

UNIVERSITY OF AMSTERDAM

MASTER COURSE

---

# Star and Planet Formation

Part I: Stars and disks

---

*Authors:*

Carsten Domink & Inga Kamp  
February 5, 2018

# Contents

<b>1</b>	<b>Preface</b>	<b>4</b>
1.1	Origins . . . . .	4
1.2	Literature and Appendices . . . . .	4
<b>2</b>	<b>Introduction and solar system context</b>	<b>5</b>
2.1	A historical perspective . . . . .	5
2.1.1	The Sun as a star . . . . .	6
2.1.2	Models of Solar System Formation . . . . .	8
2.2	Observational Constraints . . . . .	10
2.2.1	Regularity . . . . .	12
2.2.2	Composition . . . . .	12
2.2.3	Asteroid Belt . . . . .	12
2.2.4	Ages from radioactive dating . . . . .	12
2.2.5	Deuterium . . . . .	12
2.2.6	Dynamics of Small Bodies . . . . .	13
2.2.7	Angular Momentum . . . . .	13
2.2.8	Minimum Mass Solar Nebula . . . . .	13
2.3	How did the Sun form? . . . . .	14
2.3.1	Molecular Clouds . . . . .	15
2.3.2	Starless Cores . . . . .	15
2.3.3	Young Stellar Objects . . . . .	15
2.3.4	Jets & Outflows . . . . .	16
2.4	How did the planets form? . . . . .	16
2.4.1	Protoplanetary Disks . . . . .	16
2.4.2	Debris Disks . . . . .	17
2.4.3	Exoplanets . . . . .	17
<b>3</b>	<b>Molecular clouds</b>	<b>19</b>
3.1	Observational properties of molecular clouds . . . . .	19
3.1.1	Dust and gas in clouds . . . . .	20
3.1.2	Low mass star forming clouds . . . . .	21
3.1.3	High mass star forming clouds . . . . .	21
3.1.4	Cores and clumps . . . . .	22
3.1.5	The Initial Mass Function . . . . .	23
3.2	Molecular Cloud Stability . . . . .	23
3.2.1	The Jeans mass . . . . .	24
3.2.2	The problem of star formation efficiency . . . . .	24
3.2.3	The Virial theorem for clouds . . . . .	25

<b>4</b>	<b>Cloud equilibrium and Stability</b>	<b>26</b>
4.1	Hydrostatic self-gravitating spheres	26
4.1.1	Density structure	26
4.1.2	Scale-free formulation	27
4.1.3	Rotational support	30
4.1.4	Magnetic field support	31
4.1.5	Ambipolar diffusion	32
<b>5</b>	<b>Collapse of clouds</b>	<b>33</b>
5.1	Spectral Energy Distributions (SEDs)	33
5.1.1	Lada classification	33
5.1.2	Characterization	34
5.2	Shu inside-out collapse	35
5.2.1	Hydrodynamical equations	36
5.2.2	Instabilities	37
5.2.3	Opacities	39
5.2.4	Rotating collapse	39
5.2.5	Collapse in the presence of a magnetic field	40
5.3	Line profiles	42
<b>6</b>	<b>Protostars &amp; pre-main-sequence evolution</b>	<b>44</b>
6.1	Timescales	44
6.2	Early growth and collapse	45
6.3	Dust envelope	47
6.4	Stellar structure	48
6.4.1	Mass-radius relation	49
6.4.2	Deuterium burning	49
6.4.3	Lithium destruction	49
6.4.4	Hydrogen ignition	50
6.5	Magnetospheric accretion	50
6.5.1	Theory	50
6.5.2	Observations	51
6.6	Evolution in the HR diagram	51
6.6.1	Birthline	52
6.6.2	Hayashi tracks	52
6.7	PMS classification	52
6.7.1	T Tauri stars	53
6.7.2	Weak line T Tauri stars	54
6.7.3	FU Orionis stars	54
6.7.4	Herbig AeBe stars	54
<b>7</b>	<b>The formation of massive stars</b>	<b>55</b>
7.1	The formation of massive stars by accretion	55
7.2	The formation of massive stars by collision	60
7.3	Formation of the First Stars	62
<b>8</b>	<b>Viscous Accretion Disks</b>	<b>66</b>
8.1	Disk formation	67
8.2	Disk structure	67
8.2.1	Steady state disk structure	69
8.2.2	Vertical disk structure	69
8.2.3	Radial disk structure	70

8.2.4	Temperature profile of an accretion disk . . . . .	70
8.3	Disk evolution . . . . .	71
8.4	Angular momentum transport . . . . .	72
8.4.1	Turbulent viscosity . . . . .	73
8.4.2	Shakura-Sunyaev viscosity . . . . .	74
8.4.3	MRI . . . . .	74
8.4.4	Gravitational instabilities . . . . .	75
8.4.5	Dead zones . . . . .	75
8.5	FU Orionis outbursts . . . . .	76
<b>9</b>	<b>Irradiated Disks</b> . . . . .	<b>77</b>
9.1	Basic principles . . . . .	77
9.2	Spectral energy distribution . . . . .	78
9.3	Flat irradiated disks . . . . .	79
9.4	Flaring disks . . . . .	80
9.4.1	Two-layer Chiang & Goldreich disks . . . . .	80
9.4.2	Radiative equilibrium . . . . .	81
9.4.3	Inner disk radius . . . . .	86
9.4.4	SEDs . . . . .	86
9.5	Disk dispersal . . . . .	87
9.5.1	Photoevaporation . . . . .	87
9.5.2	External evaporation . . . . .	89
<b>10</b>	<b>Observations of Disks</b> . . . . .	<b>91</b>
10.1	SEDs . . . . .	92
10.1.1	Measuring the disk mass . . . . .	92
10.1.2	Surveys at mm wavelength . . . . .	93
10.1.3	IR surveys . . . . .	94
10.2	Imaging . . . . .	94
10.2.1	Imaging with (sub)mm interferometry . . . . .	95
10.2.2	Scattered light . . . . .	96
10.3	Spectroscopy . . . . .	96
10.3.1	Measuring the grain size . . . . .	98
10.3.2	Line observations with (sub-) mm interferometers . . . . .	98
10.3.3	IR line observations . . . . .	99
10.3.4	Optical line observations . . . . .	100
10.3.5	Measuring disk gas mass . . . . .	100
<b>A</b>	<b>Fluid dynamics and the equation of state</b> . . . . .	<b>101</b>
A.1	Simple equations of state . . . . .	101
A.2	Continuity equation . . . . .	102
A.3	The equation of motion under pressure . . . . .	103
A.4	Fluid motion in a gravitational potential field . . . . .	103
A.5	Fluid motion in a magnetic field . . . . .	104

# Chapter 1

## Preface

### 1.1 Origins

Astronomy stood at the beginning of the human quest for origins. The questions about where we come from and where we are headed as humans belong to the deepest questions humanity is asking. Modern Astronomy has made it possible to ask this question in the widest possible form, with theoretical physics, particles physics and cosmology studying the origin and the evolution of the universe as a whole.

This course focuses on the more immediate part of this question: *Where does our planet Earth come from*. Only in the 20th century it has become clear how closely this question is related to the origin of the Sun itself as it was realized that the formation of a star and its planetary system are happening at the same time and are closely linked. Even though both star formation and planet formation are huge areas each of which would warrant a complete master course on its own, we will cover both the formation of stars and planets in a single course in order to show the connectedness of both. This course focuses more on the physical concepts than on observations - so even if we will refer to observational results often, they will not be covered in full breadth.

### 1.2 Literature and Appendices

The course will make reference to the basic equations of hydrodynamics. The derivation of these equations is not part of the course, but appendix A will summarize the derivation that would normally be covered by a course in fluid dynamics. Along the same lines, we will make reference to and use of molecular and dust tracers that are being used to probe the interstellar cloud material and circumstellar material that is involved in the formation of stars and planets. Again, covering the detailed chemistry and radiation processes involved here is subject of another course (Interstellar Matter). In appendix ??, basic information about these aspects is summarized, as a reference for the student.

A number of books provide good reading alongside this course

1. Stahler & Palla: “The Formation of Stars”, Wiley, ISBN-13: 978-3527405596
2. Ward-Thompson & Whitworth “An Introduction to Star Formation”, Cambridge University Press, ISBN-13: 978-0521630306
3. Armitage: “Astrophysics of Planet Formation”, Cambridge University Press, ISBN-13: 978-0521887458. A precursor to this book is available for free: <http://arxiv.org/abs/astro-ph/0701485>

## Chapter 2

# Introduction and solar system context

This chapter presents a brief introduction into the course by putting our own Solar System in which we live into the broader context of star and planet formation. It will briefly touch upon the various historical models of our Solar System as well as present a very qualitative overview of star and planet formation in the context of modern research and observations.

### 2.1 A historical perspective

Before the 16th century, the Earth was thought to be at the center of the Solar System (geocentric model). Only one greek philosopher, Aristarchos of Samos, had proposed earlier around 280 BC that the planets were on circular orbits around the Sun. This was largely discarded because of the following three arguments (Aristotele's school):

1. If the Earth rotates around the Sun, birds should actually stay behind because of the movement of the Earth on its orbit.
2. If the Earth rotates around its axis (as required to explain day and night), things should fly off the spinning planet.
3. If the Earth rotates around the Sun, we should observe parallaxes for the fixed stars.

While the first two can actually be attributed to an inadequate understanding of the physics involved, the third one is true, but the effect of parallax is so small that it could not be observed at that time with the naked eye.

In the 16th century, Nicholas Copernicus (1473-1543) was the first to advance the model that all planets orbit the Sun. This change of paradigm was mainly driven by the complicated dynamical model of epicycles that were needed to explain the complex motion of the planets (e.g. Mars retrograde motion, see Fig. 2.2). Later, Johannes Kepler (1571-1630) formulated his famous laws of planetary motion based on the very accurate measurements of Tycho Brahe (1546-1601):

1. The planets revolve on elliptical orbits around the Sun, with the Sun in one focus.



Figure 2.1: This is a montage view of the Solar System with its eight planets and the Moon, the dwarf planet Pluto, some asteroids, and a comet.

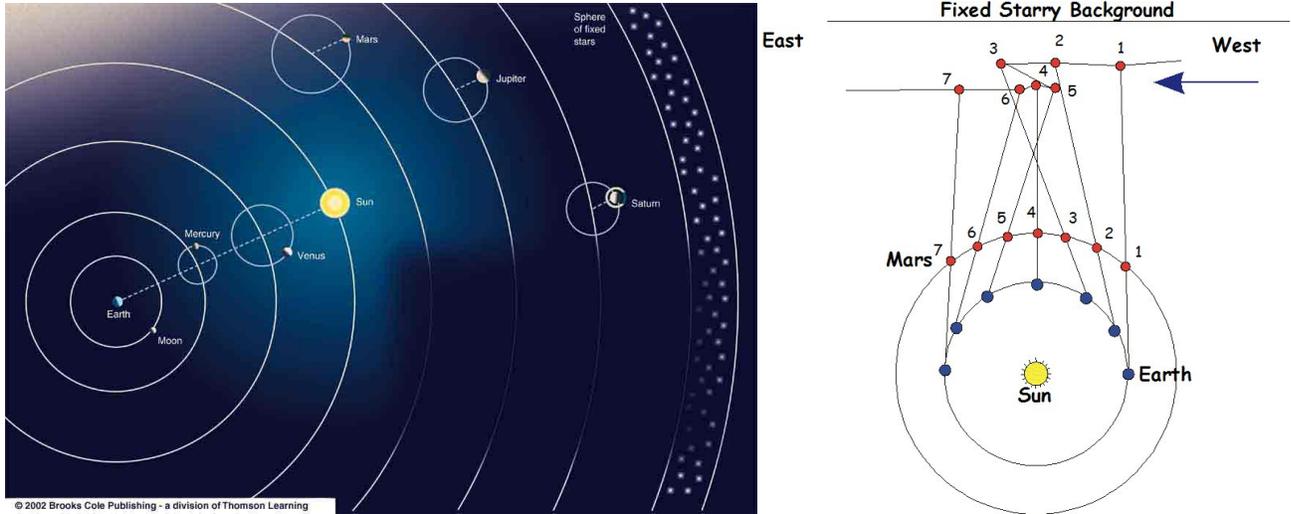


Figure 2.2: Left: The order of planets and concept of epicycles as used in the geocentric model of Ptolemy. Right: The retrograde motion of Mars.

2. The area swept out by the radius vector from the Sun to the planet per unit time  $t$  is constant

$$\frac{dA}{dt} = \frac{1}{2} \left( r^2 \frac{d\theta}{dt} \right) \quad (2.1)$$

3. The square of the orbital period  $T$  divided by the cube of the mean distance from the Sun  $a$  is the same for all planets

$$\frac{T^2}{a^3} = \frac{4\pi}{GM_{\odot}} \quad (2.2)$$

Isaac Newton (1642-1727) finally provided in the 17th century a universal theory of gravitation from which the laws of Kepler could now be derived in a mathematical fashion.

### 2.1.1 The Sun as a star

In ancient times, the Sun was often a deity and even some stars were personified or the fixed star sphere itself attributed to a deity. Pythagoras already taught in the 6th century BC that the Earth, Moon and Sun are in fact spheres and as such bodies in the heaven. Already the Greek philosophers noted that the Sun and the Stars belong to the same category; they were thought to be 'stones on fire'.

**Temperature of the Sun** The first estimates of the Sun's temperature ranged from 4 million degrees Kelvin (William Herschel) to 100 000 K (end of 19th century). It was noted at the end of the 19th century that one needed to measure the total energy received on Earth per surface area in order to estimate its surface temperature correctly. Claude Pouillet was the first to try such a measurement in 1837. He also coined the term "solar constant" for the amount of energy received on Earth per minute and per square centimeter from the Sun. We know now that this is by far not a constant, but the expression stayed. His measurement in 1838 resulted in a solar constant of  $S = 1.7633 \text{ cal cm}^{-2} \text{ min}^{-1}$ , very close to the current value of 1.94. Assuming an emissivity of the solar photosphere between 1 and 1/10, he estimated a surface temperature between 1500 and 1800 K.

John Herschel tried a similar measurement with a thermometer encased in a small tin box filled with a measured amount of water which, with the help of a black umbrella, was alternatively shaded and then

exposed to sunlight. The measured heating of the water was used to define the solar energy received on Earth. With this, he studied extensively the variation of the solar irradiance with time and height on Earth.

Only at the turning point of the 19th century, better measurements became available through actinometers. They use two different metals (e.g. bismuth and antimony) soldered together. If the soldering point is heated, a weak current can be measured and the intensity of the current can be converted into the solar irradiance. With such instruments, the surface temperature of the Sun was estimated to be  $T_{\text{eff}} = 7338$  K. The assumption is here that the Sun radiates as a black body

$$\pi F = S \left( \frac{d}{R_{\odot}} \right)^2 = \sigma T_{\text{eff}}^4, \quad (2.3)$$

where  $\sigma$  is the Stefan-Boltzmann constant,  $d = 1$  AU the distance between Earth and Sun and  $R_{\odot}$  the solar radius. The first estimates of the temperature of the Sun neglected the influence of the Earth atmosphere, i.e. the fact that the atmosphere is opaque at ultraviolet wavelength and also in the infrared and beyond — except for certain narrow windows.

The diameter of the Sun had been estimated from solar eclipses and its mass from the orbital motion of the planets using Kepler's third law. These two measurements enable an estimate of the mean density of the Sun to be  $1.4 \text{ g cm}^{-3}$ . From this low density, it became clear that the Sun must be a ball of hot gas. The Greek hypothesis that it is a stone of fire and the later hypothesis of a hot burning metal was clearly ruled out.

**Spectroscopy** The physical understanding of stars however had to await the discovery of spectroscopy. Fraunhofer was the first to study the solar spectrum and to discover about 600 dark lines in there; he designated the strongest ones with letters such as the D-lines, which we still denote as the sodium D resonance doublet (the  $D_1$  and  $D_2$  lines at 589.6 nm and 589.0 nm). The interpretation of the lines could proceed due to Kirchhoff's experiments in Heidelberg (1859), where he showed that Fraunhofer's black lines coincided with bright emission lines of evaporating sodium in the laboratory. He formulated his famous three laws

1. A hot solid body produces a simple continuous spectrum without lines.
2. A hot gas produces a spectrum with bright lines at discrete wavelengths. The position and number of lines depends on the nature of the gas.
3. If a continuous spectrum shows dark lines, it originated from a hot solid body surrounded by a gas that is cooler than the hot solid body before reaching us. Again, the number and position of the dark lines depends on the chemical nature of the surrounding cooler gas.

The second important discovery was the mathematical description of continuous electromagnetic radiation from bodies of a certain temperature, the so-called black body radiation. Planck's law

$$I_{\nu}(T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/kT} - 1} \quad (2.4)$$

describes the total energy emitted by a black body of temperature  $T$  per time interval, per frequency interval  $d\nu$  and per surface area ( $h$  is the Planck constant,  $k$  the Boltzmann constant and  $c$  the speed of light). These discoveries enabled the detailed study of astronomical objects through the analysis of their spectra.

Through the analysis of many stellar spectra besides the Sun's spectrum, it was unambiguously confirmed that the fixed stars were indeed distant Sun's. The differences in their spectra indicated that they had different masses, sizes, temperatures and chemical composition, but also most likely different ages.

**Energy source** The next step of understanding came from the analysis of the solar energy budget. By the beginning of the 20th century it became evident from radiometric dating that the Earth had an age of roughly 4.5 billion years. Hence, the Sun must have an age similar or even exceeding that. This means that the Sun — and also all other stars — need to be stable over a very long timescale. While gravity causes the

Sun to collapse, the internal pressure of the hot gas and radiation inside the Sun counteract this collapse. Hence, the Sun is in hydrostatic equilibrium and loses energy at about the same rate as it generates energy.

Herman von Helmholtz noted already in 1854 that the Sun (mass  $M_\odot$ , radius  $R_\odot$ ) could support its current level of radiation for about 20 million years if it were to use gravitational energy (gravitational constant  $G$ ) only

$$E_{\text{tot}} = \frac{3}{5} \frac{GM_\odot^2}{R_\odot} = 2.28 \times 10^{48} \text{ erg} . \quad (2.5)$$

The timescale estimated from this gravitational energy is called the Kelvin-Helmholtz timescale  $t_{\text{KH}}$ . It is much too short compared to the measured age of the Solar System. Hence, a different source of energy needed to be found.

Sir Arthur Eddington proposed in 1920 that the temperature in the core of the Sun would be high enough for nuclear fusion. The Sun could then burn hydrogen in its interior converting it into helium, an idea also proposed by J. Perrin. This was worked out in detail after nuclear science had established itself in 1938 by Bethe and von Weizsäcker. They showed which nuclear reactions could take place in the core of the Sun to burn hydrogen to helium (Fig. 2.3). In the proton-proton chain for example, two protons react to form deuterium, a positron and a neutrino. The deuterium reacts with another proton to form  $^3\text{He}$  and a  $\gamma$ -ray. Next, two  $^3\text{He}$  atoms combine to form  $^4\text{He}$  and return two protons back into the chain. The net mass balance is given by the difference between four hydrogen atoms (proton mass  $m_p$ ) and a Helium atom (two protons, two neutrons, mass  $m(^4\text{He})$ )

$$4p \rightarrow ^4\text{He} \quad (2.6)$$

$$E = \Delta mc^2 = (4m_p - m(^4\text{He})) c^2$$

This yields an energy of 26.3 MeV (or  $4.21 \times 10^{-5}$  erg) per fusion reaction. Assuming that the entire Sun consists of hydrogen and that all the hydrogen will be converted eventually into Helium, leads to a total energy of

$$E_{\text{tot}} = \frac{M_\odot}{4m_p} 4.21 \times 10^{-5} \text{ erg} = 1.25 \times 10^{52} \text{ erg} \quad (2.7)$$

which gives a lifetime for the Sun of

$$t = \frac{E_{\text{tot}}}{L_\odot} = \frac{1.25 \times 10^{52} \text{ erg}}{3.85 \times 10^{33} \text{ erg/s}} = 3.25 \times 10^{18} \text{ s} = 1.03 \times 10^{11} \text{ yr} . \quad (2.8)$$

This result is consistent with the actual measured age of the Solar System. We know now from stellar evolution models that the Sun has an age of 4.57 billion years and a core temperature of 15 million Kelvin. Its surface temperature is 5780 Kelvin.

## 2.1.2 Models of Solar System Formation

Historically, the various models of Solar System formation can be divided into four categories: models in which the Sun and the planets formed from the same material or from different material and models in which

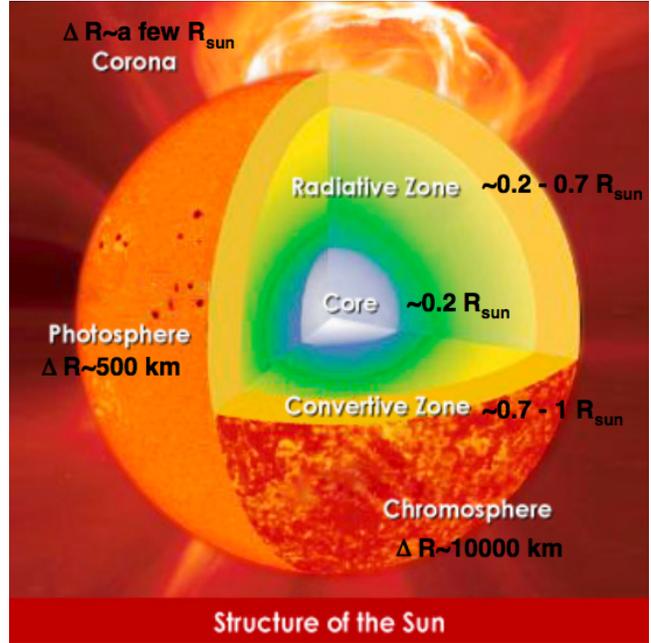


Figure 2.3: Schematic view of the interior of the Sun. The solar radius is  $R_{\text{sun}} = 6.96 \times 10^5 \text{ km}$ .

the Sun and planets formed simultaneously or at different times (Encrenaz 2004). For many centuries, the only constraint was the known orbital motion of the planets: all planets revolve on circular coplanar orbits. Only in the 20th century, knowledge of individual planetary composition, meteorites and orbital properties of Kuiper Belt objects was gathered and allowed to constrain and refine the model of the formation of our Solar System.

	<b>Planets formed from interstellar matter</b>	<b>Planets formed from stellar matter</b>
<b>The Sun and planets formed simultaneously</b>	Solar Nebula	binary star
<b>The Sun and planets did not form simultaneously</b>	capture of a passing IS cloud	capture of a passing star

Table 2.1.2 presents an overview of the categories for planet formation theories arising from the two fundamental questions whether the Sun and Planets formed together and whether they formed from the same material. We mention here only a few examples that got more scientific attention over the centuries.

**Tidal (catastrophic) theories** These theories have in common that they postulate a catastrophic event as the trigger to form the Solar System. This can be either a collision with a comet (Buffon, 18th century) or a collision with a star (Bickerton, Moulton, Chamberlain, 19th and 20th century). In these scenarios, the collision extracts material from the Sun. The material stays in a bound orbit around the Sun and then forms the planets through condensation. A later variant by Wolfson (20th century) proposes the collision between the Sun and a protostar, where the protostar gets disrupted and its material captured in orbit around the Sun. This variant better explained the measured chemical and isotopic compositions of the planets. These theories fall into the lower right box of Table 2.1.2.

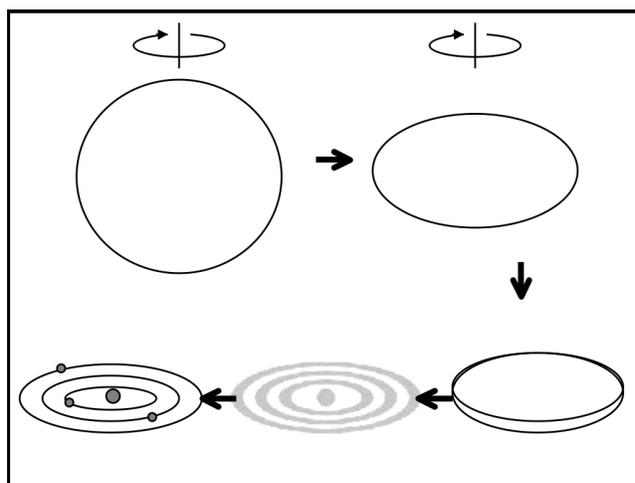


Figure 2.4: Schematic view of the nebula theory as proposed by Kant and Laplace.

**Accretion theories** These theories propose the accretion of interstellar matter from one or more clouds of material. Schmidt suggested that the presence of a second star closeby allows the captured material to end up in stable orbits and condense into the planets. Alfvén and Arrhenius even suggested that the inner terrestrial planets formed through the capture of a cloud containing refractory grains, while the outer gaseous planets formed from the capture of a pure hydrogen cloud. These theories fall into the lower left box of Table 2.1.2.

**Nebular theory** Originally, this theory was developed by Kant and Laplace in the 18th century. Fig. 2.4 illustrates schematically the basic idea: the initial interstellar cloud collapses under the influence of gravitation and angular momentum conservation leads to the formation of a flattened disk around the central mass. Within this disk, rings of material form that subsequently condensate into the individual planets and their satellites.

The two variants of the nebula theory that have been proposed to address the planet condensation problem are the massive nebula (Cameron) in which the disk contains as much mass as the Sun and becomes gravitationally unstable. The planets form then through gravitational instabilities very much like the star itself. The second variant is the low-mass nebula (Safronov) in which the disk contains only about 1% of the solar mass and the planetary cores form through collisional accretion of small dust grains and subsequent growth until kilometer-sized planetesimals. The most massive planetary cores ( $\gtrsim 10 M_{\text{Earth}}$ ) can then gravitationally attract the surrounding gas to form even more massive gaseous planets (e.g. Jupiter, Saturn). Which of these two variants truly describes the formation of the Solar System is still heavily debated. We get back to this question when we confront the theories with observations of our own Solar System in the next chapter and also with the growing evidence of extrasolar planetary systems in the last chapter.

This model – often referred to as the ‘core-accretion’ model – naturally explains why the planets orbit the Sun in the same plane and direction. However, the condensation of a planet from a ring of material is difficult as shown by Maxwell in the 19th century. The second problem is the angular momentum: the Sun contains 99.8% of the mass of the Solar System, but only 2% of the angular momentum.

## 2.2 Observational Constraints

For many centuries, our own Solar System was the only known planetary system. Planet formation could thus only be studied by looking at the present dynamical properties and composition of bodies in the Solar System, 4.5 billion years after it formed. From those observations, scientists had to derive their formation history indirectly. The situation changed with the discovery that many young stars are in fact surrounded by disks of gas and dust, the so-called protoplanetary disks. We think that these disks are the birthplaces of new planetary systems and hence they allow us to study the conditions of planet formation in situ at much earlier times in stellar evolution (typically 1-10 million years). Fig. 9.12 shows examples of such disks in the Orion star forming region and in Taurus.

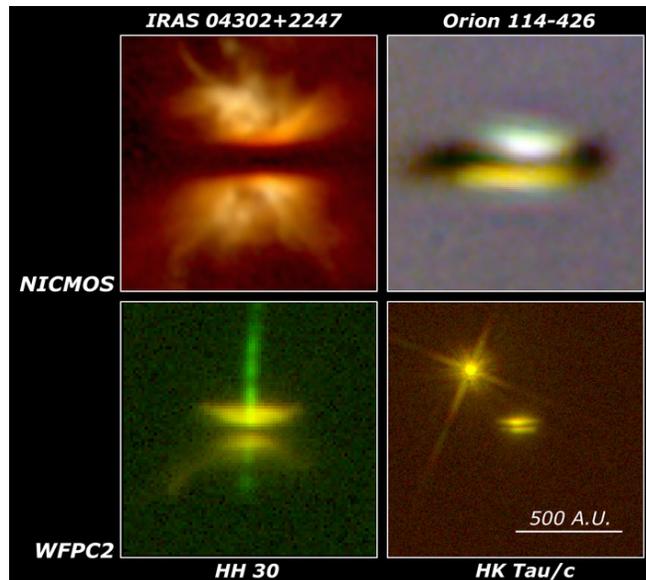


Figure 2.5: Composite of protoplanetary disks in various star forming regions viewed edge-on. The images are taken with the WFPC2 (optical) and NICMOS (near-IR) camera on board the Hubble Space Telescope.

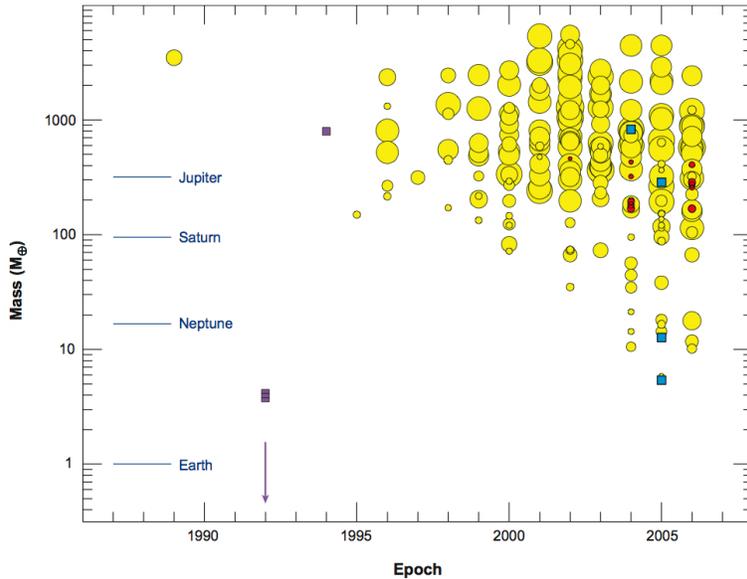


Figure 2.6: Masses of the known extrasolar planets in Earth masses as a function of the year of the discovery. Yellow and red circles denote planets discovered using the radial-velocity and transit techniques, respectively. The masses of the planets discovered using the radial-velocity technique are minimum masses. The size of the circles is proportional to the orbital period of the planet. Blue and purple squares represent the planets discovered by microlensing and pulsar-timing surveys. The purple arrow indicates the discovery of a very low-mass planet through the pulsar-timing technique. The masses of Jupiter, Saturn, Neptune, and the Earth are marked for comparison. The temporal evolution of the lightest planet detection points toward a future detection of an Earth-mass planet. Taken from Udry (2007)

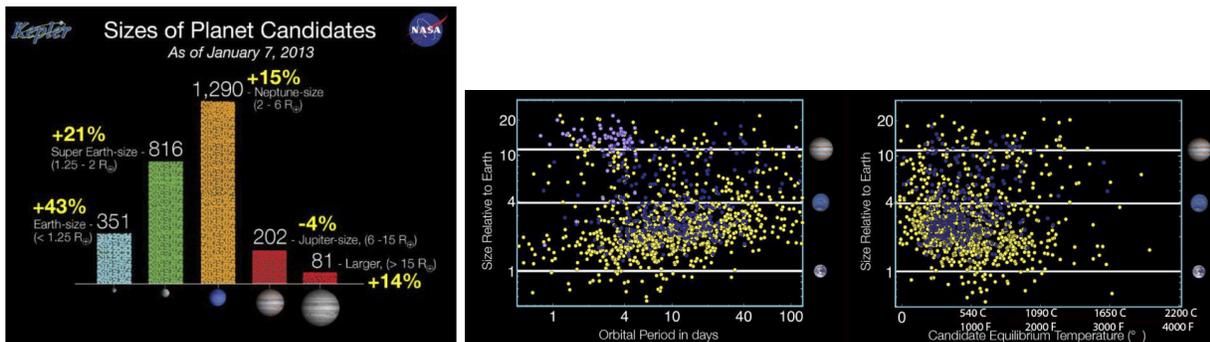


Figure 2.7: Left: Sizes of planet candidates discovered by the Kepler mission. Right: Locations and temperatures of planet candidates discovered by the Kepler mission. Credits NASA.

In addition, the growing number of exoplanets and extrasolar planetary systems discovered opens a new perspective on the statistics of planetary system architecture (Fig. 2.6). Most of the systems discovered up to now look very different from our own Solar System with massive Jupiter-like planets on short period orbits very close to the star (so-called hot Jupiters). As instruments become more sensitive and our observing time base becomes longer, we discover less exotic systems that resemble more our own. However, the finding of the exotic systems of course challenges the current standard planet formation model (core accretion). It could very well be that within the variety of planetary systems that are and will be discovered within our galaxy, there is room for more than one mode of planet formation.

In the following, we list a number of observational facts that help us to constrain models of planet formation. A more extensive overview can be found in chapter 4 of Encrenaz (2004, The Solar System).

### 2.2.1 Regularity

All planets revolve on nearly circular orbits in the same sense as the Sun and within a plane very close to the ecliptic (within  $7^\circ$ ). They also spin around axis that are within  $30^\circ$  perpendicular to the orbital plane (with the exception of Uranus) and have the same sense of rotation (all prograde except Venus and Uranus). This regularity is easily explained if the planets formed from a flat disk of material orbiting the Sun.

### 2.2.2 Composition

The differences in planetary composition, rocky planets inside 2 AU, gas giant planets beyond and icy planets at distances larger than 15 AU, is also better explained through the core-accretion model than through a model where planets form from the gaseous nebula through gravitational instabilities.

### 2.2.3 Asteroid Belt

It consists of a large number of km sized objects that are very likely left-over planetesimals that never succeeded to form a planet due to the gravitational influence of Jupiter.

### 2.2.4 Ages from radioactive dating

Radioactive dating of meteorites in our Solar System indicate that it is  $4.55 \times 10^9$  yr old. Short-lived isotopes also reveal that the first condensations from the solar nebula formed within  $10^8$  yr after the formation of the nebula and the Sun itself. Analysis of small inclusions in meteorites, the chondrules (Fig. 2.8), leads to an age estimate of  $4.56 \times 10^9$  yr. These are the oldest objects in our Solar System. The age of the oldest rocks on Earth is  $4.3 \times 10^9$  yr and that of the moon  $4.4 \times 10^9$  yr. This is clear evidence that the Sun and the planets formed at the same time and that the first aggregation phase happened very fast (less than  $10^6$  yr in the inner Solar System (within a few AU).

### 2.2.5 Deuterium

The D/H (deuterium over hydrogen) isotopic ratio in Jupiter has been found to be slightly larger than that in the interstellar medium ( $D/H = 2 \pm 0.1 \times 10^{-5}$ ). In stars, deuterium is destroyed as soon

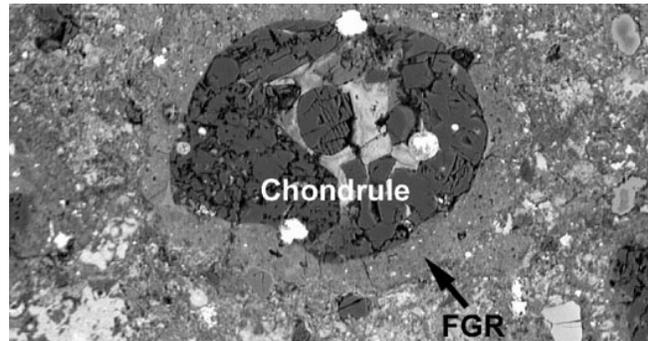


Figure 2.8: Chondrules are small silicate particles that make up large components of many of the meteorites that fall on Earth. They are believed to be the most primitive objects in the solar system. FGR stands for 'Fine Grained Rim' which can be seen and delineated.

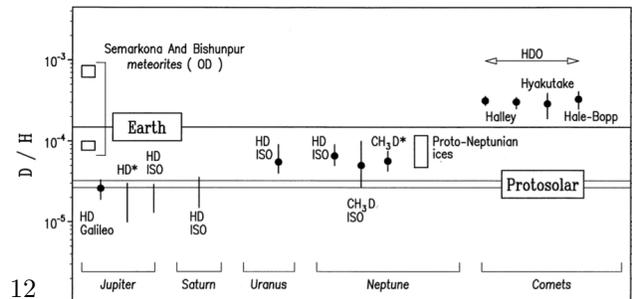


Figure 2.9: D/H isotopic ratio for bodies in the Solar System

as hydrogen burning starts. This immediately excludes the hypothesis that the planets formed from solar material. The D/H ratios found in comets also indicate some sort of deuteration that must have occurred in the ices of the low temperatures regions of the outer solar nebula (Fig. 2.9).

### 2.2.6 Dynamics of Small Bodies

The dynamics of the small bodies of the Solar System, asteroids, comets, Kuiper Belt objects is thought to carry the imprint of the dynamics of the early Solar System. Especially, the mass, eccentricities, resonances and position of the present day Kuiper Belt objects has been used recently to constrain the collisional history of the planetary system. The Nice model (Gomes et al. 2005) suggests that the gaseous planets formed much closer to each other and when Jupiter and Saturn crossed their 2:1 mean motion resonance, the system became briefly instable and Neptunus moved outwards. Placing a planetesimal reservoir at the right distance, these changes of planetary orbital parameters can explain the Late Heavy Bombardment phase (a phase of intense cratering as observed e.g. on the surfaces of the Moon and Mars), its timing ( $3.9 \times 10^9$  yr ago) as well as the currently observed properties of Kuiper Belt objects.

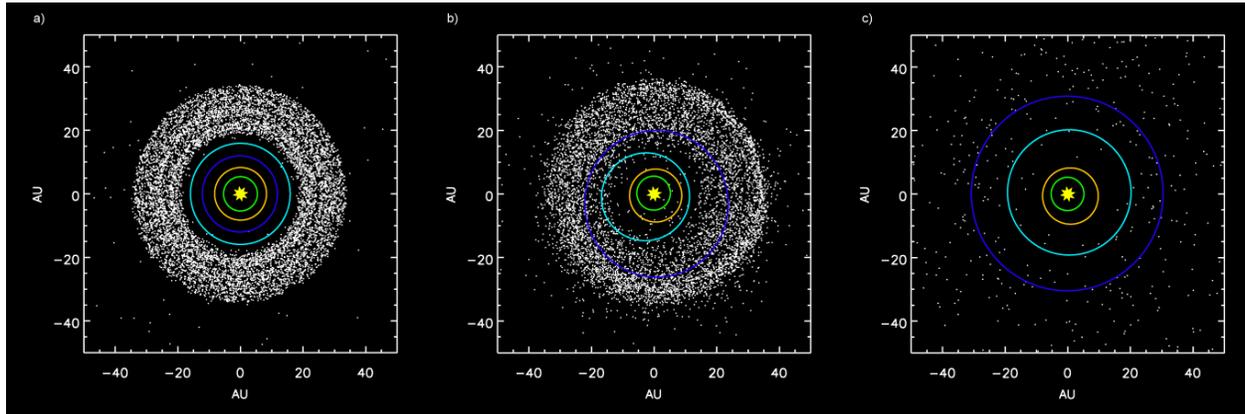


Figure 2.10: Simulation of Late Heavy Bombardment by Gomes et al. (2005): a) Before Jupiter (green) and Saturn (yellow) reach their 2:1 resonance, b) Scattering of planetesimals into the inner Solar System when resonance occurs, c) After ejection of planetesimals (Uranus: cyan, Neptunus: blue).

### 2.2.7 Angular Momentum

While the Sun contains 99.8% of the entire mass of the Solar System, 98% of its angular momentum is contained within the planets. If the Sun had acquired all the systems angular momentum it should rotate with a period of 12 hours instead of 26 days. Magnetic braking in which mass lost from the Sun was carried to large distances could have played a role in carrying along large amounts of angular momentum. The second possible mechanism is a strong solar wind in the early phases of stellar evolution.

### 2.2.8 Minimum Mass Solar Nebula

To estimate the minimum mass and structure of the solar nebula from which the planets formed, we simply assume that they have formed within their feeding zones and that the original composition of the solar nebula was primordial. Since the terrestrial planets formed only from the refractory mass in the nebula — the volatiles were lost and not incorporated — it would have taken 300 Earth masses of original nebular

material (primordial abundances) to form Earth. In this way, we can reconstruct the entire radial nebula structure, i.e. the surface density of the nebula as a function of distance  $r$  from the Sun (Fig. 2.11). This gives a power law surface density distribution  $\Sigma(r)$  with an index of  $p = -1.5$ . If we were to take into account the dynamical evolution according to the proposed Nice model, the surface density distribution would be even steeper (see Desch 2007). Kuchner (2004) has done similar estimates for extrasolar planetary systems (radial velocity method). He obtained also a steep index

of  $p$

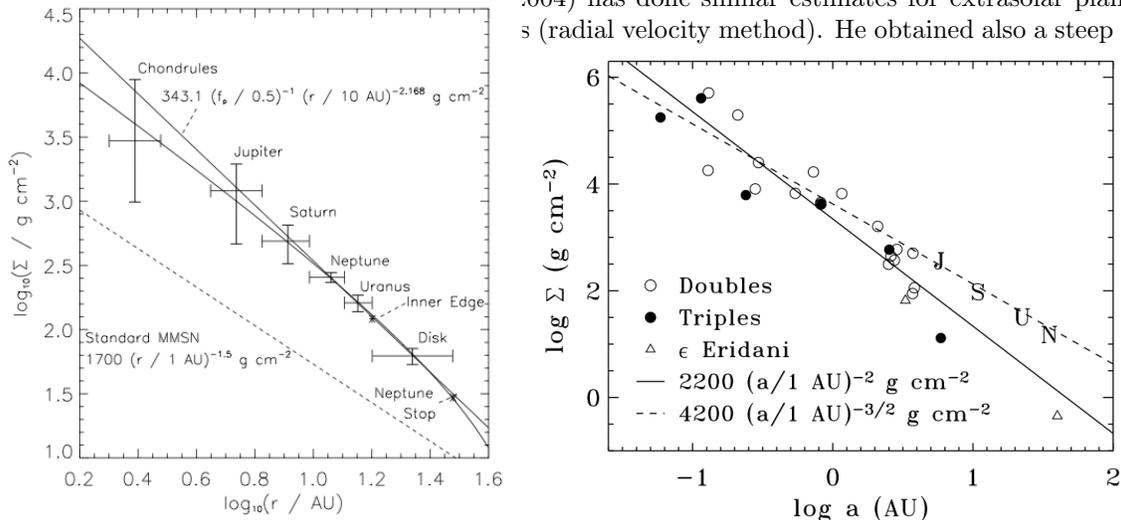


Figure 2.11: Left: Minimum mass solar nebula: surface density distribution in the solar nebula as a function of distance from the star (taken from Desch 2007). Right: Minimum mass extrasolar nebulae (taken from Kuchner 2004).

## 2.3 How did the Sun form?

Stars form from dense cores of interstellar gas and dust that become unstable under their own gravity and collapse. As the core condensation becomes denser and its density and temperature raise to the values necessary for nuclear fusion, the star enters a long stable period in its life, the main-sequence phase. After the hydrogen fuel is exhausted, stars such as the Sun become red giants and develop strong stellar winds and eject shells of material at the end of their lifetime. In this way, they return up to 10% of their mass back to the interstellar medium. The remainder of the mass forms a very compact Earth-size object, a white dwarf, that slowly cools as it ages until it is too cold to be observed anymore. This scenario forms part of the cosmic matter cycle (Fig. 2.12), in which gas and dust are recycled many times within a galaxy.

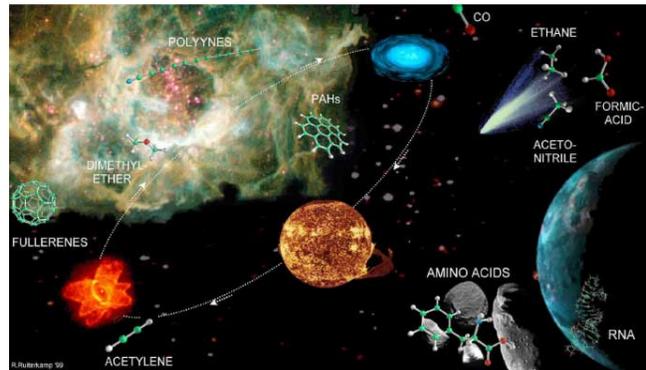


Figure 2.12: The cosmic matter cycle (figure from R. Ruiterkamp).

### 2.3.1 Molecular Clouds

Molecular clouds have typical densities of  $10^3 - 10^4$  hydrogen atoms per  $\text{cm}^3$  and temperatures of the order of 10 K. They consist of dust and — as the name says — molecular gas. Historically, the presence of molecular hydrogen was anticipated by the break-down of the linear correlation between extinction — due to IS dust — and atomic hydrogen column density ( $N_{\text{H}}$ ) — due to IS gas (Fig. 2.13). The fact that the column density does not increase beyond an extinction  $A_V$  of 1 mag, is attributed to the opaqueness of these regions that prevent destructive ultraviolet radiation to enter and thus allow molecular hydrogen and other molecules to form.

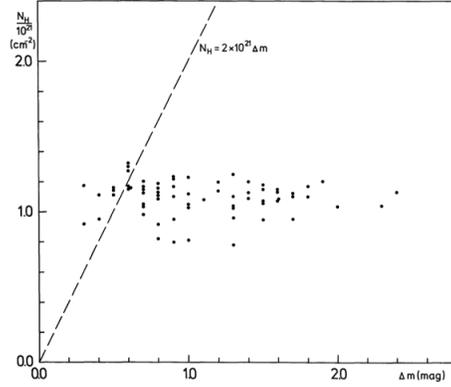


Figure 2.13: Correlation between dust extinction and atomic hydrogen column density.

### 2.3.2 Starless Cores

Objects with even higher densities and temperatures ( $10^5 - 10^6 \text{ cm}^{-3}$ , 30-100 K) that do not yet contain a protostar, are called starless cores or dense cores. Fig. 2.14 shows an example of such an object that was thought for many decades to be starless. Only recently with the advent of sensitive IR space telescopes, we could see for the first time inside these highly extinguished regions (at visual wavelength, left hand side of figure) and discover the presence of a young protostar (right hand side of the figure). Typical masses of these cores are  $10^4 M_{\odot}$  and their sizes range between 5 and 10 pc. The criterium telling us whether these cores are stable or not can be derived from a comparison of their gravitational potential and internal energy

$$\frac{3}{5} \frac{GM^2}{R} \geq \frac{3}{2} \frac{M}{m} kT, \quad (2.9)$$

where  $G$  is the gravitational constant,  $k$  the Boltzmann constant and  $M$  and  $R$  are the mass and radius of the core. The mean mass of the gas is given by  $m$  and can be assumed to be  $m = \mu m_{\text{H}} = 2.6 m_{\text{H}}$  ( $m_{\text{H}}$  being the mass of a hydrogen atom and  $\mu$  the mean molecular weight) since we deal here with molecular material (atomic matter has a mean molecular weight of  $\mu = 1.3$ ). From this equation, we can see that a core of density  $\rho$  and temperature  $T$  becomes unstable, if it is larger than

$$R_{\text{J}} = \left( \frac{15kT}{8\pi Gm\rho} \right)^{1/2}, \quad (2.10)$$

the Jeans length  $R_{\text{J}}$ .

### 2.3.3 Young Stellar Objects

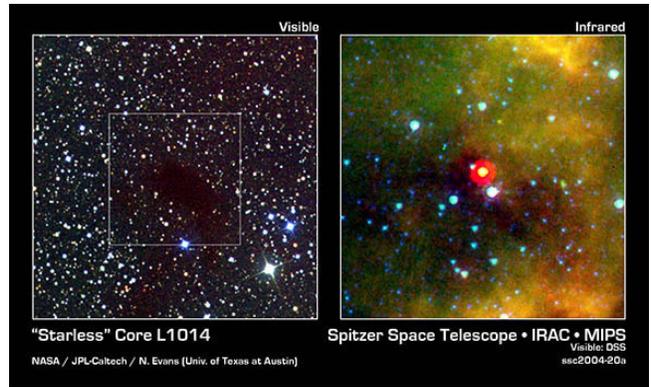


Figure 2.14: Left: The visible light image is from the Digital Sky Survey and is a B-, R-, and I-band composite image (wavelengths ranging from 0.4 to 0.7  $\mu\text{m}$ ). Right: The Spitzer image is a 3.6  $\mu\text{m}$  (blue), 8.0  $\mu\text{m}$  (green) and 24.0  $\mu\text{m}$  (red) composite image. The bright yellow object at the center of the image is the object detected in the "starless core". The red ring surrounding the object is an artifact of the reduced spatial resolution of the telescope at 24  $\mu\text{m}$ .

As the dense dark cores become unstable they form protostars, which radiate away the gravitational energy from the collapse and thereby contract even further. This first phase is mostly a free fall phase, where the collapse rate is limited by the rate at which the energy can be radiated away

$$t_{\text{ff}} \sim \sqrt{\frac{1}{G\rho}}. \quad (2.11)$$

For a  $1 M_{\odot}$  protostar, the free-fall collapse ends when its radius is  $\sim 500 R_{\odot}$ . This is also called the Hayashi-phase. In the Hertzsprung-Russell diagram, this corresponds to an evolutionary track that runs almost vertically, because the luminosity decreases without changing the temperature of the protostar much.

### 2.3.4 Jets & Outflows

In this early phase of pre-main sequence stellar evolution, jets and outflows are a typical phenomenon. Fig. 2.15 shows a few observations of jets taken with the Hubble Space Telescope. These jets are generally collimated by magnetic fields and transport away excess angular momentum, thus spinning down the protostar. They show strong emission lines of warm molecular gas ( $\sim 10^3$  K) such as  $H_2$  and CO, but also many forbidden atomic transitions.

## 2.4 How did the planets form?

The process of star formation is accompanied by the formation of a rotating flat disk, the protoplanetary disk or 'proplyds'. Such disks are now frequently observed in star forming regions such as Orion and Taurus (see Fig. 9.12). During the collapse phase, angular momentum prevents the collapse along the equator, so that the originally spherical core collapses into a flat pancake shape object with a central core, the protostar. Our current understanding is that the planets form within these protoplanetary disks and there is now growing evidence from tentative detections of young planets in such disks (TW Hya - Setiawan 2008; Fomalhaut - Kalas 2008).

### 2.4.1 Protoplanetary Disks

Originally, a protoplanetary disk contains gas and dust with a composition very similar to the

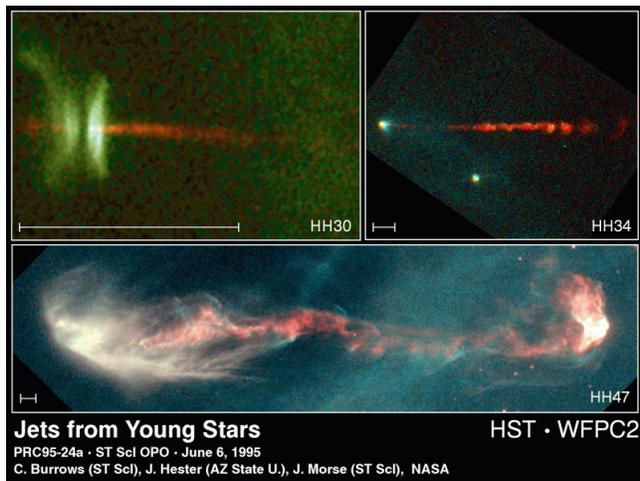


Figure 2.15: Examples of protostellar jets and outflows. These were first discovered independently by George Herbig (Lick Observatory) and Guillermo Haro (Mexican National Observatory); hence their name Herbig-Haro (HH) objects.

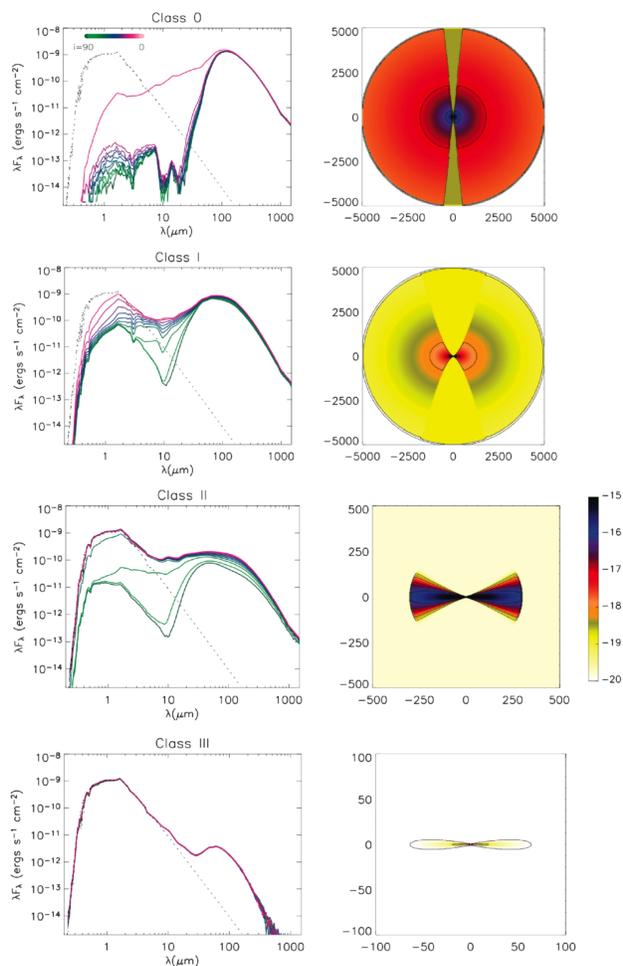


Figure 2.16: Classification of Spectral Energy Distributions of young stellar objects and disks (Whitney 2003 after Lada 1987).

parental molecular cloud. However, during its course of evolution, this material gets processed and somehow turned into larger bodies such as comets, asteroids, and planets. Since these disks are very opaque in their youngest phase, we cannot witness this process directly. However, the stellar radiation is partly reprocessed by the gas and dust in the disk and heats the matter there to typical temperatures of a few 1000 K close to the star to 10 K in the very distant regions.

What we can then observe from these protoplanetary disks is the thermal emission and also the scattered light from the dust grains in the surfaces of these disks, very much comparable to the light that we receive from stars. The stellar radiation also originates in a thin surface layer called the photosphere. Similarly, the layer of a disk that is visible to us is also often called 'disk photosphere'. This disk radiation appears in the spectral energy distribution of a young star as an excess of infrared radiation above the typical stellar photospheric level. Based on how much of the stellar luminosity is reprocessed in the disk, we can classify those objects into four classes, class 0, I, II and III (see Fig. 2.16). These four classes represent an evolutionary scenario, in which the envelope opens a cavity from class 0 to class I. Next, the envelope disappears and we are left with the naked disk (class II). The last evolutionary phase - class III - is a low mass optically thin disk (or debris disk) that generates only little IR excess radiation.

We can also observe gas emission from this disk photosphere. In fact, the first evidence for the nature of these objects came from sub-mm observations of the CO rotational transition. The resolved line profile shows a double-peaked shape that is indicative of a rotating disk (Fig. 2.17). Hence, while SEDs tell us about grain properties, disk sizes and extensions, gas lines contain additional information on disk dynamics, excitation conditions, i.e. temperatures and densities. Depending on the particular line transition that we choose to observe, we cover only a certain radial and vertical region of the disk. Near-IR lines are generally higher excitation lines (few 100 – 1000 K) and probe gas in the terrestrial planet forming zone (< few AU). Sub-mm lines originate in the much cooler (few 10 K) outer parts of the disk at distances of 100 – 1000 AU.

## 2.4.2 Debris Disks

The debris disks generally fit into the optically thin late end of the Lada classification scheme (class III). Those disks contain small micron sized dust grains which must be a result of continuous replenishment through for example collisions of larger bodies. We often think of these debris disks as young equivalents of our own Kuiper Belt. Typical ages for these systems are a few 10 – 100 Myr; hence the stars are mostly in their main-sequence evolutionary phase. The correlation between the presence of an extrasolar planetary system and a debris disk is not yet conclusive. There are a few systems that seem to have both, but current instruments are not sensitive enough to confirm or rule-out a one-to-one correlation. Fig. 2.18 shows one of the few examples, where the planet within the debris disk has been directly imaged.

## 2.4.3 Exoplanets

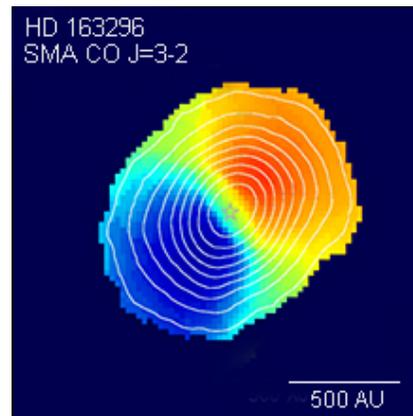


Figure 2.17: CO J=3-2 image of the disk around the star HD 163296 taken with the Submillimeter Array (SMA, credit: David Wilner). The color indicates the velocity shift of the line: blue areas of the disk are coming towards us, red areas go away from us.

Over the past 15 years, we went from one proto-type planetary system — our Solar System — to an ensemble of more than 200 planetary systems, many of which have planets on eccentric orbits and/or orbital periods of less than 1 year (Zucker & Mazeh 2001). Thus it might be expected that the variety of planetary systems is — to some extent — reflected in the much earlier stage, namely the protoplanetary disk phase. Fig. 2.6 summarizes the timeline and properties of exoplanet discoveries.

Evidence for differences in planetary system statistics between solar-type stars and lower mass stars is just starting to appear (Marcy et al. 2008). The most successful method to find extrasolar planets to date is the radial velocity method. Because the star and planet both revolve around the center of mass, the star itself makes a small movement on the sky. This motion can be detected in its spectrum through a small shift in the rest wavelength of the absorption lines (Doppler effect). From subsequent measurements of the spectrum over many years, the period and mass of the planet can be extracted. Since we do not know a priori the inclination of the system (angle  $i$ ,  $90^\circ$  is pole-on), we obtain only a lower limit of the planetary mass, namely  $m \sin i$ . Only in transiting systems, where the planet passes directly in front of the star, we can obtain additional information on the inclination and the size of the planet. This method is pursued by the two recent space missions Corot (France/ESA) and Kepler (NASA) and led to the discovery of the first terrestrial planet Corot-7b (with a mean density close to that of planet Earth).

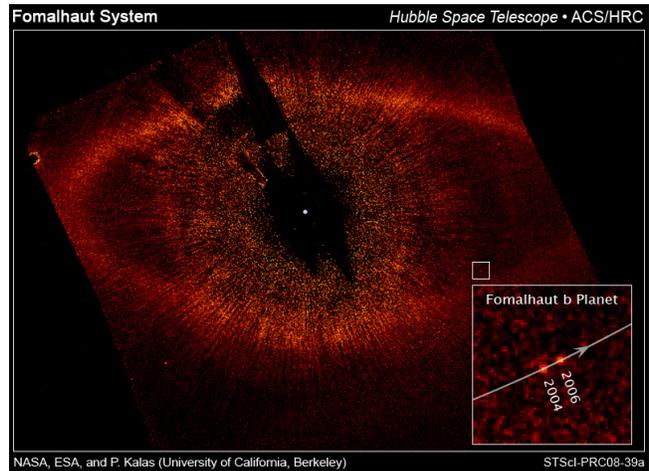


Figure 2.18: HST scattered light images of the debris disk around the A-type star Fomalhaut. Images taken at two different epochs, 2004 and 2006, show a moving object at the inner edge of the debris belt.

## Chapter 3

# Molecular clouds

This chapter lays the foundation for understanding the physics and chemistry of the material from which stars eventually form. Molecular clouds are a dense component of the interstellar medium and they comprise about 50% of the total interstellar gas of a galaxy. The following sections summarize first some observational aspects of molecular clouds, such as their sizes, appearance at various wavelength and in various tracers, and their constituents. We then move to learning some fundamental molecular physics of simple molecules such as  $\text{H}_2$  and  $\text{CO}$  and how to understand the origin of molecular lines. The third section focuses on the various heating and cooling processes in a molecular cloud that govern the energy balance of the dust and gas. This topic will be picked up in a later chapter on irradiated disks, where we will encounter a similar set of heating and cooling processes. The last section gives a brief introduction into astrochemistry, introducing the basic principles of chemistry in space. We discuss there the chemical composition of molecular clouds. Also the astrochemistry will come back at a later stage when we talk about the chemistry of protoplanetary disks.



Figure 3.1: The Eagle Nebula at a distance of 1800 pc viewed with HST/ACS in 2004. The structure is about 9.5 ly ( $\sim 3$  pc) high.

### 3.1 Observational properties of molecular clouds

A cloud is a region of the interstellar medium with densities larger than  $10 - 30 \text{ cm}^{-3}$ . Diffuse clouds can have somewhat lower densities. Clouds move through the galaxy as entities. However, they are generally not spherical, but rather irregular with filamentary structures (Fig. 3.1). Depending on their extinction,  $A_V$ , these clouds are more or less transparent to the interstellar UV radiation coming from the hot O and B stars in our galaxy. At an extinction larger than  $A_V \sim 0.5$ , most of the hydrogen is in molecular form. Table 3.1 lists the typical properties of molecular clouds found from studies within our own galaxy. We will focus in the following on some of the examples listed in the last column to illustrate the appearance and properties of these clouds.

Table 3.1: Physical properties of molecular clouds (from Stahler &amp; Palla 2004)

Cloud type	$A_V$ [mag]	$n_{\text{tot}}$ [ $\text{cm}^{-3}$ ]	$L$ [pc]	$T$ [K]	$M$ [ $M_{\odot}$ ]	Example
Diffuse	1	500	3	50	50	$\zeta$ Ophiuchi
Giant Molecular Cloud	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

### 3.1.1 Dust and gas in clouds

As the interstellar medium, molecular clouds also consist of gas and dust with a typical mass ratio of 100. The dust in these regions is typically cold and can thus be detected through its thermal emission at submm and mm wavelength. At those wavelength, the extinction is very low and the emission can be optically thin. If that is the case, the observations trace the total dust mass residing in the molecular cloud. However, for accurate mass determinations, the dust emissivity needs to be known and the typical uncertainty can be up to a factor 3.

Most of the gas is in the form of molecular hydrogen. However, as we will see later in this chapter,  $\text{H}_2$  as a symmetric molecule does not possess a permanent dipole moment. Hence, the infrared transitions are forbidden and intrinsically weak which makes it difficult to observe  $\text{H}_2$ . The second most abundant molecule is CO and that one has a rich spectrum of rotational lines that are easily excited at the low temperatures of these clouds. The ground rotational transition for example has an excitation temperature of 5 K and a wavelength of 2.6 mm. The rotational lines have very high Einstein A coefficients and low critical densities ( $n_{\text{cr}} < 10^4 \text{ cm}^{-3}$ , the critical density is the one, where collisional rates populating and depopulating a particular energy level dominate over radiative rates), which means that they are generally in LTE, but also that they are very optically thick. A way out is to use CO isotopes such as  $^{13}\text{C}^{16}\text{O}$  and  $^{12}\text{C}^{18}\text{O}$ , which are a factor 60 and 500 lower in abundance than the main isotope  $^{12}\text{C}^{16}\text{O}$ . All these lines can be easily observed with ground-based radio telescopes. Gas line observations also have the advantage that they carry kinematic information about the velocity field in the cloud (line profile and width of the line). In addition, magnetic field strength can be measured through the Zeeman splitting of molecular lines.

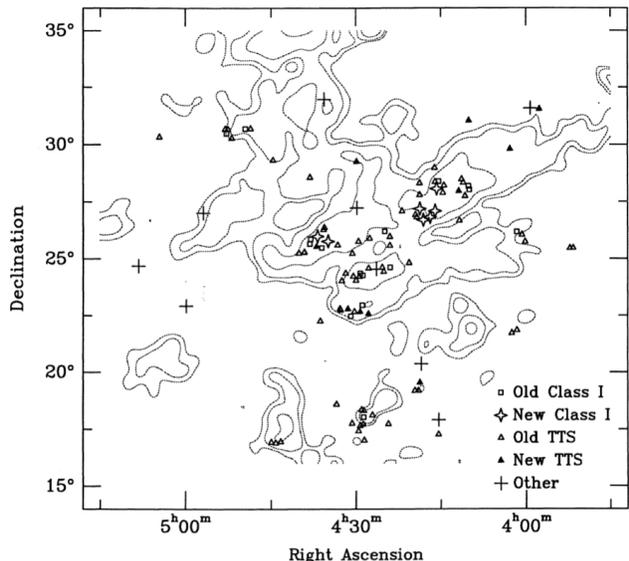


Figure 3.2: Sky map for Taurus-Auriga region. CO column densities from Ungerechts and Thaddeus (1987) are shown as dotted lines for flux densities of 3, 5, 10, 25 and 40  $\text{K km s}^{-1}$ . Open triangles indicate the positions of optically visible T Tauri stars, while core sources from the Beichman et al. (1986) and Myers et al. (1987) surveys are plotted as open squares. The new IRAS sources discussed in this paper are indicated by stars (new class I sources), filled triangles (new TTS), or large crosses (galaxies and apparent field stars). The position of some objects have been offset for clarity (caption and figure from Kenyon et al. 1990).

### 3.1.2 Low mass star forming clouds

Fig. 3.2 shows a map of the CO molecular emission ( $J=1-0$ ) in the Taurus-Auriga molecular cloud ( $d \sim 140$  pc). Taurus-Auriga is a low-mass star forming region, i.e there are no massive O and B stars forming that irradiate the molecular gas. The total gas mass in this region is  $\sim 10^4 M_{\odot}$ . The figure nicely shows how the young stars are scattered throughout the entire cloud complex. The sensitivity of this IRAS (Infrared Astronomical Satellite) survey allowed the detection of young stars brighter than  $0.5 L_{\odot}$ . More recently, Spitzer observations provided much deeper observations, hence a larger sensitivity for the detection of young stellar objects.

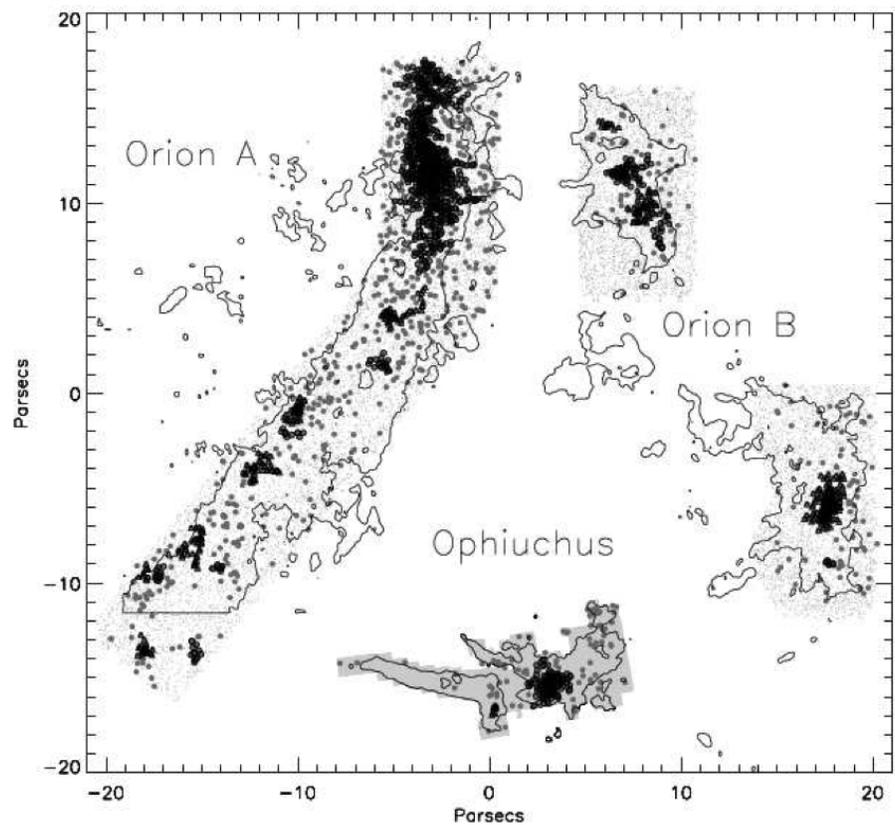


Figure 3.3: The spatial distribution of all Spitzer identified infrared excess sources from the combined IRAC and 2MASS photometry of Orion A (left), Orion B (right) and Ophiuchus (bottom center). The contours outline the Bell Labs  $^{13}\text{CO}$  maps for the Orion A and B clouds (Bally et al., 1987; Miesch and Bally, 1994), and an  $A_V$  map of Ophiuchus (Huard, 2006). The large grey dots are the sources with infrared excesses (caption and figure from Allen et al. 2007).

### 3.1.3 High mass star forming clouds

Fig. 3.3 shows the distribution of infrared-excess (likely YSOs) sources in the Orion star forming cloud ( $d \sim 400$  pc). This is a Giant Molecular Cloud (GMC), where high mass stars (O and B-type stars) have been forming within the last  $10^6$  yr. These hot stars emit strongly in the UV spectral range, thereby heating and ionizing the surrounding gas. Over time, they created the Orion nebula (M42). The nebula has recently also been surveyed by the Hubble Space Telescope to provide a legacy dataset for improving our

understanding of star formation theory (Robberto et al. 2009). The small grey dots in Fig. 3.3 show all the detections in the Spitzer 3.6 and 4.5  $\mu\text{m}$  bands with magnitudes brighter than 15 and the large grey dots are those that show an infrared excess. The black circles/triangles are sources that have been associated with a cluster (the two symbols are alternated so that neighboring clusters can be differentiated).

### 3.1.4 Cores and clumps

As the previous sections documented, molecular clouds are highly fragmented and consists of smaller entities, the clumps and cores. Clumps are generally the larger substructures in molecular clouds, having typical sizes of 1 pc. They are the precursors of stellar clusters. Cores are about a factor 10 smaller and have even higher densities. They are assumed to be gravitationally bound and will collapse to form individual stars or binaries.

The clumpy structure of molecular clouds is shown to partially originate from supersonic motions in the ISM. In some cases, the clumps seem to be rather confined by the pressure from the surrounding medium and thus not in virial equilibrium. In that sense, not every clump will eventually form stars.

The clump masses within molecular clouds have been studied via CO isotope observations. Above a certain minimum mass, the clump mass spectrum follows a power law

$$\begin{aligned} \frac{dN}{dM} &\propto M^{-1.6\dots 1.8} \\ \Rightarrow \frac{MdN}{dM} &\propto M^{-0.6\dots 0.8} \end{aligned} \quad (3.1)$$

This means that most of the mass is actually in massive clumps. Fig. 3.4 shows such a clump mass spectrum for the Rosette molecular cloud, where a minimum mass of  $\sim 30 M_{\odot}$  has been adapted. Combining this with extinction measurements, it can be shown that even though the clumps seem to fill the plane of the molecular cloud on the sky, they do not fill the volume of the cloud.

Most clumps however do not form stars. Hence, the mass spectrum of cores is more relevant in understanding the mass distribution of newly forming stars. While the star formation efficiency of an entire molecular cloud is only of the order of a few %, cores have a much higher efficiency of  $\sim 30 - 40$  %. Fig. 3.6 shows a deep 1.3 mm continuum map of  $\rho$  Ophiuchi ( $d = 140$  pc) with a resolution of 0.01 pc ( $\sim 2000$  AU). This large survey produced a core mass spectrum that can be splitted into two power law regimes, one for

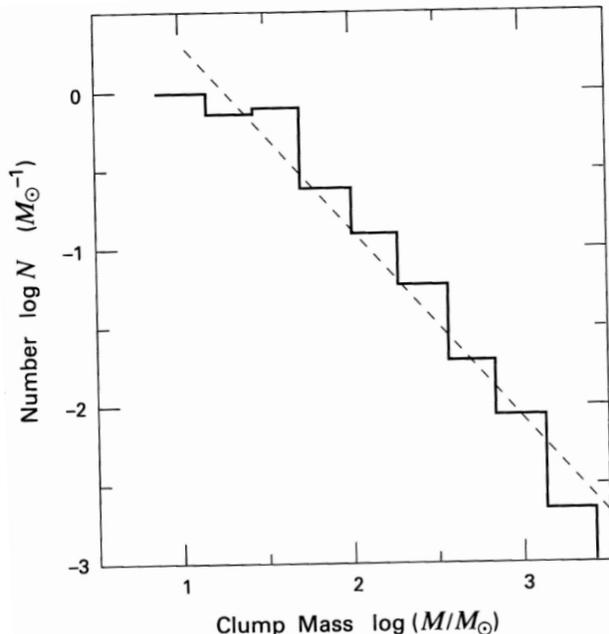


Figure 3.4: Distribution of clump masses in the Rosette Molecular Cloud (caption and figure from Stahler & Palla 2004).

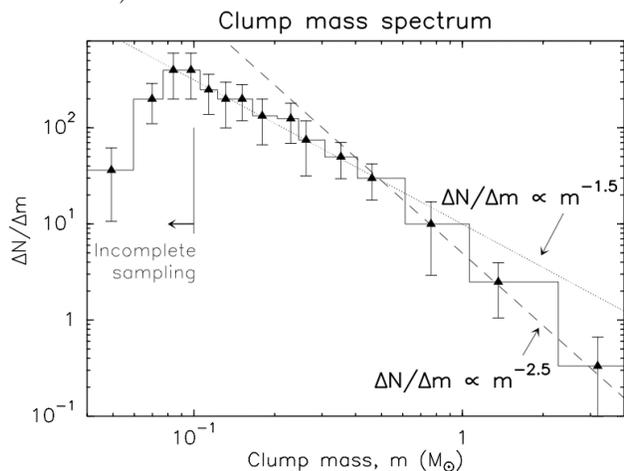


Figure 3.5: Frequency distribution of masses for 60 small-scale clumps extracted from the mosaic of Fig. 3.6 (solid line). The dotted and long-dashed lines show power laws of the form  $\Delta N/\Delta m \propto m^{-1.5}$  and  $\Delta N/\Delta m \propto m^{-2.5}$ , respectively. The error bars correspond to  $\sqrt{N}$  counting statistics (figure and caption from Motte et al. 1998).

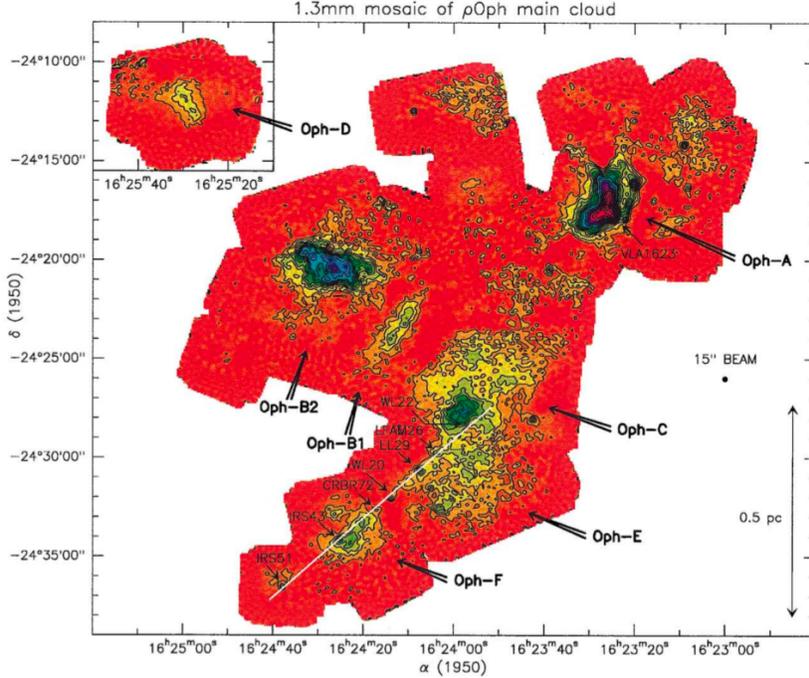


Figure 3.6: Millimeter continuum mosaic of the  $\rho$  Oph main cloud including the dense cores Oph-A, Oph-B1, Oph-B2, Oph-C, Oph-D, Oph-E, and Oph-F. The data were smoothed to an effective angular resolution of  $15''$  (HPBW). Contour levels go from 5 to 40 MJy/sr with steps of 5 MJy/sr, from 50 to 80 MJy/sr by 10 MJy/sr, 100 MJy/sr and from 120 to 280 MJy/sr with steps of 40 MJy/sr ( $1 \text{ MJy/sr} \simeq 6 \text{ mJy}/15''\text{-beam}$ ). The mean rms noise level is  $\sim 1.2 \text{ MJy/sr}$ . Note the remarkable linear chain of clumps and embedded YSOs in the southern part of the  $\text{C}^{18}\text{O}$  ridge (emphasized by a white straight line). Figure and caption from Motte et al. (1998).

$M < 0.5 M_{\odot}$  and one for  $M > 0.5 M_{\odot}$ . They have power law indices of  $-1.6$  and  $-2.1 \dots -2.5$ , respectively (see Fig. 3.5).

### 3.1.5 The Initial Mass Function

As part of our understanding of star formation, it is important to understand the mass distribution of entire groups of newly forming stars. We already know that massive hot O and B stars are much rarer than solar-type stars and that most stars in the Milky Way are in fact stars with a mass much lower than that of our Sun. We can study this quantitatively by looking for example at stars in stellar clusters. The core mass spectrum found in the previous section resembles largely the initial mass function found for newly born stars.

## 3.2 Molecular Cloud Stability

We will now try to understand the basic stability energy content and stability of molecular clouds.

### 3.2.1 The Jeans mass

Lets assume that a molecular cloud is a static and homogeneous medium that satisfies the equations of continuity, Euler's equation, and Poisson's equation

$$\frac{\partial \rho_0}{\partial t} + \rho_0 \nabla \cdot \mathbf{v}_0 = 0 \quad (3.2)$$

$$\frac{\partial (\rho_0 \mathbf{v}_0)}{\partial t} + \nabla \cdot (\rho_0 \mathbf{v}_0 \mathbf{v}_0 + P_0 + \rho_0 \nabla \Phi_0) = 0 \quad (3.3)$$

$$\nabla^2 \Phi_0 = 4\pi G \rho_0 \quad (3.4)$$

We also assume that the ideal gas law holds:  $P_0 = \rho_0 hT / (\mu m_p) = \rho_0 c_s^2$ . We assume that this medium is initially at rest, i.e.  $\mathbf{v}_0 = 0$ , and that it is infinitely extended, i.e.  $\Phi_0 = 0$ . We consider a small perturbation from the initial state

$$\rho = \rho_0 + \rho_1 \quad \mathbf{v} = \mathbf{v}_1 \quad \Phi = \Phi_1 \quad (3.5)$$

We now insert the perturbed quantities into eqs (3.2)–(4.14) and make use of the fact that the unperturbed quantities are solutions of the equations, and we neglect terms that are quadratic in the perturbations. After some algebra we find

$$\frac{\partial \rho_1}{\partial t} + \rho_0 \nabla \cdot \mathbf{v}_1 = 0 \quad (3.6)$$

$$\frac{\partial \mathbf{v}_1}{\partial t} = -\nabla \Phi_1 - \frac{\nabla P_1}{\rho_0} \quad (3.7)$$

$$\nabla^2 \Phi_1 = 4\pi G \rho_1 \quad (3.8)$$

We now take the time derivative of eq. (3.6), eliminate  $\partial \mathbf{v}_1 / \partial t$  using eq. (3.7). Then we insert eq. (3.8) and eliminate the pressure using the ideal gas law. In this way we arrive at a wave equation for the density perturbation

$$\frac{\partial^2 \rho_1}{\partial t^2} - 4\pi G \rho_0 \rho_1 + \frac{kT}{\mu m_p} \nabla^2 \rho_1 = 0 \quad (3.9)$$

which we attempt to solve using a plane wave *Ansatz*  $\rho_1 = \exp(i(\frac{2\pi x}{\lambda} - \omega t))$ . Inserting this *Ansatz* into eq.(3.9) we find the following dispersion relation:

$$\omega^2 = \left(\frac{2\pi}{\lambda}\right)^2 \left(\frac{kT}{\mu m_p}\right) - 4\pi G \rho_0 \quad (3.10)$$

This result indicates that perturbations on length scales larger than the *Jeans Length*

$$\lambda_J = \left(\frac{\pi kT}{\mu m_p G \rho_0}\right)^{1/2} \quad (3.11)$$

will grow exponentially, while smaller perturbations will be damped. Since this instability occurs in all three spatial coordinates, a sphere with diameter  $\lambda_J$  defines the maximum stable mass, the *Jeans mass*

$$M_J = \frac{\pi}{6} \rho_0 \left(\frac{\pi kT}{\mu m_p G \rho_0}\right)^{3/2} \quad (3.12)$$

### 3.2.2 The problem of star formation efficiency

The galaxy should be wildly gravitationally unstable, because all molecular clouds are above the *Jeans mass* limit derived in the previous section. If we consider the free fall time scale (we will derive it later)  $t_{\text{ff}} = \sqrt{\frac{3\pi}{32G\rho}}$

for the interstellar medium with an average density of 17 hydrogen atoms per  $\text{cm}^3$ , then we find a free fall time of  $8 \times 10^6$  years. The entire molecular gas in the milkyway (about  $2 \times 10^{10} M_\odot$ ) should be converted into stars in a free fall time scale. The expected star formation rate is therefore  $\sim 250 M_\odot/\text{yr}$ , while the observed star formation rate is only 3 solar masses per year. *Something is slowing down star formation!*

### 3.2.3 The Virial theorem for clouds

A general form of the Virial theorem can be derived from the momentum equation of magneto-hydrodynamics

$$\rho \frac{D\mathbf{v}}{Dt} = -\nabla P - \rho \nabla \Phi + \frac{1}{c} \mathbf{j} \times \mathbf{B} \quad (3.13)$$

with  $D\mathbf{v}/Dt := (\partial v/\partial t)_x + \nabla \cdot (\mathbf{v}\mathbf{v})$ . Using Maxwells equation  $\nabla \times \mathbf{B} = 4\pi \mathbf{j}/c$ , we get

$$\rho \frac{D\mathbf{v}}{Dt} = -\nabla P - \rho \nabla \Phi + \frac{1}{4\pi} (\mathbf{B} \times \nabla) \mathbf{B} - \frac{1}{8\pi} \nabla |\mathbf{B}^2| \quad (3.14)$$

This equation describes the local behavior of the fluid. Multiplying by  $\mathbf{r}$  (scalar product) and integrating over the entire volume, also using the continuity equation leads to

$$\frac{1}{2} \frac{\partial^2 I}{\partial t^2} = 2\mathcal{T} + 2\mathcal{U} + \mathcal{W} + \mathcal{M} \quad \text{with} \quad (3.15)$$

$$I = \int \rho |r^2| d^3x \quad \text{moment of inertia} \quad (3.16)$$

$$\mathcal{T} = \frac{1}{2} \int \rho |v^2| d^3x \quad \text{kinetic energy} \quad (3.17)$$

$$\mathcal{U} = \frac{3}{2} \int P d^3x \quad \text{thermal energy} \quad (3.18)$$

$$\mathcal{W} = \frac{1}{2} \int \rho \Phi d^3x \quad \text{gravitational energy} \quad (3.19)$$

$$\mathcal{M} = \frac{1}{8\pi} \int B^2 d^3x \quad \text{magnetic energy} \quad (3.20)$$

Long term stability means that the overall mass distribution and size does not change, i.e. that the moment of inertia does not change with time. As the LHS of eq. (3.15) vanishes, we find  $2\mathcal{T} + 2\mathcal{U} + \mathcal{W} + \mathcal{M} = 0$ . We can now use observed values for temperature  $T$ , kinetic (turbulent) motions  $V$ , cloud mass  $M$  and radius  $R$ , and magnetic fields  $B$ , and compare the different energies:

$$\frac{\mathcal{U}}{|\mathcal{W}|} \approx \frac{MRT}{\mu m_p} \left( \frac{R}{GM^2} \right) = 0.003 \quad \left( \frac{M}{10^5 M_\odot} \right)^{-1} \left( \frac{R}{25\text{pc}} \right) \left( \frac{T}{15\text{K}} \right) \quad (3.21)$$

$$\frac{\mathcal{M}}{|\mathcal{W}|} = \frac{B^2 R^3}{6\pi} \left( \frac{R}{GM^2} \right) = 0.3 \quad \left( \frac{B}{20\mu\text{G}} \right)^2 \left( \frac{R}{25\text{pc}} \right)^4 \left( \frac{M}{10^5 M_\odot} \right)^{-2} \quad (3.22)$$

$$\frac{\mathcal{T}}{|\mathcal{W}|} \approx \frac{1}{2} M \Delta V^2 \left( \frac{R}{GM^2} \right) = 0.5 \quad \left( \frac{\Delta V}{4\text{km/s}} \right)^2 \left( \frac{M}{10^5 M_\odot} \right)^{-1} \left( \frac{R}{25\text{pc}} \right) \quad (3.23)$$

So while thermal energies are much to low to counter gravity, both magnetic fields and kinetic (turbulent) motions contain energies that are of the same order as the gravitational energy. We conclude that magnetic fields and turbulence are keeping molecular clouds from collapsing in their entirety.

## Chapter 4

# Cloud equilibrium and Stability

After discussing the basic structure and stability of molecular clouds in the previous chapter, we now turn to small cloud cores and study the density structure of these cores and consider their stability, and their conditions that must occur for collapse to proceed.

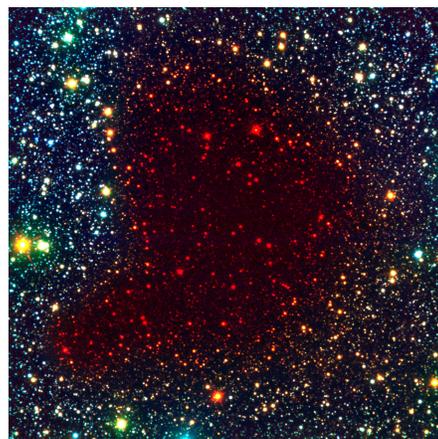
An example for a self-gravitating core is Barnard 68 at a distance of 130 pc. It has been observed in all wavelength ranges from optical to mm (e.g. Kirk et al. 2005). The size of the core is  $0.074 \times 0.047$  pc (FWHM), its temperatures as derived from the spectral energy distribution is 13 K and its extinction is 11 mag. The volume density is estimated to be  $5 \times 10^4 \text{ cm}^{-3}$  and it contains  $0.2 M_{\odot}$  of mass within the FWHM (derived from submm fluxes). Alves et al. (2001) found the density distribution of this core supercritical and hence postulate that it is at the verge of collapse and should eventually form a low mass star.

### 4.1 Hydrostatic self-gravitating spheres

In the following, we discuss the density structure of such cores — so-called Bonnor-Ebert spheres — and the role of magnetic fields and turbulence in their stability. Even though Fig. 4 shows that the shape of cores is far from spherical, we will simplify our studies by concentrating on spherical cores.

#### 4.1.1 Density structure

As the most simple and idealised case, we now turn to clouds which are isothermal (i.e. where the temperature is constant throughout the cloud), and where gravity is only balanced by pressure forces.



*B68*

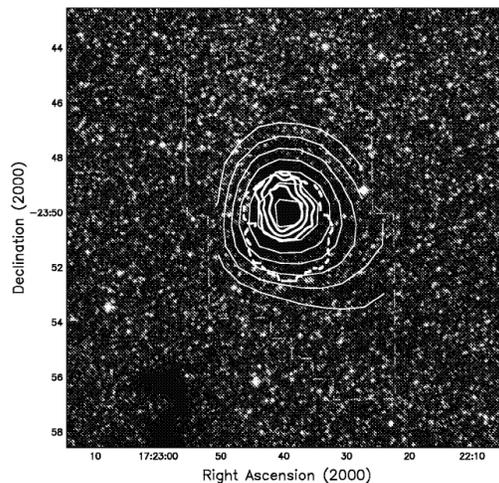


Figure 4.1: Top: ESO false color composite image of the core Barnard 68 (B, I from VLT/FORS1 and K from NTT/Sofi). Bottom: Composite SCUBA  $850 \mu\text{m}$  (thick white contours) and ISOPHOT  $200 \mu\text{m}$  (thin grey contours) map of the core Barnard 68 superposed on POSS optical (grey-scale) images. The size of the image is 16 arcmin square (figure and caption from Kirk et al. 2005)

The assumption of an isothermal state simplifies the calculations because in this case pressure gradients are solely due to density gradients, and a simple, scale-free formulation can be found. It is clear that this is an over-simplification, and that magnetic support and magneto-hydrodynamical (MHD) waves play a major role at least in larger cloud complexes. However, the basic concepts of stability and collapse can be captured here.

We start by writing down the basic equations for hydrostatic equilibrium for such a system. These are one equation for the pressure structure, an equation of state (ideal gas), and an equation for the gravitational potential:

$$\frac{dP(r)}{dr} = -\frac{GM(r)\rho(r)}{r^2} \quad (4.1)$$

$$P = \frac{\rho kT}{\mu \mp} = \rho c_s^2 \quad (4.2)$$

$$\frac{dM(r)}{dr} = 4\pi r^2 \rho(r) \quad (4.3)$$

In order to integrate these equations, one has to specify boundary conditions. We can specify  $\rho(0)$  and directly integrate the system, knowing that  $M(0) = 0$ . The integration can be terminated at a location where a specific outer pressure  $P_0$  is reached and in this way we can find the mass of the cloud for these conditions. If on the other hand we want to produce a cloud with a specific mass, we need to try many  $\rho(0)$  to match both mass and outer pressure. However, if we want to construct a cloud with a given mass and a given external pressure, then we have to construct many models with different values for  $\rho(0)$ . Fig. 4.2 shows a number of such integrations for different  $\rho(0)$ . The figure also shows that all solutions asymptotically merge into the same powerlaw in the outer region. The asymptotic solution for infinite central density is the so-called singular isothermal sphere (SIS)

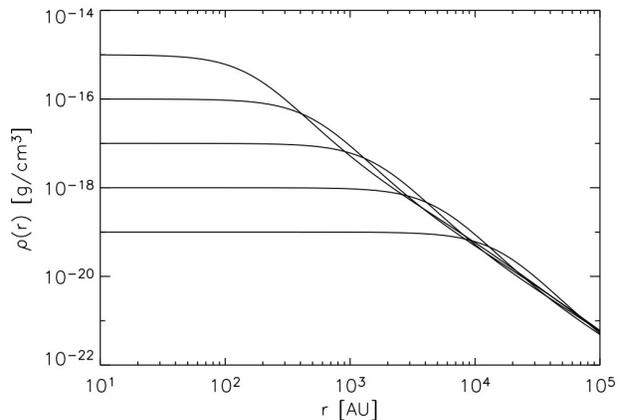


Figure 4.2: Density profiles of a hydrostatic sphere for different values of  $\rho(0)$

$$\rho(r) = \frac{c_s^2}{2\pi G r^2} \quad (4.4)$$

where the density decreases as  $r^{-2}$ . The SIS will cross our path later again, when we discuss Frank Shu's inside-out collapse solution.

In Fig. 4.2 we also see that different solutions of these equations look similar, as if they are scaled versions of each other. This indicates that it is useful to try a scale-free formulation of the problem.

#### 4.1.2 Scale-free formulation

We can use equations (4.1) and (4.2) to express the hydrostatic equilibrium in terms of the gas density alone

$$\frac{d\rho(r)}{\rho(r)dr} = \frac{d \ln \rho}{dr} = -\frac{GM(r)}{r^2 c_s^2} \quad (4.5)$$

Equation (4.5) tells us that the sum of  $(\ln \rho(r) - GM(r)/(rc_s^2))$  must be a spatial constant, so we are going to look for solutions in the form of

$$\rho(r) = \rho_c \exp(-\Psi) = \rho_c \exp\left(-\frac{GM(r)}{rc_s^2}\right) \quad (4.6)$$

where  $\rho_c$  is the central density of the cloud and where we have introduced a new dimensionless dependent variable  $\Psi$

$$\Psi = -\frac{GM(r)}{rc_s^2} \quad (4.7)$$

Comparing this definition of  $\Psi$  with Eq(4.1) leads to the relation

$$\frac{d\Psi}{dr} = \frac{GM(r)}{r^2c_s^2} = -\frac{dP(r)}{dr} \frac{1}{\rho(r)c_s^2} \quad (4.8)$$

We also introduce a dimensionless independent variable

$$\xi = \left(\frac{4\pi G\rho_c}{c_s^2}\right)^{1/2} r \quad (4.9)$$

Now we return to equation (4.3) and rewrite it as

$$\frac{1}{4\pi r^2} \frac{dM(r)}{dr} = \rho(r) \quad (4.10)$$

$$-\frac{1}{4\pi r^2} \frac{d}{dr} \left( r^2 \frac{dP(r)}{dr} \right) = \rho(r) \quad (4.11)$$

Substituting subsequently  $dP(r)/dr$  from Eq.(4.8) and then  $\xi$  we obtain

$$-\frac{1}{4\pi r^2} \frac{d}{dr} \left( r^2 \rho(r) c_s^2 \frac{d\Psi(r)}{dr} \right) = \rho(r) \quad (4.12)$$

$$-\left(\frac{4\pi G\rho_c}{c_s^2}\right) \frac{1}{4\pi \xi^2} \frac{d}{d\xi} \left( \xi^2 \frac{c_s^2}{G} \frac{d\Psi}{d\xi} \right) = \exp(-\Psi) \rho_c \quad (4.13)$$

and finally the *Lane-Emden equation*

$$\frac{1}{\xi^2} \frac{d}{d\xi} \left( \xi^2 \frac{d\Psi}{d\xi} \right) = \exp(-\Psi) \quad (4.14)$$

The fact that  $\lim_{r \rightarrow 0} M(r) = 0$  also implies that the force per unit mass

$$\frac{F(r)}{m} = \frac{GM(r)}{r^2} = c_s^2 \frac{d\Psi}{dr} \quad (4.15)$$

must be zero at the center. So we get the two necessary boundary conditions for this second-order equation:

$$\Psi(0) = 0 \quad (4.16)$$

$$\left. \frac{d\Psi}{d\xi} \right|_{\xi=0} = 0 \quad (4.17)$$

The Lane-Emden equation can then be integrated numerically to find  $\Psi$  as a function of  $\xi$  (see Fig. 4.3).

The pressure in such a cloud must fall monotonically from the center to the outside, in order to offset the weight of each added layer of material. In an isothermal cloud, this also means that the density will fall monotonically. In principle, we can follow this integration to arbitrarily large radii. However, realistic clouds will not reach zero pressure, but will eventually reach the ambient pressure  $P_0$ . For given  $c_s$ , we can immediately derive the density at the outer cloud boundary, using the ideal gas law.

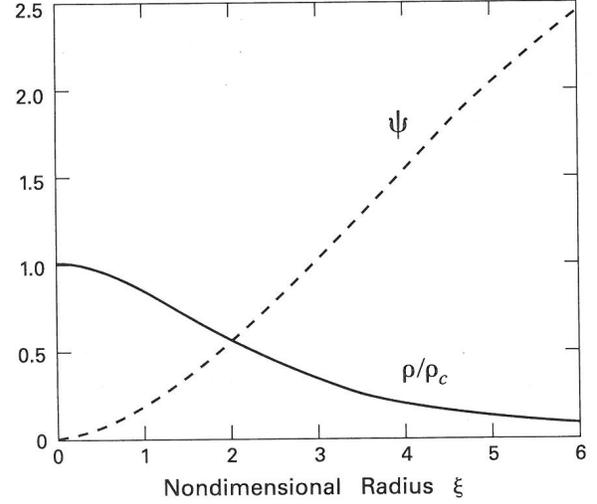


Figure 4.3: Nondimensional gravitational potential (dashed curve) and density (solid curve) in a spherical isothermal cloud, as a function of the non-dimensional radius.

Suppose now that we want to use the external pressure and the sound speed as the basic parameters of our problem. Suppose also that we specify a specific density contrast  $\rho_c/\rho_0$ . We can then read from Fig. 4.3 the corresponding dimensional radius  $\xi$  and convert it to a dimensional radius using Eq.(4.9). So we see that The solution of the Lane Emden equation is a series of models, parametrized by  $\rho_c/\rho_0$ . For  $M(r)$  we can then write

$$M(r) = \int_0^r 4\pi s^2 \rho(s) ds \quad (4.18)$$

$$= \left( \frac{1}{4\pi\rho_c} \right)^{1/2} \left( \frac{c_s^2}{G} \right)^{3/2} \int_0^{\xi_0} \exp(-\Psi) \xi^2 d\xi \quad (4.19)$$

$$= \left( \frac{1}{4\pi\rho_c} \right)^{1/2} \left( \frac{c_s^2}{G} \right)^{3/2} \xi^2 \frac{d\Psi}{d\xi} = M(\xi) \quad (4.20)$$

where we used Eq.(4.14) to replace  $\exp(-\Psi)$ .

In order to find the solution for a specific mass, we compute the mass at the outer radius  $r = r_0$  corresponding to  $\xi = \xi_0$

$$M(r_0) = \left( \frac{1}{4\pi\rho_c} \right)^{1/2} \left( \frac{c_s^2}{G} \right)^{3/2} \xi^2 \frac{d\Psi}{d\xi} \Big|_{\xi_0} = \left( \frac{c_s^2}{G} r_0 \xi_0 \right) \frac{d\Psi}{d\xi} \Big|_{\xi_0} \quad (4.21)$$

and therefore

$$r_0 = \frac{GM(r_0)}{c_s^2} \frac{1}{\xi_0 \frac{d\Psi}{d\xi} \Big|_{\xi_0}} \quad (4.22)$$

We can define a dimensionless mass  $m$  with

$$m := \frac{P_0^{1/2} G^{3/2} M}{c_s^4} \quad (4.23)$$

and insert it into Eq.(4.21). With the pressure at the outer boundary  $P_0 = \rho_0 c_s^2$ , we obtain

$$m \frac{c_s^4}{P_0^{1/2} G^{3/2}} = \left( \frac{1}{4\pi\rho_c} \right)^{1/2} \left( \frac{c_s^2}{G} \right)^{3/2} \xi^2 \frac{d\Psi}{d\xi} \Big|_{\xi_0} \quad (4.24)$$

$$m = \left( 4\pi \frac{\rho_c}{\rho_0} \right)^{-1/2} \xi_0^2 \frac{d\Psi}{d\xi} \Big|_{\xi_0} \quad (4.25)$$

$m$  is plotted as a function of the density contrast in Fig. 4.4. It first rises to a maximum value of 1.18 and then falls again. Stability analysis shows that solutions where  $dm/d(\rho_c/\rho_0) < 0$  are unstable, so that maximum  $m = 1.18$  corresponds to the maximum density contrast of a stable solution. At this point,  $\rho_c/\rho_0 = 14.1$ . Clouds with a density contrast larger than this are unstable and collapse. The mass corresponding to this critical value is called the Bonnor-Ebert mass

$$M = m \frac{c_s^4}{P_0^{1/2} G^{3/2}} \quad (4.26)$$

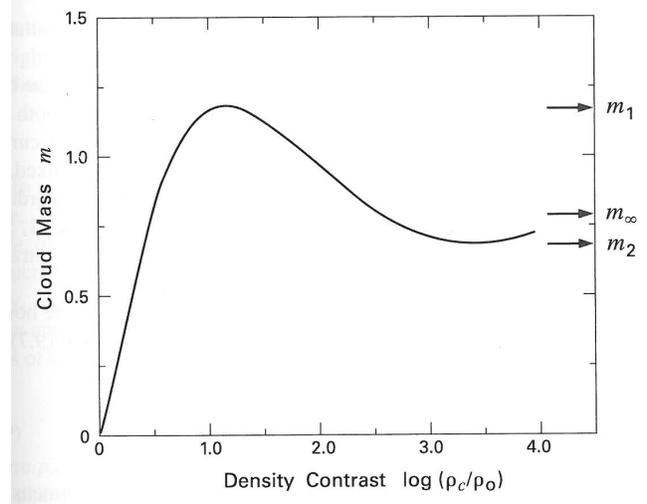


Figure 4.4: Non-dimensional mass of pressure-bounded, isothermal spheres.

Calculating the maximum  $\xi$  from Eq.(4.9)

$$\xi_{\max} = \left( \frac{4\pi G \rho_c}{c_s^2} \right)^{1/2} r_{\max} \quad (4.27)$$

for the critical density contrast yields  $\xi_{\max} = 6.5$  (Bonnor 1956). A fit to the observed density distribution of B68 yields a value  $\xi = 0.69$ , which is larger than the maximum stable value of 6.5 (Alves et al. 2001) making the core thus marginally unstable.

Ways to collapse a Bonnor-Ebert sphere include to increase the external pressure so that the density contrast in the sphere eventually exceeds this limit, or (with constant external pressure) to load mass into the sphere until the maximum mass is exceeded.

### 4.1.3 Rotational support

The velocity structure of prestellar cores can be observed with molecular line tracers such as  $\text{NH}_3$ ,  $\text{NH}_2^+$ . Fig. 4.6 shows an example of the spatial variation of the line peak  $v_{\text{LSR}}$  obtained in the  $\text{NH}_3(1,1)$  line at 1.27 cm (inversion transition). Typical velocity gradients observed in this way range from 0.3 to 10 km/s/pc (e.g. André et al. 2007). Does this mean that rotation plays a role in the gravitational stability of these cores?

To answer that question, we compare the rotational and potential energy of the core. The rotational kinetic energy of a uniform-density sphere with mass  $M$  and size  $L$  is

$$\mathcal{T} = \frac{1}{20} M L^2 \Omega^2 \quad (4.28)$$

Here,  $\Omega$  is the angular velocity of the cloud. Given the above observations, we can interpret the velocity gradient as  $\Omega$  if we assume that the cores are solid rotators with their axis perpendicular to the line of sight. The potential energy of a constant density sphere is  $6/5 GM^2/L$  (note the factor two arises from  $L$  being the diameter of the core). Hence, the ratio of the two becomes

$$\frac{\mathcal{T}}{\mathcal{W}} \approx \frac{1}{24} \frac{L^3 \Omega^2}{GM} \quad (4.29)$$

If we insert typical quantities as measured from observations, we obtain

$$\frac{\mathcal{T}}{\mathcal{W}} \approx 10^{-3} \left( \frac{\Omega}{1 \text{ km s}^{-1} \text{ pc}^{-1}} \right)^2 \left( \frac{L}{0.1 \text{ pc}} \right)^3 \left( \frac{M}{10 M_{\odot}} \right)^{-1} \quad (4.30)$$

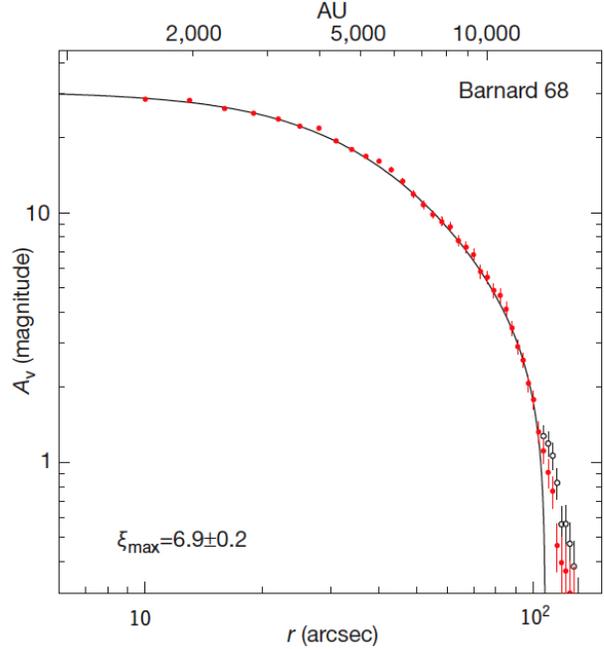


Figure 4.5: Azimuthally averaged radial dust column density profile of Barnard 68. By convention the dust column density is expressed in terms of magnitudes of visual extinction  $A_V$ . The solid line represents the best fit of a theoretical Bonnor-Ebert sphere to the data (figure and caption from Alves et al. 2001)

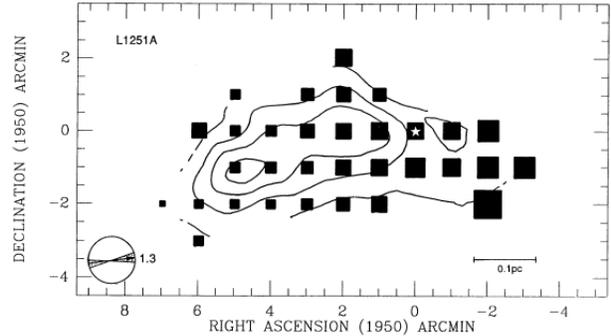


Figure 4.6: Filled squares of varying size represent the LSR velocity of the peak of the fitted  $\text{NH}_3$  line profile at each position and contours map the peak antenna temperature. The linear size of a filled square is proportional to  $v_{\text{LSR}}$  at its position (figure and caption from Goodman et al. 1993).

This clearly shows that for the typical observed cases of slowly rotating cores, the centrifugal forces will not dominate over gravitation in the initial stage of a collapse. If we denote this ratio of the two energies with  $\beta$ , numerical simulations of the collapse of an isothermal rotating cloud have shown that significant flattening only occurs above  $\beta \sim 0.1$ . If we look back again at Fig. 4.4, the qualitative shape of the curve stays, but the peak of the dimensionless mass shifts with increasing  $\beta$  to higher  $m$ . This illustrates the basic principle, namely that rotational support (centrifugal force) tends to stabilize a core against collapse.

Rotation will play a role later as we study the final stages of collapse and especially the formation of disks.

#### 4.1.4 Magnetic field support

The next best candidate supporting cores against collapse are magnetic fields. Similar as above, we can compare the magnetic energy to the gravitational energy. The magnetic energy can be derived for a constant large-scale magnetic field  $B$  as

$$\mathcal{M} = \frac{1}{8\pi} \int |B|^2 dV = \frac{1}{6} |B|^2 R^3 \quad (4.31)$$

Comparing this to the gravitational potential results in

$$\frac{\mathcal{M}}{W} = \frac{5}{18} \frac{|B|^2 R^4}{GM^2} \approx 11.7 \left( \frac{B}{20 \mu\text{G}} \right)^2 \left( \frac{R}{0.1 \text{ pc}} \right)^4 \left( \frac{M}{10 M_\odot} \right)^{-2} \quad (4.32)$$

Stahler & Palla present a table with Zeeman measurements of line-of-sight magnetic field strength  $B_{||}$  in clouds and cores (Table 4.1). A general trend seems that the magnetic field increases with volume density in the region, an indication that flux freezing — the flux lines are glued to the gas and their density increases as the physical density increases — plays a role.

Table 4.1: Zeeman measurements of magnetic fields (Table 9.1 from Stahler & Palla 2006).

Object	Type of Region	Diagnostic	$B_{  }(\mu\text{G})$
Ursa Major	Diffuse Cloud	H I	+10
L204	Dark Cloud	H I	+4
NGC2024	GMC Clump	OH	+87
B1	Dense Core	OH	-27
S106	H II Region	OH	+200
Sgr A/West	Molecular Disk	H I	-3000
W75 N	Maser	OH	+3000

When a magnetic field is present in the cloud, and when there are sufficient charges in the cloud to keep the magnetic field frozen in the gas, magnetic pressure adds to the stability of a cloud. A detailed study (see exercises) shows that the Jeans mass has to be replaced by a new critical mass

$$M_\Phi = 0.12 \frac{\Phi_M}{G^{1/2}} \approx 10^3 M_\odot \left( \frac{|B|}{30 \mu\text{G}} \right) \left( \frac{R}{2 \text{ pc}} \right)^2 \quad (4.33)$$

where  $\Phi_M$  is the magnetic flux. As we see, this mass is (other than the Jeans mass) independent of the density. So if the magnetic flux in a cloud is sufficient to keep a cloud from collapsing, than no amount of external compression will make the cloud collapse. If we turn again back to Fig. 4.4, we note once more that a magnetic field will increase the maximum  $m$  value that indicates the transition from stable to unstable solutions.

What is instead required is additional accretion of material onto the core to bring the core mass above the critical mass. Once that happens, magnetic pressure is not able to counterbalance the gravitational collapse anymore.

### 4.1.5 Ambipolar diffusion

In the presence of magnetic fields, charged species (ions and electrons) gyrate around the field lines while neutrals remain undisturbed. This creates relative drift velocities between charged and neutral species leading to frequent collisions that generate a magnetic pressure. If the ionization degree however is low, neutrals can drift efficiently across the field lines without feeling a strong pressure from the ions. Hence, mass can be redistributed with respect to the magnetic field to a degree that  $M > M_\Phi$ : the cloud can collapse. Fig. 4.7 illustrates in a sketch this effect of ambipolar diffusion in a spherical cloud.

The timescale for this ambipolar diffusion can be calculated from the typical length scale and the ionization degree of the gas  $\chi$ . In dense cores ( $n \gg 10^4 \text{ cm}^{-3}$ ), the ionization degree is largely determined by cosmic rays and can be calculated as

$$\chi = 5 \cdot 10^{-8} \left( \frac{n}{10^5 \text{ cm}^{-3}} \right)^{-1/2} \quad (4.34)$$

In very dense clouds, the ionization degree stays constant.

The timescale for neutrals to move a distance  $R$  is then given by (for a detailed derivation see Tielens 2005)

$$\tau_{\text{AD}} \approx 25 \text{ Myrs} \left( \frac{B}{3 \mu\text{G}} \right)^{-2} \left( \frac{n_{\text{H}}}{10^2 \text{ cm}^{-3}} \right)^2 \left( \frac{R}{1 \text{ pc}} \right)^2 \left( \frac{\chi}{10^{-6}} \right) \quad (4.35)$$

Filling in typical values for dense cores, we see that this timescale becomes larger than the free-fall timescale. Hence, it is difficult to argue that ambipolar diffusion is the recipe behind star formation.

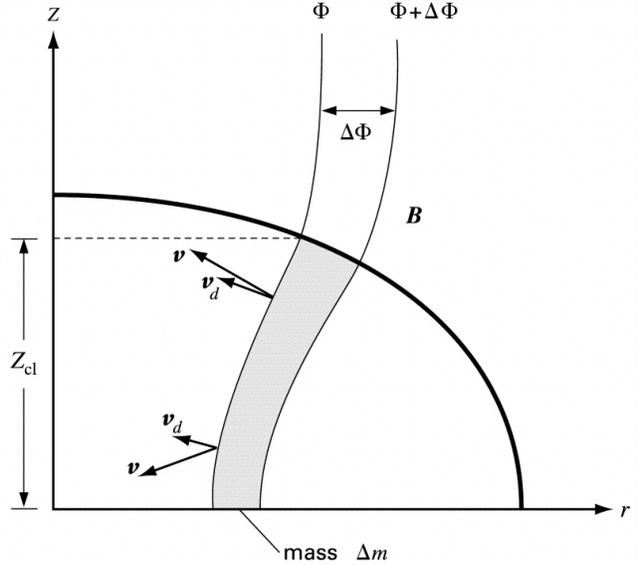


Figure 4.7: Physics of ambipolar diffusion. A portion of the cloud, of mass  $\Delta m$ , is enclosed between two flux tubes, as shown. Over time, the neutral particles drift inward relative to the field with a velocity  $v_d$ . This drift velocity differs from  $v$ , that of the neutrals in an inertial reference frame (figure and caption from Curry & Stahler 2001).

# Chapter 5

## Collapse of clouds

The challenge of star formation lies in the vast change of density and temperature that distinguishes molecular clouds — the matter from which the stars form — and the star itself. To form a star, we need to achieve an increase in density by 20 orders of magnitude and in temperature by 6 orders of magnitude. This is only possible through the self-gravitational collapse of matter.

### 5.1 Spectral Energy Distributions (SEDs)

At this point, we introduce the concept of spectral energy distributions (SEDs). Star formation and especially the earliest collapse phases can often only be studied photometrically or at very low spectral resolution ( $R < 10$ ). Detailed line profiles and/or velocity maps are often not available and hence continuum radiation provides the only clue to the nature of the young protostellar phases. Observing the star forming regions in multiple photometric bands ranging from the ultraviolet (U-band) to infrared and even sub-mm wavelengths allows us to reconstruct the total luminosity of dense cores, young stellar objects and protoplanetary disks.

#### 5.1.1 Lada classification

The shape of this energy distribution is closely related to the physical structure of the various objects. Embedded cores for example show a single-peak energy distribution that can be approximated by a single temperature black-body (Fig. 5.1). This is due to the optically thick nature of these objects where all received radiation has been re-processed by the gas and dust of the envelope. A star+disk system in later evolutionary phases displays a clear double peak

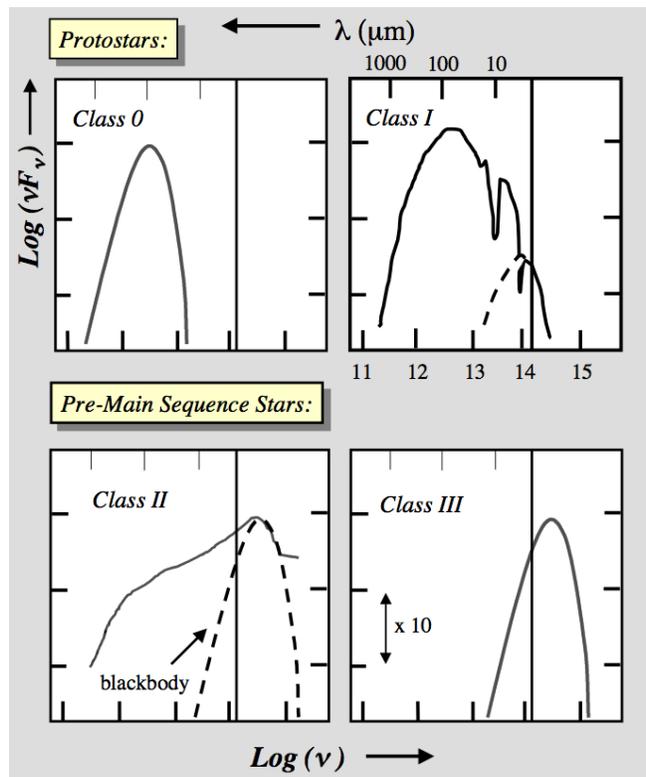


Figure 5.1: The empirical classification scheme for YSO spectral energy distributions. A vertical line appears at a wavelength of  $2.2 \mu\text{m}$  for fiducial reference in each panel. Class 0 and Class III sources have distributions whose widths are similar to single temperature black-body functions. Class II and Class I sources display infrared excess which produces energy distributions which are broader than a single blackbody function (from Lada 1999).

Table 5.1: Table with flux conversions.

To → From ↓	$S_\nu$ Jy	$F_\nu$ erg/cm <sup>2</sup> /s/Hz	$F_\lambda$ erg/cm <sup>2</sup> /s/Å	$f_E$ erg/cm <sup>2</sup> /s/eV	$f_\lambda$ erg/cm <sup>2</sup> /s/Å
$S_\nu$	$\mathbf{S}_\nu$	$10^{-23} S_n u$	$3.00 \times 10^{-5} S_n u / \lambda^2$	$1.51 \times 10^3 S_\nu / E$	$1.51 \times 10^3 S_\nu / \lambda$
$F_\nu$	$10^{23} F_\nu$	$\mathbf{F}_\nu$	$3.00 \times 10^{18} F_\nu / \lambda^2$	$1.51 \times 10^{26} F_\nu / E$	$1.51 \times 10^{26} F_\nu / \lambda$
$F_\lambda$	$3.34 \times 10^4 \lambda^2 F_\lambda$	$3.34 \times 10^{-19} \lambda^2 F_\lambda$	$\mathbf{F}_\lambda$	$4.06 \times 10^6 \lambda^3 F_\nu$	$5.03 \times 10^7 \lambda F_\lambda$
$f_E$	$6.63 \times 10^{-4} E f_E$	$6.63 \times 10^{-27} E f_E$	$1.29 \times 10^{-10} E^3 f_E$	$\mathbf{f}_E$	$8.07 \times 10^{-2} E^2 f_E$
$f_\lambda$	$6.63 \times 10^4 \lambda F_\lambda$	$6.63 \times 10^{-27} \lambda F_\lambda$	$1.99 \times 10^{-8} f_\lambda / \lambda$	$8.07 \times 10^{-2} \lambda^2 f_\lambda$	$\mathbf{f}_\lambda$

structure, with the stellar black-body peaking at optical wavelength and the cooler protoplanetary disk at infrared wavelengths (see also Fig. 2.16 in chapter 2). The stellar emission is still partly re-absorbed and scattered by the surrounding circumstellar material (disk and possibly envelope), causing veiling of the star in the UV. The energy removed in this way is reprocessed by the gas and dust in the envelope and re-emitted in the IR.

### 5.1.2 Characterization

Generally, the SED is shown as the energy density  $\nu F_\nu$  or  $\lambda F_\lambda$  emitted as a function of frequency  $\nu$  or wavelength  $\lambda$ . Here, the units can differ vastly and all combinations of units are possible in the literature, leading to significant confusion and making the comparison of results often difficult to impossible. Table 5.1 shows examples of some of the most frequently used units and their conversion factors.

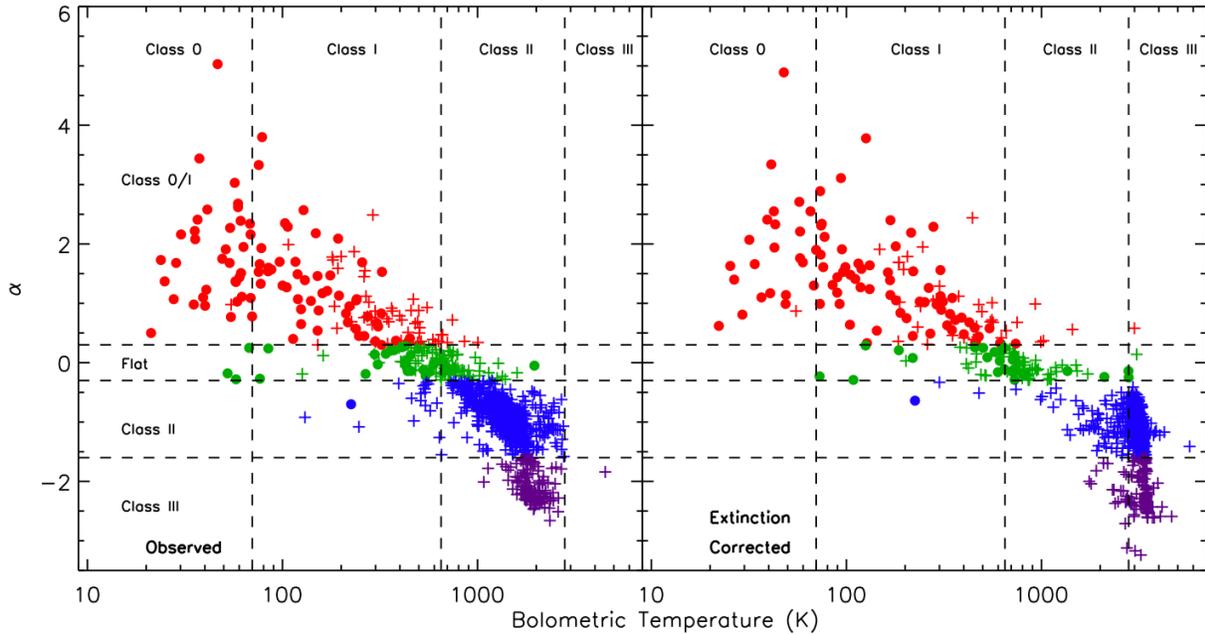


Figure 5.2: The value of  $\alpha$  is plotted versus  $T_{\text{bol}}$  for each source with well-determined values. The color code is based on the Lada class of each source, as dened by  $\alpha$ , with Class I plotted as red, flat as green, Class II as blue, and Class III as purple. Filled circles indicate sources associated with envelopes as traced by millimeter continuum emission, while plus signs indicate sources with no such associations. The right panel shows  $\alpha'$  and  $T'_{\text{bol}}$  calculated after corrections for extinction were applied. The vertical dashed lines show the boundaries between classes, as defined by Chen et al. (1995). Figure and caption taken from Evans (2009).

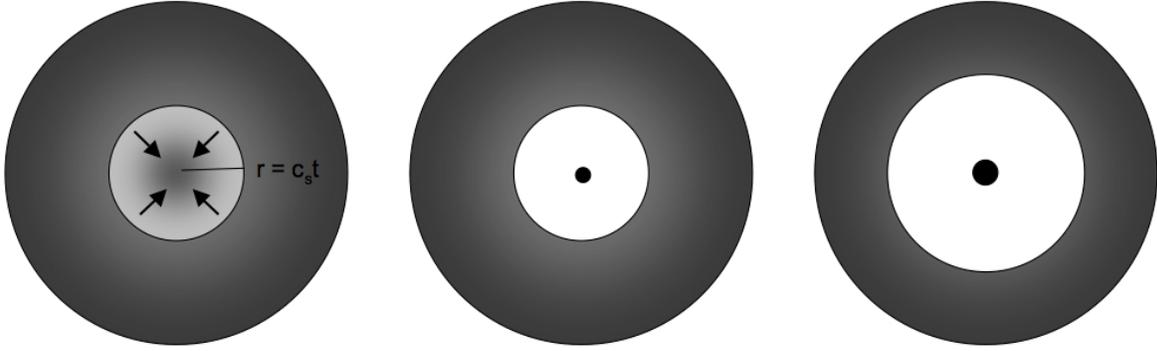


Figure 5.3: Sketch of the inside-out collapse scenario. The expansion wave travels outwards at a speed  $c_s$ .

Generally, an IR slope parameter  $\alpha_{\text{IR}}$  is defined as

$$\alpha_{\text{IR}} = -\frac{d(\nu F_\nu)}{d\nu}, \quad (5.1)$$

in the wavelength region between  $2.2 \mu\text{m}$  and several  $10 \mu\text{m}$ . Typical values for  $\alpha_{\text{IR}}$  are larger than 0 for Class I objects,  $-2 < \alpha_{\text{IR}} < 0$  for Class II sources and  $< -2$  for Class III sources. Another way of classifying young stellar objects (YSOs) is by their "bolometric temperature" which is the temperature of a black body with the same mean frequency as the SED of the YSO. Fig. 5.2 shows the results from the Spitzer Legacy Program c2d (Cores to Disks, PI: Evans). The statistics, i.e. the number of objects in each class, give an indication of the duration of the different phases of pre-main-sequence evolution. The extinction corrected estimated median lifetimes (or half-times) of Class 0 and I are 0.1, 0.44 Myr, respectively.

With the recent advent of large amounts of Spitzer SEDs, the onset of the IR excess  $\nu_{\text{onset}}$  (or  $\lambda_{\text{onset}}$ ) is used as a second parameter. It is defined as the wavelength where the measured fluxes start to be larger than the pure photospheric value. This parameter proved extremely useful in subclassifying the rather inhomogeneous Class II sources, star+disk systems. We come back to this later when we discuss the observational appearance of protoplanetary disks in Chapter 10.

## 5.2 Shu inside-out collapse

The hydrodynamical collapse of a non-isotropic sphere is highly non-homologous. The central parts of a gravitationally unstable sphere collapse faster than the outer parts. If we assume that the central parts of the core have higher densities than the outer parts, the free-fall time  $t_{\text{ff}}$ , which scales with  $\rho^{-1/2}$ , decreases towards the center. This means that the central parts collapse faster than the outer parts and the outer shells subsequently lose their pressure support, while the mass and hence the gravitational attraction inside such a shell does not change. The information of the 'lost pressure support' travels outwards as a wave with the sound speed  $c_s$ .

Frank Shu developed in his seminal paper from 1977 the so-called 'inside-out' collapse scenario and a self-similar dimensionless solution to the problem. Later numerical hydrodynamical simulations confirmed mainly four distinct phases:

1. **Free-fall phase:** The collapse is approximately isothermal and the timescale is the free-fall timescale. Efficient cooling is provided by molecular hydrogen and thermal radiation from dust grains (IR). The next two phases are adiabatic collapse phases, because the inner parts of the core become optically thick and thus cannot cool as efficiently any more.
2. **First core phase:** The first stable core of a few AU in size forms. At temperatures in excess of 1000 K, dust grains evaporate removing them as an opacity source. Once the temperature reaches 2000 K, molecular hydrogen dissociates and a second collapse phase starts. This is because the equation

of state changes (adiabatic exponent) and thus gravity cannot be counterbalanced any more by the internal pressure.

3. **Opacity phase:** The second collapse is halted once the density is high enough ( $> 10^{-2} \text{ g cm}^{-3}$ ) to provide a significant hydrogen ionization fraction, which changes the equation of state once again to a more favorable adiabatic exponent. Several more collapse phases can follow corresponding to the ionization of He and more massive atoms. At the end of phase 3, we are left with a core mass of  $\sim 0.01 M_{\odot}$ .
4. **Accretion phase:** The last phase is characterized by the entire core being optically thick and overall accretion. In this last phase, the protoplanetary disk forms. Actually, most of the material is accreted onto the star through this disk in the late collapse phase. It is also in this phase that the protostar becomes first visible in the infrared. Low mass stars ( $M_{\text{Star}} < 2 M_{\odot}$ ) finish their accretion and thus become exposed before they reach the main-sequence. They continue then to contract until their central temperature becomes high enough for hydrogen burning to start. More massive stars often reach the hydrogen burning limit and thus the main-sequence before the accretion phase ends.

### 5.2.1 Hydrodynamical equations

The basic hydrodynamical equations describing a collapsing dark core are

$$\text{Mass conservation : } \frac{\partial M_r}{\partial t} + u \frac{\partial M_r}{\partial r} = 0, \quad \frac{\partial M_r}{\partial r} = 4\pi r^2 \rho \quad (5.2)$$

$$\text{Momentum conservation : } \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial r} = -\frac{c_s^2}{\rho} \frac{\partial \rho}{\partial r} - \frac{GM_r}{r^2} \quad (5.3)$$

$$\text{Energy equation : } \frac{\partial L_r}{\partial M_r} = -\frac{\partial U_{\text{in}}}{\partial t} - P \frac{\partial(\rho^{-1})}{\partial t} \quad (5.4)$$

Here,  $r$  is the radial coordinate,  $M_r$  the mass inside a radius  $r$ ,  $u$  the infall velocity,  $c_s$  the sound speed,  $\rho$  the density,  $G$  the gravitational constant,  $L_r$  the total energy flux through the surface that includes the mass  $M_r$ , and  $U_{\text{in}}$  the internal energy. The energy flow through the surface can be derived from the basic concepts of energy transport through radiation (see e.g. Kutner, chapter 9.4.2)

$$L_r = \frac{16\pi\sigma_{\text{SB}}r^2T^3}{3\kappa_R} \frac{\partial T}{\partial r}, \quad (5.5)$$

where  $\kappa_R$  is the Rosseland mean opacity (gives the fraction of radiation absorbed per cm), and  $\sigma_{\text{SB}}$  the Stefan-Boltzmann constant.

Following Shu (1977), we can understand the inside-out collapse also as an outward traveling expansion wave (see Fig. 5.3). If we now define a new variable that travels with the expansion wave,

$$x = \frac{r}{c_s t}, \quad (5.6)$$

we can make the following 'Ansatz' for the density, mass and velocity solution using the 'to be determined' functions  $\alpha(x)$ ,  $m(x)$  and  $v(x)$

$$\rho(r, t) = \frac{\alpha(x)}{4\pi G t^2}, \quad M_r = \frac{c_s^3 t}{G} m(x), \quad u(r, t) = c_s v(x) \quad (5.7)$$

You can find the details of the derivation of these functions in the original paper by Shu. Fig. 5.4 shows the time evolution of the collapse until the expansion wave reaches the initial outer boundary. The inner part of the density profile is determined by the free-fall collapse. For  $x \ll 1$ , hence  $r \ll c_s t$ , the density and velocity in the free-fall region can be described as

$$\rho(r, t) = \frac{c_s^{3/2}}{17.96 G} \frac{1}{\sqrt{t}} \frac{1}{r^{3/2}}, \quad u(r, t) = \sqrt{\frac{2GM_*(t)}{r}}. \quad (5.8)$$

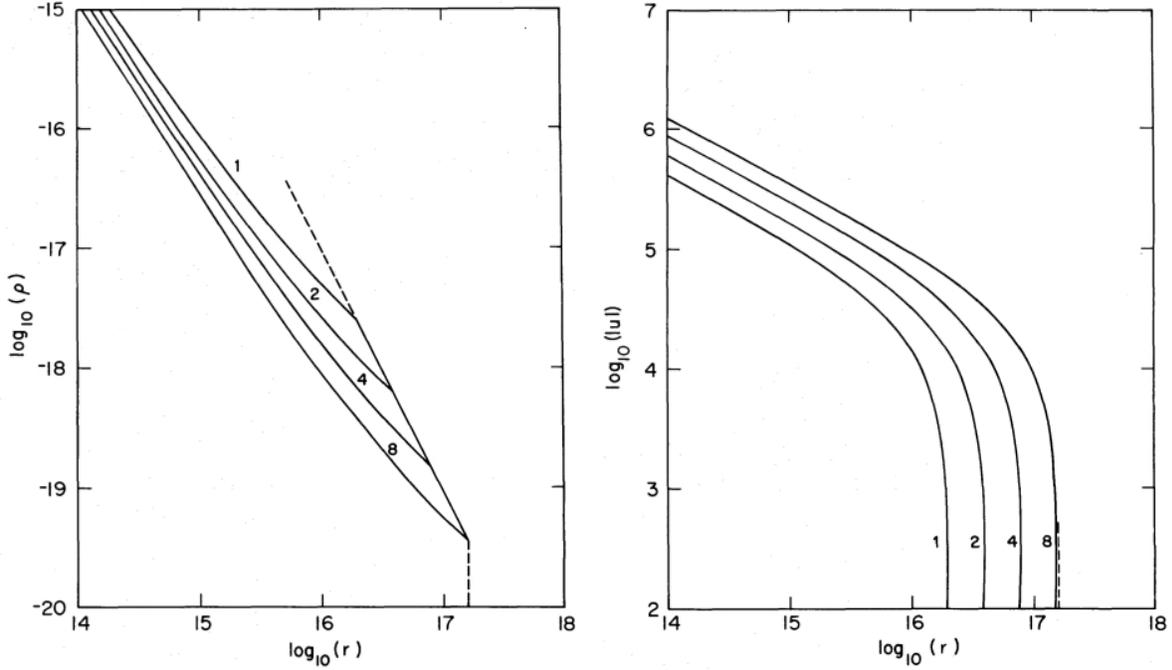


Figure 5.4: Left: Density as a function of radial coordinate for the inside-out collapse of a  $0.96 M_{\odot}$  core with an initial velocity of  $0.2 \text{ km/s}$ . The initial radius of the outer boundary is indicated by the dashed line. The different profiles correspond to  $t = 1, 2, 4, 8 \times 10^{12} \text{ s}$  ( $3.2, 6.3, 12.7, 25.4 \times 10^4 \text{ yr}$ ). Right: The corresponding velocity profiles. The plots are in cgs units.

The accretion rate  $\dot{M} = \frac{\partial M_r(0,t)}{\partial t}$  follows from Eq.(5.7)

$$\dot{M} = 0.975 \frac{c_s^3}{G} \quad (5.9)$$

and is constant. This implies that the stellar mass grows linear with time.

As we move outwards in the solution, we encounter a transition region, where matter starts to fall towards the center. The kink in the profile indicates the position of the expansion wavefront (always at  $x = 1$ ). Outside that point, we see an isothermal sphere solution with  $\rho \sim r^{-2}$ .

## 5.2.2 Instabilities

The adiabatic index  $\Gamma$  is used to describe the relation between the pressure  $P$  and the volume  $V$  of a gas during an adiabatic compression or expansion. For such a process,  $PV^{\Gamma}$  is a constant. This means that for a small adiabatic change in volume and pressure, we can write

$$\begin{aligned} \Gamma \frac{dV}{V} + \frac{dP}{P} &= 0 \\ V dP &= -\Gamma P dV \end{aligned} \quad (5.10)$$

and hence

$$\begin{aligned} d(PV) &= PdV + VdP \\ &= -(\Gamma - 1)PdV \end{aligned} \quad (5.11)$$

During an adiabatic collapse, there is no heat transfer and so the change in internal energy  $U_{\text{in}}$  of the system can be written as

$$dU_{\text{in}} = -PdV = \frac{1}{\Gamma - 1}d(PV) . \quad (5.12)$$

Hence, for a constant adiabatic index  $\Gamma$ , the relation

$$U_{\text{in}} = \frac{1}{\Gamma - 1}PV \quad (5.13)$$

holds. For a system in virial equilibrium, the average pressure  $\langle P \rangle$  can be written as one third of the density of stored gravitational energy

$$\langle P \rangle = -\frac{1}{3} \frac{U_G}{V} . \quad (5.14)$$

For a self-gravitating gas in hydrostatic equilibrium, we can thus write

$$\begin{aligned} -\frac{1}{3} \frac{U_G}{V} &= (\Gamma - 1) \frac{U_{\text{in}}}{V} \\ 3(\Gamma - 1)U_{\text{in}} + U_G &= 0 \end{aligned} \quad (5.15)$$

The total energy is

$$U_{\text{tot}} = U_{\text{in}} + U_G = (4 - 3\Gamma)U_{\text{in}} , \quad (5.16)$$

and the system is bound or stable if  $U_{\text{tot}} < 0$ , hence if  $\Gamma > 4/3$ .

On the other hand, Eq.(5.5) indicates that  $\Gamma < 5/3$ . This is the limit for radiative energy transport. If  $\Gamma$  were larger, the system would become unstable to convection and energy transport could happen that way as well.

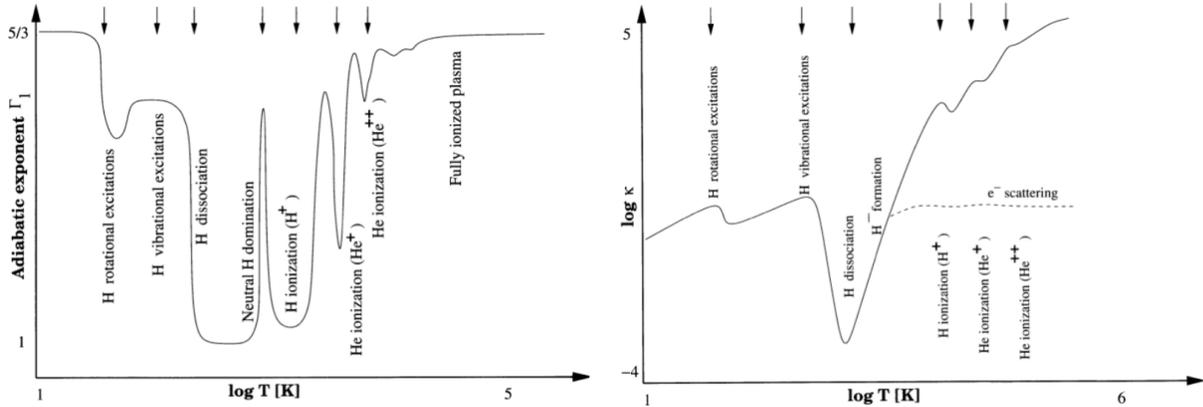


Figure 5.5: Left: The adiabatic index ( $\Gamma_1$  is called the first adiabatic index,  $\left(\frac{\partial \ln P}{\partial \ln \rho}\right)_{\text{ad}}$ ) as a function of temperature. The arrows mark the onset of specific physical processes that change the index most dramatically causing instability in the nearly formed stellar core (figure and caption from Schulz 2005). Right: The Rosseland mean opacity  $\kappa_R$  as a function of temperature taken from the same calculation as the adiabatic index. It shows the behaviour of opacity parallel to the changes in the adiabatic index and thus how it can counter drifts into instability regions (figure and caption from Schulz 2005). Comment: The leftmost three arrows should be labelled with  $\text{H}_2$  instead of H.

Fig. 5.5 shows on the left the adiabatic index as a function of temperature  $T$ . With increasing temperature, the molecules in the gas get first excited, and then they dissociate ( $\text{H}_2$  at  $\sim 2000 \text{ K}^1$ ). Above a critical temperature, which is density dependent, the gas becomes more and more ionized. The figure was computed for a gas density of  $10^{-14} \text{ g cm}^{-3}$ . With changing density, the depth and heights of these peaks and valleys in  $\Gamma$  changes.

<sup>1</sup>Under conditions of thermodynamic equilibrium, the dissociation constant  $K(\text{H}_2)$  is related to the partial pressure of the atomic constituent  $p(\text{H})$  and  $p(\text{H}_2)$  through  $K(\text{H}_2) = p^2(\text{H})/p(\text{H}_2)$ . The dissociation constant  $K$  depends only on temperature. At 2000 K,  $K(\text{H}_2) = 3.45 \text{ dyn/cm}^2$ .

### 5.2.3 Opacities

The right side of Fig. 5.5 shows the gas opacities as a function of temperature  $T$  corresponding to the change in adiabatic index shown in the left side of Fig. 5.5. The opacity is needed to prevent too much energy from leaking out. This enables the temperature and thus the pressure in the core to build up and hence hydrostatic equilibrium to be established. To calculate the Rosseland mean opacity for a certain gas mixture, we need to know the individual opacities for more than 60 million atomic, molecular transitions from neutral and ionized stages. This has largely been enabled by the nuclear bomb experiments after the second world war. Later astrophysics benefitted from this data and the largest projects were the Opacity Project (OP) and the Opacity Project at Livermore (OPAL).

The Rosseland opacity shown here contains also grain opacities at low temperature (typical condensation temperatures of grains are around 1300 K). At low temperatures and densities, changes in the opacity are due to the melting of grains ice mantles and later grain evaporation. The next changes arise from the ionization of alkali metals and then the dissociation of  $H_2$ .

Instabilities that are caused by rapid changes in opacity in connection with increasing temperature and density are called 'secular' instabilities. If  $\kappa_R$  increases upon adiabatic compression, we call these instabilities 'vibrational' instabilities. They are more complex than the dynamical instability that arises from the adiabatic index only.

### 5.2.4 Rotating collapse

The previously sketched picture of a spherical inside-out collapse (gravity versus pressure) cannot explain a series of observational phenomena that accompany star formation. These are for example the presence of protoplanetary disks, jets and outflows. While magnetic fields or rotation will not prevent the collapse, they will change the way matter flows and accretes.

We now briefly consider the impact of rotation on the protostellar collapse. Since the cloud remains hydrodynamical stable for a long time, we can assume that it rotates as a solid body. Also, we assume that the cloud rotates slowly, so that the centrifugal force is initially very small and the early stages of collapse are almost radial. In this limit, we can distinguish between an outer envelope which stays almost spherical and an inner region that gets distorted due to rotation and forms an equatorial accretion disk (Fig. 5.6).

Depending on the specific angular momentum of a gas parcel, it will either be able to accrete onto the protostellar core (very low specific angular momentum) or accrete onto the equatorial disk at some radius corresponding to its specific angular momentum. The maximum possible distance from the star at which such a gas parcel can end up is called the centrifugal radius  $r_c$ . This is also the quantity limiting the inner region and thus the size of the accretion disk. In the following paragraphs we will derive this quantity from simple principles.

If we study the trajectory of a fluid element during the infall, we can assume that the protostar + disk mass is a point source and that the only gravity and rotation play a role. Under these circumstances, the fluid element will have a bound elliptical orbit (the conic section corresponding to a negative energy — hence an ellipse or circle; unbound orbits have conic sections corresponding to a parabola or hyperbola). However,

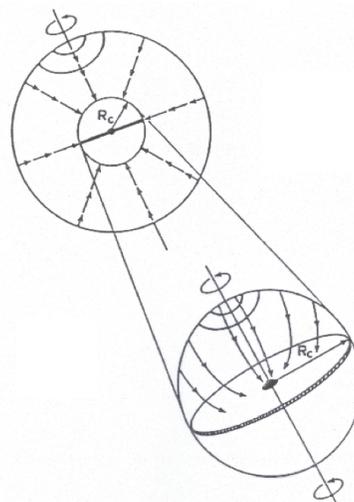


Figure 5.6: Sketch of an inside-out collapse including rotation. The outer envelope stays spherical and the inner region is affected by the rotation. The radius dividing these two zones is the centrifugal radius  $r_c$  (or here denoted as  $R_c$ ). Inside the centrifugal radius, an equatorial accretion disk forms.

since the initial energy is small compared to the energy it has when it reaches the equatorial plane, we can approximate it reasonably well by a parabola (zero-energy or unbound orbit; parabola open to the left — along x-axis — focus at (0,0))

$$r = \frac{r_{\text{eq}}}{1 + \cos \Psi} , \quad (5.17)$$

where the coordinates and quantities are introduced in Fig. 5.7. If the fluid element would reach  $r_{\text{min}}$ , the maximum orbital velocity would be

$$v_{\text{max}}^2 = \frac{2GM_*}{r_{\text{min}}} , \quad (5.18)$$

where  $M_*$  is the mass of the protostar + disk. However, the fluid element will never reach that part of the trajectory since it collides with the forming disk at  $\Psi = 90^\circ$ . The specific angular momentum can be written as

$$j^2 = r_{\text{min}}^2 v_{\text{max}}^2 = 2GM_* r_{\text{min}} . \quad (5.19)$$

Since the fluid element reaches  $r_{\text{min}}$  at  $\Psi = 0$ , Eq.(5.17) implies that  $r_{\text{min}} = r_c/2$ . Therefore, we can write the radius at which the fluid element reaches the plane as

$$r_{\text{eq}} = \frac{j^2}{GM_*} . \quad (5.20)$$

Now we can express the specific angular momentum in terms of the initial quantities  $\Omega_0$  (angular velocity) and  $\theta_0$  (inclination of orbital plane with respect to the rotation axis,  $\theta_0 = 90^\circ$  is the midplane)

$$j = R^2 \Omega_0 \sin \theta_0 = c_s^2 t^2 \Omega_0 \sin \theta_0 , \quad (5.21)$$

where  $R$  is the distance from the center at which the fluid element starts its collapse (position of the rarefaction wave at time  $t$  — its speed is  $c_s$  and hence  $R = c_s t$ ). The angular momentum is larger for higher values of  $\theta_0$  and has a maximum at  $\theta_0 = 90^\circ$  (the centrifugal radius). Inserting this into Eq.(5.20) and using the expression found for the core mass  $M_r(t)$  in Sect. 5.2.1, we can rewrite  $r_{\text{eq}}$  as

$$r_{\text{eq}} = \frac{c_s t^3 \Omega_0^2 \sin^2 \theta_0}{m_0} . \quad (5.22)$$

For  $\theta_0 = 90^\circ$ ,  $r_{\text{eq}}$  is the centrifugal radius of the disk

$$r_c \approx 7.33 \text{ AU} \left( \frac{c_s}{0.35 \text{ km/s}} \right) \left( \frac{\Omega}{10^{-14} \text{ s}^{-1}} \right)^2 \left( \frac{t}{10^5 \text{ yr}} \right)^3 . \quad (5.23)$$

For the latter expression, we assumed that  $m_0 \sim 1$  (see Sect. 5.2.1). The value  $\Omega = 10^{-14} \text{ s}^{-1}$  is close to typical observed values of a few times  $10^{-14} - 10^{-13} \text{ s}^{-1}$ . This gives us also an estimate of the initial size of the accretion disk. We will later see that the size of this initial disk grows in the later stages of star formation due to e.g. angular momentum transport (viscous spreading of the disk).

### 5.2.5 Collapse in the presence of a magnetic field

An additional complexity is introduced by the presence of a magnetic field. Typical values in molecular clouds are small, of the order of a few  $10 \mu\text{G}$ . We consider here for simplicity the non-rotating case. The magnetic field causes an additional pressure component that can counterbalance the collapse of the core. In the case of a poloidal magnetic field (Fig. 5.8), the outward moving expansion wave (Sect. 5.2.1) travels

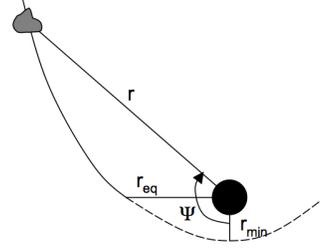


Figure 5.7: Sketch of a parabolic orbit in rotating infall. A fluid element with polar coordinates  $(r, \Psi)$  falls in and hits the equatorial disk at a radius  $r_{\text{eq}}$ . In the absence of the disk, the fluid element would have reached the smaller distance  $r_{\text{min}}$  before swinging back out again.

faster parallel to the magnetic field lines. Even though the gas flow itself slows down in the presence of a magnetic field, the expansion wave travels faster, resulting in an almost unchanged mass accretion rate  $\dot{M}$ .

In analogy to the previous section, we can now define a magnetic centrifugal radius

$$r_b = k_b G^{2/3} B_0^{4/3} c_s^{-1/3} t^{7/3} \quad (5.24)$$

$$\approx 540 \text{ AU} \left( \frac{B_0}{30 \mu\text{G}} \right)^{4/3} \left( \frac{c_s}{0.35 \text{ km/s}} \right)^{-1/3} \left( \frac{t}{10^5 \text{ yr}} \right)^{7/3}, \quad (5.25)$$

where  $k_b = 0.12$  (from numerical calculations). Typically,  $r_b$  is 1-2 orders of magnitude larger than  $r_c$ . Fig. 5.9 illustrates the collapse environment roughly  $2 \times 10^5$  yr after the onset of collapse (left) and at the end of the free-fall phase (right). This figure illustrates the formation of a pseudo-disk (caused by magnetic forces deflecting the infalling gas towards the equatorial plane) that has in the end a size of  $\sim 1000$  AU. We can picture the much smaller centrifugally supported disk inside the thick line (labelled 'inside solution'). Hence, we can imagine that the inner accretion disk is fed by the outer magnetically stabilized pseudo-disk.

Another characteristic scale is the radius  $r_m$  at which magnetic and thermal pressure are equal

$$\frac{B_0^2}{8\pi} \sim \rho c_s^2. \quad (5.26)$$

At the begin of the collapse, we can approximate the density profile as that of an isothermal sphere (see Sect. 5.2.1)

$$\rho = \frac{c_s^2}{2\pi G r^2}. \quad (5.27)$$

Inserting this in Eq.(5.26) yields

$$\frac{B_0^2}{8\pi} = \frac{c_s^4}{2\pi G r^2} \quad (5.28)$$

$$r_m = \frac{2c_s^2}{B_0 \sqrt{G}}. \quad (5.29)$$

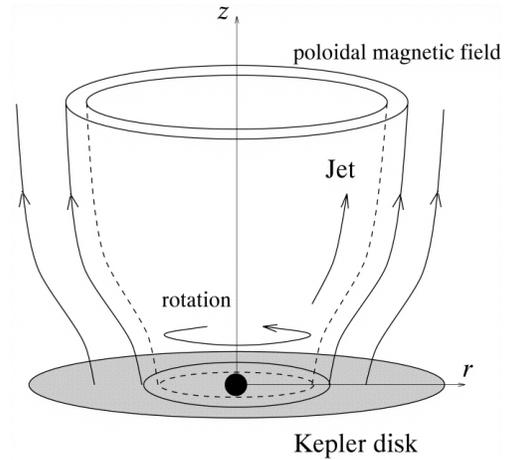


Figure 5.8: Sketch of the poloidal magnetic field during the protostellar collapse.

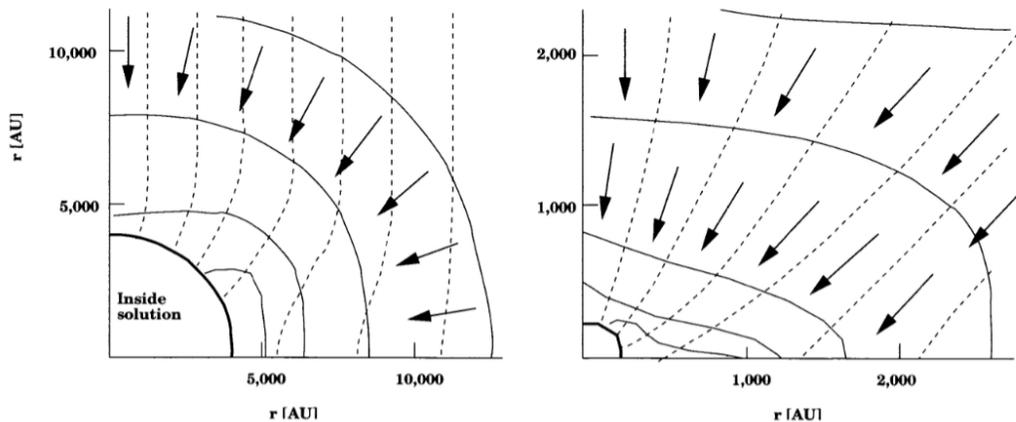


Figure 5.9: Illustration of a collapse of a magnetic spherical, isothermal sphere at about  $2 \times 10^5$  yr (left) and at the end of the free-fall phase (right). The solid lines are contours of equal density, the dashed lines are magnetic field lines, the thick arrows show the velocity field. The horizontal and vertical axis are in AU.

### 5.3 Line profiles

If spectrally resolved, line profiles carry information on the gas motion along the line of sight, its absolute values and signs. The characteristic emission profile of the protostellar collapse phase is an asymmetric double-peaked line. We can understand the formation of this profile from a simple sketch of the radial density and temperature profile of the core and its velocity structure. For that, Fig. 5.10 illustrates how the profile is composed if we assume that the collapse can be approximated by a series of uniform gas density and temperature shells. The central shells have higher densities and temperatures compared to the outer shells (see top panel). The layers in front of the protostar produce redshifted emission, while the ones behind the protostar produce blueshifted emission (the red and blue arrows indicate the infall motion of the shells - middle panel). An important pre-requisite for observing such a collapse signature is that the line is optically thick. An optically thin line is symmetric and thus indistinguishable from a normal emission line without infall/outflow. The line profile can be broken up into velocity bins corresponding to the individual layers. The layer responsible for most of the emission in a velocity bin is indicated by its number. The degree of radiation leaking into neighboring velocity bins is given by the turbulent velocity field ( $\sigma$ ). If the line is optically thick, absorption turns the line profile asymmetric as outer cooler gas shells in the redshifted wing can partially absorb emission from the hotter shells that are closer to the protostar. This is not the case for the blueshifted wing, because in that case, the gas temperature increases towards the observer (hotter inner shells are closer to us than the cool outer shells). In the optically thin case, we do not encounter absorption and hence the profile stays symmetric.

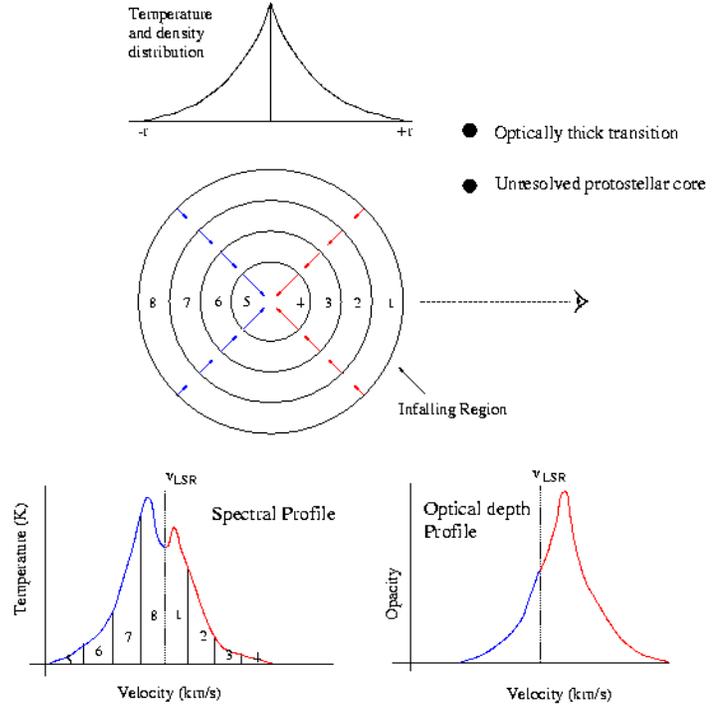


Figure 5.10: Sketch of the line profile formation in a collapsing core.

The line profile can be broken up into velocity bins corresponding to the individual layers. The layer responsible for most of the emission in a velocity bin is indicated by its number. The degree of radiation leaking into neighboring velocity bins is given by the turbulent velocity field ( $\sigma$ ). If the line is optically thick, absorption turns the line profile asymmetric as outer cooler gas shells in the redshifted wing can partially absorb emission from the hotter shells that are closer to the protostar. This is not the case for the blueshifted wing, because in that case, the gas temperature increases towards the observer (hotter inner shells are closer to us than the cool outer shells). In the optically thin case, we do not encounter absorption and hence the profile stays symmetric.

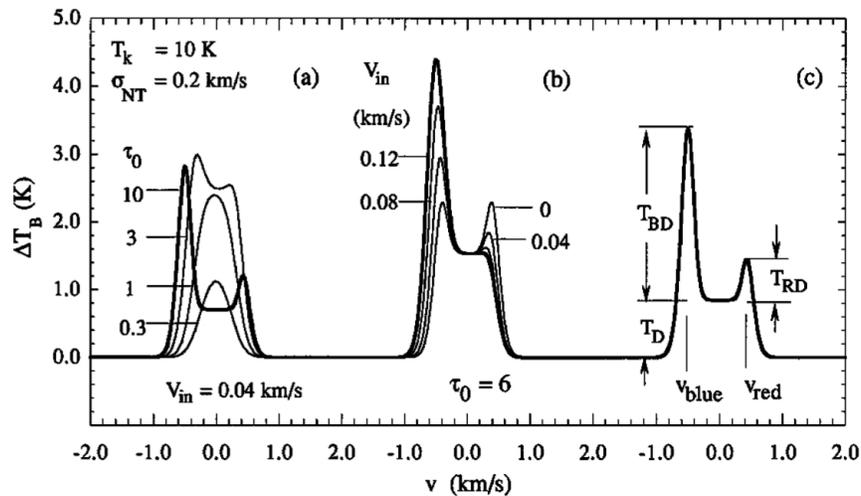


Figure 5.11: Variation of infall asymmetry (a) with peak optical depth  $\tau_0$  and (b) with infall speed  $V_{in}$  (from Myers et al. 1996). (c) illustrates the definition of various line parameters.

Fig. 5.11 illustrates this for a series of models with varying optical depth and infall velocities. In a radiative transfer model of two uniform layers with velocity dispersion  $\sigma$  and approach speed  $V_{\text{in}}$ , the line profile is symmetric for  $\tau_0 < 1$ , but its peak skews to the blue as  $\tau$  increases beyond 1. For  $V_{\text{in}} < \sigma$  and increasing  $\tau_0$ , the profile has two peaks, with increasing ratio of blue to red peak intensity. As  $V_{\text{in}}$  increases for fixed  $\tau_0 > 1$ , the blue-red intensity ratio increases until the red peak disappears into a red shoulder. The right panel of Fig. 5.11 illustrates the definition of a set of line parameters used to estimate the infall speed from the line profile

$$V_{\text{in}} \approx \frac{\sigma^2}{v_{\text{red}} - v_{\text{blue}}} \ln \left( \frac{1 + eT_{\text{BD}}/T_{\text{D}}}{1 + eT_{\text{RD}}/T_{\text{D}}} \right) \quad (5.30)$$

The infall line asymmetry is not limited to the central velocities. The asymmetry can occur in the line wings as long as the emission stays optically thick.

Fig. 5.12 illustrates the infall line profiles in the starless core L1544 (Tafalla et al. 1998). The lines vary from the high optical depth  $\text{HCO}^+$  1-0 line with strong infall asymmetry to the low optical depth  $\text{C}^{34}\text{S}$  2-1 line without infall asymmetry.

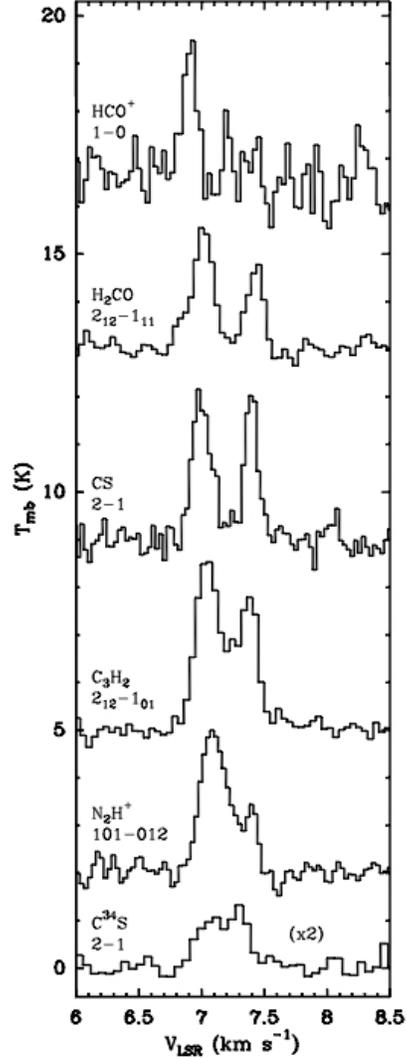


Figure 5.12: Spectral profiles of dense gas tracer lines in the starless core L1544 (Tafalla et al. 1998).

## Chapter 6

# Protostars & pre-main-sequence evolution

We have discussed the structure and collapse of cold cores in the previous chapters without putting much emphasis on the protostellar core structure. We treated the central core as a sink of mass and only considered its gravitational impact on the outer envelope. It is however clear that this inner core will at some point start to irradiate the remaining envelope from the inside. Hence, in this chapter, we focus on the evolution and detailed structure of the protostar all the way from its formation to the arrival on the main-sequence.

### 6.1 Timescales

There are three timescales relevant for the protostar: (1) the free-fall timescale of the collapsing core, (2) the Kelvin-Helmholtz timescale, (3) the accretion timescale.

The free-fall timescale was derived in chapter 4

$$t_{\text{ff}} = \sqrt{\frac{3\pi}{32G\rho}} \approx \sqrt{\frac{1}{G\rho}} , \quad (6.1)$$

and describes the time that a particle within a cloud of density  $\rho$  would take to fall a distance  $r$  (we approximated here  $\sqrt{3/2\pi} = 0.7$  to be of the order of 1 since these types of estimates are order of magnitude estimates anyhow).

The second timescale has been introduced in chapter 2

$$t_{\text{KH}} = \frac{U_{\text{grav}}}{L_*} = \frac{3 GM_*^2}{5 R_* L_*} . \quad (6.2)$$

This describes the time that a star can radiate its luminosity  $L_*$  with the sole energy source being its gravitational energy  $U_{\text{grav}}$ .

The accretion timescale can be expressed as

$$t_{\text{acc}} = \frac{M_{\text{core}}}{\dot{M}} . \quad (6.3)$$

The free-fall timescale is much shorter than the Kelvin-Helmholtz timescale and the accretion timescale. Comparing the latter two, we can distinguish two cases,  $t_{\text{KH}} > t_{\text{acc}}$  and  $t_{\text{KH}} < t_{\text{acc}}$ . In the first case, the stellar interior cannot thermally adjust to the accretion and the luminosity will be generated by accretion shocks on the surface (dissipation of kinetic energy into heat). This is the 'protostellar phase'. In the latter case, the accretion becomes inefficient and the star evolves towards the main-sequence on a thermal timescale with the stellar luminosity being generated by contraction. We call this the 'pre-main sequence phase'.

If we consider the expression for the mass accretion rate that follows from the simple Shu inside-out collapse model, we can estimate the value of accretion rates as well as the timescale for forming a  $1 M_{\odot}$  star from an isothermal 10 K envelope

$$\dot{M} \sim \frac{c_s^3}{G} \sim 1.6 \times 10^{-6} M_{\odot} \text{ yr}^{-1} \quad (6.4)$$

$$t_{\text{acc}} = 6.3 \times 10^5 \text{ yr} . \quad (6.5)$$

## 6.2 Early growth and collapse

The isothermal approximation that we used in the previous section for the core collapse breaks down when we consider the forming protostar itself. Initially, the temperature of the core is fairly low, a few 100 – 1000 K, and thus most of the energy will be radiated in the infrared. As material piles up on the central core, the matter rapidly becomes optically thick, thus preventing gravitational energy from being efficiently radiated away. This leads to an increase in central temperature and the growing gas pressure acts to stabilize the core (Fig. 6.1). The core continues to grow until it reaches a mass of  $\sim 5 \times 10^{-2} M_{\odot}$ ; it has then a size of roughly 5 AU.

From the simplest form of the virial theorem (only thermal and gravitational energy, no magnetic and kinetic energy)

$$\begin{aligned} U_{\text{grav}} &= -\frac{3}{5} \frac{GM_*^2}{R_*} \\ = -2U_{\text{therm}} &= -2 \frac{3}{2} \frac{kTM_*}{\mu m_{\text{H}}} , \end{aligned} \quad (6.6)$$

we can derive an estimate for the temperature

$$T = \frac{m}{5k} \frac{GM_*}{R_*} \approx 520 \text{ K} \left( \frac{M_*}{5 \times 10^{-2} M_{\odot}} \right) \left( \frac{R_*}{5 \text{ AU}} \right)^{-1} \quad (6.7)$$

Here,  $M_*$  and  $R_*$  denote the mass and radius of the protostellar core. We have assumed a mean molecular weight  $\mu = 2.4$  appropriate for molecular gas.

In this first phase, molecular hydrogen acts as a thermostat. At temperatures above 2000 K, collisions can dissociate  $\text{H}_2$ , a process that absorbs 4.48 eV. At the same time, the thermal energy of an  $\text{H}_2$  molecule is only 0.74 eV. Hence, during this phase, most of the gravitational contraction energy can be absorbed through the dissociation of a small fraction of the total molecular hydrogen. As a consequence, the temperature rise becomes more moderate as also shown in Fig. 6.1. The central region with atomic hydrogen starts to spread outward and this eventually leads to an unstable configuration as the adiabatic index drops below the critical value of  $4/3$ .

As the density rises further  $\rho \sim 10^{-2} \text{ g cm}^{-3}$ , hydrogen can eventually be ionized, thus pushing the adiabatic index beyond  $4/3$ . This leads to the formation of a second stable core which grows to  $0.01\text{--}0.02 M_{\odot}$  in less than 100 yr.

The protostellar mass then grows to its final value during the main accretion phase, or 'protostellar phase', where  $t_{\text{acc}} < t_{\text{KH}}$ . We can estimate the maximum size that a protostar can have as well as its accretion

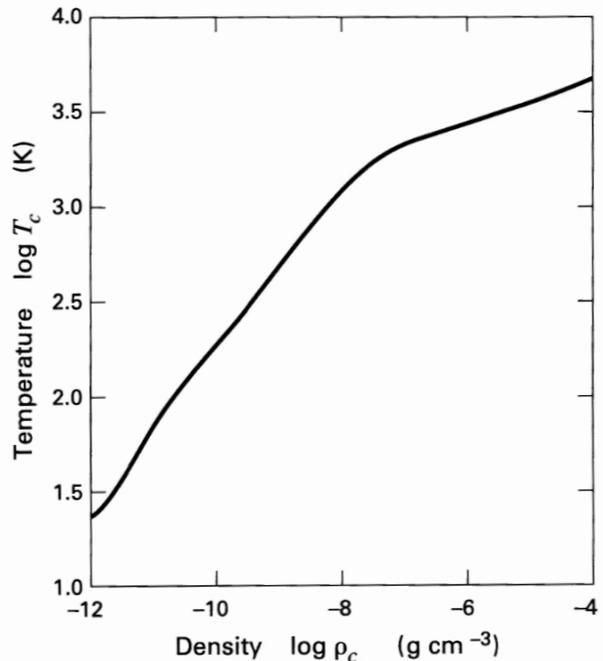


Figure 6.1: Evolution of the central temperature in the first core. The temperature is plotted as a function of central density (from Stahler & Palla 2004).

luminosity again from simple consideration of the virial theorem  $U_{\text{therm}} = -1/2 U_{\text{grav}}$ , we can write down the energy conservation equation as

$$-\frac{1}{2} \frac{3}{5} \frac{GM_*^2}{R_*} + E_{\text{int}} + L_{\text{rad}} t = 0 \quad , \quad (6.8)$$

where  $L_{\text{rad}}$  is the average luminosity escaping over the timescale  $t$  and  $E_{\text{int}}$  is the internal energy of the gas. Since the matter is initially in a cold sphere with large dimensions, the initial internal energy can be considered to be zero. However, during the contraction, molecular hydrogen gets dissociated (dissociation energy  $E_{\text{diss}} = 4.48$  eV) and hydrogen and helium get ionized (ionization energies  $E_{\text{ion}}(\text{H}) = 13.6$  eV and  $E_{\text{ion}}(\text{He}) = 75$  eV). We can consider this as internal energy  $\Delta E_{\text{int}}$  which we can thus write as

$$\Delta E_{\text{int}} = \frac{XM_*}{m_{\text{H}}} \left[ \frac{\Delta E_{\text{diss}}(\text{H}_2)}{2} + \Delta E_{\text{ion}}(\text{H}) \right] + \frac{YM_* \Delta E_{\text{ion}}(\text{He})}{4m_{\text{H}}} \quad . \quad (6.9)$$

$X$  and  $Y$  are here the mass fractions of hydrogen and helium respectively. If we ignore  $L_{\text{rad}}$ , we can derive an expression for the maximum possible radius of the protostar

$$R_{\text{max}} = \frac{3}{10} \frac{GM_*^2}{\Delta E_{\text{int}}} = 36 R_{\odot} \left( \frac{M_*}{M_{\odot}} \right) \quad . \quad (6.10)$$

The observed sizes of e.g. T Tauri stars (solar mass protostars) are generally smaller by an order of magnitude.

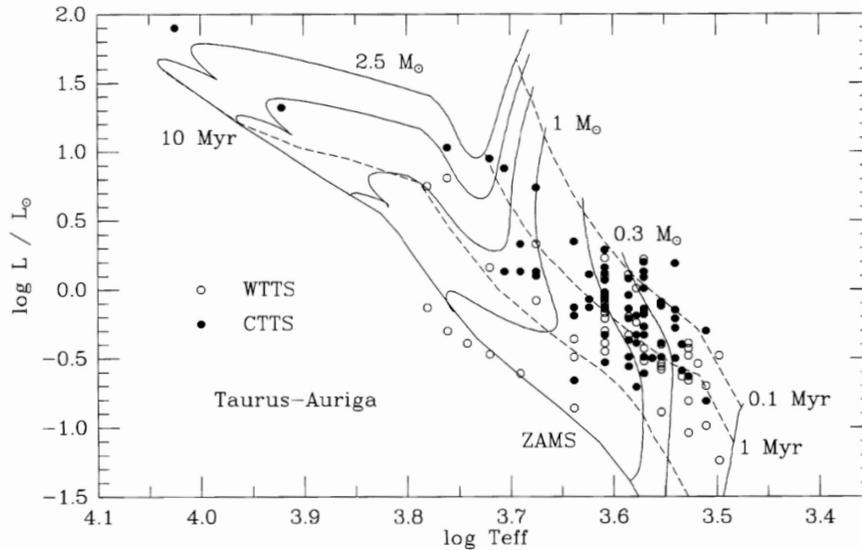


Figure 6.2: HR diagram positions of young stars lying within the Taurus-Auriga molecular cloud complex. For comparison, theoretical evolutionary tracks for pre-main-sequence stars of masses 2.5, 2.0, 1.5, 1.0, 0.5, 0.3, and 0.1  $M_{\odot}$  are shown. The dashed lines are isochrones for ages of  $10^5$ ,  $10^6$ , and  $10^7$  yr with the hydrogen-fusion zero-age main sequence (ZAMS) shown as the lowest line running from upper left to lower right. The open circles refer to weak-emission T Tauri stars (WTTS), while the filled circles denote the position of the classical T Tauri stars (CTTS) (figure and caption from Hartmann 1998).

Since we know that the true stellar radius is much smaller than the derived estimate of  $R_{\text{max}}$ , we know that the second term in Eq.(6.8) must be much smaller than the first one (if  $R = R_{\text{max}}$ , the two terms would be equal). Hence, we can neglect the second term in Eq.(6.8),  $\Delta E_{\text{int}}$  and derive an expression for the

luminosity which we assume to be close to the accretion luminosity

$$L_{\text{acc}} = \frac{3}{10} \frac{GM_* \dot{M}}{R_*} \quad (6.11)$$

$$\approx 18 L_{\odot} \left( \frac{\dot{M}}{10^{-5} M_{\odot} \text{ yr}^{-1}} \right) \left( \frac{M_*}{M_{\odot}} \right) \left( \frac{R_*}{5R_{\odot}} \right)^{-1} \quad (6.12)$$

For this, we assumed that  $\dot{M} = M_*/t$ . This equation tells us, that for a protostar with a fixed mass  $M_*$ , the energy loss through radiation will directly be connected to a contraction. If we assume for a second that the temperature of the star will stay constant (e.g. at a reference point in time, the star has the size  $R_0$  and luminosity  $L_0$ ),

$$L_* = 4\pi R_*^2 \sigma T_*^4 \quad (6.13)$$

$$T_* = \left( \frac{L_*}{4\pi\sigma R_*^2} \right)^{1/4} \approx \text{constant} \quad (6.14)$$

$$L_* = L_0 \frac{R_*^2}{R_0^2} \quad (6.15)$$

$$L_* = L_0 \left( \frac{7}{3} \frac{t L_0 R_0}{GM_*^2} \right)^{-2/3} = L_0 \left( \frac{7}{5} \frac{t}{t_{\text{KH}}} \right)^{-2/3} \quad (6.16)$$

As the star ages, it becomes thus smaller and fainter. The rate of change in luminosity and size slows with age. This is illustrated in Fig. 6.2, where the pre-main-sequence evolutionary tracks are shown in an HR diagram. The almost vertical part of the evolutionary track ( $T_* \approx \text{constant}$ ) is frequently referred to as 'Hayashi track'. A protostar is often defined as an object that derives most of its luminosity from accretion, hence  $L_* \sim L_{\text{acc}}$ .

At this point, we make a small note concerning some of the pre-factors in all these equations. Depending of which book one consults, these factors can differ by factors of  $\pi$ , 2 or 4, or anything close to it. Most of the time, these differences are irrelevant since the equations themselves contain assumptions that are often only order of magnitude assumptions. Hence, a factor 2 is never a problem. However, if you ever encounter larger discrepancies, please consult one of the authors (lecturers) to resolve the issue. It could very well be a typo. Within this syllabus, we try to be as consistent as possible in the definitions of certain quantities such as Kelvin-Helmholtz timescale, Jeans radius etc..

### 6.3 Dust envelope

We have now derived expressions for the luminosity and size of a protostar, but these are rather theoretical quantities as the surrounding envelope of material often is optically thick at UV and optical wavelength and starts to become more transparent in the infrared (see dust opacity as a function of wavelength). This means the actual surface of emission is not the protostar itself, but rather a sort of dust photosphere that we will now explore in somewhat

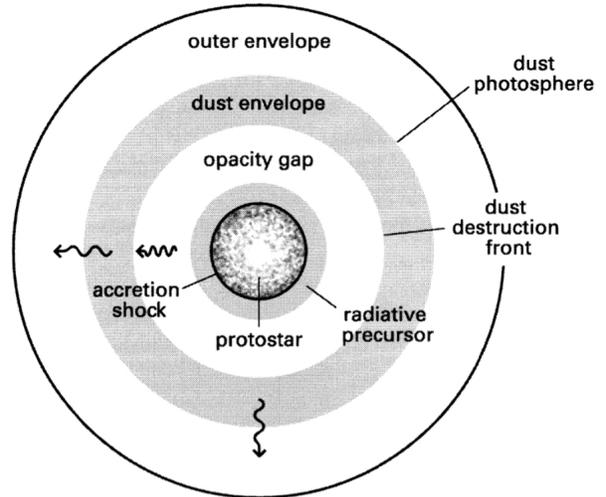


Figure 6.3: Structure of a spherical protostar and its infalling envelope. The relative dimensions of the outer regions have been greatly reduced in this sketch. Note the conversion from optical to infrared photons in the dust envelope (figure and caption from Stahler & Palla 2004).

more detail (Fig. 6.3). It also means that the UV/optical photons of the star are absorbed and scattered many times in the envelope and heat the dust there to a temperature  $T_{\text{dust}}$ . The envelope dust then emits according to its own temperature, mostly in the infrared. Hence, stellar photons are converted into infrared photons emitted by the dust. The dust photosphere is then the layer where most of the emission is coming from, hence the  $\tau_{\text{R}} \sim 1$  layer (analogy to the solar photosphere at optical wavelength;  $\tau_{\text{R}}$  is the Rosseland mean optical depth).

We can find an estimate for the size of this dust photosphere by using the following two equations

$$\rho(r)\kappa_{\text{R}}R_{\text{phot}} = 1 \quad (6.17)$$

$$L_{\text{acc}} = 4\pi R_{\text{phot}}^2 \sigma T_{\text{phot}}^4, \quad (6.18)$$

where  $\kappa_{\text{R}}$  is the Rosseland mean opacity,  $\rho$  is the density profile found from the inside-out collapse for the free-fall region, and  $T_{\text{phot}}$ ,  $R_{\text{phot}}$  are the temperature and radius of the dust photosphere. Numerical solution of this system of two equations with two unknowns yields values of  $R_{\text{phot}} = 2.1 \times 10^{14}$  cm (14 AU) and  $T_{\text{phot}} = 300$  K. So, according to Wien's law, the peak emission lies indeed in the infrared spectral region (around 50  $\mu\text{m}$ ). Of course, these mean opacities and simple approximations can only give a very crude picture of what the dust photosphere is. Full frequency-dependent radiative transfer is needed to calculate the dust temperature through an iterative (the local source function depends on the dust temperature and vice versa) or Monte-Carlo procedure.

## 6.4 Stellar structure

After this short excursion into the outer envelope, we now return to the protostellar interior and study the stellar structure equations that will tell us how mass and radius are related for these objects (similar to the mass-radius relation for main-sequence stars). We then discuss deuterium burning and lithium destruction in the stellar interior.

The basic equations of stellar structure are the equations of mass continuity, hydrostatic equilibrium, state, energy transport and energy generation

$$\frac{\partial M_r}{\partial r} = 4\pi r^2 \rho(r) \quad (6.19)$$

$$\frac{\partial P}{\partial r} = -\frac{G\rho(r)M_r}{r^2} \quad (6.20)$$

$$P(r) = \frac{\rho(r)}{\mu} RT \quad (6.21)$$

$$\frac{\partial T_r}{\partial r} = L_r \left[ \frac{\kappa\rho(r)}{16\pi r^2 \sigma T_r^3} \right] \quad (6.22)$$

$$\frac{\partial L_{\text{int}}}{\partial r} = 4\pi r^2 \rho \left( \epsilon(r) - T \frac{\partial s}{\partial t} \right) \quad (6.23)$$

where  $R$  is the gas constant,  $\mu$  the mean molecular weight,  $\epsilon(r)$  the rate of nuclear energy generation per unit mass within a shell of size  $r$  and  $s$  the specific entropy (entropy per units mass). The second term in the energy generation  $-T\partial s/\partial t$  stems from the contraction of the star.

We can solve these equations numerically by using opacities, mean molecular weights, energy generation rate and entropy tabulated as functions of  $T$  and  $\rho$ . To close the systems of equations, we also need some boundary conditions. For the center of the protostar,  $M_r$  and  $L_{\text{int}}$  vanish

$$M(0) = 0 \quad (6.24)$$

$$L_{\text{int}}(0) = 0. \quad (6.25)$$

At the surface of the protostars  $r = R_*$ , the pressure should approach the ram pressure of the infalling gas

$$P = \rho v_{\text{ff}}^2 = \frac{\dot{M}R_*^{-3/2}}{4\pi\sqrt{2GM_*}} \frac{2GM_*}{R_*} = \frac{\dot{M}}{4\pi} \cdot \left( \frac{2GM_*}{R_*^5} \right)^{1/2} \quad (6.26)$$

The last boundary condition concerns the protostars luminosity, which should be the sum of the accretion luminosity and internal luminosity (see last of Eq.(6.23))

$$L_* = L_{\text{acc}} + L_{\text{int}} . \quad (6.27)$$

### 6.4.1 Mass-radius relation

To understand pre-main sequence stellar evolutionary tracks, we like to follow the mass and radius of our protostar from the start. Initial conditions for integrating Eqs.(6.23) vary widely and so in the beginning, the solutions diverge. However, after the first doubling of the mass, the solutions are indistinguishable (see Fig. 6.4).

It is also important to note that low mass protostars (T Tauri stars) are fully convective (grey area) while higher mass protostars do form a radiative core. To understand this, we have to recall that the main opacity source in the photosphere of these protostars,  $\text{H}^-$ , is highly dependent on temperature (for  $T < 10^4$  K). The temperature dependence of the opacity is thus a sort of valve that regulates the surface temperature of the star. For a fixed mass and radius, there is a minimum allowed  $T_{\text{eff}}$ , called Hayashi temperature, implying also a minimum stellar luminosity. On the other hand, there is a maximum luminosity that can be transported by radiation ( $L_{\text{crit}}$ ). When the minimum stellar luminosity according to the Hayashi temperature exceeds  $\langle L_{\text{crit}} \rangle$ , the protostar becomes quickly fully convective. This is the case for low mass protostars. Higher mass protostars,  $2 - 8 M_{\odot}$ , have a luminosity that exceeds  $\langle L_{\text{crit}} \rangle$  and thus they can sustain a radiative core.

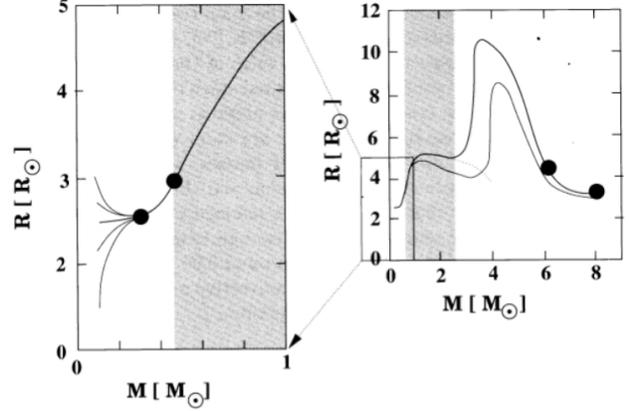


Figure 6.4: Evolution of stellar radius with mass using a constant accretion rate of  $10^{-5} M_{\odot}/\text{yr}$ . The left figure illustrates low mass pre-MS stars, the right one the entire mass range up to  $8 M_{\odot}$ . The grey area indicates where the protostars are fully convective; black dots mark the onset of deuterium burning, the point where stars are fully convective, the ignition of hydrogen in the core, and the arrival at the ZAMS. The thin line in the right panel illustrates the use of different boundary conditions (figure from Schulz 2004).

### 6.4.2 Deuterium burning

Deuterium is one of the elements generated in the Big Bang with a primordial abundance of  $[\text{D}/\text{H}] = 3 - 5 \times 10^{-5}$ . It starts burning at temperatures above  $10^6$  K. In low mass stars,  $< 1 M_{\odot}$ , mass accretion provides a steady supply of new deuterium fuel and the fusion process itself drives the protostars convection, thereby mixing the newly accreted fuel down into the core. The total amount of energy available from deuterium burning equals the gravitational energy of the protostar. Thus, deuterium regulates in that stage the mass-radius relation of the protostar (this depends on the mass accretion rate).

For higher mass stars,  $M > 2 M_{\odot}$ , the deuterium fusion depletes the core. Since new material is still accreting, the deuterium burning continues in a shell.

### 6.4.3 Lithium destruction

If the temperature in the core of the protostar reaches  $\sim 2.5 \cdot 10^6$  K, Li-burning consumes  ${}^7\text{Li}$  (the dominant isotope) through the reaction  ${}^7\text{Li}(p,\alpha){}^4\text{He}$ . The interstellar medium Li abundance is  $[\text{Li}/\text{H}] \sim 2 \cdot 10^{-9}$ , which is about an order of magnitude higher than the primordial value.

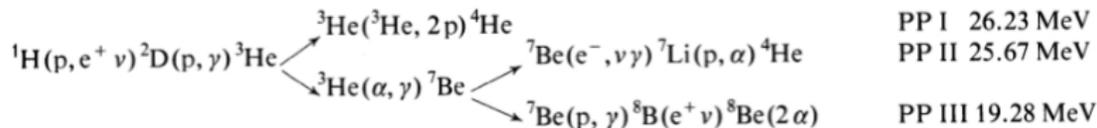
Low mass protostars never reach the central temperatures necessary for Li-burning, while more massive ones quickly exhaust their Li in the core. If the star has a mass  $< 0.9 M_{\odot}$ , it is fully convective during this period and hence destroys his entire Li reservoir. More massive stars cease to be fully convective before

all Li is consumed. Stars more massive than  $1.2 M_{\odot}$  only have shallow convection zones and hence do not deplete their Li significantly.

Measuring the Li abundance in young clusters can hence serve as a powerful test of pre-main-sequence stellar structure and evolution theory.

### 6.4.4 Hydrogen ignition

When deuterium burning ceases in the core, the star once more contracts gravitationally til it reaches a core temperature of  $10^7$  K. At those temperatures four protons can fuse into a  ${}^4\text{He}$  atom. During this process, two of the four protons decay into neutrons. The process is known as the pp-chain



At temperatures larger than  $1.5 \cdot 10^7$  K, the CNO cycle takes over converting most of the initial C, N, and O into  ${}^{14}\text{N}$ .

## 6.5 Magnetospheric accretion

Magnetic fields and rotation largely determine the accretion flow as outlined in the previous chapter (Sects. 5.2.4 and 5.2.5) and lead to the formation of an accretion disk. In the initial short collapse phase, accretion is dominated by envelope accretion onto the star. However, in later stages follows a much longer phase of disk accretion. In the latter phase, the accretion flow is predominantly through the disk onto the central protostar.

### 6.5.1 Theory

Within the accretion disk, angular momentum is transported — via some poorly understood viscosity — from the inner disk outwards to a small fraction of disk mass, so that the inner disk effectively loses angular momentum, while the outer disk gains some. This enables mass accretion from the inner disk onto the protostar and leads at the same time to a viscous spreading of the disk. Fig. 6.5 illustrates schematically how angular momentum is transported in a shearing disk. The angular velocity is decreasing outwards. If there is friction or communication between two neighboring annuli, the resulting torques will try to bring them into corotation. This accelerates gas in the outer annulus, while it decelerates gas in the inner annulus. Hence, the inner annulus loses angular momentum to the outer one. The same happens in the presence of a magnetic field, where we can illustrate the effect in form of an elastic spring connecting the two annuli. As the inner one rotates slower, the spring will be stretched out, thereby causing the inner annulus to break and lose angular momentum.

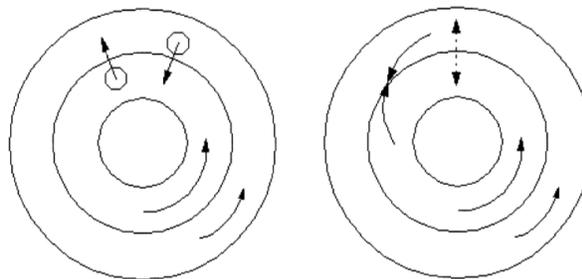


Figure 6.5: Schematic view of angular momentum transport in a shearing disk.

In a two-dimensional gaseous disk, angular momentum will be transported by turbulent motions, that is gas being mixed by turbulence among different annuli. This works as long as there is shearing orbital motion,  $d\Omega/dR \neq 0$ . Possible sources of turbulent motions are magnetic fields and convection. An alternative process for angular momentum transport are gravitational instabilities. We will get back to the details of angular momentum transport, viscosity and disks in a later chapter.

## 6.5.2 Observations

As material accretes from the disk onto the star, it hits the stellar surface, thus forming an accretion shock. Within the magnetospheric accretion model, the material flows along so-called funnel flows shaped by the connection of the stellar magnetic field with the disk (Fig. 6.6 and 6.10). The disk is truncated at a distance of several stellar radii. The accretion rates can be measured by looking at the veiling of the stellar spectrum (spectral lines) due to the accretion layer above (see also Fig. 6.9).

The stellar absorption lines will be partially filled in by the continuum emission from the hot accretion layer. By measuring this effect in a spectral line and the adjacent continuum, we can estimate the veiling  $r$  defined as the ratio between the excess  $F^E$  and stellar fluxes  $F_c^*$  as

$$r = \frac{F_l^o/F_c^o - F_l^*/F_c^*}{1 - F_l^o/F_c^o} \quad (6.28)$$

where  $l$  and  $c$  refer to the line and continuum of the veiled object ( $o$ ) or the stellar photosphere ( $*$ ) (Gullbring 1998). In this way, we can estimate the excess hot continuum emission and estimate its total luminosity  $L_{\text{hot}}$ . This luminosity is related to the accretion rate in the following way

$$L_{\text{hot}} \approx \frac{GM_*\dot{M}}{R_*} \left(1 - \frac{R_*}{R_m}\right) \quad (6.29)$$

where  $R_m$  is inner radius of the disk, the magnetospheric radius, i.e. the radius where the magnetic field of the star truncates the disk. Thus by obtaining high resolution spectra of young stars, we can estimate their accretion rates. Given the uncertainties in the details of the accretion process, these estimates should be considered as order of magnitude estimates.

## 6.6 Evolution in the HR diagram

In the previous section, we described how we can obtain the radius, mass and luminosity of the protostar (pre-main sequence star). Fig. 6.7 illustrates now the evolutionary track that pre-main sequence stars of various masses take before they reach the hydrogen burning stage, the zero-age main sequence (ZAMS, thick dashed line). Low mass stars of  $0.5 M_\odot$  take 160 Myr to reach the ZAMS, solar mass stars 50 Myr, and higher mass stars ( $3 M_\odot$ ) only 2.5 Myr.

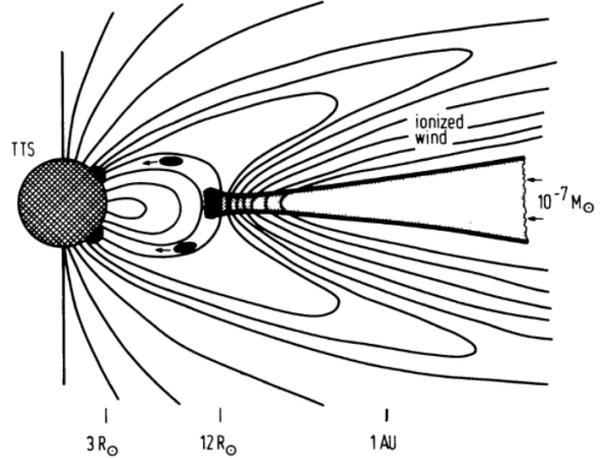


Figure 6.6: A sketch of the basic concept of magnetospheric accretion in T Tauri stars (from Camenzind 1990).

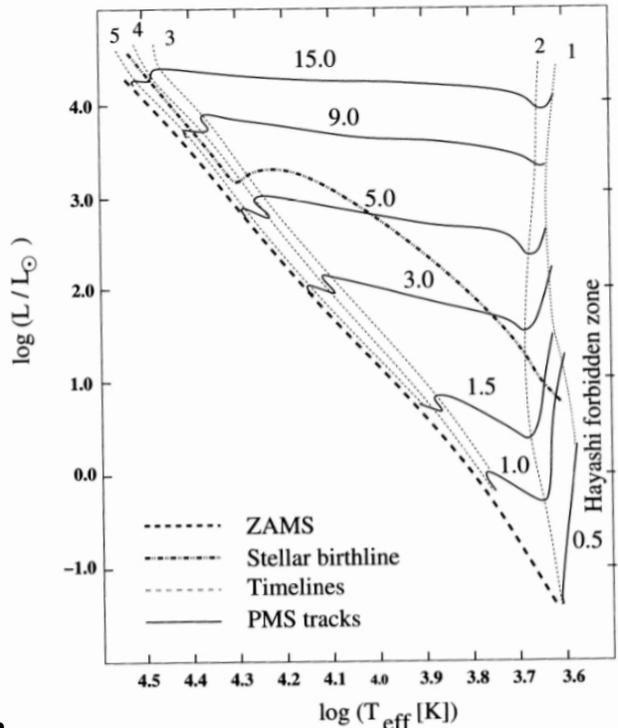


Figure 6.7: Evolutionary path in the HR diagram. Thin dashed lines indicate timelines labelled 1, 2, 3, 4, and 5. Timeline 1 is the beginning of the Hayashi track, timelines 2-5 cover the radiative track of the HR-diagram. The thick dashed line is the zero-age main sequence; the thick dash-dotted line the stellar birthline for an accretion rate of  $10^{-5} M_\odot/\text{yr}$  (figure from Schulz 2004).

### 6.6.1 Birthline

When protostars begin their quasi-hydrostatic contraction towards the ZAMS, they become optically visible. The location where this happens in the HR diagram is a well-defined line called the birthline. The star has then more or less reached its final mass. In principle, this location should coincide with the location of the youngest pre-main sequence stars observed in star forming regions. However, it depends on the assumed mass accretion rate (in the figure, this is  $10^{-5} M_{\odot}/\text{yr}$ ), with higher accretion rates shifting the line upward.

### 6.6.2 Hayashi tracks

We have encountered earlier the finding of Hayashi (1966) that a protostar of a certain radius and mass has a minimum effective temperature, the Hayashi temperature. For temperatures lower than that, the protostar is not stable and will tend to drift back to a stable configuration. The Hayashi tracks in the HR diagram mark thus asymptotes to these instabilities. As noted earlier, low mass protostars have extended Hayashi tracks, while higher mass protostars develop a radiative core and have thus very short or no Hayashi track.

## 6.7 PMS classification

Fig. 6.8 provides an overview of the various evolutionary phases of star and disk formation in the context of object classification, physical and observational properties. In the following, we discuss briefly four classes of pre-main sequence objects, three low mass object types, the classical T Tauri, weak line T Tauri, and FU Orionis stars and the intermediate mass object type, the Herbig AeBe stars.

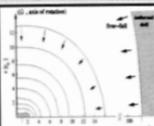
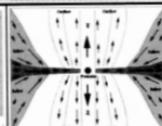
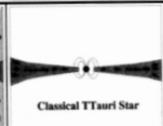
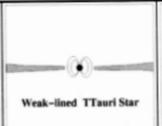
	Infalling protostar	Accreting protostar	Contracting PMS star		MS star
<b>YSO properties</b>					
<b>Phase</b>	adiabatic (A,B,C)	accretion (D) deuterium burning onset of convection	convective radiative onset of nuclear burning		convective radiative full nuclear burning
<b>Matter flows</b>	mostly infall disk & outflows form	some infall mostly accretion outflows, jets	low accretion	?	—
<b>Envelope/disk size</b>	< 10000 AU	< 1000 AU	< 400 AU	~ 100 AU	—
<b>Infall/accretion rate</b>	$10^{-4}$	$10^{-5}$	$10^{-6}$ -- $10^{-7}$	?	—
<b>Age</b>	$10^4$ - $10^5$ yr	$10^5$ yr	$10^6$ -- $10^7$ yr	$10^6$ -- $10^7$ yr	—
<b>Emission bands (except IR)</b>	thermal radio X-ray?	radio X-ray	radio optical strong X-ray	non-therm. radio optical strong X-ray	non-therm. radio optical X-ray
<b>Classes</b>	Class 0	Class I	Class II	Class III	ZAMS

Figure 6.8: The IR classification in the context with evolutionary phases and matter flow parameters (figure and caption from Schulz 2005).

### 6.7.1 T Tauri stars

Classical T Tauri stars (CTTS) are named after the prototype T Tau in the Taurus star forming region. Characteristics are strong Balmer line emission ( $H\alpha$ ), and an association with reflection nebulae and molecular clouds. These stars are low-mass pre-main sequence stars with spectral types between F and M and effective temperatures ranging from 3000 to 7000 K. They are class II sources according to their SED and the IR excess peaks between 1 and 10  $\mu\text{m}$ . The median disk mass is  $0.01 M_{\odot}$  with a huge spread around it ( $0.0003 - 1 M_{\odot}$ ). Typical disk radii are around 100 AU (see Fig 6.10 for a sketch of the star-disk system). Accretion shocks and stellar activity (coronal heating) produce strong X-rays with luminosities up to  $L_X = 10^{31} \text{ erg s}^{-1}$ . Here, accretion shocks in the free-falling flow contribute mostly to the soft X-rays, while accretion shocks on the surface only produce excess far UV radiation (see Fig. 6.9).

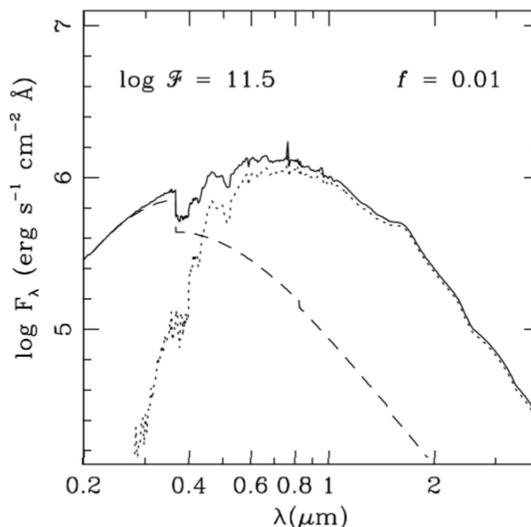


Figure 6.9: The the total spectral energy distribution (solid line) of a K7-M0 photosphere (dotted line) plus excess continuum (dashed line) for a value of  $\log F = 11.5 = \log(L_{\text{acc}}/(fR_*^2))$  and a filling factor  $f = 0.01$  (adapted from Calvet & Gullbring 1998).

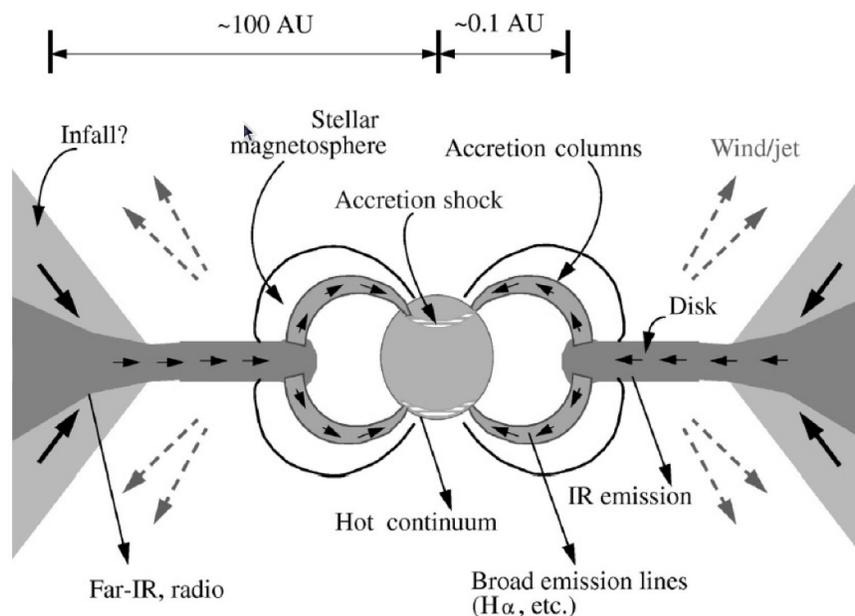


Figure 6.10: Schematic picture of accretion in T Tauri stars. The pre-main sequence star is surrounded by accreting circumstellar disk which emits at infrared, sub-mm and mm wavelength. The inner disk is disrupted by stellar magnetic field, which cause accreting material to be diverted out of the disk and fall rapidly onto the star. This magnetospheric material emits broad emission lines as it falls along the accretion columns, and produces a hot continuum when it crashed into the stellar surface at an accretion shock (figure and caption from Hartmann 1998)

### 6.7.2 Weak line T Tauri stars

Young low mass stars that are associated with star forming regions, but do not show strong  $H\alpha$  emission are generally classified as weak line T Tauri stars (WTTS). A quantitative criterium is the equivalent width of the  $H\alpha$  line to be smaller than  $10 \text{ \AA}$ . In terms of their SED classification, they often end up as class III sources due to the lack of near IR excess (absence of an inner disk). It is often assumed that they are the older counterparts of the classical T Tauri stars. Even though there may still be a tenuous disk present in these systems, there is no clear signature of accretion or optically thick dust emission. Since most of the X-ray emission from low mass stars originates from stellar activity and not from accretion, WTTS are strong X-ray sources just like the CTTS.

### 6.7.3 FU Orionis stars

FU Orionis type stars have spectral types ranging from F to G with effective temperature of  $\sim 6000\text{--}7000 \text{ K}$ . They show periodic optical outbursts of several magnitudes with a rising timescale of  $\sim 1 \text{ yr}$  and a decline timescale of  $50\text{--}100 \text{ yr}$  (see Fig. 6.11). It seems that the outbursts are connected to episodes of increased mass accretion rates from the disk (up to  $10^{-4} M_{\odot}/\text{yr}$ ), possibly related to dynamical instabilities (magneto-rotational instability). A more extensive description of the nature of these objects, the underlying accretion physics and observational properties can be found in Hartmann (1998).

### 6.7.4 Herbig AeBe stars

The stars have intermediate masses between  $2$  and  $8 M_{\odot}$ . They are the more massive counterparts of the T Tauri stars. The current definition can be summarized as stars with spectral types between B0 and early F, luminosity classes II to V, having emission lines (e.g. strong  $H\alpha$ ) in their spectra and showing an infrared excess due to the presence of circumstellar material. The pre-main-sequence evolutionary timescales of Herbig AeBe stars (also often called HAEBE stars) is much shorter than that of T Tauri stars. In the early phases, these stars are not fully convective and undergo thus a phase of thermal relaxation. Hence, they mostly lack the Hayashi part of the evolutionary track and they appear underluminous in the HR diagram. Herbig AeBe stars initially move upward in the HR diagram due to their expansion (see also right hand side of Fig. 6.4). Above stellar masses of  $4 M_{\odot}$  stars are already born fully radiative.

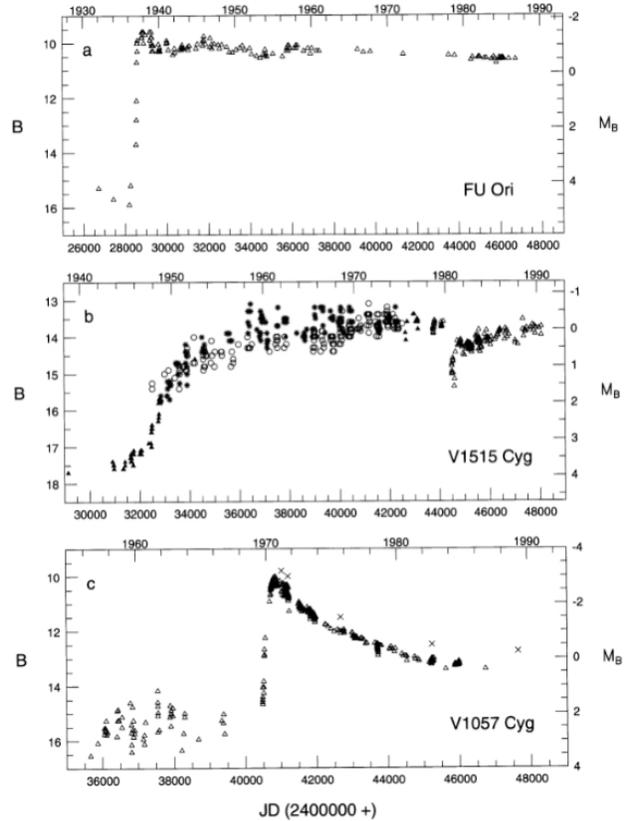


Figure 6.11: Optical (B) photometry of outbursts in three FU Ori objects (figure and caption from Hartmann 1998).

# Chapter 7

## The formation of massive stars

Stars with masses larger than  $\sim 8\text{--}9 M_{\odot}$  reach the main sequence before they become optically visible and before they have ended the main accretion phase. Note indeed in Fig. 6.7 that the birthline merges into the ZAMS at these masses. At all times in the formation process such objects are therefore severely obscured and hard to study observationally.

It is an interesting coincidence (but not more than that) that the mass limit of  $\sim 8\text{--}9 M_{\odot}$  coincides with a major change in the post-main sequence evolution of stars; such objects do not develop degenerate cores and therefore end their lives in core-collapse supernovae. Stars above  $8\text{--}9 M_{\odot}$ , i.e. stars that start out their lives as B2 stars or earlier, are referred to as *massive stars*. The most reliable mass measurements of massive stars are from those in binary systems with dynamically measured masses. In the Galaxy the most massive binaries currently known are NGC3603-A1 ( $116 \pm 31$  and  $80 \pm 16 M_{\odot}$ ; Schnurr et al. 2008), WR20a ( $83 \pm 5$  and  $82 \pm 5 M_{\odot}$ ; Rauw et al. 2005; Bonanos et al. 2004), and WR21a ( $104 \pm 10$  and  $58 \pm 4 M_{\odot}$ ; Niemela et al. 2008; Tramper et al. 2016). From analysis of the intrinsically brightest stars in massive clusters it appears that even more massive stars exist. The current record holders are located in the central cluster of the 30 Doradus region, Radcliffe 136a (R136a; see Table ?? and Fig. ??). For objects R136a1, a2, and a3, Crowther et al (2010) estimated masses of  $265^{+80}_{-35}$ ,  $195^{+35}_{-35}$ , and  $135^{+25}_{-20} M_{\odot}$ , respectively (see Fig. 7.1). Probably these stars are about 1–2 Myrs old and have lost part of their initial mass in an intense outflow or stellar wind. Correcting for this, the initial masses of the stars were probably  $320^{+100}_{-40}$ ,  $240^{+45}_{-45}$ , and  $165^{+30}_{-30} M_{\odot}$ . Note though, that these masses are much more uncertain than those derived from the orbital motion of binaries. How could such massive stars have formed?

Two main routes towards the formation of massive stars have been proposed: (1) formation via gas accretion, i.e. the same mechanism by which stars of low mass form, and (2) formation via collisions between lower mass stars, which would represent a very different formation mechanism from the bulk of the stellar population. We discuss both mechanisms.

### 7.1 The formation of massive stars by accretion

There are several major challenges to forming massive stars by accretion, which we will discuss below: fragmentation, mass-accretion rate, radiation pressure, ionization feedback, and stellar winds. We discuss each of these challenges below.

#### Fragmentation

Equation (??) provides a convenient expression for the Jeans criterion, i.e. the minimum mass necessary to initiate the collapse of a cloud due to its self-gravity. To aid the discussion, we repeat it here

$$M_J = 7.41 \left( \frac{T}{10 \text{ K}} \right)^{3/2} \left( \frac{2}{\mu} \right)^2 \left( \frac{10^4}{n} \right)^{1/2} M_{\odot}. \quad (7.1)$$

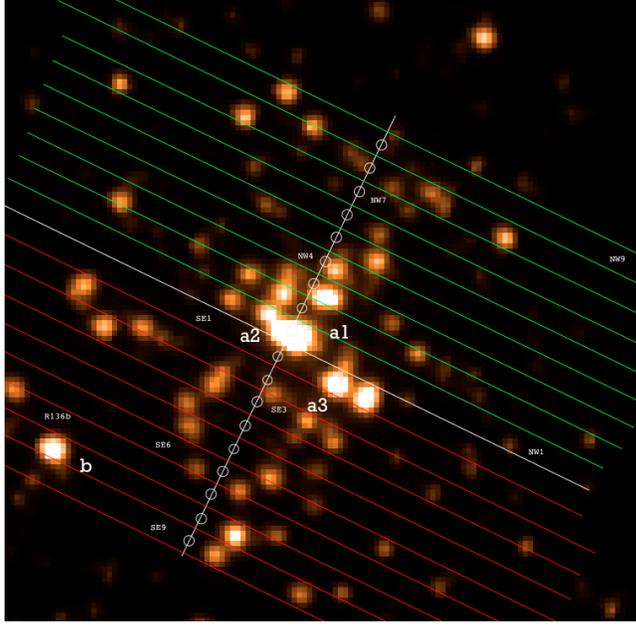


Figure 7.1: The central region Radcliff 136a of the 30 Doradus region in the Large Magellanic Cloud observed with STIS/HST. The position of stars a1, a2, and a3 are indicated. Notice that crowding effects are severe in this region and that disentangling methods need to be applied to isolate the light from each identified source. The initial masses of these stars are about  $320$ ,  $240$ , and  $165 M_{\odot}$ , identifying them as the most massive stars known. From: Paul Crowther (private communication).

For a typical dark cloud with  $T = 10$  K,  $\mu = 2.3$ , and  $n = 10^4 \text{ cm}^{-3}$  (see also Table ??)  $M_J = 5.6 M_{\odot}$ . Clearly, a larger cloud with similar local conditions would fragment immediately. What processes, then, can increase the Jeans mass, i.e. reduce the fragmentation? Fragmentation is reduced by two primary effects: radiation feedback and magnetic fields.

Radiation feedback works to reduce fragmentation by heating the gas, raising its pressure and thus its Jeans mass. For instance, if the temperature of the medium would be increased to  $200$  K, the Jeans mass equation would imply  $M_J \sim 500 M_{\odot}$ . We address the question of how to heat the medium below, when discussing the mass-accretion rate.

Magnetic fields limit fragmentation in two ways. First, they remove angular momentum. In a collapsing cloud, the densest regions have a shorter free fall time (Eq. 6.1) and collapse fastest. As the gas falls inward it will rotate faster and faster in order to conserve angular momentum. When the collapsing gas is threaded by a magnetic field, the resulting differential rotation between inner collapsing regions and outer ones twists the magnetic field lines. The twist produces a magnetic tension force (see Eqs. A.30 and ??) that transfers angular momentum from the inner to the outer regions, a process known as *magnetic breaking*. For an axi-symmetric flow one can show (e.g. Stahler & Palla 2005) that the time rate of change of the angular momentum of a fluid element at a distance  $r$  from the rotation axis due to magnetic forces is given by

$$\frac{\partial J}{\partial t} = \frac{1}{4\pi} \left[ B_r \frac{\partial}{\partial r} (r B_{\phi}) + r B_z \frac{\partial}{\partial z} B_{\phi} \right], \quad (7.2)$$

where  $\mathbf{B} = (B_r, B_{\phi}, B_z)$  is the magnetic field, and we have used cylindrical coordinates (therefore, beware of the definition of  $r$ ). For the types of magnetic field configuration produced by collapse, the net effect is to transport angular momentum outward.

This process inhibits the formation of rotationally-flattened structures such as accretion disks. This is

of consequence from the standpoint of fragmentation, because rotational flattening raises the density of the gas as it approaches the star, and dense, rotationally-flattened structures are vulnerable to the Toomre instability (see Eq. ??), in which the self-gravity of a flattened rotating structure overcomes support from thermal pressure and angular momentum, leading to fragmentation and collapse.

Second, magnetic fields provide extra pressure support that prevents regions from collapsing unless their mass is higher than the *magnetic critical mass* (see also question ???.??)

$$M_{\text{mcm}} \sim \left(\frac{5}{9G}\right)^{1/2} BR^2 = \left(\frac{5}{9\pi^2 G}\right)^{1/2} \phi, \quad (7.3)$$

where  $\phi$  is the magnetic flux. Observations indicate that star-forming cores, over a wide range of size and density scales tend to have  $\phi/M_{\text{mcm}}$  values that range between 0 and  $(\phi/M_{\text{mcm}})_{\text{crit}}$ . So, magnetic fields by themselves seem unable to prevent collapse. However, magneto-hydrodynamic simulations (by Hennebelle et al. 2011) do indicate that the support they do give leads to a reduction of the number of fragments by a factor of  $\sim 2$  for realistic levels of magnetization.

## Mass-accretion rate

The Kelvin-Helmholtz timescale Eq. (6.2) associated with a  $25 M_{\odot}$  main-sequence star of radius  $9 R_{\odot}$  and luminosity  $10^5 L_{\odot}$  is only  $2 \times 10^4$  yr. This is much shorter than the free-fall time Eq. (6.1) – typically a few times  $10^5$  yrs, consequently such massive stars reach the zero-age main sequence before finishing their mass build-up. The star can not grow faster (in mass) than the timescale set by free-fall of gas from the envelope toward the inner regions. What kind of mass-accretion rates are needed to form a massive star on a free-fall time scale? Using Eq. (6.3) we find that for a free-fall time of  $2.5 \times 10^5$  yr, an average mass-accretion rate of  $10^{-4} M_{\odot} \text{yr}^{-1}$  is needed to compile our  $25 M_{\odot}$  star. In principle, such high accretion rates can be accommodated if the temperature of the cloud is high, see the mass-accretion rate Eq. (5.9) in the Shu inside-out collapse model. Using this equation we find that a  $T \sim 160$  K is needed to achieve the high mass-accretion rate. The high temperature is associated with a high sound speed, i.e. the high  $\dot{M}_{\text{acc}}$  is possible because the speed at which material is accreted is high. So, what may heat the medium to such high temperatures? One heating source may be the early feedback provided by low-mass stars, whose luminosities are dominated by accretion rather than internal energy generation. Even a  $1 M_{\odot}$  star accreting at the relatively high rates expected in the dense regions where massive stars form could radiate strongly enough to raise the gas temperature by a factor of a few at distances of  $\sim 1000$  AU. Not only does this raise the minimum mass for gas to fragment by a factor of  $\sim 10$  (Krumholz 2006), it also increases the mass infall rate by a factor of  $\sim 10$ . Though low-mass stars may provide considerable heating, temperatures above  $\sim 100$  K are only observed – though not common – in very massive star forming regions *after* one or more massive stars have formed. Therefore it is these objects that are thought to be responsible for heating the medium to such high temperatures. Perhaps this indicates that the first or first few massive stars in a cluster are formed in a completely different way, later facilitating the formation of other massive stars through the inside-out collapse mechanisms that builds lower mass stars.

Can the central object accept all of the cloud material that is falling in? Because of conservation of angular momentum the material first settles in a disk-like structure through which material spirals in and is accreted by the central object (disks are the topic of chapter ??). For a steady disk with dimensionless Shakura & Sunyaev viscosity  $\alpha$  (Eq. ??), the accretion rate through the disk is (e.g Kratter et al. 2010)

$$\dot{M} = \frac{3\alpha a^3}{GQ} = 1.5 \times 10^{-4} \frac{\alpha}{Q} \left(\frac{T}{100 \text{ K}}\right)^{3/2} M_{\odot} \text{yr}^{-1}, \quad (7.4)$$

where  $a$  is the isothermal sound speed, and we have assumed a mean molecular weight  $\mu = 2.3$ .  $Q$  is the Toomre parameter (see Eq. ??). The Shakura & Sunyaev viscosity in the disk cannot produce  $\alpha > 1$ , and the disk cannot be gravitationally stable if  $Q < 1$ , so the accretion rate through a gravitationally-stable disk

is going to be  $\sim 10^{-4} M_{\odot} \text{yr}^{-1}$  only if the disk temperature is  $\sim 100 \text{ K}$ . Notice though that if the infalling cloud material has a high enough temperature to transport  $\sim 10^{-4} M_{\odot} \text{yr}^{-1}$  one may expect that the disk can also carry this amount, as it is not likely that the disk temperature will be substantially lower than the cloud temperature.

The process of disk heating to allow more mass to accrete has a limit: once the temperature required to stabilize the disk exceeds the dust sublimation temperature, it will not be easy to heat the disk further, and this may result in an instability so violent that the disk fragments entirely, halting further accretion (Kratter & Matzner 2006).

## Radiation pressure

The third potential problem in forming massive stars is that of radiation pressure. Material is falling in because it is subject to the gravitational acceleration  $g_{\text{N}} = GM/r^2$ . Radiation pressure is exerting an outward directed force on the in-falling material, more specifically, on the dust. This force is given by

$$g_{\text{rad}} = \frac{1}{c} \int_0^{\infty} \kappa_{\nu} F_{\nu} d\nu = \frac{\kappa}{c} \int_0^{\infty} F_{\nu} d\nu = \frac{\kappa}{c} F_{\text{tot}} = \frac{\kappa}{c} \frac{L}{4\pi r^2}, \quad (7.5)$$

where

$$\kappa = \frac{\int_0^{\infty} \kappa_{\nu} F_{\nu} d\nu}{\int_0^{\infty} F_{\nu} d\nu} \quad (7.6)$$

is the flux mean opacity in  $\text{cm}^2 \text{gr}^{-1}$ .  $F_{\nu}$  is the flux and  $F_{\text{tot}}$  is the total (i.e. frequency integrated) flux. Since the radial dependence for  $g_{\text{N}}$  and  $g_{\text{rad}}$  is the same, the net force will be inward only if

$$\frac{L}{M} < \frac{4\pi Gc}{\kappa} = 2500 \left( \frac{\kappa}{5 \text{ cm}^2 \text{ g}^{-1}} \right)^{-1} \frac{L_{\odot}}{M_{\odot}}. \quad (7.7)$$

All stars above  $\sim 20M_{\odot}$  have  $L/M > 2500L_{\odot}/M_{\odot}$ , so the question arises: why doesn't radiation pressure expel circumstellar material and prevent stars from growing to masses substantially larger than  $\sim 20M_{\odot}$ .

In the derivation of the infall criterium Eq. 7.7 it is implicitly assumed that the material falls in radially, i.e. that infall is a 1D phenomenon. However, it is inherently a 3D phenomenon. For instance, in a medium in which the dust (and the gas) is 'clumped', part of the photons created by the central star and disk pass through the envelope without being able to exert radiation pressure on the dust grains. In an initially turbulent cloud, Rayleigh-Taylor instabilities may produce optically thick 'fingers' of high optical depth material that can not be stopped by radiation pressure. Though this may be helpful to let matter pass through the envelope, it appears improbable that such a porous medium can exist in the inner region of the accretion disk, where matter is concentrated. Near the star however, an optically thick circumstellar disk will 'shield' the disk material from radiation pressure. In a sense, the disk beams the radiation field in the polar directions, weakening the radiation pressure on the disk but creating an outflow cavity or chimney in those polar directions through which radiation escapes.

## Ionization feedback

A fourth potential problem in forming massive stars is photo-ionization. Galactic molecular clouds generally have escape speeds below  $10 \text{ km sec}^{-1}$  (e.g. Heyer et al. 2009), the sound speed in  $\sim 10^4 \text{ K}$  gas – i.e. in gas where hydrogen is ionized. As a result, if the gas in a star-forming region becomes ionized, the gas pressure may drive a thermal wind that will choke off accretion. This process is thought to be a major factor in limiting the star-forming efficiency of giant molecular clouds.

However, it is much less clear whether photo-ionization can limit the formation of individual massive stars. Two arguments suggest that it may not be an issue. First, Walmsley (1995) noted that the higher the mass inflow rate, the higher the density of matter around the star, and thus the higher the recombination rate and the smaller the Strömgen radius. If the radius of the ionized region is small enough that the escape

speed from its outer edge is  $> 10 \text{ km sec}^{-1}$ , then photo-ionized gas will not be able to flow away in a thermal wind and escape. For an accretion flow in free-fall onto a star, this criterium will be matched if the accretion rate satisfies

$$\dot{M}_{\text{acc}} > \frac{8\pi\mu GMQ_{\circ}}{2.2\alpha_B \ln(v_{\text{esc}}/a)} \sim 4 \times 10^{-5} \left(\frac{M}{100 M_{\odot}}\right)^{1/2} \left(\frac{Q_{\circ}}{10^{49} \text{ s}^{-1}}\right)^{1/2} M_{\odot}\text{yr}^{-1}, \quad (7.8)$$

where  $M$  is the stellar mass,  $Q_{\circ}$  is the number of hydrogen ionizing photons emitted by the star per second,  $\alpha_B$  is the case B recombination coefficient, and  $v_{\text{esc}}$  is the escape speed from the stellar surface. The factor 2.2 in the denominator arises from the assumption that He is singly ionized. The second right hand side uses  $v_{\text{esc}} = 1000 \text{ km sec}^{-1}$  and  $a = 10 \text{ km sec}^{-1}$ . Thus an accretion rate of  $\sim 10^{-4} M_{\odot}\text{yr}^{-1}$  is sufficient to allow continuing accretion. Given the dense, compact environments in which massive stars appear to form, such high accretion rates are entirely expected.

The second argument concerns the natural 3D geometry of the process. The accretion disk and dense filaments have very high recombination rates (because they are dense) and thus resist being photo-ionized. In low-density polar regions this self-shielding does not work. The gas will become photo-ionized and is blown out, feeding back into the interstellar medium. However, the bulk of the material is contained in dense regions, where it continues to accrete. The net effect is thus an increasing mass of the central object.

## Stellar winds

The final potential challenge for the formation of massive stars by accretion is stellar winds. Once the surface temperature of stars exceed  $\sim 25\,000 \text{ K}$ , they begin to develop super-sonic radiatively driven outflows. Zero-age main sequence stars reach this temperature at a mass of  $\sim 40 M_{\odot}$ , and stars of this mass have such short Kelvin-Helmholtz timescales that, even if they are rapidly accreting, their radii and surface temperatures during formation are likely to be close to their ZAMS values (Hosokawa & Omukai 2009). The momentum by these winds is about half that of the stellar radiation field, and so if the direct stellar radiation field cannot stop accretion then the momentum carried by stellar winds will not either.

However, winds might yet be important, because the wind terminal velocity (reached already close to the surface) is quite large,  $\sim 1000 \text{ km sec}^{-1}$ . As a result, when the winds shock against the dense accretion flow, their post-shock temperature can be  $> 10^7 \text{ K}$ . This is well past the peak of the cooling curve (Weaver et al. 2007), and thus can not cool efficiently by emitting radiation. Should it become trapped, this hot gas could exert a pressure that is far greater than what would be suggested by its launch momentum. If this were to happen, it is possible that the stellar wind gas might be able to interfere with accretion. Again, we should consider possible 3D effects, i.e. clumps and filaments, that allow the (shocked) wind gas to leak out through openings in the surrounding gas rather than becoming trapped and building up a large pressure. If the medium is sufficiently porous, it is unlikely that stellar winds set significant limits on the masses to which stars can grow by accretion.

## Accretion models for massive star formation

In section ?? we introduced two variants of massive star formation theories, *monolithic collapse* and *competitive accretion*.

In the *monolithic collapse* model single molecular cloud cores of  $\sim 0.1 \text{ pc}$  would result in the formation of a single, or perhaps a low-order multiple, massive star provided that the mass of the core is large enough. The cloud core has presumably formed by condensation from a much larger molecular cloud with a mass of thousands of solar masses, but once the core has formed, the resulting collapse is little influenced by the remainder of the material in the clump. Thus low-mass stars would form from individual low-mass cores, and high-mass stars from high-mass cores. The entire mass of the core would not necessarily go into the final star; feedback processes may cause the final mass to be less than the core mass.

The second model, *competitive accretion*, is based on the premise that star formation is controlled not by collapse of individual cores, but by the overall collapse of a much larger region, containing initially gas

with several thousand  $M_{\odot}$  (or more). Individual fragments (cores) form at low mass and compete for the accretion of the remaining gas; also there can be interactions between the various fragments. One might expect a massive star to form in the center, since as the cloud collapses and develops a gravitational potential well, a considerable amount of gas can be funnelled toward the center. Massive (proto-)stars hence compete quite successfully for the available gas, and to some extent deprive lower mass siblings from the gas needed to grow as well.

We now introduce a third flavour of massive star formation, fragmentation-induced starvation.

*Fragmentation-induced starvation* is, to some extent, an added physical consideration to the two theories mentioned above. Its basic principle is that the star accretes from a rotating gravitationally unstable disk-like structure. Gravitational instabilities in such a disk lead to fragmentation of the accretion flow and the formation of companions around the central massive star. These companions subsequently compete with the initial high-mass star for the same common gas reservoir and limit its mass growth. In other words: they starve the central star of in-falling material. This is in contrast to the competitive accretion model, where the massive stars are never hindered in growth by the low-mass stars in the cluster.

Figure 7.2 shows results from a 3D radiation-hydrodynamical simulation of massive star formation that includes heating by ionising and non-ionising radiation by Peters et al. (2010, ApJ 725, 134) for a volume of size  $0.24 \text{ pc} \times 0.24 \text{ pc} \times 0.015 \text{ pc}$ . One ‘observes’ the continuous formation of new sink particles (i.e. proto-stellar companions) in an expanding region around the central star. The heating by stellar radiation can suppress instability locally, but shielding by the dense filaments prevents the whole disk from becoming stable and restricts the heating to small regions near the center of the disk that are surrounded by filaments. This shielding makes it possible for star formation to progress radially outward despite heating by the stars. The disk remains sufficiently cool at the inner edge for gravitational instability to set in and star formation to proceed inside-out in the disk plane. So, the effect of the filamentary structures in the disk is twofold: (1) they are so dense that they render the disk unstable locally, and (2) because of their high density, they can effectively shield the radiation from the stellar configuration (the primary star and its companions) near the center of the disk, so that radiative heating does not stabilise the outer-parts of the disk.

In the first frame at 0.621 Myr the disk contains about  $35 M_{\odot}$  of gas. The total disk mass remains about constant while star formation proceeds radially outward. To illustrate the tendency of star formation to occur at increasingly larger disk radii, the left panel of Fig. 7.3 shows the disk radii at which new sink particles form as a function of time. Accretion heating in the initial phase suppresses sink formation at small radii. The companions slowly spiral outward with time, so that at  $\tau \sim 0.67 \text{ Myr}$  their radiation can be shielded by filaments in the disk. Within these filaments, the gas then cools until local collapse sets in and two sink particles form near the center of the disk. Once the filament has dissolved, the gas heats up again and no further sink particles form in the inner disk region. Eventually, companions can be found up to  $\sim 20\,000 \text{ AU}$  from the first source. The simulation produces a total of 25 sinks for a combined mass of  $125.6 M_{\odot}$  of which the most massive star has a mass  $M_{\text{max}} = 23.4 M_{\odot}$ . The individual accretion history of the sink sources is shown in the right panel of Fig. 7.3.

## 7.2 The formation of massive stars by collision

Another quite drastic scenario is that gravitational collapse indeed does not lead to the formation of objects more massive than some ten solar masses, and that more massive objects are the result of coalescence of low- and intermediate-mass stars. Some support for this idea could come from the observation that massive stars do originate in fairly dense clusters and that stars are seen to ‘run away’ from young clusters. The latter may be the result from dynamical ejection when two stars experience a close encounter and in which typically the less massive star is ejected<sup>1</sup>.

The problem with coalescence is that it requires very high stellar densities to be effective. In a system containing stars of mass  $M_{\star}$  and radius  $R_{\star}$  that has a stellar density  $n_{\star}$  and where the velocity distribution

<sup>1</sup>In older clusters stars may be ejected from binary systems in which the initially most massive component experiences its supernova explosion.

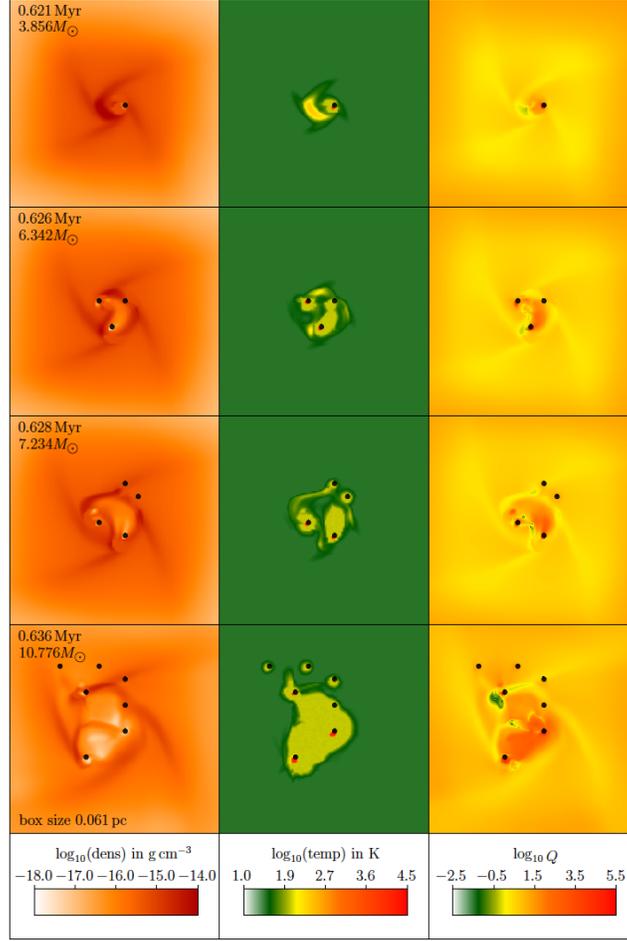


Figure 7.2: The initial phases of disk instability in the fragmentation-induced starvation scenario. The panels show slices of density, temperature, and Toomre  $Q$ -parameter (see Eq. ??) in the disk plane for four different times. Each frame shows the simulation time and the mass of the most massive star in the simulation. The black dots indicate the positions of sink particles. The stellar radiation initially heats up the disk locally, which enhances the stability. The dense filaments, however, shield the radiation, and the cold material within and behind the filaments becomes unstable again. From: Peters et al. 2010, ApJ 725, 134.

of the objects is according to Maxwell with a 1D velocity dispersion  $\sigma$ , the collision time is given by (e.g. Binney & Tremaine 1987, equation 8-125 or Freitag et al. 2006, MNRAS 368, 121, equation 3)

$$\tau_{\text{coll}} \simeq 2.1 \times 10^{12} \left( \frac{10^6 \text{ pc}^{-3}}{n_{\star}} \right) \left( \frac{\sigma}{30 \text{ km s}^{-1}} \right) \left( \frac{R_{\odot}}{R_{\star}} \right) \left( \frac{M_{\odot}}{M_{\star}} \right) \text{ yr.} \quad (7.9)$$

Inserting values typical for the central cluster Radcliffe 136 of the 30 Doradus region (see figure 7.1), where  $n \sim 1.5 \times 10^4$  (Selman & Melnick 2013, A&A 552, A94, adopting a typical core cluster size of 0.4 pc) and  $\sigma \sim 5 \text{ km sec}^{-1}$  (Hénault-Brunet et al. 2012, A&A 545, L1), and pretending all stars have a mass  $M_{\star} = 1 M_{\odot}$  and radius  $R_{\star} = 1 R_{\odot}$ , yields  $\tau_{\text{coll}} \simeq 2.5 \times 10^{13} \text{ yr}$ . At the present-day evolutionary state of the cluster collisions are thus negligible.

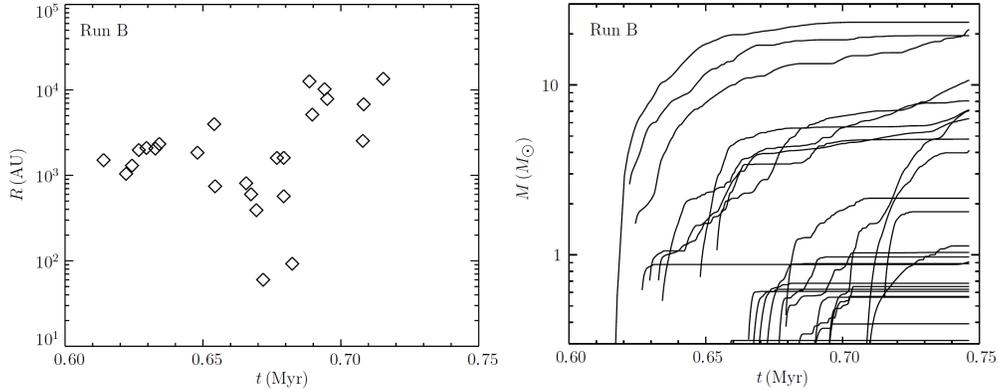


Figure 7.3: *Left.* Radius of sink formation as a function of time for the fragmentation-induced starvation scenario. Sink formation gradually occurs at larger disk radii. The accretion heating by the first stars to form suppresses sink formation at small disk radii until relatively late in the system evolution. *Right.* Individual accretion history of the sink sources, showing the stellar masses as a function of time for all sink particles. From: Peters et al. 2010, ApJ 725, 134.

## Gas accretion-driven collision models

The situation may, however, have been quite different early on in the cluster evolution. Right now, most of the gas that did not end up in stars has been blown out. This may have expanded the size of the core by a factor of ten, implying a  $10^3$  shorter collision time during the early phases of formation. The velocity dispersion will, however, have been higher, with typical values of tens of  $\text{km sec}^{-1}$ . Though these considerations help to bring down the collision time, the collision efficiency would be much higher if the merging would happen during an earlier phase of evolution, i.e. during the stage when the future stars are still clumps (of size  $\sim 10 \text{ AU} \simeq 2000 R_\odot$ ) or are surrounded by dense proto-stellar disks (of size  $\sim 100 \text{ AU} \simeq 20000 R_\odot$ ). Taking all these considerations into account, the collision time may be reduced to  $\sim 0.5\text{--}5 \text{ Myr}$ . This is still a long time interval, as for stellar mergers to be efficient in making massive stars, the timescale  $\tau_{\text{coll}}$  should be of the order of  $0.1\text{--}1 \text{ Myr}$ . Perhaps a further reduction of  $\tau_{\text{coll}}$  can be achieved if the competition for gas in the context of competitive accretion is also considered. This would allow massive cores or proto-stars to grow faster, becoming more massive prior to merging events. The mass scaling in Eq. 7.9 may then imply a reduction of the collision time by another factor a few.

So perhaps a combination of merging, starting already in the early core phase, and competitive accretion may be the road towards massive star formation. It seems likely to assume that near the end of the merger phase some, or many, spiraling, massive stars do not complete their merger, but remain in tight orbits. Indeed, it has been established that the fraction of close binaries (with periods less than 1500 days) among massive stars is high – about 70 percent – and that many of these binaries are very compact (Sana et al. 2012, Science 337, 444). About 1/3th of these binaries orbit each other in less than six days when arriving on the main sequence. These periods are so short that these stars will merge during their lifetime as a result of stellar evolution. Indeed, the formation process of massive stars may start in the cloud core phase and continue all the way into the main sequence phase.

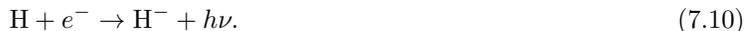
## 7.3 Formation of the First Stars

The discussion in this section closely follows and summarizes the excellent and accessible review by Bromm et al. (2009, Nature 459, 49).

The cold dark-matter (CDM) model of cosmic evolution posits that structure grew hierarchically, such

that objects of small size formed first and then merged to form increasingly larger systems. Within this model dark-matter ‘mini-haloes’ formed first through the action of gravity alone, probably a few hundred million years after the Big Bang (corresponding to redshifts  $z \approx 20\text{--}30$ ) in an era referred to as the ‘dark ages’. In simulations of structure formation these dark matter mini-haloes, with virial temperature  $T \sim 1000$  K and mass  $M \sim 10^6 M_\odot$ , are identified as the sites where the first stars formed. They correspond to  $3\sigma$  to  $4\sigma$  peaks in the cosmic density field. Such high-density peaks are expected to be strongly clustered, suggesting that feedback effects from the first stars are important in determining the fate of the surrounding primordial gas clouds. It is for this reason that two classes of population III stars, i.e. stars composed of no elements heavier than helium (ignoring the small amount of lithium produced in baryogenesis), are defined. First generation *population III.1 stars* form from initial conditions determined entirely by cosmological parameters, and second generation *population III.2 stars*, which originate from material that was influenced by nearby population III.1 stars.

For star formation to begin a sufficient amount of gas must accumulate in a dark matter mini-halo. This gas was essentially atomic in nature, with only small traces of molecular hydrogen. These molecules can be produced via a sequence of reactions, starting with the formation of the hydrogen anion  $\text{H}^-$  by radiative association



This is followed by the formation of  $\text{H}_2$  by associative detachment



This channel toward  $\text{H}_2$  production is, however, not efficient and a fractional abundance of  $\text{H}_2$  of  $10^{-5}$  is expected. Hydrogen or helium atoms can not efficiently cool the primordial gas by emitting radiation because they have excitation energies that are high, the primary cooling channel being collisional excitation of the hydrogen  $1s \rightarrow 2p$  transition followed by an emission of a  $\text{Ly}\alpha$  photon. The minimum temperature that can be reached in this way is  $\sim 10^4$  K. The lowest accessible rotational energy level of  $\text{H}_2$  has an excitation energy of 510 K, implying that no cooling mechanism is available to lower the temperature of the gas below a few hundred Kelvin. This is a fundamental difference with gas clouds in the local universe, where dust particles and molecules such as CO can cool the gas to temperature of  $\sim 10$  K.

The minimum mass at the onset of collapse is determined by the Jeans mass. Rewriting Eq. ?? to suit our needs, we get

$$M_J \sim 1000 \left( \frac{T}{200 \text{ K}} \right)^{3/2} \left( \frac{10^4}{n} \right)^{1/2} M_\odot, \quad (7.12)$$

where we have adopted  $\mu = 1.3$ . Note that the critical density of the para- $\text{H}_2$  ( $J = 2 - 0$ ) transition (of 510 K) – above which cooling becomes less efficient – is only  $10 \text{ cm}^{-3}$ , suggesting Jeans masses of thousands of solar masses. Whether the collapse leads to the formation of a single object or a multiple systems is presently unclear. Simulations based on self-consistent cosmological initial conditions seem to favor a single object, however those starting from non-cosmological initial conditions have yielded multiple cloud cores. Simulations assuming a high initial degree of spin lead to disk formation and its subsequent break-up.

Although we can fairly accurately estimate the mass triggering the first runaway collapse, it does not really tell us what the mass of the forming star(s) will be. The final mass is determined by both the mass of the cloud of which it forms and by the efficiency of feedback processes that occur during the evolution of the proto-star. These feedback processes may differ in several important ways in massive primordial stars from those of massive stars in the present-day universe. First, primordial gas does not contain dust and therefore radiative forces on the gas are much weaker. Second, it is generally assumed that magnetic fields are not important in primordial gas. Such fields in the present-day universe help to drive powerful outflows and remove angular momentum (see section 7.1). Third, primordial stars are much hotter than contemporary stars of the same mass, resulting in significantly greater ionizing luminosities.

Once the first massive proto-star is formed, be it in isolation or as part of a multiple system, its far-ultraviolet radiation in the Lyman-Werner bands (in the range of 11.2 to 13.6 eV) will photo-dissociate all the  $\text{H}_2$  in the parent gas cloud. This reduces the cooling rate, but simulations do not indicate that this stops

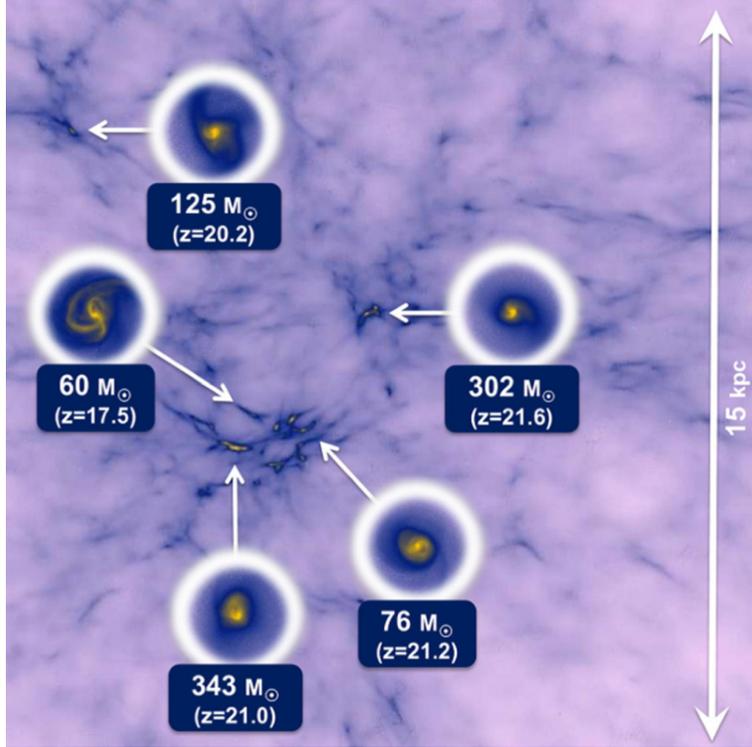


Figure 7.4: Projected gas density distribution at redshift  $z = 25$  in a radiation hydrodynamical simulation of primordial star formation in a cosmological context. Five primordial star-forming clouds in a cube of 15 kpc on a side are shown. The circles provide a zoom-in to the central 1 pc region of the clouds at the respective formation epochs. The redshifts of these epochs and the masses of the first stars formed in these clouds are given in the blue boxes. From: Hirano et al. 2014, ApJ 781, 60.

the accretion. Radiation pressure (in the Lyman- $\alpha$  line) can reverse the infall in the polar regions once the proto-star has grown to  $20\text{--}30 M_{\odot}$ , but cannot significantly reduce the accretion rate. Once the proto-star reaches  $50\text{--}100 M_{\odot}$ , the expansion of the H II region can reduce the accretion rate, but accretion can continue in the equatorial plane. It is ultimately photo evaporation-driven mass loss from the disk that is thought to halt the accretion and fix the final mass of the stars. For reasonable conditions simulations favor final masses of  $10\text{--}1000 M_{\odot}$ , depending on the entropy and angular momentum of the natal cloud and the mass-accretion rate (e.g. Hirano et al. 2014; see also Fig. 7.4).

### The role of the first stars in the formation of the first galaxies

Cosmological simulations aimed at producing the first galaxies quite generally show that the first stars form before the first galaxies do, and that feedback effects from these stars play a key role in determining the initial conditions for the formation of the first galaxies and of subsequent star formation in these primordial galaxies. Lyman-Werner radiation and hydrogen-ionizing radiation emitted by the first stars may have heated and pressurized the ambient medium (on kpc scales), reducing the density, and overall delaying the formation of the second generation of stars by as much as 100 Myr. If radiation driven winds of population III stars are weak and mass eruptions are absent, such that the pre-supernova masses remain high, stars with initial masses in the range  $25 \lesssim M \lesssim 140 M_{\odot}$  and  $M \gtrsim 260 M_{\odot}$ , end their lives by collapsing into black holes with relatively little ejection of heavy elements. Population III stars in the range  $140\text{--}260 M_{\odot}$  explode

as pair-instability supernovae, leaving no compact remnant. The nucleosynthetic yields of such explosions, defined as the heavy-element mass fraction, may reach 0.5. If this leads to metallicities in the surrounding gas in excess of a critical value of  $Z_{\text{crit}} \sim 10^{-4} Z_{\odot}$  it may be enough to increase the cooling efficiency such that low-mass (population II) stars may form in a next event of star formation.

It is possible that at least one primordial star has formed in a region that is destined to eventually become a first galaxy. However, current simulations do not yet fully reach the required resolution and physical realism to test this. At the present time, a promising idea to pursue seems to be an atomic cooling halo – i.e. a halo in which the Lyman-Werner ultraviolet flux from population III.1 stars in the surroundings exceeds a critical threshold, effectively destroying  $\text{H}_2$  – with  $\sim 10^8 M_{\odot}$  and a virial temperature greater than  $\sim 10^4 \text{ K}$ . The dynamical time for a first-galaxy halo to assemble is estimate at  $\sim 10^8 \text{ yr}$ , i.e. of the same order as the delay time between the first and second generation of star formation. The time it takes for the chemically enriched material ejected by pair-creation population III.1 supernova to disseminate and be re-incorporated in a new (population III.2) star or galaxy formation event may also reach up to  $10^8 \text{ yr}$ . Recent cosmological simulations indicate that – in marked contrast to the situation in dark matter mini-haloes – star formation in the first galaxies was strongly influenced by gravitationally driven supersonic turbulence generated during the virialization process. This, together with the chemical enrichment, may have lead to the formation of the first stellar clusters. It remains however an open question as to whether the first galaxies were the sites where globular clusters formed, which are the oldest star clusters known.

Discoveries of quasars at redshifts of  $z \sim 6 - 7$  that are powered by black holes with masses of  $\sim 10^9 M_{\odot}$  show that supermassive black holes (SMBH) already came into existence within 1 Gyr after the Big Bang. The current paradigm implies that these objects have grown, starting from a much smaller seed, via matter accretion and, to a lesser extent by merging with other compact objects (see e.g. Natarajan 2011). To allow growth to the observed masses, the initial seeds must have been  $\sim 400 M_{\odot}$  or more. Such objects may have been formed in supernova explosions in the high-mass tail of the first stars.

An alternative route toward the formation of SMBH is that of ‘direct collapse black hole (DCBH) formation’ by gas gravitational collapse – proceeding at rates of  $\gtrsim 0.1 - 1 M_{\odot} \text{ yr}^{-1}$ , i.e.  $\sim 100$  times larger than for standard metal-free star formation – in dark matter haloes with virial temperatures  $T \gtrsim 10^4 \text{ K}$  (e.g. Dijkstra et al. 2014). Given the strong temperature sensitivity of the atomic cooling process the gas collapses almost isothermally, thermostating at  $T \sim 8000 \text{ K}$ . Under these conditions, it seems that gas fragmentation into sub-clumps is almost completely inhibited and collapse proceeds to very high densities unimpeded. To conclude, DCBH formation requires the following conditions on the host haloes: (1) virial temperature  $> 10^4 \text{ K}$  to ensure high accretion rates allowed by atomic cooling; (2) gas metallicities  $Z < Z_{\text{crit}}$  to prevent fragmentation into clumps induced by heavy elements and dust cooling; (3) exposure to a Lyman-Werner ultraviolet radiation field from a population of first stars that is strong enough to strongly depress  $\text{H}_2$  abundances.

We conclude that though much is still unclear as to the mechanism of galaxy formation, the role of the first stars in the shaping of these systems seems pivotal.

## Chapter 8

# Viscous Accretion Disks

The following four chapters focus on the physics and chemistry of protoplanetary disks. As outlined in the introduction, the original idea of a protosolar nebula — the so-called 'Urnebel' — dates back to Kant & Laplace in the 18th century. The confirmation that such disks indeed exist and that they are a common by-product of star formation dates back to the late 80's and early 90's, when the first mm images (Beckwith et al. 1986, Sargent & Beckwith 1987, Rodriguez et al. 1992) and spectroscopy (e.g. Koerner et al. 1993) of such disks were taken (Fig. 10.1). The mm dust emission revealed the extended structures around young stars and the double peaked line profiles were proof of the regular Keplerian rotation pattern. The launch of the Hubble Space Telescope in 1990 opened a new window for research on protoplanetary disks. The high spatial resolution from space enabled the detailed imaging of the Orion nebula, a star forming region at a distance of 450 pc (ODell et al. 1992). These images were taken with the Wide Field Camera in several optical narrow band filters such as  $H\alpha$ , [O III], [O I] and [S II]. This data shows not only that such protoplanetary disks are ubiquitous around newly formed stars (50% of stars show disk detection), but reveals also the impact of disk irradiation and erosion by nearby hot O and B stars.

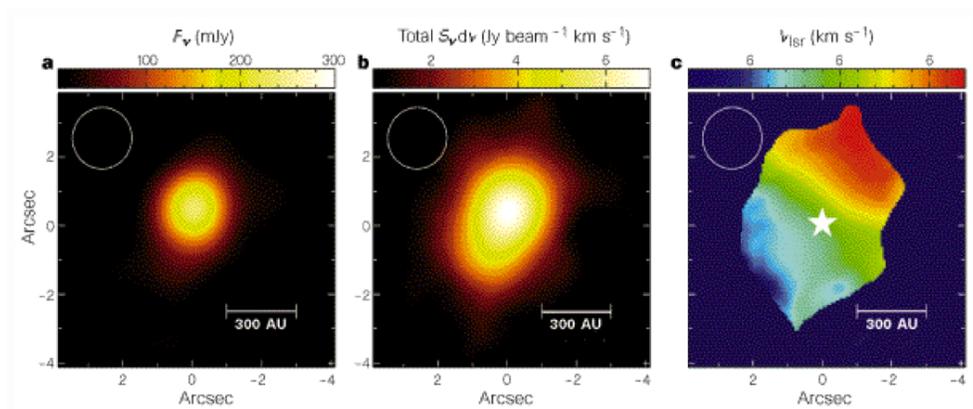


Figure 8.1: In each panel, the full-width at half-maximum of the synthesized beam is shown at top left, and offsets in arcseconds from the stellar position are indicated along the vertical and horizontal axes. (a) Thermal continuum from dust grains, at wavelength  $\lambda = 1.3$  mm. The flux density ( $F$ ) scale is shown at the top of the panel. (b) CO(2-1) emission, integrated across the full velocity range (+2.4 to +7.6  $\text{km s}^{-1}$ ) in which emission is measured above the  $3\sigma$  level. The scale at the top of the panel provides the integrated intensities ( $Sdv$ ). (c) Intensity-weighted mean gas velocities at each spatial point of the structure shown in b. The stellar (systemic) velocity,  $V_{\text{lsr}} \approx 5.1 \text{ km s}^{-1}$ , is represented by green. Blue-shifted (approaching) and red-shifted (receding) velocity components are shown as blue and red, respectively. The star symbol indicates the position of MWC480 (figure and caption from Mannings et al. 1997).

In Sect. 6.5, we described already briefly the formation of an accretion disk that accompanies the protostellar collapse. This chapter will focus on the formation of the disk, the radial and vertical structure of the disk, the transport of angular momentum — a prerequisite for accretion of material onto the star — and jets & outflows. Many of the basics of accretion disks are covered in the review article by Pringle (1981). An extensive description is also given in chapter 3 of Armitage (2010).

## 8.1 Disk formation

Sect. 5.2.4 describes how a flat disk forms during the collapse of a rotating cloud. Since the simple collapse problem outlined there is spherically symmetric and the rotation itself axisymmetric, the trajectories of infalling material from the top are mirrored at the equatorial plane. The infalling fluid elements, which follow parabolic trajectories, thus collide in the equatorial plane with matter coming from the other side of the cloud. An accretion shock forms where the material collides and this accretion shock extends over the entire centrifugal radius  $r_c$ . If the excess heat energy from the shock can be efficiently radiated away, a thin disk structure forms in the equatorial plane (Fig. 8.2).

The disk is supported against gravity by its rotation, but also to a lesser extent by the gas pressure gradient. To first order we can thus approximate the orbital velocity of the gas by the Keplerian orbital speed

$$v_\Phi = \Omega_K r = \sqrt{\frac{GM_*}{r}} . \quad (8.1)$$

We call  $\Omega_K$  the Kepler frequency

$$\Omega_K = \sqrt{\frac{GM_*}{r^3}} . \quad (8.2)$$

In the shock and subsequent cooling process, the vertical component of the infalling velocity is dissipated, so that the material within the disk is left with the parallel component of its infalling velocity  $v_r$  (see Fig. 8.2). However, this remaining radial velocity is not necessarily equal to the Keplerian velocity at the point where the material falls onto the disk. This causes some additional mixing and angular momentum transport within the disk upon until its angular momentum and mass is redistributed in such a way that we end up with a disk in nearly Keplerian rotation. The point at which the infalling material hits the disk depends entirely on its initial angular momentum, so its position within the cloud.

## 8.2 Disk structure

For deriving the general equations of disk structure, we follow the motion of an annulus of the disk and use the vertically integrated form in all equations, i.e. the surface density  $\Sigma$  instead of the physical density  $\rho$ . Matter can flow into the annulus and out of it with a velocity  $v_r$ .

We consider in the following an axisymmetric disk and assume that the surface density can be written in the general form of  $\Sigma(r, t)$ , which is an integral of the mass density  $\rho(r, z)$  over the disk height  $z$

$$\Sigma(r, t) = \int_{-\infty}^{\infty} \rho(r, z) dz . \quad (8.3)$$

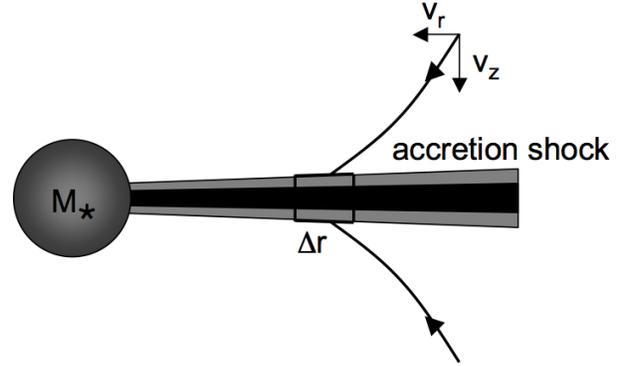


Figure 8.2: Schematic view of disk formation during the collapse of a rotating spherical cloud.

Then the vertically integrated continuity equation reads as

$$r \frac{\partial \Sigma}{\partial t} + \frac{\partial}{\partial r} (r \Sigma v_r) = 0 \quad (8.4)$$

and the vertically integrated conservation of orbital angular momentum becomes

$$r \frac{\partial}{\partial t} (r^2 \Omega \Sigma) + \frac{\partial}{\partial r} (r^2 \Omega \cdot r \Sigma v_r) = \frac{1}{2\pi} \frac{\partial G}{\partial r} . \quad (8.5)$$

The two velocity components are  $v_r$  and  $v_\phi = r\Omega$ . The angular momentum of a thin annulus of the disk is  $2\pi r \Delta r \Sigma r^2 \Omega$ , where  $\Delta r$  is the width of the annulus. The term on the right hand side of Equation 8.5 arises from the viscous torques in the disk.

We briefly revisit here the concept of a torque. A force that acts on the mass center of a body will cause a straight motion. If the same force acts on a point off the mass center, the body will start to rotate as well. The torque is then the product of the force  $F$  and the lever, i.e. the distance from the mass center  $r$

$$G = rF \sin \theta \quad (8.6)$$

where  $\theta$  is the angle between the force vector and the lever vector. The accretion disk is not rotating as a solid body and so we can apply the concept of torque on each annulus in the disk. The neighboring annulus will exert a force on it proportional to the difference in orbital velocities, the orbital velocity gradient  $d\Omega/dr$ . The force is called the shear force or viscous force.

The rate of shearing due to the orbital velocity gradient can be written as

$$A = r \frac{d\Omega}{dr} , \quad (8.7)$$

which has the units of  $s^{-1}$ . Without describing yet the details of the viscosity or angular momentum transport, we note that

$$G = 2\pi r \cdot \nu \Sigma r \frac{d\Omega}{dr} \cdot r = 2\pi r^3 \nu \Sigma \frac{d\Omega}{dr} . \quad (8.8)$$

Here,  $\nu$  is the kinematic viscosity and  $G$  is the torque exerted on an annulus in the disk. It is the product of the circumference, the viscous force per unit length ( $\nu \Sigma A = \nu \Sigma r d\Omega/dr$ ) and the lever  $r$ .

Using the expression for the torque  $G$ , we can rewrite Eq.(8.5) as

$$\frac{\partial}{\partial t} (r^2 \Omega \Sigma) + \frac{1}{r} \frac{\partial}{\partial r} (r^2 \Omega \cdot r \Sigma v_r) = \frac{1}{r} \frac{\partial}{\partial r} \left( \nu \Sigma r^3 \frac{d\Omega}{dr} \right) . \quad (8.9)$$

We can now use the continuity equation (Eq.8.4) to eliminate the time dependence and simplify this to

$$\Sigma v_r \frac{\partial (r^2 \Omega)}{\partial r} = \frac{1}{r} \frac{\partial}{\partial r} \left( \nu \Sigma r^3 \frac{d\Omega}{dr} \right) . \quad (8.10)$$

The derivatives of  $\Omega$  can be worked out, if we assume that the orbital frequency equals the Kepler frequency  $\Omega_K$ . Then, this equation yields an expression for the radial velocity in the disk

$$v_r = -\frac{3}{\Sigma \sqrt{r}} \frac{\partial}{\partial r} (\nu \Sigma \sqrt{r}) . \quad (8.11)$$

Alternatively, we can insert Eq.(8.10) into Eq.(8.4) to eliminate  $v_r$  and obtain

$$\frac{\partial \Sigma}{\partial t} = \frac{1}{r} \frac{\partial}{\partial r} \left[ \left( \frac{d}{dr} (r^2 \Omega) \right)^{-1} \frac{\partial}{\partial r} \left( \nu \Sigma r^3 \left( -\frac{d\Omega}{dr} \right) \right) \right] . \quad (8.12)$$

If we assume again that  $\Omega = \Omega_K$ , we can simplify this to

$$\frac{\partial \Sigma}{\partial t} = \frac{3}{r} \frac{\partial}{\partial r} \left[ \sqrt{r} \frac{\partial}{\partial r} (\nu \Sigma \sqrt{r}) \right] . \quad (8.13)$$

### 8.2.1 Steady state disk structure

In steady state, the disk structure follows from radial and angular momentum conservation and the assumption that the vertical component of gravity from the star is balanced by the vertical gas pressure gradient (vertical hydrostatic equilibrium). We study for this a disk annulus at a distance  $r$  from the star. Material can flow into and out of the annulus with the velocity  $v_r$  in radial direction. The radial momentum conservation equation for a steady-state flow is

$$v_r \frac{\partial v_r}{\partial r} - \frac{v_\phi^2}{r} + \frac{1}{\rho_{\text{gas}}} \frac{\partial P}{\partial r} + \frac{G M_*}{r^2} = 0, \quad (8.14)$$

where  $v_r$  is the radial velocity of the gas,  $v_\phi$  is the circular velocity, and  $M_*$  is the mass of the central star. The gas sound speed is defined as  $c_s^2 = \partial P / \partial \rho_{\text{gas}}$ , where  $P$  is the gas pressure and  $\rho_{\text{gas}}$  is the gas density. For an ideal gas,  $c_s = \sqrt{k T_g / (\mu m_p)}$ . Here,  $k$  is the Boltzmann constant,  $T_g$  is the gas temperature,  $\mu$  is the mean molecular weight of the gas, and  $m_p$  is the proton mass. The four terms in the above equation arise from radial mass flow, centrifugal force, gas pressure, and gravity. Since pressure typically decreases with increasing radius, the third term is nearly always negative; effectively, gas pressure resists the gravitational force, resulting in gas rotating at sub-Keplerian orbital velocities.

Following (Eq. 8.5) and using the expression for the torque  $G$ , the angular momentum conservation equation becomes

$$\frac{\partial}{\partial r} (r v_r \Sigma r^2 \Omega) = \frac{\partial}{\partial r} \left( \nu \Sigma r^3 \frac{d\Omega}{dr} \right), \quad (8.15)$$

where the left-hand side is the radial change in angular momentum and the right-hand side arises from the viscous torques. Integrating this equation yields the following expression

$$\nu \Sigma \frac{d\Omega}{dr} = \Sigma v_r \Omega + \frac{C}{2\pi r^3}. \quad (8.16)$$

Evaluating this equation at the point where the shear,  $r d\Omega/dr$ , vanishes, yields an expression for the integration constant

$$C = -2\pi \Sigma v_r \Omega r^3 = -(2\pi r \Sigma v_r) r^2 \Omega = \dot{M} r^2 \Omega. \quad (8.17)$$

Here,  $\dot{M} = 2\pi r \Sigma (-v_r)$ , because in an accretion flow, the radial velocity is inward and thus negative. For simplicity, we assume that the disk extends down to the star. Then, the shear vanishes close to the star for

$$C = \dot{M} \sqrt{G M_* R_*} \quad (8.18)$$

The integration constant can thus be seen as the influx of angular momentum through the disk. With this and  $\Omega = \Omega_K = \sqrt{G M_* / r^3}$ , the disk surface density becomes

$$\Sigma = \frac{\dot{M}}{3\pi\nu} \left( 1 - \sqrt{\frac{R_*}{r}} \right), \quad (8.19)$$

for radii much larger than the stellar radius.

### 8.2.2 Vertical disk structure

The vertical ( $z$ -direction) disk structure is found by solving the equation of hydrostatic equilibrium,

$$\frac{1}{\rho_{\text{gas}}} \frac{\partial P}{\partial z} = \frac{\partial}{\partial z} \left( \frac{G M_*}{\sqrt{r^2 + z^2}} \right). \quad (8.20)$$

If the disk is vertically thin ( $z \ll r$ ) and the gas temperature does not depend on  $z$  (i.e. a vertically isothermal disk), Equation 8.20 can be integrated to give the vertical density structure,

$$\rho_{\text{gas}}(r, z) = \rho_c(r) e^{-z^2 / (2 H_{\text{gas}}^2)}, \quad (8.21)$$

where  $\rho_c(r)$  is the density at the disk midplane. The gas scale height is

$$H_{\text{gas}} = \sqrt{kT_c r^3 / (\mu m_p G M_*)}, \quad (8.22)$$

where  $T_c$  is the midplane gas temperature. This equation shows that the gas scale height is the ratio of the gas sound speed to the angular velocity ( $H_{\text{gas}} = c_s/\Omega$ ).

### 8.2.3 Radial disk structure

Having found an expression for the scale height, we return to the radial disk structure and write Equation 8.19 using  $r \gg R_*$  as

$$\Sigma = \frac{\mu m_p \sqrt{G M_*}}{3 \pi k} \frac{\dot{M}}{\alpha T_c r^{3/2}}. \quad (8.23)$$

Here we used the relation  $\nu = \alpha c_s h$  which will be introduced later in Sect. 8.4.2 after a detailed discussion of the viscosity in disks. For a disk with a simple power-law midplane temperature profile,  $T_c \propto r^{-q}$ , the surface density is proportional to  $r^{q-3/2}$ , thus a simple power law of radius. Further, the midplane density may be written as

$$\rho_c \simeq \Sigma / H_{\text{gas}} = \rho_{\text{in}} (r/r_{\text{in}})^\epsilon, \quad (8.24)$$

where  $\rho_{\text{in}}$  is the density at the inner disk radius and the exponent  $\epsilon$  equals  $\frac{3}{2}q - 3$ . Assuming a simple power law for the temperature profile is a first guess, but is there actually a way to determine the temperature profile within an accretion disk?

### 8.2.4 Temperature profile of an accretion disk

In an accretion disk, the energy generation is predominantly through viscous torques,  $G$ . The net torque on a disk annulus is equal to the difference between the torque on the outer and inner surface

$$G(r + \Delta r) - G(r) = \frac{\partial G}{\partial r} \Delta r. \quad (8.25)$$

The energy generated by this torque is then

$$\Omega \frac{\partial G}{\partial r} \Delta r = \left[ \frac{\partial}{\partial r} (G\Omega) - G \frac{d\Omega}{dr} \right] \Delta r \quad (8.26)$$

The first term is simply  $G\Omega|_{r_{\text{out}}} - G\Omega|_{r_{\text{in}}}$  and thus given by the disk boundary conditions. The second term describes the local energy dissipation and thus the heat generation in the disk. We assume that this energy is radiated away and thus the rate of dissipation per disk surface area ( $2\pi r \Delta r$ ; additional factor 2 due to the two sides of the disk) is

$$\begin{aligned} \dot{E} &= -\frac{G}{4\pi r} \frac{d\Omega}{dr} \\ &= -\frac{1}{2} \nu \Sigma \left( r \frac{d\Omega}{dr} \right)^2 \\ &= \frac{9}{8} \nu \Sigma \Omega^2 \end{aligned} \quad (8.27)$$

For this, we assumed that the disk has a Keplerian rotation profile  $\Omega = \sqrt{GM/r^3}$ . Using the relation found between the mass accretion rate and the viscosity,  $\dot{M} = 3\pi\nu\Sigma$  for  $r \gg R_*$ , we can eliminate the viscosity from this equation

$$\dot{E} = \frac{3}{8\pi} \dot{M} \Omega^2. \quad (8.28)$$

If the disk emits as a black body, we can write  $\dot{E}$  as  $\sigma T_{\text{disk}}^4$  with  $\sigma$  being the Stefan-Boltzmann constant. This yields a temperature profile of

$$T_{\text{disk}} = \left( \frac{3}{8\pi\sigma} \dot{M}\Omega^2 \right)^{1/4}, \quad (8.29)$$

thus a  $r^{-3/4}$  profile. We see that the disk temperature does not depend on the viscosity. The implicit assumption here is that the viscosity of the disk corresponds to its accretion rate, an observable quantity. Hence we do not need to know the details of the angular momentum transport to describe the disk structure. Using a typical observed accretion rate for young T Tauri stars ( $M_* = 1 M_\odot$ ) of  $10^{-7} M_\odot/\text{yr}$ , we obtain a disk temperature of 150 K at a distance of 1 AU.

### 8.3 Disk evolution

Using Equation 8.4, we have eliminated  $v_r$  in Equation 8.5 and obtained from the latter an equation that describes the temporal evolution of the surface density of the disk

$$\frac{\partial \Sigma}{\partial t} = \frac{3}{r} \frac{\partial}{\partial r} \left[ \sqrt{r} \frac{\partial}{\partial r} (\nu \Sigma \sqrt{r}) \right]. \quad (8.30)$$

We can see that this equation has the form of a diffusion equation

$$\frac{\partial f}{\partial t} = D \frac{\partial^2 f}{\partial x^2}, \quad (8.31)$$

with a diffusion coefficient  $D$ . Fig. 8.3 illustrates the viscous evolution of the surface density of a ring of matter in the disk. Initially at  $\tau = 0$ , all matter resides in a narrow ring of mass  $m$  at radius  $r_0$  (radial distance from the star)

$$\Sigma(r, t = 0) = \frac{m}{2\pi r_0} \delta(r - r_0), \quad (8.32)$$

with the Dirac delta function  $\delta(r - r_0)$ . With time, the mass initially contained in a narrow ring, spreads out to both sides of the initial ring. Most of the mass moves inward, while some smaller amount of mass moves out to larger radii to take up the angular momentum.

If we parametrize the viscosity as a power law of radius

$$\nu \propto r^\gamma, \quad (8.33)$$

we can obtain the self-similar solution of a spreading disk also derived by Lynden-Bell & Pringle (1974). We assume that the initial surface density profile is that of a steady-state disk with an exponential cutoff at  $r = r_1$  (the scaling radius)

$$\Sigma(\tilde{r}, t = 0) = \frac{C}{3\pi\nu(r_1)\tilde{r}^\gamma} \exp(-\tilde{r}^{2-\gamma}), \quad (8.34)$$

where  $\tilde{r} \equiv r/r_1$ . Without the details of derivation, the solution has then the form

$$\Sigma(\tilde{r}, t) = \frac{C}{3\pi\nu(r_1)\tilde{r}^\gamma} \theta^{-(5/2-\gamma)/(2-\gamma)} \exp\left(-\frac{\tilde{r}^{2-\gamma}}{\theta}\right). \quad (8.35)$$

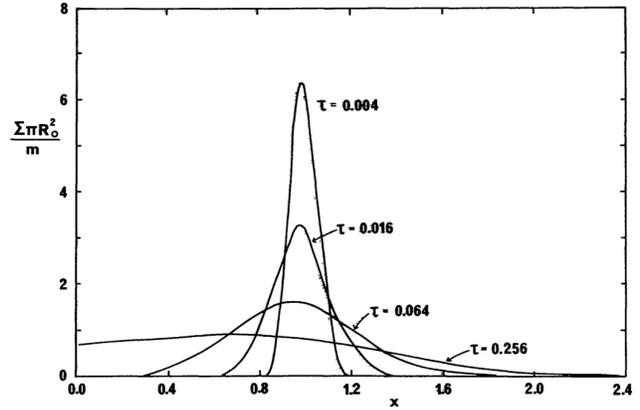


Figure 8.3: The viscous evolution of a ring of matter of mass  $m$ . The surface density  $\Sigma$  is shown as a function of the dimensionless radius  $x = r/r_0$ , where  $r_0$  is the initial radius of the ring, and of dimensionless time  $\tau = 12\nu t/r_0^2$ , where  $\nu$  is the viscosity (caption and figure from Pringle 1981).

Here, the scaled time variable  $\theta$  is related to  $t_s$ , the viscous scaling time, through

$$\theta = \frac{t}{t_s} + 1 \quad (8.36)$$

$$t_s = \frac{1}{3(2-\gamma)^2} \frac{r_1^2}{\nu(r_1)}.$$

The solution is plotted in Fig. 8.4 for three subsequent timesteps  $\theta = 2, 4$ , and 8. As time evolves, the mass of the disk decreases. A large fraction of the mass is accreted onto the star, while a small fraction moves outwards (viscous spreading of the disk) taking along all the angular momentum.

A recent simulation of the disk formation and viscous spreading phase has been done by Hueso & Guillot (2005). Fig. 8.5 shows the viscous evolution of a star + accretion disk system with realistic input parameters. The figure shows thick grey lines superimposed on each plotted quantity. Each one shows the range of time for which that quantity agrees with the available observations. This example shows how an 800 AU disk can be formed by viscous diffusion of an initially much smaller disk, with a centrifugal radius  $r_c = 11$  AU. At the end of the simulation, the star has almost acquired its final mass (here  $0.5 M_\odot$  at 10 Myr). The disk mass grows until a few times  $10^5$  yr. After that the initial cloud has largely dispersed and now the disk material slowly accretes onto the star.

## 8.4 Angular momentum transport

The specific angular momentum, which is the angular momentum per unit mass, stored in a  $1 M_\odot$  disk with a size of 10 AU is  $3 \times 10^{53}$  cm<sup>2</sup>/s. On the other hand, a  $1 M_\odot$  star rotating at break-up velocity has a specific angular momentum of  $6 \times 10^{51}$  cm<sup>2</sup>/s. Hence, the original angular momentum of the disk is 50 times higher than the maximum allowed for a star. Since angular momentum is strictly conserved, there needs to be a process that actually transports angular momentum away to prevent it from accumulating on the star. The main possibilities are a torque from the external medium (e.g. magnetic fields), viscosity inside the disk transporting angular momentum to the outer disk, disk winds taking angular momentum away. In the following, we discuss the basic principle of angular momentum transport within the disk.

If we assume for a moment that the disk rotates with Keplerian speed and consider only two particles

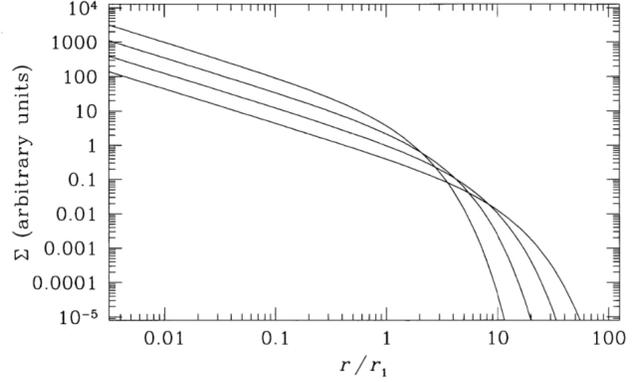


Figure 8.4: The self-similar solution to the disk evolution equation is plotted for a viscosity  $\nu \propto r$ . The initial surface density tracks the profile for a steady-state disk ( $\Sigma \propto r^{-1}$ , Eq. 8.23 for  $T_c \propto r^{-1/2}$ ) at small radii, before cutting off exponentially beyond  $r = r_1$ . The curves show the surface density at the initial value of the scaled time variable  $\theta = 1$  and at subsequent times  $\theta = 2, \theta = 4$ , and  $\theta = 8$  (caption and figure from Armitage 2010).

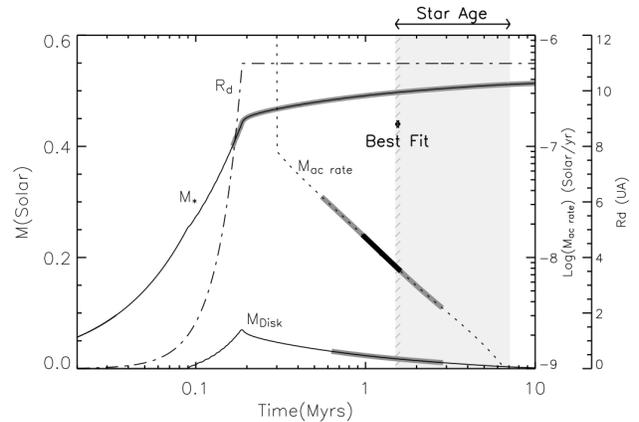


Figure 8.5: Evolution of star mass  $M_*$  and disk mass  $M_{\text{disk}}$  as a function of time with masses in solar units (corresponding axis to the left) for a viscosity  $\alpha = 0.01$ . The accretion rate onto the central star is shown as a dotted line (corresponding axis: right to the right). The evolution of the centrifugal radius  $r_c$  (labelled for some reason  $R_d$  in the plot) is shown as a dash-dotted line (corresponding axis: far-right and units are AU not UA). Gray curves and the hashed region indicate time sequences when selected observational constraints are varied (from Hueso & Guillot 2005).

with masses  $m_1$  and  $m_2$ , we can write their energy and angular momentum as

$$E = -\frac{GM_*}{2} \left( \frac{m_1}{r_1} + \frac{m_2}{r_2} \right) \quad (8.37)$$

$$\begin{aligned} J &= m_1 v_{\Phi,1} r_1 + m_2 v_{\Phi,2} r_2 \\ &= \sqrt{GM_*} (m_1 \sqrt{r_1} + m_2 \sqrt{r_2}) \quad , \end{aligned} \quad (8.38)$$

where  $r_1$  and  $r_2$  are the corresponding distances of the two masses from the central star with mass  $M_*$ . From the conservation of angular momentum of the total system, it follows that a small change in orbit for one of the masses requires a corresponding change for the other one

$$\begin{aligned} \frac{\partial J}{\partial r} &= -m_1 \frac{1}{2\sqrt{r_1}} \\ \Delta J_1 &= -m_1 \frac{\Delta r_1}{2\sqrt{r_1}} \\ m_1 \frac{\Delta r_1}{\sqrt{r_1}} &= -m_2 \frac{\Delta r_2}{\sqrt{r_2}} \end{aligned} \quad (8.39)$$

The same of course holds true if we consider two neighboring annuli in the disk.

This process requires a way of connecting the two particles (or rings of material) with each other. So, what remains to be identified is the source of this coupling. We can consider for example the friction between the two annuli to be caused by their difference in rotation speed (shear motion). However, since we deal here with a gas, diffusion plays a role in transporting gas in both directions, inward and outward. We can hence consider small turbulent random motions as a cause of radial mixing of material and thus as a cause of coupling the various annuli with the disk.

### 8.4.1 Turbulent viscosity

We can derive a simple estimate of the molecular viscosity  $\nu_m$  through

$$\nu_m \approx \lambda c_s \quad , \quad (8.40)$$

where  $\lambda$  is the mean free path of the molecules given by  $\frac{1}{(n\sigma_m)}$ , the product of gas particle density  $n$  and collisional cross section  $\sigma_m$  between the molecules. If we approximate the latter by the typical physical size of molecules ( $2 \times 10^{-15}$  cm<sup>2</sup>), we obtain for a typical distance of 10 AU in the disk a molecular viscosity of

$$\nu_m = \frac{1}{n\sigma_m} c_s = \frac{5 \times 10^4}{10^{12} \times 2 \times 10^{-15}} \text{ cm}^2/\text{s} = 2.5 \times 10^7 \text{ cm}^2/\text{s} \quad . \quad (8.41)$$

If we estimate the corresponding viscous timescale at  $r = 10$  AU

$$t_\nu = \frac{r^2}{\nu_m} \approx 3 \times 10^{13} \text{ yr} \quad (8.42)$$

we immediately see, that we can rule out molecular viscosity as the main source of turbulent motions in these accretion disks. The typical disk evolutionary timescale is of the order of a few million years, so more than ten million times shorter than this molecular viscosity timescale.

The Reynolds number is used in gas dynamics to describe the ratio of inertial forces (resistance to change or motion) to viscous forces (glue) and thus also to define whether a fluid is laminar or turbulent. The Reynolds number  $\text{Re}$  of a typical accretion disk can be estimated as

$$\text{Re} = \frac{VL}{\nu_m} \quad , \quad (8.43)$$

where  $V$  and  $L$  are characteristic velocity and length scales in the disk, i.e.  $U = c_s = 0.5$  km/s at 10 AU and  $L = h = 0.05r = 0.5$  AU, where  $h$  is the scale height of the gas in the disk (we get back to the vertical disk structure later). Filling in these numbers, we obtain a Reynolds number of  $Re \approx 10^{10}$ . Hence, in the presence of some instability, the disk is highly turbulent. The Reynolds number also indicates the ratio between the largest and smallest scales of fluid motion (sometimes referred to as eddies). The largest scale fluid motion is set by the geometry of the disk (e.g. its scale height), while the smallest scales are here a factor  $\sim 10^{10}$  smaller.

### 8.4.2 Shakura-Sunyaev viscosity

If molecular viscosity is not large enough to provide a source of turbulence that drives the angular momentum transport in disks, what else can it be? We need to identify possible instabilities that can cause a turbulence larger than that caused by the molecular viscosity. Before we go into that discussion, we present here a very successful idea of parametrizing the viscosity without identifying its source. The idea goes back to Shakura & Sunyaev (1973) who first proposed the so momentum transport.

Since the largest turbulent scales will be set by the geometry of the flow, we can use the vertical scale height of the gas as a representative scale. Along the same dimensional analysis scheme, we can use the sound speed  $c_s$  as the characteristic velocity of the turbulent motions. Hence, we can write the viscosity  $\nu$  as

$$\nu = \alpha c_s h . \quad (8.44)$$

We can now express the viscosity in terms of the disk parameters and hence estimate the efficiency of angular momentum transport and thus mass accretion onto the central star. It is also very obvious to estimate how large  $\alpha$  should be to reproduce the observed timescale of disk evolution. The viscous timescale can be expressed as

$$\tau = \frac{r^2}{\nu} = \left(\frac{h}{r}\right)^{-2} \frac{1}{\alpha \Omega} . \quad (8.45)$$

Here, we can fill in 1 Myr as the typical evolutionary timescale at 50 AU. In addition, we assume that the disks are indeed very thin and  $h/r \sim 0.05$ . This yields an  $\alpha$  of 0.02.

### 8.4.3 MRI

In the presence of magnetic fields, the field lines act like springs connecting different annuli within the disk. Fig. 8.6 illustrates the basic principle. If a weak pull exists between two masses in adjacent annuli of the disk,  $m_i$  and  $m_o$ , the inner mass element will lose angular momentum and hence drift even further inward, while the outer mass element gains angular momentum and drifts further outward. The new orbital velocity of the inner mass element is even higher than before (presuming that  $\Omega$  decreases with  $r$ ), while the new orbital velocity of the outer element is even smaller. This effect thus enhances the velocity difference between the two masses in adjacent annuli, giving rise to an instability. The instability is generally referred to as magneto-rotational instability and has been described first by Balbus & Hawley (hence we often refer to it

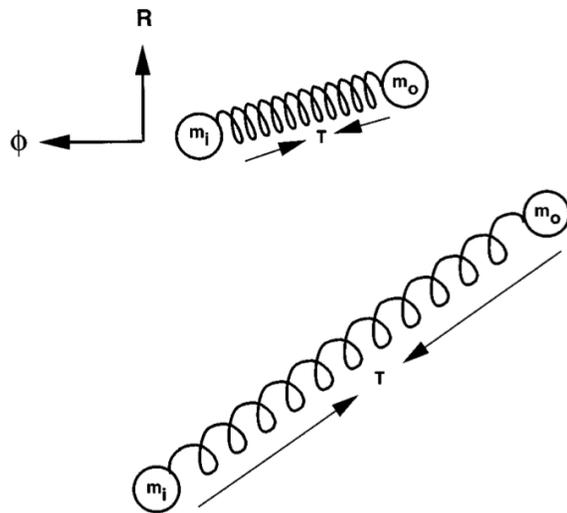


Figure 8.6: Two masses in orbit connected by a weak spring. The spring exerts tension force  $T$  resulting in a transfer of angular momentum from the inner mass  $m_i$  to the outer mass  $m_o$ . If the spring is weak, the transfer results in an instability as  $m_i$  loses angular momentum, drops through more rapidly rotating inner orbits, and moves further ahead. The outer mass  $m_o$  gains angular momentum, moves through slower outer orbits, and drops further behind. The spring tension increases and the process runs away (figure and caption from Balbus & Hawley 1998).

as Balbus-Hawley instability). A prerequisite for a magnetized disk to be linearly unstable is that the orbital velocity decreases with radius

$$\frac{d}{dr}(\Omega) < 0, \quad (8.46)$$

a condition clearly satisfied in Keplerian disks. In addition, the disk needs to be ionized to a certain degree since neutral gas does not couple efficiently to the magnetic field lines. The critical ionization degree is  $n_e/n_{\text{tot}} \sim 10^{-12}$  to sustain turbulence generation by the magneto-rotational instability (Sano & Stone 2002). We get back to the last point in the section on dead zones.

It is very difficult to assess the amount of viscosity generated by MRI. The only way of measuring  $\alpha$  would be through numerical magneto-hydrodynamical simulations. Disentangling physical and numerical effects becomes then a challenge. However, from simulations it seems that  $\alpha$  could be consistent with a value of  $\sim 10^{-2}$  as inferred from observations.

#### 8.4.4 Gravitational instabilities

For massive disks, self-gravity can become important and introduce another sort of instability. In general, self-gravity will tend to locally clump material together. This process will be counteracted by shear motions and also local gas pressure. It can be shown that gravity wins this competition if the "clumping" timescale is short compared to the time scales on which sound waves cross a clump or shear can destroy it. The three timescales can be written for a clump of size  $\Delta r$  and mass  $m \sim \pi(\Delta r)^2\Sigma$  as

$$t_{\text{ff}} \sim \sqrt{\frac{(\Delta r)^3}{Gm}} \sim \sqrt{\frac{\Delta r}{\pi G\Sigma}} \quad (8.47)$$

$$t_p = \frac{\Delta r}{c_s} \quad (8.48)$$

$$t_{\text{shear}} = \frac{1}{r} \left( \frac{d\Omega}{dr} \right)^{-1} \sim \Omega^{-1} \quad (8.49)$$

Comparing these timescales, the "Toomre Q" parameter (Toomre 1964) indicating whether the disk is stable can then be written as

$$Q = \frac{c_s\Omega}{\pi G\Sigma}. \quad (8.50)$$

The disk becomes unstable against self-gravity at  $Q < 1$ .

#### 8.4.5 Dead zones

Dead zones arise in disks as soon as the ionization degree drops below the critical one  $n_e/n_{\text{tot}} \sim 10^{-12}$ . Sources of ionization are collisional ionization, thermal ionization (only in the inner disk rim exposed to the stellar radiation), X-ray and cosmic ray ionization. The low level of ionization arising from cosmic rays or X-rays deep inside the disk is sufficient to sustain turbulence through MRI. Cosmic rays can penetrate column densities of 50 g/cm<sup>2</sup> and thus reach the midplane except for the inner few AU in a very massive disk with a steep surface density profile. X-rays penetrate column densities of  $\sim 100$  g/cm<sup>2</sup>. However, stellar X-rays (coming from the corona) will penetrate the disk radially and quickly reach these column densities. Only X-rays coming from high above the disk (e.g. from a stellar jet or from the accretion flow onto the star) can be expected to reach the midplane also at larger radii.

The existence of a dead zone has consequences for the accretion flow through the disk. Since viscosity drops within the dead zone, material streaming in from larger radii will accumulate in the dead zone. This has consequences for the density and pressure profile in the disk, causing e.g. a local inversion in the pressure gradient. Such a pressure gradient inversion can lead to the accumulation of small dust particles, increasing locally the dust:gas mass ratio in the disk. We get back to this point in the chapters on planetesimal formation.

## 8.5 FU Orionis outbursts

These objects have been previously discussed in Sect. 6 and we briefly summarize some basic information. FU Orionis stars are variable pre-main sequence stars that undergo optical outbursts of several magnitudes. The outbursts show fast rises on timescales of 1 yr and slow decays with timescales of the order of 50-100 yr (see Fig 6.11). These systems are thought to undergo episodic mass accretion from the inner disk with accretion rates being as high as  $10^{-4} M_{\odot}/\text{yr}$ . How can we explain these episodic high accretion rates within the just developed concept of accretion disks?

The basic idea is that the inner disk is extremely opaque and thus cannot efficiently radiate energy away that has been inserted into the gas through the viscous accretion process. The inner disk could then be hotter than  $10^4$  K. This also implies that the original assumption of a thin disk ( $z \ll r$ ) breaks down and the disk has a very large scale height  $h/r \sim 0.4$ . Under these circumstances, the inner 1 AU of the disk can have enough mass to fuel an FU Orionis outburst event. The event itself could be triggered by thermal instabilities that are due to the temperature dependence of the gas opacity. This can be illustrated using a curve that describes the loci of thermal equilibrium in the  $T$  versus  $\Sigma$  plane of the disk. This curve has the shape of an 'S' with the kink at the point where the opacity causes the thermal instability (Fig. 8.7). As more material piles up in the disk, the locus starts shifting to the right on the lower branch on the 'S curve'. The disk is stable, because a small temperature perturbation can be damped due to efficient cooling. As material piles up in the inner disk, the disk simply moves upward in that branch. As it reaches the inflection point A, a small positive temperature perturbation leads to the disk being in an unstable regime and it has to jump to the upper arm of the 'S curve' (position B). There, the surface density and hence mass accretion rate is much higher and eventually, the enhanced accretion onto the star will drain material away from the disk, thereby moving it down the upper arm of the 'S curve' to point C. At that stage, the only stable solution becomes again the low accretion solution on the lower arm of the 'S curve' (point D). The timescale of such thermal instabilities  $t_{th}$  is shorter than the viscous timescale  $t_{\nu}$

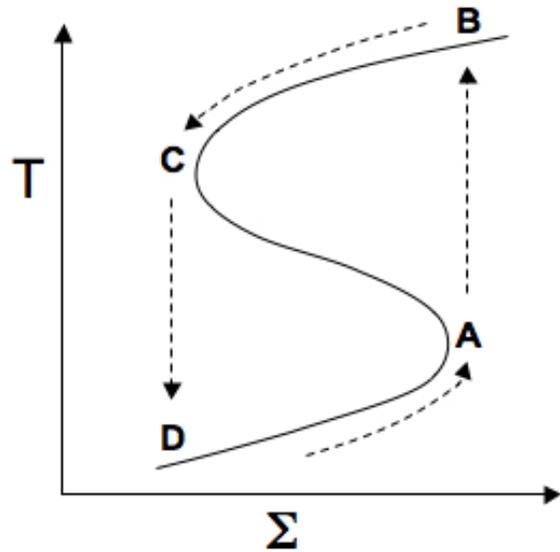


Figure 8.7: Schematic 'S curve' for the disk thermal instability (adapted from Hartmann 1998, Fig.7.11).

As more material piles up in the disk, the locus starts shifting to the right on the lower branch on the 'S curve'. The disk is stable, because a small temperature perturbation can be damped due to efficient cooling. As material piles up in the inner disk, the disk simply moves upward in that branch. As it reaches the inflection point A, a small positive temperature perturbation leads to the disk being in an unstable regime and it has to jump to the upper arm of the 'S curve' (position B). There, the surface density and hence mass accretion rate is much higher and eventually, the enhanced accretion onto the star will drain material away from the disk, thereby moving it down the upper arm of the 'S curve' to point C. At that stage, the only stable solution becomes again the low accretion solution on the lower arm of the 'S curve' (point D). The timescale of such thermal instabilities  $t_{th}$  is shorter than the viscous timescale  $t_{\nu}$

$$t_{th} \sim \left(\frac{h}{r}\right)^2 t_{\nu} , \quad (8.51)$$

and thus much closer to the typical 1 yr rise seen in the FU Orionis outburst events.

# Chapter 9

## Irradiated Disks

In the last chapter, we studied the dynamical evolution of protoplanetary disks, the transport of angular momentum and hydrostatic solutions to the vertically integrated hydrodynamical equations. That chapter described mainly the early stages of disk evolution when accretion rates are high and the disk still assembles a substantial fraction of its final mass.

In this chapter, we will study the somewhat later stages, when disks have acquired their final mass and evolve viscously by angular momentum transport. We focus especially on the vertical structure of the disks. The key physics setting the density distribution in vertical direction are the vertical component of stellar gravity — acting to compress the gas towards the midplane — and the temperature of the gas set by irradiation — providing the pressure that vertically expands the disk. In the absence of high accretion rates, which would provide an additional heating term in the midplane of the disk, the equilibrium between these two forces determines the vertical disk structure.

### 9.1 Basic principles

Most of the protoplanetary disks that we observe (class II and III sources) have ages  $> 1$  Myr and the accretion rates are generally below  $10^{-7} M_{\odot}/\text{yr}$ . Without specifying the mechanism behind turbulence, we assume that angular momentum transport is therefore slow and leads to a constant inward flow of matter. The outer disks generally contains most of the mass. Hence, we can assume that there exists a stationary state, where the material falling onto the star is continuously replenished from the large outer reservoir. In the following, we consider such stationary disks with typical sizes of a few 100 AU.

The disk temperature is set by an equilibrium between heating and cooling. The disk cools by thermal emission from the dust grains at infrared wavelengths. This radiation is what is observed as infrared dust continuum radiation from such disks. Line cooling is only a minor coolant, and only plays a role for  $T_{\text{gas}}$  when gas and dust are thermally decoupled. Dust grains can be heated in part by radiation from other grains

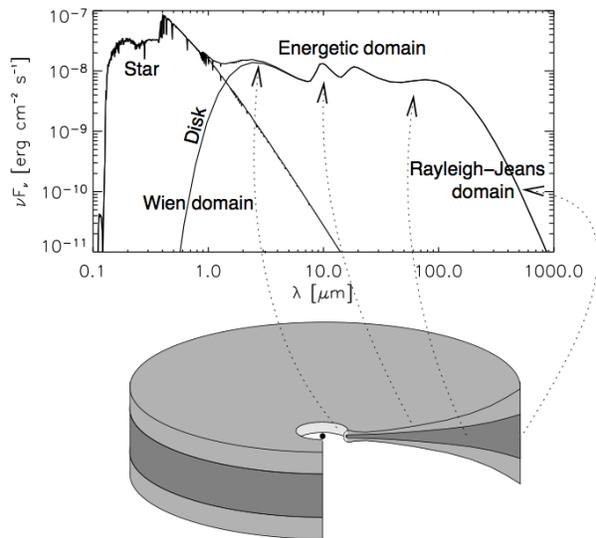


Figure 9.1: Build-up of the SED of a flaring protoplanetary disk and the origin of various components: the near-infrared bump comes from the inner rim, the infrared dust features from the warm surface layer, and the underlying continuum from the deeper (cooler) disk regions. Typically the near- and mid-infrared emission comes from small radii, while the far-infrared comes from the outer disk regions. The (sub-)millimeter emission mostly comes from the midplane of the outer disk. Scattering is not included here (Figure and caption from Dullemond et al. (2007)).

in the disk. The iterative absorption and re-emission of infrared radiation by dust grains in the disk causes the radiation to propagate through the disk in a diffusive way. Net energy input comes from absorption of direct stellar light in the disks surface layers, and from viscous dissipation of gravitational energy in the disk due to accretion. For most disks around CTTS's and Herbig Ae/Be stars the heating by stellar radiation is dominant over the viscous heating (except in the very inner regions). Only for strongly accreting disks does the latter dominate.

Fig. 9.2 illustrates the disk regions (radial distance from the star) where irradiation or accretion dominate the energy balance and hence the dust (gas) temperature. Low accretion rates ( $\dot{M} < 10^{-8} M_{\odot}/\text{yr}$ ) are irrelevant for the disk structure. As the accretion rate increases, first the inner disk and then also the outer disk becomes dominated by the accretion process. Viscous heating is then the dominant heat source for the gas (dust). Due to the generally higher disk temperatures that follow from the higher accretion rates, the dust sublimation radius moves slightly outwards.

## 9.2 Spectral energy distribution

Once the temperature structure is determined, the spectral energy distribution (SED) can be computed. The name stems from the fact that we plot  $\nu F_{\nu}$  instead of the flux  $F_{\nu}$ , and hence is the unit "energy emitted per surface area per second".

The observable thermal emission of a dusty disk model consists of three wavelength regions (see Fig. 9.1). The main portion of the energy is emitted in a wavelength range depending on the minimum and maximum temperature of the dust in the disk. We call this the "energetic domain" of the SED, which typically ranges from  $1.5 \mu\text{m}$  to about  $100 \mu\text{m}$ . At shorter wavelength the SED turns over into the "Wien domain". At longer wavelengths the SED turns over into the "Rayleigh-Jeans domain", a steep, nearly power law profile with a slope depending on grain properties and disk optical depth. Differences in disk geometry are mainly reflected in the energetic domain of the SED, while the submm and mm fluxes probe the disk mass.

The short wavelength regime, the UV and optical, is dominated by the stellar radiation and thus reflects the Planck function corresponding to the stellar blackbody temperature. In the very long wavelength regime, the Rayleigh-Jeans region, the slope of the SED is given by the tail of the Planck function

$$\nu F_{\nu} \sim \nu^3 \quad (9.1)$$

and the slope is independent of temperature. The middle part of the SED is a superposition of many different black bodies with a range of temperatures, the so-called multi-color region. We can try to connect the radial temperature gradient in the disk to the observed slope of the SED in this multi-color region. The SED can be expressed as

$$\nu F_{\nu} \sim S F_{\text{max}} \quad , \quad (9.2)$$

where  $S$  is the emitting surface and  $F_{\text{max}}$  the peak energy of the distribution. Suppose the temperature profile of the disk can be expressed as a power law

$$T(r) \sim r^{-q} \Rightarrow r \sim T^{-1/q} \quad , \quad (9.3)$$

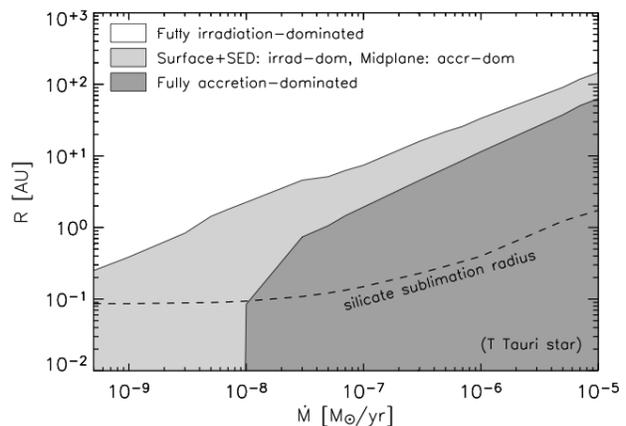


Figure 9.2: Characteristic radii of a disk around a  $0.9 M_{\odot}$  star with  $T_{*} = 4000 \text{ K}$  and  $R_{*} = 1.9 R_{\odot}$  at different accretion rates. The silicate sublimation radius is the dust inner rim. Figure based on models by DAlessio et al. (1998) and taken from the PPV review by Dullemond et al. (2007).

then we can express the emitting area as

$$S \sim r dr \sim r^2 \sim T^{-2/q} \sim \nu^{-2/q} . \quad (9.4)$$

The location of the peak emission is given by Wien's law and the peak energy of the Planck function is

$$F_{\max} \sim T^4 \sim \nu^4 . \quad (9.5)$$

Using these relations, we can derive an approximation for the slope of the multi-color region

$$\nu F_\nu \sim S F_{\max} \sim \nu^{-2/q} \nu^4 = \nu^{(4q-2)/q} . \quad (9.6)$$

If the disk has a finite optical depth, the slope of this multi-color region remains unchanged. The Rayleigh-Jeans slope however, changes depending on the slope of the dust opacity. Suppose that the opacity  $\kappa_\nu$  can be parametrized in the long wavelength regime as

$$\kappa_\nu \sim \nu^\beta . \quad (9.7)$$

Then the intensity modifies to

$$I_\nu(r) = (1 - e^{-\tau_\nu}) B_\nu(T) \approx \tau_\nu B_\nu \approx \kappa_\nu B_\nu . \quad (9.8)$$

Hence, the energy  $\nu F_\nu$  becomes

$$\nu F_\nu \approx \kappa_\nu \nu B_\nu \approx \nu^{3+\beta} . \quad (9.9)$$

We recall from the previous chapter the temperature profile of an accretion disk

$$T_{\text{disk}} = \left( \frac{3}{8\pi\sigma} \dot{M} \Omega^2 \right)^{1/4} \sim r^{-3/4} . \quad (9.10)$$

Inserting this into Eq.(9.6), we obtain  $\nu F_\nu \sim \nu^{4/3}$ . If we compare this with the observed SED of two young stars, HD104237 and AB Aurigae (Fig. 9.3), we see that the slope of this flat disk does not fit the observed SED. Also, the required mass accretion rates of several times  $10^7 M_\odot/\text{yr}$  disagree with estimates from the veiling detected in the UV/optical part of the spectrum.

### 9.3 Flat irradiated disks

It is quite challenging to solve the entire disk structure according to the above principles. Early disk models were therefore often based on strong simplifications. An example of such a model is a perfectly at disk being irradiated by the star due to the stars non-negligible size. The stellar radiation impinges onto the at disk under an irradiation angle  $\phi \approx 0.4R_*/r$  (with  $R_*$  the stellar radius). Neglecting viscous dissipation, the effective temperature of the disk is set by a balance between their radiated flux

$$\frac{1}{2} \phi \frac{L_*}{4\pi r^2} , \quad (9.11)$$

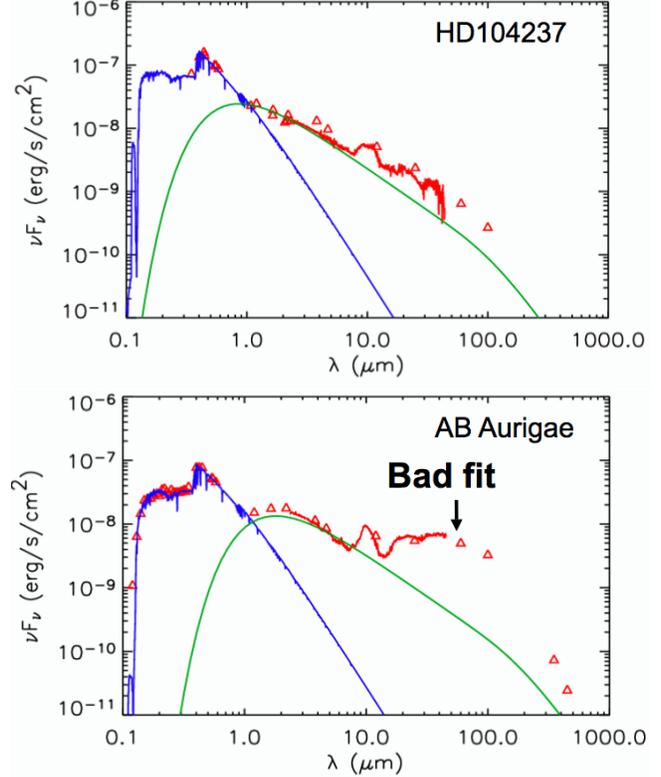


Figure 9.3: Observed (red) SED of two Herbig AeBe stars: HD104237 (top) and AB Aurigae (bottom). The blue line is a fit to the stellar spectrum and the green line is the best fit found from an accretion disk model (temperature profile from Eq.(9.10)) with an accretion rate of  $\dot{M} = 2 \times 10^{-7} M_\odot/\text{yr}$  and  $\dot{M} = 7 \times 10^{-7} M_\odot/\text{yr}$  for the two stars respectively.

with  $L_*$  being the stellar luminosity. We assume that the disk is vertically isothermal and re-emits as a blackbody

$$I_\nu = B_\nu(T(r)) . \quad (9.12)$$

Hence, the total cooling is  $\sigma T_*^4$ , which yields  $T \propto r^{-3/4}$ .

If we take an annulus of radius  $r$  and width  $dr$ , it covers for an observer a solid angle

$$d\Omega = \frac{2\pi r dr}{d^2} \cos i . \quad (9.13)$$

Here,  $i$  is the inclination angle of the disk with  $i = 0$  being a face-on disk. The emitted flux from that annulus is then

$$F_\nu = I_\nu d\Omega \quad (9.14)$$

and the total flux is the integral over all annuli

$$F_\nu = \frac{2\pi \cos i}{d^2} \int_{r_{\text{in}}}^{r_{\text{out}}} B_\nu(T(r)) r dr \quad (9.15)$$

Since the temperature profile has the same radial dependency as the accretion disk solution, the problem of fitting observed SEDs seen in the previous section remains unsolved. We will hence focus in the following on a different physical disk model, the so-called "flaring" disks that offer a solution to the SED slope problem.

## 9.4 Flaring disks

It was recognized by Kenyon & Hartmann (1987) that a natural explanation for the strong far-infrared flux (i.e. shallow SED slope) of most sources is a flaring (bowl-shaped) geometry of the disks surface, as depicted in Fig. 9.1. The flaring geometry allows the disk to capture a significant portion of the stellar radiation at large radii where the disk is cool, thereby boosting the mid- to far-infrared emission. The flaring geometry adds an extra term to the irradiation angle (Chiang & Goldreich 1997)

$$\phi \approx 0.4 \frac{R_*}{r} + r \frac{\partial}{\partial r} \left( \frac{H_s}{r} \right) , \quad (9.16)$$

where  $H_s$  is the height above the midplane where the disk becomes optically thick to the impinging stellar radiation.

### 9.4.1 Two-layer Chiang & Goldreich disks

The two-layer disk model assumes that the disk can be divided into a hot surface that is directly irradiated by the star and a cooler interior that is vertically isothermal. We will check this simple assumption at a later stage when we consider the results from detailed two-dimensional radiative transfer calculations.

As for the at disks, the thermal balance determines the temperature of the disk  $T_{\text{disk}}$ , but this now depends strongly on the shape of the disk:  $H_s(r)$ . The flaring angle as introduced above can be expressed as

$$\alpha = r \frac{\partial}{\partial r} \left( \frac{H_s}{r} \right) \sim \xi \frac{H_s}{r} \quad (9.17)$$

and the irradiation flux can be written as

$$F_{\text{irr}} = \alpha \frac{L_*}{4\pi r^2} . \quad (9.18)$$

With the cooling flux  $F_{\text{cool}} = \sigma T_{\text{disk}}^4$ , we obtain

$$T_{\text{disk}}^4 = \frac{\xi H_s L_*}{4\sigma \pi r^3} . \quad (9.19)$$

The pressure scale height  $H_p$ , on the other hand, is given by the ratio of the sound speed to the orbital speed  $c_s/\Omega$  and thus depends on the midplane temperature  $T_c$

$$H_p = \sqrt{\frac{kT_c r^3}{\mu m_p GM_*}} \quad (9.20)$$

If we set  $T_c = T_{\text{disk}}$  and if the ratio  $\chi = H_s/H_p$  is known, then the system of equations is closed and can be solved (see appendix Chiang et al. 2001). For the special case that  $\chi$  is constant, we can simplify Eq.(9.19) to

$$T_c^4 = \frac{\xi \chi H_p L_*}{4\sigma \pi r^3} \quad (9.21)$$

and use Eq.(9.20) to see how the disk surface curves as a function of radius

$$\begin{aligned} H_p^8 &= \left(\frac{k}{\mu m_p GM_*}\right)^4 r^{12} T_c^4 \\ H_p^7 &= \left(\frac{k}{\mu m_p GM_*}\right)^4 r^9 \frac{\xi \chi L_*}{4\sigma \pi} \end{aligned} \quad (9.22)$$

which can be expressed as  $H_p = C^{1/7} r^{9/7}$  with  $C$  absorbing all constant terms in the expression. Since we assumed that the surface scale height is a constant fraction of the pressure scale height, this also means that  $H_s \sim r^{9/7}$ , a-posteriori confirming that the disk indeed has a bowl shape. In general, though,  $\chi$  must be computed numerically, and depends on the dust opacity of the disk upper layers. The resulting temperature profile is typically about  $T \sim r^{-0.5}$ .

Fig. 9.4 shows a typical SED calculated from a pure two-layer Chiang & Goldreich disk model. The SED also shows the silicate emission feature at  $10 \mu\text{m}$  from the hot surface.

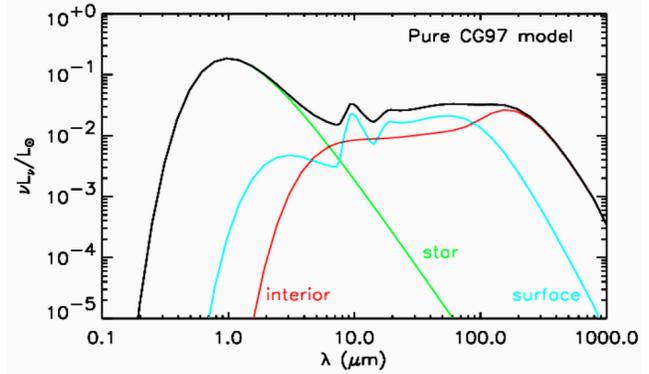


Figure 9.4: SED of a Chiang & Goldreich two-layer model. The cyan line is the thermal emission from the hot disk surface, while the red line denotes the thermal emission from vertically isothermal disk interior. The stellar blackbody is given by the green line.

## 9.4.2 Radiative equilibrium

The concept of radiative equilibrium (or thermal balance) was introduced already in Sect. ?? . Fig. ?? illustrates the basic concept. We will discuss here the more general case of arbitrary grain sizes and size distributions and also the main differences in the dust and gas thermal balance.

### Dust

The thermal balance equation for a grain of radius  $a$  with a frequency dependent absorption efficiency  $\epsilon_\nu$  reads

$$\begin{aligned} Q_+ &= Q_- \\ \pi a^2 \int_0^\infty F_\nu \epsilon_\nu d\nu &= 4\pi a^2 \int_0^\infty \pi B_\nu(T_{\text{dust}}) \epsilon_\nu d\nu \end{aligned} \quad (9.23)$$

We can now look at the general case of a disk with a specific grain size distribution  $dn \sim a^{-3.5} da$ . This is a typical size distribution as it results from a collisional cascade breaking up larger dust grains into smaller entities. If we know the absorption efficiencies as a function of wavelength for each grain size, we can solve the above thermal balance equation numerically and obtain for each grain size a temperature. Alternatively, we can sum the opacity of all dust grains within a certain volume to an average opacity and thus obtain a

representative "dust" temperature. The latter approach is fine if we compute for example spectral energy distributions.

We have to keep in mind, that very small dust grains (VSGs) are stochastically heated. This means that a single photon absorbed by such a VSG will cause a temperature spike within the grain and the grain will take a certain time to re-emit that energy and thus cool down. This is also sometimes referred to as the "duty-cycle" of a grain. In strong radiative environments, multiple photon events can take place before the grain has time to cool. In the most extreme case, this can even lead to the destruction of the VSG.

Most of the time, we use either simple analytic opacities or opacities measured in the laboratory. The following two sections discuss the two approaches very briefly. Most of it is taken from a review on 'Dust in protoplanetary disks' by Henning & Meeus (2009).

## Dust Opacities

The interaction of a radiation field with a system of solid particles can be described by their absorption and scattering cross sections  $C_{\text{abs}}$  and  $C_{\text{sca}}$ , respectively. They describe what fraction of the incoming radiation is absorbed or scattered by a dust particle. The extinction cross section  $C_{\text{ext}}$  is given by

$$C_{\text{ext}} = C_{\text{abs}} + C_{\text{sca}} \quad (9.24)$$

The cross sections depend on the chemical and structural properties of the solid particles, ranging from the atomic scale (chemical composition, crystal and defect structure), to the mesoscopic scale (porosity and inhomogeneities, mantles, surface states), and finally the macroscopic morphology (size and shape distribution, agglomeration, coalescence). Instead of the extinction cross section, the mass extinction coefficient  $\kappa_m$ , defined as the extinction cross section per unit particle mass, is often used to characterize the extinction of light. This quantity is actually more appropriate in describing how much light is removed from the incoming radiation field by a fixed mass of particles. For a spherical particle,

$$\kappa_m = \frac{3Q_{\text{ext}}}{4a\delta} \quad , \quad (9.25)$$

where  $a$  is the particle radius,  $\delta$  the material density, and  $Q_{\text{ext}}$  the extinction efficiency (extinction cross section per geometrical cross section,  $C_{\text{ext}}/\pi a^2$ ).

The qualitative features of the absorption and scattering of light strongly depend on the ratio between the wavelength of the incident light  $\lambda$  and the size of the particle (for a spherical particle, the radius  $a$ ). We can distinguish three cases:

1. Geometrical optics (Size parameter  $x = 2\pi a/\lambda \gg 1$ ): The propagation of light is described by rays which are reflected and refracted at the surface of the scatterer and finally transmitted, according to Snell's law and the Fresnel formulae. The scattering of a wave incident on a particle can be described as a combination of a reflected and a transmitted wave. For absorbing materials, light can penetrate only within the skin depth. Scattering, therefore, is mainly a surface effect and the absorption cross section becomes proportional to the area of the particle as the radius increases. In this case, the mass absorption coefficient for a sphere scales roughly as  $1/a$ . We should note that for very large size parameters, the extinction efficiency  $Q_{\text{ext}}$  approaches the limiting value 2.
2. Wave optics ( $\lambda \sim a$ ): The angular and wavelength dependence of the scattered radiation is dominated by interferences and resonances. For spherical particles, this is the domain of Gustav Mie's (1908) scattering theory, which is often applied in astrophysics.
3. Rayleigh limit (Size parameter  $x = 2\pi a/\lambda \ll 1$ ): If, in addition, we have  $|m|x \ll 1$ , where  $m = n + ik$  is the (complex) refractive index of the particle, we are in the quasi-static limit. Then, both the incident and the internal field can be regarded as static fields. In this regime, phase shifts over the particle size are negligible. For non-magnetic particles this implies that it is generally sufficient to only consider the dipolar electric mode. The interaction of infrared and (sub)millimeter radiation with sub-micron sized

grains can generally be considered as good examples of the quasi-static case. However, particles with high imaginary parts of the refractive index (metals, semiconductors, crystalline grains) and particles of somewhat larger sizes can easily violate the conditions for the quasi-static limit, even at infrared wavelengths.

### Analytic dust opacities

In the case of large black body grains, we can assume simple analytic cross sections derived from Mie theory. The basic idea is that the grains absorb very efficiently in the wavelength regime  $\lambda \ll 2\pi a$ , so when the wavelength of radiation is much smaller than the size of the grain. At longer wavelength, the efficiency  $\epsilon_\lambda$  decreases linearly with  $\lambda$

$$\epsilon_\lambda = \begin{cases} \frac{2\pi a}{\lambda} & \text{for } \lambda > 2\pi a \\ 1 & \text{for } \lambda \leq 2\pi a \end{cases} \quad (9.26)$$

### Laboratory dust opacities

The opacities of real grains are more complex and Fig. 9.5 shows as an example the absorption efficiency  $Q_{\text{abs}}$  (same as  $\epsilon$  used above) for silicate grains (astronomical silicate - often represented by the optical constants of olivine) calculated from optical constants measured in the laboratory as a function of wavelength. The various curves are labelled by the grain size  $a$  in  $\mu\text{m}$ . Large grains,  $a = 1 \mu\text{m}$ , are very efficient absorbers over the entire wavelength range displayed here,  $0.03 < \lambda < 1 \mu\text{m}$ . Smaller grains display a more complex wavelength dependent behavior with resonances and peaks. It can be clearly seen that the critical wavelength at which grains become less efficient absorbers depends on the grain size. For smaller grains, this break point lies at shorter wavelength, e.g. for  $0.1 \mu\text{m}$  grains, the absorption efficiency has dropped to 0.1 around  $6000 \text{ \AA}$ . This corresponds roughly to  $2\pi\lambda$

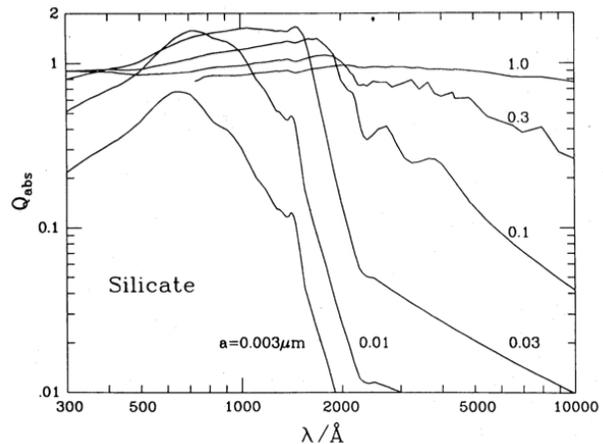


Figure 9.5: Absorption efficiency  $Q_{\text{abs}}$  (same as  $\epsilon$  above) as a function of wavelength for spherical grains of 'astronomical silicate'. Curves are labelled by the grain radius  $a$  (Figure from Draine & Lee 1984).

### Dust temperature profile

In a numerical model, the disk is often split into one-dimensional vertical slices. Each slice then forms a separate 1D problem (often referred to as 1+1D models). We assume here that the slices are independent from each other, i.e. there is no energy or information exchange in radial direction. This reduction to 1D makes the solution of the dust thermal balance very fast and computationally inexpensive. However, it assumes a priori an irradiation angle  $\phi$  and thus produces either flat or flared disk geometries.

In reality, the disk problem is rather two-dimensional or even three-dimensional in the presence of spiral structures or other non-axisymmetric density perturbations. Solving the full two-dimensional radiative transfer is more challenging, but it also allows to re-calculate the hydrostatic equilibrium and hence the irradiation angle self-consistently. This requires an iterative loop between solving the hydrostatic equilibrium and performing the 2D radiative transfer.

The results from these more sophisticated models largely confirm the simple 1+1D results. However, they also show new structure solutions, for example so called 'self-shadowed' disks, where the inner disk rim can be vertically extended (and optically thick) in such a way that it casts a shadow on the disk behind it.

This suppresses part of the near-IR flux (Fig. 9.9). This will be discussed in some more detail in the next section.

A closer look at the physics of an irradiation-dominated disk (be it flat or flared) reveals that its surface temperature is generally higher than its interior temperature. Dust grains in the surface layers are directly exposed to the stellar radiation, and are therefore hotter than dust grains residing deep in the disk which only see the infrared emission by other dust grains. The temperature difference is typically a factor of 2 – 4 for non/weakly-accreting disks (see curve labeled -9 in Fig. 9.6).

For non-negligible accretion, on the other hand, the disk is heated from inside as well, producing a temperature minimum somewhere between the equatorial plane and the surface layer (see other curves in Fig. 9.6). Because of the disks high optical depth, radiation cannot penetrate efficiently to the midplane. These dense layers are instead primarily heated by local viscous dissipation (Frank et al. 1992). The viscous heating rate is

$$\Gamma_{\text{vis}} = \frac{9}{4} \rho_g \nu \Omega^2 \text{ s}^{-1} \text{ cm}^{-3} , \quad (9.27)$$

where  $\nu$  is here again the viscosity.

## Gas

The gas heating/cooling balance is far more complex, due to the multitude of cooling lines (atomic and molecular) and intricate coupling between the chemical and thermal balance. Collisions between gas particles and dust grains thermally couple the gas to the dust ( $T_{\text{gas}} = T_{\text{dust}}$ ) only if radiative processes become ineffective due to high optical depth ( $\tau \gg 1$ ) and/or densities become very large ( $n \gtrsim 10^7 \text{ cm}^{-3}$ ).

Dust heating and cooling is a continuum process, while gas heating and cooling mostly occurs at discrete wavelengths (line absorption/emission) except in a few cases like  $\text{H}^-$ . In the following, we briefly summarize the most relevant heating and cooling processes of the gas. They largely resemble those found for molecular clouds and the interstellar medium, but there are a few extra ones that become only relevant at the high densities of disks and/or under very high irradiation conditions (i.e. close to the star).

Table 9.1 provides an overview of the various heating and cooling processes that are important for setting the gas thermal balance in protoplanetary disks. Radiative heating processes such as photoelectric, PAH heating, C ionisation,  $\text{H}_2$  photodissociation and X-ray heating dominate in the surface regions of the disks, the optically thin disk atmospheres, where radiation can either directly reach or can be efficiently scattered.

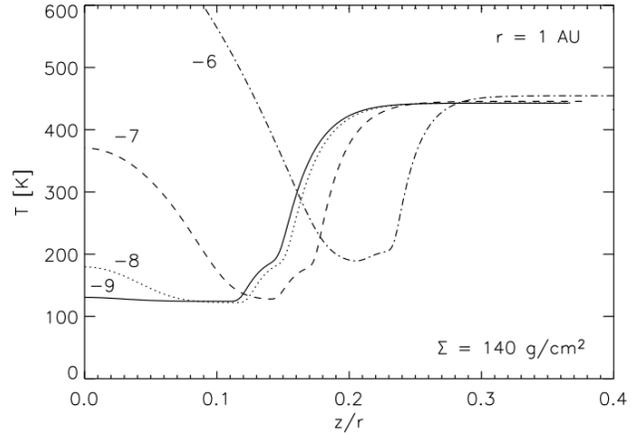


Figure 9.6: Vertical temperature distribution of an irradiated  $\alpha$ -disk at 1 AU, for a fixed  $\Sigma$  (chosen to be that of a disk model with  $\dot{M} = 10^{-8} M_{\odot}/\text{yr}$  for  $\alpha = 0.01$ ), but varying  $\dot{M}$ , computed using the models of D’Alessio et al. (1998). The labels of the curves denote the 10-log of the accretion rate in  $M_{\odot}/\text{yr}$ . Figure and caption taken from the PPV review by Dullemond et al. (2007).

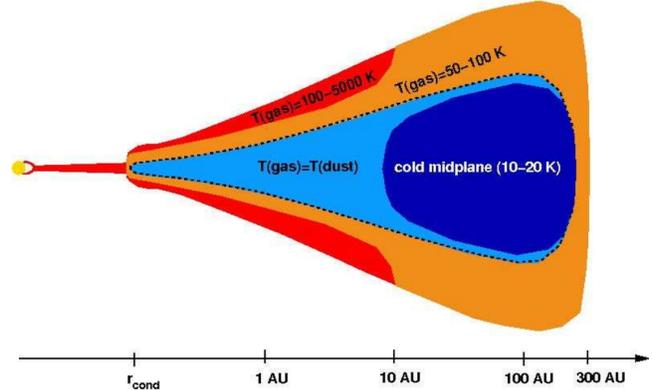


Figure 9.7: Schematic gas temperature structure of a protoplanetary disk. The disk surface is heated by the photoelectric effect on grains (FUV) and/or by X-ray ionization ( $r < 10 \text{ AU}$ ). As the radiation is attenuated through optical depth and/or distance, the gas becomes cooler. Below a continuum optical depth of  $\tau \sim 1$ , shown by the dashed line, gas and dust can efficiently couple through collisions. A prominent feature of most models is a hot layer just above the  $\tau \sim 1$  surface. Figure and caption taken from the Exoplanet book chapter by Roberge & Kamp (2010).

Table 9.1: Heating and cooling processes of the gas in protoplanetary disks. Thermal accomodation on dust grains is a heating or cooling term depending on the sign of the temperature difference  $T_{\text{gas}} - T_{\text{dust}}$ .

Heating	Cooling
photoelectric heating	O I, C II, C I fine-structure lines
PAH heating	PAH recombination cooling
viscous ( $\alpha$ ) heating	CO/OH ro-vibrational lines
cosmic rays	o/p H <sub>2</sub> O rotational lines
C photo-ionisation	o/p H <sub>2</sub> quadrupole lines
collisional de-excitation of H <sub>2</sub>	Si II, S II, Fe II semi-forbidden lines
H <sub>2</sub> formation on grains	Lyman $\alpha$
H <sub>2</sub> photodissociation	Mg II h&k lines
IR background line heating	O I 6300 Å line
X-ray heating	
thermal accomodation on grains	

Viscous heating dominates the high density midplane regions and depends via the viscosity  $\nu$  on the mass accretion rate (Eq.(9.27)). Fig. 9.6 illustrates at which radii this process dominates over irradiation heating depending on the mass accretion rate.

Fine structure line cooling dominates mostly in the low density disk atmospheres. These transitions have often very low critical densities and determine the gas temperature even at the outer disk radius. In the inner higher density regions, ro-vibrational line cooling and also electronic transitions become more important in setting the gas thermal balance. Gas-grain collisions can act both as heating or cooling term depending on the sign of the temperature difference. Whenever gas-grain thermal accomodation dominates the energy balance, gas and dust temperatures are equal. This happens mostly at high densities and large optical depth.

## Gas temperatures

Down to a continuum optical depth of  $\sim 1$ , the gas temperature in the disk surface decouples from the dust temperature and ranges from a few thousand to a few hundred Kelvin (e.g. Kamp & Dullemond 2004; Jonkheid et al. 2004, Nomura & Millar 2005). The dominant heating source in the presence of ultraviolet radiation ( $\lambda < 3000 \text{ \AA}$ ) is the photoelectric effect on grains, either small polycyclic aromatic hydrocarbons (PAHs) or larger grains. The ejected electrons transfer their kinetic energy to the gas through collisions. X-rays also heat the gas via the generation of energetic electrons through primary and secondary ionization of the gas. This heating plays an important role at fairly high X-ray luminosities ( $L_X \sim 10^{30} - 10^{31} \text{ erg/s}$ ; Glassgold et al. 2004). At lower luminosities ( $\sim \text{few } 10^{29} \text{ erg/s}$ ), their impact is restricted to intermediate depths at larger radii ( $r > 10 \text{ AU}$ ). While far UV (FUV, photon energies  $6 < h\nu < 13.6 \text{ eV}$ ) radiation typically gets absorbed within column densities of  $10^{21} \text{ cm}^{-2}$ , X-rays can penetrate roughly a factor 100 deeper and start to dominate the heating in layers where FUV processes are largely suppressed. However, the general importance and penetration depth of X-rays depends largely on the position of the X-ray source with respect to the disk surface. The scattering of UV photons on dust grains seems to be more efficient than X-ray scattering. If the UV and X-ray source is the star itself, UV heating will dominate down to larger optical depth. If the X-ray source is located above the disk, those more energetic photons can penetrate much deeper due to their larger irradiation angle.

Fig. 9.7 illustrates the two-dimensional temperature structure of a gas disk. In the optically thick layers of the disk, collisions efficiently couple the gas and dust temperatures. This produces a temperature profile that is roughly isothermal in the vertical direction and declines approximately as  $r^{-0.5}$  in the radial direction, as predicted by the simple two-layer flared disk model of Chiang & Goldreich (1997). The temperature profiles in the upper disk layers, however, are more complex. Cooling of the surface layers proceeds through atomic emission, mainly the [C II] and [O I] fine structure lines (for gas temperatures of a few hundred Kelvins), Fe II electronic transitions, Lyman  $\alpha$ , and O I 6300 Å (for temperatures of a few thousand Kelvins).

While photoelectric heating scales roughly as density squared, the cooling lines often form out of local thermodynamic equilibrium (LTE; level populations follow from the Boltzmann and Saha equations) and thus scale linearly with  $n$ . Together with the weak dependence of the photoelectric heating process on gas temperature, this can lead to strong temperature gradients. Most gas disk models display a hot temperature layer at intermediate depth, just above the height at which the gas and dust thermally couple. The position of this layer strongly depends on the  $\text{H}_2$  abundance. In protoplanetary disks,  $\text{H}_2$  forms on grain surfaces and its abundance is related to assumed grain properties, such as surface area, sticking, and opacity. As soon as the Lyman and Werner  $\text{H}_2$  absorption bands in the UV become optically thick (self-shielding),  $\text{H}_2$  formation on dust grain surfaces outweighs photodissociation and the kinetic energy with which the  $\text{H}_2$  molecule is released from the grain surface becomes an important energy source in the gas. The sensitivity of the gas temperature on the location of the  $\text{H}/\text{H}_2$  boundary has been noted by Kamp and Dullemond (2004) and later by Woods and Willacy (2008).

### 9.4.3 Inner disk radius

The inner disk radius for models of irradiated disks is often set to the dust condensation radius. One should keep in mind that this has two reasons, a practical one and a physical one: (1) a practical aspect of the irradiated disk models is that they use the dust opacities to obtain the dust temperature assuming radiative equilibrium. The gas temperature is simply assumed to be well coupled to that of the dust. Hence, the disk vertical structure — which is set by the gas pressure — is derived from that of the dust. (2) It is very likely that the disk structure will change around the dust condensation radius, because the elemental composition of the gas changes, and the disk structure will entirely be set by gas opacities. In fact very little modelling has been done on the detailed structure of these inner gaseous disks. Observations of this inner gas disk are scarce (mostly indirect evidence from spectral lines and interferometry) and new larger facilities (e.g. ELT's) that can spatially resolve these regions are necessary to push the models into this regime.

The dusty part of the disk can be expected to have a relatively abrupt inner edge at about 0.5 AU for a  $50 L_\odot$  star (scaling roughly with  $\sqrt{L_\star}$ ). If the gas inward of this dust inner rim is optically thin, which appears to be mostly the case (Muzerolle et al., 2004), then this dust inner rim is illuminated by the star at a  $\sim 90$  degree angle, and is hence expected to be much hotter than the rest of the disk behind it which is irradiated under a shallow angle  $\phi \ll 1$  (Natta et al. 2001). Consequently it must be hydrostatically puffed-up, although this is still under debate.

Natta et al. (2001) showed that the emission from such a hot inner rim can explain the near-infrared bump seen in almost all Herbig Ae/Be star SEDs (see e.g. Meeus et al. 2001). This is a natural explanation, since dust sublimation occurs typically around 1500 K, and a 1500 K blackbody bump fits reasonably well to the near-infrared bumps in those sources (Fig. 9.8).

### 9.4.4 SEDs

Because of the shallow incidence angle of the stellar radiation  $\phi \ll 1$ , the vertical optical depth of the warm surface layer is very low. The layer produces optically thin emission at a temperature higher than the effective temperature of the disk. The frequency-integrated flux of this emission is the same as that from the disk interior. As a consequence, the thermal radiation from these surface layers produces dust features

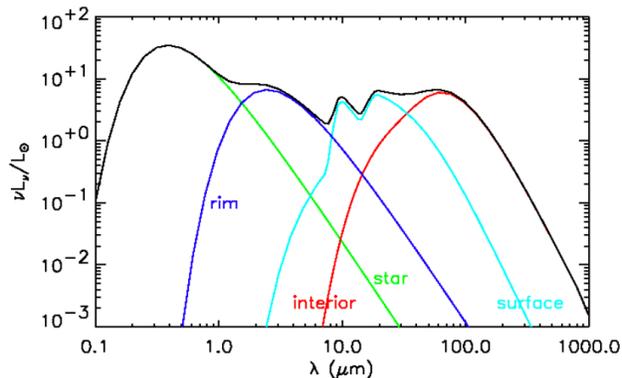


Figure 9.8: SED of a 1+1D disk model with a puffed-up inner rim. The contribution of the star, the disk surface, disk interior and inner rim is denoted by the various colors. The inner rim is indeed very well approximated by a single 1500 K blackbody.

in emission. This is exactly what is seen in nearly all non-edge-on T Tauri and Herbig Ae/Be star spectra (see Fig. 9.3), indicating that these disks are nearly always dominated by irradiation.

Fig 9.9 shows the effect of geometry on the calculated SED (without scattering). Highly flared disks show strong mid-IR fluxes, while flat or 'self-shadowed' disks show very weak mid-IR fluxes (dash-dotted line). A higher rim means a larger emitting area (at roughly the same temperature of 1500 K) and hence slightly higher near-IR fluxes. If the inner rim is artificially moved to a larger radius, the grains there have temperatures much lower than the dust condensation temperature and the near-IR excess is absent.

## 9.5 Disk dispersal

We have seen in the last chapter that angular momentum transport affects the evolution of disks, but we have so far assumed that disk mass is only lost through accretion onto the central star. Observations tell us that this mass accretion rate decreases as a function of time. However, disk evolution can also be influenced by other dynamical processes such as disk winds and jets driven by stellar magnetic fields, dynamical encounters with nearby stars and photoevaporation by either the central star or an external source. Hollenbach et al. (2000) discuss all these processes in detail and conclude that for most systems, viscous evolution and photoevaporation are the dominant processes that shape the overall disk evolution. The other processes discussed can dominate over short timescales or in small regions of the disk (such as magnetically-launched jets very close to the star), or can be important for a small fraction of stars (such as tidal stripping by nearby stars, which affects  $\sim 10\%$  of the stars in the Orion nebula cluster), but do not dominate the evolution of most T Tauri disks. The following sections present a summary of photoevaporation by the central star or an external source and briefly describe how disk evolution proceeds in the presence of these two phenomena.

### 9.5.1 Photoevaporation

The above section has shown that in the surface layers of the disk the gas temperature can become very high, greatly exceeding the dust temperature. The warm surface gas can flow off the disk and escape the gravity of the star. Since the heating process responsible for these high temperatures is the radiation from the central star or a nearby O-star, this process is called photoevaporation. The viscous evolution (i.e. accretion and spreading) of the disk, discussed in Chapter 8, can be strongly affected by this photoevaporation process.

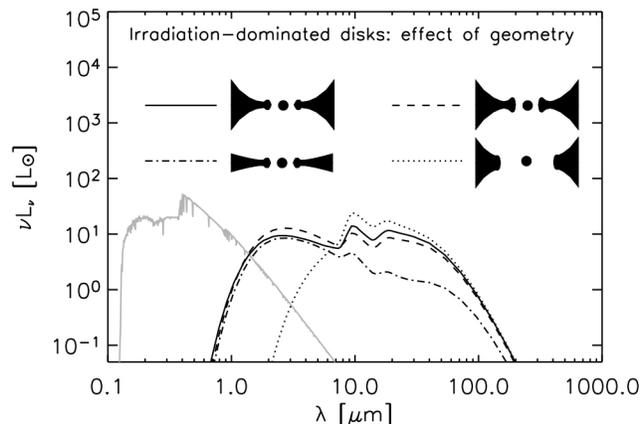


Figure 9.9: Overall SED shape for non-accreting disks with stellar irradiation, computed using the 2-D radiative transfer tools from Dullemond & Dominik (2004a). The stellar spectrum is added in grey-scale. Scattered light is not included in these SEDs. Solid line is normal flaring disk with inner dust rim; dashed line is when the rim is made higher; dot-dashed line is when the flaring is reduced (or when the disk becomes self-shadowed); dotted line is when the inner rim is at  $10\times$  larger radius. Figure and caption taken from the PPV review by Dullemond et al. (2007).

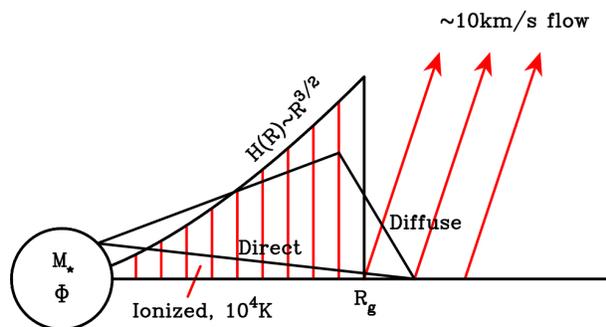


Figure 9.10: Schematic illustration of a photoevaporative disk wind. Inside the gravitational radius  $r_g$  an ionized disk atmosphere is formed, with scale-height  $H(R)$ . Outside  $r_g$  material is evaporated away from the disk. (Figure adapted from Hollenbach et al. 1994, 2000).

Typically, it significantly shortens the lifetime of a disk compared to pure viscous evolution. Photoevaporation can also create inner holes or truncate the outer disk. This has relevance to observations of such disks, such as the percentage of young stars with infrared excess versus their age, or the inferred large inner holes of some disks. It has also far-reaching consequences for the formation of planets, as we will discuss in some of the later chapters. We will discuss here photoevaporation by the central star and turn to external sources in the next section.

### EUV evaporation

EUV photons ionize the hydrogen in the very upper layers of the disk and heat it to a temperature of  $\sim 10^4$  K, independent of radius. FUV photons penetrate deeper into the disk and heat the gas to  $T_g \sim 100 - 5000$  K, depending on the intensity of the FUV flux, the gas density and the chemistry. Whether the EUV or FUV heating is enough to drive an evaporative flow depends on how the resulting thermal speed (or sound speed) compares to the local escape speed from the gravitationally bound system. A characteristic radius for thermal evaporation is the gravitational radius  $r_g$  (Fig. 9.10), where the sound speed equals the escape speed

$$r_g = \frac{GM_* \mu m_p}{kT} \sim 100 \text{ AU} \left( \frac{T_g}{1000 \text{ K}} \right)^{-1} \left( \frac{M_*}{M_\odot} \right) \quad (9.28)$$

Early analytic models made the simple assumption that photoevaporation occurred for  $r > r_g$ , and that the warm surface was gravitationally bound for  $r < r_g$ . However, a closer look at the gas dynamics shows that this division happens not at  $r_g$  but at about  $0.1 - 0.2 r_g$  (Liffman 2003, Adams et al. 2004, Font et al. 2004), and that this division is not entirely sharp. In other words, photoevaporation happens mostly outside of the critical radius  $r_{cr} \sim 0.15 r_g$ , though a weak evaporation occurs inside of  $r_{cr}$ .

With  $T \sim 10^4$  K, the critical radius for EUV-induced photoevaporation is  $r_{cr}(\text{EUV}) \sim 1 - 2 (M_*/M_\odot)$  AU. However, there is no fixed  $r_{cr}(\text{FUV})$  because the FUV-heated gas has temperatures that depend on FUV flux and gas density, i.e., on  $r$  and  $z$ . Therefore,  $r_{cr}(\text{FUV})$  depends on  $r$  and  $z$ , and may range from  $3 - 150$  AU for solar mass stars.

Hollenbach et al. (1994) first outlined the essential physics of EUV-induced flows by the central star and presented an approximate analytic solution to the mass loss rate for a disk larger than  $r_g$ . The basic physics is the Strömgen relation  $\Phi_{\text{EUV}} \approx \alpha_r n_e^2 r^3$ , where  $\alpha_r$  is the hydrogen recombination coefficient and  $n_e$  is the electron density in the ionized surface gas. This sets the hydrogen nucleus (proton) number density at the base of the flow  $n_b \sim \Phi_{\text{EUV}}^{1/2}$ , and therefore an identical proportionality for the mass loss rate

$$\dot{M}_{\text{EUV}} \sim 4 \times 10^{-10} \left( \frac{\Phi_{\text{EUV}}}{10^{41} \text{ s}^{-1}} \right)^{0.5} \left( \frac{M_*}{M_\odot} \right)^{0.5} \quad (9.29)$$

in units of  $M_\odot/\text{yr}$ . This result applies for both high and low mass central stars, and is valid for a weak stellar wind. The effect of a strong stellar wind is such that the ram pressure reduces the scale height of the atmosphere above the disk and the EUV photons are allowed to penetrate more easily to larger radii. This increases the mass-loss rate from the outer parts of the disk. It is noteworthy that the diffuse EUV field, caused by recombining electrons and protons in the disks ionized atmosphere inside  $r_{cr}$ , controls the EUV-induced mass-loss rates for disks with no or small inner holes ( $< r_{cr}$ ). This effect negates any potential for self-shadowing of the EUV by the disk.

### FUV evaporation

Although FUV-heated layers have lower temperatures than the EUV-heated skin they are at higher densities and may equally well initiate the flow and determine the mass flux as EUV photons (see Johnstone et al. 1998 for a similar situation for externally illuminated disks). Gorti and Hollenbach (2009) find that the FUV-photoevaporative flow typically originates at vertical heights where  $T_g \sim 100 - 200$  K, yielding  $r_{cr} \sim 50 - 100$  AU. For  $r > 50$  AU, the FUV photoevaporation dominates. On the other hand, EUV photons (with  $r_{cr} \sim 1$  AU) affect the planet forming regions at  $r \ll 50$  AU more than the FUV photons.

## X-ray evaporation

Young stars also emit X-rays and we have seen in Sect. 9.4.2 how this affects the gas temperature in the disk surface. It is thus conceivable, that we have next to EUV and FUV photoevaporation also an evaporation driven by X-rays. Alexander et al. (2004), Gorti & Hollenbach (2009) and Ercolano et al. (2009) obtain different conclusions on the relevance of X-ray photoevaporation. While the first two groups dismiss it, the last one finds that it can dominate the photoevaporative mass loss in the 10-40 AU regime (10 times higher than EUV driven mass loss rates). Clearly more work is needed to settle the relative role of the various photon energy regimes in the disk photoevaporation process.

## Photoevaporation and viscous evolution

Let us consider a disk that is actively accreting onto the star (see Chapter 8). In general, if the photoevaporation drills a hole somewhere in the disk or eats its way from outside in, the forces of viscous spreading tend to move matter toward these photoevaporation regions, which can accelerate the dissipation of the disk. If the disk has a steady accretion rate  $\dot{M}$ , then a gap forms once  $\dot{M}_{\text{evap}} \sim r^2 \dot{\Sigma}$  exceeds  $\dot{M}$ . Since  $r^2 \dot{\Sigma} \sim r^{-0.5}$  for EUV photoevaporation beyond  $r_{cr}$ , the gap first forms at the minimum radius ( $\sim r_{cr}$ ) and then works its way outward.

The simulation shown in Fig. 9.11 starts from a given disk structure of about  $0.05 M_{\odot}$  (marked with Start in the figure). Initially the disk accretes and viscously spreads (solid lines). At  $t = 6 \times 10^6$  yr the photoevaporation starts affecting the disk. Once the EUV-photoevaporation has drilled a gap in the disk at  $\sim 1$  AU, the destruction of the disk goes very rapidly (dashed lines). The inner disk accretes onto the star on an inner disk viscous timescale of  $\sim 10^5$  yr. Once the inner hole is produced, the diffuse flux from the atmosphere of the inner disk is removed and the attenuation of the direct flux by this same atmosphere is also removed. This enhances the EUV photoevaporation rate by the direct EUV flux from the star, and the effect magnifies as the inner hole grows as  $\dot{M}_{\text{EUV}} \sim r_{\text{inner}}^{1/2}$ , again derivable from a simple Strömgen criterion. This leads to a rapid erosion of the outer disk from inside out. In this model the disk viscosity spreads to  $> 1000$  AU; however, FUV-photoevaporation (not included) will likely truncate the outer disk.

The disk evolution can be split into three phases: (1) the viscous phase which lasts a few Myr, (2) the gap-opening phase which is shorter than  $10^5$  yr and (3) the disk clearing phase which lasts a few  $10^5$  yr. During the viscous evolution phase, most of the disk mass is accreted onto the star and the disk spreads to large radii. Photoevaporation is negligible. During the second phase, the photoevaporative wind drills a gap into the disk. This stops mass supply from the outer disk, and the inner disk accretes onto the star on its viscous timescale. During the clearing phase (third phase), the inner disk is optically thin to irradiating and ionizing photons and evaporates from inside out.

## 9.5.2 External evaporation

In crowded regions of star formation, especially in the presence of high mass stars, external photoevaporation can also play a significant role. This effect has been observed e.g. for the 'proplyds' in the Orion nebula (Fig. 9.12, O'Dell et al. 1993). These silhouette disks around low mass stars display cometary-shaped nebulae that point away from the illuminating massive star. They are also bright in typical nebulae emission lines such as  $H\alpha$  and  $[N\text{II}]$ . It became clear that these disks were photoevaporated by the nearby bright

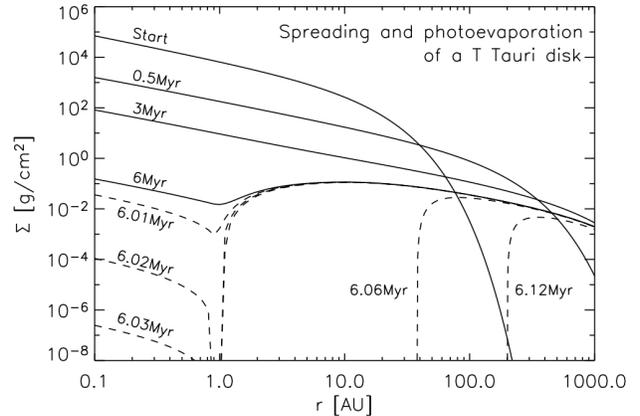


Figure 9.11: Evolution of the surface density of a EUV-photoevaporating disk (Figure adapted from Alexander et al. 2006b). The simulation assumes that evaporation occurs outside  $r_{cr}$ .

O stars. The EUV photons of these stars are absorbed within a thin layer of atomic hydrogen gas above the disk surface. Hence, the photoevaporative flow is mainly driven by the non-ionizing (hydrogen) FUV photons. The cometary-shaped nebulae are explained by a combination of the photoevaporative flow from the disk surface and the dynamic interaction with the O star winds. Typical mass loss rates are of the order of  $10^{-7} M_{\odot}/\text{yr}$ , but can differ substantially depending on the distance from the illuminating source.

Even though one might be lead to the conclusion that this strong photoevaporation leads to extremely short lifetimes for disks, this is not necessarily true. The gravitational radius for FUV photoevaporation is much larger than that for EUV photoevaporation, because the flow is driven at substantially lower temperatures, typically below 1000 K (see Eq.(9.28)). Since material inside  $\sim 0.15 r_g$  is mostly bound (only weak photoevaporative flow), disks of that size can survive over much longer timescales. There is still a discrepancy between observed mass loss rates for these externally evaporated disks and those predicted by models. Only the low mass loss rates predicted by hydrodynamical models can explain the current lifetime of these systems.

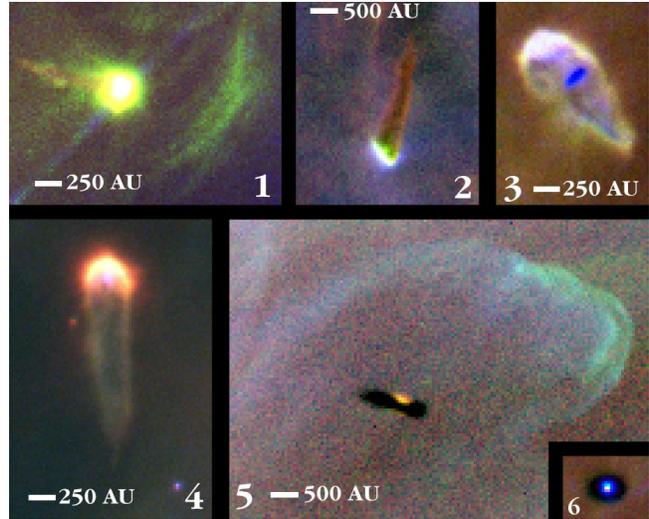


Figure 9.12: HST image of the protoplanetary disks in the Orion nebula.

## Chapter 10

# Observations of Disks

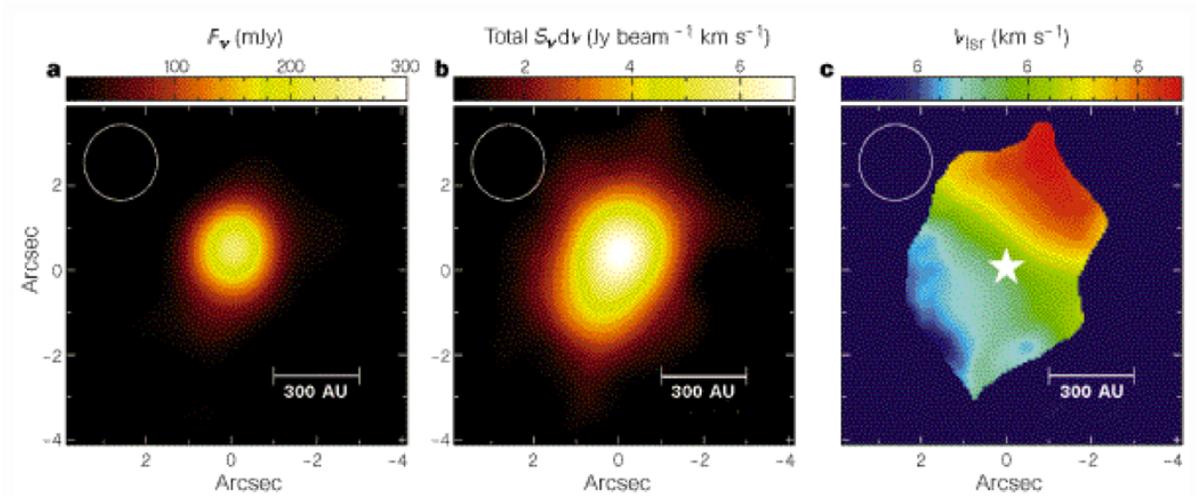


Figure 10.1: Continuum images and line maps of the intermediate mass Herbig Ae star MWC 480: 1.3 mm continuum emission (left), CO J=2-1 integrated intensity map (middle) and CO J=2-1 mean velocities (right) (Figure from Mannings & Sargent 1997).

After several more theoretically oriented chapters on disks, we focus here on the observational material and how it can test the various models and scenarios of disk evolution outlined before. Due to the wealth of observations, especially in the IR wavelength regime over the past 15 years, we can only present a selective review and do not claim completeness anywhere.

Even though theories of star formation have long predicted the existence of protoplanetary disks, observational evidence of their presence came only in the 80s, first only indirectly inferred from the presence of optical jets and bipolar outflows that point towards an axially, and not spherically symmetric distribution of material. Additional indirect evidence can come from polarization measurements that showed jets and outflows to be aligned with the magnetic field of the young stellar object (YSO). Spectroscopically, asymmetric forbidden line profiles (e.g. [O I] 6300 Å) provided additional dynamic information: the blue-shifted component is clearly observed, while the red-shifted emission is occulted by an optically thick disk with dimensions of 10-100 AU. Another indirect evidence for disks is the observation of IR excesses around young stars (e.g. with the IRAS, ISO and Spitzer satellites all operating at infrared wavelengths). However, the interpretation of an SED is to some extent degenerate and does not carry dynamic information of e.g. the disk rotation. The IR excess could be clearly attributed to a flat disk-like structure since a spherical envelope containing  $0.01 M_{\odot}$  of gas and dust would totally obscure the stellar object ( $A_V \sim 500$  mag for a 50 AU sphere). Typically observed extinctions towards T Tauri stars are smaller than a few mag. However, disentangling the disk SED contribution from that of tenuous remnant envelope material requires generally additional observations of the source such as e.g. imaging to fix parameters such as disk extension, inclination and grain properties.

Only sub-mm observations of rotational lines such as CO J=2-1 provided an unambiguous proof that the observed IR and mm-excess is associated with a disk of gas and dust in Keplerian rotation (Fig. 10.1).

A second independent proof are optical scattered light images of disks or near-IR thermal images. Prominent examples are the HST images of the proplyds in Orion, scattered light images of disks around Herbig AeBe stars or debris disks. Adaptive optics and 8-10 m class telescopes have opened this avenue also for ground based telescopes. Fig. 10.2 shows a VLT/ISAAC image of the 'flying saucer', an edge-on disk at the periphery of the  $\rho$  Ophiuchi dark cloud ( $d = 140$  pc). The dark dust lane seen in the image corresponds to an optically thick dust disk with an extension of 300 AU.

In the following sections, we will discuss how observational surveys help us to understand disk properties and evolution in a statistical way. We will also discuss a few individual extremely well studied objects where we can test details in our understanding of disk physics and chemistry.

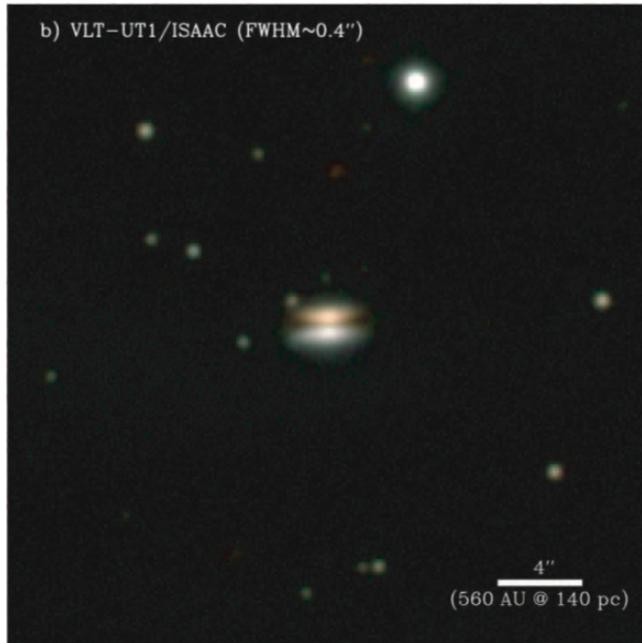


Figure 10.2: The color composite image combines J (blue), H (green), and  $K_s$  (red), with logarithmic color stretch. The white balance is obtained from the brightest background star. The image size is  $30'' \times 30''$ . The image orientation is north on the top and east on the left. The observations were obtained with the VLT-UT1/ISAAC. The pixel size is  $0.148''$ , the seeing FWHM  $\sim 0.4''$  (Figure and caption from Grosso et al. 2003).

## 10.1 SEDs

Spectral energy distributions of protoplanetary disks have been discussed extensively already in earlier chapters. We restrict the discussion here to physical quantities that can be derived from the SED as well as evolutionary scenarios that can be tested.

### 10.1.1 Measuring the disk mass

At very long wavelengths, mm-regime, even massive protoplanetary disks become optically thin. At least the outer disk beyond 50 AU, which generally contains most of the disk mass. This allows us to measure the disk dust mass directly from mm-fluxes.

The flux can be expressed as

$$F_\nu \sim \frac{A_d}{d^2} B_\nu(T_{\text{dust}}) (1 - e^{-\tau_\nu}) \quad , \quad (10.1)$$

where  $A_d$  is the projected area of the disk,  $d$  the distance to the object, and  $\tau_\nu$  the optical depth at frequency  $\nu$ . For small optical depth in the Rayleigh Jeans limit, we can approximate this to

$$\begin{aligned} F_\nu &\sim \frac{A_d}{d^2} k T_{\text{dust}} \nu^2 \tau_\nu \\ &\sim \frac{A_d}{d^2} k T_{\text{dust}} \nu^2 \frac{M_{\text{dust}}}{A_d} \quad . \end{aligned} \quad (10.2)$$

This assumes that the 'average' disk optical depth is directly proportional to the 'average' disk column density. Using power law opacities at these long wavelength  $\kappa_\nu \sim \kappa_0 (\nu/\nu_0)^{-\beta}$  (see also Sect. 9.2), we can write the relation between mm-flux and disk mass as

$$F_\nu \sim \kappa_0 \nu^{2+\beta} T_{\text{dust}} M_{\text{dust}} \quad . \quad (10.3)$$

If we can estimate the typical dust temperature from the SED, we can obtain an estimate of the disk mass from the mm-flux. Since the mm-flux is dominated by emission from larger grains, this measurement is often biased towards large grains and depends on the assumption of the power law opacity ( $\kappa_0$  and  $\beta$ ). Even though this provides good estimates of the outer disk mass contained in large grains, we still need to employ some conversion factor to turn this into a gas mass. Often, one assumes the canonical factor 100 as the mass ratio between gas and dust. However, from various disk evolution scenarios as well as from observations, we know that this ratio is not constant over time and that it can even vary with location in the disk, e.g. surface versus interior and inner disk versus outer disk. We will discuss below alternative methods to measure the gas mass of the disk from e.g.  $^{13}\text{CO}$  lines.

### 10.1.2 Surveys at mm wavelength

From single-dish mm-surveys such as the one carried out by Beckwith et al. (1990), we know that in the Taurus star forming region  $\sim 40\%$  of classical T Tauri stars and only less than  $30\%$  of weak-line T Tauri stars have disks. A similar survey was carried out by André & Montmerle (1994) for the Ophiuchus star forming region. They found a detection rate of  $\sim 15\%$  for class III sources and  $\sim 60 - 70\%$  for class II and I sources. Osterloh & Beckwith (1995) found in a larger survey that more than  $50\%$  of the classical T Tauri stars were detected as opposed to only  $\sim 15\%$  of the weak-line T Tauri stars.

Fig. 10.3 shows results of the recent very sensitive ( $3\sigma = 5 \times 10^{-4} M_\odot$ ) sub-mm survey of Andrews & Williams (2005) of 153 targets in the Taurus-Auriga star forming region. The median disk contains  $\sim 0.5\%$  of the mass of the central star; the typical mass is  $0.005 M_\odot$ . This is a factor of  $2 - 3$  less than the expected minimum mass solar nebula (Weidenschilling 1977).

The submillimeter SED follows a power-law  $F_\nu \sim \nu^n$  with  $n \approx 2$  ( $n$  is also referred to as the submm color; in the previous section, we have used  $n = 2 + \beta$ , where  $\beta$  is the power law index of the opacity). This is much lower than the  $n \approx 4$  seen in the ISM and can be caused e.g. by optically thick emission. Under these conditions, we cannot directly translate  $n$  into the opacity power law index  $\beta$  using  $\beta = n - 2$ , but we first have to estimate the fraction of optically thick to optically thin emission  $\Delta$  from the best fit model. Correcting for that, we can write the opacity index  $\beta$  as  $(n - 2)(1 + \Delta)$  and typically obtain values of  $\beta \sim 1$ . The low value of  $n$  could also be associated with an altered disk opacity due to the collisional growth of dust grains. Fig. 10.3 shows significant decrease in the masses and submillimeter colors ( $n$ ) of disks along an empirical evolution sequence based on the infrared SED. These trends may be due to changes in disk structure, a change in the submillimeter emission efficiency due to particle growth, or both.

From fitting the full spectral energy distribution from IR to mm wavelength, we obtain an estimate of the temperature profile of disks. We can parametrize the profile as a power law  $T_{\text{disk}} \sim T_1 (r/1 \text{ AU})^{-q}$  and determine  $q$  from the best fit. This way, we find that  $T_1$  ranges between 50 and 400 K (peak around 200 K) and the power law exponent  $q$  between 0.4 and 0.75 (Fig. 10.4). From the theory outlined in Chapter 9, we know that  $T_{\text{disk}} \sim r^{-0.5}$  for a flared disk and

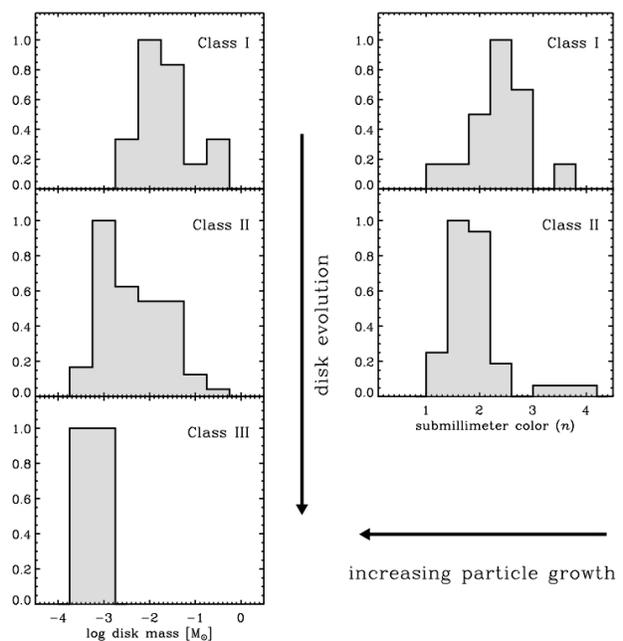


Figure 10.3: Normalized histograms of disk masses (left) and submillimeter colors ( $n$ ; right) from a multiwavelength photometry survey of Taurus-Auriga members (Andrews and Williams 2005). The different panels represent disk evolution classifications based on the infrared SED. There is empirical evidence for decreasing disk masses and increasing particle growth as disks evolve (Figure and caption from Andrews & Williams 2008).

$T_{\text{disk}} \sim r^{-0.75}$  for a flat disk.

### 10.1.3 IR surveys

Near-IR surveys ( $2.2 - 10 \mu\text{m}$ ) are on the other hand more sensitive to the hot dust in the inner disk, typically inside  $\sim 1$  AU for T Tauri stars. Skrutskie et al. (1990) performed a survey of 83 stars in the Taurus Auriga star forming region and found an inner disk frequency of  $\sim 50\%$ . Distinguishing between optically thick and thin disks at  $2.2 \mu\text{m}$  shows a clear trend with age. The fraction of optically thin versus thick disks increases strongly after a few Myr. Only a few disks are present around objects older than 10 Myr.

Later surveys including those with the Spitzer satellite focus on clusters and moving groups, because their ages are better determined. Hernandez et al. (2008) present their own survey of the  $\gamma$  Velorum cluster in the context of a compilation of several such surveys (e.g. NGC 2024 and Trapezium: Haisch et al. 2001; Chamaeleon I: Gomez & Kenyon 2001;  $\lambda$  Orionis: Barrado y Navascués et al. 2007; NGC 2362: Dahm & Hillenbrand 2007). Fig. 10.5 shows the disk frequency around stars of spectral type K5 or later in a wide range of clusters with ages between 0.3 and 11 Myr. The disappearance of inner disks occurs within a few Myr. This corresponds to the timescale within which CAI's and Chondrules must have formed within our own Solar System (most likely also in the very inner disk). From radioactive dating, we can infer that the asteroid Vesta for example formed within the first 5 Myr of the Solar System.

Evolutionary classification based on SEDs alone is difficult and often ambiguous. A small fraction of stars ( $< 10\%$ ) show excess emission at mid- and far-infrared wavelengths but lack near-infrared emission. The conclusion is that they contain cold dust in the outer region but the inner region ( $< \text{few AU}$ ) is cleared of small hot dust grains. Disks that show these characteristics are sometimes referred to as 'transitional' or 'cold' disks. However, there are a number of physical explanations (evolutionary scenarios) for this type of SED. Possible explanations are: (1) coagulation of small particles into larger particles, (2) inside-out clearing of the disk through photoevaporation, (3) dynamical clearing through the gravitational influence of planet-sized bodies, (4) a close binary object. In scenario 1, the coagulation of small grains into larger grains would change  $\kappa_\nu$  while keeping the mass density unchanged. The other three scenarios most likely decrease the mass density of the disk, at least in the inner regions. Scenario 3 would offer the possibility for an indirect proof of planet formation in these systems. Observational evidence from e.g. additional imaging and polarization observations shows that all 4 scenarios indeed occur.

## 10.2 Imaging

We can either observe the disks in scattered light or in their thermal emission. Since the scattered light originates at the surface of the disk, it is very sensitive to the disk structure, i.e. flat or flared disks. Only the latter show extended emission out to large radii because the entire surface is irradiated by the central

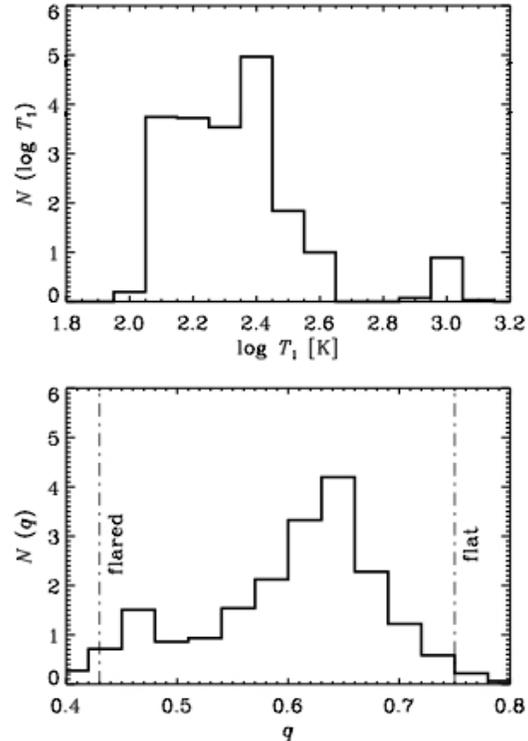


Figure 10.4: Distributions of the parameters describing the radial variation of disk temperatures for this sample:  $\log T_1$ , the temperature at 1 AU (top); and  $q$ , the radial power-law index (bottom). The bottom panel labels the idealized  $q$  values for flared ( $q \sim 0.43$ ) and flat ( $q \sim 0.75$ ) disks (Figure and caption from Andrews & Williams 2007).

star. Also, scattered light observations are often limited in their spatial resolution and require a coronagraph that blocks the direct stellar light. A phantastic instrument for this type of observations is HST with the Advanced Camera for Surveys (ACS) that is equipped with a coronagraph. Because it is located in space, it has a stable Point Spread Function (PSF) that enables correction of the diffraction spikes using reference stars of the same color as the target star.

Thermal images in the IR can be done from the ground in certain atmospheric windows,  $J$ ,  $H$  and  $K$  (1.20, 1.62, 2.20  $\mu\text{m}$ ) in the near-IR and  $L$ ,  $M$  and  $N$  band (3.50, 5.00, 9.00  $\mu\text{m}$ ) at slightly longer wavelength. In the mid-IR, we must rely on space-based observations that in the past lacked the spatial resolution necessary to resolve the disks (ISO as well as Spitzer had an 80 cm dish). Only Herschel with its 3.5 m mirror finally provides a few resolution elements per disk for the nearest star forming regions.

Imaging at long wavelength, e.g in the mm-regime provides a spatial resolution of  $\sim 11''$  (compared to  $1'$  with the infrared satellites IRAS and ISO and comparable to what the 3.5 m Herschel satellite now achieves in the far-IR and submm). The advantage is that these observations can be easily performed from the ground and that we can obtain a direct estimate of the disk mass (see above).

### 10.2.1 Imaging with (sub)mm interferometry

Single dish observations most of the time yield only a measurement of the total (sub)mm continuum flux. With The new millimeter interferometers such as the SMA, CARMA, Plateau de Bure, or the Nobeyama Radio Telescope, we can directly image disks at arcsec resolution and resolve the flux density distribution. This enables us to directly study the spatial disk structure.

From continuum sub-mm interferometry, we can infer the disk size and radial surface density gradient  $\Sigma \sim r^{-p}$ , i.e. the outer radius of the dust disk and  $p$ . This can be compared to predictions from accretion disks and viscous evolution models. Fig. 10.6 shows the size as a function of age (collected from the literature and thus inhomogeneous in nature). These disks seem to have a typical level of turbulent viscosity that is consistent with  $\alpha \approx 0.01$ . It must be noted however, that there is no obvious trend of disk radius evolution at ages lower than 10 Myr. While viscous disk evolution would predict a spreading and thus increase of outer disk radius with time, the estimated  $R_d$  values in Fig. 10.6 do not confirm such a simple scenario.

The values of outer disk radius  $R_d$  measured here technically represent some characteristic radius beyond which the temperature/density conditions in the disk do not produce substantial submillimeter emission compared to the noise levels in the data. True disk sizes could be larger, but would have to be measured with more optically thick tracers, such as molecular line emission or optical silhouettes (such as the Orion proplyds). Finally, as a caution it should be emphasized that elliptical Gaussians fit to interferometer data tend to underestimate disk sizes. The problem is really one of contrast, so that the level of discrepancy increases dramatically for more centrally concentrated surface brightness distributions (e.g. for higher values of  $p$ ).

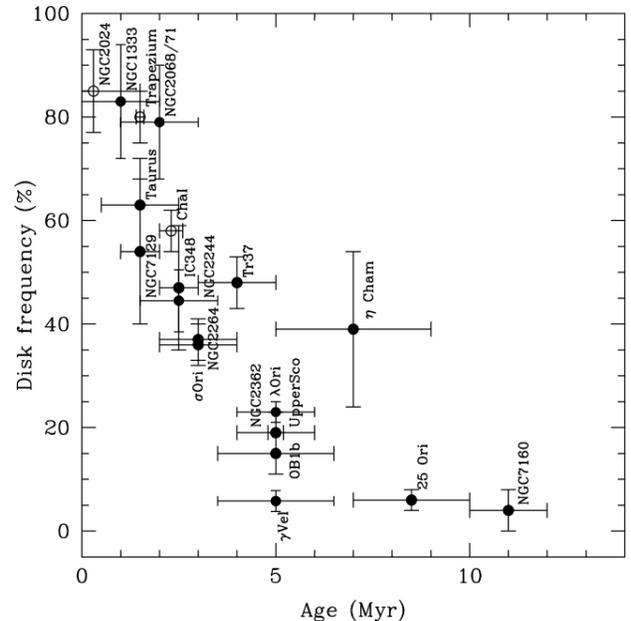


Figure 10.5: Fraction of stars with near-infrared disk emission as a function of the age of the stellar group. Open circles represent the disk frequency for stars in the T Tauri (TTS) mass range ( $\sim K5$  or later), derived using  $JHK_L$  observations: NGC 2024 and Trapezium (Haisch et al. 2001), and Chameleon I (Gómez & Kenyon 2001). Solid symbols represent the disk frequency calculated for stars in the TTS mass range using Spitzer data (Figure and caption from Hernandez et al. 2008).

## 10.2.2 Scattered light

Of the order of 30 disks have been imaged in scattered light at optical and near-IR wavelength. Since these images are spatially resolving the disks, they provide information on inclination, mass opacity, outer radius, radial dependence of scale height (edge-on), and spatial inhomogeneities (spiral arms). The light from the central object scatters on the dust grains in the disk surface (around  $\tau \sim 1$ ) and thus provides information on the dust material sitting there as well as on the shape of the surface. Multi wavelength observations are hence very powerful in deriving dust properties at the disk surface.

HST has been proven a powerful instrument for this type of research and scattered light images of young protoplanetary disks as well as low mass debris disks have been taken using the STIS coronagraph (UV/Optical) and the ACS coronagraph (optical/near-IR). Fig. 10.7 shows four Herbig disks observed with the STIS coronagraph onboard the Hubble Space Telescope, Especially in the case of AB Aur, we can see large scale non-axisymmetric structure (spiral arms). From surveying larger samples of Herbig disks, it seems that the STIS detections correlate with the Polycyclic Aromatic Hydrocarbon (PAH) emission features. This suggests a correlation with disk aring and an anticorrelation with the degree of dust settling toward the midplane.

A nice example of a VLT-UT1/ISAAC ground-based observations of an edge-on disk is shown in Fig. 10.2. It shows nicely how we can study the surface shape, i.e the flaring of the disk surface, in the case of edge-on objects.

## 10.3 Spectroscopy

Spectroscopy of disks provides information on the gas and dust in the disks. At high spectral resolution, we can obtain gas line profiles which contain a wealth of physical, chemical and kinematic information. At low spectral resolution, we either obtain integrated line fluxes, where we loose the kinematic and also the direct spatial information, or we study dust features that tell us about the dust sizes, shapes and mineralogy.

Interpreting gas lines is intrinsically more difficult than dust observations since we often have to consider a complex interplay between the chemical abundance structure in the disk, varying excitation

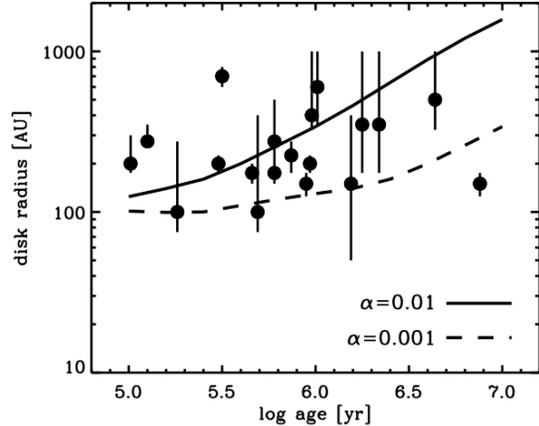


Figure 10.6: The disk radius  $R_d$  as a function of stellar age. Overlaid are the expected behaviors for a fiducial viscous accretion disk model with two values of the viscosity parameter,  $\alpha = 0.01$  (solid) and  $0.001$  (dashed). The constraints on disk structure from resolved submillimeter data can potentially help determine the typical level of turbulent viscosity in these disks, which sets the internal timescale for their evolution (Figure and caption from Andrews & Williams 2008).

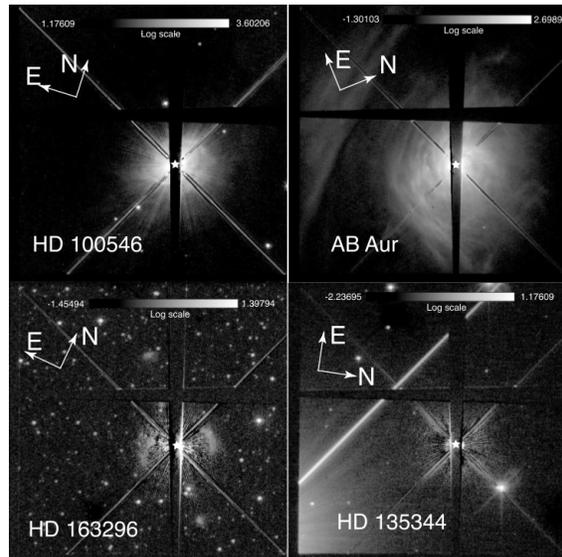


Figure 10.7: PSF-subtracted STIS imagery displayed with a logarithmic stretch. Background stars in the PSF template observations produce negative holes in the images. All stars are shown over a  $25'' \times 25''$  field of view. The approximate stellar location is indicated by a star symbol; the actual stellar location is at the crossing of the diffraction spikes (figure and caption from Grady et al. 2005).

Table 10.1: Chemical characteristics of YSO's (table from van Dishoeck 2007)

Component	Size <sup>a</sup> (")	Chemical characteristics	Submillimeter diagnostics	Infrared diagnostics
Pre-stellar cores	60	Low-T chemistry, Heavy freeze-out	Ions, deuterated, simple mol. (N <sub>2</sub> H <sup>+</sup> , H <sub>2</sub> D <sup>+</sup> )	Simple ices (H <sub>2</sub> O, CO <sub>2</sub> , CO)
Outer cold envelope	20	Low-T chemistry, Heavy freeze-out	Ions, deuterated, simple mol. (N <sub>2</sub> H <sup>+</sup> , HCO <sup>+</sup> , H <sub>2</sub> CO...)	Ices (H <sub>2</sub> O, CO <sub>2</sub> , CO, CH <sub>3</sub> OH?)
Inner warm envelope	1–2	Evaporation, X-rays (CO <sup>+</sup> , OH <sup>+</sup> , ...)	High T <sub>ex</sub> (H <sub>2</sub> CO, CH <sub>3</sub> OH, CO, CO <sub>2</sub> )	High gas/solid, High T <sub>ex</sub> , Heated ices
Hot core	≤ 1	Evaporation, High-T chemistry	H <sub>2</sub> O, Complex organics (CH <sub>3</sub> OCH <sub>3</sub> , CH <sub>3</sub> CN)	Hot gas (HCN, NH <sub>3</sub> , HNCO)
Outflow: direct impact	<1–20	Shock chemistry, Sputtering cores	H <sub>2</sub> O, OH, Si- and S-species (SiO, SO <sub>2</sub> )	H <sub>2</sub> , Atomic lines ([O I], [Si II], [S I])
Outflow: gentle impact	<1–20	Sputtering ices	Ice products (CH <sub>3</sub> OH, H <sub>2</sub> O, ..)	H <sub>2</sub>
Outer disk: surface	~2	UV irradiation, Ion-molecule	Ions, radicals (CN/HCN, HCO <sup>+</sup> )	PAHs, heated ices (NH <sub>4</sub> <sup>+</sup> )
Outer disk: midplane	~2	Heavy freeze-out,	Deuterated mol. (DCO <sup>+</sup> , H <sub>2</sub> D <sup>+</sup> )	...
Inner disk	≤0.02	Evaporation, High-T chemistry, X-rays	...	Silicates, PAHs, [Ne II], CO, H <sub>2</sub> , hot gas (H <sub>2</sub> O, C <sub>2</sub> H <sub>2</sub> , HCN)

<sup>a</sup> Typical angular size for a low-mass YSO at 150 pc, or a high-mass YSO at 3 kpc.

conditions (gas temperatures, densities, local continuum radiation field) and radiative transfer. However, the reward for undertaking the additional effort is the kinematic information that can be only gained from spectrally resolved line profiles. Dust continuum observations only provide indirect evidence of what happens to the major disk component, the gas. For this, we often make very simple assumptions such as a 'global' gas-to-dust mass ratio that does not change as a function of distance and height in the disk. Given the possibility of more complex disk evolutionary scenarios outlined in the previous chapters, it is time to abandon these assumptions and to turn to gas observations themselves to learn about the gas physics and evolution in the disk.

In the following we illustrate in a few examples what we learn from spectroscopy. We first summarize what can be learned from dust features at low spectral resolution. Then we turn to gas line observations at various wavelength, ranging from the sub-mm that probes the outer cold part of the disk to the near-IR that probes the terrestrial planet forming zone. Simple (symmetric) molecules in the mid-infrared (5 – 28  $\mu\text{m}$ ) probe the warm terrestrial planet forming zone out to  $\sim 10$  AU in disks around T Tauri stars. The far-IR region up to 350  $\mu\text{m}$  contains many lines of water, OH, hot CO and the atomic fine structure lines of O I and C II. Depending on the excitation temperature of these lines, they probe a range of gas temperatures in the few 100 K range. The sub-mm (0.3-7 mm) contains many lines of cold CO gas (low rotational lines) and a forest of lines from simple and more complex molecules (including O<sub>2</sub>). These lines probe the outer disk beyond 50 AU. For the diagnostic range of the various lines, see also Fig. ???. Table 10.1 summarizes the diagnostic value of various infrared and submm lines for the chemistry in various phases of star formation.

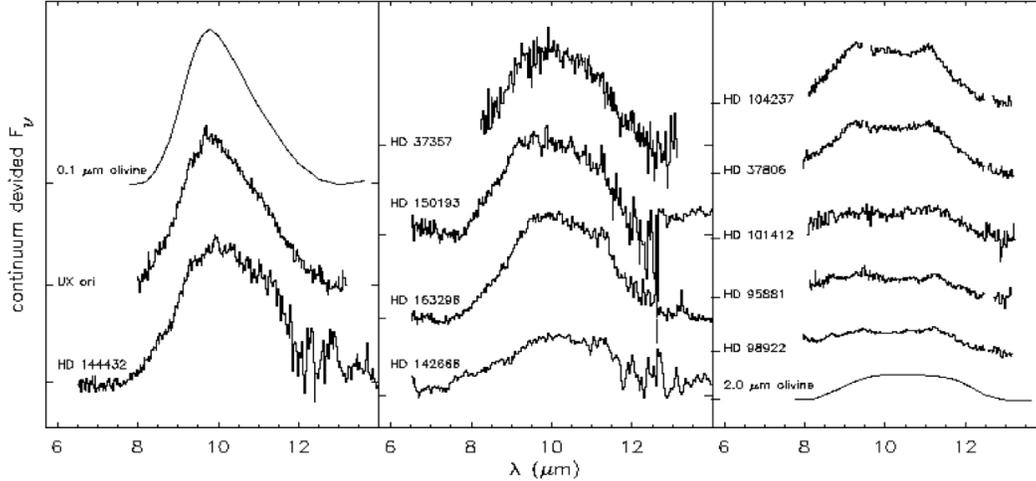


Figure 10.8: The  $10\ \mu\text{m}$  silicate feature in a number of disks around Herbig Ae/Be stars. For comparison, the laboratory shapes for a  $0.1\ \mu\text{m}$  and  $2.0\ \mu\text{m}$  olivine grain are shown as well. The features are ordered according to their shape in a sequence of growing grain size.

### 10.3.1 Measuring the grain size

The shape of the  $10\ \mu\text{m}$  silicate feature depends strongly on grain size and can thus be used to proof grain growth in the disks around young stars.

### 10.3.2 Line observations with (sub-) mm interferometers

In addition to imaging, spectral line observations with interferometers provide disk kinematics at high spatial resolution, and allow to distinguish the disk component from an infalling envelope. Beckwith & Sargent (1993) provide in their seminal paper the basic explanation of the double-peaked line profiles observed from disks around young stars. We illustrate here only the special case of an edge-on disk in Keplerian rotation. This is the most simple case and the velocity pattern seen from above looks like a dipole pattern. These are the line-of-sight velocities and the emitted line photons will be shifted accordingly to the red and blue, leading to a broad line profile centered around the stellar velocity (or zero in the star's rest frame). Fig. 10.9 relates this velocity dipole pattern to the observed double peaked line profile. The high velocities, i.e. the center of the line profile is dominated by the inner disk which contains less surface area. The lowest velocities correspond to the outer disk and the dip around the central velocity is caused by the fact that the disk has a finite outer radius, i.e. the lowest velocity bins do not show closed loops and hence also lack emitting surface area.

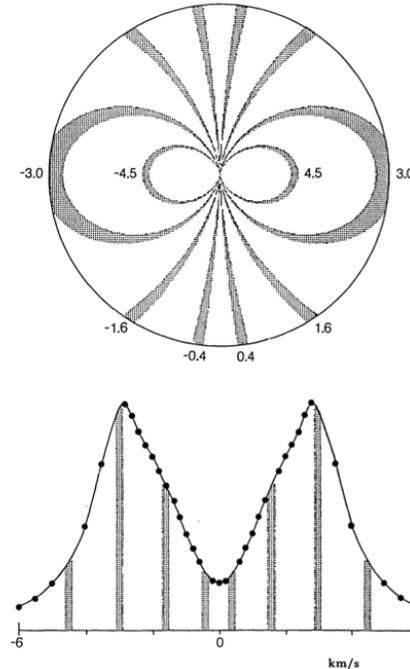


Figure 10.9: Top: Dipole velocity pattern of a Keplerian edge-on disk as seen by an observer in the plane of the page; a number of velocity intervals are indicated as shaded areas with the mean velocity annotated in km/s. Bottom: Resulting double-peaked line profile. The shaded areas correspond to the velocity bins in the top panel (Figure from Beckwith & Sargent 1993).

For optically thick lines and disks with sharp outer edges, the peak separation is a measure of the outer disk radius. The peaks occur at the projected Keplerian velocities  $\pm v_{\text{out}} \sin \theta$ , with  $v_{\text{out}}$  the Keplerian velocity at the outer disk radius and  $\theta$  the disk inclination with respect to the observer ( $\theta = 0$  is face-on).

Sub-mm line observations can be used to measure the rotation of the disks and to confirm their Keplerian rotation. With high enough accuracy, the velocity pattern can be even used to measure the mass of the central star, a quantity that is otherwise only directly measurable for binary stars. Most molecules other than CO have only been detected in spatially unresolved single-dish submillimeter spectra from which disk-averaged abundances can be derived (e.g. Dutrey et al. 1997, Thi et al. 2004). Chemical images of disks have so far been limited to just a few pixels across a handful of disks in a few lines (e.g. Qi et al. 2003, Dutrey et al. 2007).

The data obtained so far support the layered chemical structure predicted in chemical disk models. We can for example use a series of CO lines from the main isotope as well as  $^{13}\text{CO}$  to probe the vertical disk structure. These lines become optically thick at very different vertical column densities and can thus establish for example the presence of vertical temperature gradients (Fig. 10.10). In addition, we can use detailed two-dimensional disk models to fix parameters such as the disk scale height and outer radius. CO observations are especially sensitive to the outer disk radius. Dent et al. (2005) used this to show that younger disks (3-7 Myr) are on average larger than older disks ( $> 7$  Myr). However, statistics are generally small due to the limited sensitivity of current facilities. ALMA will clearly revolutionize this field in the near future.

### 10.3.3 IR line observations

Observation with the Spitzer Space Telescope have revealed highly abundant and hot (300 – 700 K) HCN and  $\text{C}_2\text{H}_2$  (acetylene) in the inner few AU of two young near edge-on disks (Lahuis et al. 2006, Gibb et al. 2007). More recently, Spitzer and ground-based Keck and VLT data have revealed surprisingly strong mid-IR emission lines of hot  $\text{H}_2\text{O}$  ( $\sim 800$  K), together with lines of OH, HCN,  $\text{C}_2\text{H}_2$  and/or  $\text{CO}_2$ , toward a number of disks. This emission must be originating from the inner AU of the disk (Carr & Najita 2008; Salyk et al. 2008).

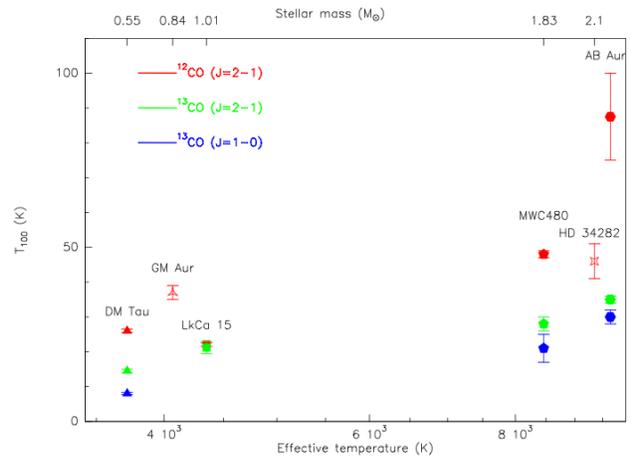


Figure 10.10: Temperatures derived from the CO isotopes versus effective temperature of the central star. From left to right, the sources are DM Tau (lled triangles), GM Aur (empty stars), LkCa 15 (lled squares), MWC 480 (lled pentagons), HD 34282 (empty stars) and AB Aur (lled hexagons). From top to bottom,  $^{12}\text{CO}$  J=2-1,  $^{13}\text{CO}$  J=2-1,  $^{13}\text{CO}$  J=1-0 temperatures (figure and caption from Pietu et al. 2007).

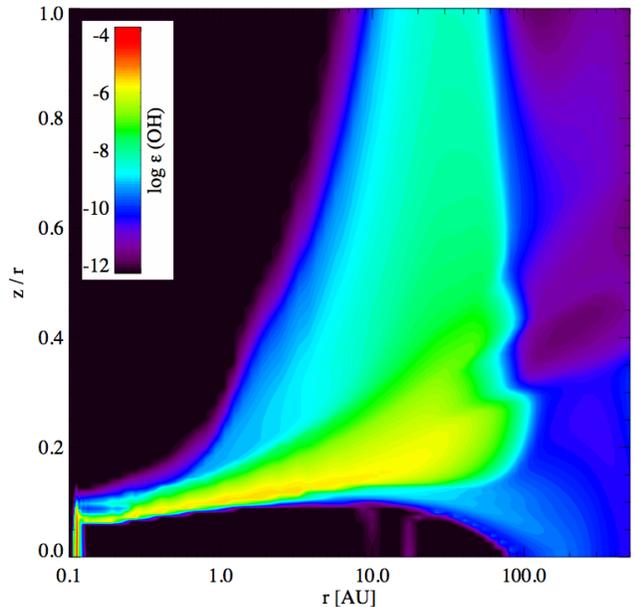


Figure 10.11: Abundance distribution of OH in the disk around a Herbig star. The vertical axis shows  $z/r$ , the ratio between height in the disk and radial distance.

### 10.3.4 Optical line observations

The forbidden [O I] 6300 Å line is observed towards a number of Herbig Ae/Be stars. In some cases, the line shows the double-peaked profile that is characteristic for a disk origin. The line width shows that the emission originates typically within 10-40 AU. Two possible excitation mechanisms are discussed: (1) thermal emission, maybe in non-LTE or (2) OH photodissociation. The current disk models have problems explaining the level of emission that is observed. In the scenario, where OH photodissociation leads to the preferential population of the  $^1D_2$  level, unusually high OH abundances are required to explain the total [O I] 6300 Å emission. The spatial distribution of the OH in the chemical models however seems to agree with the observed one.

### 10.3.5 Measuring disk gas mass

If we can find a species that is the dominant carrier of atomic oxygen or carbon throughout the entire disk and if this line proves to be optically thin and thus probe the total vertical column density of gas through the disk, we have some hope that we can measure the disk gas mass. If the species happens to be only the dominant carrier in a particular radial or vertical disk regime, we need to piece together the entire disk mass adding up all other potential carriers in the other disk regimes as well.

A potential species to trace the total disk gas mass could be  $^{13}\text{CO}$  or  $\text{C}^{18}\text{O}$ . Being optically thin down to the midplane (Fig. 10.12), the only problem are: (1) the disk surfaces, in which CO is efficiently photodissociated and the dominant carbon carrier would be either neutral or ionized carbon; (2) the cold outer disk regions, where CO might efficiently freeze-out onto cold dust grains and be in the form of ice. Panic et al. (2008) have used this technique to estimate the gas mass in the disk around a Herbig star. Since dust temperatures are above the freeze-out temperature for CO out to large radii ( $\sim 200$  AU), and the lines used were proven to be optically thin, this would be the most optimal case.

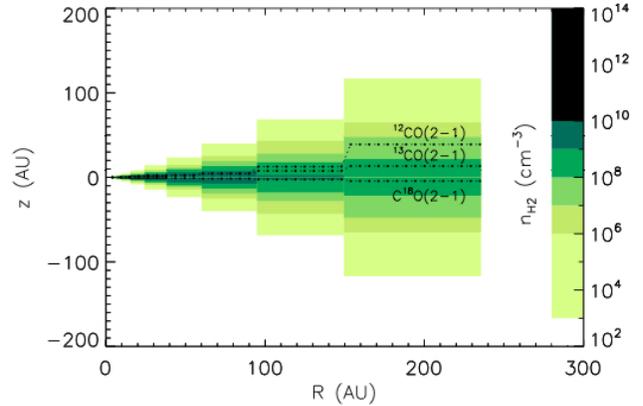


Figure 10.12: Density structure of the best-fit disk model shown as colour scale. The disk is viewed by us from almost straight above. The three dotted lines show the heights above which 90% of the integrated line emission of  $^{12}\text{CO}$ ,  $^{13}\text{CO}$  and  $\text{C}^{18}\text{O}$  J=2-1 originates. For the  $\text{C}^{18}\text{O}$  emission this line lies below the disk midplane (figure and caption from Panic et al. 2008).

# Appendix A

## Fluid dynamics and the equation of state

### A.1 Simple equations of state

The behavior of a gas under different physical conditions is described by the equation of state of the gas. The simplest assumption we can adopt is that a molecular cloud behaves as an ideal gas, obeying the equation

$$p = nkT = \frac{\rho kT}{\mu m_{\text{H}}} = \frac{\rho \mathcal{R}T}{\mu}, \quad (\text{A.1})$$

where  $p$  is the gas pressure,  $n$  is the particle density,  $\rho$  is the density,  $T$  is the temperature, and  $\mu$  is the mean molecular weight of the gas in units of  $m_{\text{H}}$ , which would be two if the cloud were entirely composed of molecular hydrogen. It holds that  $\rho = \mu m_{\text{H}} n$ . The constants have their usual meaning;  $\mathcal{R} = k/m_{\text{H}}$  is the gas constant. It is sometimes convenient when treating astrophysical problems to assume that the gas temperature and chemical composition remain constant. This is known as the isothermal assumption, under which equation A.1 becomes

$$\frac{p}{\rho} = \frac{kT}{\mu m_{\text{H}}} = \frac{\mathcal{R}T}{\mu} = a_{\circ}^2 = \text{constant}, \quad (\text{A.2})$$

where  $a_{\circ}$  is the isothermal sound speed in the gas at temperature  $T$ .

The next simplest assumption we can adopt is the adiabatic equation of state

$$p \rho^{-\gamma} = K, \quad (\text{A.3})$$

where  $K$  is constant and  $\gamma$  is the ratio of specific heats at constant pressure,  $c_p$ , and volume,  $c_v$ , respectively:  $\gamma = c_p/c_v$ . An adiabatic process is any process occurring without gain or loss of heat within a system, i.e. during the process the system is thermodynamically isolated.

The adiabatic index is fairly straightforward for a gas that is pure in the sense that it consists (or is dominated by) one type of particle. In that case it can be related to the degrees of freedom  $f$  of the molecule or atom/ion,

$$\gamma = 1 + \frac{2}{f}. \quad (\text{A.4})$$

For a monoatomic gas, for instance, with three (translational) degrees of freedom,  $\gamma = 5/3$ . For a diatomic gas, two rotational degrees of freedom may be added, yielding  $\gamma = 7/5$ . At high temperatures a vibrational degree of freedom may be added. For gasses that consist of different component, for instance a gas that is partly molecular and partly atomic and or ionic, it is more complex to determine the value of  $\gamma$ .

## A.2 Continuity equation

The material in an astrophysical medium may not be static or stationary, and for a general description we must take this into account. To do so, we use the basic equations of aero/fluid mechanics. In general, there are two methods commonly used to model gas/fluid flow. One method is to use a fixed set of coordinates in space and calculate the parameters of the gas as it flows through the coordinate frame. This is known as the Eulerian method. An alternative is to choose a set of coordinates fixed to a particle of the gas, moving with that particle, and to calculate the varying parameters in that coordinate frame (referred to as co-moving coordinates). This is known as the Lagrangian method.

Assume an arbitrary volume  $V$  contained by a closed surface  $S$ .  $V$  and  $S$  are fixed in space, so we are here adopting the Eulerian viewpoint. The mass flowing out of  $V$  per unit time through the element of area  $d\mathbf{S}$  of the surface is given by

$$\rho \mathbf{v} \cdot d\mathbf{S}, \quad (\text{A.5})$$

where  $\mathbf{v}$  is the macroscopic (i.e. systemic) velocity in  $\text{cm sec}^{-1}$  and  $\rho$  the total mass density in  $\text{gr cm}^{-3}$ , and so the net rate at which mass flows out of  $V$  through  $S$  is given by

$$\oint_S \rho \mathbf{v} \cdot d\mathbf{S} = \int_V \nabla \cdot (\rho \mathbf{v}) dV, \quad (\text{A.6})$$

where we have obtained the right-hand side by invoking Gauss's divergence theorem. The rate at which the mass in  $V$  decreases is given by

$$-\frac{\partial}{\partial t} \left( \int_V \rho dV \right) = \int_V \left( -\frac{\partial \rho}{\partial t} \right) dV, \quad (\text{A.7})$$

where we can take  $\partial/\partial t$  inside the integral because  $V$  is fixed in space. Obviously, the rate at which the mass in  $V$  decreases, i.e. eq. (A.7), must equal the rate at which mass flows out of  $V$ , across  $S$ , so

$$\int_V \left[ \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) \right] dV = 0, \quad (\text{A.8})$$

and since the volume  $V$  is arbitrary, it follows that

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \quad (\text{A.9})$$

everywhere. This is known as the *continuity equation*. We can expand this equation and rewrite it as

$$\frac{\partial \rho}{\partial t} + \rho \nabla \cdot \mathbf{v} + \mathbf{v} \cdot \nabla \rho = \frac{\partial \rho}{\partial t} + \rho \frac{\partial v^i}{\partial x^i} + v^i \frac{\partial \rho}{\partial x^i} = 0 \quad (\text{A.10})$$

where the double occurrence of index  $i$  in both terms in the second right-hand-side implies a summation over the components of the coordinate system.

We stress that  $\partial\rho/\partial t$  is the Eulerian time derivative of the density, i.e. the rate of change of density at a fixed point in space. If we want the Lagrangian time derivative of the density, i.e. the rate of change of density moving with the fluid, we must include the contribution due to the displacement,  $d\mathbf{r} = \mathbf{v}dt$ , which occurs during the time interval  $dt$ . The net density change is

$$d\rho = \frac{\partial \rho}{\partial t} dt + d\mathbf{r} \cdot \nabla \rho, \quad (\text{A.11})$$

and hence the Lagrangian time derivative of the density is

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \mathbf{v} \cdot \nabla \rho = -\rho (\nabla \cdot \mathbf{v}), \quad (\text{A.12})$$

where the final expression is obtained by substituting from equation (A.10). The Lagrangian time derivative is sometimes called the co-moving time derivative or fluid-frame derivative of the density.

## Mass-loss rate and mass-accretion rate

For a stationary one-dimensional radial spherical flow the mass continuity equation reduces to

$$\frac{1}{r^2} \frac{\partial(r^2 \rho v)}{\partial r} = 0 \quad (\text{A.13})$$

from which we find

$$4\pi r^2 \rho(r) v(r) = \text{constant} \equiv \dot{M} \quad (\text{A.14})$$

where  $\dot{M}$  is the *mass-loss rate* through a spherical surface of radius  $r$ , and  $v$  is the velocity in the radial direction. If the velocity  $v$  is negative, i.e. directed towards the center of the sphere,  $\dot{M}$  is to be interpreted as the *mass-accretion rate*. The cgs unit of mass loss is  $\text{gr sec}^{-1}$ , although it is custom to use the unit  $M_\odot$  per year:  $1 M_\odot \text{yr}^{-1} = 6.303 \times 10^{25} \text{gr sec}^{-1}$ .

### A.3 The equation of motion under pressure

Consider again the volume of gas  $V$ . If the gas in this volume has pressure  $p(\mathbf{r}, t)$ , then the total force acting on the volume is the sum of the external pressure on the surface. This is given by the surface intergral

$$- \oint_S p dS. \quad (\text{A.15})$$

Transforming this into a volume integral using Gauss's divergence theorem, the net pressure force exerted on the arbitrary volume  $V$  is

$$- \oint_S p dS = - \int_V \nabla p dV, \quad (\text{A.16})$$

and hence the net pressure force per unit volume is simply  $-\nabla p$ .

The equation of motion of this volume can be derived by equating the force per unit volume with the mass per unit volume multiplied by its acceleration. This is simply Newton's third law. The mass per unit volume is defined as the density  $\rho$  and the acceleration is the time derivative of the velocity, i.e.  $d\mathbf{v}/dt$ . So we have

$$-\nabla p = \rho \frac{d\mathbf{v}}{dt}, \quad (\text{A.17})$$

and hence

$$\frac{d\mathbf{v}}{dt} = -\frac{\nabla p}{\rho}. \quad (\text{A.18})$$

Here  $d\mathbf{v}/dt$  is the co-moving acceleration of the gas, so equation A.18 is the Lagrangian formulation of the equation of motion. The Eulerian formulation is obtained by substituting for  $d\mathbf{v}/dt$  using

$$\frac{d\mathbf{v}}{dt} = \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v}. \quad (\text{A.19})$$

### A.4 Fluid motion in a gravitational potential field

If the gas is in a gravitational potential field  $\Phi$ , where any gas element suffers an acceleration  $\mathbf{g} = -\nabla\Phi$  (for instance towards a mass  $M$ ), an extra term must be included in (A.18) to account for this. The force  $\mathbf{f}_g$  on unit volume due to a gravitational acceleration  $\mathbf{g}$  is simply  $\rho \mathbf{g}$ , and so (A.18) becomes

$$\frac{d\mathbf{v}}{dt} = \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\frac{\nabla p}{\rho} + \mathbf{g} \quad (\text{A.20})$$

This is sometimes referred to as Euler's equation. Note the notation  $(\mathbf{v} \cdot \nabla) \mathbf{v}$ . This has by definition the components  $v_i = \mathbf{v} \cdot (\nabla v_i)$ , for  $i = 1, 2, 3$ , where  $\mathbf{v} = (v_1, v_2, v_3)$ . In spherical symmetry, with  $\mathbf{v} = (v, 0, 0)$ , this reduces to  $v \partial v / \partial r$ .

## A.5 Fluid motion in a magnetic field

There is ample observational evidence that molecular clouds are permeated by magnetic fields. Electric currents in the fluid can give rise to a force acting on the gas, which then needs to be taken account in the equation of motion. We begin with two of Maxwell's equations. Firstly we use Ampère's law

$$\nabla \times \mathbf{B} = \frac{4\pi}{c} \mathbf{J} - \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t}, \quad (\text{A.21})$$

where  $\mathbf{B}$  is the magnetic flux density,  $\mathbf{J}$  is the current density,  $\mathbf{E}$  is the electric field strength, and  $c$  is the speed of light in a vacuum. Secondly we use Ohm's law, which for a fluid of conductivity  $\sigma$  becomes

$$\mathbf{J} = \sigma \left[ \mathbf{E} + \frac{1}{c} (\mathbf{v} \times \mathbf{B}) \right]. \quad (\text{A.22})$$

In this equation the first term on the right-hand side describes the conductivity of a static fluid and the second term incorporates the motion of that fluid at velocity  $\mathbf{v}$ . In the limit of infinite conductivity we have

$$\mathbf{E} + \frac{1}{c} (\mathbf{v} \times \mathbf{B}) = 0. \quad (\text{A.23})$$

In a conducting fluid there are electric currents which can flow and give a force per unit volume  $\mathbf{f}_m$  given by

$$\mathbf{f}_m = \frac{1}{c} (\mathbf{J} \times \mathbf{B}). \quad (\text{A.24})$$

We can eliminate  $\mathbf{J}$  from this force equation by using Ampère's law, and remembering that in the limit of high conductivity, such as the ISM, the only electric fields present are those that are induced by motions in the gas with velocity  $v \ll c$ , and hence the second term on the right in Ampère's law can be ignored, simplifying it to the magnetostatic case

$$\nabla \times \mathbf{B} = \frac{4\pi}{c} \mathbf{J}. \quad (\text{A.25})$$

Substituting into the force equation we have

$$\mathbf{f}_m = \frac{1}{4\pi} (\nabla \times \mathbf{B}) \times \mathbf{B} = -\frac{1}{4\pi} \mathbf{B} \times (\nabla \times \mathbf{B}). \quad (\text{A.26})$$

We can simplify this equation using the vector identity

$$\frac{1}{2} \nabla (\mathbf{B} \cdot \mathbf{B}) = (\mathbf{B} \cdot \nabla) \mathbf{B} + \mathbf{B} \times (\nabla \times \mathbf{B}), \quad (\text{A.27})$$

and thus

$$\mathbf{B} \times (\nabla \times \mathbf{B}) = \frac{1}{2} \nabla (\mathbf{B} \cdot \mathbf{B}) - (\mathbf{B} \cdot \nabla) \mathbf{B}. \quad (\text{A.28})$$

Then the force equation becomes

$$\mathbf{f}_m = -\nabla \left( \frac{B^2}{8\pi} \right) + \frac{1}{4\pi} (\mathbf{B} \cdot \nabla) \mathbf{B}. \quad (\text{A.29})$$

We can now write a new version of the equation of motion, also accounting for magnetic forces (in which, to preserve generality, we employ the gravitational potential  $\Phi$ )

$$\rho \frac{d\mathbf{v}}{dt} = -\nabla p - \rho \nabla \Phi - \nabla \left( \frac{B^2}{8\pi} \right) + \frac{1}{4\pi} (\mathbf{B} \cdot \nabla) \mathbf{B}. \quad (\text{A.30})$$

We can obtain an intuitive physical feel for the meaning of the magnetic terms in equation A.30. Notice that the first magnetic term on the right-hand side enters in an identical way to the gas pressure. Hence this term is known as the *magnetic pressure*. Any gradient in the magnetic pressure results in a net force on

the fluid, just as in the gas of the gas pressure. Hence a region of the ISM with a high  $\mathbf{B}$  value tends to be over-pressured relative to neighboring regions with lower  $\mathbf{B}$ , and will tend to expand.

The second magnetic term on the right-hand side of equation A.30 can be understood as follows. We can write  $\mathbf{B} = Bs$ , where  $\mathbf{s}$  is a unit vector in the direction of the field. Then we have

$$(\mathbf{B} \cdot \nabla) \mathbf{B} = Bs \frac{d(Bs)}{dx} = B^2 \frac{ds}{dx} + Bs^2 \frac{dB}{dx}, \quad (\text{A.31})$$

where  $x$  is the direction along the field. For a constant magnetic field,  $dB/dx$  is zero and the second term vanishes. Furthermore, if the field lines are straight, then  $ds/dx$  is zero and the first term vanishes. Hence this term clearly relates to how ‘bent’ the magnetic field lines are. It is found that the more bent the field lines, the stronger the restoring force. The term is therefore sometimes referred to as the *magnetic tension*.

# Appendix B

## Molecules and dust

### B.1 Line emission from molecules

The first molecules were detected in space in the late 1930's through their absorption lines in the spectra of background stars. Among the first were CH, CH<sup>+</sup> and CN. As laboratory experiments provided more accurate line frequencies, the field exploded leading to what we call now Astrochemistry. The chemical viewpoint of how these molecules form in the first place will be covered in Sect. ?? . Here we focus first on understanding the spectra of molecules.

Figure ?? provides an overview table of some of the most relevant molecules in space. It shows the wavelengths at which these molecules are observed, the type of transition and also the temperatures required for exciting those transitions. The last column indicates which type of information those transitions carry. Column 7 gives the transition probability  $A_{ul}$  (The Einstein A coefficient for spontaneous emission) for the line and column 8 lists the critical density. The latter is defined as

$$n_{\text{crit}} = \frac{A_{ul}}{\gamma_{ul}} , \quad (\text{B.1})$$

which is the ratio between downward radiative rates ( $A_{ul}$ ) and collisionally induced downward transitions ( $\gamma_{ul}$ , where  $u$  denotes the upper level and  $l$  the lower level). This critical density indicates the minimum density required to achieve LTE level populations, i.e. energy levels populated according to the Boltzmann equation at the local ambient temperature. Hence, at densities  $n > n_{\text{crit}}$ , the de-population of the upper level occurs preferentially through collisions. In a two-level system without radiative excitation, the equation of statistical equilibrium would read as

$$n_i n_x \gamma_{lu} = n_x n_u \gamma_{ul} + n_u A_{ul} . \quad (\text{B.2})$$

Here,  $n_i$  is the density of the molecule we consider and  $n_x$  is the density of the collision partner (e.g. electrons, neutral or molecular hydrogen). The two terms on the right hand side are due to collisional de-excitation and radiative de-excitation. The critical density is then defined as the one where these two are equal.

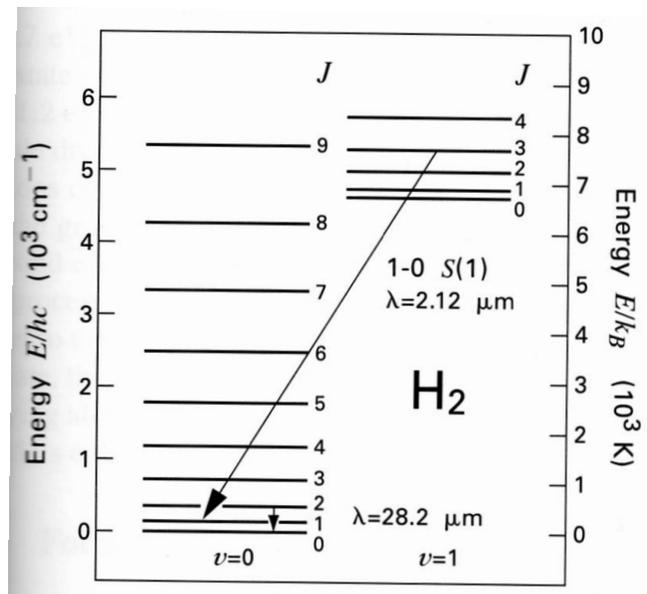


Figure B.2: Rotational levels of H<sub>2</sub> for the first two vibrational states. Within the  $v = 0$  state, the  $J = 2 - 0$  transition at  $28.2 \mu\text{m}$  is displayed. Also shown is the transition giving the  $1 - 0$  S(1) ro-vibrational line at  $1.12 \mu\text{m}$ . Note that two different energy scales are used (caption and figure from Stahler & Palla 2004).

Figure B.1: Some astrophysically relevant molecules (taken from Stahler & Palla 2004).

molecule	abundance <sup>a</sup>	transition	type	$\lambda$	$T_o^b$ (K)	$A_{ul}$ (s <sup>-1</sup> )	$n_{crit}^c$ (cm <sup>-3</sup> )	comments
H <sub>2</sub>	1	1→0 S(1)	vibrational	2.1 $\mu$ m	6600	$8.5 \times 10^{-7}$	$7.8 \times 10^7$	shock tracer
CO	$8 \times 10^{-5}$	J= 1 → 0	rotational	2.6 mm	5.5	$7.5 \times 10^{-8}$	$3.0 \times 10^3$	low density probe
OH	$3 \times 10^{-7}$	<sup>2</sup> $\Pi_{3/2}; J=3/2$	$\Lambda$ -doubling	18 cm	0.08	$7.2 \times 10^{-11}$	$1.4 \times 10^0$	magnetic field probe
NH <sub>3</sub>	$2 \times 10^{-8}$	(J,K)=(1,1)	inversion	1.3 cm	1.1	$1.7 \times 10^{-7}$	$1.9 \times 10^4$	temperature probe
H <sub>2</sub> CO	$2 \times 10^{-8}$	2 <sub>12</sub> →1 <sub>11</sub>	rotational	2.1 mm	6.9	$5.3 \times 10^{-5}$	$1.3 \times 10^6$	high density probe
CS	$1 \times 10^{-8}$	J= 2 →1	rotational	3.1 mm	4.6	$1.7 \times 10^{-5}$	$4.2 \times 10^5$	high density probe
HCO <sup>+</sup>	$8 \times 10^{-9}$	J= 1 → 0	rotational	3.4 mm	4.3	$5.5 \times 10^{-5}$	$1.5 \times 10^5$	tracer of ionization
H <sub>2</sub> O		6 <sub>16</sub> →5 <sub>23</sub>	rotational	1.3 cm	1.1	$1.9 \times 10^{-9}$	$1.4 \times 10^3$	maser
//	$<7 \times 10^{-8}$	1 <sub>10</sub> →1 <sub>11</sub>	rotational	527 $\mu$ m	27.3	$3.5 \times 10^{-3}$	$1.7 \times 10^7$	warm gas probe

<sup>a</sup> number density of main isotope relative to hydrogen, as measured in the dense core TMC-1

<sup>b</sup> equivalent temperature of the transition energy;  $T_o \equiv \Delta E_{ul}/k_B$

<sup>c</sup> evaluated at T=10 K, except for H<sub>2</sub> (T=2000 K) and H<sub>2</sub>O at 527  $\mu$ m (T=20 K)

Homonuclear molecules, i.e. molecules that consist of two or more equal atoms (e.g. H<sub>2</sub>), do not possess a permanent dipole moment. The dipole moment is generated through the slight difference in electronegativity between the constituents of a molecule. This slight difference in electronegativity allows one of the atoms to pull the 'cloud of electrons' closer than the other, hence inducing a slight shift in charge.

A molecule has various degrees of freedom. It can rotate, vibrate or be electronically excited. Its total energy is thus

$$E_{tot} = E_{rot} + E_{vib} + E_{elect} \quad . \quad (B.3)$$

Generally, the rotational levels have the lowest energies, followed by the vibrational energy levels and then by the electronic states. However, at some point the highly rotational excited levels of a particular vibrational state can overlap with the lower rotational levels of a higher vibration state. Hence, the final term schema can be rather complex (see Fig. ??).

The energy distance between the electronic states is typically of the order of  $10^5$  K. In the case of H<sub>2</sub>, the ground electronic state contains 14 vibrational states (Fig. ??).

## B.2 Rotational lines

For simplicity, we consider here first diatomic molecules that we picture as rigid rotators. If we think of a diatomic molecule as a dumbbell oriented along the x-axis and fixed at its center of mass, it can rotate in two different planes, the (x,z)-plane and the (x,y)-plane. In classical mechanics, a dumbbell rotating about his axis through the center of mass has the possible energy states

$$E_{rot} = \frac{J^2}{2I} \quad (B.4)$$

where  $I$  is the moment of inertia and  $J$  is the angular momentum. In quantum mechanics (without

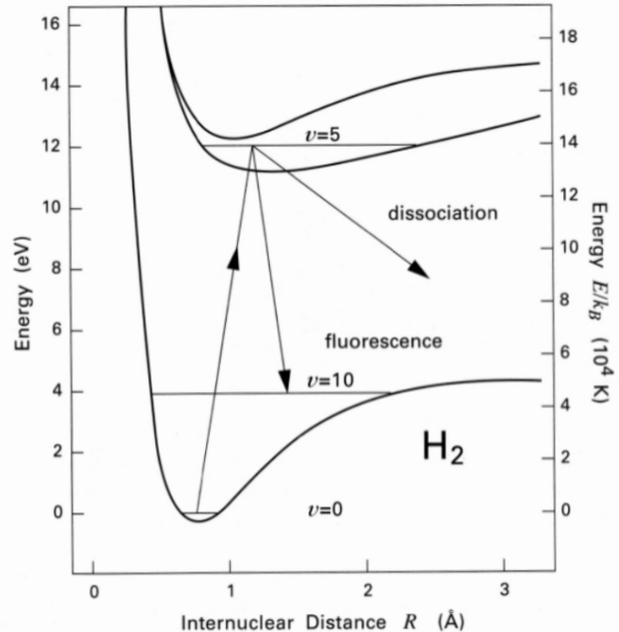


Figure B.3: Potential energy of H<sub>2</sub> as a function of internuclear separation. The three solid curves correspond to the ground and first two excited electronic states. The horizontal lines represent vibrational energy levels. Arrows depict photo-excitation into the Lyman band, followed by either fluorescence or dissociation (caption and figure from Stahler & Palla 2004).

Table B.1: Molecular constants: reduced mass  $\mu$ , rotational constant  $B$ , force constant  $k$ , dissociation energy  $E_{\text{diss}}$ .

Molecule	B [GHz]	$\mu$	B [ $\text{cm}^{-1}$ ]	internuclear distance [ $\text{\AA}$ ]	$k$ [N/m]	$E_{\text{diss}}$ [eV]
H <sub>2</sub>	1824.2	0.504	60.85	0.741	$5.2 \times 10^2$	4.478
<sup>12</sup> C <sup>16</sup> O	57.636	6.856	1.931	1.128		11.092
<sup>13</sup> C <sup>16</sup> O	55.101	7.172	1.846	1.128		
SiO	21.161	10.177	0.721	1.510		8.26

further derivation), this reads as

$$E_{\text{rot}} = \frac{\hbar^2}{2I} J(J+1) = BJ(J+1) \quad (\text{B.5})$$

where  $\hbar$  is the Planck constant  $h/2\pi$  and  $J$  is now a dimensionless quantum number, called the rotational quantum number. It can only take integer values of 0, 1, 2, 3, ... We merge  $h/(8\pi^2 I)$  into the rotational constant of a molecule  $B$ . It has the units of frequency or classically also  $\text{cm}^{-1}$ . Table ?? shows values of some representative molecules.  $I$  is the moment of inertia

$$I = \mu \langle r^2 \rangle, \quad (\text{B.6})$$

where  $r$  is the bond length and  $\mu$  is the reduced mass of the molecule, here expressed for a diatomic molecule,

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad (\text{B.7})$$

Let us first consider diatomic linear molecules such as CO. Their rotational spectra are easily understood. A dipole radiative transition occurs for  $\Delta J = 1$

$$\Delta E = Bh [J(J+1) - (J-1)J] = 2BhJ. \quad (\text{B.8})$$

The transitions  $J \rightarrow J+1$  are labelled as the R-branch, while the transitions  $J \rightarrow J-1$  are labelled the P-branch. The nomenclature here is R(0)= 0  $\rightarrow$  1, R(1)= 1  $\rightarrow$  2, etc. and the same holds for the P-branch starting with P(1)= 1  $\rightarrow$  0. From these formulae, it is clear that the energy spacing is linear with  $J$  and the frequency separation is  $\Delta\nu = 2B$ . The lowest rotational transitions of CO are

$$\begin{aligned} J = 1 - 0 & \quad \lambda = 2.6 \text{ mm} \quad \nu = 116 \text{ GhZ} \\ J = 2 - 1 & \quad \lambda = 1.3 \text{ mm} \quad \nu = 232 \text{ GhZ} \\ J = 3 - 2 & \quad \lambda = 0.87 \text{ mm} \quad \nu = 347 \text{ GhZ} \end{aligned} \quad (\text{B.9})$$

Once we go to higher rotational levels, the approximation of the rigid rotator becomes less good and there will occur deviations from centrifugal distortions.

The CO rotational lines are the brightest molecular rotational transitions from space. Thus they are frequently used in ground based observations of dense molecular environments. On a side note, if we measure the rotational spectrum very accurately, we can turn the exercise around and estimate the rotational constant  $B$  and hence the bond length from the spacing of the rotational lines.

### B.3 Vibrational lines

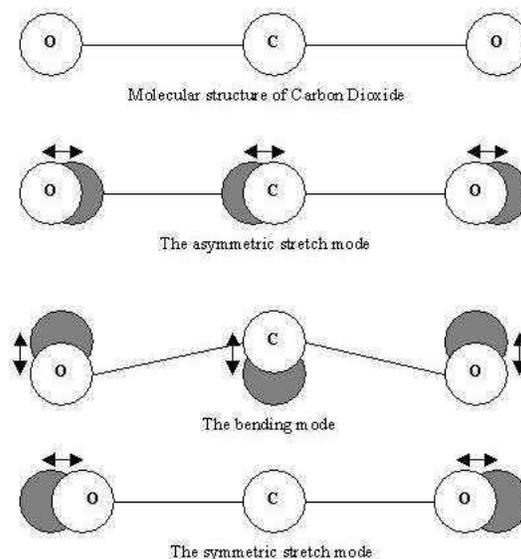


Figure B.4: Vibrational modes in the linear CO<sub>2</sub> molecule.

The vibrational levels of a molecule lie at higher energies than the rotational levels. Fig. ?? illustrates the possible vibrational modes in the linear CO<sub>2</sub> molecule. We can consider the bonds between the atoms in the molecules as springs between the masses that allow them to move in one of the three vibrational modes: the symmetric stretching mode, the asymmetric stretching mode, and the bending mode. In the symmetric stretching mode, the carbon atom remains fixed while the two oxygen atoms move closer to and farther from the carbon atom. In the bending mode the central carbon atom moves up and down while the two outer oxygen atoms move up and down in the opposite direction. Finally, in the asymmetric stretch mode, all three atoms move left to right; one bond contracts while the other expands.

If we solve the Schrödinger equation for a diatomic vibrating-spring model, we obtain the following expression for the spacing of the vibrational energy levels

$$E(v) = h\nu_e\left(v + \frac{1}{2}\right) - \nu_e x_e\left(v + \frac{1}{2}\right)^2, \quad (\text{B.10})$$

where  $v$  is the vibrational quantum number (values of 0, 1, 2, ...) and  $\nu_e$  is the fundamental frequency. Here, the first term corresponds to the harmonic oscillator approximation and the second term describes the anharmonic effect on the spring as it stretches ( $x_e$  being the anharmonic constant). The fundamental difference to the rotational energy levels is that the ground vibrational state has a non-zero energy, namely  $1/2h\nu_e$ . The fundamental frequency is given by

$$\nu_e = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}}, \quad (\text{B.11})$$

where  $k$  is the force constant for the chemical bond, typically 500 N/m, and  $\mu$  is again the reduced mass of the molecule. For vibrational transitions, selection rules are not as hard as for rotational transitions and so we observe various bands belonging to  $\Delta v = \pm 1, \pm 2, \pm 3, \dots$ . However, the transitions become weaker with increasing change in vibrational quantum number.

To make things more complex, a vibrating molecule can rotate at the same time. Each vibrational level contains a ladder of rotational levels that will be populated according to the excitation conditions in which the molecule sits (often, but not necessarily local thermodynamic equilibrium). The spectrum is then a ro-vibrational spectrum, where during a transition the vibrational as well as the rotational quantum number can change.

Vibrational transitions of simple molecules such as CO, CO<sub>2</sub> and H<sub>2</sub>O often lie in the infrared spectral range. There are two prominent windows accessible from Earth, 1 – 5  $\mu\text{m}$  and 8 – 20  $\mu\text{m}$ .

## B.4 Electronic transitions

The UV and visual spectroscopy of molecules corresponds to electronic transitions, that is electron rearrangements in the molecules. These are least constraining as they also change the strength of molecular bonds and can enable chemical reactions. An electronically excited molecule can become a reactive molecule.

Each electronic state has the full suite of vibrational states which is subsequently split up into the rotational ladder.

## B.5 Local thermodynamic equilibrium

The strength of these molecular transitions depends on their transition probability, the Einstein  $A$  coefficient, and on the population of the involved levels. In general, allowed lines are stronger than forbidden lines, reflected in their large difference of Einstein  $A$  coefficients. The levels are populated and de-populated by collisions and line transitions.

In many environments, collisions dominate the population and de-population of levels and thus the level populations follow from the Boltzmann equation

$$\frac{n_i}{n_j} = \frac{g_i}{g_j} e^{-\Delta E/kT} \quad (\text{B.12})$$

where  $n_i$  and  $n_j$  are the two level population numbers,  $g_i$  and  $g_j$  their respective statistical weights and  $\Delta E$  the energy difference between them. Collisions are said to thermalize the level population numbers to the corresponding local temperature  $T$  of the gas.

Rotational level population numbers are often in LTE, because the energy spacing between them is small ( $\Delta E \ll \text{few eV}$ ) and so even at low temperatures collision partners carry enough energy to excite them. The statistical weight  $g_i$  of a rotational level is given by  $(2J + 1)$ . Since the energy spacing is linear, but the statistical weight increases with  $J$ . For a each temperature, level population numbers peak around a fixed rotational quantum number and this is reflected in the line strength of the associated P- and R-branches.

The critical density introduced above is a measure of the density at which the LTE approximation holds. If the particle density is larger than  $n_{\text{crit}}$ , the level population numbers can be approximated using the Boltzmann equation and the local gas temperature. If  $n < n_{\text{crit}}$ , the levels will be subthermally excited, meaning that the temperature derived from the level excitation temperature,  $T_{\text{exc}}$  — as derived from the population numbers — will be lower than the actual gas temperature, the kinetic temperature  $T_{\text{kin}}$ .

## B.6 Statistical equilibrium

If LTE is no longer a valid approximation, we need to calculate the level population numbers from the equations of statistical equilibrium, i.e. we consider all collisional and radiative processes that populate and de-populate a level  $n_i$ . Hence, we can study the temporal change of population for each level

$$\begin{aligned} \frac{dn_i}{dt} = & \sum_{j>i} n_j (A_{ji} + B_{ji}P(\nu_{ji})) + \sum_{j<i} n_j B_{ij}P(\nu_{ij}) + n_i \sum_{j\neq i} n_j C_{ji} \\ & - n_i \sum_{j<i} (A_{ij} + B_{ij}P(\nu_{ij})) - n_i \sum_{j>i} B_{ij}P(\nu_{ij}) - n_i \sum_{j\neq i} n_j C_{ij} . \end{aligned} \quad (\text{B.13})$$

Here, the first three terms describe the population of level  $i$  through spontaneous and stimulated emission from higher levels, absorption from lower levels and collisions from all levels. The stimulated emission is described by the Einstein  $B$  coefficient, which is related through the expression

$$B_{ji} = \frac{c^2}{2h\nu^3} A_{ji} \quad (\text{B.14})$$

to the Einstein  $A$  coefficient. The relation between the stimulated emission and the absorption coefficient  $B_{ij}$  is

$$g_i B_{ij} = g_j B_{ji} \quad (\text{B.15})$$

The last three terms describe the de-population of level  $i$  via spontaneous and stimulated emission, absorption and collisions. Solving these equations requires a large amount of atomic/molecular data including accurate energy levels, transitions probabilities and collisional cross sections. Especially the latter are generally very difficult to find and often only crudely approximated. The accuracy of radiative transfer and line formation calculations depends largely on the level of accuracy of this atomic/molecular data.

## B.7 Molecular line observations

We can use molecules as tracers of physical conditions, e.g. gas density and temperature. The amount of flux emitted in a particular line can be written as

$$F_x(\nu_{ji}) = N_j A_{ji} h \nu_{ji} \beta (\tau_{ji}) \Omega_{\text{source}} \quad , \quad (\text{B.16})$$

where  $N_j$  is the column density of the molecule  $x$  being in the excited upper state  $j$  of the line transition with frequency  $\nu_{ji}$ .  $\beta$  is the escape probability of that line photon and the value of  $\beta$  depends obviously on the optical depth of that particular line  $\tau_{ji}$ .  $\Omega_{\text{source}}$  is the solid angle of the source (e.g. the molecular cloud) on the sky. The difficulties in using molecular lines as physical tracers becomes immediately clear: we need to know the level population of the upper energy level — which requires often accurate knowledge of molecular collision cross sections — and we need to understand the optical depth effects, here simplified in form of the escape probability  $\beta$ . The last uncertainty is the abundance of the molecular species itself. However, in some cases, we can use simplifications that allow us to use certain molecular lines as density or temperature indicators.

One such example is the CO molecule. It is a very robust molecule that is only photodissociated at the skin of the cloud (Photon Dominated Regions - PDRs). Molecular chemistry (see later sections) locks at high extinction all carbon in the form of CO. Hence, in the interior of molecular clouds, we can approximate the CO abundance to be almost constant  $\epsilon(\text{CO}) = 10^{-4}$ . The only exception occurs at low temperatures  $T < 20$  K, where CO can efficiently freeze out onto the cold dust grain surfaces. Since the molecule has very low critical densities for the low rotational transitions, we can assume LTE for the level populations of the rotational levels. To mitigate the problem of optical depth, we often use isotopes of CO which have up to a factor 1500 lower abundances.

Lines generally trace material in the density regime close to the critical density. The latter scales with  $\nu^3$ , i.e.  $J^3$  for the rotational lines. Hence, higher rotational lines trace warmer and denser gas. If we choose the molecule (dipole moment) and the line transition (rotational quantum number  $J$ ) well, we can probe gas in the entire temperature range between 10 and 1000 K and in the density range between  $10^2$  and  $10^{10}$   $\text{cm}^{-3}$ . In the following, we give a few examples of specific lines and their diagnostic power.

Fig. ?? illustrates the entire parameter space of temperature and density and outlines the molecule and transition best used in each regime. The  $J=1-0$ ,  $J=2-1$  CO lines are good temperature indicators for densities  $10^{2.5} < n(\text{H}_2) < 10^4$   $\text{cm}^{-3}$ . By using a combination of rotational lines, ranging from high to low optical depth, one can probe the physical conditions in the cloud as a function of depth. The highly optical thick lines of the main isotope originate close to the cloud surface, while the optically thin isotope lines probe down into the core of the molecular cloud.

At higher densities,  $\text{NH}_3$  can be used instead of CO. The oscillation of the nitrogen atom through the hydrogen plane (Fig. ??) causes an inversion transition at microwave frequencies. The main line is at 1.27 cm. The right hand side of the figure shows  $\text{NH}_3$  emission of the Ophiuchus F core.

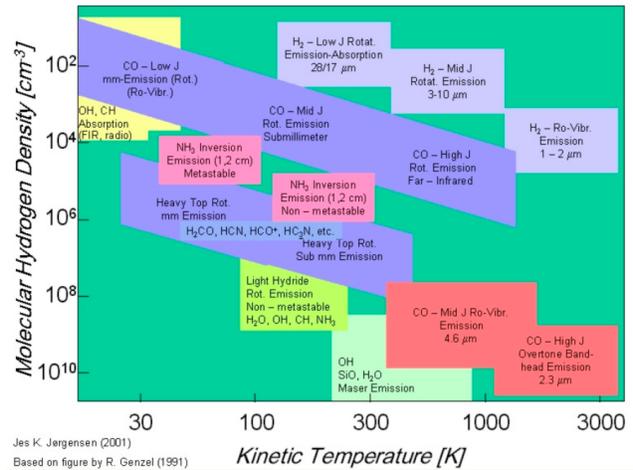


Figure B.5: Molecules as diagnostics as a function of gas temperature and density.

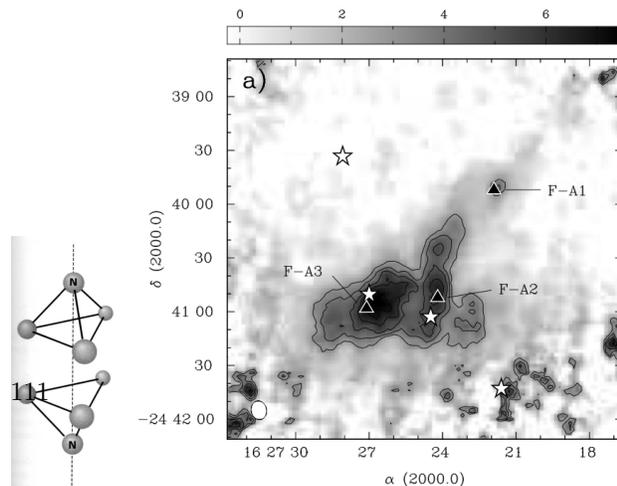


Figure B.6: Left: inversion transition of  $\text{NH}_3$ . Right: map of  $\text{NH}_3$  emission towards the Ophiuchus star forming core, Oph F. Stars denote locations of Class I pro-

Stars denote locations of Class I protostars, while triangles indicate the positions of NH<sub>3</sub> clumps. There is good correspondence between the two.

At very low densities, typical for diffuse clouds, the hydroxyl radical, OH, can be used as a tracer. Its ground state hyperfine transitions lie around 18 cm and are thus easily seen in absorption against the background continuum emission. Its rotational levels have very high excitation temperatures and critical densities and are thus often detected in shocks.

## B.8 Temperature structure

Molecular clouds consist of gas and dust with a typical mass ratio of 100. If the density and extinction are high, gas and dust temperature are well coupled through collisions and equal. Otherwise, both phases settle to their own equilibrium temperature.

## B.9 Dust temperature

In the case of dust, the temperature is set by radiative equilibrium, i.e. the radiation energy absorbed by each dust grain  $Q_+$  is re-emitted according to its temperature  $Q_-$  (Fig. ??). The timescale for reaching this equilibrium is generally short compared to the collisional energy exchange with the dust particles. In the optically thin case, we can write this balance as

$$Q_+ = Q_- \quad (\text{B.17})$$

$$\pi a^2 \int_0^\infty F_\nu \epsilon_\nu d\nu = 4\pi a^2 \int_0^\infty \pi B_\nu(T_{\text{dust}}) \epsilon_\nu d\nu .$$

Here,  $a$  is the radius of the dust grain,  $\epsilon_\nu$  its frequency dependent absorption efficiency. We have assumed here, that the grains radiate as black bodies and thus absorption and emission efficiencies are equal and we can use the Planck function to characterize the emitted spectrum of the dust grain.  $F_\nu$  is the external radiation field that the grain 'sees'. For certain grain size limits, we can solve this equation analytically. For that, we assume that the external radiation field can be characterized through a black body of temperature  $T_*$ . For large grains, the opacity is grey and we can write

$$T_{\text{dust}} = \sqrt{\frac{R_*}{2r}} T_* . \quad (\text{B.18})$$

This equation can for example be used in approximating the temperatures of planets in our Solar System. In the limit of micron-sized dust grains, the opacity is high ( $\epsilon = 1$ ) at short wavelength (compared to the size of the grains) and follows a power law  $1/\lambda$  at long wavelengths. Hence, we obtain

$$T_{\text{dust}} = 324 \left( \frac{L_*}{L_\odot} \right)^{0.2} \left( \frac{a}{\mu\text{m}} \right)^{-0.2} \left( \frac{r}{\text{AU}} \right)^{-0.4} . \quad (\text{B.19})$$

We use here the unit AU for the distance to the radiation source, because this formula is most useful in the context of debris disks. In molecular clouds, dust grains are typically much smaller than micron sized. And there, we have to solve the dust radiative equilibrium numerically. In addition, we have to take into account optical depth effects.

## Dust radiative equilibrium

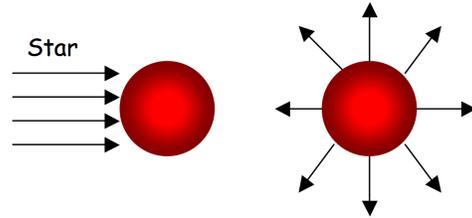


Figure B.7: Energy balance for a dust grain.

## B.10 Gas temperature

Just as for the dust, the gas temperature can be found from the equilibrium between heating and cooling processes

$$\sum_i \Gamma_i = \sum_j \Lambda_j , \quad (\text{B.20})$$

where the complication arises from the fact that there is not a single heating mechanisms or cooling mechanism, but a series of processes. Depending on the particular physical and chemical environment, we may be lucky to identify *the* most important heating and *the* most important cooling process and thus solve the formula analytically. But in most cases, we cannot identify a single process and even if, the process itself is often intimately coupled to the chemical structure of the cloud through the abundances of certain molecules.

## B.11 Heating processes

Heating of the cloud occurs through cosmic rays and radiation. Cosmic rays consist of relativistic protons and a mixture of heavy elements such as iron and electrons. Cosmic rays span an enormous range in energies between 10 and  $10^{14}$  MeV. They originate in supernova explosions.

A cosmic ray that travels into a molecular cloud interacts with the nuclei and electrons of the elements there. The excitation of the nuclei decays through emission of  $\gamma$ -rays that escape the cloud. The electronic excitation leads to dissociation, ionization, and emission of UV photons. The most likely process is ionization of molecular hydrogen — which is often the most abundant species inside the molecular clouds —



The ejected electron carries a large amount of kinetic energy, typically 30 eV, and can cause secondary ionization of molecular hydrogen as well as collisional excitation. The latter process is heating the gas. The heating rate per unit volume can be written as

$$\Gamma_{\text{CR}}(\text{H}_2) = \zeta(\text{H}_2)n_{\text{H}_2}\Delta E(\text{H}_2) . \quad (\text{B.22})$$

Here,  $\zeta(\text{H}_2)$  is the ionization rate of molecular hydrogen. The amount of heat released per ionization is  $\Delta E(\text{H}_2) = 7.0$  eV. The reason for  $\Delta E$  being smaller than the typical 30 eV is in fact the secondary ionization. We can write down a similar formula for neutral hydrogen using  $\zeta(\text{H I})$  and  $\Delta E(\text{H I}) = 6.0$  eV.

The dominant heating process at the surface of molecular clouds is photoelectric heating (Fig. ??). UV radiation hitting a dust grain ejects an electron which transmits its kinetic energy to the gas through subsequent collisions. The typical work function of a neutral grain is  $W \sim 6$  eV. The liberated photoelectron has a certain probability to leave the grain and must also overcome the charge potential of the grain (in case it is not neutral). Typical electron energies are of the order of 1 eV. The efficiency of ejecting a photoelectron for each incoming photon is close to one for very small grains. Only in grains that are larger than the typical mean free path of the electron inside the grain, this efficiency drops significantly. The reason for this is that the photoelectron has a higher probability to be re-absorbed inside the grain. We can write the heating rate as

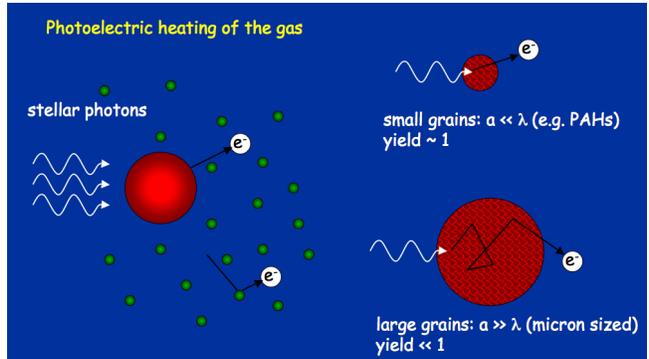


Figure B.8: Photoelectric effect on interstellar and circumstellar dust grains.

$$\Gamma_{\text{PE}} = 4\pi n_{\text{dust}}\sigma_{\text{dust}}\epsilon_{\text{PE}} \int_0^{\nu_{\text{w}}} J_{\nu} d\nu , \quad (\text{B.23})$$

where  $n_{\text{dust}}$  and  $\sigma_{\text{dust}}$  are the dust grain number density and the grain surface area.  $\epsilon_{\text{PE}}$  is the efficiency of ejecting a photoelectron. The interstellar radiation field is integrated down to the frequency  $\nu_{\text{W}}$ , which is the threshold frequency corresponding to the work function  $W$ .

## B.12 Cooling processes

Most of the cooling in molecular clouds is due to collisional ionization of atoms or molecules and subsequent radiation of a photon that escapes the cloud. One of the most efficient molecules in the cooling of molecular clouds is CO through its rotational lines. At the surface of clouds, where molecules are photodissociated, most of the cooling is due to fine structure lines of neutral oxygen and ionized carbon, [O I] 63 and 145  $\mu\text{m}$  and [C II] 157  $\mu\text{m}$ .

Dust grains can also cool the gas if the dust temperature is lower than that of the gas. In that case, collisions between gas particles and dust grains lead to an exchange of energy which does not affect the dust radiative equilibrium (see note above on the timescale of dust radiative equilibrium)

$$\Lambda_{\text{gas-grain}} = \frac{3}{2} k_B (T_{\text{gas}} - T_{\text{dust}}) \frac{n_{\text{dust}}}{t_{\text{coll}}} . \quad (\text{B.24})$$

Here,  $k_B$  is the Boltzmann constant and  $t_{\text{coll}}$  is the collision timescale. We see from this formula that the rate turns into a heating rate if  $T_{\text{dust}} > T_{\text{gas}}$ .

## B.13 Cloud temperatures

We see from the above processes of heating and cooling that most heating rates scale linearly with density, while cooling processes scale quadratic in density (at least under LTE conditions where collisions dominate the level population). Hence, denser environments will be colder, i.e. the molecular clouds are cooler than diffuse clouds.

At the surface of the molecular clouds, molecules are efficiently photodissociated and the ambient radiation field can easily penetrate up to an extinction of  $A_V \sim 1$ . This surface is often called a photon dominated region (PDR). The MC surface is mostly atomic consisting of atomic hydrogen, neutral oxygen and ionized carbon. Most metals with a low ionization energy such as carbon, silicon, magnesium, iron etc. are efficiently ionized by the interstellar radiation field. Hydrogen and oxygen have higher ionization potentials of 13.60 and 13.62 eV. Thus at the surface of the cloud, the most relevant heating process is photoelectric heating while cooling proceeds through the [C II] fine structure line at 157  $\mu\text{m}$ . We can thus approximate the gas temperature by equating those two processes

$$\Gamma_{\text{PE}} = \Lambda_{\text{C II}} , \quad (\text{B.25})$$

yielding an approximate expression for the gas temperature at the surface

$$T_{\text{gas}} = \frac{40 \text{ K}}{2.0 + \log(n_{\text{H}}/10^3 \text{ cm}^{-3})} . \quad (\text{B.26})$$

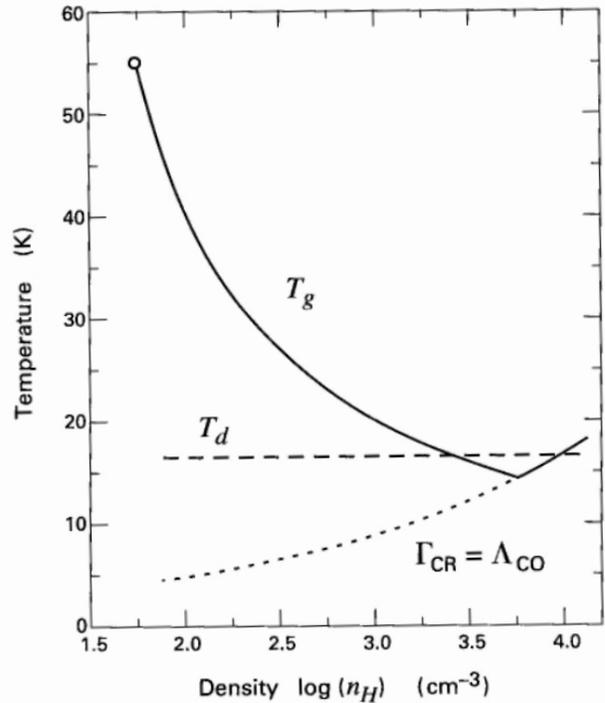


Figure B.9: Temperature profiles in the lower density region of a molecular cloud. The solid curve represents the gas temperature, while the dashed line refers to dust grains. The lower dotted curve results from balancing cosmic-ray heating and CO cooling in the gas (figure and caption from Stahler & Palla 2004).

Deep inside molecular clouds, cosmic ray heating is the only remaining heating process — photons cannot reach here due to the high extinction. The cooling is predominantly by CO rotational lines. We can thus find the gas temperature from

$$\Gamma_{\text{CR}} = \Lambda_{\text{COrot}} . \quad (\text{B.27})$$

Typical gas temperatures are of the order of 10 K. Fig. ?? illustrates the typical temperature profile derived for a molecular cloud.

## B.14 Chemistry

Just as level populations can be approximated through LTE (local thermodynamic equilibrium) or SE (statistical equilibrium), chemistry can assume various equilibria.

The most constraining one is thermodynamic equilibrium in which each reaction is counterbalanced by its own backreaction, leading to the concept of equilibrium constants  $K$  describing these as a function of temperature

$$K = \frac{k_{\text{forward}}}{k_{\text{backward}}} . \quad (\text{B.28})$$

These constants depend only on temperature and this type of equilibrium is typically found in very dense environments such as stellar atmospheres, planetary atmospheres and the densest inner regions of protoplanetary disks.

The alternative is stationary chemistry, which assumes that the density of each species reaches a stationary value, i.e. does not change with time. For gas phase chemistry, this can be described with a rate equation approach, where the time variation in the average density of a species is given by the sum of various production and loss rates. A generic rate equation for species  $i$  is

$$\begin{aligned} \frac{dn_i}{dt} &= P_i - L_i \\ &= \sum_{jl} k_{ijl}(T_g)n_jn_l + \sum_j (\Gamma_{ij}(r, z) + \zeta_{ij}(r, z)) - n_i \left[ \sum_{jl} k_{jil}(T_g)n_l + \sum_j (\Gamma_{ji}(r, z) + \zeta_{ji}(r, z)) \right] , \end{aligned} \quad (\text{B.29})$$

where  $n_i$  is the average volume density of species  $i$  and  $P_i$  and  $L_i$  are the chemical production and loss rates for that species. The first and third sums are the production and loss terms for species  $i$  through chemical reactions with species  $j$  and  $l$ , at rates  $k_{ijl}$  and  $k_{jil}$ . The second and fourth sums are the production and loss terms for species  $i$  through photoreactions involving species  $j$  (at rates  $\Gamma_{ij}$  and  $\Gamma_{ji}$ ) and cosmic ray reactions involving species  $j$  (at rates  $\zeta_{ij}$  and  $\zeta_{ji}$ ).

On the other hand, grain surface chemistry is stochastic in nature and strongly depends on the surface coverage of the species involved in the reaction at any instant. Hence, it is much better described using a master equation approach instead of a rate equation,

$$\frac{d}{dt}P(i_1 \dots i_N) = \sum (\text{accretion} + \text{evaporation} + \text{surface reactions}) ,$$

where  $P$  is the probability that  $i$  particles of species 1 to  $N$  are on the grain surface. Caselli et al. (1998) and Caselli et al. (2002) find that the rate equation approach can be modified for grain surface chemistry by taking into account the accretion timescale for a species as well as its migration timescale on the grain surface. This enables efficient simultaneous treatment of gas- and dust-phase chemistry.

Under some astrophysical conditions — low densities, low temperatures — , chemical equilibrium is hardly reached on timescales of  $10^5 - 10^6$  yr. This is especially relevant for diffuse and molecular clouds. In that case, we have to solve the time-dependent rate equations.

## B.15 The formation of molecules

Under typical molecular cloud conditions (temperature and density), chemistry proceeds through two body interaction. If we first consider the collision of two atoms. The formation of a molecule requires the excess energy needs to be either transmitted to a third body or radiated away. A third collision partner is not an option given the low density of interstellar space. An exception is the formation of molecular hydrogen on grain surfaces, where the grain acts as the third body. Alternatively, a photon can carry away the excess energy in a reaction type called radiative association. Here, the molecule is formed in an excited state that decays radiatively to the ground state. However, the most important type of reaction in interstellar chemistry is ion-molecule reactions



The rate coefficient for such reactions is of the order of  $k \sim 10^{-9} \text{ cm}^3/\text{s}$  and depends only weakly on temperature. The reason for the large rate coefficient is that the charged particle induces a dipole moment in the neutral one, thereby enhancing the cross section of the reaction over the geometric one (Coulomb interaction). The counterbalancing process is often dissociative recombination with free electrons



The electron recombines with the positive molecule creating an excited unstable neutral molecule. In most cases, the molecule autoionizes and returns the electron to the gas phase. However, if the constituents separate before autoionization occurs, the molecule dissociates into two neutral species. The typical rate coefficient is of the order of  $10^{-7} \text{ cm}^3/\text{s}$  for temperatures around 100 K and increases slowly with decreasing temperature.

## B.16 Chemical networks

In order to understand molecule formation, we need to take into account all possible reactions among all possible species. Given the sheer amount of atoms and molecules, this sounds like a hopeless endeavor. However, there exist large molecular databases that contain collections of molecular reactions that can occur under interstellar densities and temperatures. One of them is the UMIST database.

Selecting a number of atoms and molecules — based on abundance and experience — and using such a database allows us to build a chemical reaction network that interconnects all those species through chemical reactions. We can then solve for the abundances of all species using either a stationary approximation  $dn_i/dt = 0$  or the time-dependant approach.

In this way, we find for example that under the typical conditions of dense cores, all carbon is turned into CO on timescales of  $10^6 \text{ yr}$ .

## B.17 Chemical structure of molecular clouds

The surface of a molecular cloud resembles a PDR and shows the typical H/H<sub>2</sub> and C<sup>+</sup>/C/CO transition with increasing  $A_V$  (Fig. ??). The location where hydrogen turns molecular can be simply found from the equilibrium between H<sub>2</sub> formation on grain surfaces and H<sub>2</sub> photodissociation by interstellar UV photons

$$R_{\text{form}} n(\text{H}) n_{\text{H}} = \zeta(N_{\text{H}_2}) n(\text{H}_2) \quad . \quad (B.32)$$

The photodissociation process of H<sub>2</sub> proceeds via discrete bands. Once these H<sub>2</sub> bands become optically thick, photodissociation is largely suppressed

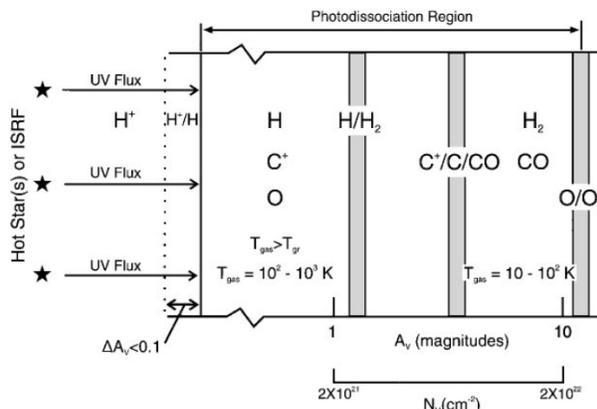
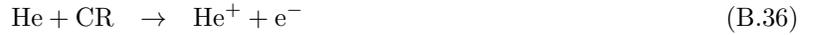


Figure B.10: A schematic diagram of a photodissociation region (PDR). The PDR is illuminated from the left and extends from the predominantly atomic surface region to the point where O<sub>2</sub> is not appreciably photodissociated ( $A_V \sim 10$  visual magnitude). Hence, the PDR includes gas whose hydrogen is mainly H<sub>2</sub> and whose carbon is mostly CO. Large columns of warm O, C, C<sup>+</sup>, and CO and vibrationally excited H<sub>2</sub> are produced in

by self-shielding. Hence, the photodissociation rate  $\zeta$  depends to first order on the column density of  $\text{H}_2$ ,  $N_{\text{H}_2}$ . Assuming now that the total hydrogen number density is  $n_{\text{H}} = n(\text{H}) + 2n(\text{H}_2)$ , we can work out the molecular fraction as a function of column density into the cloud

$$\frac{n(\text{H}_2)}{n(\text{H})} = \frac{R_{\text{form}}n_{\text{H}}}{2R_{\text{form}}n_{\text{H}} + \zeta(N_{\text{H}_2})} . \quad (\text{B.33})$$

The ionization degree inside the molecular cloud is low, because CR are the only remaining ionization source via the following processes



The secondary electrons are energetic and cause additional ionization and also excitation of molecular hydrogen. When  $\text{H}_2$  decays to the ground state, secondary UV photons are generated that can cause additional ionization deep inside the cloud. Dust grains can act as a sink of electrons and become negatively charged. Typical electron densities in the cloud interior are of the order of  $10^{-7} n_{\text{H}}$ .

