Control of star formation by supersonic turbulence

Mordecai-Mark Mac Low*

Department of Astrophysics, American Museum of Natural History, 79th Street at Central Park West, New York, New York 10024-5192, USA

Ralf S. Klessen†

Astrophysikalisches Institut Potsdam, An der Sternwarte 16, D-14482 Potsdam, Germany and UCO/Lick Observatory, University of California, Santa Cruz, California 95064, USA

(Published 7 January 2004)

Understanding the formation of stars in galaxies is central to much of modern astrophysics. However, a quantitative prediction of the star formation rate and the initial distribution of stellar masses remains elusive. For several decades it has been thought that the star formation process is primarily controlled by the interplay between gravity and magnetostatic support, modulated by neutral-ion drift (known as ambipolar diffusion in astrophysics). Recently, however, both observational and numerical work has begun to suggest that supersonic turbulent flows rather than static magnetic fields control star formation. To some extent, this represents a return to ideas popular before the importance of magnetic fields to the interstellar gas was fully appreciated. This review gives a historical overview of the successes and problems of both the classical dynamical theory and the standard theory of magnetostatic support, from both observational and theoretical perspectives. The outline of a new theory relying on control by driven supersonic turbulence is then presented. Numerical models demonstrate that, although supersonic turbulence can provide global support, it nevertheless produces density enhancements that allow local collapse. Inefficient, isolated star formation is a hallmark of turbulent support, while efficient, clustered star formation occurs in its absence. The consequences of this theory are then explored for both local star formation and galactic-scale star formation. It suggests that individual star-forming cores are likely not quasistatic objects, but dynamically collapsing. Accretion onto these objects varies depending on the properties of the surrounding turbulent flow; numerical models agree with observations showing decreasing rates. The initial mass distribution of stars may also be determined by the turbulent flow. Molecular clouds appear to be transient objects forming and dissolving in the larger-scale turbulent flow, or else quickly collapsing into regions of violent star formation. Global star formation in galaxies appears to be controlled by the same balance between gravity and turbulence as small-scale star formation, although modulated by cooling and differential rotation. The dominant driving mechanism in star-forming regions of galaxies appears to be supernovae, while elsewhere coupling of rotation to the gas through magnetic fields or gravity may be important.

CONTENTS

I. Introduction
   A. Overview
   B. Turbulence
   C. Outline

II. Observations
   A. Composition of molecular clouds
   B. Density and velocity structure of molecular clouds
   C. Support of molecular clouds
   D. Scaling relations for molecular clouds
   E. Protostellar cores
      1. From cores to stars
      2. Properties of protostellar cores
   F. The observed initial mass function

III. Historical Development
   A. Classical dynamical theory
   B. Problems with classical theory
   C. Standard theory of isolated star formation
   D. Problems with standard theory
      1. Singular isothermal spheres
      2. Observations of clouds and cores
         a. Magnetic support
         b. Infall motions
         c. Density profiles
         d. Chemical ages
      3. Protostars and young stars
         a. Accretion rates
         b. Embedded objects
         c. Stellar ages

IV. Toward a New Paradigm
   A. Maintenance of supersonic motions
   B. Turbulence in self-gravitating gas
   C. A numerical approach
   D. Global collapse
   E. Local collapse in globally stable regions
   F. Effects of magnetic fields
   G. Promotion and prevention of local collapse
   H. The time scales of star formation
      I. Scales of interstellar turbulence
      J. Termination of local star formation
      K. Outline of a new theory of star formation

V. Local Star Formation
   A. Star formation in molecular clouds
   B. Protostellar core models
   C. Binary formation
   D. Dynamical interactions in clusters
   E. Accretion rates
   F. Initial mass function
      1. Models of the IMF
      2. Turbulent fragmentation example
   G. Initial mass function

VI. Galactic-Scale Star Formation
   A. Formation and lifetime of molecular clouds
   B. When is star formation efficient?
      1. Overview
      2. Gravitational instabilities in galactic disks
   C. Driving mechanisms

*Electronic address: mordecai@amnh.org
†Electronic address: rklessen@aip.de

0034-6861/2004/76(1)/125(70)/$40.00 ©2004 The American Physical Society
I. INTRODUCTION

A. Overview

Stars are important. They are the dominant source of radiation (with competition from the cosmic microwave background and from accretion onto black holes, which themselves probably formed from stars), and of all chemical elements heavier than the H, He, and Li that made up the primordial gas. The Earth itself consists mainly of these heavier elements, called metals in astronomical terminology. Metals are produced by nuclear fusion in the interior of stars, with the heaviest elements produced during the passage of the final supernova shockwave through the most massive stars. To reach the chemical abundances observed today in our solar system, the material had to go through many cycles of stellar birth and death. In a literal sense, we are star dust.

Stars are also our primary source of astronomical information and hence are essential for our understanding of the universe and the physical processes that govern its evolution. At optical wavelengths almost all natural light we observe in the sky originates from stars. During the day this is obvious, but it is also true at night. The Moon, the second brightest object in the sky, reflects light from our Sun, as do the planets, while virtually every other extraterrestrial source of visible light is a star or collection of stars. Throughout the millennia, these objects have been the observational targets of traditional astronomy and define the celestial landscape, the constellations.

When we look at the sky on a clear night, we can also note dark patches of obscuration along the band of the Milky Way. These are clouds of dust and gas that block the light from stars further away. For roughly the last century we have known that these clouds give birth to stars. The advent of new observational instruments made it possible to observe astronomical objects at wavelengths ranging from γ rays to radio frequencies. Especially useful for studying the dark clouds are radio, submillimeter, and far-infrared wavelengths, at which they are transparent. Observations now show that all star formation occurring in the Milky Way is associated with the dark clouds of molecular hydrogen and dust.

Stars are common. The mass of the Galactic disk plus bulge is about \(6 \times 10^{10} M_\odot\) (e.g., Dehnen and Binney, 1998), where \(1 M_\odot = 1.99 \times 10^{33} \text{ g}\) is the mass of our Sun. Thus there are of order \(10^{12}\) stars in the Milky Way, assuming standard values for the stellar mass distribution (e.g., Kroupa, 2002). Stars form continuously. Roughly 10% of the disk mass of the Milky Way is in the form of gas, which is forming stars at a rate of about \(1 M_\odot \text{ yr}^{-1}\). Although stars dominate the baryonic mass in the Galaxy, dark matter determines the overall mass budget: invisible material that reveals its presence only by its contribution to the gravitational potential. The dark-matter halo of our Galaxy is about ten times more massive than gas and stars together. At larger scales this imbalance is even more pronounced. Stars are estimated to make up only 0.4% of the total mass of the universe (Lanzetta, Yahil, and Fernandez-Soto, 1996) and about 17% of the total baryonic mass (Walker et al., 1991).

Mass is the most important parameter determining the evolution of individual stars. Massive stars with high pressures at their centers have strong nuclear fusion reactions, making them short lived but very luminous, while low-mass stars are long lived but extremely faint. For example, a star with \(5 M_\odot\) only lives for \(2.5 \times 10^7\) yr, while a star with \(0.2 M_\odot\) survives for \(1.2 \times 10^{13}\) yr, orders of magnitude longer than the current age of the universe. For comparison the Sun, with an age of \(4.5 \times 10^9\) yr, has reached approximately half of its life span. The relationship between mass and luminosity is quite steep, with roughly \(L \propto M^{3.2}\) (Kippenhahn and Weigert, 1990). During its short life a \(5 M_\odot\) star will shine with a luminosity of \(1.5 \times 10^6 L_\odot\), while the luminosity of an \(0.2 M_\odot\) star is only \(\sim 10^{-3} L_\odot\). For reference, the luminosity of the Sun is \(L_\odot = 3.85 \times 10^{33} \text{ erg s}^{-1}\).

The light from star-forming external galaxies in the visible and blue wavebands is dominated by young, massive stars. This is the reason why we observe beautiful spiral patterns in many disk galaxies, like NGC 4622 shown in Fig. 1, as spiral density waves lead to gas compression and subsequent star formation at the wave locations. Massive stars dominate the optical emission from external galaxies. In their brief lifetimes, massive stars do not have sufficient time to disperse in the galactic disk, so they still trace the characteristics of the instability that triggered their formation. Hence understanding the dynamical properties of galaxies requires an understanding of how, where, and under which conditions stars form.

In a simple approach, galaxies can be seen as gravitational potential wells containing gas that has been able to cool radiatively in less than the current age of the universe. In the absence of any hindrance, the gas then collapses gravitationally to form stars in a free-fall time (Jeans, 1902),

\[
\tau_{ff} = \left( \frac{3 \pi}{32 G \rho} \right)^{1/2} = 140 \text{ Myr} \left( \frac{n}{0.1 \text{ cm}^{-3}} \right)^{-1/2},
\]

where \(\tau_{ff}\) is the free-fall time, \(G\) is the gravitational constant, and \(n\) is the number density of gas.
where \( n \) is the number density of the gas. Interstellar gas in the Milky Way consists of one part He for every ten parts H. The mass density \( \rho = \mu n \), where we take the Galactic value for the mean mass per particle in neutral atomic gas of \( \mu = 2.11 \times 10^{-24} \) g, and \( G \) is the gravitational constant. The free-fall time \( \tau_f \) is very short compared to the age of the Milky Way, about \( 10^{10} \) yr. However, gas remains in the Galaxy and stars continue to form from gas that must already have been cooled below its virial temperature for many billions of years. What physical processes regulate the rate at which gas turns into stars? Another way of asking the question is, what prevented the Galactic gas from forming stars at an extremely high rate immediately after it first cooled and then being completely used up?

Observations of the star formation history of the universe demonstrate that stars did indeed form more vigorously in the past than today (see, for example, Lilly et al., 1996; Madau et al., 1996; Baldry et al., 2002; Lanzetta et al., 2002), with as much as 80% of star formation in the universe being complete by redshift \( z = 1 \), less than half of the current age of 13 Gyr. What mechanisms allowed rapid star formation in the past, but reduce its rate today?

The clouds of gas and dust in which stars form are dense enough, and well enough protected from dissociating UV radiation by self-shielding and dust scattering in their surface layers, for hydrogen to be mostly in molecular form in their interior. The density and velocity structure of these molecular clouds is extremely complex and follows hierarchical scaling relations that appear to be determined by supersonic turbulent motions (Blitz and Williams, 1999). Molecular clouds are large, and their masses exceed the threshold for gravitational collapse by far when taking only thermal pressure into account. Just like galaxies as a whole, naively speaking, they should be contracting rapidly and forming stars at a very high rate. This is generally not observed. We can define the star formation efficiency of a region as

\[
\epsilon_{SF} = \frac{M_\odot}{M} \frac{\tau}{M}
\]

(2)

where \( M_\odot \) is the star formation rate, \( \tau \) is the lifetime of the region, and \( M \) is the total gas mass in the region (Elmegreen and Efremov, 1997). The star formation efficiency of molecular clouds in the solar neighborhood is estimated to be of the order of a few percent (Zuckerman and Evans, 1974).

For many years it was thought that support by magnetic pressure against gravitational collapse offered the best explanation for the slow rate of star formation. In this theory, developed by Shu (1977; and see Shu, Adams, and Lizano, 1987), Mouschovias (1976; and see Mouschovias, 1991b, 1991c), Nakano (1976), and others, interstellar magnetic field prevents the collapse of gas clumps with insufficient mass-to-flux ratio, leaving dense cores in magnetohydrostatic equilibrium. The magnetic field couples only to electrically charged ions in the gas, though, so neutral atoms can only be supported by the field if they collide frequently with ions. The diffuse interstellar medium (ISM), with number densities \( n \approx 1 \) cm\(^{-3} \) (see Ferrière, 2001 for a general review of ISM properties), remains highly ionized, enough that neutral-ion collisional coupling is very efficient (as we discuss below in Sec. III.C). In dense cores, where \( n > 10^5 \) cm\(^{-3} \), ionization fractions drop below parts per ten million. Neutral-ion collisions no longer couple the neutrals tightly to the magnetic field, so the neutrals can diffuse through the field. This neutral-ion drift allows gravitational collapse to proceed in the face of magnetostatic support, but on a time scale as much as an order of magnitude longer than the free-fall time, drawing out the star formation process.

In this paper we review a body of work that suggests that magnetohydrostatic support modulated by neutral-ion drift fails to explain the star formation rate and indeed appears inconsistent with observations of star-forming regions. Instead, we suggest that control of molecular cloud formation and subsequent support by supersonic turbulence is both sufficient to explain star formation rates and more consistent with observations. Our review focuses on how gravitationally collapsing regions form. The recent comprehensive review by Larson (2003) goes into more detail on the final stages of disk accretion and protostellar evolution.

B. Turbulence

At this point, we need to discuss briefly the concept of turbulence and the differences between supersonic, compressible (and magnetized) turbulence and the more
commonly studied incompressible turbulence. We mean by turbulence, in the end, nothing more than the gas flow resulting from random motions at many scales. We furthermore shall use in our discussion only the very general properties and scaling relations of turbulent flows, focusing mainly on effects of compressibility. For a more detailed discussion of the complex statistical characteristics of turbulence, we refer the reader to the book by Lesieur (1997).

Most studies of turbulence treat incompressible turbulence, characteristic of most terrestrial applications. Root-mean-square (rms) velocities are subsonic, and the density remains almost constant. Dissipation of energy occurs primarily in the smallest vortices, where the dynamical scale $\ell$ is shorter than the length on which viscosity acts $\ell_{\text{visc}}$. Kolmogorov (1941a) described a heuristic theory based on dimensional analysis that captures the basic behavior of incompressible turbulence surprisingly well, although subsequent work has refined the details substantially. He assumed turbulence driven on a large scale $L$, forming eddies at that scale. These eddies interact to form slightly smaller eddies, transferring some of their energy to the smaller scale. The smaller eddies in turn form even smaller ones, until energy has cascaded all the way down to the dissipation scale $\ell_{\text{visc}}$.

In order to maintain a steady state, equal amounts of energy must be transferred from each scale in the cascade to the next, and eventually dissipated, at a rate

$$
\dot{E} = \eta v^3/L,
$$

where $\eta$ is a constant determined empirically. This leads to a power-law distribution of kinetic energy $E \propto v^2 \propto k^{-11/3}$, where $k = 2\pi/\ell$ is the wave number, and density does not enter because of the assumption of incompressibility. Most of the energy remains near the driving scale, while energy drops off steeply below $\ell_{\text{visc}}$. Because of the apparently local nature of the cascade in wave-number space, the viscosity only determines the behavior of the energy distribution at the bottom of the cascade below $\ell_{\text{visc}}$, while the driving only determines the behavior near the top of the cascade at and above $L$.

The region in between is known as the inertial range, in which energy transfers from one scale to the next without influence from driving or viscosity. The behavior of the flow in the inertial range can be studied regardless of the actual scale at which $L$ and $\ell_{\text{visc}}$ lie, so long as they are well separated. One statistical description of incompressible turbulent flow, the structure functions $S_p(\vec{r}) = \langle (v(\vec{x}) - v(\vec{x} + \vec{r}))^p \rangle$, has been successfully modeled by assuming that dissipation occurs in filamentary vortex tubes (She and Leveque, 1994).

Gas flows in the ISM, however, vary from this idealized picture in three important ways. First, they are highly compressible, with Mach numbers $M$ ranging from order unity in the warm ($10^4$ K), diffuse ISM, up to as high as 50 in cold ($10$ K), dense molecular clouds. Second, the equation of state of the gas is very soft due to radiative cooling, so that pressure $P \propto \rho^\gamma$ with the polytropic index falling in the range $0.4 < \gamma < 1.2$ as a function of density and temperature (see, for example, Scalo et al., 1998; Ballesteros-Paredes, Vázquez-Semadeni, and Scalo, 1999; Spaans and Silk, 2000). Third, the driving of the turbulence is not uniform, but rather comes from blast waves and other inhomogeneous processes.

Supersonic flows in highly compressible gas create strong density perturbations. Early attempts to understand turbulence in the ISM (von Weizsäcker, 1943, 1951; Chandrasekhar, 1949) were based on insights drawn from incompressible turbulence. An attempt to analytically derive the density spectrum and resulting gravitational collapse criterion was first made by Chandrasekhar (1951a, 1951b). This work was followed up by several authors, culminating in work by Sasao (1973) on density fluctuations in self-gravitating media whose interest has only recently been appreciated. Larson (1981) qualitatively applied the basic idea of density fluctuations driven by supersonic turbulence to the problem of star formation. Bonazzola et al. (1992) used a renormalization-group technique to examine how the slope of the turbulent velocity spectrum could influence gravitational collapse. This approach was combined with low-resolution numerical models to derive an effective adiabatic index for subsonic compressible turbulence by Panis and Pérault (1998). Adding to the complexity of the problem, the strong density inhomogeneities observed in the ISM can be caused not only by compressible turbulence, but also by thermal phase transitions (Field, Goldsmith, and Habing, 1969; McKee and Ostriker, 1977; Wolfire et al., 1995) or gravitational collapse (Kim and Ostriker, 2001).

In supersonic turbulence, shock waves offer additional possibilities for dissipation. Shock waves can also transfer energy between widely separated scales, removing the local nature of the turbulent cascade typical of incompressible turbulence. The spectrum may shift only slightly, however, as the Fourier transform of a step function representative of a perfect shock wave is $k^{-2}$. Integrating in three dimensions over an ensemble of shocks, one finds the differential energy spectrum $E(k) = \rho u^2(k) k^2 dk \propto k^{-2} dk$. This is just the compressible energy spectrum reported by Porter and Woodward (1992) and Porter, Pouquet, and Woodward (1992, 1994). They also found that, even in supersonic turbulence, the shock waves do not dissipate all the energy, as rotational motions continue to contain a substantial fraction of the kinetic energy, which is then dissipated in small vortices. Boldyrev (2002) has proposed a theory of velocity structure-function scaling based on the work of She and Leveque (1994) using the assumption that dissipation in supersonic turbulence primarily occurs in sheetlike shocks, rather than linear filaments at the centers of vortex tubes. The first comparisons to numerical models show good agreement with this model (Boldyrev, Nordlund, and Padoan, 2002a), and it has been extended to the density structure functions by Boldyrev, Nordlund, and Padoan (2002b).
physical context have been discussed by Avillez and Mac Low (2002) and Klessen and Lin (2003).

The driving of interstellar turbulence is neither uniform nor homogeneous. Controversy still reigns over the most important energy sources at different scales, but we make the argument in Sec. VI.C that isolated and correlated supernovae dominate. However, it is not yet understood at what scales expanding, interacting blast waves contribute to turbulence. Analytic estimates have been made based on the radii of the blast waves at late times (Norman and Ferrara, 1996), but never confirmed with numerical models (much less experiment). Indeed, the thickness of the blast waves may be more important than the radii.

Finally, the interstellar gas is magnetized. Although magnetic-field strengths are difficult to measure, with Zeeman line splitting being the best quantitative method, it appears that fields within an order of magnitude of equipartition with thermal pressure and turbulent motions are pervasive in the diffuse ISM, most likely maintained by a dynamo driven by the motions of the interstellar gas (Ferrière, 1992). A model for the distribution of energy and the scaling behavior of strongly magnetized, incompressible turbulence based on the interaction of shear Alfvén waves is given by Goldreich and Sridhar (1995, 1997) and Ng and Bhattacharjee (1996). They found that an anisotropic Kolmogorov spectrum $k^{-5/3}$ best describes the one-dimensional (1D) energy spectrum, rather than the $k^{-3/2}$ spectrum first proposed by Iroshnikov (1963) and Kraichnan (1965). These results have been confirmed by Verma et al. (1996) using numerical models, and by Verma (1999) using a renormalization-group approach. The scaling properties of the structure functions of such turbulence were derived from the work of She and Leveque (1994) by Müller and Biskamp (2000; also see Biskamp and Müller, 2000) by assuming that dissipation occurs in current sheets. A theory of weakly compressible turbulence applicable in particular to small scales in the ISM has been derived by Lithwick and Goldreich (2001), but little progress has been made towards analytic models of strongly compressible magnetohydrodynamic (MHD) turbulence with $M \gg 1$. See, however, the reviews by Cho, Lazarian, and Vishniac (2002), and Cho and Lazarian (2003). In particular, an analytic theory of the nonlinear density fluctuations characteristic of such turbulence remains lacking.

C. Outline

With the above in mind, we suggest that stellar birth is regulated by interstellar turbulence and its interplay with gravity. Turbulence, even if strong enough to counterbalance gravity on global scales, will usually provoke collapse on smaller scales. Supersonic turbulence establishes a complex network of interacting shocks, where converging flows generate regions of high density. This density enhancement can be sufficient for gravitational instability. Collapse sets in. However, the random flow that creates local density enhancements also may disperse them again. Hence the efficiency of star formation [Eq. (2)] depends strongly on the properties of the underlying turbulent velocity field, on its driving length scale and strength relative to gravitational attraction. This principle holds for star formation throughout all scales considered in this review, ranging from small star-forming regions up to galaxies as a whole.

To lay out this picture of star formation in more detail, we first outline the observed properties of star-forming interstellar clouds and the distribution of stellar masses that form there in Sec. II. We then critically discuss the historical development of star formation theory in Sec. III. We begin the section by describing the classical dynamic theory, and then move on to the so-called standard theory, in which the star formation process is controlled by magnetic fields. After describing the theoretical and observational problems that both approaches have, we present work in Sec. IV that leads us to the argument that star formation is controlled by the interplay between gravity and supersonic turbulence. The theory is applied to individual star-forming regions in Sec. V, where we investigate the implications for stellar clusters, protostellar cores (the direct progenitors of individual stars), binary stars, protostellar mass accretion, and the subsequent distribution of stellar masses. In Sec. VI, we discuss the control of star formation by supersonic turbulence on galactic scales. We first examine the formation and destruction of star-forming molecular clouds in light of models of turbulent flow. We then ask, When is star formation efficient in galaxies? We review the energetics of the possible mechanisms that generate and maintain supersonic turbulence in the interstellar medium, and come to the conclusion that supernova explosions accompanying the death of massive stars are the most likely agents. Then we briefly apply the theory to various types of galaxies, ranging from low-surface-brightness galaxies to massive star bursts. Finally, in Sec. VII, we summarize, and describe unsolved problems open for future research.

II. OBSERVATIONS

All present-day star formation takes place in molecular clouds (see, for example, Blitz, 1993; Williams, Blitz, and McKee, 2000), so it is vital to understand the properties, dynamical evolution, and fragmentation of molecular clouds in order to understand star formation. We begin this section by describing the composition (Sec. II.A) and density and velocity structure (Sec. II.B) of molecular clouds. We then discuss turbulent support of clouds against gravitational collapse (Sec. II.C) and introduce the observed scaling relations and their relation to the turbulent flow (Sec. II.D). Finally, we describe observations of protostellar cores (Sec. II.E) and of the initial mass function of stars (Sec. II.F).

A. Composition of molecular clouds

Molecular clouds are density enhancements in the interstellar gas dominated by molecular H$_2$ rather than the
atomic H typical of the rest of the ISM (Ferrière, 2001), mainly because they are opaque to the UV radiation that elsewhere dissociates the molecules. In the plane of the Milky Way, interstellar gas has been extensively reprocessed by stars, so the metallicity \(^1\) is close to the solar value \(Z_{\odot}\), while in other galaxies with lower star formation rates, the metallicity can be as little as \(10^{-3}Z_{\odot}\). The refractory elements condense into dust grains, while others form molecules. The properties of the dust grains change as the temperature drops within the cloud, probably due to the freezing of volatiles such as water and ammonia (Goodman et al., 1995). This has important consequences for the radiation transport properties and the optical depth of the clouds. The presence of heavier elements such as carbon, nitrogen, and oxygen determines the heating and cooling processes in molecular clouds (Genzel, 1991). In addition, continuum emission from dust and emission and absorption lines emitted by molecules formed from these elements are the main observational tracers of cloud structure, as cold molecular hydrogen is very difficult to observe. Radio and submillimeter telescopes mostly concentrate on the thermal continuum from dust and the rotational transition lines of carbon, oxygen, and nitrogen molecules (e.g., CO, \(\text{NH}_3\), or \(\text{H}_2\text{O}\)). By now, several hundred different molecules have been identified in the interstellar gas. An overview of the application of different molecules as tracers for different physical conditions can be found in the reviews by van Dishoeck et al. (1993), Langer et al. (2000), and van Dishoeck and Hogerheijde (2000).

### B. Density and velocity structure of molecular clouds

Emission-line observations of molecular clouds reveal clumps and filaments on all scales accessible by present-day telescopes. Typical parameters of different regions in molecular clouds are listed in Table I, adapted from Cernicharo (1991). The mass spectrum of clumps in molecular clouds appears to be well described by a power law, indicating self-similarity: there is no natural mass or size scale between the lower and upper limits of the observations. The largest molecular structures considered to be single objects are giant molecular clouds (GMC’s), which have masses of \(10^5 - 10^6 M_{\odot}\) and extend over a few tens of parsecs. The smallest observed structures are protostellar cores, with masses of a few solar masses or less and sizes of \(\lesssim 0.1\) pc, and less dense clumps of similar size. The volume filling factor \(\langle n \rangle / n\) (where \(n\) is the local density, while \(\langle n \rangle\) is the average density of the cloud) of dense clumps, even denser subclumps, and so on, is rather small, ranging from 10% down to 0.1% at densities of \(n > 10^4\) \(\text{cm}^{-3}\) (McKee, 1999). Star formation always occurs in the densest regions within a cloud, so only a small fraction of molecular cloud matter is actually involved in building up stars, while the bulk of the material remains at lower densities.

The density structure of molecular clouds is best inferred from the column density of dust, which can be observed either via its thermal emission at millimeter wavelengths in dense regions (Motte, André, and Neri, 1998; Testi and Sargent, 1998) or via its extinction of background stars in the infrared, if a uniform screen of background stars is present (Lada et al., 1994; Alves, Lada, and Lada, 2001). Deriving density and mass from thermal emission requires modeling the temperature profile, which depends on optically thick radiative transfer through uncertain density distributions. Infrared extinction, on the other hand, requires only suitable background stars. Reliance on the near-IR color excess to measure column densities ensures a much greater dynamic range than optical extinction. This method has been further developed by Cambresy et al. (2002), who use an adaptive grid to extract maximum information from nonuniform background star fields. It turns out that the higher the column density in a region, the higher the variation in extinction among stars behind

<table>
<thead>
<tr>
<th>Size (pc)</th>
<th>10–60</th>
<th>2–20</th>
<th>0.1–2</th>
<th>(\leq 0.1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density ([n(H_2)/\text{cm}^3])</td>
<td>100–500</td>
<td>(10^2–10^4)</td>
<td>(10^3–10^5)</td>
<td>&gt;10^5</td>
</tr>
<tr>
<td>Mass ((M_{\odot}))</td>
<td>(10^4–10^6)</td>
<td>(10^2–10^4)</td>
<td>(10–10^5)</td>
<td>0.1–10</td>
</tr>
<tr>
<td>Linewidth (km s(^{-1}))</td>
<td>5–15</td>
<td>1–10</td>
<td>0.3–3</td>
<td>0.1–0.7</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>7–15</td>
<td>10–30</td>
<td>10–30</td>
<td>7–15</td>
</tr>
<tr>
<td>Examples</td>
<td>W51, W3, M17, Orion-Monoceros, Taurus-Auriga-Perseus complex</td>
<td>L1641, L1630, W33, W3A, B227, L1495, L1529</td>
<td>see Sec. II.E</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Protostellar cores in the “prestellar” phase, i.e., before the formation of the protostar in its interior.

\(^1\)Metallicity in astrophysics is usually defined as the fraction of heavy elements relative to hydrogen. It averages over local variations in the abundance of the different elements caused by varying chemical enrichment histories.
that region (Lada et al., 1994). Padoan and Nordlund (1999) demonstrated this to be consistent with a super-Alfvénic turbulent flow, while Alves et al. (2001) modeled it with a single cylindrical filament with density $\rho \propto r^{-2}$. Because turbulence forms many filaments, it is not clear that these two descriptions are actually contradictory (Padoan, 2001), although the identification of a single filament would then suggest that a minimum scale for the turbulence has been identified.

A more general technique is emission in optically thin spectral lines. The best candidates are $^{13}$CO and C$^{18}$O, though CO freezes out in the very densest regions (with visual extinctions above $A_V \sim 10$ magnitudes; see Alves, Lada, and Lada, 1999). CO observations are therefore only sensitive to gas at relatively low densities, $n \lesssim 10^3$ cm$^{-3}$, and are limited in dynamic range to at most two decades of column density. The colder the gas, the lower the column density at which lines will become optically thick. Nevertheless, the development of sensitive radio receivers in the 1980s first made it feasible to map an entire molecular cloud region with high spatial and spectral resolution to obtain quantitative information about the overall density structure.

The hierarchy of clumps and filaments spans all observable scales (Falgarone, Puget, and Perault, 1992; Falgarone and Phillips, 1996; Wiesemeyer et al., 1997), extending down to individual protostars studied with millimeter-wavelength interferometry (Ward-Thompson et al., 1994; Langer et al., 1995; Gueth et al., 1997; Motte et al., 1998; Testi and Sargent, 1998; Ward-Thompson, Motte, and André, 1999; Bacmann et al., 2000; Motte et al., 2001). This is illustrated in Fig. 2, which shows $^{13}$CO, $^{12}$CO, and C$^{18}$O maps of a region in the Cygnus OB7 complex at three levels of successively higher resolution (from Falgarone et al., 1992). At each level, the molecular cloud appears clumpy and highly structured. When observed with higher resolution, each clump breaks up into a filamentary network of smaller clumps. Unresolved features exist even at the highest resolution. The ensemble of clumps identified in this survey covers a mass range from about 1$M_\odot$ up to a few 100$M_\odot$ and densities $50<n(H_2)<10^4$ cm$^{-3}$. These values are typical for all studies of cloud clump structure, with higher densities being reached primarily in protostellar cores.

The distribution of clump masses is consistent with a power law of the form

$$\frac{dN}{dm} \propto m^\alpha,$$

with $-1.3<\alpha<-1.9$ in molecular line studies (Carr, 1987; Stutzki and Güsten, 1990; Lada, Bally, and Stark, 1991; Williams, de Geus, and Blitz, 1994; Onishi et al., 1996; Heithausen et al., 1998; Kramer, 1998). Dust continuum studies, which pick out the regions of highest column density, find steeper values of $-1.9<\alpha<-2.5$ (Motte et al., 1998; Testi and Sargent, 1998; also see the discussion in Ossenkopf, Klessen, and Heitsch, 2001), similar to the stellar mass spectrum. The power-law mass spectrum is often interpreted as a manifestation of fractal density structure (see, for example, Elmegreen and Falgarone, 1996). However, the full physical meaning remains unclear. In most studies molecular cloud clumps are determined either by a Gaussian decomposition scheme (Stutzki and Güsten, 1990) or by the attempt to define (and separate) clumps following density peaks (Williams et al., 1994). There is no one-to-one correspondence between the identified clumps in either method, however. Furthermore, molecular clouds are only seen in projection, so one only measures column density instead of volume density. It remains unproven that all regions of high density also have high column density, and vice versa. Even when velocity information is taken into account, the real 3D structure of the cloud remains elusive. In particular, it can be demonstrated in models of interstellar turbulence that single clumps identified in simulated observational cubes (position-position-velocity space) tend to separate into multiple clumps in real 3D space (Ostriker et al., 2001; Ballesteros-Paredes and Mac Low, 2002). This effect acts over regions of velocity width similar to the velocity dispersion, enough to confound clumps even in clouds showing large-scale velocity gradients. These projection effects leave clump mass spectra as poor statistical tools for characterizing molecular cloud structure.

Other means of quantifying the structural and dynamical properties of molecular clouds involve correlations and probability distribution functions of dynamical variables. Two-point correlation functions have been studied by many authors, including Scalo (1984), Kleiner and Dickman (1987), Kitamura et al. (1993), Miesch and Bally (1994), LaRosa, Shore, and Magnani (1999), and Ballesteros-Paredes, Vázquez-Semadeni, and Goodman (2002), while other studies have concentrated on analyzing the probability distribution functions of the column density in observations, both physical and in computational models, and of dynamical observables such as the centroid velocities of molecular lines and their differences. The density probability distribution function has been used to characterize numerical simulations of the interstellar medium by Vázquez-Semadeni (1994), Padoan, Nordlund, and Jones (1997), Passot and Vázquez-Semadeni (1998), Scalo et al. (1998), and Klessen (2000). Velocity probability distribution functions for several star-forming molecular clouds have been determined by Miesch and Scalo (1995) and Miesch, Scalo, and Bally (1999). Lis et al. (1996, 1998) analyzed snapshots of a numerical simulation of mildly supersonic, decaying turbulence (without self-gravity) by Porter, Pouquet, and Woodward (1994) and applied the method to observations of the ρ-Ophiuchus cloud. The observed probability distribution functions exhibit strong non-Gaussian features, often being nearly exponential, with possible evidence for power-law tails in the outer parts. Further methods to quantify molecular cloud structure involve spectral correlation (Rosolowsky et al., 1999), principal component analysis (Heyer and Schloerb, 1997), or...
pseudometric methods used to describe and rank cloud complexity (Adams and Wiseman 1994; Wiseman and Adams, 1994).

A technique especially sensitive to the amount of structure on different spatial scales is wavelet analysis (Gill and Henriksen, 1990; Langer, Wilson, and Anderson, 1993). In particular, the $\Delta$ variance, introduced by Stutzki et al. (1998), provides a good separation of noise and observational artifacts from the real cloud structure. For isotropic systems its slope is directly related to the spectral index of the corresponding Fourier power spectrum. It can be applied in an equivalent way both to observational data and to gas-dynamic and MHD turbulence simulations, allowing a direct comparison, as discussed by Mac Low and Ossenkopf (2000), Bensch, Stutzki, and Ossenkopf (2001), and Ossenkopf and Mac Low (2002). They find that the structure of low-density gas in molecular clouds is dominated by large-scale modes and, equivalently, the velocity field by large-scale motions. This means that molecular cloud turbulence is likely to be driven from the outside, by sources acting external to the cloud on scales of at least several tens of parsec (Ossenkopf and Mac Low, 2002).

The observational findings are different, however, when focusing on high-density gas in star-forming regions. In this case, the $\Delta$ variance clearly shows that the

FIG. 2. Maps of the molecular gas in the Cygnus OB7 complex: (a) Large-scale map of the $^{13}$CO ($J=1\rightarrow0$) emission. The first level and the contour spacing are 0.25 K. (b) Map of the same transition line of a subregion with higher resolution (first contour level and spacing are 0.3 K). Both maps were obtained with the Bordeaux telescope. (c) $^{12}$CO ($J=1\rightarrow0$) and (d) $^{13}$CO ($J=1\rightarrow0$) emission from the most transparent part of the field. (e) $^{13}$CO ($J=1\rightarrow0$) and (f) C$^{18}$O ($J=1\rightarrow0$) emission from the most opaque field. (g) $^{13}$CO ($J=1\rightarrow0$) and (h) C$^{18}$O ($J=1\rightarrow0$) emission from a filamentary region with medium density. The indicated linear sizes are given for a distance to Cygnus OB7 of 750 pc. From Falgarone et al., 1992.
density structure is dominated by individual protostellar cores at the smallest resolved scales (Ossenkopf et al., 2001). This effect is best seen in dust emission because it is able to trace large density contrasts. Alternatively, dust extinction maps may also prove to be useful in this context (see Alves et al., 2001 for the Bok globule B68; or Padoan, Cambresy, and Langer, 2002 for the Taurus molecular cloud). As CO line emission maps mostly trace the tenuous gas between dense cores, they miss the small-scale features and pick up the overall density structure, which is dominated by large-scale modes (Ossenkopf et al., 2001).

C. Support of molecular clouds

Molecular clouds are cold (Cernicharo, 1991). The kinetic temperature inferred from molecular line ratios is typically about 10 K for dark, quiescent clouds and dense cores in GMCs that are shielded from UV radiation by high column densities of dust, while it can reach 50–100 K in regions heated by UV radiation from high-mass stars. For example, the temperature of gas and dust behind the Trapezium cluster in Orion is about 50 K. In cold regions, the only heat sources are cosmic rays and dissipation of turbulence, while cooling comes from emission from dust and abundant molecular species. The thermal structure of the gas is related to its density distribution and its chemical abundance, so it is remarkable that over a wide range of gas densities and metallicities the equilibrium temperature remains almost constant in a small range around \( T \approx 10 \) K (Goldsmith and Langer, 1978; Goldsmith, 2001). In the absence of strong UV irradiation, the approximation of isothermality only breaks down when the cloud becomes dense enough to be opaque to cooling radiation, so that heat can no longer be radiated away efficiently. This occurs at gas density \( n(H_2) > 10^{10} \text{ cm}^{-3}. \) The equation of state then moves from isothermal, with polytropic exponent \( \gamma = 1, \) to adiabatic, with \( \gamma = 7/5 \) being appropriate for molecular hydrogen (see, for example, Tohline, 1982, and references therein).

Despite their low temperatures, the densities in molecular clouds are so high that their pressures exceed the average interstellar pressure by an order of magnitude or more. Typical interstellar pressures lie around \( 10^{-13} \text{ erg cm}^{-3} \) (Jenkins and Shaya, 1979; Bowyer et al., 1995), while at a temperature of 10 K and a density of \( 10^3 \text{ cm}^{-3}, \) the pressure in a typical molecular cloud exceeds \( 10^{-12} \text{ erg cm}^{-3}. \) Gravitational confinement was traditionally cited to explain the high pressures observed in GMCs (Kutner et al., 1977; Elmegreen, Lada, and Dickinson, 1979; Blitz, 1993; Williams et al., 2000). Their masses certainly exceed by orders of magnitude the critical mass for gravitational stability \( M_f \) defined by Eq. (14) in Sec. III, computed from their average density and temperature. However, if only thermal pressure opposed gravitational attraction, they should be collapsing and very efficiently forming stars on a free-fall time scale [Eq. (1)]. That is not the case. Within molecular clouds, low-mass gas clumps appear highly transient and pressure confined rather than being bound by self-gravity. Self-gravity appears to dominate only in the most massive individual cores, where star formation actually is observed (Williams, Blitz, and Stark, 1995; Yonekura et al., 1997; Kawamura et al., 1998; Simon et al., 2001).

In the short lifetimes of molecular clouds (Sec. VI.A) they likely never reach a state of dynamical equilibrium (Ballesteros-Paredes, Hartmann, and Vázquez-Semadeni, 1999; Elmegreen, 2000b). This is in contrast to the classical picture, which sees molecular clouds as long-lived equilibrium structures (Blitz and Shu, 1980). The overall star formation efficiency [Eq. (2)] on scales of molecular clouds as a whole is low in our Galaxy, of the order of 10% or smaller (Zuckerman and Evans, 1974). Only a small fraction of molecular cloud material associated with the highest-density regions is actually forming stars. The bulk of observed molecular cloud material is inactive, in a more tenous state between individual star-forming regions.

Except on the scales of isolated protostellar cores, the observed linewidths are always wider than implied by the excitation temperature of the molecules. This is interpreted as the result of bulk motion associated with turbulence. We shall argue in this review that it is this interstellar turbulence that determines the lifetime and fate of molecular clouds and so their ability to collapse and form stars.

Magnetic fields have long been discussed as a stabilizing agent in molecular clouds. However, magnetic fields with average field strength of 10 \( \mu \text{G} \) (Verschuur, 1995a, 1995b; Troland et al., 1996; Crutcher, 1999) cannot stabilize molecular clouds as a whole. This is particularly true on the scale of individual protostars, where magnetic fields appear too weak to impede gravitational collapse in essentially all cases observed (see Sec. III.D). Furthermore, magnetic fields cannot prevent turbulent velocity fields from decaying quickly (see the discussion in Sec. IV.A).

Molecular clouds appear to be transient features of the turbulent flow of the interstellar medium (Ballesteros-Paredes, Hartmann, and Vázquez-Semadeni, 1999). Just as Lyman-\( \alpha \) clouds in the intergalactic medium were shown to be transient objects formed in the larger-scale cosmological flow (Cen et al., 1994; Zhang, Anninos, and Norman, 1995), rather than stable objects in gravitational equilibrium (Ikeuchi, 1986; Rees, 1986), molecular clouds may never reach an equilibrium configuration. The high pressures seen in molecular clouds can be produced by ram pressure from converging supersonic flows in the ISM (see Sec. VI.A). So long as the flow persists, it confines the cloud and supplies turbulent energy. When the flow ends, the cloud begins to expand at its sound speed, eventually dissipating into the ISM (Vázquez-Semadeni, Shadmehri, and Ballesteros-Paredes, 2002). Further shocks may help this process along.

D. Scaling relations for molecular clouds

Observations of molecular clouds exhibit correlations between various properties, such as clump size, velocity...
dispersion, density, and mass. Larson (1981) first noted, using data from several different molecular cloud surveys, that the density $\rho$ and the velocity dispersion $\sigma$ appear to scale with the cloud size $R$ as

$$\rho \sim R^\alpha,$$  \hspace{1cm} (5)

$$\sigma \sim R^\beta,$$  \hspace{1cm} (6)

with $\alpha$ and $\beta$ being constant scaling exponents. Many studies have been done of the scaling properties of molecular clouds. The most commonly quoted values of the exponents are $\alpha \approx -1.15 \pm 0.15$ and $\beta \approx 0.4 \pm 0.1$ (Dame et al., 1986; Myers and Goodman, 1988; Falgarone et al., 1992; Fuller and Myers, 1992; Wood, Myers, and Daugherty, 1994; Caselli and Myers, 1995). However, the validity of these scaling relations is the subject of strong controversy, and significantly discrepant values have been reported by Carr (1987) and Loren (1989), for example.

The above standard values are often interpreted in terms of the virial theorem (Larson, 1981; Caselli and Myers, 1995). If one assumes virial equilibrium, Larson’s relations [Eqs. (5) and (6)] are not independent. For $\alpha = -1$, which implies constant column density, a value of $\beta = 0.5$ suggests equipartition between self-gravity and the turbulent velocity dispersion, such that the ratio between kinetic and potential energy is constant, with $E_{\text{kin}}/E_{\text{pot}} = \sigma^2 R/(2GM) \approx 1/2$. Note that, for any arbitrarily chosen value of the density scaling exponent $\alpha$, a corresponding value of $\beta$ obeying equipartition can always be found (Vázquez-Semadeni and Gazol, 1995). Equipartition is usually interpreted as indicating virial equilibrium in a static object. However, Ballesteros-Paredes, Vázquez-Semadeni, and Scalo (1999) pointed out that in a dynamic, turbulent environment, the other terms of the virial equation (McKee and Zweibel, 1992) can have values as large as, or larger than, the internal kinetic and potential energy. In particular, the changing shape of the cloud will change its moment of inertia, and turbulent flows will produce large fluxes of kinetic energy through the surface of the cloud. As a result, equipartition between internal kinetic and potential energy does not necessarily imply virial equilibrium.

Kegel (1989) and Scalo (1990) proposed that the density-size relation may be a mere artifact of the limited dynamic range in the observations, rather than reflecting a real property of interstellar clouds. In particular, in the case of molecular line data, the observations are restricted to column densities large enough for the tracer molecule to be shielded against photodissociating UV radiation, but small enough for the lines to remain optically thin. With limited integration times, most CO surveys tend to select objects in an even smaller range of column densities, giving roughly constant column density, which automatically implies $\rho \sim R^{-1}$. Surveys with longer integration times, and therefore larger dynamic ranges, seem to exhibit an increasingly large scatter in density-size plots, as seen, for example, in the data of Falgarone et al. (1992). Results from numerical simulations, which are free from observational bias, indicate the same trend (Vázquez-Semadeni, Ballesteros-Paredes, and Rodríguez, 1997). Three-dimensional simulations of supersonic turbulence (Mac Low, 1999) were used by Ballesteros-Paredes and Mac Low (2002) to perform a comparison of clumps measured in physical space to clumps observed in position-position-velocity space. They found no relation between density and size in physical space, but a clear trend of $\rho \sim R^{-1}$ in the simulated observations, caused simply by the tendency of clump-finding algorithms to pick out clumps with column densities close to the local peak values. Also, for clumps within molecular clouds, the structures identified in CO often do not correspond to those derived from higher-density tracers (see Langer et al., 1995, Bergin et al., 1997, Motte et al., 1998 for observational discussion, and Ballesteros-Paredes and Mac Low, 2002 for theoretical discussion). In summary, the existence of a physical density-size relation appears doubtful.

The velocity-size relation appears less prone to observational artifacts. Although some measurements of molecular clouds do not seem to exhibit this correlation (e.g., Loren, 1989; Plume et al., 1997), it does appear to be a real property of the cloud. It is often explained using the standard (though incomplete) argument of virial equilibrium. In supersonic turbulent flows, however, the scaling relation is a natural consequence of the characteristic energy spectrum $E(\nu) \sim \nu^{-2}$ in an ensemble of shocks, even in the complete absence of self-gravity (Ballesteros-Paredes and Mac Low, 2002; Boldyrev, Nordlund, and Padoan, 2002a, 2002b; Ossenkopf and Mac Low, 2002). Larger scales carry more energy, leading to a relation between velocity dispersion and size that empirically reproduces the observed relation. Thus, although the velocity-size relation probably does exist, its presence does not argue for virial equilibrium or even energy equipartition, but rather for the presence of a supersonic turbulent cascade.

E. Protostellar cores

1. From cores to stars

Protostellar cores are the direct precursors of stars. The transformation of cloud cores into stars can be conveniently subdivided into four observationally motivated phases (Shu et al., 1987; André et al., 2000).

(a) The prestellar phase describes the isothermal gravitational contraction of molecular cloud cores before the formation of the central protostar. Prestellar cores are cold and are best observed in molecular lines or dust emission. The isothermal collapse phase ends when the inner parts reach densities of $n(H_2) \approx 10^{14}$ cm$^{-3}$. Then the gas and dust become optically thick, so the heat generated by the collapse can no longer freely radiate away (Tohline, 1982). The central region begins to heat up, and contraction pauses. As the temperature increases to $T \approx 2000$ K, molecular hydrogen begins to dissociate, absorbing energy. The core becomes unstable again and collapse sets in anew. Most of the released gravitational energy goes into the dis-
sociation of H₂ so that the temperature rises only slowly. This situation is similar to the first isothermal collapse phase. When all molecules in the core are dissociated, the temperature rises sharply and pressure gradients again halt the collapse. This second hydrostatic object is the true protostar.

(b) The cloud core then enters the class 0 phase of evolution, in which the central protostar grows in mass by the accretion of infalling material from the outer parts of the original cloud core. Higher-angular-momentum material first falls onto a disk and then gets transported inwards by viscous processes. In this phase, star and disk are deeply embedded in an envelope of gas and dust. The mass of the envelope $M_{\text{env}}$ greatly exceeds the total mass $M_*$ of star and disk together. The main contribution to the total luminosity is accretion, and the system is best observed at submillimeter and infrared wavelengths.

(c) At later times, powerful protostellar outflows develop that clear out the envelope along the rotational axis. This is the class I phase, during which the system is observable in infrared and optical wavebands, and for which $M_{\text{env}} \ll M_*$. In optical light the central protostar is only visible when one looks along the outflow direction.

(d) In the class II phase, the envelope has disappeared, because all available gas has either been accreted or dispersed by the outflow. The protostar no longer accretes, and it enters the classical pre-main-sequence contraction phase. It still is surrounded by a tenous disk of gas and dust with a mass of order $10^{-3}$ that of the star. The disk adds an infrared excess to the spectral energy distribution of the system, which is dominated by the stellar Planck spectrum at visible wavelengths (Beckwith, 1999). This is the stage during which planets are believed to form (Lissauer, 1993; Ruden, 1999). Protostellar systems in this stage are commonly called T Tauri stars (Bertout, 1989). As time evolves further, the disk becomes more and more depleted until only a tenuous dusty disk of debris remains that is long lived and can last (i.e., continuously reform from collisions of planetesimals) into and throughout the stellar main-sequence phase (Zuckerman, 2001).

Detailed calculations of all phases of dynamical collapse assuming spherical symmetry are presented by Masunaga, Miyama, and Inutsuka (1998), Masunaga and Inutsuka (2000a, 2000b), Wuchterl and Klessen (2001), and Wuchterl and Tscharnutzer (2003).

2. Properties of protostellar cores

A number of small, dense molecular cores have been identified by low-angular-resolution, molecular line surveys of nearby dark clouds (Benson and Myers, 1989; Myers et al., 1991), as illustrated in Fig. 3. About half of them are associated with protostars, i.e., they are in the class 0 or class I phase of evolution, as inferred from the presence of low-luminosity sources observed by the Infrared Astronomical Satellite (IRAS) and CO outflows while the other half are observed to still be in their prestellar phase (Beichman et al., 1986; André et al., 2000).

One of the most notable properties of the sampled cores is their very narrow linewidths. These are very close to the linewidths expected for thermal broadening alone and, as a result, many of the cores appear approximately gravitationally virialized (Myers, 1983). They are thought either to be in the very early stage of gravitational collapse or to have subsonic turbulence supporting the clump. A comparison of the linewidths of cores with embedded protostellar objects (i.e., with associated IRAS sources) and the starless cores reveals a substantial difference. Typically, cores with infrared sources exhibit broader lines, which suggests the presence of a considerable turbulent component not present in starless cores. This may be caused by the central protostar's feeding back energy and momentum into its surrounding envelope. Molecular outflows associated with many of the sources may be a direct indication of this process.

The advent of a new generation of infrared detectors and powerful receivers in the radio and submillimeter wavebands in the late 1990s made it possible to determine the radial column density profiles of prestellar cores with high sensitivity and resolution (Ward-Thompson et al., 1994, 1999; André et al., 1996; Motte, André, and Neri, 1998; Bacmann et al., 2000; Motte and André, 2001). These studies show that starless cores typically have flat inner density profiles out to radii of a few hundredths of a parsec, followed by a radial decline of roughly $\rho \propto r^{-2}$ and possibly a sharp outer edge at radii 0.05–0.3 pc (André et al., 2000). This is illustrated in Fig. 4, which shows the observed column density of the starless core L1689B derived from combining mid-infrared absorption maps with 1.3-mm dust continuum emission maps (from Bacmann et al., 2000). Similar profiles have been derived independently from dust extinction studies (Lada et al., 1994; Alves et al., 2001). Protostellar cores often are elongated or cometary shaped and appear to be parts of filamentary structures that connect several objects.

The various theoretical approaches to explaining the observed core properties are discussed and compared in Sec. V.B.

F. The observed initial mass function

Hydrogen-burning stars can only exist in a finite mass range

$$0.08 \leq m \leq 100,$$

where the dimensionless mass $m=M/(1M_\odot)$. Objects with $m \leq 0.08$ do not have central temperatures and pressures high enough for hydrogen fusion to occur. If they are larger than about ten times the mass of Jupiter, $m > 0.01$, they are called brown dwarfs, or more generally substellar objects (Burrows et al., 1993; Laughlin and Bodenheimer, 1993; or for a review, Burrows et al.,...
Stars with $m > 100$, on the other hand, blow themselves apart by radiation pressure (Phillips, 1994).

It is complicated and laborious to estimate the initial mass function (IMF) in our Galaxy empirically. The first such determination from the solar neighborhood (Salthpeter, 1955) showed that the number $j(m)dm$ of stars with masses in the range $m$ to $m + dm$ can be approximated by a power-law relation,

$$j \sim m^{-\alpha} dm,$$

with index $\alpha \approx 2.35$ for stars in the mass range $0.4 \leq m \leq 10$. However, approximation of the IMF with a single power law is too simple. Miller and Scalo (1979) introduced a log-normal functional form, again to describe the IMF for Galactic field stars in the vicinity of the Sun,

$$\log_{10} j \sim \log_{10} m \sim \log_{10} \left( \frac{m}{m_0} \right),$$

with $m_0 = 0.23$, $\sigma = 0.42$, $A = 0.1$.

The IMF can also be estimated, probably more directly, by studying individual young star clusters. Typical examples are given in Fig. 5 (taken from Kroupa, 2002), which plots the mass function derived from star counts in the Trapezium Cluster in Orion (Hillenbrand and

---

**FIG. 3.** Intensity contours at half maximum of 16 dense cores in dark clouds, in the 1.3 cm $J$(1,1) lines of NH$_3$, in the 3.0 mm $J$=2→1 line of CS, and in the 2.7 mm $J$=1→0 line of C$^{18}$O. A linear scale of 0.2 pc is indicated in each individual map, and associated protostars are specified by a cross. From Myers et al., 1991.

**FIG. 4.** Radial column density profile of the prestellar core L1689B derived from combined infrared absorption and 1.3-mm continuum emission maps. Crosses show the observed values with the corresponding statistical errors, while the total uncertainties in the method are indicated by the dashed lines. For comparison, the solid line denotes the best-fitting Bonnor-Ebert sphere and the dotted line the column density profile of a singular isothermal sphere. The observed profile is well reproduced by an unstable Bonnor-Ebert sphere with a density contrast of $\sim 50$. See Bacmann et al. (2000) for further details.
FIG. 5. (Color in online edition) The measured stellar mass function $\xi$ as a function of logarithmic mass $\log_{10} m$ in the Orion nebular cluster (upper circles), the Pleiades (triangles connected by line), and the cluster M35 (lower circles). None of the mass functions is corrected for unresolved multiple stellar systems. The average initial stellar mass function derived from Galactic field stars in the solar neighborhood is shown as a line with the associated uncertainty range indicated by the hatched area. From Kroupa, 2002.

Carpenter, 2000), in the Pleiades (Hambly et al., 1999), and in the cluster M35 (Barrado y Navascués et al., 2001).

The most popular approach to approximating the IMF empirically is to use a multiple-component power law of the form of Eq. (8) with the following parameters (Scalo, 1998; Kroupa, 2002):

$$\xi(m) = \begin{cases} 
0.26 m^{-0.3} & \text{for } 0.01 \leq m < 0.08 \\
0.035 m^{-1.3} & \text{for } 0.08 \leq m < 0.5 \\
0.019 m^{-2.3} & \text{for } 0.5 \leq m < \infty.
\end{cases}$$

(11)

This representation of the IMF is statistically corrected for binary and multiple stellar systems too close to be resolved, but too far apart to be detected spectroscopically. Neglecting these systems overestimates the masses of stars, as well as reducing inferred stellar densities. These mass overestimates influence the derived stellar mass distribution, underestimating the number of low-mass stars. The IMF may steepen further towards high stellar masses and a fourth component could be defined with $\xi(m) = 0.019 m^{-2.7}$ for $m > 1.0$, thus arriving at the IMF proposed by Kroupa, Tout, and Gilmore (1993). In Eq. (11), the exponents for masses $m < 0.5$ are very uncertain due to the difficulty of detecting and determining the masses of very young low-mass stars. The exponent for 0.08 < $m$ < 0.5 could vary between $-0.7$ and $-1.8$, and the value in the substellar regime is even less certain.

There are some indications that the slope of the mass spectrum obtained from field stars may be slightly shallower than the one obtained from observing stellar clusters (Scalo, 1998). The reason for this difference is unknown. It is somehow surprising, given the fact most field stars appear to come from dissolved clusters (Adams and Myers, 2001). It is possible that the field star IMF is inaccurate because of incorrect assumptions about past star formation rates and age dependences for the stellar scale height. Both issues are either known or irrelevant for the IMF derived from cluster surveys. On the other hand, the cluster surveys could have failed to include low-mass stars due to extinction or crowding. It has also been claimed that the IMF may vary between different stellar clusters (Scalo, 1998), as the measured exponent $\alpha$ in each mass interval exhibits considerable scatter when comparing different star-forming regions. This is illustrated in Fig. 6, which is again taken from Kroupa (2002). This scatter, however, may be entirely due to effects related to the dynamical evolution of stellar clusters (Kroupa, 2001).

Despite these differences in detail, all IMF determinations share the same basic features, and it appears reasonable to say that the basic shape of the IMF is a universal property common to all star-forming regions in the present-day Galaxy, perhaps with some intrinsic scatter. There still may be some dependency on the metallicity of the star-forming gas, but changes in the IMF do not seem to be gross even in that case. There is no compelling evidence for qualitatively different behavior such as truncation at the low- or high-mass end.

III. HISTORICAL DEVELOPMENT

Stars form from gravitational contraction of gas and dust in molecular clouds. A first estimate of the stability of such a system against gravitational collapse can be made by simply considering its energy balance. For instability to occur, gravitational attraction must overcome the combined action of all dispersive or resistive forces. In the simplest case, the absolute value of the potential energy of a system in virial equilibrium is exactly twice the total kinetic energy, $E_{pot} + 2E_{kin} = 0$. If $E_{pot} + 2E_{kin} < 0$ the system collapses, while for $E_{pot} + 2E_{kin} > 0$ it expands. This estimate can easily be extended by including the surface terms and additional physical forces (see further discussion in Sec. II.D). In particular, taking magnetic fields into account may become important for describing interstellar clouds (Chandrasekhar and Fermi, 1953b; see also McKee et al., 1993, for a more recent discussion). In the presence of turbulence, the total ki-
the wave number \( k \) arrived a relation between the oscillation frequency and the gravitational potential 

\[
\omega^2 - c_s^2 k^2 + 4 \pi G \rho_0 = 0, \tag{12}
\]

where \( c_s \) is the isothermal sound speed, \( G \) the gravitational constant, and \( \rho_0 \) the initial mass density. The derivation neglects viscous effects and assumes that the linearized version of the Poisson equation describes only the relation between the perturbed potential and the perturbed density (neglecting the potential of the homogeneous solution, the so-called “Jeans swindle”; see, for example, Binney and Tremaine, 1997). The third term in Eq. (12) is responsible for the existence of decaying and growing modes, as pure sound waves stem from the dispersion relation \( \omega^2 - c_s^2 k^2 = 0 \). Perturbations are unstable against gravitational contraction if their wave number is below a critical value, the Jeans wave number \( k_J \), i.e., if

\[
k^2 < k_J^2 = \frac{4 \pi G \rho_0}{c_s^2}, \tag{13}
\]

or equivalently, if the wavelength of the perturbation exceeds a critical size given by \( \lambda_j = 2 \pi k_J^{-1} \). Assuming the perturbation is spherical with diameter \( \lambda_j \), this directly translates into a mass limit

\[
M_j = \frac{4 \pi}{3} \rho_0 \left( \frac{\lambda_j}{2} \right)^3 = \frac{\pi}{6} \left( \frac{\pi G}{c_s^3} \right)^{3/2} \rho_0^{-1/2} c_s^3. \tag{14}
\]

All perturbations exceeding the Jeans mass \( M_J \) will collapse under their own weight. For isothermal gas \( c_s^2 \propto T \), so \( M_J \propto \rho_0^{-1/2} T^{3/2} \). The critical mass \( M_J \) decreases when the density \( \rho_0 \) grows or when the temperature \( T \) sinks.

The Jeans instability has a simple physical interpretation in terms of the energy budget. The energy density of a sound wave is positive. However, its gravitational energy is negative, because the enhanced attraction in the compressed regions outweighs the reduced attraction in the dilated regions. The instability sets in at the wavelength \( \lambda_j \), where the net energy density becomes negative. The perturbation grows, allowing the energy to decrease further. For a fundamental derivation of this instability from the canonical ensemble in statistical physics, see Semelin, Sánchez, and de Vega (2001) and de Vega and Sánchez (2002a, 2002b). In isothermal gas, there is no mechanism that prevents complete collapse. In reality, however, during the collapse of molecular gas clumps, the opacity increases and at densities of \( n(H_2) \approx 10^{10} \text{ cm}^{-3} \) the equation of state becomes adiabatic rather than isothermal. Then collapse proceeds slower. Finally at very high central densities (\( \rho \approx 1 \text{ g cm}^{-3} \)) fusion sets in. This energy source leads to a new equilibrium (Tohline, 1982): a new star is born.

Attempts to include the effect of turbulent motions in the star formation process were already being made in the middle of the twentieth century by von Weizsäcker (1943, 1951) based on Heisenberg’s (1948a, 1948b) concept of turbulence. He also considered the production of interstellar clouds from the shocks and density fluctuations in compressible turbulence. A more quantitative theory was proposed by Chandrasekhar (1951a, 1951b), who investigated the effect of microturbulence in the subsonic regime. In this approach the scales of interest, e.g., for gravitational collapse, greatly exceed the outer scale of the turbulence. If turbulence is isotropic (and more or less incompressible), it contributes to the pressure on large scales, and Chandrasekhar derived a dispersion relation similar to Eq. (12) except for the introduction of an effective sound speed,

\[
c_{s, \text{eff}}^2 = c_s^2 + 1/3 (v^2), \tag{15}
\]
where \( \langle v^2 \rangle \) is the rms velocity dispersion due to turbulent motions.

Sasao (1973) noted that Chandrasekhar's derivation neglected the effect of the inertia of the turbulent flow in forming density enhancements while focusing only on the effective turbulent pressure. The developments through the mid-eighties are reviewed by Scalo (1986). Both Sasao (1973) and Chandrasekhar (1951a, 1951b) made the microturbulent assumption that the outer scale of the turbulence is smaller than that of the turbulent clouds. However, the outer scales of molecular cloud turbulence typically exceed, or are at least comparable to the size of the system (Ossenkopf and Mac Low, 2002), so the assumption of microturbulence is invalid. Bonazzola et al. (1987) therefore suggested a wavelength-dependent effective sound speed \( c_{\text{eff}}^2(k) = c_s^2 + 1/3 \rho v^2(k) \) for Eq. (12). In this description, the stability of the system depends not only on the total amount of energy, but also on the wavelength distribution of the energy, since \( v^2(k) \) depends on the turbulent power spectrum. A similar approach was also adopted by Vázquez-Semadeni and Gazol (1995), who added Larson's (1981) empirical scaling relations to the analysis.

An elaborate investigation of the stability of turbulent, self-gravitating gas was undertaken by Bonazzola et al. (1992), who used renormalization-group theory to derive a dispersion relation with a generalized, wave-number-dependent, effective sound speed and an effective kinetic viscosity that together account for turbulence at all wavelengths shorter than the one in question. According to their analysis, turbulence with a power spectrum steeper than \( P(k) \propto 1/k^\beta \) can support a region against collapse at large scales, and below the thermal Jeans scale, but not in between. On the other hand, they claim that turbulence with a shallower slope, as is expected for incompressible turbulence (Kolmogorov 1941a, 1941b), Burgers turbulence (Lesieur, 1997, p. 238), or shock-dominated flows (Passot, Pouquet, and Woodward, 1988), cannot support clouds against collapse at scales larger than the thermal Jeans wavelength. It may even be possible to describe the equilibrium state of self-gravitating gas as an inherently inhomogeneous thermodynamic critical point (de Vega, Sánchez, and Combes, 1996a, 1996b; de Vega and Sánchez, 2000). This may render all applications of incompressible turbulence to the theory of star formation meaningless. In fact, it is the main goal of this review to introduce and stress the importance of compressional effects in supersonic turbulence for determining the outcome of star formation.

In order to do that, we need to recapitulate the development of our understanding of the star formation process over the last few decades. We begin with the classical dynamical theory (Sec. III.A) and describe the problems that it encountered in its original form (Sec. III.B). In particular the time-scale problem led astrophysicists to think about the influence of magnetic fields. This line of reasoning resulted in the construction of the paradigm of magnetically mediated star formation, which we discuss in Sec. III.C. However, it became clear that this so-called “standard theory” had a variety of serious shortcomings (Sec. III.D). These led to the rejuvenation of the earlier dynamical concepts of star formation and their reconsideration in the modern framework of compressible supersonic turbulence, which we discuss in Sec. IV.

### A. Classical dynamical theory

The classical dynamical theory focuses on the interplay between self-gravity on the one side and pressure gradients on the other. Turbulence is taken into account, but only on microscopic scales significantly smaller than the collapse scales. In this microturbulent regime, random gas motions yield an isotropic pressure that can be absorbed into the equations of motion by defining an effective sound speed as in Eq. (15). The dynamical behavior of the system remains unchanged, and we do not distinguish between the effective and thermal sound speed \( c_s \) in this and the following two sections.

Because of the importance of gravitational instability for stellar birth, Jeans' (1902) pioneering work triggered numerous attempts to derive solutions to the collapse problem, both analytically and numerically. Particularly noteworthy are the studies by Bonnor (1956) and Ebert (1957), who independently derived analytical solutions for the equilibrium structure of spherical density perturbations in self-gravitating, isothermal, ideal gases, as well as a criterion for gravitational collapse. See Lombardi and Bertin (2001) for a recent analysis, and studies by Schmitz (1983, 1984, 1986, 1988) and Schmitz and Ebert (1986, 1987) for the treatment of rotation and generalized, polytropic equations of state. It has been argued recently that this may be a good description for the density distribution in quiescent molecular cloud cores just before they begin to collapse and form stars (Bacmann et al., 2000; Alves, Lada, and Lada, 2001). The first numerical calculations of protostellar collapse became possible in the late 1960s (e.g., Bodenheimer and Sweigart, 1968; Larson, 1969; Penston, 1969a, 1969b). They showed that gravitational contraction proceeds in a highly nonhomologous manner, contrary to what had previously been assumed (Hayashi, 1966).

This is illustrated in Fig. 7, which shows the radial density distribution of a protostellar core at various stages of the isothermal collapse phase. The gas sphere initially follows a Bonnor-Ebert critical density profile but has four times more mass than allowed in an equilibrium state. Therefore it is gravitationally unstable and begins to collapse. As the inner part has no pressure support, it falls freely. As matter moves inwards, the density in the interior grows, while the density decreases in the outer parts. This builds up pressure gradients in the outer parts, where contraction is retarded from free fall. In the interior, however, the collapse remains in approximate free fall. Thus it actually speeds up, because the free-fall time scale scales with density as \( \tau_{\text{ff}} \propto \rho^{-1/2} \). Changes in the density structure occur in a smaller and smaller region near the center and on shorter and shorter time scales, while practically nothing happens in the outer parts. As a result the overall matter distribu-
tion develops a strong central peak, approaching \( \rho \propto r^{-2} \). This is the density profile of an isothermal sphere. The establishment of a central singularity corresponds to the formation of a protostar that grows in mass by accreting the remaining envelope until the reservoir of gas is exhausted.

It was Larson (1969) who realized that the dynamical evolution in the initial isothermal collapse phase can be described by an analytical similarity solution. This was independently discovered by Penston (1969b) and later extended by Hunter (1977) into the regime after the protostar has formed. This so called Larson-Penston solution describes the isothermal collapse of homogeneous ideal gas spheres initially at rest. Its properties are summarized in Table II. This solution makes two important predictions. The first is the occurrence of supersonic infall velocities that extend over the entire collapsing core. Before the formation of the central protostar, the infall velocity tends towards \(-3.3c_s\), while afterwards it approaches free-fall collapse in the center with \( v \propto r^{-1/2} \), while still maintaining \( v = -3.3c_s \) in the outer envelope (Hunter, 1977). Second, the Larson-Penston solution predicts constant protostellar accretion rates \( M = 47c_s^3/G \).

In general, the dynamical models conceptually allow for time-varying protostellar mass accretion rates, if the gradient of the density profile of a collapsing cloud core varies with radius. In particular, if the core has a density profile with a flat inner region and then a decrease outwards, as is observed in low-mass cores (see Sec. III.D), then \( M \) has a high initial peak, while the flat core is accreted and later declines as the lower-density outer-envelope material is falling in (Ogino et al., 1999). The time evolution of \( M \) for the collapse of a sphere with a generalized Plummer (1911) profile, with parameters fit to the protostellar core L1544, is illustrated in Fig. 8 (see Whitworth and Ward-Thompson, 2001). Plummer-type spheres have flat inner density profiles with radius \( R_0 \) and density \( \rho_0 \) followed by an outer power-law decline,

\[
\rho(r) = \rho_0 \left( \frac{R_0}{R_0^2 + r^2} \right)^{\eta},
\]

where \( \eta = 5 \) is the classical Plummer sphere, while \( \eta = 4 \) is adopted by Whitworth and Ward-Thompson (2001) to reproduce observed cloud cores. Such a profile has the basic properties of a Larson-Penston sphere in mid-collapse.

The dynamical properties of the Larson-Penston solution set it clearly apart from the inside-out collapse model (Shu, 1977) derived for magnetically mediated star formation (Sec. III.C). One-dimensional numerical simulations of the dynamical collapse of homogeneous, isothermal spheres typically demonstrate global conver-

<table>
<thead>
<tr>
<th>TABLE II. Properties of the Larson-Penston solution of isothermal collapse.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Before core formation</strong></td>
</tr>
<tr>
<td>( t&lt;0 )</td>
</tr>
<tr>
<td>Density profile</td>
</tr>
<tr>
<td>( (r_0 \rightarrow 0 \text{ as } t \rightarrow 0^-) )</td>
</tr>
<tr>
<td>flattened isothermal sphere</td>
</tr>
<tr>
<td>Velocity profile</td>
</tr>
<tr>
<td>( v \approx -3.3c_s ), ( r \rightarrow \infty )</td>
</tr>
<tr>
<td>Accretion rate</td>
</tr>
</tbody>
</table>
and synchrotron emission (Davies and Shuter, 1963). Lower values from Zeeman measurements of H I (Troland and Heiles, 1986) and from measurements of pulsar rotation and dispersion measures (Rand and Kulkarni, 1989; Rand and Lyne, 1994) comparable to the modern value of around 3 \( \mu G \) only gradually became accepted over the next two decades. Even now, measurements of synchrotron emission leave open the possibility that there is a stronger disordered field in the Milky Way, although their interpretation depends critically on the assumption of equipartition between magnetic-field energy and other forms (Beck, 2001).

The presence of a field, especially one much stronger than 3 \( \mu G \), formed a major problem for the classical theory of star formation. To see why, let us consider the behavior of a field in a region of isothermal, gravitational collapse (Mestel and Spitzer, 1956; Spitzer, 1968). If we neglect all surface terms except thermal pressure \( P_0 \) (a questionable assumption, as shown by Ballesteros-Paredes, Vázquez-Semadeni, and Scalo, 1999, but the usual one at the time) and assume that the field \( \vec{B} \) is uniform and passes through a spherical region of average density \( \rho \) and radius \( R \), we can write the virial equation as (Spitzer, 1968)

\[
4 \pi R^3 P_0 = \frac{3}{5} \frac{M k_B T}{\mu} - \frac{1}{R} \frac{3}{5} GM^2 - \frac{1}{3} R^4 B^2, \tag{17}
\]

where \( M = (4/3) \pi R^3 \rho \) is the mass of the region, \( k_B \) is Boltzmann’s constant, \( T \) is the temperature of the region, and \( \mu \) is the mean mass per particle. So long as the ionization is sufficiently high for the field to be frozen to the matter, the flux through the cloud \( \Phi = \pi R^2 B \) must remain constant. Therefore the opposition to collapse due to magnetic energy given by the last term on the right-hand side of Eq. (17) will remain constant during collapse. If it cannot prevent collapse at the beginning, it remains unable to do so as the field is compressed.

If we write the radius \( R \) in terms of the mass and density of the region, we can rewrite the two terms in parentheses on the right-hand side of Eq. (17) to show that gravitational attraction can only overwhelm magnetic repulsion if

\[
M > M_{\sigma} = \frac{5^{3/2}}{48 \pi} \frac{B^3}{G^{5/2} \rho} = (4 \times 10^6 \, M_\odot) \left( \frac{n}{1 \, \text{cm}^{-3}} \right)^{-2} \left( \frac{B}{3 \, \mu G} \right)^3, \tag{18}
\]

where the numerical constant is correct for a uniform sphere, and the number density \( n \) is computed with mean mass per particle \( \mu = 2.11 \times 10^{-24} \, \text{g cm}^{-3} \). Mouschovias and Spitzer (1976) noted that the critical mass can also be written in terms of a critical mass-to-flux ratio

\[
\left( \frac{M}{\Phi} \right)_{\sigma} = \frac{3}{5 \pi} \left( \frac{5}{G} \right)^{1/2} = 490 \, \text{g G}^{-1} \text{cm}^{-2}, \tag{19}
\]
where the constant $\zeta = 0.53$ for uniform spheres (or flattened systems, as shown by Strittmatter, 1966) is used in the final equality. Assuming a constant mass-to-flux ratio in a region results in $\zeta = 0.3$ (Nakano and Nakamura, 1978). For a typical interstellar field of $3 \mu G$, the critical surface density for collapse is $7M_\odot \text{pc}^{-2}$, corresponding to a number density of $230 \text{cm}^{-3}$ in a layer of thickness $1 \text{pc} \approx 3.09 \times 10^{18} \text{cm}$. A cloud is termed subcritical if it is magnetostatically stable and supercritical if it is not.

The very large value for the magnetic critical mass in the diffuse ISM given by Eq. (18) forms a crucial objection to the classical theory of star formation. Even if such a large mass could be assembled, how could it fragment into objects with stellar masses of $0.01-100M_\odot$, when the critical mass should remain invariant under uniform spherical gravitational collapse?

Two further objections to the classical theory were also prominent. First was the embarrassingly high rate of star formation predicted by a model governed by gravitational instability, in which objects should collapse on roughly the free-fall time scale, Eq. (1), orders of magnitude shorter than the ages of typical galaxies.

Second was the gap between the angular momentum contained in a parcel of gas participating in rotation in a galactic disk and the much smaller angular momentum contained in stars (Spitzer, 1968; Bodenheimer, 1995). The disk of the Milky Way rotates with angular velocity $V$. The galactic disk and the much smaller angular momentum contained in a parcel of gas participating in rotation in a roughly the free-fall time scale, Eq. (1), orders of magnitude shorter than the ages of typical galaxies.

Ambipolar diffusion can solve the question of how magnetically supported gas can fragment if it allows neutral gas to gravitationally condense across field lines. The local density can then increase without also increasing the magnetic field, thus lowering the critical mass for gravitational collapse $M_c$ given by Eq. (18). This can also be interpreted as increasing the local mass-to-flux ratio towards the critical value given by Eq. (19).

The time scale $\tau_{\text{AD}}$ on which this occurs can be derived by considering the relative drift velocity of neutrals and ions $v_D = \vec{v}_i - \vec{v}_n$ under the influence of the magnetic field $\vec{B}$ (Spitzer, 1968). So long as the ionization fraction is small and we do not care about instabilities such as that found by Wardle (1990), the inertia and pressure of the ions may be neglected. The ion momentum equation then reduces in the steady-state case to a balance between Lorentz forces and ion-neutral drag,

$$\frac{1}{4\pi} (\nabla \times \vec{B}) \times \vec{B} = \alpha p_i \rho_n (\vec{v}_i - \vec{v}_n), \quad (20)$$

where the coupling coefficient is ($\alpha$ (Smith and Mac Low, 1997)

$$\alpha = (\sigma v)/(m_i + m_n) \approx 9.2 \times 10^{13} \text{ cm}^3 \text{s}^{-1} \text{g}^{-1}, \quad (21)$$

with $m_i$ and $m_n$ the mean mass per particle for the ions and neutrals, and $\rho_i$ and $\rho_n$ the ion and neutral densities. Typical values in molecular clouds are $m_i = 10m_{\text{H}}$ and $m_n = (7/3)m_{\text{H}}$. The value of $\alpha$ is roughly independent of the mean velocity, as the ion-neutral cross section $\sigma$ scales inversely with velocity in the regime of interest (Osterbrock, 1961; Draine, 1980). To estimate the typical time scale, consider drift occurring across a cylindrical region of radius $R$, with a typical bend in the field also of order $R$ so the Lorentz force can be estimated as roughly $B^2/4\pi R$. Then the ambipolar diffusion time scale can be derived by solving for $v_D$ in Eq. (20) to be

\[2\text{In plasma physics, the term ambipolar diffusion is applied to ions and electrons held together electrostatically rather than magnetically, while drifting together out of neutral gas.}\]
For ambipolar diffusion to solve the magnetic flux problem on an astrophysically relevant time scale, the ionization fraction \( x \) must be extremely small. With the direct observation of dense molecular gas (Palmer and Zuckerman, 1967; Zuckerman and Palmer, 1974) more than a decade after the original proposal by Mestel and Spitzer (1956), such low ionization fractions came to seem plausible. Nakano (1976, 1979) and Elmegreen (1979) computed the detailed ionization balance of molecular clouds for reasonable cosmic-ray ionization rates, showing that at densities greater than \( 10^4 \text{ cm}^{-3} \), the ionization fraction was roughly (Elmegreen, 1979)

\[
x \approx 5 \times 10^{-5} \left( \frac{n}{10^5 \text{ cm}^{-3}} \right)^{-1/2},
\]

(23)

becoming constant at densities higher than \( 10^7 \text{ cm}^{-3} \) or so. Below densities of \( 10^4 \text{ cm}^{-3} \), the ionization increases because of the external UV radiation field, and the gas is tightly coupled to the magnetic field.

With typical molecular cloud parameters \( \tau_{\text{AD}} \) is of order \( 10^7 \text{ yr} \) [Eq. (22)]. The ambipolar diffusion time scale \( \tau_{\text{AD}} \) is thus about 10–20 times longer than the corresponding dynamical time scale \( \tau_{\text{d}} \) of the system (McKee et al., 1993). The delay induced by waiting for ambipolar diffusion to occur was taken as a way to explain the low star formation rates observed in normal galaxies, as well as the long lifetimes of molecular clouds, which at that time were thought to be about 30–100 Myr (Blitz and Shu, 1980; Solomon et al., 1987). See Sec. VI.A, however, for arguments that they are under 10 Myr.

These considerations led to the investigation of star formation models based on ambipolar diffusion as a dominant physical process rather than relying solely on gas-dynamical collapse. In particular, Shu (1977) proposed the self-similar collapse of initially quasistatic singular isothermal spheres as the most likely description of the star formation process. He assumed that ambipolar diffusion in a magnetically subcritical, isothermal cloud core would lead to the buildup of a quasistatic \( 1/r^2 \)-density structure that contracts on time scales of the order of \( \tau_{\text{AD}} \). This evolutionary phase is denoted quasi-static because \( \tau_{\text{AD}} \gg \tau_{\text{d}} \). Ambipolar diffusion is supposed to lead eventually to the formation of a singularity in the central density, at which point the system becomes unstable and undergoes inside-out collapse. During collapse this model assumes that magnetic fields are no longer dynamically important and they are subsequently ignored in the original formulation of the theory. A rarefaction wave moves outward with the speed of sound, with the cloud material behind the wave falling freely onto the core and matter ahead still being at rest.

The Shu (1977) model predicts constant mass accretion onto the central protostar at a rate \( M \approx 0.975 c_s^3/\sqrt{G} \). This is significantly below the values derived for Larson-Penston collapse. In the latter case the entire system is collapsing dynamically and delivers mass to the center very efficiently, while in the former case in-fall mass transport is comparatively inefficient as the cloud envelope remains at rest until reached by the rarefaction wave. The density structure of the inside-out collapse, however, is essentially indistinguishable from the predictions of dynamical collapse. To observationally differentiate between the two models one needs to obtain kinematical data and determine the magnitude and spatial extent of infall motions with high accuracy. The basic predictions of inside-out collapse are summarized in Table III. As singular isothermal spheres by definition have infinite mass, the growth of the central protostar is taken to come to a halt when feedback processes (like bipolar outflows, stellar winds, etc.) become important and terminate further infall.

Largely within the framework of the standard theory, numerous analytical extensions to the original inside-out collapse model have been proposed. The stability of isothermal gas clouds with rotation, for example, has been investigated by Schmitz (1983, 1984, 1986), Terebey, Shu, and Cassen (1984), Schmitz and Ebert (1986, 1987), Inutsuka and Miyama (1992), Nakamura, Hanawa, and Nakano (1995), and Tsuribe and Inutsuka (1999b).

The effects of magnetic fields on the equilibrium structure of clouds and later during the collapse phase (where they have been neglected in the original inside-out scenario) are considered by Schmitz (1987), Baureis, Ebert, and Schmitz (1989), Tomisaka, Ikeuchi, and Nakamura (1988a, 1988b, 1989a, 1990b), Tomisaka (1991, 1995, 1996a, 1996b), Galli and Shu (1993a, 1993b), and Shu et al. (2000). The proposed picture is that ambipolar diffusion of initially subcritical cores that are threaded by uniform magnetic fields will lead to the buildup of disklke structures with constant mass-to-flux ratio. These disks are called isopedic. The mass-to-flux

\[\tau_{\text{AD}} = \frac{R}{v_D} = \frac{4 \pi \alpha \rho \rho_n R}{(\nabla \times B) \times B} \]

\[= \frac{4 \pi \alpha \rho \rho_n R^2}{B^2} \]

\[= (25 \text{ Myr}) \left( \frac{B}{3 \mu G} \right)^{-2} \left( \frac{n_n}{10^2 \text{ cm}^{-3}} \right)^2 \times \left( \frac{R}{1 \text{ pc}} \right)^2 \left( \frac{x}{10^{-6}} \right). \]
ratio increases steadily with time. As it exceeds the maximum value consistent with magnetostatic equilibrium, the entire core becomes supercritical and begins to collapse from the inside out with the mass-to-flux ratio assumed to remain approximately constant. It can be shown (Shu and Li, 1997) that, for isopedic disks, the forces due to magnetic tension are just a scaled version of the disk’s self-gravity with opposite sign (i.e., obstructing gravitational collapse), and that the magnetic pressure scales as the gas pressure (although the proportionality factor in general is spatially varying except in special cases). These findings allow the application of many results derived for unmagnetized disks to the magnetized regime, with only slight modifications to the equations. One application of this result is that for isopedic disks the derived mass accretion rate is just a scaled version of the original Shu (1977) rate, i.e., $M = (1 + H_0) c_s^2 / G$, with the dimensionless parameter $H_0$ depending on the effective mass-to-flux ratio.

However, the basic assumption of constant mass-to-flux ratio during the collapse phase appears inconsistent with detailed numerical calculations of ambipolar diffusion processes (see Sec. III.D.1). In these computations the mass-to-flux ratio in the central region increases more rapidly than in the outer parts of the cloud. This leads to a separation into a dynamically collapsing inner core with $(M/\Phi)_n > 1$ and an outer envelope with $(M/\Phi)_n < 1$ that is still held up by the magnetic field. The parameter $(M/\Phi)_n$ is the dimensionless mass-to-flux ratio normalized to the critical value given by Eq. (19). The isopedic description may therefore only be valid in the central region with $(M/\Phi)_n > 1$.

The presence of strong magnetic fields was suggested by Arons and Max (1975) as a way to explain the universally observed (Zuckerman and Palmer, 1974) presence of hypersonic random motions in molecular clouds. They noted that linear Alfvén waves have no dissipation associated with them, as they are purely transverse. In a cloud with Alfvén speed $v_A = B/(4 \pi \rho)^{1/2}$ much greater than the sound speed $c_s$, such Alfvén waves could produce the observed motions without necessarily forming strong shocks. This was generally, though incorrectly, interpreted to mean that these waves could therefore survive from the formation of the cloud, explaining the observations without reference to further energy input into the cloud. The actual work acknowledged that ambipolar diffusion would still dissipate these waves (Kulsrud and Pearce, 1969; Zweibel and Josafatsson, 1983) at a rate substantial enough to require energy input from a driving source to maintain the observed motions.

Strong magnetic fields furthermore provided a mechanism to reduce the angular momentum in collapsing molecular clouds through magnetic braking. Initially this was treated assuming that clouds were rigid rotating spheres (Ebert, von Hoerner, and Temesváry, 1960), but was accurately calculated by Mouschovias and Paleologou (1979, 1980) for both perpendicular and parallel cases. They showed that the criterion for braking to be effective was essentially that the outgoing helical Alfvén waves from the rotating cloud couple to a mass of gas equal to the mass in the cloud. Mouschovias and Paleologou (1980) show that this leads to a characteristic deceleration time for a parallel rotator of density $\rho$ and thickness $H$ embedded in a medium of density $\rho_0$ and Alfvén velocity $v_A = B/(4 \pi \rho_0)^{1/2}$ of

$$\tau_\parallel = (\rho/\rho_0)(H/2v_A)^{1/2},$$

and a characteristic time for a perpendicular rotator with radius $R$,

$$\tau_\perp = \frac{1}{2} \left[ \left( 1 + \frac{\rho}{\rho_0} \right)^{1/2} - 1 \right] \frac{R}{v_A}.$$

For typical molecular cloud parameters, these times can be less than the free-fall time, leading to efficient transfer of angular momentum away from collapsing cores. This may help to resolve the angular momentum problem in star formation described in Sec. III.B.

### D. Problems with standard theory

During the 1980s the theory of magnetically mediated star formation discussed in the previous section was widely advocated and generally accepted as the standard theory of low-mass star formation, almost completely replacing the earlier dynamical models. However, despite its success and intellectual beauty, the picture of magnetically mediated star formation occurring on time scales an order of magnitude longer than the free-fall time scale suffers from a series of severe observational and theoretical shortcomings. This became obvious in the 1990s with improved numerical simulations and the advent of powerful new observational techniques, especially at submillimeter and infrared wavelengths. Critical summaries are given by Whitworth et al. (1996) and Nakano (1998).

The standard theory introduces an artificial dichotomy to the star formation process. It suggests that low-mass stars form from low-mass, magnetically subcritical cores, whereas high-mass stars and stellar clusters form from magnetically supercritical cloud cores (Shu et al., 1987; Lizano and Shu, 1989). This distinction became necessary when it was understood that the formation of very massive stars or stellar clusters cannot be regulated by magnetic fields and ambipolar diffusion processes (see Sec. III.D.3). We shall argue in Sec. IV that this is true for low-mass stars also, and therefore that star formation is not mediated by magnetic fields on any scale, but instead is controlled by interstellar turbulence (Sec. IV.K). The new theory gives a unified description of both low-mass and high-mass star formation, thus removing the undesired artificial dichotomy introduced by the standard theory.

Before we introduce the new theory of star formation based on interstellar turbulence, we need to analyze in detail the properties and shortcomings of the theory we seek to replace. We begin with the theoretical considerations that make the inside-out collapse of quasistatic, singular, isothermal spheres unlikely to be an accurate description of stellar birth. We then discuss the disagreement of the theory with observations.
1. Singular isothermal spheres

The collapse of singular isothermal spheres is the astrophysically most unlikely and unstable member of a large family of self-similar solutions to the 1D collapse problem. Ever since the studies by Bonnor (1956) and Ebert (1957), and by Larson (1969) and Penston (1969a), much attention in the star formation community has been focused on finding astrophysically relevant, analytic, asymptotic solutions to this problem (see Sec. III.A). The standard solution derived by Shu (1977) considers evolution of initially singular, isothermal spheres as they leave equilibrium. His findings subsequently were extended by Hunter (1977, 1986). Whitworth and Summers (1985) demonstrated that all solutions to the isothermal collapse problem are members of a two-parameter family with the Larson-Penston-type solutions (collapse of spheres with uniform central density) and the Shu-type solutions (expansion-wave collapse of singular spheres) populating extreme ends of parameter space. The solution set has been extended to include a polytropic equation of state (Suto and Silk, 1988), shocks (Tsai and Hsu, 1995), and cylinder and disklike geometries (Inutsuka and Miyama, 1992; Nakamura, Hanawa, and Nakano, 1995). In addition, mathematical generalization using a Lagrangian formulation has been proposed by Henriksen (1989; see also Henriksen, André, and Bontemps, 1997).

Of all proposed initial configurations for protostellar collapse, quasistatic, singular, isothermal spheres seem to be the most difficult to realize in nature. Stable equilibria for self-gravitating, spherical, isothermal gas clouds embedded in an external medium of given pressure are only possible up to a density contrast of \( \rho_c/\rho_s \approx 14 \) between the cloud center and surface. More centrally concentrated clouds can only reach unstable equilibrium states. Hence all evolutionary paths that could yield a central singularity lead through instability, so collapse will set in long before a \( 1/r^2 \) density profile is established at small radii \( r \) (Whitworth et al., 1996; see also Silk and Suto, 1988, and Hanawa and Nakayama, 1997).

External perturbations also tend to break spherical symmetry in the innermost region and flatten the overall density profile at small radii. The resulting behavior in the central region then more closely resembles the Larson-Penston description of collapse. Similar behavior is found if outward propagating shocks are considered (Tsai and Hsu, 1995). As a consequence, the existence of physical processes that are able to produce singular, isothermal, equilibrium spheres in nature is highly questionable.

The original proposal of ambipolar diffusion processes in magnetostatically supported gas does not yield the desired result either. Ambipolar diffusion in magnetically supported gas clouds results in a dynamical Larson-Penston-type collapse of the central region where magnetic support is lost, while the outer part is still held up primarily by the field (and develops a \( 1/r^2 \) density profile). Mass is fed to the center not by an outward moving expansion wave, but by ambipolar diffusion in the outer envelope. The proposal that singular isothermal spheres may form through ambipolar diffusion processes in magnetically subcritical cores has been extensively studied by Mouschovias and collaborators in a series of numerical simulations with ever increasing accuracy and astrophysical detail. The numerical results indicate that the decoupling between matter and magnetic fields occurs over several orders of magnitude in density, becoming important at \( n(H_2) > 10^{10} \text{ cm}^{-3} \). There is no single critical density below which matter is fully coupled to the field and above which it is not, although ambipolar diffusion is indeed the dominant physical decoupling process (Desch and Mouschovias, 2001). As a consequence of ambipolar diffusion, initially subcritical gas clumps separate into a central nucleus that becomes both thermally and magnetically supercritical and an extended envelope that is still held up magnetostatically. The central region goes into rapid collapse, sweeping up much of its residual magnetic flux with it (Basu, 1997).

Star formation from singular isothermal spheres is also biased against binary formation. The collapse of rotating singular isothermal spheres very likely will result in the formation of single stars, as the central protostellar object forms very early and rapidly increases in mass with respect to a simultaneously forming and growing rotationally supported protostellar disk (Tsuribe and Inutsuka, 1999a, 1999b). By contrast, the collapse of cloud cores with flat inner density profiles will deliver a much smaller fraction of mass directly into the central protostar within a free-fall time. More matter will go first into a rotationally supported disklike structure. These disks tend to be more massive with respect to the central protostar in a Larson-Penston-type collapse. Compared to collapsing singular isothermal spheres, they are more likely to become unstable to subfragmentation, resulting in the formation of binary or higher-order stellar systems (see the review by Bodenheimer et al., 2000). Since the majority of stars seems to form as part of a binary or higher-order system (Mathieu et al., 2000), star formation in nature appears incompatible with collapse from initial conditions with a strong central peak (Whitworth et al., 1996).

2. Observations of clouds and cores

Before we consider the observational evidence against the standard theory of magnetically mediated star formation, let us recapitulate its basic predictions as introduced in Sec. III.C. The theory predicts (a) constant accretion rates and (b) infall motions that are confined to regions that have been passed by a rarefaction wave that

---

moves outwards with the speed of sound, while the parts of a core that lie further out remain static. The theory furthermore relies on (c) the presence of magnetic field strong enough to hold up the gas in molecular cloud cores sufficiently long, so it predicts that cores should be magnetically subcritical during most of their lifetimes. In the following we demonstrate that all these predictions appear to be contradicted by observations.

### a. Magnetic support

In his critical review of the standard theory of star formation, Nakano (1998) pointed out that no convincing magnetically subcritical core had been found up to that time. Similar conclusions still hold today. All magnetic-field measurements are consistent with cores' being magnetically supercritical or at most marginally critical.

When the theory was formulated in the late 1970s and 1980s, accurate measurements of magnetic-field strength in molecular clouds and cloud cores did not exist or were highly uncertain. Consequently magnetic fields in molecular clouds were essentially assumed to have the properties necessary for the theory to work and to circumvent the observational problems associated with the classical dynamical theory (Sec. III.B). In particular, the field was thought to have a strong fluctuating component associated with magnetohydrodynamic waves that give rise to the superthermal linewidths ubiquitously observed in molecular cloud material, as well as offering nonthermal support against self-gravity.

Even today, accurate determinations of magnetic-field strengths in molecular cloud cores are rare. Most field estimates rely on measuring the Zeeman splitting in molecular lines, typically OH, which is observationally challenging (Crutcher et al., 1993). The Zeeman effect has only been detected above the 3σ significance level in a few dozen clouds, while the number of nondetections or upper limits is considerably larger. For a compilation of field strengths in low-mass cores see Crutcher (1999) or more recent work by Bourke et al. (2001). Their basic results are summarized in Fig. 9, which plots the observed line-of-sight magnetic field $B_{\text{los}}$ against the column density $N(\text{H}_2)$ determined from CO measurements.

More detailed interferometric measurements may shift some of the data points closer to the critical value or even slightly into the subcritical regime (Crutcher, 2003), but Nakano’s (1998) objection still remains valid. There are no observations of strongly subcritical cores, and the inferred potential energies typically exceed, or at most are in approximate equipartition with, the magnetic energies. This means that even in the nominally subcritical cases dynamical collapse will set in quickly, as it requires very little ambipolar diffusion to reach a critical mass-to-flux ratio (Ciolek and Basu, 2001). If both nondetections and upper limits are taken into account, magnetic fields are on average too weak to prevent or significantly retard the gravitational collapse of cores. The basic assumption of the standard theory of magnetically mediated star formation therefore seems at odds with the observational facts.

This contradiction could, in principle, be weakened by making the extreme geometrical assumption that all cores are highly flattened, essentially sheetlike objects (Shu et al., 1999). Flux-to-mass ratios can then be derived that come closer to the critical value of equilibrium between magnetic pressure and gravity. But even for sheetlike cores, the average flux-to-mass ratio lies some-

![Fig. 9. Line-of-sight magnetic-field strength $B_{\text{los}}$ vs column density $N(\text{H}_2)$ for various molecular cloud cores. Squares are observations of Bourke et al. (2001) and circles are observations summarized by Crutcher (1999), Crutcher and Troland (2000), and Sarma et al. (2000). Large symbols represent clear detections of the Zeeman effect, whereas small ones are 3σ upper limits to the field strength. The lines drawn for the upper limits connect 3σ to 1σ limits. To guide the eye, lines of constant flux-to-mass ratio $(\Phi/M)_n$ are given, normalized to the critical value, i.e., to the inverse of Eq. (19). The observed line-of-sight component $B_{\text{los}}$ of the field is statistically deprojected to obtain the absolute value of the field $B$. The upper panel (a) assumes spherical core geometry, while the lower panel (b) assumes a sheetlike geometry. A value of $(\Phi/M)_n < 1$ corresponds to a magnetically supercritical core with magnetic-field strengths too weak to support against gravitational contraction, while $(\Phi/M)_n > 1$ allows magnetic support as required by the standard theory. Note that almost all observed cores are magnetically supercritical. This is evident when assuming spherical symmetry, but even in the case of sheetlike protostellar cores the average ratio is $(\langle \Phi/M \rangle)_n = 0.4$ when considering the 1σ upper limits. This is significantly lower than the critical value. The one clear exception is RCW57, but it has two velocity components, leaving its Zeeman value in some doubt. For further discussion see Bourke et al. (2001), from which this figure was reproduced.](image-url)
what below the critical value when taking all measurements into account, including the upper limits at the 1σ level (Bourke et al., 2001). In addition, highly flattened morphologies appear inconsistent with the observed density structure of cores. They typically appear as roundish objects (like the dark globule B68 described by Alves et al., 2001) and more likely are moderately prolate (with axis ratios of about 2:1) than highly oblate (with axis ratios ∼6:1) when statistically deprojected (Myers et al., 1991; Ryden, 1996; Curry, 2002; however, some authors do prefer the oblate interpretation—see Li and Shu, 1996; Jones, Basu, and Dubinski, 2001).

Bertoldi and McKee (1992) argued that the very massive clumps that form stellar clusters need to be magnetically supercritical. This conclusion was extended to low-mass cloud cores by Nakano (1998). He noted that clumps and cores in molecular clouds are generally observed as regions of significantly larger column density than the cloud as a whole (Benson and Myers, 1989; Tatematsu et al., 1993). If a core contained a strongly subcritical magnetic field, it would need to be confined by the mean cloud pressure or mean magnetic field in the cloud. Otherwise it would quickly expand and disappear. Calculations of the collapse of strongly subcritical cores such as those by Ciolek and Mouschovias (1994) fix the magnetic field at the outer boundaries, artificially confining the cloud. Under the assumption of virial equilibrium, typical values for the mean pressure and mean magnetic field in molecular clouds demand column densities in magnetostatic cores that are comparable to those in the ambient molecular cloud material. This contradicts the observed large contrast in column density between cloud cores and the rest of the cloud, giving additional evidence that most low-mass cores are also magnetically supercritical and collapsing.

b. Infall motions

Protostellar infall motions are observed on scales larger than, and with velocities greater than, predicted by the standard theory. One of the basic assumptions of the standard theory is the existence of a long-lasting quasistatic phase in protostellar evolution while ambipolar diffusion acts. Once ambipolar diffusion establishes the central singularity, a rarefaction wave expands transsonically. Gas beyond the rarefaction wave remains at rest. Therefore the theory predicts that protostellar cores (cloud cores without central protostars; see André et al., 2000 for a discussion) should show no signatures of infall motions, and that protostellar cores at later stages of evolution should exhibit collapse motions confined to their central regions. This can be tested by mapping molecular cloud cores at the same time in optically thin and thick lines. Inward motions can be inferred from asymmetry of optically thick lines, if the zero point of the velocity frame is determined from the optically thin lines. This procedure allows the separation of infall signatures from those of rotation and possible outflows (Myers et al., 1996).

One of the best studied examples is the apparently starless core L1544, which exhibits infall asymmetries (implying velocities up to 0.1 km s⁻¹) that are too extended (∼0.1 pc) to be consistent with inside-out collapse (Tafalla et al., 1998; Williams et al., 1999). Similar conclusions can be derived for a variety of other sources (see the review by Myers, Evans, and Ohashi, 2000, or the extended survey for infall motions in prestellar cores by Lee, Myers, and Tafalla, 1999, 2001). Typical observed contraction velocities in the prestellar phase are between 0.05 and 0.1 km s⁻¹, corresponding to mass infall rates ranging from a few 10⁻⁶ to a few 10⁻⁵ M⊙ yr⁻¹. The sizes of the infalling regions (e.g., as measured in CS) typically exceed the sizes of the corresponding cores as measured in high-density tracers like N₂H⁺ by a factor of 2–3. Even the dark globule B335, which was considered “a theorists dream” (Myers et al., 2000) and which was thought to match standard theory very well (Zhou et al., 1993) is also consistent with Larson-Penston collapse (Masunaga and Inutsuka, 2000a, 2000b) when analyzed using improved radiation transfer techniques but relying on single-dish data only. The core, however, exhibits considerable substructure and complexity (clumps, outflows, etc.) when observed with high spatial and spectral resolution using interferometry (Wilner et al., 2000). This raises questions about the applicability of any 1D, isothermal collapse model to real cloud cores.

Extended inward motions are a common feature in prestellar cores, and they appear to be a necessary ingredient for the formation of stars as predicted by dynamical theories (Secs. III.A and IV.K).

c. Density profiles

Observed prestellar cores have flat inner density profiles, as described in Sec. II.E. The basis of the Shu (1977) model is the singular isothermal sphere: the theory assumes radial density profiles ρ ∝ 1/r² at all radii r as starting conditions of protostellar collapse. High-resolution mapping of the density profiles of prestellar cores provides the most direct evidence against singular isothermal spheres as initial conditions of protostellar collapse.

d. Chemical ages

The chemical age of substructure in molecular clouds, as derived from observations of chemical abundances (also see Sec. VI.A), is much smaller than the ambipolar diffusion time. This poses a time-scale argument against magnetically regulated star formation. The comparison of multimolecule observations of cloud cores with time-dependent chemical models indicates typical ages of about 10^5 years (see the reviews by van Dishoeck et al., 1993; van Dishoeck and Blake, 1998; and Langer et al., 2000). This is orders of magnitude shorter than the time scales of up to 10^7 yr required for ambipolar diffusion to become important as required by the standard model.
3. Protostars and young stars

a. Accretion rates

Observed protostellar accretion rates decline with time, in contradiction to the constant rates predicted by the standard model. As matter falls onto the central protostar, the accretion luminosity $L_{\text{bol}}$ is a direct measure of the accretion rate as $\dot{M}$ of the object is a direct measure of the accretion rate as well as reasonable estimates of $M_*$. The fact that most of the matter first falls onto a protostellar disk, where it gets transported inwards on a viscous time scale before it is able to accrete onto the star, does not alter the expected overall luminosity by much (Hartmann, 1998).

During the early phases of protostellar collapse, while the mass $M_{\text{env}}$ of the infalling envelope exceeds the mass $M_*$ of the central protostar, the accretion luminosity $L_{\text{acc}}$ far exceeds the intrinsic luminosity $L_*$ of the young star. Hence the observed bolometric luminosity $L_{\text{bol}}$ of the object is a direct measure of the accretion rate as long as reasonable estimates of $M_*$ and $R_*$ can be obtained. Determinations of bolometric temperature $T_{\text{bol}}$ and luminosity $L_{\text{bol}}$ therefore should provide a fair estimate of the evolutionary stage of a protostellar core (Chen et al., 1995; Myers et al., 1998). Scenarios in which the accretion rate decreases with time and increases with total mass of the collapsing cloud fragment yield qualitatively better agreement with the observations than do models with constant accretion rate (André et al., 2000; see, however, Jayawardhana, Hartmann, and Calvet, 2001, for an alternative interpretation based on environmental conditions). A comparison of observational data with theoretical models in which $M_*$ decreases exponentially with time is shown in Fig. 10.

A closely related method for estimating the accretion rate $\dot{M}_*$ is determination of protostellar outflow strengths (Bontemps et al., 1996). Most embedded young protostars have powerful molecular outflows (Richer et al., 2000), while outflow strength decreases towards later evolutionary stages. At the end of the main accretion phase, the bolometric luminosity of protostars $L_{\text{bol}}$ strongly correlates with the momentum flux $F_{\text{CO}}$ (Cabrit and Bertout, 1992). Furthermore, $F_{\text{CO}}$ correlates well with $M_{\text{env}}$ for all protostellar cloud cores (Bontemps et al., 1996; Henning and Launhardt, 1998; Hogerheijde et al., 1998). This result is independent of the $F_{\text{CO}}^{-1}L_{\text{bol}}$ relation and most likely results from a progressive decrease of outflow power during the main accretion phase. With the linear correlation between outflow mass loss and protostellar accretion rate (Hartigan et al., 1995) these observations therefore suggest stellar accretion rates $\dot{M}_*$ that decrease with time. This is illustrated in Fig. 11, which compares the observed values of the normalized outflow flux and the normalized envelope mass for a sample of $\sim 40$ protostellar cores with a simplified dynamical collapse model with decreasing accretion rate $\dot{M}_*$ (Henriksen et al., 1997). The model describes the data relatively well, as opposed to models of constant $\dot{M}_*$.

![FIG. 10. Plot of envelope mass $M_{\text{env}}$ vs bolometric luminosity $L_{\text{bol}}$ for a sample of cores containing protostars in the main accretion phase with masses $M_*$, from Andrés and Montmerle (1994) and Saraceno et al. (1996). Open circles are objects with $M_{\text{env}}>M_*$ and filled circles are in the later evolutionary stage where $M_{\text{env}}<M_*$. The evolutionary tracks shown assume a bound initial configuration of finite mass that has $L_{\text{bol}}=GM_*/R_*$ with $L_*$ from Stahler (1988), and assume that both $M_{\text{env}}$ and $M_*=M_{\text{env}}/\tau (\tau=10^5 \text{ yr})$ decline exponentially with time (Bontemps et al., 1996, also Myers et al., 1998). Exponentially declining $\dot{M}_*$ shows better agreement with the data than do constant accretion rates. Small arrows are plotted on the tracks every $10^4 \text{ yr}$, and large arrows when $50$% and $90$% of the total mass is accreted onto the central protostar. The dashed and dotted lines indicate the transition from $M_{\text{env}}>M_*$ to $M_{\text{env}}<M_*$ using two different relations, $M_* \propto L_{\text{bol}}$ and $M_* \propto L_{\text{bol}}^2$, respectively, indicating the range proposed in the literature (André and Montmerle, 1994, or Bontemps et al., 1996). The latter relation is suggested by the accretion scenario adopted in the tracks. The figure is adapted from André et al. (2000).]
that the timespan over which cores increase their central density \( n(\text{H}_2) \) from \( \sim 10^4 \) to \( \sim 10^7 \) cm\(^{-3} \) is about the same as that from \( n(\text{H}_2) \approx 10^5 \) cm\(^{-3} \) to the formation of the central protostar. This clearly disagrees with standard ambipolar diffusion models (Ciolek and Mouschovias, 1994), which predict a duration six times longer. Ciolek and Basu (2000) were indeed able to accurately model infall in L1544 using an ambipolar diffusion model, but they did so by using initial conditions that were already almost supercritical, so that very little ambipolar diffusion had to occur before dynamical collapse would set in. Ciolek and Basu (2001) quantify the central density required to match the observations and conclude that observed prestellar cores are either already supercritical or just about to be. These observations reveal that already in the prestellar phase the time scales of core contraction are determined by fast dynamical processes rather than by slow ambipolar diffusion.

c. Stellar ages

If the contraction time of individual cloud cores in the prestellar phase is determined by ambipolar diffusion, then the age spread in a newly formed group or cluster should considerably exceed the dynamical time scale. Within a star-forming region, high-density protostellar cores will evolve and form central protostars faster than their low-density counterparts, so the age distribution is roughly determined by the evolution time of the lowest-density condensation.

However, the observed age spread in star clusters is very short. For example, in the Orion Trapezium cluster, a dense cluster of a few thousand stars, it is less than 10\(^6\) years (Prosser et al., 1994; Hillenbrand, 1997; Hillenbrand and Hartmann, 1998), and the same holds for L1641 (Hodapp and Deane, 1993). These age spreads are comparable to the dynamical time in these clusters. Similar conclusions can be obtained for Taurus (Hartmann, 2001), NGC1333 (Bally et al., 1996; Lada et al., 1996), NGC6531 (Forbes, 1996), and a variety of other clusters (see Palia and Stahler, 1999; Elmegreen et al., 2000; Hartmann, 2001).

Larger regions form stars for a longer time. This correlation suggests that typical star formation times correspond to about two to three turbulent crossing times in that region (Efremov and Elmegreen, 1998a). This is very fast compared to the ambipolar diffusion time scale, which is about ten crossing times in a uniform medium with cosmic-ray ionization (Shu et al., 1987) and is even longer if stellar UV sources contribute to the ionization (Myers and Kheronsky, 1995) or if the cloud is very clumpy (Elmegreen and Combes, 1992). Magnetic fields therefore appear unable to regulate star formation on the scales of stellar clusters.

IV. TOWARD A NEW PARADIGM

In this section we suggest that self-gravity acting in a supersonic, turbulent flow can lead to behavior consistent with observations of star formation, returning to the pioneering ideas of Larson (1981) in a more quantitative fashion. We first examine in Sec. IV.A whether magnetic fields can maintain the supersonic motions observed in molecular clouds. We then study the behavior of self-gravitating, turbulent gas, beginning in Sec. IV.B. We conclude, in Sec. IV.G, that turbulence often inhibits collapse without preventing it entirely, and we discuss the relation between turbulence and star formation in the subsequent subsections. Finally, we outline our conclusions in Sec. IV.K.

A. Maintenance of supersonic motions

We first consider the question of how to maintain the observed supersonic motions. As described above in Sec. III.C, magnetohydrodynamic waves were once generally thought to provide the means to prevent the dissipation of interstellar turbulence. However, numerical models have now shown that they do not. One-dimensional simulations of decaying, compressible, isothermal, magnetized turbulence by Gammie and Ostriker (1996) showed quick decay of kinetic energy \( E_{\text{kin}} \) in the absence of driving, but found that the quantitative decay rate depended strongly on initial and boundary conditions because of the low dimensionality. Mac Low et al. (1998), Stone, Ostriker, and Gammie
from large scales (Sec. II.B), while the effective value of $\lambda_J$ is not entirely clear in a strongly inhomogeneous medium. If $\kappa < 3 - 4$, then undriven turbulence will decay in a free-fall time. As we discussed in Sec. VI.A, the observational evidence suggests that clouds are a few free-fall times old, on average, though perhaps not more than two or three, so there may be continuing energy input into the clouds. This energy input may come from the same compressive motions that have been suggested to form the clouds (Ballesteros-Paredes, Hartmann, and Vázquez-Semadeni, 1999).

**B. Turbulence in self-gravitating gas**

This leads to the question of what effect supersonic turbulence has on self-gravitating clouds. Can turbulence alone delay gravitational collapse beyond a free-fall time? In Secs. I.B and III.A, we summarized analytic approaches to this question and pointed out that, aside from Sasao (1973), they were all based on the assumption that the turbulent flow is close to incompressible. Sasao (1973) concluded that compressible turbulence with a Kolmogorov spectrum would show collapse at roughly the Jeans scale.

Numerical models of highly compressible, self-gravitating turbulence have shown the importance of density fluctuations generated by the turbulence to understanding support against gravity. Early models by Bonazolla et al. (1987), Passot et al. (1988), and Léorat, Passot, and Pouquet (1990) used low-resolution calculations ($32^2 - 64^2$ collocation points) with a two-dimensional spectral code to support their analytical results. The gas-dynamical studies by Vázquez-Semadeni, Passot, and Pouquet (1995) and Ballesteros-Paredes, Vázquez-Semadeni, and Scalo (1999) were also restricted to two dimensions, and were focused on the ISM at kiloparsec scales rather than on molecular clouds, although they were performed with far higher resolution (up to $800 \times 800$ points). Magnetic fields were introduced in these models by Passot, Vázquez-Semadeni, and Pouquet (1995) and extended to three dimensions with self-gravity (though at only $64^3$ resolution) by Vázquez-Semadeni, Passot, and Pouquet (1996). One-dimensional computations focused on molecular clouds, including both MHD and self-gravity, were presented by Gammie and Ostriker (1996) and Balsara, Crutcher, and Pouquet (2001). Ostriker, Gammie, and Stone (1999) extended their work to 2.5 dimensions.

These early methods at low resolution, low dimension, or both, suggested two conclusions important for this review. First, gravitational collapse, even in the presence of magnetic fields, does not generate sufficient turbulence to markedly slow further collapse. Second, turbulent support against gravitational collapse may act globally, while still allowing local collapse. More recent three-dimensional, high-resolution computations by Klessen et al. (1998, 2000), Klessen (2000), Klessen and Burkert (2000, 2001), and Heitsch, Mac Low, and...
Klessen (2001) have now confirmed and extended these earlier results. In the following, we use these calculations to draw consequences for the theory of star formation.

C. A numerical approach

Klessen et al. (2000) and Heitsch, Mac Low, and Klessen (2001) applied two different numerical methods: ZEUS-3D, an Eulerian MHD code (Stone and Norman, 1992a, 1992b; Clarke and Norman, 1994; Hawley and Stone, 1995), and an implementation of smoothed particle hydrodynamics (SPH), a Lagrangian hydrodynamics method using particles as an unstructured grid (Benz, 1990; Monaghan, 1992). Klessen et al. (1998), Klessen (2000), and Klessen and Burkert (2000, 2001) used only SPH computations.

SPH can resolve very high density contrasts because it increases the particle concentration in regions of high density and thus the effective spatial resolution, making it well suited for computing collapse problems. By the same token, though, it resolves low-density regions poorly. Shock structures tend to be broadened by the averaging kernel in the absence of adaptive techniques. The correct numerical treatment of gravitational collapse requires the resolution of the local Jeans mass at every stage of the collapse (Bate and Burkert, 1997). In the models described here, once an object with density beyond the resolution limit of the code has formed in the center of a collapsing gas clump, it is replaced by a “sink” particle (Bate, Bonnell, and Price, 1995). Replacing high-density cores and keeping track of their further evolution in a consistent way prevents the time step from becoming prohibitively small. This allows modeling of the collapse of a large number of cores until the over-

FIG. 13. Density cubes in a model of supersonic turbulence driven on intermediate scales, with wave numbers $k = 3 - 4$. Four different evolutionary stages are shown: (a) just before gravity is turned on; (b) when the first collapsed cores are formed and have accreted $M_*=5\%$ of the mass; (c) when the mass in dense cores is $M_*=25\%$; and (d) when $M_*=50\%$. Time is measured in units of the global system free-fall time scale $t_{ff}$, and dark dots indicate the locations of protostars. From Klessen et al., 2000.
all gas reservoir becomes exhausted.

ZEUS-3D, conversely, gives equal resolution in all regions, resolving shocks equally well everywhere, as well as allowing the inclusion of magnetic fields (see Sec. IV.F). On the other hand, collapsing regions cannot be followed to scales less than one or two grid zones. Once again the resolution required to follow gravitational collapse must be considered. For a grid-based simulation, the criterion given by Truelove et al. (1997) holds. Equivalent to the SPH resolution criterion, the mass contained in one grid zone has to be rather smaller than the local Jeans mass throughout the computation. In the models described here, this criterion is satisfied until gravitational collapse is underway.

The computations discussed here are done on periodic cubes, with an isothermal equation of state, using up to 256^3 zones (with one model at 512^3 zones) or 80^3 SPH particles. To generate turbulent flows, Gaussian velocity fluctuations are introduced with power only in a narrow interval $k - 1 \leq |\vec{k}| \leq k$, where $k = L / \lambda_J$ counts the number of driving wavelengths $\lambda_J$ in the box (Mac Low et al., 1998). This offers a simple approximation to driving by mechanisms that act on a single scale. To drive the turbulence, this fixed pattern is normalized to maintain constant kinetic-energy input rate $\dot{E}_\text{kin} = \Delta E / \Delta t$ (Mac Low, 1999). Self-gravity is turned on only after the turbulence reaches a state of dynamical equilibrium.

D. Global collapse

First, we examine the question of whether gravitational collapse can itself generate enough turbulence to prevent further collapse. Gas-dynamical SPH models initialized at rest with Gaussian density perturbations show fast collapse, with the first collapsed objects forming in a single free-fall time (Klessen, Burkert, and Bate, 1998; Klessen and Burkert, 2000, 2001; Bate et al., 2002a, 2002b). Models set up with a freely decaying turbulent velocity field behave similarly (Klessen, 2000). Further accretion of gas onto collapsed objects then occurs over the next free-fall time, defining the predicted spread of stellar ages in a freely collapsing system. The turbulence generated by collapse (or virialization) fails to prevent further collapse, contrary to previous suggestions (e.g., by Elmegreen, 1993). Such a mechanism only works for thermal pressure support in systems such as galaxy cluster halos, where energy is lost inefficiently, while turbulence dissipates energy quite efficiently [Eq. (26)].

Models of freely collapsing, magnetized gas remain to be done, but models of self-gravitating, decaying, magnetized turbulence by Balsara, Ward-Thompson, and Crutcher (2001) using an MHD code incorporating a Riemann solver suggest that the presence of magnetic fields does not markedly extend collapse time scales. They further show that accretion down filaments aligned with magnetic field lines onto cores occurs readily. This allows high mass-to-flux ratios to be maintained even at small scales, which is necessary for supercritical collapse to continue in a magnetized medium after fragmentation occurs.

E. Local collapse in globally stable regions

Second, we examine whether continuously driven turbulence can support against gravitational collapse. The models of driven, self-gravitating turbulence by Klessen et al. (2000) and Heitsch, Mac Low, and Klessen (2001) described in Sec. IV.C show that local collapse occurs even when the turbulent velocity field carries enough energy to counterbalance gravitational contraction on global scales. This idea was first suggested by Hunter (1979), who used the virial theorem to make his case. Later support for it was offered by 2D gas-dynamical computations by Léorat et al. (1990). An example of local collapse in a globally supported cloud is given in Fig. 13. A hallmark of global turbulent support is isolated, inefficient, local collapse.

Local collapse in a globally stabilized cloud is not predicted by any of the analytic models for turbulent, self-gravitating gas, as discussed by Klessen et al. (2000). The resolution to this apparent paradox lies in the requirement that any substantial turbulent support come from supersonic flows, as otherwise pressure support would be at least equally important. However, supersonic flows compress the gas in shocks. In isothermal gas with density $\rho$ the postshock gas has density $\rho' = M^2 \rho$, where $M$ is the Mach number of the shock. The turbulent Jeans
length $\lambda_d \approx \rho^{-1/2}$ in these density enhancements, so it drops by a factor of $\mathcal{M}$ in isothermal shocks.

Klessen et al. (2000) demonstrated that supersonic turbulence can completely prevent collapse only when it can support not just the average density, but also the shocked, high-density regions, as shown in Fig. 14. This basic point was earlier made by Elmegreen (1993) and Vázquez-Semadeni et al. (1995). Two criteria must be fulfilled in these regions. The rms velocity must be high enough, and the driving wavelength $\lambda_d = \lambda_J(\rho')$ small enough. If these two criteria are not met, the localized high-density regions collapse, although the surrounding flow remains turbulently supported.

The length scale and strength of energy injection into the system determine the structure of the turbulent flow and therefore the locations at which stars are most likely to form. Large-scale driving leads to large coherent shock structures [Fig. 15(a)]. Local collapse occurs predominantly in these filaments and layers of shocked gas (Klessen et al., 2000). Increasing the driving scale produces larger and more massive structures that can become gravitationally unstable. Hence the star formation efficiency [Eq. (2)] increases. The same is true for weaker driving. Reducing the turbulent kinetic energy means that more and larger volumes exceed the Jeans criterion for gravitational instability. The more massive the unstable region is, the more stars it will form. Dense clusters or associations of stars build up, either in the complete absence of energy input, or when small-scale turbulence is too weak to support large volumes, or when large-scale turbulence sweeps up large masses of gas that collapse. In all of these cases star formation is highly efficient and proceeds on a free-fall time scale (Klessen et al., 1998; Klessen and Burkert, 2000, 2001).

These points are illustrated in Fig. 16, which compares the distribution of protostars in a model of freely decaying turbulence to the distribution in a model of turbulence driven both at large scale and strongly enough to formally support against collapse. Both scenarios lead to star formation in aggregates and clusters. However, the figure suggests a possible way to distinguish between the two. Decaying turbulence typically results in a bound cluster of stars, while stellar aggregates associated with large-scale, coherent, shock fronts often have supervirial velocity dispersions that result in their quick dispersal. The numerical models discussed here do not include feedback from the newly formed stars, however. Ionization and outflows may retard or impede further star formation or gas accretion, possibly preventing a bound cluster from forming even in the case of freely decaying turbulence.

The efficiency of collapse depends on the properties of the supporting turbulence. It could be regulated by the amount of gas available for collapse on scales where turbulence turns from supersonic to subsonic (Vázquez-Semadeni, Ballesteros-Paredes, and Klessen, 2003). Sufficiently strong driving on short enough scales can prevent local collapse for arbitrarily long periods of time, but such strong driving may be rather difficult to arrange in a real molecular cloud. If we assume that stellar driving sources have an effective wavelength close to their separation, then the condition that driving act on scales smaller then the Jeans wavelength in typical shock-generated gas clumps requires the presence of an extraordinarily large number of stars evenly distributed throughout the cloud, with typical separation 0.1 pc in Taurus, or only 350 AU$^4$ in Orion. This is not observed.

$^4$One astronomical unit is the mean radius of Earth’s orbit around the Sun, 1 AU$=1.5\times10^{13}$ cm.
Very small driving scales seem to be at odds with the observed large-scale velocity fields in at least some molecular clouds (Ossenkopf and Mac Low, 2002).

Triggering of star formation by compression has been discussed at least since the work of Elmegreen and Lada (1977) on star formation in the gas swept up by expanding H II regions. The turbulent compressions described here do indeed trigger local star formation in globally supported regions. However, in the absence of the flow, global collapse would cause more vigorous star formation. Indeed, Elmegreen and Lada (1977) noted themselves that the time for gravitational instability to occur in the shocked, compressed layer was actually somewhat longer than the Jeans time in the undisturbed cloud. We discuss this issue further in Sec. VI.D.2.

The domination of cloud structure by large-scale modes leads to the formation of stars in groups and clusters. When stars form in groups, their velocities initially reflect the turbulent velocity field of the gas from which they formed. However, as more and more mass accumulates in protostars, their mutual gravitational interaction becomes increasingly important, beginning to determine the dynamical state of the system, which then behaves more and more like a collisional N-body system, where close encounters occur frequently (see Sec. V.D).

![Projected positions of protostars in models of turbulence driven at large scale (model B1 with driving wave number $k = 1 - 2$) and freely decaying turbulence at two different times. The upper panels show projections into (a) the $x$-$y$ plane and (b) the $x$-$z$ plane at an evolutionary stage when $M_\ast \approx 20\%$ of the total mass has been accreted. The gravitational potential is still dominated by nonaccreted gas, and the two models are statistically indistinguishable. In both cases protostars form in a group. However, only decaying turbulence leads to a bound cluster. In the model of driven turbulence, the group disperses quickly, and stars end up widely distributed throughout the computational volume. This is illustrated in the lower panels (c) and (d), which show the same projections of the system at later times when $M_\ast \approx 65\%$. From Klessen et al., 2000.](image)
F. Effects of magnetic fields

So far, we have concentrated on the effects of purely gas-dynamical turbulence. How does the picture discussed here change if we consider the presence of magnetic fields? Magnetic fields have been suggested to support molecular clouds well enough to prevent gravitationally unstable regions from collapsing (McKee, 1999), either magnetostatically or dynamically through MHD waves.

Assuming ideal MHD, a self-gravitating cloud of mass $M$ permeated by a uniform flux $\Phi$ is stable unless the mass-to-flux ratio exceeds the value given by Eq. (19). Without any other mechanism of support, such as turbulence acting along the field lines, a magnetostatically supported cloud collapses to a sheet, which is then supported against further collapse. Fiege and Pudritz (1999) found an equilibrium configuration of helical field that could support a filament, rather than a sheet, from fragmenting and collapsing. Such configurations do not appear in numerical models of turbulent molecular clouds, however, suggesting that reaching this stable equilibrium is difficult.

Investigation of support by MHD waves concentrates mostly on the effect of Alfvén waves, as they (i) are not as subject to damping as magnetosonic waves and (ii) can exert a force along the mean field, as shown by Dewar (1970) and Shu et al. (1987). This is because Alfvén waves are transverse waves, so they cause perturbations $\delta B$ perpendicular to the mean magnetic field $B$. McKee and Zweibel (1995) argue that Alfvén waves can even lead to an isotropic pressure, assuming that the waves are neither damped nor driven. However, in order to support a region against self-gravity, the waves would have to propagate outwardly, because inwardly propagating waves would only further compress the cloud. Thus this mechanism requires a negative radial gradient in wave sources in the cloud (Shu et al., 1987).

It can be demonstrated (Heitsch, Mac Low, and Klessen, 2001) that supersonic turbulence does not cause a magnetostatically supported region to collapse, and vice versa, that in the absence of magnetostatic support, MHD waves cannot completely prevent collapse, although they can retard it to some degree. The case of a subcritical region with $M < M_{\text{cr}}$ is illustrated in Fig. 17. Indeed, sheets form, roughly perpendicular to the field lines. This is because the turbulent driving can shift the sheets along the field lines without changing the mass-to-flux ratio. The sheets do not collapse further, because the shock waves cannot sweep gas across field lines, and the entire region is initially supported magnetostatically.

A supercritical cloud with $M > M_{\text{cr}}$ could only be stabilized by MHD wave pressure. This is insufficient to completely prevent gravitational collapse, as shown in Fig. 18. Collapse occurs in all models of unmagnetized and magnetized turbulence regardless of the numerical resolution and magnetic field strength as long as the system is magnetically supercritical. This is shown quantitatively in Fig. 19. Increasing the resolution makes itself felt in different ways in gas-dynamical and MHD models. In the gas-dynamical case, higher resolution results in thinner shocks and thus higher peak densities. These higher density peaks form cores with deeper potential wells that accrete more mass and are more stable against disruption. Higher resolution in the MHD models, on the other hand, better resolves short-wavelength MHD waves, which apparently can delay collapse, but not prevent it. This result extends to models with 512$^3$ zones.

![FIG. 17. Two-dimensional slice through a cube of magnetostatically supported, self-gravitating turbulence driven at large scale. The upper panel shows velocity vectors and the lower panel magnetic field vectors. The initial magnetic field is along the z direction, i.e., vertically oriented in all plots presented. The field is strong enough in this case not only to prevent the cloud from collapsing perpendicular to the field lines, but even to suppress turbulent motions in the cloud. The turbulence barely affects the mean field. The density grayscale is given in the colorbar, in model units. The time shown is $t = 5.5 t_{\text{ff}}$. From Heitsch, Mac Low, and Klessen, 2001.](image-url)
The delay of local collapse seen in our magnetized simulations is caused mainly by weakly magnetized turbulence acting to decrease density enhancements due to shock interactions.

**G. Promotion and prevention of local collapse**

Highly compressible turbulence both promotes and prevents collapse. Its net effect is to inhibit collapse globally, while perhaps promoting it locally. This can be seen by examining the dependence of the Jeans mass $M_J \propto \rho^{-1/2} c_s^2$, Eq. (14), on the rms turbulent velocity $v_{\text{rms}}$. If we follow the classical picture that treats turbulence as an additional pressure (Chandrasekhar, 1951a, 1951b), then we define $c_{s,\text{eff}} = c_s + v_{\text{rms}}^2/3$, giving the Jeans mass a dependence on velocity of $v_{\text{rms}}^3$. However, compressible turbulence in an isothermal medium causes local density enhancements that increase the density by $\mathcal{M}^2 v_{\text{rms}}^2$, adding a dependence $1/v_{\text{rms}}^2$. Combining these two effects, we find that

$$M_J \propto v_{\text{rms}}^2$$

for $v_{\text{rms}} \gg c_s$, so that ultimately turbulence does inhibit collapse. However, there is a broad intermediate region, especially for long-wavelength driving, where local collapse occurs despite global support, as shown in Fig. 15, which can be directly compared with Fig. 13(b).

The total mass and lifetime of a Jeans-unstable fluctuation determine whether it will actually collapse. Roughly speaking, the lifetime of an unstable clump is determined by the interval between two successive passing shocks: the first creates it, while the second one, if strong enough, may disrupt the clump again (Klein, McKee, and Colella, 1994; Mac Low et al., 1994). If the time interval between the two shocks is sufficiently long, however, an unstable clump can contract to high enough densities to effectively decouple from the ambient gas.
flow and becomes able to survive the encounter with further shock fronts (Krebs and Hillebrandt, 1983). Then it continues to accrete from the surrounding gas, forming a dense core.

A more detailed understanding of how local collapse proceeds comes from examining the full time history of accretion for different models (Fig. 14). The cessation of strong accretion onto cores occurs long before all gas has been accreted, with the mass fraction at which this occurs depending on the properties of the turbulence (Vázquez-Semadeni et al., 2003). This is because the time that dense cores spend in shock-compressed, high-density regions decreases with increasing driving wave number and increasing driving strength. In the case of long-wavelength driving, cores form coherently in high-density regions associated with one or two large shock fronts that can accumulate a considerable fraction of the total mass of the system, while in the case of short-wavelength driving, the network of shocks is tightly knit, and cores form in smaller clumps and remain in them for shorter times.

H. The time scales of star formation

Turbulent control of star formation predicts that stellar clusters form predominantly in regions that are insufficiently supported by turbulence or where only large-scale driving is active. In the absence of driving, molecular cloud turbulence decays on order of the free-fall time scale \( \tau_{\text{ff}} \) [Eq. (26)], which is roughly the time scale for dense star clusters to form. Even in the presence of support from large-scale driving, substantial collapse still occurs within a few free-fall time scales [Figs. 14 and 20(a)]. If the dense cores followed in these models continue to collapse to quickly build up stellar objects in their centers, then this directly implies the star formation time scale. Therefore the age distribution will be roughly \( \tau_{\text{ff}} \) for stellar clusters that form coherently with high star formation efficiency. When scaled to low densities, say \( n(H_2) \approx 10^2 \text{ cm}^{-3} \) and \( T \approx 10 \text{ K} \), the global free-fall time scale [Eq. (1)] in the models is \( \tau_{\text{ff}} = 3.3 \text{ Myr} \).

If star-forming clouds such as Taurus indeed have ages of order \( \tau_{\text{ff}} \), as suggested by Ballesteros-Paredes, Hartmann, and Vázquez-Semadeni (1999), then the star formation time scale computed here is quite consistent with the low star formation efficiencies seen in Taurus (Leisawitz et al., 1989; Palla and Stahler, 2000; Hartmann, 2001), as the cloud simply has not had time to form many stars. In the case of high-density regions, \( n(H_2) \approx 10^2 \text{ cm}^{-3} \) and \( T \approx 10 \text{ K} \), the dynamical evolution proceeds much faster and the corresponding free-fall time scale drops to \( \tau_{\text{ff}} \approx 10^3 \text{ yr} \). These values are indeed supported by observational data such as the formation time of the Orion Trapezium cluster. It is inferred to have formed from gas of density \( n(H_2) \approx 10^3 \text{ cm}^{-3} \) and is estimated to be less than \( 10^6 \text{ yr} \) old (Hilbornbrand and Hartmann, 1998). The age spread in the models increases with increasing driving wave number \( k \) and increasing effective turbulent Jeans mass \( \langle M_{\text{J}} \rangle_{\text{turb}} \), as shown in Fig. 20. Long periods of core formation for globally supported clouds appear consistent with the low efficiencies of star formation in regions of isolated star formation, such as Taurus, even if they are rather young objects with ages of order \( \tau_{\text{ff}} \).

I. Scales of interstellar turbulence

Turbulence has self-similar properties only on scales between the driving and dissipation scales. What are these scales for interstellar turbulence? The driving scale is determined by what stirs the turbulence. We discuss different driving mechanisms in Sec. VI.C, where we conclude that supernovae are likely dominant in star-forming galaxies. Norman and Ferrara (1996) attempted to estimate from general analytic arguments the driving scale of an ensemble of blast waves from supernovae and superbubbles. However, their description remains to be tested against nonlinear models. An outer limit to the turbulent cascade in disk galaxies is given by the scale height. If indeed molecular clouds are created at least in part by converging large-scale flows generated by the collective influence of recurring supernova explosions in the gaseous disk of our Galaxy, as we argue in Sec. VI.A, then the extent of the Galactic disk is indeed the true upper scale of turbulence in the Milky Way. For
individual molecular clouds this means that turbulent energy is fed in at scales well above the size of the cloud itself. This picture of molecular cloud turbulence being driven by large-scale, external sources is supported by the observation that density and velocity structure shows power-law scaling extending up to the largest scales observed in all clouds that have been analyzed (Ossenkopf and Mac Low, 2002).

In a purely gas-dynamic system the dissipation scale is the scale at which molecular viscosity becomes important. In interstellar clouds the situation may be different. Zweibel and Josafatsson (1983) showed that ambipolar diffusion (discussed in Sec. III.C) is the most important dissipation mechanism in typical molecular clouds with very low ionization fractions $x = \rho_i / \rho_n$, where $\rho_i$ is the density of ions, $\rho_n$ is the density of neutrals, and the total density $\rho = \rho_i + \rho_n$. An ambipolar diffusion coefficient with units of viscosity can be defined as

$$\lambda_{AD} = \frac{v_A^2}{\nu_{ni}},$$

where $v_A^2 = B^2/4\pi \rho_n$ approximates the effective Alfvén speed for the coupled neutrals and ions if $\rho_n \gg \rho_i$, and $\nu_{ni} = \alpha \rho_i$ is the rate at which each neutral is hit by ions, where the coupling constant $\alpha$ is given by Eq. (21). Zweibel and Brandenburg (1997) define an ambipolar diffusion Reynolds number by analogy with the normal viscous Reynolds number as

$$R_{AD} = \frac{\bar{L} \bar{V}}{\lambda_{AD}} = \frac{\bar{M}_A \bar{v}_{ni}}{v_A},$$

which must fall below unity for ambipolar diffusion to be important (also see Balsara, 1996), where $\bar{L}$ and $\bar{V}$ are the characteristic length and velocity scales, and $\bar{M}_A = \bar{V}/v_A$ is the characteristic Alfvén Mach number. In our situation we again can take the rms velocity as a typical value for $\bar{V}$. By setting $R_{AD} = 1$, we can derive a critical length scale below which ambipolar diffusion is important,

$$\bar{L}_{cr} = \frac{v_A}{\bar{M}_A \nu_{ni}} \times \left( \frac{B}{10 \mu G} \right) \left( \frac{\bar{M}_A}{10^{-6}} \right)^{-1} \times \left( \frac{n_n}{10^7 \text{ cm}^{-3}} \right)^{-3/2},$$

with the magnetic-field strength $B$, the ionization fraction $x$, and the neutral number density $n_n$. Here we have taken $\rho_n = \mu n_n$, with $\mu = 2.36 m_\text{H}$. This is consistent with typical sizes of protostellar cores (Bacmann et al., 2000), if we assume that ionization and magnetic

FIG. 20. Masses of individual protostars as function of time in smoothed particle hydrodynamics models: (a) $B_1$ driven at large scales with driving wave number $k=1–2$; (b) $B_2$ with intermediate-scale driving ($k=3–4$); and (c) $B_3$ with small-scale driving ($k=7–8$). For the sake of clarity, only every other object is shown in (a) and (b), whereas in (c) the evolution of every single object is plotted. Time is given in units of the global free-fall time $\tau_g$. Note the different time scale in each plot. In the depicted time interval, models $B_1$ and $B_2$ reach a core mass fraction $M_\bullet = 70\%$, and both form roughly 50 cores. Model $B_3$ reaches $M_\bullet = 35\%$ and forms only 25 cores. Figure (d) compares the distributions of formation times. The age spread increases with decreasing driving scale, showing that clustered core formation should lead to a coeval stellar population, whereas a distributed stellar population should exhibit considerable age spread. From Klessen et al., 2000.
field both depend on the density of the region and follow the empirical laws $n = 3 \times 10^{-3} \text{ cm}^{-3} (n_d/10^3 \text{ cm}^{-3})^{1/2}$ (Mouschovias, 1991b) and $B = 30 \mu \text{G} (n_d/10^3 \text{ cm}^{-3})^{1/2}$ (Crutcher, 1999). Balsara (1996) notes that there are wave families that can survive below $L_{\text{cr}}$ and that resemble gas-dynamic sound waves. This means that this scale may determine where the magnetic field becomes uniform, but not necessarily where the gas-dynamic turbulent cascade cuts off.

J. Termination of local star formation

It remains quite unclear what terminates stellar birth on scales of individual star-forming regions, and even whether these processes are the primary factor determining the overall efficiency of star formation in a molecular cloud [Eq. (2)]. Three main possibilities exist. First, feedback from the stars themselves in the form of ionizing radiation and stellar outflows may heat and stir up surrounding gas sufficiently to prevent further collapse and accretion. Second, accretion might peter out either when all the high-density, gravitationally unstable gas in the region has been accreted in individual stars, or after a more dynamical period of competitive accretion, leaving any remaining gas to be dispersed by the background turbulent flow. Third, background flows may sweep through, destroying the cloud, perhaps in the same way that it was created. Most likely the astrophysical truth lies in some combination of all three possibilities.

If a stellar cluster formed in a molecular cloud contains OB stars, then the radiation field and stellar winds from these high-mass stars strongly influence the surrounding cloud material. The UV flux ionizes gas out beyond the local star-forming region. Ionization heats the gas, raising its Jeans mass, and possibly preventing further protostellar mass growth or new star formation. The termination of accretion by stellar feedback has been suggested at least since the calculations of ionization by Oort and Spitzer (1955), Whitworth (1979) and Yorke et al. (1989) computed the destructive effects of individual blister H II regions on molecular clouds, while in a series of papers, Franco et al. (1994), Rodriguez-Gaspar et al. (1995), and Diaz-Miller et al. (1998) concluded that indeed the ionization from massive stars may limit the star formation efficiency [Eq. (2)] of molecular clouds to about 5%. Matzner (2002) analytically modeled the effects of ionization on molecular clouds, concluding as well that turbulence driven by H II regions could support and eventually destroy molecular clouds. The key question facing these models is whether H II region expansion couples efficiently to clumpy, inhomogeneous molecular clouds, a question probably best addressed with numerical simulations.

Bipolar outflows are a different manifestation of protostellar feedback and may also strongly modify the properties of star-forming regions (Norman and Silk, 1980; Lada and Gautier, 1982; Adams and Fatuzzo, 1996). Matzner and McKee (2000) modeled the ability of bipolar outflows to terminate low-mass star formation, finding that they can limit star formation efficiencies to 30–50 %, although they are ineffective in more massive regions. How important these processes are compared to simple exhaustion of available reservoirs of dense gas (Klessen et al., 2000; Vázquez-Semadeni et al., 2003) remains an important question.

The models relying on exhaustion of the reservoir of dense gas argue that only dense gas will actually collapse and that only a small fraction of the total available gas reaches sufficiently high densities, due to cooling (Elmegreen and Parravano, 1994; Schaye, 2002), gravitational collapse and turbulent triggering (Elmegreen, 2002b), or both (Wada, Meurer, and Norman, 2002). This of course pushes the question of local star formation efficiency up to larger scales, which may indeed be the correct place to ask it.

Other models focus on competitive accretion in local star formation, showing that the distribution of masses in a single group or cluster can be well explained by assuming that star formation is fairly efficient in the dense core, but that stars that randomly start out slightly heavier tend to fall towards the center of the core and accrete disproportionately more gas (Bonnell et al., 1997, 2001a). These models have recently been called into question by the observation that the stars in lower-density young groups in Serpens simply have not had the time to engage in competitive accretion, but still have a normal IMF (Olmi and Testi, 2002).

Finally, star formation in dense clouds created by turbulent flows may be terminated by the same flows that created them. Ballesteros-Paredes, Hartmann, and Vázquez-Semadeni (1999) suggested that the coordination of star formation over large molecular clouds and the lack of post-T Tauri stars with ages greater than about 10 Myr, tightly associated with those clouds, could be explained by their formation in a larger-scale turbulent flow. Hartmann et al. (2001) make the detailed argument that these flows may disrupt the clouds after a relatively short time, limiting their star formation efficiency that way. The extensive evidence for short sequences of cluster ages (Blauw, 1964; Walborn and Parker, 1992; Efremov and Elmegreen, 1998b) is often attributed to sequential triggering (Elmegreen and Lada, 1977) by shock fronts expanding into existing clouds. An additional mechanism for producing such sequences may be the sequential formation of clouds by the larger-scale flow. Below, in Sec. VLC we argue that field supernovae are the most likely driver for the background turbulence, at least in the star-forming regions of galaxies. Supernovae associated with any particular star-forming region will not be energetically important, although they may produce locally significant compressions.

K. Outline of a new theory of star formation

The support of star-forming clouds by supersonic turbulence can explain many of the same observations successfully explained by the standard theory, while also addressing the inconsistencies between observation and
the standard theory described in the previous section. The key point that is new in our argument is that supersonic turbulence produces strong density fluctuations in the interstellar gas, sweeping gas up from large regions into dense sheets and filaments, and does so even in the presence of magnetic fields. Supersonic turbulence decays quickly, but so long as it is maintained by input of energy from some driver it can support regions against gravitational collapse.

Such support comes at a cost, however. The very turbulent flows that support the region produce density enhancements in which the Jeans mass drops as $M_J \propto \rho^{-1/2}$ [Eq. (14)], and the magnetic critical mass above which magnetic fields can no longer support against that collapse drops even faster, as $M_{c\omega} \propto \rho^{-2}$ [Eq. (18)]. For local collapse to actually result in the formation of stars, Jeans-unstable, shock-generated, density fluctuations must collapse to sufficiently high densities on time scales shorter than the typical time interval between two successive shock passages. Only then can they decouple from the ambient flow and survive subsequent shock interactions. The shorter the time between shock passages, the less likely these fluctuations are to survive. Hence the time scale and efficiency of protostellar core formation depend strongly on the wavelength and strength of the driving source, and the accretion histories of individual protostars are strongly time varying (Sec. V.E). Global support by supersonic turbulence thus tends to produce local collapse and low-rate star formation, exactly as seen in low-mass star formation regions characteristic of the disks of spiral galaxies. Conversely, lack of turbulent support results in regions that collapse freely. In gas-dynamic simulations, freely collapsing gas forms a web of density enhancements in which star formation can proceed efficiently, as seen in regions of massive star formation and starbursts.

The regulation of the star formation rate then occurs not just at the scale of individual star-forming cores through ambipolar diffusion balancing magnetostatic support, but rather at all scales via the dynamical processes that determine whether regions of gas become unstable to prompt gravitational collapse. Efficient star formation occurs in collapsing regions; apparent inefficiency occurs when a region is turbulently supported and only small subregions get compressed sufficiently to collapse. The star formation rate is determined by the balance between turbulent support and local density, and is a continuous function of the strength of turbulent support for any given region. Fast and efficient star formation is the natural behavior of gas lacking sufficient turbulent support for its local density.

Regions that are gravitationally unstable in this picture collapse quickly, on the free-fall time scale. They never pass through a quasiequilibrium state as envisioned by the standard model. Large-scale density enhancements such as molecular clouds could be caused either by gravitational collapse, or by ram pressure from turbulence. If collapse does not succeed, the same large-scale turbulence that formed molecular clouds can destroy them again.

V. LOCAL STAR FORMATION

In this section we apply the theoretical picture of Sec. IV to observations of individual star-forming regions. We show how the efficiency and time and length scales of star formation depend on the properties of turbulence (Sec. V.A), followed by a discussion of the properties of protostellar cores (Sec. II.E), the immediate progenitors of individual stars. We then speculate about the formation of binary stars (Sec. V.C) and stress the importance of the dynamical interaction between protostellar cores and their competition for mass growth in dense, deeply embedded clusters (Sec. V.D). This implies strongly time-varying protostellar mass accretion rates (Sec. V.E). Finally, we discuss the consequences of the probabilistic processes of turbulence and stochastic mass accretion for the resulting stellar initial mass function (Sec. V.F).

A. Star formation in molecular clouds

Not only does all star formation occur in molecular clouds, but all giant molecular clouds appear to form stars. At least, all those surveyed within distances less than 3 kpc form stars (Blitz, 1993; Williams et al., 2000), except possibly the Maddalena and Thaddeus (1985) cloud (Lee, Snell, and Dickman, 1996; Williams and Blitz, 1998), and this last cloud may have formed just recently.

The star formation process in molecular clouds appears to be fast. Once the collapse of a cloud region sets in, it rapidly forms an entire cluster of stars within $10^6$ yr or less. This is indicated by the young stars associated with star-forming regions, typically T Tauri stars with ages less than $10^6$ yr (see, for example, Gomez et al., 1992; Greene and Meyer, 1995; Carpenter et al., 1997; Hartmann, 2001) and by the small age spread in more evolved stellar clusters (Hillenbrand, 1997; Palla and Stahler, 1999, 2000). Star clusters in the Milky Way also exhibit an amazing degree of chemical homogeneity (in the case of the Pleiades, see Wilden et al., 2002), implying that the gas out of which these stars formed must have been chemically well mixed initially (see also Avillez and Mac Low, 2002; Klessen and Lin, 2003).

Star-forming molecular clouds in our Galaxy vary enormously in size and mass. In small, low-density clouds stars form with low efficiency, more or less in isolation or scattered around in small groups of up to a few dozen members. Denser and more massive clouds may build up stars in associations and clusters of a few hundred members. This appears to be the most common mode of star formation in the solar neighborhood (Adams and Myers, 2001). Examples of star formation in small groups and associations are found in the Taurus-Aurigae molecular cloud (Hartmann, 2002). Young stellar groups with a few hundred members form in the Chamaeleon I dark cloud (Persi et al., 2000) or ρ-Ophiuchi (Bontemps et al., 2001). Each of these clouds is at a distance of about 130–160 pc from the Sun. Many nearby star-forming regions have been associated with a
The formation of dense rich clusters with thousands of stars is rare. The closest molecular cloud where this happens is the Orion Nebula Cluster in L1641 (Hillenbrand, 1997; Hillenbrand and Hartmann, 1998), which lies at a distance of ~450 pc. A rich cluster somewhat further away is associated with the Monoceros R2 cloud (Carpenter et al., 1997) at a distance of ~830 pc. The cluster NGC 3603 is roughly ten times more massive than the Orion Nebula Cluster. It lies in the Carina region, at about 7 kpc distance. It contains about a dozen O stars and is the nearest object analogous to a starburst knot (Brandl et al., 1999; Moffat et al., 2002). To find star-forming regions building up hundreds of O stars one has to look towards giant extragalactic \( M_{\odot} \) regions, the nearest of which is 30 Doradus in the Large Magellanic Cloud, a satellite galaxy of our Milky Way at a distance at 55 kpc (for an overview see the book edited by Chu et al., 1999). The giant star-forming region 30 Doradus is thought to contain up to a hundred thousand young stars, including more than 400 O stars (Hunter et al., 1995; Walborn et al., 1999). Even more massive star-forming regions are associated with tidal knots in interacting galaxies, as observed in the Antennae (NGC 4038/8; see Zhang, Fall, and Whitmore, 2001) or as in forming regions are associated with tidal knots in interacting galaxies, as observed in the Antennae (NGC 4038/8; see Zhang, Fall, and Whitmore, 2001) or as inferred for starburst galaxies at high redshift (Sanders and Mirabel, 1996).

This sequence demonstrates that the star formation process spans many orders of magnitude in scale, ranging from isolated single stars (\( M = 1M_{\odot} \)) to ultraluminous starburst galaxies with masses of several \( 10^{11}M_{\odot} \) and star formation rates of \( 10^2 - 10^3 M_{\odot} \) yr\(^{-1}\); for comparison the present-day rate in the Milky Way is about \( 1M_{\odot} \) yr\(^{-1}\). This enormous variety of star-forming regions appears to be controlled by the competition between self-gravity and the turbulent velocity field in interstellar gas (see Sec. VI.D).

The control of star formation by supersonic turbulence gives rise to a continuous but articulated picture. There may not be physically distinct modes of star formation, but qualitatively different behaviors do appear over the range of possible turbulent flows. The apparent dichotomy between a clustered and an isolated mode of star formation, as discussed by Lada (1992) for L1630 and Strom, Strom, and Merrill (1993) for L1941, disappears if a different balance between turbulent strength and gravity holds at the relevant length scales in these different clouds.

Turbulent flows tend to have hierarchical structure (She and Leveque, 1994), which may explain the hierarchical distribution of stars in star-forming regions shown by statistical studies of the distribution of neighboring stars in young stellar clusters (Larson, 1995; Simon, 1997; Bate, Clarke, and McCaughrean, 1998; Nakajima et al., 1998; Gladwin et al., 1999; Klessen and Kroupa, 2001). Hierarchical clustering seems to be a common feature of all star-forming regions (Efremov and Elmegreen, 1998a). It may be a natural outcome of turbulent fragmentation (Padoan and Nordlund, 2002).

**B. Protostellar core models**

The observed properties of molecular cloud cores as discussed in Sec. II.E.2 can be compared with gas clumps identified in numerical models of interstellar cloud turbulence. Like their observed counterparts, the model cores are generally highly distorted and triaxial. Depending on the projection angle, they often appear extremely elongated, being part of a filamentary structure that may connect several objects. Figure 21 plots a sample of model cores from Klessen and Burkert (2000). Those which have already formed a protostellar object in their interior are shown on the left, while starless cores without central protostars are shown on the right. Note the similarity to the appearance of observed protostellar cores (Fig. 3). The model clumps are clearly elongated. The ratios between the semimajor and the semiminor axis measured at the second contour level are typically between 2:1 and 4:1. However, there are significant deviations from simple triaxial shapes. As a general trend, high-density contour levels typically are regular and smooth, because there the gas is mostly influenced by pressure and gravitational forces. On the other hand, the lowest contour level samples gas that is strongly influenced by environmental effects. Hence it appears patchy and irregular. The location of the protostar is not necessarily identical with the center of mass of the core, especially when it is irregularly shaped.

The surface density profiles of observed protostellar cores are often interpreted in terms of equilibrium Bonnor-Ebert spheres (Ebert, 1955; Bonnor, 1956). The best example is the Bok globule B68 (Alves et al., 2001). However, it is a natural prediction of turbulent fragmentation calculations that stars form from cloud cores with density profiles that are constant in the inner region, then exhibit an approximate power-law falloff, and that appear finally truncated at some maximum radius. Such a configuration can easily be misinterpreted as Bonnor-Ebert sphere, as illustrated in Fig. 22. Indeed, Ballesteros-Paredes, Klessen, and Vázquez-Semadeni (2003) argue that turbulent fragmentation produces cores that in about 60% of all cases fit well to Bonnor-Ebert profiles, of which most imply stable equilibrium conditions. However, supersonic turbulence does not create hydrostatic equilibrium configurations. Instead, the density structure is transient and dynamically evolving, as the different contributions to virial equilibrium do not balance (Vázquez-Semadeni et al., 2003). None of the cores analyzed by Ballesteros-Paredes et al. (2002) are in true equilibrium, and the physical properties inferred from fitting Bonnor-Ebert profiles generally deviate from the real values.

Besides the direct comparison of projected surface density maps, there is additional evidence supporting the idea of the turbulent origin of the structure and kinematics of molecular cloud cores and clouds as a whole. This includes comparisons of numerical models of supersonic turbulence to stellar extinction measurements (Padoan et al., 1997), Zeeman splitting measurements (Padoan and Nordlund, 1999), polarization maps
(Heitsch, Zweibel, et al., 2001; Ostriker, Stone, and Gammie, 2001; Padoan, Goodman, et al., 2001), determination of the velocity structure of dense cores and their immediate environment (Padoan, Juvela, et al., 2001), and various other statistical measures of the structure and dynamics of observed clouds (as mentioned in Sec. II.B).

The density profiles of observed cores can also be fit by invoking confinement by helical magnetic fields (Fiege and Pudritz, 2000a, 2000b). Helical field structures unwind, though, as magnetic tension forces straighten field lines. Therefore these models require external forces to continuously exert strong torques in order to twist the field lines. These would necessarily drive strong flows, making it impossible to achieve the static equilibrium configurations required by the model.

Other models that have been proposed to describe the

FIG. 21. Protostellar cores from a model of clustered star formation. The left side shows protostellar cores with collapsed central objects (indicated by a black dot), the right side “starless” cores without central protostars. Cores are numbered according to their peak density. Surface density contours are spaced logarithmically with two contour levels spanning one decade, $\log_{10} \Delta \rho = 0.5$. The lowest contour is a factor of $10^{0.5}$ above the mean density. From Klessen and Burkert, 2000.
properties of protostellar cores are based on quasistatic equilibrium conditions (e.g., with a composite polytropic equation of state, as discussed by Curry and McKee, 2000), invoke thermal instability (e.g., Yoshii and Saito, 1980; Gilden, 1984b; Graziani and Black, 1987; Burkert and Lin, 2000), gravitational instability through ambipolar diffusion (e.g., Basu and Mouschovias, 1994; Nakamura, Hanawa, and Nakano, 1995; Ciolek and Basu, 2000; Indebetouw and Zweibel, 2000) or nonlinear Alfvén waves (e.g., Carlberg and Pudritz, 1990; Elmegreen, 1990, 1997a, 1999b), or rely on clump collisions (e.g., Gilden, 1984a; Kimura and Tosa, 1996; Murray and Lin, 1996). Models based on supersonic turbulence as discussed here appear to be most consistent with observational data.

C. Binary formation

In order to study binary formation, one must understand stellar collapse in multiple dimensions. Dynamical
modeling in 2D has the advantage of speed compared to 3D simulations and therefore allows for the inclusion of a larger number of physical processes while reaching higher spatial resolution. The disadvantage of 2D models is that only axisymmetric perturbations can be studied. Initial attempts to study collapse in 3D were reported by Cook and Harlow (1978), Boss and Bodenheimer (1979), Boss (1980b), Rozyczka et al. (1980), and Tohline (1980). Since these early studies, numerical simulations of the collapse of isolated isothermal objects have been extended, for example, to include highly oblate cores (Boss, 1996), elongated filamentary cloud cores (Bastien et al., 1991; Tomisaka, 1995, 1996a; Inutsuka and Miyama, 1997), differential rotation (Boss and Myhill, 1995), and different density distributions for the initial spherical cloud configuration with or without barlike perturbations. Whereas 1D spherical collapse models could only treat the formation of single stars, the 2D and 3D calculations show that the formation of binary and higher-order multiple stellar systems can be described in terms of the classical dynamical theory. Indeed, they show that it is a likely outcome of protostellar collapse and molecular cloud fragmentation. For a comprehensive overview see Bodenheimer et al. (2000).

The observed fraction of binary and multiple stars relative to single stars is about 50% for the field star population in the solar neighborhood. This has been determined for all known F7–G9 dwarf stars within 22 pc from the Sun by Duquennoy and Mayor (1991) and for M dwarfs out to similar distances by Fischer and Marcy (1992; see also Leinert et al., 1997). The binary fraction for pre-main sequence stars appears to be at least equally high (see, for example, Köhler and Leinert, 1998, or Table I in Mathieu et al., 2000). These findings put strong constraints on the theory of star formation, as any reasonable model needs to explain the observed high number of binary and multiple stellar systems. It has long been suggested that subfragmentation and multiple star formation is a natural outcome of isothermal collapse (Hoyle, 1953). However, stability analyses show that the growth time of small perturbations in the isothermal phase is typically small compared to the collapse time scale itself (Silk and Suto, 1988; Hanawa and Nakayama, 1997).

Hence, in order to form multiple stellar systems, either perturbations to the collapsing core must be external and strong, or subfragmentation must occur at a later, nonisothermal phase of collapse, after a protostellar disk has formed. This disk may become gravitationally unstable if the surface density exceeds a critical value given by the epicyclic frequency and the sound speed (Safranov, 1960; Toomre, 1964; see derivation in Sec. VI.B.2), allowing fragmentation into multiple objects (as summarized by Bodenheimer et al., 2000).

Contracting gas clumps with strong external perturbation occur naturally in turbulent molecular clouds or when stars form in clusters. While collapsing to form or feed protostars, clumps may lose or gain mass from interaction with the ambient turbulent flow (Klessen et al., 2000). In a dense cluster environment, collapsing clumps may merge to form larger clumps containing multiple protostellar cores that subsequently compete with each other for accretion from the common gas environment (Murray and Lin, 1996; Bonnell et al., 1997; Klessen and Burkert, 2000, 2001). Strong external perturbations and capture through clump merger leads to wide binaries or multiple stellar systems. Stellar aggregates with more than two stars are dynamically unstable, so protostars may be ejected against the gas-rich environment they accrete from. This not only terminates their mass growth, but leaves the remaining stars behind more strongly bound. These dynamical effects can transform the original wide binaries into close binaries (see also Kroupa, 1995a, 1995b, 1995c). Binary stars that form through disk fragmentation are close binaries right from the beginning, as typical sizes of protostellar disks are of order of a few hundred AU (Bate, Bonnell, and Bromm, 2002b).

Magnetic fields above all influence the development of close binaries, as magnetic braking acts during the collapse of protostellar cores (Sec. III.C). Preliminary 3D models of the collapse of a magnetized, rotating cloud, described in Balsara (2001), already demonstrate that braking can drain enough angular momentum to prevent the formation of a binary from a core that would form a binary in the absence of magnetic field. This effect is not captured by the simple inclusion of an additional pressure term to model the magnetic field, as
done by Boss (2000, 2002), as it is the magnetic tension that brakes rotation. The observed prevalence of binaries suggests that the magnetic field must indeed decouple from collapsing cores at some stage, presumably by ambipolar diffusion. However, quantitative models of this process remain needed.

D. Dynamical interactions in clusters

Star-forming regions can differ enormously in scale and density as a consequence of supersonic turbulence (as discussed in Sec. V.A). Stars almost never form in isolation, but instead in groups and clusters. The number density of protostars and protostellar cores in rich compact clusters can be high enough for mutual dynamical interactions to become important. This introduces a further degree of stochasticity to the star formation process in dense clusters beyond the statistical chaos associated with turbulence and turbulent fragmentation in the first place.

When a molecular cloud region of a few hundred solar masses or more coherently becomes gravitationally unstable, it contracts and builds up a dense cluster of embedded protostars within one or two free-fall time scales. While contracting individually to build up a star in their interior, protostellar gas clumps still follow the global flow patterns. They stream towards a common center of attraction, undergo further fragmentation, or merge together. The time scales for clump mergers and clump collapse are comparable. Merged clumps therefore can contain multiple protostars that now compete with each other for further accretion. They are embedded in the same limited and rapidly changing reservoir of contracting gas. As the cores are dragged along with the global gas flow, a dense cluster of accreting protostellar cores quickly builds up. As in dense stellar clusters, the dynamical evolution is subject to the complex gravitational interaction between the cluster members. Close encounters or even collisions occur, drastically altering the protostellar orbits. Triple or higher-order systems often form. They are generally unstable, so a large fraction of protostellar cores are expelled from the parental cloud. The expected complexity of protostellar dynamics already in the deeply embedded phase of evolution is illustrated in Fig. 23, which shows trajectories of five accreting protostars in a calculation of molecular cloud fragmentation and clustered star formation by Klessen and Burkert (2000).

The effects of mutual dynamical interaction of protostellar cores in the embedded phase of star cluster formation have been investigated by a variety of authors. Here, we list some basic results:

(a) Close encounters in nascent star clusters strongly influence the accretion disk expected to surround every protostar. These disks can be tidally truncated or even disrupted. This influences mass accretion through the disk, modifies the ability to subfragment and form a binary star and the probability of planet formation (Clarke and Pringle, 1991; Murray and Clarke, 1993; McDonald and Clarke, 1995; Hall et al., 1996; Bonnell, Smith, et al., 2001; Kroupa and Burkert, 2001; Scally and Clarke, 2001; Smith and Bonnell, 2001). In particular, Ida, Larwood, and Burkert (2000) note that an early stellar encounter may explain features of our own solar system, namely, the high eccentricities and inclinations observed in the outer part of the Edgeworth-Kuiper Belt at distances larger than 42 AU.

(b) Stellar systems with more than two members are in general unstable. In a triple system, for example, the lowest-mass member has the highest probability of being expelled. If this happens during the embedded phase, the protostar leaves a region of high-density gas. This terminates further mass growth and sets its final mass. Thus the dynamical processes have important consequences for the resulting stellar mass spectrum in dense stellar clusters (see Sec. V.F). Ejected objects can travel quite far, and indeed this has been suggested to account for the so-called “runaway” T Tauri stars found in x-ray

FIG. 23. Example of protostellar interactions in an embedded nascent star cluster: (a) The projected trajectories of five accreting cores in a numerical model of star cluster formation by Klessen and Burkert (2000). We highlight two events in the evolutionary sequence: (b) the formation of an unstable triple system at the beginning of cluster formation with the lowest mass member being expelled from the cluster, and (c) binary hardening in a close encounter together with subsequent acceleration of the resulting close binary due to another distant encounter during late evolution. The corresponding parts of the orbital paths are enlarged by a factor of 6. For simplicity, neither the trajectories of other cores in the cluster nor the distribution of gas are shown. Numbers next to the trajectories identify the protostellar cores. For further detail see Klessen and Burkert (2000).
observation in the vicinities of star-forming molecular clouds (Sterzik and Durison, 1995, 1998; Smith et al., 1997; Klessen and Burkert, 2000; for observations, see Neuhauser et al., 1995, or Wichmann et al., 1997). However, many of these stars could not have traveled to their observed positions in their own lifetimes if they were formed in the currently star-forming cloud. It appears more likely that the observed extended stellar population is associated with clouds that dispersed long ago.

(c) Dynamical interaction leads to mass segregation. Star clusters evolve towards equipartition. For massive stars this means that they have on average smaller velocities than low-mass stars (in order to keep the kinetic energy $E_{\text{kin}}=1/2 m v^2$ roughly constant). Thus massive stars sink towards the cluster center, while low-mass stars predominantly populate large cluster radii (Kroupa, 1995a, 1995b, 1995c). This holds already for nascent star clusters in the embedded phase (Bonnell and Davies, 1998).

(d) Dynamical interaction and competition for mass accretion lead to highly time-variable protostellar mass growth rates. This is discussed in more detail in Sec. V.E. (e) The radii of stars in the pre-main-sequence contraction phase are several times larger than stellar radii on the main sequence (for a review on pre-main-sequence evolution see, for example, Palla, 2000, 2002). Stellar collisions are therefore more likely to occur during the very early evolution of star clusters. During the embedded phase the encounter probability is further increased by gas drag and dynamical friction.

Collisions in dense protostellar clusters have been proposed as mechanism to produce massive stars (Bonnell, Bate, and Zinnecker, 1998; Stahler, Palla, and Ho, 2000; Bonnell and Bate, 2002). The formation of massive stars has long been considered a puzzle in theoretical astrophysics, because 1D calculations predict that for stars above $\sim 10M_\odot$ the radiation pressure acting on the infalling dust grains will be strong enough to halt or even reverse further mass accretion (Yorke and Krügel, 1977; Wolfire and Cassinelli, 1987; Palla, 2000, 2002). However, detailed 2D calculations by Yorke and Sonnhalter (2002) demonstrate that in the more realistic scenario of mass growth via an accretion disk the radiation barrier may be overcome. Mass can accrete from the disk onto the star along the equator while radiation is able to escape along the polar direction. Massive stars thus may form via the same processes as ordinary low-mass stars (also see McKee and Tan, 2002). Collisional processes need not be invoked.

E. Accretion rates

When a gravitationally unstable gas clump collapses onto a central star, it follows the observationally well-determined sequence described in Sec. II.E. In the main accretion phase (class 0), the release of gravitational energy by accretion dominates the energy budget. Hence protostars exhibit large IR and submillimeter luminosities and drive powerful outflows. Both phenomena can be used to estimate the protostellar mass accretion rate $\dot{M}$ (André and Montmerle, 1994; Bontemps et al., 1996; Henrichsen, André, and Bontemps, 1997). These observations suggest that $\dot{M}$ varies strongly, and declines with time after the class-0 phase (Sec. III.D.3). The estimated lifetimes are a few tens of thousands of years for the class-0 and a few hundreds of thousands of years for the class-I phase.

Most models of protostellar core collapse concentrate on isolated objects, whether the models are analytic (Larson, 1969; Penston, 1969a; Basu, 1997; Hunter, 1977; Henrichsen et al., 1997) or numerical (Foster and Chevalier, 1993; Tomisaka, 1996a, 1996b; Ogino, Tomisaka, and Nakamura, 1999; Wuchterl and Tscharnuter, 2003). A typical accretion history is shown in Fig. 10. However, stars predominantly form in groups and clusters. Numerical studies that investigate the effect of the cluster environment on protostellar mass accretion rates have been reported by Bonnell et al. (1998, 2001a, 2001b), Klessen and Burkert (2000, 2001), Klessen et al. (2000), Heitsch, Mac Low, and Klessen (2001), and Klessen (2001a). These numerical models suggest the following predictions about protostellar accretion in dense clusters:

(a) Protostellar accretion rates from turbulent fragmentation in a dense cluster environment are strongly time variable. This is illustrated in Fig. 24 for 49 randomly selected cores from the model of Klessen (2001a).

(b) The typical density profiles of gas clumps that give birth to protostars indeed exhibit a flat inner core, followed by a density falloff $\rho \propto r^{-2}$, and are truncated at some finite radius, which in the dense centers of clusters often is due to tidal interaction with neighboring cores (see Secs. II.E and V.D). As a result, the modeled accretion rates agree well with the observations. A short-lived initial phase of strong accretion occurs when the flat inner part of the prestellar clump collapses, corresponding to class 0. If the cores remain isolated and unperturbed, the mass growth rate gradually declines in time as the outer envelope accretes, giving class I. Once the truncation radius is reached, accretion fades and the object enters class II. However, collapse does not start from rest for the density fluctuations considered here, so accretion rates exceed the values predicted by models of isolated objects, even for objects in the simulations far from their nearest neighbors.

(c) The mass accretion rates of protostellar cores in a dense cluster also deviate strongly from the rates of isolated cores because of mergers and competition between cores, as discussed above (Sec. V.D). Mergers drastically change the density and velocity structure of cores, so the predictions for isolated cores no longer hold. Furthermore these new, larger cores contain multiple protostars that subsequently compete with each other for accretion from a common gas reservoir. The most massive protostar in a clump accretes more matter than its competitors (see also Bonnell et al., 1997, 2001a, 2001b; Klessen and Burkert, 2000). Its accretion rate is enhanced through clump mergers, whereas the accretion rate of low-mass cores typically decreases. Many temporary accretion peaks in the wake of clump mergers are visible in Fig. 24. The small aggregates of cores that build up
are dynamically unstable, so low-mass cores may be ejected. As they leave the high-density environment, accretion terminates and their final mass is reached.

(d) The most massive protostars begin to form first and continue to accrete at a high rate throughout the entire cluster evolution. As the most massive gas clumps tend to have the largest density contrast, they are the first to collapse, forming the center of the nascent cluster. These protostars are fed at a high rate and gain mass very quickly. McKee and Tan (2002) argue on analytical grounds that they may form in as little as 10^5 yr, even in the absence of competitive accretion. As their parental cores merge with others, more gas is fed into their “sphere of influence.” They are able to maintain or even increase the accretion rate when competing with lower-mass objects (e.g., cores 1 and 8 in Fig. 24). Low-mass stars tend to form somewhat later in the dynamical evolution of the system (as indicated by the absolute formation times in Fig. 24; also Fig. 8 in Klessen and Burkert, 2000) and typically have only short periods of high accretion.

(e) As high-mass stars form in massive cores, while low-mass stars form in less massive cores, the stellar population in clusters should be mass segregated right from the start. High-mass stars form in the center, lower-mass stars tend to form towards the cluster outskirts. This agrees with recent observations of the cluster NGC 330 in the Small Magellanic Cloud (Sirianni et al., 2002). Dynamical effects during the embedded phase of star cluster evolution enhance this initial segregation even further [see (c) in Sec. V.D].

(f) Individual cores in a cluster environment form and evolve through a sequence of highly probabilistic events, so their accretion histories differ even if they accumulate the same final mass. Accretion rates for protostars of a certain mass can only be determined in a statistical sense. Klessen (2001a) suggests that an exponentially declining rate with a peak value of a few 10^{-5} M_☉ yr^{-1}, a time constant in the range 0.5–2.5 × 10^5 yr, and a cutoff related to gas dispersal from the cluster offers a reasonable fit to the average protostellar mass growth in dense embedded clusters, but with large variations.

F. Initial mass function

Knowledge of the distribution of stellar masses at birth, described by the initial mass function (IMF), is necessary to understand many astrophysical phenomena, but no analytic derivation of the observed IMF has yet stood the test of time. In fact, it appears likely that a fully deterministic theory for the IMF does not exist. Rather, any viable theory must take into account the probabilistic nature of the turbulent process of star formation, which is inevitably highly stochastic and indeterminate. We gave a brief overview of the observational constraints on the IMF in Sec. II.F, and we here review models for it.

1. Models of the IMF

Existing models to explain the IMF can be divided into five major groups. In the first group feedback from the stars themselves determines their masses. Silk (1995) suggests that stellar masses are limited by the feedback from both ionization and protostellar outflows. Nakano, Hasegawa, and Norman (1995) describe a model in which stellar masses are sometimes limited by the mass scales of the formative medium and sometimes by stellar feedback. Adams and Fatuzzo (1996) apply the central-limit theorem to the hypothesis that many independent physical variables contribute to the stellar masses to derive a log-normal IMF regulated by protostellar feedback. However, for the overwhelming majority of stars with masses M ≤ 5 M_☉, protostellar feedback (i.e., winds, radiation, and outflows) are unlikely to be strong.
enough to halt mass accretion, as shown by detailed protostellar collapse calculations (Wuchterl and Klessen, 2001; Wuchterl and Tscharnuter, 2003).

In the second group of models, initial and environmental conditions determine the IMF. In this picture, the structural properties of molecular clouds determine the mass distribution of Jeans-unstable gas clumps, and the clump properties determine the mass of the stars that form within. If one assumes a fixed star formation efficiency for individual clumps, there is a one-to-one correspondence between the molecular cloud structure and the final IMF. The idea that fragmentation of clouds leads directly to the IMF dates back to Hoyle (1953) and later Larson (1973). More recently, this concept has been extended by Larson (1992, 1995) to include the observed fractal and hierarchical structure of molecular clouds. Indeed, random sampling from a fractal cloud seems to be able to reproduce the basic features of the observed IMF (Elmegreen and Mathieu 1983; Elmegreen, 1997a, 1997b, 1999a, 2000a, 2000c, 2002a). A related approach is to see the IMF as a domain packing problem (Richtler, 1994).

The hypothesis that stellar masses are determined by clump masses in molecular clouds is supported by observations of the dust continuum emission of protostellar condensations in the Serpens, ρ-Ophiuchi, and Orion star-forming regions (Motte et al., 1998, 2001; Testi and Sargent, 1998; Johnstone et al., 2000, 2001). These protostellar cores are thought to be in a phase immediately before a star forms in their interior. Their mass distribution resembles the stellar IMF reasonably well, suggesting a close correspondence between protostellar clump masses and stellar masses, leaving little room for stellar feedback processes, competitive accretion, or collisions to act to determine the stellar mass spectrum.

A third group of models relies on competitive coagulation or accretion processes to determine the IMF. This has a long tradition and dates back to investigations by Oort (1954) and Field and Saslaw (1965), but the interest in this concept continues to the present day (Silk and Takahashi, 1979; Lejeune and Bastien, 1986; Price and Podsiadlowski, 1995; Murray and Lin, 1996; Bonnell et al., 2001a, 2001b; Durisen, Sterzik, and Pickett, 2001). Stellar collisions require very high stellar densities, however, for which observational evidence and theoretical mechanisms remain scarce.

Fourth, there are models that connect the supersonic turbulent motions in molecular clouds to the IMF. In particular, there are a series of attempts to find an analytical relation between the stellar mass spectrum and statistical properties of interstellar turbulence (e.g., Larson, 1981; Fleck, 1982; Hunter and Fleck, 1982; Elmegreen, 1993; Padoan, 1995; Padoan et al., 1997; Myers, 2000; Padoan and Nordlund, 2002). However, properties such as the probability distribution of density in supersonic turbulence in the absence of gravity have never successfully been shown to have a definite relationship to the final results of gravitational collapse (Padoan et al., 1997). Even the more sophisticated model of Padoan and Nordlund (2002) does not take into account that not single but multiple compressions and rarefactions determine the density structure of supersonic turbulence (Passot and Vázquez-Semadeni, 1998, 2003) and that higher-mass clumps will fragment into multiple protostars. Furthermore, such models neglect the effects of competitive accretion in dense cluster environments (Sec. V.E), which may be important for determining the upper end of the IMF.

Finally, there is a more statistical approach. Larson (1973) and Zinnecker (1984, 1990) argued that whenever a large set of parameters is involved in determining the masses of stars, invoking the central-limit theorem of statistics naturally leads to a log-normal stellar mass spectrum (Adams and Fatuzzo, 1996, made similar arguments).

Regardless of the detailed physical processes involved, the common theme in all of these models is the probabilistic nature of star formation. It appears impossible to predict the formation of specific individual objects. Only the fate of an ensemble of stars can be described ab initio. The implication is that the star formation process can only be understood within the framework of a probabilistic theory.

2. Turbulent fragmentation example

To illustrate some of the issues discussed above, we examine the mass spectra of gas clumps and collapsed cores from models of self-gravitating, isothermal, supersonic turbulence driven with different wavelengths (Klessen, 2001b). In the absence of magnetic fields and more accurate equations of state, these models can only be illustrative, not definitive, but nevertheless they offer insight into the processes acting to form the IMF. Figure 25 shows that, before local collapse begins to occur, the clump mass spectrum is not well described by a single power law. During subsequent evolution, as clumps merge and grow larger, the mass spectrum extends towards higher masses, approaching a power law with slope \( \alpha \approx -1.5 \). Local collapse sets in, resulting in the formation of dense cores, most quickly in the freely collapsing model. The influence of gravity on the clump mass distribution weakens when turbulence dominates over gravitational contraction on the global scale, as in the other three models. The more the turbulent energy dominates over gravity, the more the spectrum resembles the initial case of pure gas-dynamic turbulence. This suggests that the clump mass spectrum in molecular clouds will be shallower in regions where gravity dominates over turbulent energy. This may explain the observed range of slopes for the clump mass spectrum in different molecular cloud regions (Sec. II.B).

Like the distribution of Jeans-unstable clumps, the mass spectrum of dense protostellar cores resembles a log-normal in the models without turbulent support and with long-wavelength turbulent driving, with a peak at roughly the average thermal Jeans mass \( m_\text{J} \) of the system. These models also predict initial mass segregation \([e] \text{ in Sec. V.E} \). The protostellar clusters discussed here
contain only between 50 and 100 cores. This allows for comparison with the IMF only around the characteristic mass scale, typically about 1 \( M_{\odot} \), since the numbers are too small to study the very low- and high-mass ends of the distribution. Focusing on low-mass star formation, however, Bate, Bonnell, and Bromm (2002a) demonstrate that brown dwarfs are a natural and frequent outcome of turbulent fragmentation. In this model, brown dwarfs form when dense molecular gas fragments into unstable multiple systems that eject their smallest members from the dense gas before they have been able to accrete to stellar masses (also see Reipurth and Clarke, 2001). Numerical models with sufficient dynamic range to treat the full range of stellar masses [Eq. (7)] remain to be done.

VI. GALACTIC-SCALE STAR FORMATION

How do the mechanisms that control local star formation determine the global rate and distribution of star formation in galaxies? In this section we begin (Sec. VI.A) by examining how molecular clouds form from the interstellar medium. We then outline in Sec. VI.B what determines the efficiency of star formation. We ar-
gue that the balance between the density of available gas and its turbulent velocity determines where star formation will occur, and how efficiently. Even when the turbulent velocity in a region is relatively high, if the density in that region is also high, the region may still not be supported against gravitational collapse and prompt star formation. Therefore any mechanism that increases the local density without simultaneously increasing the turbulent velocity sufficiently can lead to star formation, via molecular cloud formation. Most mechanisms that increase local density appear to be external to the star formation process, however. Accretion during initial galaxy formation, interactions and collisions between galaxies, spiral gravitational instabilities of galactic disks, and bar formation are major examples. In this review we cannot do justice to the vast literature on galactic dynamics and the interactions that determine the density distribution in galaxies. We do, however, examine what physical mechanisms control turbulent velocity dispersion in Sec. VI.C. Finally, in Sec. VI.D we briefly speculate on how turbulent control of star formation may help explain objects with very different star formation properties, including low-surface-brightness galaxies, normal galactic disks, globular clusters, galactic nuclei, and primordial dwarf galaxies.

A. Formation and lifetime of molecular clouds

How do molecular clouds form? Any explanation must account for their low star formation efficiencies and broad linewidths. Molecular hydrogen forms on dust grains at a rate calculated by Hollenbach, Werner, and Salpeter (1971) to be

$$t_{\text{form}} = \left(1.5\times10^9 \text{ yr} \right) \left( \frac{n}{1 \text{ cm}^{-3}} \right)^{-1},$$

(31)

where $n$ is the number density of gas particles. Recent experimental work by Pirronello, Biham et al. (1997), Pirronello, Liu et al. (1997), and Pirronello et al. (1999) on molecular hydrogen formation on graphite and olivine suggests that these rates may be strongly temperature dependent, so that the Hollenbach et al. (1971) result may be a lower limit to the formation time. However, the same group reports that molecule formation is rather more efficient on amorphous ices (Manicolo et al., 2001) such as would be expected on grain surfaces deep within dark clouds, so that the rates computed by Hollenbach et al. (1971) may be reached after all. Further experimental investigation of molecule formation appears necessary.

When molecular clouds were first discovered, they were thought to have lifetimes of over 100 Myr (Scoville and Hersh, 1979) because of their apparent predominance in the inner galaxy. These estimates were shown by Blitz and Shu (1980) to depend upon too high a conversion factor between CO and H$_2$ masses. They revised the estimated lifetime down to roughly 30 Myr based on the association of clouds with spiral arms, apparent ages of associated stars, and overall star formation rate in the Galaxy.

Chemical equilibrium models of dense cores in molecular clouds (as reviewed, for example, by Irvine, Goldsmith, and Hjalmarson, 1986) showed disagreements with observed abundances in a number of molecules. These cores would take as long as 10 Myr to reach equilibrium, which could still occur in the standard model. However, Prasad, Heere, and Tarafdar (1991) demonstrated that the abundances of the different species agreed much better with the results at times of less than 1 Myr from time-dependent models of the chemical evolution of collapsing cores (also see Sec. III.D.2). Bergin, Goldsmith, et al. (1997), Bergin and Langer (1997), Pratap et al. (1997), and Aikawa et al. (2001) came to similar conclusions from careful comparison of several different cloud cores to extensive chemical model networks. Aikawa et al. (2001) and Saito et al. (2002) also studied deuterium fractionation, again finding short lifetimes.

Ballesteros-Paredes, Hartmann, and Vázquez-Semadeni (1999) argue for a lifetime of less than 10 Myr for molecular clouds as a whole. They base their argument on the notable lack of a population of 5–20-Myr-old stars in molecular clouds. Stars in the clouds typically have ages under 3–5 Myr, judging from their position on pre-main-sequence evolutionary tracks in a Hertzsprung-Russell diagram (D’Antona and Mazzitelli, 1994; Swenson et al., 1994; with discrepancies resolved by Stauffer, Hartmann, and Barrado y Navascués, 1995). Older weak-line T Tauri stars identified by x-ray surveys with Einstein (Walter et al., 1988) and ROSAT (Neuhäuser et al., 1995) are dispersed over a region as much as 70 pc away from molecular gas, suggesting that they were not formed in the currently observed gas (Feigelson, 1996). Leisawitz, Bash, and Thaddeus (1989), Fukui et al. (1999), and Elmegreen (2000b) have made similar arguments based on the observation that only stellar clusters with ages under about 10 Myr are associated with substantial amounts of molecular gas in the Milky Way and the Large Magellanic Cloud.

For these short lifetimes to be plausible, either molecule formation must proceed quickly, and therefore at high densities, or observed molecular clouds must be formed from preexisting molecular gas, as suggested by Pringle, Allen, and Lubow (2001). A plausible place for fast formation of H$_2$ at high density is the shock compressed layers naturally produced in a supernova-driven ISM, as shown in Fig. 26 from Avillez and Breitschwerdt (2003). Similar morphologies have been seen in many other global simulations of the ISM, including those by Rosen, Bregman, and Norman (1993), Rosen and Bregman (1995), Rosen, Bregman, and Kelson (1996), Korpi et al. (1999), Wada and Norman (1999, 2001), and Avillez (2000). Mac Low (2000) reviews these earlier simulations. Mac Low et al. (2001) showed that pressures in the ISM are broadly distributed, with peak pressures in cool gas ($T<10^3$ K) as much as an order of magnitude above the average because of shock compressions (also see Passot and Vázquez-Semadeni, 2003). This gas is swept up from ionized $10^4$-K gas, so between cooling and compression its density has already been
increased by two orders of magnitude from \( n = 1 \text{ cm}^{-3} \) to \( n = 100 \text{ cm}^{-3} \). These simulations did not include a correct cooling curve below \( 10^4 \text{ K} \), so further cooling could not occur even if physically appropriate, but it would be expected.

We can understand this compression quantitatively. The sound speed in the warm gas is \((8.1 \text{ km s}^{-1})(T/10^4 \text{ K})^{1/2}\), taking into account the mean mass per particle \( \mu = 2.11 \times 10^{-24} \text{ g} \) for gas 90% H and 10% He by number. The typical velocity dispersion for this gas is 10–12 km s\(^{-1}\) (Dickey, Hanson, and Helou, 1990; Dickey and Lockman, 1990), so that shocks with Mach numbers \( M = 2 - 3 \) are moderately frequent. Temperatures in these shocks reach values of \( T \approx 10^5 \text{ K} \), which is close to the peak of the interstellar cooling curve (Dalgarno and McCray, 1972; Raymond, Cox, and Smith, 1976), so the gas cools quickly back to \( 10^3 \text{ K} \). The density behind an isothermal shock is \( \rho_1 = M^2 \rho_0 \), where \( \rho_0 \) is the preshock density, so order-of-magnitude density enhancements occur easily. The optically thin radiative cooling rate \( \Lambda(T) \) drops off at \( 10^4 \text{ K} \) as H atoms no longer radiate efficiently (Dalgarno and McCray, 1972; Spaans and Norman, 1997), but the radiative cooling \( L \propto n^2 \Lambda(T) \). Therefore density enhancements strongly increase the ability to cool. Hennebelle and Pérault (1999) show that such shock compressions can trigger the isobaric thermal instability (Field et al., 1969; Wolfire et al., 1995), reducing temperatures to of order 100 K or less. Heiles (2001) observes a broad range of temperatures for neutral hydrogen from below 100 K to a few thousand K. The reduction in temperature by two orders of magnitude from \( 10^4 \) to 100 K raises the density correspondingly. Combined with the initial isothermal shock compression, this results in a total of as much as three orders of magnitude of compression. Gas that started at densities somewhat higher than average, say at \( 10^3 \text{ cm}^{-3} \), can be compressed to densities of \( 10^5 \text{ cm}^{-3} \), enough to reduce \( \text{H}_2 \) formation times to a few hundred thousand years.

Koyama and Inutsuka (2000) have demonstrated numerically that shock-confined layers do indeed quickly develop high enough densities to form \( \text{H}_2 \) in under a million years, using 1D computations including heating and cooling rates from Wolfire et al. (1995) and \( \text{H}_2 \) formation and dissociation. In Fig. 27 we show the parameter space in which they find \( \text{H}_2 \) formation to be efficient. Hartmann, Ballesteros-Paredes, and Bergin (2001) make a more general argument for rapid \( \text{H}_2 \) formation, based in part on lower-resolution, 2D simulations described by Passot, Vázquez-Semadeni, and Pouquet (1995) that could not fully resolve realistic densities like those of Koyama and Inutsuka (2000), but do include larger-scale flows showing that the initial conditions for the 1D models are quite reasonable. Hartmann et al. (2001) further argue that the self-shielding against the background UV field also required for \( \text{H}_2 \) formation will become important at approximately the same column densities required to become gravitationally unstable.

Shock-confined layers were shown numerically to be unstable by Hunter et al. (1986) in the context of colliding spherical density enhancements, and by Stevens, Blondin, and Pollack (1992) in the context of colliding stellar winds. Vishniac (1994) demonstrated analytically that isothermal, shock-confined layers are subject to a nonlinear thin-shell instability. The physical mechanism can be seen by considering a shocked layer perturbed sinusoidally. The ram pressure on either side of the layer acts parallel to the incoming flow and thus at an angle to the surface of the perturbed layer. Momentum is deposited in the layer with a component parallel to the surface, which drives material towards extrema in the layer, causing the perturbation to grow. A numerical study by Blondin and Marks (1996) in two dimensions demonstrated that the nonlinear thin-shell instability saturates in a thick layer of trans-sonic turbulence when the flows

---

**FIG. 26.** Log of number density in a cut through the galactic plane from a 3D supernova-driven model of the ISM with resolution of 1.25 pc, including radiative cooling and the vertical gravitational field of the stellar disk, as described by Avillez (2000), Avillez and Mac Low (2002), and Avillez and Breitschwerdt (2003). High-density, shock-confined regions are naturally produced by intersecting supernova-shocks from field supernovae.

**FIG. 27.** Shock velocities \( V_d \) and preshock number densities \( n \) at which the cold postshock layer is more than 8% molecular, taken from 1D simulations by Koyama and Inutsuka (2000) that include \( \text{H}_2 \) formation and dissociation, and realistic heating and cooling functions from Wolfire et al. (1995).
become sufficiently chaotic that the surface no longer rests at a substantial angle to the normal of the incoming flow.

Thermal instability will act in conjunction with shock confinement (see Sec. VI.B). Goldsmith (1970) and Schwarz, McCray, and Stein (1972) first computed the nonlinear development of the thermal instability, demonstrating that shock waves form during the dynamical collapse of nonlinear regions. Hennebelle and Pérault (1999) demonstrated that shock compression can trigger thermal instability in otherwise stable regions in the diffuse ISM, even in the presence of magnetic fields (Hennebelle and Pérault, 2000), so that compressions much greater than the isothermal factor of $M^2$ can occur (see the quantitative discussion by Vázquez-Semadeni et al., 1996).

These cold, dense layers are themselves subject to dynamical instabilities, as has recently been shown in 2D computations by Koyama and Inutsuka (2002). The instabilities they found are caused by some combination of thermal instability and mechanisms very similar to the nonlinear thin-shell instability (Vishniac, 1994) for the isothermal case. Figure 28 shows another example of these instabilities from a numerical study by Walder and Folini (2000). These dynamical instabilities can drive strongly supersonic motions in the cold, dense layer. If that layer is dense enough for molecule formation to proceed quickly, those molecules will show strongly supersonic linewidths on all but the very smallest scales, as seen in the models of Koyama and Inutsuka (2002), in agreement with the observations of molecular clouds. It remains to be shown whether this scenario can quantitatively explain the full ensemble of molecular clouds observed in the solar neighborhood or elsewhere in our own and external galaxies.

The final destruction of molecular clouds then proceeds from a combination of several effects. First, once the external turbulent compression has passed, they will begin to freely expand (Vázquez-Semadeni et al., 2002), but only at the sound speed of the cold gas of $0.2\, \text{km}\, \text{s}^{-1}$, or roughly a parsec every 5 Myr. Second, the same turbulent flows that formed them may again tear them apart. As the density decreases in either of these cases, background dissociating radiation will tend to destroy the molecules (McKee, 1989). Third, radiation from stars forming in the cloud may heat and dissociate the molecular gas, reducing its density and preventing it from forming further stars (Matzner, 2002, and references therein).

B. When is star formation efficient?

1. Overview

Observers have documented a surprisingly strong connection between the star formation rate and the local velocity dispersion, column density, and rotational velocity of disk galaxies (Kennicutt, 1998a; Martin and Kennicutt, 2001). A global Schmidt (1959) law relates star formation rate per unit area to gas surface density as

$$\Sigma_{\text{SFR}} = A \Sigma_{\text{gas}}^N,$$

where a value of $N=1.4\pm0.05$ can be derived from the observations (Kennicutt, 1989, 1998b). Star formation also cuts off sharply at some radius in most star-forming galaxies (Kennicutt, 1989; Martin and Kennicutt, 2001), which also appears related to the gas surface density.

The Schmidt law can be interpreted as reflecting star formation on a free-fall time scale, so that (following Wong and Blitz, 2002, for example) the star formation rate per unit volume of gas with density $\rho$ is

$$\dot{\rho}_{\text{SF}} = \varepsilon_{\text{SF}} \frac{\rho}{\tau_{\text{ff}}} \frac{\rho}{(G\rho)^{1/2}} \rho^{1.5},$$

where $\varepsilon_{\text{SF}}$ is an efficiency factor observed to be substantially less than unity. [It differs from the star formation efficiency $\varepsilon_{\text{SF}}$ defined in Eq. (2) only in that we here compare to the free-fall lifetime rather than the total lifetime of the system.]

The connection between magnetically controlled small-scale star formation and large-scale star formation is not clear in the standard theory. Shu, Adams, and Lizano (1987) did indeed suggest that OB associations were formed by freely collapsing gas that had overwhelmed the local magnetic field, but that still implied that the star formation rate was controlled by the details of the magnetic-field structure, which in turn is presumably controlled by the galactic dynamo. The connection appears clearer, though, if turbulence, as represented by the velocity dispersion, controls the star formation rate. The same physical mechanisms control star formation at all scales. Regions that are globally supported by turbulence still engage in inefficient star formation, but the
frequency of regions of efficient star formation determines the overall star formation rate in a feedback loop.

The big open question in this area remains the importance of radiative cooling for efficient star formation, either on its own or induced by turbulent compression. Is cooling, and indeed molecule formation, necessary for gravitational collapse to begin, or is it rather a result of already occurring collapse in gravitationally unstable gas? Certainly there are situations in which cooling will make the difference between gravitational stability and instability, but are those just marginal cases or the primary driver for star formation in galaxies?

In Fig. 29 we outline a unified picture that depends on turbulence and cooling to control the star formation rate. After describing the different elements of this picture, we discuss the steps that we think will be needed to move from this cartoon to a quantitative theory of the star formation rate. The factor that determines the star formation rate above any other is whether the gas is sufficiently dense to be gravitationally unstable without additional cooling. Galactic dynamics and interactions with other galaxies and the surrounding intergalactic gas determine the average gas densities in different regions of a galaxy. The gravitational instability criterion here includes both turbulent motions and galactic shear, as well as magnetic fields. If gravitational instability sets in at large scale, collapse will continue so long as sufficient cooling mechanisms exist to prevent the temperature of the gas from rising (effective adiabatic index \( \gamma_{\text{eff}} \leq 1 \)). Molecular clouds can form in less than 10^5 yr, as the gas passes through densities of 10^4 cm^{-3} or higher, as an incidental effect of the collapse. A starburst results, with stars forming efficiently in compact clusters. The size of the gravitationally unstable region determines the size of the starburst.

If turbulent support, rather than thermal support, prevents the gas from immediately collapsing, compression-induced cooling can become important. Supersonic turbulence compresses some fraction of the gas strongly. As most cooling mechanisms depend on the gas density nonlinearly, the compressed regions cool quickly. When these regions reach densities of order 10^4 cm^{-3}, again molecule formation occurs, allowing the gas to cool to even lower temperatures (see Sec. VI.A). These cold regions then can become gravitationally unstable and collapse, if allowed by the local turbulence. Triggering by nearby star formation events (Elmegreen and Lada, 1977) represents a special case of this mode (see Sec. VI.D.2). This mechanism is less efficient than prompt gravitational instability, as much of the gas is not compressed enough to form molecules. It is, however, more efficient in regions of higher average density. Galactic dynamics again determines the local average density and so, in the end, the star formation efficiency in this regime as well.

If turbulence even in the cooled regions supports the gas against general gravitational collapse, isolated, low-rate star formation can still occur locally in regions further compressed by the turbulence. This may describe regions of low-mass star formation like the Taurus clouds. On the other hand, if the cooled gas begins to collapse gravitationally, locally efficient star formation can occur. The size of the gravitationally unstable region then really determines whether a group, OB association, or bound cluster eventually forms. Star formation in regions like Orion may result from this branch.

2. Gravitational instabilities in galactic disks

Now let us consider the conditions under which gravitational instability will set in. On galactic scales, the Jeans instability criterion for gravitational instability must be modified to include the additional support offered by the shear coming from differential rotation, as well as the effects of magnetic fields. The gravitational potential of the stars can also contribute to gravitational instability on large scales. Which factor determines the onset of gravitational instability remains unknown. Five that have been proposed are the temperature of the cold phase, the surface density, the local shear, the presence of magnetic fields, and the velocity dispersion, in different combinations.

We can heuristically derive the Toomre (1964) criterion for stability of a rotating, thin disk with uniform velocity dispersion \( \sigma \) and surface density \( \Sigma \) using time-scale arguments (Schaye, 2002). First consider the Jeans criterion for instability in a thin disk, which requires that the time scale for collapse of a perturbation of size \( \lambda \),

\[
t_{\text{coll}} = \sqrt{\lambda / G \Sigma},
\]

be shorter than the time required for the gas to respond to the collapse, the sound crossing time,

\[
t_{\text{sc}} = \lambda / c_s.
\]

This implies that gravitational stability requires perturbations with size

\[
\lambda < c_s^2 / G \Sigma.
\]

Similarly, in a disk rotating differentially, a perturbation will spin around itself, generating centrifugal motions that can also support against gravitational collapse. This will be effective if the collapse time scale \( t_{\text{coll}} \) exceeds...
the rotational period \( t_{\text{rot}} = 2\pi/\kappa \), where \( \kappa \) is the epicyclic frequency, so that stable perturbations have

\[
\lambda > 4\pi^2 G \Sigma / \kappa^2.
\]  

(37)

A regime of gravitational instability occurs if there are wavelengths that lie between the regimes of pressure and rotational support, with

\[
\frac{c_s^2}{G \Sigma} < \lambda < \frac{4\pi^2 G \Sigma}{\kappa^2}.
\]  

(38)

This will occur if

\[
Q = \frac{c_s\kappa}{2\pi G \Sigma} < 1,
\]  

(39)

which is the Toomre criterion for gravitational instability to within a factor of 2. The full criterion from a linear analysis of the equations of motion of gas in a shearing disk gives a factor of \( \pi \) in the denominator (Safronov, 1960; Goldreich and Lynden-Bell, 1965), while a kinetic theory approach appropriate for a collisionless stellar system gives a factor of 3.36 (Toomre, 1964).

Kennicutt (1989) and Martin and Kennicutt (2001) have demonstrated that the Toomre criterion generally can explain the location of the edge of the star-forming disk in galaxies, although they must introduce a correction factor \( a = 0.69 \pm 0.2 \) into the left-hand side of Eq. (39). Schaye (2002) notes that this factor should be corrected to \( a = 0.53 \) to account for the use of the velocity dispersion rather than sound velocity, and the exact Toomre criterion for a stellar rather than a gas disk.

The Toomre criterion given in Eq. (39) was derived for a pure gas disk with uniform temperature and velocity dispersion and no magnetic field. Relaxation of each of these assumptions modifies the criterion, and indeed each has been argued by different authors to be the controlling factor in determining star formation thresholds.

Stars in a gas disk respond as a collisionless fluid to density perturbations large compared to their mean separation. Jog and Solomon (1984a) computed the Toomre instability in a disk composed of gas and stars, and found it to always be more unstable than either component considered individually. Both components contribute to the growth of density perturbations, allowing gravitational collapse to occur more easily. Taking into account both gas (subscript \( g \)) and stars (subscript \( r \)), instability occurs when

\[
2\pi G k \left( \frac{\Sigma_r}{\kappa^2 + k^2 c_s^2 r_s} + \frac{\Sigma_g}{\kappa^2 + k^2 c_s^2 g_s} \right) > 1,
\]  

(40)

where \( k = 2\pi/\lambda \) is the wave number of the perturbation considered. Jog and Solomon (1984b) and Romeo (1992) extended this model to include the effect of the finite thickness of the disk. Elmegreen (1995) was able with some effort to derive an effective Toomre parameter that includes the effects of both stars and gas, but that can only be analytically computed in the thin-disk limit. To compute it, independent measures of the velocity dispersion of the stars and of the gas are, of course, needed. Jog (1996) numerically computed the effective stability parameter for a wide range of values of stellar and gas disk parameters. The contribution of the stellar disk may alone be sufficient to explain the correction factors found by Kennicutt (1989) and Martin and Kennicutt (2001).

Magnetic fields offer direct support against collapse through their magnetic pressure and tension. However, Chandrasekhar (1954) and Lynden-Bell (1966) were the first to note that they can also have the less expected effect of destabilization of a rotating system. The magnetic field in this case acts to brake the shear that would otherwise prevent collapse, redistributing angular momentum and allowing collapse to occur down field lines. Elmegreen (1987) performed a linear analysis of the growth rate of gravitational instability in a rotating, magnetized disk, which was extended by Fan and Lou (1997) to follow the excitation of the different modes. Kim and Ostriker (2001) determined when a magnetic field acts to prevent or to promote collapse. When shear is strong, as it is in the parts of galactic disks with flat rotation curves, and the field is moderate or weak, with plasma \( \beta \leq 1 \), swing amplification stabilized by magnetic pressure dominates. Sufficiently unstable disks, with Toomre \( Q \leq 1.0–1.1 \) (depending on field strength), collapse due to nonlinear secondary instabilities despite magnetic stabilization. On the other hand, if shear is weak, and fields are stronger (\( \beta > 1 \)), magnetic tension forces act against epicyclic motions, reducing their stabilizing effect, and producing magneto-Jeans instabilities along the field lines. This leads to large regions of gravitational collapse. In the outer parts of disks, the collapse rate from swing amplification is so slow that additional effects such as spiral arm amplification may be important to drive the formation of observed regions of star formation. Kim and Ostriker (2002) show that the introduction of spiral arms indeed produce featherlike features similar to those observed in spiral galaxies, with masses comparable to the largest star-forming regions. These results suggest that the presence of magnetic fields can actually enhance the star formation rate in some parts of galactic disks.

The temperature and the velocity of the coldest gas in a multiphase interstellar medium at any point in the disk may be the determining factor for gravitational instability, rather than some average temperature. Schaye (2002) suggests that the sharp rise in temperature associated with the lack of molecular gas causes the sharp drop in the star formation rate at the edges of disk galaxies. This reverses the argument of Elmegreen and Parravano (1994), who suggested that the lack of gravitational instability prevents cooling. Schaye (2002) derives the disk surface density required to allow molecule formation in the presence of the intergalactic ultraviolet background field and suggests that this is consistent with the observed threshold column densities. However, Martin and Kennicutt (2001) show a wide variation in the atomic gas fraction at the critical radius [see their Fig. 9(a)], calling this idea into question.

The balance between gravitation and local shear is argued by Hunter, Elmegreen, and Baker (1998) to be a better criterion than the Toomre (1964) criterion, which
balances gravitation against Coriolis forces. Effectively this substitutes the Oort A constant (Binney and Tremaine, 1987) for the epicyclic frequency $\kappa$ in Eq. (39). The difference is small (of order 10%) in galaxies with flat rotation curves, but can lower the critical density substantially in galaxies with rising rotation curves, such as dwarf galaxies.

3. Thermal instability

Thermal instability has been the organizing principle behind the most influential models of the ISM (Field et al., 1969; McKee and Ostriker, 1977; Wolfire et al., 1995). Under the assumption of approximate pressure and thermal equilibrium, thermal instability can explain the widely varying densities observed in the ISM. It cannot explain the order-of-magnitude higher pressures observed in molecular clouds, though, so it was thought that most molecular clouds must be confined by their own self-gravity. Turbulent pressure fluctuations in a medium with effective adiabatic index less than unity (that is, one that cools when compressed, like the ISM) can provide an alternative explanation for both pressure and density fluctuations. Although thermal instability exists, it does not necessarily act as the primary structuring agent nor therefore as the determining factor for the star formation rate.

Thermal instability occurs when small perturbations from thermal equilibrium grow. The dependence on density $\rho$ and temperature $T$ of the heat-loss function $L = \lambda - T$, the sum of the rate of specific energy loss minus gain, determines whether instability occurs. Parker (1953) derived the isochoric instability condition, while Field (1965) pointed out that cooling inevitably causes density changes, either due to dynamical flows if the region is not isobaric, or due to pressure changes if it is. He then derived the isobaric instability condition. The alternative of dynamical compression in a region large enough to be unable to maintain isobaric conditions has recently received renewed attention, as described below.

The isobaric instability condition derived by Field (1965) is

$$\left( \frac{\partial L}{\partial T} \right)_p = \left( \frac{\partial L}{\partial T} \right)_\rho - \frac{\rho_0}{T_0} \left( \frac{\partial L}{\partial \rho} \right)_T < 0,$$  \hspace{1cm} (41)

where $\rho_0$ and $T_0$ are the equilibrium values. Optically thin radiative cooling in the interstellar medium gives a cooling function that can be expressed as a piecewise power law $L = \rho^2 T^\beta$, where $\beta$ gives the value for a temperature range $T_{1} < T < T_{2}$, while photoelectric heating is independent of temperature. Isobaric instability occurs when $\beta_1 < 1$, while isochoric instability only occurs with $\beta_1 = 0$ (Field, 1965).

In interstellar gas cooling with equilibrium ionization, there are two temperature ranges subject to thermal instability (Pikel’ner, 1968; Field et al., 1969). In the standard picture of the three-phase interstellar medium governed by thermal instability (McKee and Ostriker, 1977), the higher of these, with temperatures $10^{15} < T < 10^{7}$ K (Raymond, Cox, and Smith, 1976), separates hot gas from the warm ionized medium. The lower range of $10^{12} < T < 10^{4.5}$ K (Fig. 3(a) of Wolfire et al., 1995) separates the warm neutral medium from the cold neutral medium. Cooling of gas out of ionization equilibrium has been studied in a series of papers by Spaans (1996; Spaans and Norman, 1997; Spaans and van Dishoeck, 1997; Spaans and Carollo, 1998) as described by Spaans and Silk (2000). The effective adiabatic index depends quite strongly on the details of the local chemical, dynamical, and radiation environment, in addition to the pressure and temperature of the gas. Although regions of thermal instability occur, the pressures and temperatures may depend strongly on the details of the radiative transfer in a turbulent medium, the local chemical abundances, and other factors.

When thermal instability occurs, it can drive strong motions that dynamically compress the gas nonlinearly. Thereafter neither the isobaric nor the isochoric instability conditions hold, and the structure of the gas is determined by the combination of dynamics and thermodynamics (Meerson, 1996; Burkert and Lin, 2000; Lynden-Bell and Tout, 2001; Kritsuk and Norman, 2002a; Sánchez-Salcedo et al., 2002), Vázquez-Semadeni, Gazol, and Scalo (2000) examined the behavior of thermal instability in the presence of driven turbulence, magnetic fields, and Coriolis forces and concluded that the structuring effect of the turbulence overwhelmed that of thermal instability in a realistic environment. Gazol et al. (2001) and Sánchez-Salcedo, Vázquez-Semadeni, and Gazol (2002) found that about half of the gas in a turbulent environment actually has temperatures falling in the thermally unstable region, and emphasized that a bimodal temperature distribution may simply be a reflection of the gas cooling function, not a signature of a discontinuous phase transition. Mac Low et al. (2003) examined supernova-driven turbulence and found a broad distribution of pressures, which were more important than thermal instability in producing a broad range of densities in the interstellar gas.

Heiles (2001) confirmed the suggestions of Dickey, Salpeter, and Terzian (1978) and Mebold et al. (1982) that substantial amounts of gas lie out of thermal equilibrium. This has provided observational support for a picture in which turbulent flows rather than thermal instability dominate structure formation prior to gravitational collapse. Heiles (2001) measured the temperature of gas along lines of sight through the warm and cold neutral medium by comparing absorption and emission profiles of the H I 21-cm fine-structure line. He found that nearly half of the warm neutral clouds measured showed temperatures that are unstable according to the application of the isobaric instability condition, Eq. (41), to the Wolfire et al. (1995) equilibrium ionization phase diagram.

Although the heating and cooling of the gas clearly plays an important role in the star formation process, the presence or absence of an isobaric instability may be less important than the effective adiabatic index, or...
similar measures of the behavior of the gas on compression, in determining its ultimate ability to form stars.

C. Driving mechanisms

Both support against gravity and maintenance of observed motions appear to depend on continued driving of the turbulence, which has kinetic energy density $e = (1/2) \rho v_{\text{rms}}^2$. Mac Low (1999, 2002) estimates that the dissipation rate for isothermal, supersonic turbulence is

$$\dot{e} = - (1/2) \rho v_{\text{rms}}^2 L_d \left( \frac{\rho}{\text{g cm}^{-3}} \right) \left( \frac{\Omega}{\text{s}^{-1}} \right) \left( \frac{L_d}{\text{pc}} \right) \left( \frac{v_{\text{rms}}}{10 \text{ km s}^{-1}} \right)^3 \left( \frac{\mu}{100 \text{ pc}} \right)^{-1},$$

where $L_d$ is the driving scale, which we have somewhat arbitrarily taken to be 100 pc (though it could well be smaller), and we have assumed a mean mass per particle $\mu = 2.11 \times 10^{-24} \text{ g}$. The dissipation time for turbulent kinetic energy is

$$\tau_d = e/\dot{e} = L_d/v_{\text{rms}} \left( \frac{L_d}{100 \text{ pc}} \right) \left( \frac{v_{\text{rms}}}{10 \text{ km s}^{-1}} \right)^{-1},$$

which is just the crossing time for the turbulent flow across the driving scale (Elmegreen, 2000b). What, then, is the energy source for this driving? We here review the energy input rates for a number of possible mechanisms.

1. Magnetorotational instabilities

One energy source for interstellar turbulence that has long been considered is shear from galactic rotation (Fleck, 1981). However, the question of how to couple from the large scales of galactic rotation to smaller scales remained open. Sellwood and Balbus (1999) suggested that the magnetorotational instability (Balbus and Hawley, 1991, 1998) could couple the large and small scales efficiently. The instability generates Maxwell stresses (a positive correlation between radial $B_r$ and azimuthal $B_\phi$ magnetic-field components) that transfer energy from shear into turbulent motions at a rate

$$\dot{\varepsilon} = - T_{R\Phi} d(\Omega/d \ln R) = T_{R\Phi} \Omega,$$

where the last equality holds for a flat rotation curve. (Sellwood and Balbus, 1999). Numerical models suggest that the Maxwell stress tensor is $T_{R\Phi} = 0.6 B^2/(8 \pi)$ (Hawley, Gammie, and Balbus, 1995). For the Milky Way, the value of the rotation rate recommended by the IAU is $\Omega = (220 \text{ Myr})^{-1} = 1.4 \times 10^{-16} \text{ rad s}^{-1}$, though this may be as much as 15% below the true value (Olling and Merrifield, 1998, 2000). The magnetorotational instability may thus contribute energy at a rate

$$\dot{\varepsilon} = (3 \times 10^{-29} \text{ erg cm}^{-3} \text{ s}^{-1}) \left( \frac{B}{3 \mu G} \right)^2 \left( \frac{\Omega}{(220 \text{ Myr})^{-1}} \right).$$

For parameters appropriate to the H I disk of a sample small galaxy, NGC 1058, including $\rho = 10^{-24} \text{ g cm}^{-3}$, Sellwood and Balbus (1999) find that the magnetic field required to produce the observed velocity dispersion of 6 km s$^{-1}$ is roughly 3 $\mu$G, a reasonable value for such a galaxy. This instability may provide a base value for the velocity dispersion, below which no galaxy will fall. If that is sufficient to prevent collapse, little or no star formation will occur, producing something like a low-surface-brightness galaxy with large amounts of H I and few stars. This may also apply to the outer disk of our own Milky Way and other star-forming galaxies.

2. Gravitational instabilities

 Motions coming from gravitational collapse have often been suggested as a local driving mechanism in molecular clouds, but fail due to the quick decay of the turbulence (Sec. IV.A). If the turbulence decays in less than a free-fall time, as suggested by Eq. (26), then it cannot delay collapse for substantially longer than a free-fall time.

On the galactic scale, spiral structure can drive turbulence in gas disks. Roberts (1969) first demonstrated that shocks would form in gas flowing through spiral arms formed by gravitational instabilities in the stellar disk (Lin and Shu, 1964; Lin, Yuan, and Shu, 1969). These shocks were studied in thin disks by Tubbs (1980) and Soukoup and Yuan (1981), who found few vertical motions. More recently, it has been realized that in a more realistic thick disk, the spiral shock will take on some properties of a hydraulic bore, with gas passing through a sudden vertical jump at the position of the shock (Martos and Cox, 1998; Gómez and Cox, 2002). Behind the shock, downward flows of as much as 20 km s$^{-1}$ appear (Gómez and Cox, 2002). Some portion of this flow will contribute to interstellar turbulence. However, the observed presence of interstellar turbulence in irregular galaxies without spiral arms, as well as in the outer regions of spiral galaxies beyond the regions where the arms extend, suggest that this cannot be the only mechanism driving turbulence. A more quantitative estimate of the energy density contributed by spiral arm driving has not yet been done.

The interaction between rotational shear and gravitation can, at least briefly, drive turbulence in a galactic disk, even in the absence of spiral arms. Vollmer and Beckert (2002) describe the consequences of assuming that this effect fully supplies the energy to drive the observed turbulent flow, without demonstrating explicitly that this is the case. They base their assumption on 2D, high-resolution (subparsec zones) numerical models described in a series of papers by Wada and Norman (1999, 2001), Wada, Spaans, and Kim (2000), and Wada, Meurer, and Norman (2002). However, these numerical models all share two limitations: they do not include the dominant stellar component, and gravitational collapse cannot occur beneath the grid scale. The computed filaments of dense gas are thus artificially supported and would actually continue to collapse to form stars, rather
than driving turbulence in dense disks. Sánchez-Salcedo (2001) gives a detailed critique of these models. In very-low-density disks, where even the dense filaments remained Toomre stable, this mechanism might operate, however.

Wada et al. (2002) estimated the energy input from this mechanism using Eq. (44), where, in the absence of significant Maxwell stresses, the stress tensor is given by the Newton stresses resulting from correlations in the gravitational velocity \( u_G \) as \( T_{p\Phi} = (p u_G u_G \rho) \) (Lynden-Bell and Kalnajs, 1972). However, the Newton stresses will add energy only if a positive correlation between radial and azimuthal gravitational forces exists, which is not demonstrated by Wada et al. (2002). Nevertheless, they estimate the order of magnitude of the energy input from Newton stresses as

\[
\dot{E} \approx G (\Sigma_g / H)^2 \lambda^3 \Omega
\]

\[
\approx (4 \times 10^{-29} \text{ erg cm}^{-3} \text{s}^{-1})
\times \left( \frac{\Sigma_g}{10 M_\odot \text{ pc}^{-2}} \right)^2 \left( \frac{H}{100 \text{ pc}} \right)^{-2}
\times \left( \frac{\lambda}{100 \text{ pc}} \right)^2 \left( \frac{\Omega}{(220 \text{ Myr})^{-1}} \right),
\]

(46)

where \( G \) is the gravitational constant, \( \Sigma_g \) the density of gas, \( H \) the scale height of the gas, \( \lambda \) the length scale of turbulent perturbations, and \( \Omega \) the angular velocity of the disk. Values chosen are appropriate for the Milky Way. This is two orders of magnitude below the value required to maintain interstellar turbulence [Eq. (42)].

3. Protostellar outflows

Protostellar jets and outflows are a popular suspect for the energy source of the observed turbulence in molecular clouds. We can estimate their average energy input rate into the overall ISM, following McKee (1989), by assuming that some fraction \( f_w \) of the mass accreted onto a star during its formation is expelled in a wind traveling at roughly the escape velocity. Shu et al. (1988) argue that \( f_w \approx 0.4 \) and that most of the mass is ejected close to the stellar surface, where the escape velocity is

\[
v_{esc} = \left( \frac{2GM}{R} \right)^{1/2}
\]

\[
= (200 \text{ km s}^{-1}) \left( \frac{M}{1 M_\odot} \right)^{1/2} \left( \frac{R}{10 R_\odot} \right)^{-1/2},
\]

(47)

with scaling appropriate for a solar-type protostar with radius \( R = 10 R_\odot \). Observations of neutral atomic winds from protostars show outflow velocities of roughly this value (Lizano et al., 1988; Giovanardi, et al. 2000).

The total energy input from protostellar winds will substantially exceed the amount that can be transferred to the turbulence because of radiative cooling at the wind termination shock. We represent the fraction of energy lost there by \( \eta_w \). A reasonable upper limit to the energy loss is offered by assuming fully effective radiation and momentum conservation, so that

\[
\eta_w \leq \frac{v_{rms}}{v_w} = 0.05 \left( \frac{v_{rms}}{10 \text{ km s}^{-1}} \right)^{-1} \left( \frac{200 \text{ km s}^{-1}}{v_w} \right),
\]

(48)

where \( v_{rms} \) is the rms velocity of the turbulence, and we have assumed that the flow is coupled to the turbulence at typical velocities for the diffuse ISM. If we assumed that most of the energy went into driving dense gas, the efficiency would be lower, as typical rms velocities for CO outflows are only \( 1-2 \text{ km s}^{-1} \). The energy injection rate is

\[
\dot{E} = \frac{1}{2} f_w \eta_w \frac{\Sigma_\ast}{H} v_w^2
\]

\[
= (2 \times 10^{-28} \text{ erg cm}^{-3} \text{s}^{-1}) \left( \frac{H}{200 \text{ pc}} \right)^{-1} \left( \frac{f_w}{0.4} \right)
\]

\[
\times \left( \frac{v_{rms}}{200 \text{ km s}^{-1}} \right) \left( \frac{v_w}{10 \text{ km s}^{-1}} \right)
\]

\[
\times \left( \frac{\Sigma_\ast}{4.5 \times 10^{-9} M_\odot \text{ pc}^{-2} \text{yr}^{-1}} \right),
\]

(49)

where \( \Sigma_\ast \) is the surface density of star formation, and \( H \) is the scale height of the star-forming disk. The scaling value used for \( \Sigma_\ast \) is the solar neighborhood value (McKee, 1989).

Although protostellar jets and winds are indeed quite energetic, they deposit most of their energy into low-density gas (Henning, 1989), as is shown by the observation of multiparsec long jets extending completely out of molecular clouds (Bally and Devine, 1994). Furthermore, observed motions of molecular gas show increasing power on scales all the way up to and perhaps beyond the largest scale of molecular cloud complexes (Ossenkopf and Mac Low, 2002). It is hard to see how such large scales could be driven by protostars embedded in the clouds.

4. Massive stars

In active star-forming galaxies, massive stars probably dominate the driving. They could do so through ionizing radiation and stellar winds from O stars, or clustered and field supernova explosions, predominantly from B stars no longer associated with their parent gas. The supernovae appear likely to be most important, as we now show.

a. Stellar winds

First, we consider stellar winds. The total energy input from a line-driven stellar wind over the main-sequence lifetime of an early O star can equal the energy from its supernova explosion, and its Wolf-Rayet wind can be even more powerful. However, the mass-loss rate from stellar winds drops as roughly the sixth power of the star’s luminosity if we take into account that stellar luminosity varies as the fourth power of stellar mass (Vink, de Koter, and Lamers, 2000), while the powerful
Wolf-Rayet winds (Nugis and Lamers, 2000) last only $10^5$ yr or so, so only the very most massive stars contribute substantial energy from stellar winds. The energy from supernova explosions, on the other hand, remains nearly constant down to the least massive star that can explode. As there are far more lower-mass stars than massive stars, with a Salpeter initial mass function giving a power law in mass of $\alpha = -2.35$ [Eq. (8)], supernova explosions inevitably dominate over stellar winds after the first few million years of the lifetime of an OB association.

b. Ionizing radiation

Next, we consider ionizing radiation from OB stars. The total amount of energy contained in ionizing radiation is vast. Abbott (1982) estimates the integrated luminosity of ionizing radiation in the disk of the Milky Way to be

$$\dot{e} = 1.5 \times 10^{-24} \text{ erg s}^{-1} \text{ cm}^{-3}. \quad (50)$$

However, only a small fraction of this total energy goes to driving interstellar motions.

Ionizing radiation contributes to interstellar turbulence in two ways. First, it ionizes the diffuse interstellar gas, heating it to $7000$–$10000$ K and adding energy to it. As this gas cools, it contracts due to thermal instabilities, driving turbulent flows, as modeled by Kritsuk and Norman (2002a, 2002b). They modeled the flow in a cooling instability after a sudden increase in heating by a factor of 5, and found that a flow with peak thermal energy of $E_{\text{th}}$ gains a peak kinetic energy of roughly $E_{\text{kin}} = \eta_t E_{\text{th}}$, with $\eta_t = 0.07$. Parravano, Hollenbach, and McKee (2003) find that the local UV radiation field, and thus the photoelectric heating rate, increases by a factor of 2–3 due to the formation of a nearby OB association every 100–200 Myr. However, substantial motions only lasted about 1 Myr after a heating event in the model by Kritsuk and Norman (2002b). We can estimate the energy input from this mechanism on average by taking the kinetic-energy input from the heating event and dividing by the typical time $\tau_{\text{OB}}$ between heating events. If we take the thermal energy to be that of $n = 1$ cm$^{-3}$ gas at $10^4$ K (perhaps a bit higher than typical), we find that

$$\dot{e} = \frac{3}{2} n k T \eta_c / \tau_{\text{OB}}$$

$$= (5 \times 10^{-29} \text{ erg cm}^{-3} \text{ s}^{-1}) \left(\frac{n}{1 \text{ cm}^{-3}}\right) \left(\frac{T}{10^4 \text{ K}}\right)$$

$$\times \left(\frac{\eta_c}{0.07}\right) \left(\frac{\tau_{\text{OB}}}{100 \text{ Myr}}\right)^{-1}. \quad (51)$$

Although comparable to some other proposed energy sources discussed here, this mechanism appears unlikely to be as important as the supernova explosions from the same OB stars, as discussed below.

The second way that ionization drives turbulence is through driving the supersonic expansion of H II regions after photoionization heating raises their pressures above that of the surrounding neutral gas. Matzner (2002) computes the momentum input from the expansion of an individual H II region into a surrounding molecular cloud, as a function of the cloud mass and the ionizing luminosity of the central OB association. By integrating over the H II region luminosity function derived by McKee and Williams (1997), he finds that the average momentum input from a Galactic region is

$$\langle \delta p \rangle = (260 \text{ km s}^{-1})$$

$$\times \left(\frac{N_H}{1.5 \times 10^{22} \text{ cm}^{-2}}\right)^{-3/14} \left(\frac{M_{cl}}{10^6 M_\odot}\right)^{1/14} \langle M_\star \rangle. \quad (52)$$

The column density $N_H$ is scaled to the mean value for Galactic molecular clouds (Solomon et al., 1987), which varies little as cloud mass $M_{cl}$ changes. The mean stellar mass per cluster in the Galaxy is $\langle M_\star \rangle = 440 M_\odot$ (Matzner, 2002).

The number of OB associations contributing substantial amounts of energy can be drawn from the McKee and Williams (1997) cluster luminosity function,

$$N(> S_{49}) = 6.1 \left(\frac{10^8}{S_{49} - 1}\right), \quad (53)$$

where $N$ is the number of associations with ionizing photon luminosity exceeding $S_{49} = S/(10^{49} \text{ s}^{-1})$. The luminosity function is rather flat below $S_{49} = 2.4$, the theoretical luminosity of the highest-mass single star considered (120$M_\odot$), so taking its value at $S_{49} = 1$ is about right, giving $N(>1) \approx 650$ clusters.

To derive an energy input rate per unit volume $\dot{e}$ from the mean momentum input per cluster $\langle \delta p \rangle$, we need to estimate the average velocity of momentum input $v_i$, the time over which it occurs $t_i$, and the volume $V$ under consideration. Typically expansion will not occur supersonically with respect to the interior, so $v_i \ll c_{s,i}$, where $c_{s,i} = 10$ km s$^{-1}$ is the sound speed of the ionized gas. McKee and Williams (1997) argue that clusters go through about five generations of massive star formation, in which each generation lasts $\langle t_i \rangle \approx 3.7$ Myr. The scale height for massive clusters is $h_{\text{c}} \sim 100$ pc (Bronfman et al., 2000), and the radius of the star-forming disk is roughly $R_{sf} \sim 15$ kpc, so the relevant volume $V = 2 \pi R_{sf}^2 h_{\text{c}}$. The energy input rate from H II regions is then

$$\dot{e} = \frac{\langle \delta p \rangle}{V t_i} N(>1) v_i$$

$$= (3 \times 10^{-30} \text{ erg s}^{-1} \text{ cm}^{-3})$$

$$\times \left(\frac{N_H}{1.5 \times 10^{22} \text{ cm}^{-2}}\right)^{-3/14} \left(\frac{M_{cl}}{10^6 M_\odot}\right)^{1/14} \left(\frac{M_\star}{440 M_\odot}\right) \left(\frac{N(>1)}{650}\right) \left(\frac{v_i}{10 \text{ km s}^{-1}}\right)$$

$$\times \left(\frac{H_{c}}{100 \text{ pc}}\right)^{-1} \left(\frac{R_{sf}}{15 \text{ kpc}}\right)^{-2} \left(\frac{t_i}{18.5 \text{ Myr}}\right)^{-1}. \quad (54)$$
where all the scalings are appropriate for the Milky Way, as discussed above. Nearly all of the energy in ionizing radiation goes towards maintaining the ionization and temperature of the diffuse medium, and hardly any towards driving turbulence. Flows of ionized gas may be important very close to young clusters and may terminate star formation locally (Sec. IVJ), but do not appear to contribute significantly on a global scale.

c. Supernovae

The largest contribution from massive stars to interstellar turbulence comes from supernova explosions. To estimate their energy input rate, we begin by finding the supernova rate in the Galaxy \( \sigma_{SN} \). Cappellaro et al. (1999) estimate the total supernova rate in supernova units to be \( 0.72 \pm 0.21 \) SNe for galaxies of type S0a-b and \( 1.21 \pm 0.37 \) SNe for galaxies of type Sbc-d, where 1 SNe = 1 SN (100 yr)^{-1} \( (10^5 L_B/L_\odot)^{-1} \), and \( L_B \) is the blue luminosity of the Galaxy. Taking the Milky Way as lying between Sb and Sbc, we estimate \( \sigma_{SN} \approx 1 \) SNe. Using a Galactic luminosity of \( L_B = 2 \times 10^{10} L_\odot \), we find a supernova rate of \( (50 \text{ yr})^{-1} \), which agrees well with the estimate in Eq. (A4) of McKee (1989). If we use the same scale height \( H_c \) and star-forming radius \( R_{sf} \), as above, we can compute the energy input rate from supernova explosions with energy \( E_{SN} = 10^{51} \text{ erg} \) to be

\[
\dot{e} = \frac{\sigma_{SN} \eta_{SN} E_{SN}}{\pi R_{sf}^2 H_c} = (3 \times 10^{-26} \text{ erg s}^{-1} \text{ cm}^{-3}) \left( \frac{\eta_{SN}}{0.1} \right) \left( \frac{\sigma_{SN}}{1 \text{ SNe}} \right) \left( \frac{100 \text{ pc}}{H_c} \right)^{-1} \left( \frac{15 \text{ kpc}}{R_{sf}} \right)^{-2} \left( \frac{E_{SN}}{10^{51} \text{ erg}} \right).
\]

The efficiency of energy transfer from supernova blast waves to the interstellar gas \( \eta_{SN} \) depends on the strength of radiative cooling in the initial shock, which will be much stronger in the absence of a surrounding superbubble (Heiles, 1990). Substantial amounts of energy can escape in the vertical direction in superbubbles as well, however. Norman and Ferrara (1996) make an analytic estimate of the effectiveness of driving by supernova remnants and superbubbles. The scaling factor \( \eta_{SN} = 0.1 \) used here was derived by Thornton et al. (1998) from detailed, 1D, numerical simulations of supernova expanding in a uniform ISM. It can alternatively be drawn from momentum conservation arguments [Eq. (48)], comparing a typical expansion velocity of 100 km s^{-1} to typical interstellar turbulence velocity of 10 km s^{-1}. Multidimensional models of the interactions of multiple supernova remnants (Avillez, 2000) are required to better determine the effective scaling factor.

Supernova driving appears to be powerful enough to maintain the turbulence even with the dissipation rates estimated in Eq. (42). It provides a large-scale self-regulation mechanism for star formation in disks with sufficient gas density to collapse despite the velocity dispersion produced by the magnetorotational instability.

As star formation increases in such galaxies, the number of OB stars increases, ultimately increasing the supernova rate and thus the velocity dispersion, which restrains further star formation.

D. Applications

The theory of star formation controlled by supersonic turbulence offers a unified approach to a wide range of astrophysical objects. In this section we discuss several illustrative scenarios, moving from low to high star formation efficiencies.

1. Low-surface-brightness galaxies

Low-surface-brightness galaxies have large fractions of their baryonic mass in gas, whether they have masses typical of massive (Schombert et al., 1992; McGaugh and de Blok, 1997) or dwarf galaxies (Schombert, McGaugh, and Eder, 2001). Nevertheless, their star formation rates lie well below typical values for high-surface-brightness galaxies (van der Hulst et al., 1993; McGaugh and de Blok, 1997). Their rotation curves have been derived from both \( \text{H}^\alpha \) measurements (van der Hulst et al., 1993; de Blok, McGaugh, and van der Hulst, 1996), and higher-resolution \( \text{H}^\alpha \) measurements (Swaters, Madore, and Trenthall, 2000; McGaugh, Rubin, and de Blok, 2001; Matthews and Gallagher, 2002), which may sometimes disagree with the \( \text{H}^\alpha \) in the innermost regions (Swaters et al., 2000), but are in generally good agreement (McGaugh et al., 2001). They have lower gas and stellar surface densities than high-surface-brightness galaxies (van der Hulst et al., 1987; de Blok and McGaugh, 1996). The question of whether their disks have surface densities lying below the Kennicutt (1989) threshold for star formation has been studied using rotation curves derived from \( \text{H}^\alpha \) measurements for both massive (van der Hulst et al., 1993) and dwarf (van Zee et al., 1997) galaxies.

In the case of massive galaxies, surface densities beneath the Kennicutt (1989) threshold do indeed appear to explain the lack of star formation (van der Hulst et al., 1993). The moderate levels of turbulence required to maintain the observed velocity dispersions may be produced by magnetorotational instabilities (Sellwood and Balbus, 1999). Other explanations for the lack of star formation, such as an inability to form molecular hydrogen (Gerritsen and de Blok, 1999) or to cool it (Mihos, Spaans, and McGaugh, 1999), were derived from numerical models that did not include magnetic effects and thus had no source of support other than thermal pressure to counteract gravitational collapse and star formation. If magnetorotational instability is the dominant support mechanism, then star formation will not be suppressed in the center, where the rotational shear drops. This is, in fact, where star formation is found in low-surface-brightness galaxies.

In the case of dwarf galaxies (Hunter, 1997), the situation appears to be slightly more complex. On the one hand, van Zee et al. (1997) demonstrate that the surface density in a sample of low-surface-brightness dwarf gal-

Rev. Mod. Phys., Vol. 76, No. 1, January 2004
axies falls systematically below the Kennicutt threshold, with star formation indeed observed in regions that approach the threshold, and van Zee, Skillman, and Salzer (1998) show that blue compact dwarf galaxies have surface densities exceeding the threshold in their centers, where star formation occurs. On the other hand, Hunter, Elmegreen, and Baker (1998) argue that a criterion based on local shear correlates better with the observations, especially in galaxies with rising rotation curves. Furthermore, another factor that may contribute to the star formation histories of dwarf galaxies is that starbursts in the smaller ones (under $10^8 M_\odot$) can actually push all the gas well out into the halo, from whence it will take some hundreds of millions of years to collect back in the center (Mac Low and Ferrara, 1999). This last scenario may be consistent with observations in some galaxies, as summarized by Simpson and Gottesman (2000).

2. Galactic disks

In normal galactic disks, where supernovae appear to dominate the driving of the turbulence, we speculate that most regions will have a star formation rate just sufficient to produce turbulence that can balance the local surface density in a self-regulating fashion. However, as spiral arms or other dynamical features increase the local density, this balance would fail, leading to higher local star formation rates. Because the increase in star formation rate as turbulence is overwhelmed is not sudden but gradual, the enhanced star formation in spiral arms and similar structures should not globally approach starburst rates except when density rises sharply. Locally, however, even relatively small regions can reach starburstlike star formation efficiencies if they exceed the local threshold for turbulent support and begin to collapse freely. A classic example of this is the massive star formation region NGC 3603, which locally resembles a starburst knot, even though the Milky Way globally does not have a large star formation rate. On a smaller scale, even the Trapezium cluster in Orion seems to have formed with an efficiency of $\lesssim 50\%$ (Hillenbrand and Hartmann, 1998).

Triggering of star formation by compressive shocks from nearby star-forming regions (Elmegreen and Lada, 1977) is a special case of global support from turbulence leading to local collapse. Although prompt blast waves from winds and early supernovae of OB association can compress nearby gas and induce collapse, most of the energy from that association is released at later times as the less massive B stars explode, driving the larger-scale interstellar turbulence that provides support against general collapse. The instances of apparent triggering seen both in linear sequences of OB associations (see, for example, Blaauw, 1964) and in shells (e.g., Walborn and Parker, 1992; Efremov and Elmegreen, 1998b; Kamaya, 1998; Barbá et al., 2003) may represent this prompt triggering. It seems unlikely, however, that this prompt triggering will dominate large-scale star formation as first suggested by Gerola and Seiden (1978) and since developed by Neukirch and Feitzinger (1988), Korchagin et al. (1995), and Nomura and Kamaya (2001). Compression due to supersonic turbulent flows is suggested to be the main mechanism leading to stellar birth in gas-rich dwarf galaxies without spiral density waves, such as Holmberg II (Stewart et al., 2000). However, the rate of compression-induced star formation is small compared to the rates expected for global collapse, which is effectively prevented by the same turbulent flows.

3. Globular clusters

Globular clusters may simply be the upper end of the range of normal cluster formation. Whitmore (2003) reviews evidence showing that young clusters have a power-law distribution reaching up to globular cluster mass ranges. The luminosity function for old globular clusters is log normal, which Fall and Zhang (2001) attribute to the evaporation of smaller clusters by two-body relaxation and the destruction of the largest clusters by dynamical interactions with the background galaxy (also see Vesperini, 2000, 2001). Fall and Zhang (2001) suggest that the power-law distribution of young clusters is related to the power-law distribution of molecular clouds masses found by Harris and Pudritz (1994). However, numerical models of gravitational collapse tend to produce mass distributions that appear more log-normal, and are not closely related in shape to the underlying mass distributions of density peaks (Klessen et al., 2000; Klessen, 2001b). It remains unknown whether cluster masses are determined by the same processes as the masses of individual collapsing objects, but the simulations do not include any physics that would limit them to one scale and not the other. Further investigation of this question will be interesting.

4. Galactic nuclei

In galaxies with low star formation rates, the galactic nucleus is often the only region with substantial star formation occurring. As rotation curves approach solid body in the centers of galaxies, magnetorotational instabilities die away, leaving less turbulent support and perhaps greater opportunity for star formation. In more massive galaxies, gas is often funneled towards the center by bars and other disk instabilities, again increasing the local density sufficiently to overwhelm local turbulence and drive star formation.

Hunter et al. (1998) and Schaye (2002) note that central regions of galaxies have normal star formation despite having surface densities that appear to be stable according to the Toomre criterion. This could be due to reduced turbulence in these regions decreasing the surface density required for efficient star formation. The radial dependence of the velocity dispersion is difficult to determine, because H I observations with sufficient velocity resolution to measure typical turbulent linewidths of $6 - 12$ km s$^{-1}$ have rather low spatial resolution, with just a few beams across the galaxy. Most calculations of the critical surface density, therefore,
assume a constant value of the turbulent velocity dispersion, which may well be incorrect (Wong and Blitz, 2002).

As an alternative, or perhaps additional explanation, Kim and Ostriker (2001) point out that the magneto-Italiana instability acts strongly in the centers of galaxies. The magnetic tension from strong magnetic fields can reduce or eliminate the stabilizing effects from Coriolis forces in these low-shear regions, effectively reducing the problem to a 2D Jeans stability problem along the field lines.

5. Primordial dwarfs

In the complete absence of metals, cooling becomes much more difficult. Thermal pressure supports gas that accumulates in dark-matter halos until the local Jeans mass is exceeded. The first objects that can collapse are the ones that can cool from H\textsubscript{2} formation through gas phase reactions. Abel, Bryan, and Norman (2000, 2002) and Bromm, Coppi, and Larson (1999) have computed models of the collapse of these first objects. Abel et al. (2000, 2002) used realistic cosmological initial conditions and found that inevitably a single star formed at the highest density peak before substantial collapse had occurred elsewhere in the galaxy. Bromm et al. (1999) used a flat-top density perturbation that was able to fragment in many places simultaneously, due to its artificial symmetry.

Li, Klessen, and Mac Low (2003) suggest that the lack of fragmentation seen by Abel et al. (2000, 2002) may be due to the relatively stiff equation of state of metal-free gas. Li et al. (2003) found that fragmentation of gravitationally collapsing gas is strongly influenced by the polytropic index γ of the gas, with fragmentation continuously decreasing from γ\sim 0.2 to γ\sim 1.3. The limited cooling available to primordial gas even with significant molecular fraction may raise its polytropic index sufficiently to suppress fragmentation. Abel et al. (2000, 2002) argue that the resulting stars are likely to have masses exceeding 100M\odot, leading to prompt supernova explosions with accompanying metal pollution and radiative dissociation of H\textsubscript{2}.

6. Starburst galaxies

Starburst galaxies convert gas into stars at such enormous rates that the time scale to exhaust the available material becomes short compared to the age of the universe (see the review by Sanders and Mirabel, 1996). Starbursts typically last for a few tens or hundreds of millions of years. However, they may occur several times during the life of a galaxy. The star formation rates in starburst galaxies can be as high as 1000M\odot yr\textsuperscript{-1} (Kennicutt, 1998b), some three orders of magnitude above the current rate of the Milky Way. Starburst galaxies are rare in the local universe, but rapidly increase in frequency at larger lookback times, suggesting that starbursts are characteristic of early galaxy evolution at high redshifts. The strongest starbursts occur in galactic nuclei or circumnuclear regions.

However, in interacting galaxies, strong star formation is also triggered far away from the nucleus in the overlapping regions, in spiral arms, or sometimes even in tidal tails. In these interactions a significant number of super star clusters form, which may be the progenitors of present-day globular clusters (Whitmore et al., 1999; Zhang and Fall, 1999), or even compact elliptical galaxies (Fellhauser and Kroupa, 2002). The Antennae galaxy, the product of a major merger of the spiral galaxies NGC 4038 and 4039, is a famous example where star formation is most intense in the overlap region between the two galaxies (Whitmore and Schweizer, 1995). Merging events always seem to be associated with the most massive and luminous starburst galaxies, the ultraluminous IR galaxies identified by Sanders and Mirabel (1996).

Gentler minor mergers can also trigger starbursts. Such an event disturbs but does not disrupt the primary galaxy. It recovers from the interaction without dramatic changes in its overall morphology. This could explain the origin of lower-mass, luminous, blue, compact galaxies, which often show very little or no sign of interaction (see, for example, van Zee, Salzer, and Skillman, 2001). Alternative triggers of the starburst phenomenon that have been suggested for these galaxies include bar instabilities in the galactic disk (Shlosman, Begelman, and Frank, 1990) or the compressional effects of multiple supernovae and winds from massive stars (Heckman, Armus, and Miley, 1990), which then would lead to a very localized burst of star formation.

Regardless of the details of the different starburst triggering mechanisms, they all focus gas into a concentrated region quickly enough to overwhelm the local turbulence and any additional turbulence driven by newly formed stars. Combes (2001) argues that this can only be accomplished by gravitational torques on the gas. We suggest that starburst galaxies are just extreme examples of the continuum of star formation phenomena, with gravity overwhelming support from turbulent gas motions on kiloparsec scales rather than the parsec scales of individual OB associations, or the even smaller scales of low-mass star formation.

VII. CONCLUSIONS

A. Summary

The formation of stars represents the triumph of gravity over a succession of opponents. These include thermal pressure, turbulent flows, magnetic flux, and angular momentum. For several decades, magnetic fields were thought to dominate the resistance against gravity, with star formation occurring quasistatically as ambipolar diffusion allows collapse of neutral gas towards the center of magnetically supported cores. However, a growing body of observational evidence suggests that when stars do form they do so quickly and dynamically, with gravitational collapse occurring at a rate controlled by supersonic turbulence driven at scales of order a hundred parsecs. Such turbulence can explain the superthermal
linewidths and self-similar structure observed in star-forming clouds, while magnetic fields fail to do so. The varying balance between turbulence and gravity then provides a natural explanation for the widely varying star formation rates seen at both cloud and galactic scales.

Scattered, inefficient star formation is a signpost of turbulent support, while clustered, efficient star formation occurs in regions lacking support. In this picture, gravity has already won in observed dense protostellar cores: dynamical collapse seems to explain their observed properties better than the alternatives. On the other hand, molecular clouds as a whole may be able to form from turbulent compression, rather than being dominated by self-gravity. The mass distribution of stars then depends at least partly on the density and velocity structure resulting from the turbulence, perhaps explaining the apparent local variations of the stellar initial mass function despite its broad universality.

We began by summarizing observations of the structure and properties of molecular clouds and the mass distribution of stars that form in them in Sec. II. Observations of self-similar structure in molecular clouds (Sects. II.A and II.B) seem to indicate that interstellar turbulence is driven on scales substantially larger than the clouds themselves (see also Sec. VI.C). Molecular clouds appear actually to be transient objects with lifetimes of several million years that form and dissolve in the larger-scale turbulent flow. Some well-known descriptions of the clouds like Larson’s (1981) size-linewidth relation may be natural consequences of the turbulent gas flow observed in projection (Sec. II.D). However, Larson’s mass-radius relation appears to be an observational artifact, suggesting that most molecular clouds are not in virial equilibrium.

In Sec. III we gave a historical overview of our understanding of star formation, beginning with the classical dynamical theory of star formation (Sec. III.A), which already included turbulent flows, but only in the microturbulent approximation, treating them as an addition to the thermal pressure. We then turned to the development of the standard theory of star formation (Sec. III.C), which was motivated by growing understanding of the importance of the interstellar magnetic field in the 1960s and 1970s, as discussed in Sec. III.B.

The standard theory relies on ion-neutral drift, also known as ambipolar diffusion, to solve the magnetic flux problem for protostellar cores, which until recently were thought to be initially magnetostatically supported. At the same time magnetic tension acts to brake rotating cores, thus solving the angular momentum problem as well. The time scale for ambipolar diffusion to remove enough magnetic flux from the cores for gravitational collapse to set in can exceed the free-fall time by as much as an order of magnitude, suggesting that magnetic support could also explain low observed star formation rates. Finally, magnetic fields were also invoked to explain observed supersonic motions as Alfvén waves.

Recently, however, both observational and theoretical results have begun to cast doubt on the standard theory. In Sec. III.D we summarized theoretical limitations of the singular isothermal sphere model that forms the basis for many of the practical applications of the standard theory. We then discussed several observational findings that put the fundamental assumptions of that theory into question. The observed magnetic-field strengths in molecular cloud cores appear too weak to support against gravitational collapse. At the same time, the infall motions measured around star-forming cores extend too broadly, while the central density profiles of cores are flatter than expected for isothermal spheres. Furthermore, the chemically derived ages of cloud cores are comparable to the free-fall time instead of the much longer ambipolar diffusion time scale. Observations of young stellar objects also appear discordant. Accretion rates appear to decrease rather than remaining constant, far more embedded objects have been detected in cloud cores than predicted, and the spread of stellar ages in young clusters does not approach the ambipolar diffusion time.

New theoretical and numerical studies of turbulence that point beyond the standard theory while looking back to the classical dynamical theory for inspiration have now emerged (Sec. IV). Numerical models show that supersonic turbulence decays rapidly, in roughly a crossing time of the region under consideration, regardless of magnetic-field strength. Under molecular cloud conditions, such turbulence decays in less than a free-fall time. This implies that the turbulence in star-forming clouds needs to be continuously driven in order to maintain the observed motions. Driven turbulence has long been thought capable of supporting gas against gravitational collapse. A numerical test demonstrated that turbulence indeed can offer global support, but at the same time can lead to local collapse on small scales. In strongly compressible turbulence, gravitational collapse occurs in the density enhancements produced by shocks. The rate of local collapse depends strongly on the strength and driving scale of the turbulence. This gives a natural explanation for widely varying star formation rates. Magnetic fields not strong enough to provide static support make a quantitative but not a qualitative difference. They are capable of reducing the collapse rate somewhat, but not of preventing collapse altogether. They may still act to transfer angular momentum so long as they are coupled to the gas, however.

We outlined the shape of the new theory in Sec. IV.K. Rather than relying on quasistatic evolution of magnetostatically supported objects, it suggests that supersonic turbulence controls star formation. Inefficient, isolated star formation is a hallmark of turbulent support, while efficient, clustered star formation occurs in its absence. When stars form, they do so dynamically, collapsing in the local free-fall time. The initial conditions of clusters appear largely determined by the properties of the turbulent gas, as is the rate of mass accretion onto these objects. The balance between turbulent support and local density then determines the star formation rate. Tur-
bulent support is provided by some combination of supernovae and galactic rotation, along with possible contributions from other processes. Local density is determined by galactic dynamics including galaxy interactions, along with the balance between heating and cooling in a region. The initial mass function is at least partly determined by the initial distribution of density resulting from turbulent flows, although a contribution from stellar feedback and interactions with nearby stars cannot yet be ruled out.

We explored the implications of the control of star formation by supersonic turbulence at the scale of individual stars and stellar clusters in Sec. V. We examined how turbulent fragmentation determines the star-forming properties of molecular clouds (Sec. V.A), and then turned to discuss protostellar cores (Sec. II.E), binary stars (Sec. V.C), and stellar clusters (Sec. V.D) in particular. Strongly time-varying protostellar mass growth rates may result as a natural consequence of competitive accretion in nascent embedded clusters (Sec. V.E). Turbulent models predict protostellar mass distributions (Sec. V.F) that appear roughly consistent with the observed stellar mass spectrum (Sec. II.F), although more work needs to be done to arrive at a full understanding of the origin of stellar masses.

The same balance between turbulence and gravity that seems to determine the efficiency of star formation in molecular clouds also works at galactic scales, as we discussed in Sec. VI. The transient nature of molecular clouds suggests that they form and are dispersed in either of two ways. One possibility is that they form during large-scale gravitational collapse and are dispersed quickly thereafter by radiation and supernovae from the resulting violent internal star formation. The other possibility is that large-scale turbulent flows in galactic disks compress and cool gas. These same flows will continue to drive the turbulent motions observed within the clouds. Some combination of turbulent flow, free expansion at the sound speed of the cloud, and dissociating radiation from internal star formation will then be responsible for their destruction on a time scale of 5–10 Myr (Sec. VI.A).

Having considered the formation of molecular clouds from the interstellar gas, we then discussed in Sec. VI.B the role of differential rotation and thermal instability competing and cooperating with turbulence to determine the overall star formation efficiency. We examined the physical mechanisms that could drive the interstellar turbulence, focusing on the energy available from each mechanism in Sec. VI.C. In star-forming regions of disks, supernovae appear to overwhelm all other possibilities. In outer disks and low-surface-brightness galaxies, on the other hand, the situation is not so clear: magnetorotational or gravitational instabilities look most likely to drive the observed flows, but further work is required on these regions. Finally, in Sec. VI.D, we gave examples of how this picture may apply to different types of objects, including low-surface-brightness, normal, and starburst galaxies, as well as galactic nuclei and globular clusters. We argued that efficient star formation occurs at all scales when gravity overwhelms turbulence, with the result ranging from a single low-mass star at the very smallest scale to a starburst at the very largest scale.

B. Future research problems

Although the outline of a new theory of star formation has emerged, it is by no means complete. The ultimate goal of a predictive, quantitative theory of the star formation rate and stellar initial mass function remains elusive. It may be that the problem is intrinsically so complex, like terrestrial climate, that no single solution exists, but only a series of temporary, quasisteady states. Certainly, though, our understanding of the details of the star formation process can be improved. Eventually, coupled models capturing different scales will be necessary to follow the interaction of the turbulent cascade with the thermodynamics, chemistry, and opacity of the gas at different densities. We can identify several major questions that summarize the outstanding problems. As we merely want to summarize these open issues in star formation, we refrain from giving an in-depth discussion and the associated references, which may largely be found in the body of the review.

How can we describe turbulence driven by astrophysical processes? There is really no single driving scale, because of the nonuniformity of explosions and perhaps of other drivers. However, a good description of the structure around the driving scales remains essential, as the largest perturbations lie at the largest scales in any turbulent flow. This description remains to be found. The length of a self-similar turbulent cascade also depends on the scales on which the driving acts. The self-similarity of a turbulent cascade is further perturbed by the drastic changes in the equation of state that occur as increasing densities lead first to stronger radiative cooling and then to the reduction of heating by the exclusion of first ionizing radiation, and then cosmic rays. Finally, at small scales, diffusion and dissipation mechanisms determine the structure. Although ambipolar diffusion probably limits the production of small-scale magnetic-field structures, there is increasing theoretical support for additional density and velocity structure at scales below the ambipolar diffusion cutoff, whose interaction with self-gravity needs to be investigated.

What determines the masses of individual stars? One factor must be the size of the initial reservoirs of collapsing gas, determined by turbulent fragmentation in the complex flow just described. Subsequent accretion from the turbulent gas, perhaps in competition with other stars, or even by collisions between either protostellar cores or stars, could also be important, but must still be shown to occur, especially in a magnetized medium. The properties of protostellar objects depend on the time history of the accretion. Feedback from the newly formed star itself, or from its neighbors, in the form of radiation pressure, ionizing radiation, or stellar winds and jets, may yet prove to be another bounding term on stellar mass.
At what scales does the conservation of angular momentum and magnetic flux fail? That they must fail is clear from the vast discrepancy between galactic and stellar values. Protostellar jets almost certainly form when magnetic fields redistribute angular momentum away from accreting gas. This demonstrates that the conservation of flux and angular momentum must be coupled at least at small scales. However, the observational hint that molecular cloud cores may be lacking substantial flux from the galactic value suggests that flux may already be lost at rather large scales and low densities. Conversely, the prevalence of binary stars suggests that magnetic braking cannot be completely efficient at draining angular momentum from collapsing protostars, and indicates that ambipolar diffusion or some other process limits the effectiveness of braking. This has not yet been modeled.

What determines the initial conditions of stellar groups and clusters? The spatial distribution, initial velocity dispersion, and binary distribution of stars of different masses in a stellar cluster or association are all determined at least partly by the properties of the turbulent flow from which the stars formed. A quantitative analytic model for the retardation of collapse by hydrodynamical or MHD turbulence remains needed. It further remains unknown how much the final properties of a stellar group or cluster depend on the initial state of the turbulence and how much they depend on the properties of gravitationally collapsing gas. The influence of magnetic fields on these properties also remains almost unexplored, although the ability of the field to redistribute angular momentum suggests that they must play at least some role.

What controls the distribution and metallicity of gas in star-forming galaxies? At the largest scale, gas follows the potential of a galaxy just as do all its other constituents. The dissipative nature of gas can allow it to quickly shed angular momentum in disturbed potentials and fall to the centers of galaxies, triggering starbursts. Even in normal galaxies, gravitational instability may determine the location of the largest concentrations of gas available for star formation. How important is turbulence in determining the location and properties of molecular clouds formed from that gas? Are the molecular clouds destroyed again by the same turbulent flow that created them, or do they decouple from the flow, only to be destroyed by star formation within them? How slowly do turbulent flows mix chemical inhomogeneities, and can the scatter of metallicities apparent in stars of apparently equal age be explained by the process?

Where and how fast do stars form in galaxies? The existence of the empirical Schmidt law relating gas column density to star formation rate, with a threshold at low column density, still needs to be definitively explained. Can the threshold be caused by a universal minimum level of turbulence or by a minimum column density, below which it is difficult for gas to cool? In either case, examination of low-metallicity and dwarf galaxies may well provide examples of objects sufficiently different from massive disk galaxies in both cooling and rotation to demonstrate one or the other of these possibilities.

What determines the star formation efficiency of galaxies? The relative importance of turbulence, rotation, gravitational instability, and thermal instability remains unresolved. At this scale, turbulence can only play an instrumental role, transmitting the influence of whatever drives it to the interstellar gas. One possibility is that galaxies are essentially self-regulated, with supernovae from recent star formation determining the level of turbulence and thus the ongoing star formation rate. Another possibility is that a thermal or rotational bottleneck to star formation exists, and that galaxies actually form stars just as fast as they are able, more or less regardless of the strength of the turbulence in most reasonable regimes. Finding observational and theoretical means to distinguish between these scenarios represents the great challenge for understanding the large-scale behavior of star formation in galaxies.

ACKNOWLEDGMENTS

We have benefited from long-term collaborations, discussions, and exchange of ideas and results with a large number of people. We particularly mention (in alphabetical order) M. A. de Avillez, J. Ballesteros-Paredes, P. Bodenheimer, A. Burkert, B. G. Elmegreen, C. Gammie, L. Hartmann, F. Heitsch, P. Kroupa, D. N. C. Lin, C. F. McKee, A. Nordlund, V. Ossenkopf, E. C. Ostriker, P. Padoan, F. H. Shu, M. D. Smith, J. M. Stone, E. Vázquez-Semadeni, H. Zinnecker, and E. G. Zweibel. This review also benefited from two detailed anonymous reviews and extended comments from E. Falgarone, F. Heitsch, M. K. R. Joung, J. Krogli, A. Lazarian, J. M. D. Smith, and E. Vázquez-Semadeni. Finally, we thank V. Trimble for her encouragement. M.-M.M.L. was supported by CAREER Grant No. AST 99-85392 from the US National Science Foundation, and by the US National Aeronautics and Space Administration (NASA) Astrophysics Theory Program under Grant No. NAG5-10103. R.S.K. was supported by the Emmy Noether Program of the Deutsche Forschungsgemeinschaft (Grant No. KL1358/1) and by the NASA Astrophysics Theory Program through the Center for Star Formation Studies at NASA's Ames Research Center, UC Berkeley, and UC Santa Cruz. In preparation of this work we have made extensive use of the NASA Astrophysical Data System Abstract Service.

REFERENCES
