1 - u_0$ and $v_2 \equiv u_2 - u_0$ are, respectively, the upstream and downstream velocities *relative to the shock front*. In § 8.4, we referred to v_1 as the "shock speed", $V_{\rm shock}$.

Applying equation (F.4) to (F.1), we obtain

$$\rho_2\, v_2 \;=\; \rho_1\, v_1 \;. \tag{F.5}$$

Turning to equation (F.2), we find, after using (F.5), that

$$P_2 \,+\, \rho_2\, v_2^2 \;=\; P_1 \,+ \rho_1\, v_1^2 \;. \tag{F.6}$$

Finally, equation (F.3) becomes

$$\frac{1}{2}\, v_2^2 \,+\, \epsilon_2 \,+\, \frac{P_2}{\rho_2} \;=\; \frac{1}{2}\, v_1^2 \,+\, \epsilon_1 \,+\, \frac{P_1}{\rho_1} \;. \tag{F.7}$$

Equations (F.5)–(F.7) constitute the Rankine-Hugoniot jump conditions for an "adiabatic" (*i. e.*, non-radiating) shock. In general, the complete solution of a flow problem with a shock requires knowledge of seven quantities – the velocity plus two thermodynamic variables (*e. g.*, P and ρ) both upstream and downstream, together with the velocity of the shock itself. If the 3 variables on one side of the shock are specified, application of the 3 jump conditions still leaves one more unknown quantity. For example, in the problem of spherical protostellar accretion shocks (Chapter 11), we specify the upstream state of the gas and take the velocity of the shock front (*i. e.*, the protostellar surface) to be essentially zero.

What is the temperature increase across a shock? If the fluid of interest is a perfect gas, then the internal energy ϵ is given by $1/(\gamma - 1)P/\rho$, where γ is the ratio of specific heats at constant pressure and volume, respectively. Substituting this relation into (F.7) leads to

$$\frac{1}{2}v_2^2 - \frac{1}{2}v_1^2 = \frac{\gamma}{\gamma - 1}\left(\frac{P_1}{\rho_1} - \frac{P_2}{\rho_2}\right) . \tag{F.8}$$

Equations (F.5) and (F.6) together imply that

$$v_2 - v_1 = \frac{P_1 - P_2}{\rho_1 v_1} . \tag{F.9}$$

Multiplying equation (F.9) by $1/2(v_2 + v_1)$ and combining the result with (F.5) and (F.8), we obtain

$$\frac{P_2 - P_1}{2}\left(\frac{1}{\rho_1} + \frac{1}{\rho_2}\right) = \frac{\gamma}{\gamma - 1}\left(\frac{P_2}{\rho_2} - \frac{P_1}{\rho_1}\right) ,$$

from which we can solve for ρ_2/ρ_1 in terms of P_2/P_1:

$$\begin{aligned}\frac{\rho_2}{\rho_1} &= \frac{(\gamma+1)P_2/P_1 + (\gamma-1)}{(\gamma-1)P_2/P_1 + (\gamma+1)} \\ &= \frac{v_1}{v_2} ,\end{aligned} \tag{F.10}$$

where the last equality follows from equation (F.5). For an ideal gas, T_2/T_1 is equal to $P_2\rho_1/P_1\rho_2$, so that

$$\frac{T_2}{T_1} = \frac{(\gamma-1)(P_2/P_1)^2 + (\gamma+1)P_2/P_1}{(\gamma+1)P_2/P_1 + (\gamma+1)} . \tag{F.11}$$

For very weak shocks ($P_2 \approx P_1$), T_2/T_1 is close to unity, but equation (F.11) shows that this ratio increases without bound as P_2/P_1 becomes very large. On the other hand, equation (F.10) implies that, in the same limit, ρ_2/ρ_1 approaches $(\gamma + 1)/(\gamma - 1)$, which is equal to 4 in the case of a monoatomic, ideal gas ($\gamma = 5/3$).

In equations (F.10) and (F.11), we have used the ratio P_2/P_1 as our measure of shock strength. Alternatively, we could employ the *upstream Mach number* $M_1 \equiv v_1/a_1$, where a_1 is the adiabatic sound speed ahead of the shock. For a perfect gas, we have $a_1 = (\gamma P_1/\rho_1)^{1/2}$. To relate M_1 to P_2/P_1, we first divide (F.7) by v_1, and then use (F.10) to express v_2/v_1 in terms of P_2/P_1. The result can be expressed as

$$\rho_1 v_1^2 = \frac{P_1}{2}\left[(\gamma+1)P_2/P_1 + (\gamma-1)\right] . \tag{F.12}$$

But we also know that

$$\rho_1 a^2 = \gamma P_1 . \tag{F.13}$$

Dividing (F.12) by (F.13) yields the desired relation:

$$M_1^2 = \frac{1}{2\gamma}\left[(\gamma+1)P_2/P_1 + (\gamma-1)\right] . \tag{F.14}$$

We can now use equation (F.14) to recast (F.10) and (F.11) in terms of M_1:

$$\frac{\rho_2}{\rho_1} = \frac{v_1}{v_2} = \frac{(\gamma+1)\,M_1^2}{(\gamma-1)\,M_1^2 + 2} \tag{F.15}$$

and

$$\frac{T_2}{T_1} = \frac{[2\,\gamma\,M_1^2 - (\gamma-1)][(\gamma-1)\,M_1^2 + 2]}{(\gamma+1)^2\,M_1^2} \; . \tag{F.16}$$

If the postshock temperature is sufficiently high, the fluid can lose energy through radiation. As discussed in § 8.4, this loss occurs over a relaxation region much broader than the particle mean free path characterizing the shock front itself. Nevertheless, it is sometimes appropriate, as in the protostellar case, to view the entire region as a single, radiating shock front. The jump conditions (F.5) and (F.6) are still valid across this broadened front, but the energy relation (F.7) is not. To see the required modification, we return to equation (F.3), which now becomes

$$\frac{d}{dt}\int_{x_1(t)}^{x_3(t)} \rho\,(u^2/2 + \epsilon)\,dx = P_1\,u_1 - P_3\,u_3 - 2\,F_{\rm rad} \; ,$$

where $F_{\rm rad}$ is the flux radiated in either the $+$ or $-x$-direction, and where the postshock point $x_3(t)$ lies beyond the relaxation region (see Figure 8.9). Applying the same limiting procedure as before, we find that

$$\frac{1}{2}\,v_3^2 + \epsilon_3 + \frac{P_3}{\rho_3} = \frac{1}{2}\,v_1^2 + \epsilon_1 + \frac{P_1}{\rho_1} - \frac{2\,F_{\rm rad}}{\rho_1\,v_1} \; . \tag{F.17}$$

Note that the compression ratio ρ_3/ρ_1 can now be much higher than ρ_2/ρ_1 for the adiabatic shock. Suppose, for example, that the postshock gas were to cool to the preshock temperature T_1. Then the relation $a_3 = a_1$, coupled with (F.5) and (F.6), implies that $\rho_3/\rho_1 = \mathcal{M}_1^2$. This last term, which can be arbitrarily high, is the square of the *isothermal Mach number*, *i. e.*, the ratio of v_1 to the isothermal sound speed $a_T \equiv (P/\rho)^{1/2}$.

G Radiative Diffusion and Stellar Opacity

In a very optically thick medium, such as the interior of a star, the specific intensity I_ν is very nearly the Planckian function $B_\nu(T)$. The temperature T is also the same as the kinetic gas value. To see why, we first recall from Appendix C that I_ν for any spectral line approaches $B_\nu(T_{\rm ex})$ in this limit, where $T_{\rm ex}$ is the excitation temperature of the relevant transition. We further note, from Appendix B, that $T_{\rm ex}$ becomes $T_{\rm kin}$ for densities $n_{\rm tot} \gg n_{\rm crit}$. Both these results are restatements of the fact that LTE prevails under these conditions.

On the other hand, I_ν cannot be *precisely* $B_\nu(T)$. Since the latter is isotropic, the monochromatic flux F_ν, given by the integral in equation (2.17), would vanish identically. Suppose the net flux is locally in the z-direction, and consider the variation of $I_\nu(z,\theta)$ with respect to θ, the angle between the photon propagation direction $\hat{\boldsymbol{n}}$ and $\hat{\boldsymbol{z}}$ (see Figure G.1). If $\mu = \cos\,\theta$, then the oblique path length $\Delta s = \Delta z/\mu$. Thus, we may rewrite the radiative transfer equation (2.20) as

$$\mu\,\frac{\partial\,I_\nu(z,\mu)}{\partial\,z} = -\alpha_\nu\,I_\nu + j_\nu \; . \tag{G.1}$$

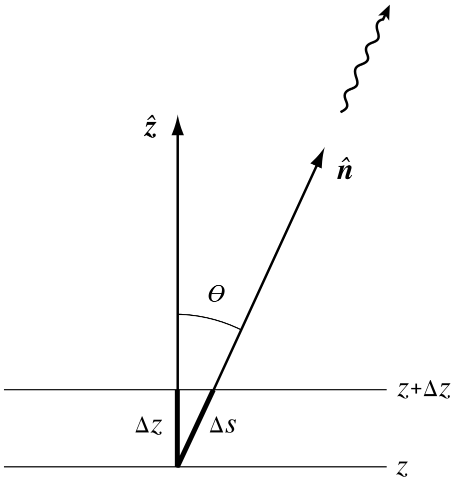


Figure G.1 Variation of radiative intensity with angle. Although the net flux is in the $\hat{z}$-direction, we consider the specific intensity in the $\hat{n}$-direction, inclined at the angle θ.

We assume the emission to be thermal, so that j_ν is given by Kirkhoff's Law, equation (E.20). After replacing α_ν by $\rho\kappa_\nu$, we may rewrite equation (G.1) as

$$I_\nu(z,\mu) \;=\; B_\nu \;-\; \frac{\mu}{\rho\,\kappa_\nu}\,\frac{\partial\, I_\nu}{\partial\, z}\;. \tag{G.2}$$

In pure LTE, I_ν would have no spatial variation; equation (G.2) then confirms that $I_\nu = B_\nu$ exactly. For small departures from this state, we may substitute B_ν for I_ν in the derivative. Noting that B_ν is a function of T alone, we obtain the desired angular dependence:

$$I_\nu(z,\mu) \;\approx\; B_\nu(T) \;-\; \frac{\mu}{\rho\,\kappa_\nu}\,\frac{dB_\nu}{dT}\,\frac{dT}{dz}\;. \tag{G.3}$$

When we use *this* form of I_ν in equation (2.17) for F_ν, the isotropic part still vanishes. However, the second part survives integration to yield

$$\begin{aligned} F_\nu(z) \;&=\; \int I_\nu\,\mu\,d\Omega \\ &=\; 2\,\pi\int_{-1}^{+1} d\mu\,\mu\,I_\nu \\ &=\; -\frac{4\,\pi}{3\,\rho\,\kappa}\,\frac{dB_\nu}{dT}\,\frac{dT}{dz}\;. \end{aligned} \tag{G.4}$$

We are often more interested in the frequency-integrated quantity $F_{\rm bol}$, which is also denoted $F_{\rm rad}$ in this context. We have

$$\begin{aligned} F_{\rm rad} \;&=\; \int_0^\infty F_\nu\,d\nu \\ &-\; \frac{4\,\pi}{3\,\rho}\,\frac{dT}{dz}\int_0^\infty \frac{1}{\kappa_\nu}\,\frac{dB_\nu}{dT}\,d\nu\;. \end{aligned} \tag{G.5}$$

If we define the *Rosseland mean opacity* κ through the relation

$$\frac{1}{\kappa} \;\equiv\; \frac{\int_0^\infty \kappa_\nu^{-1}\,dB_\nu/dT\,d\nu}{\int_0^\infty dB_\nu/dT\,d\nu}\;, \tag{G.6}$$

then (G.5) becomes the *radiative diffusion equation*:

$$F_{\rm rad} \;=\; -\frac{16\,\sigma_B\,T^3}{3\,\rho\,\kappa}\,\frac{dT}{dz} \;. \tag{G.7}$$

Here, we have utilized

$$\begin{aligned}\int_0^\infty \frac{dB_\nu}{dT}\,d\nu \;&=\; \frac{d}{dT}\int_0^\infty B_\nu\,d\nu \\ &=\; \frac{c}{4\,\pi}\,\frac{d}{dT}\int_0^\infty u_\nu\,d\nu \;,\end{aligned}$$

where u_ν is the energy density of the blackbody radiation field (recall equation (2.18)). Since $\int_0^\infty u_\nu d\nu = 4\sigma_B T^4/c$, we obtain the desired result.

Although our derivation strictly applies within the interior of a star, equation (G.7) gives a useful expression for the temperature variation near the stellar surface. In this case, $F_{\rm rad}$ may be considered a spatial constant. Integrating from the tenuous, outer regime inward, we find

$$\sigma_B\,T^4 \;=\; \frac{3\,F_{\rm rad}}{4}\,\tau \,+\, \sigma_B\,T_\circ^4 \;. \tag{G.8}$$

Here, the optical depth τ increases deeper into the atmosphere:

$$\tau \;\equiv\; \int_z^\infty \rho\,\kappa\,dz \;.$$

The boundary temperature $T_\circ$ is the value at $\tau = 0$. To derive an expression for $T_\circ$, we adopt the *Eddington approximation*. We assume that I_ν at the top of the atmosphere is isotropic in the outward direction, *i. e.*, for $0 \leq \mu \leq 1$, and is zero otherwise. From equations (2.17) and (2.18), we find that

$$F_\nu \;=\; \frac{c}{2}\,u_\nu$$

at that location. Integration over all frequencies then yields

$$F_{\rm rad} \;=\; 2\,\sigma_B\,T_\circ^4 \;. \tag{G.9}$$

Finally, substitution of equation (G.9) into (G.8) gives the desired run of temperature:

$$\sigma_B\,T^4 \;=\; \frac{3\,F_{\rm rad}}{4}\left(\tau \,+\, \frac{2}{3}\right) \;. \tag{G.10}$$

The stellar photosphere may be defined as that location where the local gas temperature T equals $T_{\rm eff}$. The latter is related to the flux through equation (2.35), which we write as

$$F_{\rm rad} \;=\; \sigma_B\,T_{\rm eff}^4 \;.$$

Equation (G.10) then implies that the photosphere has an optical depth of 2/3.

Using the radiative diffusion equation requires knowledge of κ, which is a function of ρ and T for a gas of fixed chemical composition. It is sometimes useful to approximate the frequency dependence of the monochromatic opacity by a power law, over some specified interval:

$$\begin{aligned}\kappa_\nu \;&\approx\; \kappa_\circ\left(\frac{\nu}{\nu_\circ}\right)^n \\ &=\; \kappa_\circ\left(\frac{k_B\,T}{h\,\nu_\circ}\right)^n x^n \;,\end{aligned} \tag{G.11}$$

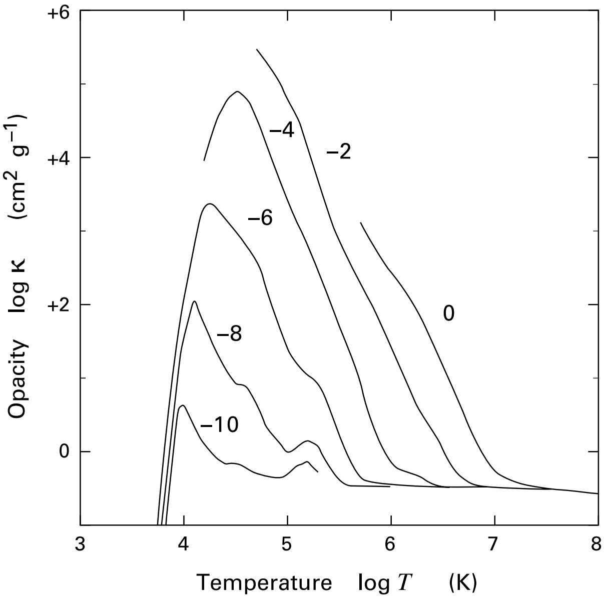


Figure G.2 The Rosseland mean opacity for a gas of solar composition, shown as a function of temperature. The numbers labeling each curve are values of mass density, displayed as $\log\rho$.

where $x \equiv h\nu/k_B T$. Suppose we let $\partial B_\nu/\partial T\, d\nu = A\, f(x)\, dx$, where A is the appropriate dimensional factor. Substituting this form into the definition of equation (G.6), we find

$$\frac{1}{\kappa} = \frac{1}{\kappa_\circ}\left(\frac{h\,\nu_\circ}{k_B\,T}\right)^n \frac{\int x^{-n}\, f(x)\, dx}{\int f(x)\, dx}\ . \tag{G.12}$$

The quotient of integrals, both of whose limits we have left unspecified, is a pure number. We see that κ is proportional to T^n in the frequency regime of interest.

For more precise numerical work, the Rosseland mean opacity must be obtained numerically and has been tabulated by many researchers over the years. Generally speaking, $\kappa(\rho, T)$ is well established at the higher stellar temperatures, while its accurate determination at photospheric conditions in the lowest-mass stars is a continuing enterprise of considerable importance. Figure G.2 displays modern results for κ as a function of temperature, for a range of densities. It is clear that the curves share a number of generic features, which reflect the salient physical processes underlying the extinction.

The peak in opacity somewhat above 10^4 K is principally due to the photoionization of hydrogen. This is a *bound-free* transition, in that the electron goes from the lowest, $n = 1$ orbital into the continuum. At lower temperatures, the opacity falls steeply because too few photons in the Planckian distribution have energies exceeding 13.6 eV, or even the 10.2 eV required to promote the hydrogenic electron from $n = 1$ to $n = 2$. The latter is an example of a *bound-bound* transition.

For temperatures as low as 5000 K, most free electrons in the gas come from more easily ionized metals. A small fraction of atomic hydrogen captures extra electrons to become H$^-$ ions. With its ionization potential of 0.75 eV, this species can be more readily photoionized by the ambient radiation field and provides the chief opacity source in the outer layers of late-type dwarfs and pre-main-sequence stars. The decline in opacity with falling temperature halts temporarily near 10^3 K, where the dominant process is now bound-bound transitions in molecules. The complex behavior of $\kappa(\rho, T)$ in this regime is not included in Figure G.2. Nor is the absorption and scattering by dust grains, which dominates for $T \lesssim 10^3$ K.

At temperatures significantly above 10^4 K, the opacity declines and varies approximately as $\rho\, T^{-7/2}$, the so-called *Kramers law*. Initially, the main extinction mechanism is still bound-free absorption, not only of hydrogen and helium, but also of metals such as carbon, nitrogen, and oxygen. Photons of higher energy transfer momentum with less efficiency to the bound electrons. The monochromatic opacity thus declines with frequency, and the Rosseland mean has its characteristic falloff with temperature. The Kramers-law behavior applies also at the higher densities and temperatures where *free-free* transitions prevail. Here, photons are absorbed by free electrons, whose recoil is taken up by nearby metallic ions. For the two lowest densities shown in Figure G.2, the opacity exhibits a localized peak near 10^5 K, the *Z-bump*. This feature stems from bound-bound transitions involving the inner electrons of heavier metals, principally iron.

The opacity at the highest temperatures is due to *electron scattering*. A free electron in the now fully ionized plasma absorbs a photon and reemits it without change of frequency. For photon energies less than $m_e c^2$, the electron rest energy, one may employ classical theory to analyze the process. Here, one pictures scattering as dipole radiation emanating from an electron oscillating under the influence of a plane, electromagnetic wave. The total power radiated divided by the incident flux is the *Thomson cross section* for a single electron, $\sigma_T = (8\pi/3) r_\circ^2 = 6.65 \times 10^{-25}\,\mathrm{cm}^{-2}$. The quantity $r_\circ \equiv e^2/m_e c^2 = 2.82 \times 10^{-13}\,\mathrm{cm}$ is the *electron classical radius*, obtained by equating its rest mass to the electrostatic potential energy. Note that the expression for σ_T does not involve the wave frequency. We obtain the opacity through multiplying σ_T by the number of free electrons per gram, a figure which is independent of density and temperature once the gas is fully ionized. The Rosseland mean opacity thus attains the constant value of 0.3 cm^2 gm^{-1} displayed in Figure G.2.

H Derivation of Binary Star Relations

To obtain equation (12.17), we examine more closely the orbit of either star within the binary system depicted in Figure 12.6a. Figure H.1 displays the ellipse, whose focus (again shown as a cross) lies at the binary's center of mass. The star attains its greatest speed $V_{\max}$ at the smallest distance, $r_{\min}$, from the focus. Conversely, the lowest speed $V_{\min}$ occurs at the largest radius $r_{\max}$. These two distances are related to the semi-major axis a by

$$r_{\min} = a\,(1 - e) \tag{H.1a}$$

$$r_{\max} = a\,(1 + e)\ , \tag{H.1b}$$

where e is the eccentricity.

We now invoke conservation of angular momentum for the entire system. The velocity and radius vectors are perpendicular at the extreme points. Using subscripts to distinguish the primary and secondary stars, we have

$$\begin{aligned} &M_1\,(V_{\max})_1\,(r_{\min})_1 + M_2\,(V_{\max})_2\,(r_{\min})_2 = \\ &M_1\,(V_{\min})_1\,(r_{\max})_1 + M_2\,(V_{\min})_2\,(r_{\max})_2\ . \end{aligned} \tag{H.2}$$

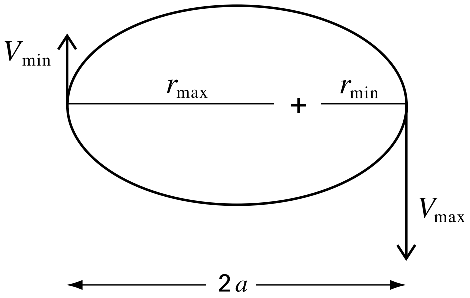


Figure H.1 The elliptical orbit of a star within a binary system. The semi-major axis of the ellipse has length a, and the focus is at the cross. When the star reaches the minimum and maximum radial distances, it attains its greatest and least speeds, respectively.

Recall, from equation (12.14), that the radii and velocities in the two orbits scale inversely with the respective masses. Thus, we also have

$$\frac{(r_{\rm min})_2}{(r_{\rm min})_1} \,=\, \frac{(r_{\rm max})_2}{(r_{\rm max})_1} \,=\, \frac{(V_{\rm min})_2}{(V_{\rm min})_1} \,=\, \frac{(V_{\rm max})_2}{(V_{\rm max})_1} \,=\, \frac{M_1}{M_2}\ . \tag{H.3}$$

Applying these relations to equation (H.2) simplifies it to

$$(V_{\rm max})_1\,(r_{\rm min})_1 \,=\, (V_{\rm min})_1\,(r_{\rm max})_1\ . \tag{H.4}$$

In a similar fashion, energy conservation for the binary reads

$$\begin{aligned} -\frac{G\,M_1\,M_2}{(r_{\rm min})_1 + (r_{\rm min})_2} \,+\, \frac{1}{2}\,M_1\,(V_{\rm max})_1^2 \,+\, \frac{1}{2}\,M_2\,(V_{\rm max})_2^2 \,=\\ -\frac{G\,M_1\,M_2}{(r_{\rm max})_1 + (r_{\rm max})_2} \,+\, \frac{1}{2}\,M_1\,(V_{\rm min})_1^2 \,+\, \frac{1}{2}\,M_2\,(V_{\rm min})_2^2\ . \end{aligned} \tag{H.5}$$

We again use equation (H.3) to eliminate all radii and speeds of the secondary. We thus find

$$G\,M_2\left[\frac{1}{(r_{\rm min})_1} \,-\, \frac{1}{(r_{\rm max})_1}\right] \,=\, \frac{1}{2}\left(1 \,+\, \frac{M_1}{M_2}\right)^2 \left[(V_{\rm max})_1^2 \,-\, (V_{\rm min})_1^2\right]\ .$$

Combining this equation with (H.4) yields

$$G\,M_2 \,=\, \frac{1}{2}\left(1 \,+\, \frac{M_1}{M_2}\right)^2 (V_{\rm max})_1^2\,\left[(r_{\rm max})_1 \,+\, (r_{\rm min})_1\right]\left(\frac{r_{\rm min}}{r_{\rm max}}\right)_1\ .$$

We then use equation (H.1) to obtain

$$\frac{G\,M_{\rm tot}}{a_1} \,=\, \left(1 \,+\, \frac{M_1}{M_2}\right)^3 \left(\frac{1 \,-\, e}{1 \,+\, e}\right)\,(V_{\rm max})_1^2\ . \tag{H.6}$$

Here, $M_{\rm tot} \,=\, M_1 \,+\, M_2$. Finally, we utilize Kepler's third law, equation (12.15). The combination of (12.15) and (H.6) gives an expression for $(V_{\rm max})_1$:

$$(V_{\rm max})_1 \,=\, \frac{2\,\pi\,a_1}{P}\left(\frac{1 \,+\, e}{1 \,-\, e}\right)^{1/2}\ . \tag{H.7}$$

Consider now the radial (line-of-sight) velocity of the primary star. This quantity attains the maximum value $[V_{\rm cm} \,+\, (V_{\rm max})_1]\,\sin\,i$. Here, $V_{\rm cm}$ is the velocity of the binary's center of mass, and i is the inclination angle depicted in Figure 12.6b. Similarly, the minimum value is $[V_{\rm cm} \,-\, (V_{\rm min})_1]\,\sin\,i$. Let the amplitude K_1 be half the total range in radial velocity. Using equations (H.1) and (H.3), we find

$$\begin{aligned} K_1 \,&=\, \frac{1}{2}\,[(V_{\rm max})_1 \,+\, (V_{\rm min})_1]\,\sin\,i \\ &=\, \frac{(V_{\rm max})_1}{1\,+\,e}\,\sin\,i\ . \end{aligned}$$

We evaluate $(V_{\rm max})_1$ from equation (H.7) and obtain

$$K_1 \,=\, \frac{2\,\pi\,a_1}{P}\,(1\,-\,e^2)^{-1/2}\,\sin\,i\ , \tag{H.8}$$

which is equivalent to equation (12.17).

To derive equation (12.19) for the mass function $f(M)$, we first recast (H.8) into the form

$$K_1^3\,P^3 \,=\, (2\,\pi)^3\,a_1^3\,(1\,-\,e^2)^{-3/2}\,\sin^3\,i\ . \tag{H.9}$$

Squaring equation (12.15) gives an expression for P^2, which we combine with this last relation to yield

$$\begin{aligned} K_1^3\,P \,&=\, \frac{2\,\pi\,G\,(M_1\,+\,M_2)\,a_1^3}{(a_1\,+\,a_2)^3}\,(1\,-\,e^2)^{-3/2}\,\sin^3\,i \\ &=\, \frac{2\,\pi\,G\,M_2^3\,(1\,-\,e^2)^{-3/2}\,\sin^3\,i}{(M_1\,+\,M_2)^2}\ . \end{aligned}$$

Here, we have also used $a_1/a_2 \,=\, M_2/M_1$. We may rearrange the last equation as

$$\frac{M_2^3\,\sin^3\,i}{(M_1\,+\,M_2)^2} \,=\, \frac{1}{2\,\pi\,G}\,(1\,-\,e^2)^{3/2}\,K_1^3\,P\ , \tag{H.10}$$

which is also equation (12.19).

I Evaluation of a Polytropic Integral

In Chapter 16, we encounter the following nondimensional integral:

$$\mathcal{I} \,=\, \int_0^{\xi_\circ} \xi^2\,\theta^{5/2}\,d\xi\ . \tag{I.1}$$

Here, $\theta(\xi)$ is the Lame-Emden function of index $n = 3/2$. This function obeys equation (16.19):

$$\frac{1}{\xi^2}\,\frac{d}{d\xi}\left(\xi^2\,\frac{d\theta}{d\xi}\right) \,=\, -\theta^{3/2}\ , \tag{I.2}$$

subject to the initial conditions

$$\theta(0) \,=\, 1 \tag{I.3a}$$

$$\theta'(0) \,=\, 0\ . \tag{I.3b}$$

We have denoted by $\xi_\circ$ the point where $\theta(\xi)$ falls to zero.

To evaluate $\mathcal{I}$, we first use equation (I.2) to replace the integrand:

$$\mathcal{I} = -\int_0^{\xi_\circ} \theta \frac{\partial}{\partial \xi}\left(\xi^2 \frac{\partial \theta}{\partial \xi}\right) d\xi \ . \tag{I.4}$$

We then integrate by parts:

$$\mathcal{I} = -\int_0^{\xi_\circ} \frac{\partial}{\partial \xi}\left(\xi^2 \theta \frac{\partial \theta}{\partial \xi}\right) d\xi + \int_0^{\xi_\circ} \xi^2 \left(\frac{\partial \theta}{\partial \xi}\right)^2 d\xi \tag{I.5a}$$

$$= \int_0^{\xi_\circ} \left(\xi^2 \frac{\partial \theta}{\partial \xi}\right)^2 \xi^{-2} \, d\xi \ . \tag{I.5b}$$

Here, the first term in equation (I.5a) vanishes, since $\theta(\xi_\circ) \equiv 0$. We now integrate by parts again:

$$\mathcal{I} = -\int_0^{\xi_\circ} \frac{\partial}{\partial \xi}\left[\left(\xi^2 \frac{\partial \theta}{\partial \xi}\right)^2 \xi^{-1}\right] d\xi + \int_0^{\xi_\circ} 2\,\xi \frac{\partial \theta}{\partial \xi} \frac{\partial}{\partial \xi}\left(\xi^2 \frac{\partial \theta}{\partial \xi}\right) d\xi \tag{I.6a}$$

$$= -\xi_\circ^3 \left(\frac{\partial \theta}{\partial \xi}\right)_\circ^2 - 2\int_0^{\xi_\circ} \xi^3 \frac{\partial \theta}{\partial \xi} \theta^{3/2} \, d\xi \ , \tag{I.6b}$$

$$= -\xi_\circ^3 \left(\frac{\partial \theta}{\partial \xi}\right)_\circ^2 - \frac{4}{5}\int_0^{\xi_\circ} \xi^3 \frac{\partial}{\partial \xi} \theta^{5/2} \, d\xi \ , \tag{I.6c}$$

where we have again applied equation (I.2) to the second righthand term in equation (I.6a).

Consider finally the remaining integral within equation (I.6c). If we integrate it by parts, we find

$$\mathcal{I} = -\xi_\circ^3 \left(\frac{\partial \theta}{\partial \xi}\right)_\circ^2 - \frac{4}{5}\int_0^{\xi_\circ} \frac{\partial}{\partial \xi}\left(\xi^3 \theta^{5/2}\right) d\xi + \frac{12}{5}\int_0^{\xi_\circ} \xi^2 \theta^{5/2} \, d\xi \tag{I.7a}$$

$$= -\xi_\circ^3 \left(\frac{\partial \theta}{\partial \xi}\right)_\circ^2 + \frac{12}{5}\mathcal{I} \ , \tag{I.7b}$$

since the second term in equation (I.7a) vanishes. Solving equation (I.7b) for $\mathcal{I}$, we find

$$\mathcal{I} = \frac{5}{7}\xi_\circ^3 \left(\frac{\partial \theta}{\partial \xi}\right)_\circ^2 \ , \tag{I.8}$$

which is identical to equation (16.27) in the text.

The Formation of Stars
Steven W. Stahler and Francesco Palla

Sources for Tables, Figures, and Plates

The following abbreviations have been used for journals and reviews:

AA	Astronomy and Astrophysics
AAS	Astronomy and Astrophysics Supplements
AJ	Astronomical Journal
ApJ	Astrophysical Journal
ApJSS	Astrophysical Journal Supplement Series
ApSS	Astrophysics and Space Science
ARAA	Annual Reviews of Astronomy and Astrophysics
AREPS	Annual Reviews of Earth and Planetary Sciences
AstL	Astronomy Letters
Astron. Rep.	Astronomy Reports
BAN	Bulletin of the Astronomical Institute of the Netherlands
Can. J. Phys.	Canadian Journal of Physics
Fund. Cosm. Phys.	Fundamentals of Cosmic Physics
Mem. Soc. Astron. It.	Memorie della Società Astronomica Italiana
MNRAS	Monthly Notices of the Royal Astronomical Society
PASJ	Publications of the Astronomical Society of Japan
PASP	Publications of the Astronomical Society of the Pacific
Phys. Rev. Lett.	Physical Review Letters
Planet. Sp. Sci.	Planetary and Space Science
Prog. Theor. Phys.	Progress of Theoretical Physics
Rep. Prog. Phys.	Reports of Progress in Physics
Sp. Sci. Rev.	Space Science Reviews
ZAp	Zeitschrift für Astrophysik

Chapter 1

Table 1.1 - Data from § 9.6 and § 9.7 of Lang 1992, *Astrophysical Data: Planets and Stars*, (New York: Springer-Verlag), p. 132. For stars with $M_* \leq 0.8\,M_\odot$, data are from the theoretical calculations of Baraffe et al. 1997, AA, 337, 403.
Figure 1.2 - *left panel*: Combination of Figures 6 and 7 of Maddalena et al. 1986, ApJ, 303, 375. *right panel*: From Figure 2 of Lada 1992, ApJ, 393, L25.
Figure 1.3 - From Lang 1992, *Astrophysical Data: Planets and Stars*, (New York: Springer-Verlag), p. 365.
Figure 1.4 - Photograph courtesy of C. Beichman, Infrared Processing and Analysis Center.
Figure 1.5 - *left panel*: Hubble Space Telescope photograph by C. R. O'Dell, courtesy of Space Telescope Science Institute. *right panel*: Photograph courtesy of M. McCaughrean.

The Formation of Stars. Steven W. Stahler and Francesco Palla

ISBN: 3-527-40559-3

Figure 1.6 - Main sketch adapted from Figure 1 of Zuckerman 1973, ApJ, 183, 863. Insert from Figure 1 of Gordon & Churchwell 1970, AA, 9, 307.
Figure 1.7 - From Figure 1 of Cameron et al. 1992, in *Progress in Telescope and Instrumentation Technologies*, ed. M.-H. Ulrich (Munich: ESO Conf. and Workshop Proc.), p. 705.
Figure 1.8 - From Plate 5 of Barnard 1927, *A Photographic Atlas of Selected Regions of the Milky Way*, Carnegie Inst. of Washington Publ. 247, (Washington: Carnegie Inst.).
Figure 1.9 - From Figure 4 of Ungerechts & Thaddeus 1987, ApJS, 63, 645.
Figure 1.10 - *Upper panel*: From Plate L19 of Mizuno et al. 1995, ApJ, 445, L161. *Lower panel*: From Figure 5 of Abergel et al. 1994, ApJ, 423, L59.
Figure 1.11 - From Figure 1 of Kenyon et al. 1990, AJ, 99, 869.
Figure 1.12 - From Figure 14 of Benson & Myers 1989, ApJS, 71, 89.
Figure 1.13 - From Figure 1 of Pound & Bally 1991, ApJ, 383, 705.
Figure 1.14 - From Figure 1-1 of Binney & Tremaine 1989, *Galactic Dynamics*, (Princeton: Princeton U. Press), p. 10.
Figure 1.15 - From Figure 13 of Palla & Stahler 1993, ApJ, 418, 414; Figure 1c of Sweigart & Gross 1978, ApJS, 36, 405; Figure 2 of Schönberner 1983, ApJ, 272, 708.
Figure 1.18 - Based on a calculation by the authors.
Figure 1.19 - *Upper panel*: Photograph courtesy of M. Hauser, COBE Science Working Group. *Lower panel*: Lund Observatory.
Figure 1.20 - From Plate 1 of Elmegreen 1981, ApJS, 47, 229.
Figure 1.21 - From Figure 4 of Garcia-Burillo et al. 1993, AA, 274, 123.

Chapter 2

Table 2.1 - Data for the Solar System from Table 1 of Anders & Grevesse 1989, Geochim. et Cosmochim. Acta 53, 197; results for M42 are from Rubin et al. 1991, ApJ, 374, 564.
Table 2.2 - The total mass in H_2 is from Dame 1993, in *Back to the Galaxy*, ed. S. Holt and F. Verter (New York: AIP), p. 267. Dame also gives the total mass in the warm and cold neutral components, citing HI surveys. The *relative* masses of these two components are from E. Corbelli 2002 (personal communication) and Haynes & Broeils 1997, in *The Insterstellar Medium in Galaxies*, ed. J. M. Van der Hulst (Dordrecht: Kluwer), p. 75. Filling factors and other data are from Dopita & Sutherland 2002, *Astrophysics of the Diffuse Universe* (Berlin: Springer-Verlag).
Figure 2.2 - From Figure 13 of Radhakrishnan et al. 1972, ApJS, 24, 15.
Figure 2.3 - Combination of Figure 1 of Dame 1993, in *Back to the Galaxy*, ed. S. Holt and F. Verter, (New York: AIP), p. 267; Figure 5 of Lockman 1990, in *Radio Recombination Lines: Twenty-five Years of Investigation*, ed. M. A. Gordon and R. L. Sorochencko (Dordrecht: Kluwer), p. 225. The central density of H_2 is from Blitz et al. 1995, AA, 143, 267.
Figure 2.4 - From Figure 2.3 of Wilson & Pauls 1984, AA, 138, 225.
Figure 2.5 - From Figures 3(a) and 3(d) of Wolfire et al. 1995, ApJ, 443, 152.
Figure 2.7 - From Figure 3.3 of Whittet 1992, *Dust in the Galactic Environment* (Bristol: Institute of Physics Publishing).
Figure 2.14 - From Figure 5 of Sellgren 1984, ApJ, 277, 623.
Figure 2.15 - From Figure 5 of Puget & Leger 1989, ARAA, 27, 161.

Chapter 3

Figure 3.1 - From Figure 1 of Solomon & Rivolo 1987, in *The Galaxy*, ed. Gilmore and Carswell, (Dordrecht: Reidel), p. 105.

Figure 3.2 - From Figure 2(a) of Dame 1993, in *Back to the Galaxy*, ed. S. Holt & F. Verter, (New York: AIP), p. 272.
Figure 3.3 - Combination of Figure 2(a) of Blitz & Thaddeus 1980, ApJ, 241, 676 and Figure 28 of Raimond 1966, Bull. Astr. Inst. Neth., 18, 191.
Figure 3.4 - From Figure 2 of Blitz & Stark 1986, ApJ, 300, L89.
Figure 3.5 - From Figure 3 of Blitz & Stark 1986, ApJ, 300, L89.
Figure 3.6 - Constructed from Table 2 in Williams & Blitz 1994, ApJ, 428, 693.
Figure 3.7 - From Figure 3 of Knapp 1974, AJ, 79, 527.
Figure 3.8 - From Figure 47 of Leisawitz 1989, ApJS, 70, 731.
Figure 3.9 - From Figure 3 of Myers & Goodman 1988, ApJ, 329, 392.
Figure 3.10 - From Figure 9 of Solomon & Rivolo 1987, in *The Galaxy*, ed. Gilmore & Carswell, (Dordrecht: Reidel), p. 124.
Figure 3.11 - From Figure 2 of Fuller & Myers 1993, ApJ, 418, 273.
Figure 3.12 - From Figure 1a of Myers et al. 1991, ApJ, 376, 561.
Figure 3.13 - From Figure 41 of Benson & Myers 1989, ApJS, 71, 89.
Figure 3.16 - (a) From Figure 27 of Benson & Myers 1989, ApJ, 71, 89. (b) From Figure 1 of Ward-Thompson et al. 1999, MNRAS, 305, 143; figure supplied by P. André.
Figure 3.17 - Combination of Figure 7 of Loren 1989, ApJ, 338, 902 and Figure 8 of Goodman et al. 1990, ApJ, 359, 363.
Figure 3.18 - From Figure 3a of Goodman et al. 1993, ApJ, 406, 528.
Figure 3.19 - From Figure 4 of Bok 1977, PASP, 89, 597.
Figure 3.20 - From Figure 1 of Frerking et al. 1987, ApJ, 313, 320.
Figure 3.21 - Constructed from Table 3 of Frerking et al. 1987, ApJ, 313, 320.

Chapter 4

Table 4.1 - Data from Racine 1968, AJ, 73, 233 and Herbst 1975, AJ, 80, 503. Note that distances to several systems have been updated.
Table 4.2 - Data from de Zeeuw et al. 1999, AJ, 117, 354 and, for Orion, from Brown et al. 1994, AA, 289, 101.
Table 4.3 - From Mermilliod, J.-C. 1995, in *Information and On-Line Data in Astronomy*, ed. D. Egret and M. A. Albrecht (Dordrecht: Reidel), p. 127. For updated, electronic versions of this information, see http://obswww.unige.ch/webda/ and http://cfa∼www.harvard.edu ∼stauffer/opencl/.
Figure 4.1 - From Figure 1 and Tables 3 and 4 of Koornneef 1983, AA, 128, 84.
Figure 4.2 - From Figure 3 of Lada & Lada 1995, AJ, 109, 1682.
Figure 4.3 - From Figures 3a, 3c, and 3d of Wilking et al. 1989, ApJ, 340, 823.
Figure 4.4 - From unpublished data by Barsony, Ward-Thompson, & O‘Linger 1997; figure courtesy of J. O‘Linger.
Figure 4.5 - Constructed from Table 3 of Wilking et al. 1989, ApJ, 340, 823, and from Table 1 of Greene et al. 1994 ApJ, 434, 614.
Figure 4.6 - Constructed from Table 3 of Wilking et al. 1989, ApJ, 340, 823.
Figure 4.8 - From Figure 8 of Gomez et al. 1993, AJ, 105, 1927.
Figure 4.9 - From (a) Figure 15 of Kenyon & Hartmann 1995, ApJS, 101, 117; (b) Table 3 of Hughes et al. 1994, AJ, 108, 1071; (c) Table 6 of Lawson et al. 1996, MNRAS, 280, 1071, but adjusted to the distance of 215 pc advocated by Gauvin & Strom 1992, ApJ, 385, 217; (d) Figure 4 of Cohen & Kuhi 1979, ApJS, 41, 743, and Figure 3 of Penston 1964, Observatory, 84, 141.
Figure 4.10 - From Figure 2 of Murphy et al. 1986, AA, 167, 234.
Figure 4.11 - From Figure 1 of Alcalá et al. 1995, AAS, 114, 109.

Figure 4.12 - Combination of Figure 1 of Herbig 1954, ApJ, 119, 483 and Figure 1 of Margulis & Lada 1986, ApJ, 309, L87.
Figure 4.13 - (a) Constructed from Table A4 of Kenyon & Hartmann 1995, ApJS, 101, 117. (b) Constructed from Table 4 of Prusti et al. 1992, MNRAS, 254, 361. The initial luminosity function is based on Figure 1b of Basu & Rana 1992, ApJ, 393, 373, together with equation (4.6) of the text.
Figure 4.14 - From Figure 1 of de Zeeuw et al. 1999, AJ, 117, 354.
Figure 4.15 - Combination of our Figures 3.15 and 4.9, together with Figure 1 of Blaauw 1991, in *The Physics of Star Formation*, ed. C. J. Lada and N. D. Kylafis (Dordrecht: Kluwer), p. 125.
Figure 4.16 - From Figures 3 and 4 of Blaauw 1978, in *Problems of Physics and Evolution of the Universe*, ed. L. Mirzoyan, (Yerevan USSR), p. 101.
Figure 4.17 - From Figure 6b of de Geus et al. 1989, AA, 216, 44.
Figure 4.18 - From Figure 7 of Blaauw 1991, in *The Physics of Star Formation*, ed. C. J. Lada and N. D. Kylafis (Dordrecht: Kluwer), p. 125.
Figure 4.19 - From Figure 10 of Churchwell 1991, in *The Physics of Star Formation*, ed. C. J. Lada and N. D. Kylafis (Dordrecht: Kluwer), p. 221.
Figure 4.20 - Based on data in Meynet et al. 1993, AAS, 98, 477, and in Stauffer 1984, ApJ, 280, 189, together with additional data supplied by J.-C. Mermilliod.
Figure 4.21 - From Figure 2 of Basu & Rana 1992, ApJ, 393, 373.
Figure 4.22 - From Figure 4 of Basu & Rana 1992, ApJ, 393, 373.
Figure 4.23 - (a) From Figure 9 of Hambly et al. 1991, MNRAS, 253, 1; and data in Moraux et al. 2003, AA, 400, 891. (b) From Figure 10 of Hillenbrand et al. 1993, AJ, 106, 1906.
Plate 1 - Optical image from R. Probst, National Optical Astronomy Observatories. Near-infrared image from Figure 1 of Haisch et al 2001, AJ, 121, 1512.
Plate 2 - Photograph from Subaru Telescope, National Astronomical Observatory of Japan.
Plates 3 and 4 - Photographs courtesy of H. Zinnecker and M. McCaughrean.
Plate 5 - Image taken by L. Testi at the Telescopio Nazionale Galileo, La Palma, Spain.
Plate 6 - Photograph courtesy of J. Rayner, H. Zinnecker, and M. McCaughrean.
Plate 7 - From Figure 1 of Brandl et al. 1999, AA, 352, L69; photograph courtesy of B. Brandl.
Plate 8 - Photograph from European Southern Observatory.

Chapter 5

Table 5.1 - Relative abundances in TMC-1 are from Ohishi et al. 1992, in *Astrochemistry of Cosmic Phenomena*, ed. P. D. Singh, (Dordrecht: Kluwer), p. 171. The collision rates used to compute n_{crit} are in Danby et al. 1988, MNRAS, 235, 299.
Figure 5.2 - From Figure 9 of Van Dishoeck et al. 1993, in *Protostars and Planets III*, ed. E. H. Levy and J. I. Lunine, (Tucson: U. of Arizona Press), p. 163.
Figure 5.3 - From Figure 2 of Lis & Carlstrom 1994, ApJ, 424, 189.
Figure 5.4 - Constructed from data in Table 39 of Herzberg 1950, *Molecular Spectra and Molecular Structure: I. Diatomic Molecules*, (New York: Van Nostrand), p. 532.
Figure 5.5 - From Figure 1 of Shull & Beckwith 1982, ARAA, 20, 163; and data in Table 39 of Herzberg 1950, *Molecular Spectra and Molecular Structure: I. Diatomic Molecules*, (New York: Van Nostrand), p. 532.
Figure 5.6 - Constructed from data in Table 39 of Herzberg 1950, *Molecular Spectra and Molecular Structure: I. Diatomic Molecules*, (New York: Van Nostrand), p. 522.
Figure 5.7 - Based on an unpublished calculation by W. Welch.
Figure 5.8 - From Figure 1 of Scoville et al. 1979, ApJ, 232, L121.
Figure 5.9 - From Figure 2 of Carr & Tokunaga 1992, ApJ, 393, L67.

Figure 5.12 - From Figure 1 of Ho & Townes 1983, ARAA, 21, 239.
Figure 5.13 - The function U is the Newton-Thomas potential, given in Townes & Schawlow, *Microwave Spectroscopy*, (New York: McGraw-Hill), p. 306.
Figure 5.14 - From Figure 2 of Ho & Townes 1983, ARAA, 21, 239.
Figure 5.16 - Constructed from data supplied by G. Melnick.
Figure 5.18 - Combination of Figure 9-9 of Moran 1976, in *Frontiers of Astrophysics*, ed. E. H. Avrett (Cambridge: Harvard U. Press), p. 406; and Figure 1 of Destombes et al. 1977, AA, 60, 55.

Chapter 6

Figure 6.1 - From Figure 1 of Frerking et al. 1982, ApJ, 262, 590.
Figure 6.3 - From Figure 4-3 of Mihalas & Binney 1981, *Galactic Astronomy*, (San Francisco: Freeman), p. 216.
Figure 6.5 - From Figure 2 of Benson & Myers 1980, ApJ, 242, L87.
Figure 6.6 - From Figure 2 of Heiles et al. 1993, in *Protostars and Planets III*, ed. E. H. Levy and J. I. Lunine (Tucson: U. of Arizona Press), p. 279.
Figure 6.7 - From Figure 1 of Goodman 1989, unpublished PhD. thesis, Harvard University.
Figure 6.8 - From Figure 10-5 of Powell & Crasemann 1961, *Quantum Mechanics*, (Reading: Addison-Wesley), p. 352.
Figure 6.9 - Lower panel from Figure 1 of Goodman et al. 1989, ApJ, 338, L61.

Chapter 7

Figure 7.1 - Combination of Figure 1 of Wefel 1992, in *The Astronomy and Astrophysics Encyclopedia*, ed. S. P. Maran, (New York: Van Nostrand Reinhold), p. 134; and Figure 9a of Seo et al. 1991, ApJ, 378, 763.
Figure 7.4 - From Figure 1 of Black 1994, in *The First Symposium on the Infrared Cirrus and Diffuse Interstellar Clouds*, eds. R. M. Cutri & W. B. Latter, (San Francisco: ASP), p. 360.
Figure 7.8 - From Figure 14.1 of Melnick 1990, in *Molecular Astrophysics*, ed. T. W. Hartquist, (Cambridge: Cambridge U. Press), p. 273.
Figure 7.9 - Based on a numerical calculation by the authors.

Chapter 8

Figures 8.2, 8.4, and 8.6 - Based on calculations by the authors.
Figures 8.5 - From Figure 1 of Tielens & Hollenbach 1985, ApJ, 291, 722.
Figure 8.8 - Combination of optical photo, courtesy of K. Birkle, and CII emission contours from Figure 1 of Zhou et al. 1993, ApJ, 419, 190.
Figure 8.9 - From Figure 1 of Hasegawa et al. 1987, ApJ, 318, L77.
Figure 8.10 - From Figure 3 of Shull & McKee 1979, ApJ, 227, 131; and Figure 1 of Hollenbach & McKee 1989, ApJ, 342, 406; supplemented with a numerical calculation by the authors.
Figure 8.12 - From Figure 1 of Wardle 1999, ApJ, 525, L101.

Chapter 9

Table 9.1 - From data in Myers et al. 1995, ApJ, 442, 177; Heiles 1988, ApJ, 324, 321; Crutcher et al. 1999, ApJ, 275, 285; Goodman et al. 1989, ApJ, 338, L61; Roberts et al. 1995, ApJ, 442, 208; Plante et al. 1995, ApJ, 445, L113; Baart et al. 1986, MNRAS, 219, 145.

Figures 9.1 and 9.2 - From the tabulation of Chandrasekhar & Wares 1949, ApJ, 109, 551.
Figure 9.7 - From Figures 1 and 10 of Stahler 1983, ApJ, 268, 165.
Figure 9.8 - From Figure 9 of Stahler 1983, ApJ, 268, 165.
Figure 9.9 - From Figure 7 of Stahler 1983, ApJ, 268, 165.
Figure 9.12 - From Figure 1d and 1g of Tomisaka et al. 1988, ApJ, 335, 239.
Figure 9.13 - Combination of Figures 4 and 6 of Tomisaka et al. 1988, ApJ, 335, 239.

Chapter 10

Figure 10.3 - From Figure 2 of Fiedler & Mouschovias 1993, ApJ, 415, 680.
Figure 10.4 - From Figure 1a of Fiedler & Mouschovias 1993, ApJ, 415, 680.
Figure 10.5 - From Figure 1a of Foster & Chevalier 1993, ApJ, 416, 303.
Figure 10.6 - From Figures 3a and 3b of Foster & Chevalier 1993, ApJ, 416, 303.
Figure 10.9 - From Figure 12 of Black & Scott 1982, ApJ, 263, 696.
Figure 10.10 - From Figure 15.2 of Parker 1979, *Cosmical Magnetic Fields*, (Oxford: Clarendon Press), p. 395.
Figure 10.11 - From Figure 3 of Mestel & Strittmatter 1967, MNRAS, 137, 95.
Figure 10.18 - Plotted by F. Wilkin; see also Figure 2 of Cassen & Moosman 1981, Icarus, 48, 353.

Chapter 11

Figure 11.1 - From an unpublished calculation by H. Masanuga, S. M. Miyama, & S.-I. Inutsuka. Figure courtesy of S.-I. Inutsuka.
Figure 11.2 - From Figure 1 of Stahler et al. 1980, ApJ, 241, 637.
Figure 11.3 - From Figure 1d of Kenyon et al. 1993, ApJ, 414, 676.
Figure 11.4 - From Figure 10 of Chick et al. 1996, ApJ, 461, 956.
Figure 11.5 - Based on Figure 5 of Stahler et al. 1980, ApJ, 241, 637.
Figure 11.6 - From Figure 5 of Stahler 1988, ApJ, 332, 804.
Figure 11.8 - From Figure 7 of Stahler 1988, ApJ, 332, 804.
Figure 11.9 - Constructed from numerical calculations described in Palla & Stahler 1991, ApJ, 375, 288.
Figure 11.12 - Based on Figure 7 of Stahler et al. 1994, ApJ, 431, 341.
Figure 11.13 - From Figure 11 of Stahler et al. 1994, ApJ, 431, 341.
Figure 11.16 - From Figure 5 of Laughlin & Bodenheimer 1994, ApJ, 436, 335.
Figure 11.17 - From Figure 9 of Laughlin & Bodenheimer 1994, ApJ, 436, 335.
Figure 11.18 - From Figure 5 of Palla & Stahler 1991, ApJ, 375, 288.
Figure 11.19 - From Figure 7 of Palla & Stahler 1991, ApJ, 375, 288.
Figure 11.20 - From Figure 3 of Stahler & Walter 1993, in *Protostars and Planets III*, ed. E. H. Levy and J. I. Lunine (Tucson: U. of Arizona), p. 405.
Figure 11.21 - From Figure 1 of Palla & Stahler 1991, ApJ, 375, 288.
Figure 11.22 - Adapted from Figure 6.19 of Rolfs & Rodney 1988, *Cauldrons in the Cosmos* (Chicago: U. of Chicago), p. 366.
Figure 11.24 - Empirical data from Figure 1 of Ladd et al. 1991, ApJ, 366, 203. Profiles based on numerical calculations by the authors.
Figure 11.25 - Empirical data from Figure 4 of Chandler et al. 1990, MNRAS, 243, 330. Theoretical curve based on calculations by the authors.
Figure 11.26 - From panels (b) and (c) of Figure 19 in Zhou et al. 1994, ApJ, 421, 854.

Chapter 12

Figure 12.3 - From Figures 6 and 11 of Miyama et al. 1984, ApJ, 279, 621.
Figure 12.4 - From Figure 16 of Stahler 1983, ApJ, 268, 165.
Figure 12.5 - From Figure 10d of Monaghan 1994, ApJ, 420, 692.
Figure 12.7 - The G-dwarf histogram is based on Figure 7 of Duquennoy & Mayor 1991, AA, 248, 485; M-dwarf results adopted from Figure 2b of Fischer & Marcy 1992, ApJ, 396, 178.
Figure 12.8 - From Figure 10 of Duquennoy & Mayor 1991, AA, 248, 485; and Figure 3 of Fischer & Marcy 1992, ApJ, 396, 178.
Figure 12.9 - From Figure 5 of Duquennoy & Mayor 1991, AA, 248, 485.
Figure 12.10 - From Figure 2 of Herbst et al. 1997, AJ, 114, 744.
Figure 12.11 - From Figure 1 of Chen et al. 1990, ApJ, 357, 224.
Figure 12.12 - Panels (a) and (b) are from Figure 3a and Figure 3d, respectively, of Mathieu et al. 1989, AJ, 98, 987; panel (c) is from Figures 1 and 3 of Casey et al. 1998 AJ, 115, 1617.
Figure 12.13 - From Figure 1 of Mathieu 1994, ARAA, 32, 465.
Figure 12.14 - From Figure 2a of Mathieu 1994, ARAA, 32, 465.
Figure 12.15 - From Figure 1g of Koresko et al. 1997, ApJ, 480, 741.
Figure 12.16 - Upper panel is from data in Casey et al. 1998, AJ, 115, 1617; lower panel is adapted from Figure 2 of Welty 1995, AJ, 110, 776.
Figure 12.17 - From data in Kenyon & Hartmann 1995, ApJS, 101, 117; and Brandner & Zinnecker 1997, AA, 321, 220.
Figure 12.18 - From Figure 4 of Durisen et al. 1986, ApJ, 305, 281.
Figure 12.19 - From Figures 1c and 2a of Bonnell et al. 1996, MNRAS, 279, 121.
Figure 12.20 - From Figure 1b of Clarke & Pringle 1993, MNRAS, 261, 190.
Figure 12.22 - From Figure 2 of Bonnell & Bate 1994, MNRAS, 269, L45.
Figure 12.23 - From Figure 5 of Fletcher & Stahler 1994, ApJ, 435, 313.
Figure 12.24 - From Figures 6, 7, and 8 of Fletcher & Stahler 1994, ApJ, 435, 329.
Figure 12.25 - From Figure 17 of Fletcher & Stahler 1994, ApJ, 435, 329, which also used data from Wilking et al. 1989, ApJ, 340, 823.
Figure 12.26 - Based on data from Kenyon & Hartmann 1995, ApJS, 101, 117; and Hughes et al. 1994, AJ, 108, 1071.
Figure 12.27 - Unpublished data supplied by J. Bally and L. Hillenbrand.
Figure 12.28 - From Figure 2 of Palla & Stahler 1999, ApJ, 525, 772.
Figure 12.29 - From Figure 3 of Palla & Stahler 1999, ApJ, 525, 772.
Figure 12.30 - From Figure 5 of Palla & Stahler 1999, ApJ, 525, 772.

Chapter 13

Figure 13.1 - From Figure 2 of Reipurth 1989, Nature, 340, 2.
Figure 13.2 - From Figure 1 of Heathcote et al. 1996, AJ, 112, 1141; figure supplied by J. Morse.
Figure 13.3 - From Figure 5a of Hartigan et al. 1987, ApJ, 316, 323.
Figure 13.4 - The bowshock curve (*left side*) and theoretical line profile (*right side*) are based on calculations by the authors.
Figure 13.5 - From Figures 3 and 8 of Eislöffel & Mundt 1992, AA, 263, 292.
Figure 13.6 - Photograph courtesy of M. McCaughrean.
Figure 13.7 - From Figure 1 of Devine et al. 1997, AJ, 114, 2095; figure supplied by D. Devine.
Figure 13.8 - From Figure 4 of Rodriguez et al. 1990, ApJ, 352, 645.
Figure 13.9 - From Figure 1 of Snell et al. 1980, ApJ, 239, L17.

Figure 13.10 - From Figure 1 of Tafalla et al. 1993, ApJ, 415, L139; figure supplied by M. Tafalla.
Figure 13.11 - From Figure 5 of Moriarty-Schieven et al. 1987, ApJ, 319, 742.
Figure 13.12 - *left panel*: From Figure 17 of Levreault 1988, ApJS, 67, 283; *right panel*: from Figure 5 of André et al. 1990, AA, 236, 180.
Figure 13.13 - From Figure 2 of Masson & Chernin 1992, ApJ, 387, L47 (L1551/IRS 5; NGC 2071), and P. André, personal communication (VLA 1623).
Figure 13.14 - Based on data in Cabrit & Bertout 1992, AA, 261, 274 and André et al. 2000, in *Protostars and Planets IV*.
Figure 13.15 - From Figure 2 of Bachiller et al. 1993, ApJ, 417, L45.
Figure 13.16 - Based on a calculation by the authors.
Figure 13.19 - From Figure 6 of Goodson et al. 1997, ApJ, 489, 199.
Figure 13.20 - From Figure 1 of Hartigan & Raymond 1993, ApJ, 409, 705.
Figure 13.22 - *bottom panel*: From Figure 2d of Tenorio-Tagle et al. 1988, AA, 202, 256.
Plate 9 - From Figure 2 of Stapelfeldt et al. 1998, ApJ, 116, 372; figure supplied by K. Stapelfeldt.
Plate 10 - From Figure 12 of Reipurth et al. 1997, AJ, 114, 757; figure supplied by J. Morse.
Plate 11 - From Figure 18 of Reipurth et al. 1997, AJ, 114, 757; figure supplied by P. Hartigan.
Plate 12 - From Figure 4 of Gueth & Guilloteau 1999, AA, 343, 57.
Plate 13 - From Figure 2b of Wiseman & Ho 1996, Nature, 382, 139; figure supplied by J. Wiseman.

Chapter 14

Figure 14.1 - *Upper panel*: From Figure 4 of Welch et al. 1987, Science, 238, 1550. *Lower panel*: From Figures 1 and 3 of Dreher et al. 1984, ApJ, 283, 632.
Figure 14.2 - From Figure 2 of Walker et al. 1982, ApJ, 255, 128.
Figure 14.3 - From Figure 1 of Gwinn 1994, ApJ, 429, 241.
Figure 14.4 - From Figure 24 of Gaume & Mutel 1987, ApJS, 65, 193. For the H_2O maser positions, Table 2 of Forster & Caswell 1999, AAS, 137, 43.
Figure 14.5 - From Figure 2 of Claussen et al. 1996, ApJS, 106, 111.
Figure 14.8 - Based on a calculation by the authors.
Figure 14.11 - From Figure 4 of Torrelles et al. 1997, ApJ, 489, 744.
Figure 14.12 - Constructed from Figure 4 of Bally et al. 1993, ApJ, 418, 322 and Figure 2 of Chernin 1995, ApJ, 440, L99.
Figure 14.13 - From Figure 5 of Gwinn et al. 1992, ApJ, 393, 149.
Figure 14.14 - From Figure 1a of Greenhill et al. 1998, Nature, 396, 650.

Chapter 15

Table 15.1 - From data in Diaz-Miller et al 198, ApJ, 501, 192.
Figure 15.3 - From Figure 3 of Yorke 1986, ARAA, 24, 49.
Figure 15.4 - *Upper panel:* From Figure 4 of Felli & Harten 1981, AA, 100, 42. *Lower panel:* From Figure 2 of Garay et al. 1998, ApJ, 501, 710.
Figure 15.5 - Hubble Space Telescope photograph by J. Hester and P. Scowen; image courtesy of Hubble Space Telescope Science Institute.
Figure 15.6 - From Figure 91 of Wood & Churchwell 1989, ApJS, 69, 831. Data points at 350 µm and 7 mm are from Hunter et al. 2000, AJ, 119, 2711.
Figure 15.7 - From Figures 2a, 14a, 42, and 75 of Wood & Churchwell 1989, ApJS, 69, 831.
Figure 15.8 - From Figure 1 of Kurtz et al. 2000, in *Protostars and Planets IV*, eds. V. Mannings, A. P. Boss, and S. S. Russell (Tucson: U. of Arizona Press), p. 299.

Figure 15.9 - *Upper panel:* From Figure 1 of Moscadelli et al. 2000, AA, 360, 663 *Lower panel:* From Figure 8 of Cesaroni et al. 1997, AA, 325, 725.
Figure 15.11 - From Figure 3 of Leitherer & Robert 1991, ApJ, 377, 629.
Figure 15.12 - *Upper panel:* From Figure 3 of Kudritzki et al. 1996, MPE Report, 263, 9. *Lower panel:* From Figure 10 of Berghöfer et al. 1996, AA, 306, 899.
Figure 15.14 - From Figure 1 of Devine et al. 1999, AJ, 117, 2919, Figure 1 of Shepherd & Kurtz 1999, ApJ, 523, 690, and Figure 1 of Shepherd et al. 1998, ApJ, 507, 861.
Figure 15.15 - From Figure 11 of Henning et al. 2000, AA, 353, 211.
Figure 15.16 - From Figure 2 of White & Becker 1985, ApJ, 297, 677.
Figure 15.17 - Based on Figure 1 of Hollenbach et al. 1994, ApJ, 428, 654.
Figure 15.19 - *Left panel:* From Figure 7b of Bally et al. 2000, AJ, 119, 2919, *Right panel:* From Figure 1 of McCaughrean & O'Dell 1996, AJ, 111, 1977.
Figure 15.21 - From Figure 2 of White et al. 1997, AA, 323, 931.
Figure 15.23 - Based on calculations by the authors.
Figure 15.24 - From Figure 3 of Sugitani et al. 1995, ApJ, 455, L39.
Figure 15.25 - From Figure 4 of Lefloch & Lazareff 1994, AA, 289, 559.
Figure 15.26 - Optical photograph from the Digital Sky Survey, courtesy of the Hubble Space Telescope Science Institute. Near-infrared image courtesy of M. Burton. X-ray image courtesy of F., Seward.
Figure 15.27 - From Figures 2 and 3 of Lindblad et al. 1973, AA, 24, 309.
Figure 15.28 - From Figures 2d and 2e of Boss 1995, ApJ, 439, 224.
Figure 15.29 - From Figure 1 of Lee et al. 1977, ApJ, 211, L107.

Chapter 16

Figure 16.1 - Based on data in Hillenbrand 1997, AJ, 113, 1733.
Figure 16.2 - Herbig Ae/Be stars are from Berilli et al. 1992, ApJ, 398, 254; Testi et al. 1999, AA, 342, 515. Sources of CO outflows are from Levreault 1988, ApJ, 330, 897.
Figure 16.3 - From Figure 11 of Stahler 1988, ApJ, 332, 804.
Figure 16.5 - Based on Figure 2 of Stahler 1988, PASP, 100, 1474.
Figures 16.7 and 16.8 - Based on calculations in Palla & Stahler 1999, ApJ, 525, 772.
Figure 16.9 - Figure supplied by G. Basri.
Figure 16.10 - Based on the computed lithium abundances in Siess et al. 2000, AA, 358, 593.
Figure 16.11 - Data supplied by S. Randich.
Figures 16.12 and 16.13 - From a calculation by I. Baraffe.
Figure 16.14 - From Figure 1 of Golimowski et al. 1998, AJ, 115, 2579; figure supplied by B. Oppenheimer.
Figure 16.15 - M stars from Leggett 1992, ApJS, 82, 351; Gizis et al. 2000, AJ, 120, 1085. L dwarfs from Kirkpatrick et al. 2000, AJ, 120, 447. For the T dwarfs, see references listed in Burgasser 2001, in *Galactic Structure and the Interstellar Medium*, ed. M. Bicay and C. Woodward (San Francisco: ASP), p. 411.
Figure 16.16 - Figure supplied by G. Basri.
Figure 16.17 - Based on the data in Table II of Fukuda 1982, PASP, 94, 271.
Figure 16.18 - Velocity for NGC 6475 is from Jeffries et al. 1998, MNRAS, 300, 550. All other points are from Bouvier 1997, Mem. Soc. Astr. It., 68, 885.
Figure 16.19 - Based on data in Queloz et al. 1998, AA, 335, 183 and Terndrup et al. 2000, AJ, 119, 1303. Figure prepared with the assistance of K. Stassun.
Figure 16.20 - Based on data in Stassun et al. 1999, AJ, 117, 2941; Herbst et al. 2000, AJ, 119, 261; and Rhode et al. 2001, AJ, 122, 3258. Figure prepared with the assistance of K. Stassun.

Chapter 17

Table 17.1 - Constructed using the NIST Atomic Spectra Database, available at http://physics.nist.gov/cgi-bin/AtData/lines_form.

Figure 17.1 - Figure courtesy of C. Johns-Krull and J. Valenti.

Figure 17.3 - Based on data from Kenyon & Hartmann 1995, ApJS, 101, 117 and Herbig & Bell 1988, Lick Obs Bull, 1111 (for T Tauri stars), and Herbig 1985, ApJ, 289, 269 (for main-sequence stars).

Figure 17.4 - From Figure 7c of Hartigan 1991, ApJ, 382, 617.

Figure 17.5 - Based on data in Kenyon & Hartmann 1995, ApJS, 101, 117.

Figure 17.6 - Based on data in Herbig & Bell 1988, Lick Obs Bull, 1111 and Hartigan et al. 1995, ApJ, 452, 736.

Figure 17.7 - From Figure 4 of Tsuboi et al. 1998, ApJ, 503, 894.

Figure 17.8 - From Figure 15 of Beristain et al. 2001, ApJ, 551, 1037.

Figure 17.9 - From Figure 1 of Lavalley-Fouquet et al. 2000, AA, 356, L41.

Figure 17.10 - From Figure 1 of Hartmann & Raymond 1989, ApJ, 337, 903.

Figure 17.11 - Based on data in Hartigan et al. 1995, ApJ, 452, 736.

Figure 17.12 - From Figure 8 of Krautter et al. 1990, AA, 236, 416.

Figure 17.14 - Data on BP Tau from Kenyon & Hartmann 1995, ApJS, 101, 117; Beckwith et al. 1990, AJ, 99, 924. Data on GM Aur from Kenyon & Hartmann 1995, ApJS, 101, 117; Weaver et al. 1992, ApJS, 78, 239; Beckwith & Sargent 1991, ApJ, 381, 250; Dutrey et al. 1998, AA, 338, L63.

Figure 17.15 - From Figure 1 of Simon et al. 2000, ApJ, 545, 1034.

Figure 17.16 - From Figure 1 of Guilloteau et al. 1999, AA, 348, 570.

Figure 17.19 - Based on the data maintained by W. Herbst at $\mathrm{http://www.astro.wesleyan.edu/\sim bill/research/ttauri.html}$.

Figure 17.21 - From Figure 13 of Hatzes et al. 1995, ApJ, 451, 784.

Figure 17.22 - From Figure 4 of Ibragimov 1997, AstL, 23, 103.

Figure 17.23 - From Figure 14 of Herbig, 1977, ApJ, 217, 693.

Figure 17.24 - From Figure 1 of Haisch et al. 2001, ApJ, 553, L153.

Figure 17.25 - From Figure 6b of Stelzer & Neuhaüser 2001, AA, 377, 538.

Figure 17.26 - Theoretical lithium depletion curve from Martín et al. 1997, AA, 321, 492 and Zapatero-Osorio et al. 2002, AA, 384, 937. Data on the Taurus stars refer to weak-lined members, and are from Martín et al. 1994, AA, 282, 503. Pleiades data kindly supplied by S. Randich.

Figure 17.27 - Figure supplied by J. Greaves.

Chapter 18

Table 18.1 - Numerical data from Palla & Stahler 1991, ApJ, 375, 288.

Table 18.2 - The inner edges of the debris disks are from: Grady et al. 2001, AJ, 122, 3396 (HD 100546); Koerner et al. 1998, ApJ, 503, L83 (HR 4796A); Heap et al. 2000, ApJ, 539, 435 (β Pic); Holland et al. 1998, Nature, 392, 788 (Fomalhaut); Wilner et al. 2002, ApJ, 569, L115 (Vega).

Figure 18.1 - Stellar data from Sung et al. 2000, ApJ, 120, 333. The pre-main-sequence tracks are from Palla & Stahler 1999, ApJ, 525, 772, and the post-main-sequence isochrone from Schaller et al. 1992, AAS, 96, 269.

Figure 18.2 - From Figure 1 of Fuente et al. 2001, AA, 366, 873.

Figure 18.3 - Pre-main-sequence rotational speeds from Finkenzeller 1985, AA, 151, 340; Böhm & Catala 1995, AA, 301, 155; Mora et al. 2001, AA, 378, 116. Main-sequence data from Table II of Fukuda 1982, PASP, 94, 271.

Figure 18.4 - Spectrum of Herbig Be star from Figure 5 of Tjin A Djie et al. 2001, MNRAS, 325, 1441. We also thank P. Manoj for help with this figure. Main-sequence B6 spectrum from Figure 2b of Jacoby et al. 1984, ApJS, 56, 257.
Figure 18.5 - From Figure 2 of Grinin et al. 1994, AA, 292, 165.
Figure 18.7 - From Figure 8 of Palla & Stahler 1993, ApJ, 418, 414.
Figure 18.8 - Based on calculation in Palla & Stahler 1999, ApJ, 525, 772.
Figure 18.10 - From Figure 3 of Marconi et al. 2001, AA, 372, L21.
Figure 18.11 - From Figure 3 of Marconi & Palla 1998, ApJ, 507, L141.
Figure 18.12 - Optical and infrared data for all stars from Hillenbrand et al. 1992, ApJ, 397, 613, with additional far-infrared data for AB Aur from Meeus et al. 2001, AA, 365, 476. Submillimeter data from Mannings 1994, MNRAS, 267, 361.
Figure 18.13 - From Figure 1 of Finkenzeller & Mundt 1984, AAS, 55,109.
Figure 18.14 - From Figure 4 of Praderie et al. 1986, AA, 303, 311.
Figure 18.15 - From Figure 6a of Praderie et al. 1986, AA, 303, 311.
Figure 18.16 - From Figure 3 of Mannings & Sargent, 1997, ApJ, 490, 792.
Figure 18.17 - From Figure 4 of Bouwman et al. 2001, AA, 375, 950. Original figures kindly supplied by G. Meeus.
Figure 18.18 - Data for HD 100546 and β Pic from Figure 2 of van den Ancker et al. 1997, AA, 324, L33. Optical and near-infrared fluxes from Vega are from Malfait et al. 1998, AA, 331, 211; longer-wavelength data are from Figure 1 of Backman & Paresce 1993, in *Protostars and Planets III*, ed. E. H. Levy & J. I. Lunine (Tuscon: U. of Arizona), p. 1253.
Figure 18.19 - The image of HR 4796A is from Figure 1 of Schneider et al. 1999, ApJ, 513, L127, while the β Pic figure was kindly supplied by P. Kalas. The bottom two images are both from Figure 1 of Holland et al. 1998, Nature, 392, 788.
Figure 18.21 - From Figure 1 of Beust 1994, in *Circumstellar Dust Disks and Planet Formation,*, eds. R. Ferlet & A. Vidal-Madjar (Gif-sur-Yvette: Editions Frontières), p. 35.

Chapter 19

Table 19.l - Most data are from S. Van den Bergh 2000, *The Galaxies of the Local Group*, (New York: Cambridge U. Press). Additional references for molecular gas masses are: Mizuno et al. 2001, PASJ, 53, 971 (Large Magellanic Cloud); Mizuno et al. 2001, PASJ, 53, L35 (Small Magellanic Cloud); Young 2000, AJ, 120, 2460 (NGC 185); Heyer 2003 (personal communication) (M31); Corbelli 2003, MNRAS, 342, 199 (M33).
Figure 19.1 - Data for HII regions from Güsten & Mezger 1982, Vistas Astron, 26, 159. Data for pulsars and supernova remnants from Lyne et al. 1985, MNRAS, 213, 613 and Leahy & Wu 1989, PASP, 101, 607, respectively.
Figure 19.2 - From Figure 7 of Glass 1994, in *The Nuclei of Nearby Galaxies*, ed. R. Genzel & A. I. Harris (Dordrecht: Kluwer), p. 209.
Figure 19.3 - All images from the NASA/IPAC Extragalactic Database, available at http://nedwww.ipac.caltech/edu.
Figure 19.4 - From Figure 5 of Roberts & Haynes 1994, ARAA, 32, 115.
Figure 19.5 - From Figure 3 of Gavazzi et al. 2002, ApJ, 576, 135.
Figure 19.6 - Filled circles are from Table 2 of Kennicutt & Kent 1983, AJ, 88, 1094. Open circles are from Table 4 of Gavazzi et al. 2002, AA, 396, 449. Original data kindly provided by G. Gavazzi.
Figure 19.7 - From Figure 15 of Kennicutt 1998, in *Starbursts: Triggers, Nature, and Evolution*, ed. B. Guiderdoni & A. Kembhavi (Berlin: Springer-Verlag), p. 1.
Figure 19.8 - From Figure 11 of Kennicutt 1989, ApJ, 344, 685.

Figure 19.9 - From Figure 1a of Skelton et al. 1999, PASP, 111, 465.
Figure 19.10 - Color-magnitude diagram from Figure 1 of Gallart et al. 1999, AJ, 118, 2245. Star formation history from Figure 4 of Hernandez et al. 2000, MNRAS, 317, 831.
Figure 19.11 - Color-magnitude diagram from Figure 4 of Martinez-Delgado et al. 1998, AJ, 115, 1462. Star formation history from Figure 15 of Martinez-Delgado et al. 1999, AJ, 118, 2229.
Figure 19.12 - From Figure 1a of O'Connell et al. 1994, ApJ, 433, 65.
Figure 19.13 - From Figure 1 of Larson & Tinsley 1978, ApJ, 219, 46.
Figure 19.14 - Spectral energy distributions are from Silva et al. 1998, ApJ, 509, 103: Figures 10 (M51), 6 (M82), and 9 (Arp 220).
Figure 19.15 - Optical image from Figure 1 of Sanders et al. 1988, ApJ, 325, 74. Near-infrared image from Figure 3 of Scoville et al 1998, ApJ, 492, L107.
Figure 19.16 - Optical image from the NASA/IPAC Extragalactic Database. Near-infrared (2.12 μm) image courtesy of B. Brandl and the Cornell WIRC team, and Palomar Observatory.
Figure 19.17 - From Figure 43 of Barnes 1998, in *Galaxies: Interactions and Induced Star Formation*, ed. Kennicutt, Schweizer, & Barnes (Berlin: Springer), p. 275.
Figure 19.18 - Image of NGC 3351 from Figure 1 of Colina et al. 1997, ApJ, 484, L41. Image of NGC 1300 from Figure 1 of Aguerri et al. 2000, AA, 361, 841.
Figure 19.19 - From data compiled by Rocha-Pinto & Maciel 1996, MNRAS, 279, 447.
Figure 19.20 - From Figure 1 of Glazebrook et al. 1995, MNRAS, 275, L19.
Figure 19.21 - From Figure 9 of Steidel et al. 1999, ApJ, 519, 1.
Figure 19.22 - From Figure 1 of Dickinson 1998, in *The Hubble Deep Field*, ed. M. Livio, S. M. Fall, & P. Madau (New York: Cambridge U. Press), p. 219.
Figure 19.23 - From Figure 4 of Galli & Palla 1998, AA, 335, 403.
Figure 19.24 - Based on an unpublished calculation by D. Galli and F. Palla.
Figure 19.25 - From Figure 1 of Omukai & Palla 2001, ApJ, 561, L55.
Figure 19.26 - Upper panel from Figure 3 of Lyons et al. 1995, AJ, 110, 1544. Lower Panel from Figure 1 of Becker et al. 2001, AJ, 122, 2850.

Appendix G

Figure G.2 - From the tabulation of Rogers & Iglesias 1992, ApJS, 79, 507.

All figures from AA, AAS, and BAN are reproduced with authorization of the European Southern Observatory. Figures from ApJ, ApJS, AJ, and Icarus are reproduced courtesy of The American Astronomical Society.

The Formation of Stars
Steven W. Stahler and Francesco Palla

Index of Astronomical Objects

Subject Index