Star Formation in the Galaxy: An Observational Overview

Charles J. LADA*

Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA

The problems of star and planet formation are among the most important challenges facing modern astrophysical research. Stars and their accompanying planetary systems are continuously being formed in the Galaxy enabling direct observation and investigation of the star forming process. However, stars form invisibly deep within cold and dark molecular clouds. Observations of these stellar birth sites at infrared and millimeter wavelengths from space and the ground have resulted in considerable progress toward a physical understanding of stellar origins. In this contribution I will review the empirical basis for our current understanding of the process of star formation with an emphasis on the origin of low mass (sunlike) stars.

§1. Introduction

Stars are the fundamental objects of the Cosmos. They are the homes of planetary systems and they provide the energy necessary for the development and maintenance of life. The evolution of stars drives the evolution of all stellar systems including clusters and galaxies. Stellar Evolution also controls the chemical evolution of the universe. The physical natures and life histories of stars are well described by the theory of stellar structure and evolution, one of the great achievements of 20th century astronomy. According to this theory stars are self-gravitating balls of (mostly) hydrogen gas that act as thermonuclear furnaces to convert the primary product of the big bang (hydrogen) into the heavier elements of the periodic table. For most of its life a star maintains a state of stable equilibrium in which the inward force of gravity is balanced by the outward force of pressure. This internal pressure is generated by the energy released in the nuclear reactions which burn hydrogen at the star’s center. These nuclear reactions are also the source of the star’s luminosity. During the time the star burns hydrogen in its core it maintains a fixed radius and luminosity and consequently a constant surface temperature. The exact value of the equilibrium radius, luminosity and surface temperature of a star depends almost exclusively on one parameter, the star’s mass. The more massive the star, the greater its luminosity, size and surface temperature. Together, hydrogen burning stars of varying mass form a well defined locus of points in the observable luminosity-effective temperature plane (the HR diagram). This locus of points is called the main sequence and the hydrogen burning phase of a star’s life is known as the main sequence phase. Main sequence stars range in mass from about 0.08 to 100 M⊙ (solar masses). Stars with smaller masses have insufficient weight to raise their central temperatures enough to enable hydrogen fusion. (Such objects are referred to as brown dwarfs.) Stars with larger masses are presumably too luminous to hold on to their outer atmospheres.

* E-mail: clada@cfa.harvard.edu.
Once a star exhausts its supply of hydrogen in its central core, nuclear reactions there plummet and the helium-rich core contracts unable to support itself against gravity. The contraction of the core releases large amounts of gravitational potential energy that cause increased hydrogen burning in a shell around the core. For a solar type star this eventually leads to the expansion of the star’s outer layers which is accompanied by a drastic increase in the star’s luminosity. Thus begins the post-main sequence phase of stellar evolution during which the star evolves away from the main sequence while keeping its temperature roughly constant in a manner prescribed by Hayashi (Hayasi & Hoshi 1961). The inner core both gains mass (from the continual production of helium in the hydrogen burning shell) and contracts until fusion reactions involving helium restore a temporary equilibrium in which gravity is again balanced by internal pressure. After this phase such stars will eject their outer atmospheres producing planetary nebula and leaving behind a white dwarf stellar remnant. More massive stars experience a more complex post-main sequence evolution during which they will fuse heavier elements in their cores and eventually end their lives in catastrophic supernova explosions which violently eject much of their original mass, enriched by heavy elements, into the interstellar medium leaving behind exotic remnants, namely neutron stars and black holes.

Despite its spectacular success in explaining the life histories and deaths of stars, the theory of stellar evolution is incomplete in a very fundamental aspect. It is not able to account for the origin of stars. Knowledge of the physical mechanism for the formation of stars is essential for understanding the evolution of the galaxies and the universe from the earliest times after the big bang to the current epoch of cosmic history. Development of a theory of star formation is also crucial for understanding the origin of planetary systems which, in turn, is important for evaluating the possibility of biology beyond the solar system.

The inability of the theory of stellar evolution to explain star formation likely points to the inherent complexity of the physical process itself. Consequently construction of a theory of star formation must require a strong foundation of empirical data or observation. The empirical study of star formation is greatly facilitated by a fundamental property of the universe. Namely that star formation has been a continuous and ongoing process which in our galaxy extends into the present epoch. Consequently, the physical process of star formation can be investigated by direct observation. The realization of ongoing star formation in our galaxy was an important milestone in twentieth century astronomy and as such merits some further discussion. By the middle of the twentieth century the theory of stellar structure and evolution had demonstrated that certain luminous stars, OB stars, burn their nuclear fuel at such prodigious rates that they can live for only a small fraction of the lifetime of our galaxy. The very existence of such stars clearly indicated that star formation has occurred in the present epoch of Galactic history. In 1947 the Armenian astronomer V. A. Ambartsumian showed that such OB stars were almost always members of stellar groupings he termed OB Associations. The space densities of stars in OB associations were well below the threshold necessary to prevent their disruption by Galactic tidal forces. Ambartsumian calculated dynamical ages for the associations that were much less than the age of the galaxy. These dynamical
ages turned out to be in good agreement with the nuclear ages of the stars and independently provided evidence that star formation is still an active process in the Galaxy. The discovery of the interstellar medium of gas and dust during the early part of the twentieth century provided a crucial piece of corroborating evidence in support of the concept of present epoch Galactic star formation. Subsequent observations of interstellar material established that clouds of interstellar gas and dust had roughly stellar composition and were considerably more massive than a single star or group of stars. This revealed that the raw material to make new stars was relatively abundant in the Galaxy. These three pieces of evidence, 1) stellar evolution theory, 2) expanding OB associations and 3) the interstellar medium, constitute three basic “proofs” of ongoing star formation in the Milky Way. This concept was further bolstered by the work of Walker (1954) who found a large population of low mass stars to lie above the main sequence in the HR diagram of the young cluster NGC 2264, consistent with the predictions of Henyey et al. (1955) and later Hayashi (1961) who showed that the locations of these stars in the HR diagram indicated that they were contracting pre-main sequence stars that had not yet initiated fusion reactions in their cores.

For much of the last 50 years direct observation of the star formation process and the development of a theory to explain it, have been severely hampered by the fact that most stars form in dark clouds and during their formative stages are invisible optically. Fortunately, advances in observational technology over the last quarter century opened the infrared and millimeter-wave windows to astronomical investigation and enabled direct observations of star forming regions and this has significantly expanded our knowledge of the star formation process. As a result the foundations for a coherent theory of star formation and early evolution are being laid.

§2. Giant clouds and dense cores: The cradles of star formation

2.1. The basic properties of GMCs

When observed at a dark site on a clear moonless night the Milky Way is truly an impressive sight. One of its most prominent characteristics is that it is split down the middle by a dark obscuring band. The band consists of the superposition of many interstellar dark clouds which contain tiny opaque dust grains that very effectively absorb and scatter the background starlight. These dark clouds are the sites of star and planet formation in the Galaxy. Millimeter-wave observations in the 1970s demonstrated that these clouds were primarily molecular clouds made up of very cold molecular hydrogen gas (H2). Indeed with temperatures of between 10–50 K, these objects are the coldest objects in the universe. Early molecular-line surveys performed with the CO molecule (the most easily detected species in such clouds) showed that the molecular mass of the Galaxy was about $2 \times 10^9 M_\odot$, concentrated tightly within the galactic plane and dominated by Giant Molecular Clouds (GMCs) (e.g., Clemens et al. 1988).

The physical properties of GMCs have been elucidated in many reviews (e.g.,
Blitz 1991; 1993; McKee 1999; Evans 1999) and here I only will briefly review their primary characteristics. For more details the reader is referred to the references cited above. With sizes ranging from 20–100 pc, these objects are the largest structures in the Milky Way and with masses between $10^4$–$10^6 \, M_\odot$ they rival globular clusters as the most massive objects in the Galaxy. The mean densities, $\rho$, of GMCs are approximately $50$–$100 \, cm^{-3}$. Given their cold temperatures this corresponds to a Jeans mass, $M_J \approx 20 \, M_\odot$, so GMCs contain thousands of Jeans masses and are clearly gravitationally bound if not highly unstable entities. Measurements of molecular linewidths indicate cloud velocity dispersions of $2$–$3 \, km \, s^{-1}$. The velocity fields of these objects are highly supersonic given the cloud sound speeds, $c_s \sim 0.2 \, km \, s^{-1}$. These supersonic motions for the most part are not due to systematic mass motions (e.g., collapse or expansion) of the cloudy material, rather they appear to be turbulent in nature. For example, the clouds obey a linewidth-size relation such that $\Delta v \sim R^{-p}$, where $p \approx .38$, similar to the $0.33$ expected for incompressible turbulent flows (Larson 1981). The clouds are also permeated by interstellar magnetic fields with measured strengths of order $10 \, \mu G$ corresponding to A1fven speeds of $1$–$2 \, km \, s^{-1}$, comparable to the observed linewidths. Given the magnetic field strength and density of GMCs their magnetically critical mass is $M_B \sim 5 \times 10^5 \, M_\odot$ (McKee 1999). So although GMCs contain many (thermal) Jeans masses, their magnetic fields appear sufficiently strong to support them against catastrophic collapse.

The ages of GMCs are very uncertain and the subject of long term controversy and discussion (e.g., Blitz & Shu 1980). From association with young clusters the observations of Leisawitz et al. (1989) suggest an age of $5$–$10 \, Myr$ for GMCs and this is likely the best estimate available, although other workers have suggested ages as small as $2$–$3 \, Myr$ (Hartmann 2001) and as large as $100 \, Myr$ (Scoville & Hersh 1979). The timescale for gravitational collapse of a cloud is the free-fall time, $\tau_{ff} = \sqrt{3\pi/32G\rho}$, which for typical parameters of GMCs is about $2$–$4 \, Myr$. The best estimates of cloud lifetimes suggests that cloud ages can be either comparable to or up to five times greater than their collapse times. Clearly knowing which of these alternatives is the case is of critical importance to understanding the evolution of such complexes.

2.2. The dense cores

One of the most important physical characteristics of GMCs is that they are highly structured. In particular, they contain dense gas in the form of identifiable cores which are more or less discrete entities. The mean density of these dense cores is typically $\sim 10^4 \, cm^{-3}$ but the peak densities within them can range to values as high as $10^6 \, cm^{-3}$. The fraction of GMC mass at densities $> 10^4 \, cm^{-3}$ is found to range between roughly $1$–$10\%$ (e.g. Lombardi & Alves 2005, Kato et al. 1999; Lada et al. 1991). The variation in this fraction is possibly due to the evolutionary state of the cloud, young clouds with little star formation have a small fraction of dense gas while more evolved clouds with active star formation have a relatively high fraction (Kato et al. 1999; Blitz & Williams 1999).

The masses of dense cores range from $\sim 1$–$1000 \, M_\odot$. The frequency distribution of these dense core masses in molecular clouds appears to follow a power-law such
that $dn/dm \sim m^{-1.5}$ (e.g., Lada et al. 1991; Nozawa et al. 1991; Williams et al. 1994; Hara et al. 1999, etc.). The slope of this power law relation is such that most of the mass of dense molecular gas in a GMC is tied up in the most massive (i.e., $M \geq 100 \, M_\odot$) cores. The velocity dispersions in the massive dense cores tend to be supersonic, dominated by turbulent motions. However, for the lower mass ($\leq 5 \, M_\odot$) cores the velocity dispersions are significantly smaller and often can characterized by subsonic turbulence (e.g., Myers 1983) and in some cases are even dominated by thermal motions and pressure (e.g., Lada et al. 2003). There is also a tendency for the higher mass cores to be warmer as well as more turbulent than the lower mass cores (Jijina et al. 1999).

§3. Embedded stellar populations: The yield of star formation in GMCs

3.1. The stellar content of GMCs

Knowledge of the content of protostars and young stellar objects (YSOs) within a GMC or any molecular cloud is important for understanding how the mechanism of star formation proceeds within it. Since these objects are often buried deep within the clouds they tend to be invisible optically and only detectable at infrared wavelengths where the opacity of dust is smaller by an order of magnitude or more compared to optical wavelengths. The younger or less evolved the star is, the more heavily it is buried and the more difficult to detect. Getting a census of star forming activity in GMCs requires extensive infrared surveys from both the ground and space.

In the early 1980s the IRAS satellite produced the first sensitive surveys of mid- to far-infrared (i.e. $\lambda > 10 \, \mu m$) emission from GMCs. By the end of the decade the development of imaging array detectors enabled the first thorough near-infrared (1–2 $\mu m$) surveys of these clouds. Together these surveys provided the first systematic inventories of the embedded populations of YSOs within GMCs, that is, the yield of the star formation process in the molecular gas. Three fundamental results pertaining to star formation were revealed by these surveys:

1) - Stars form in dense ($n(H_2) \geq 10^4 \, cm^{-3}$) molecular gas,
2) - Most stars form in embedded clusters, and
3) - The overall star formation efficiency in GMCs is quite low, of order 1–3% or less.

The finding of a low global efficiency for star formation (e.g., Myers et al. 1986; Lada 1987) has significant implications for understanding the nature of the star formation process. It indicates that most of the mass of a molecular cloud does not participate in star formation and must exist in a relatively stable and long-lived (compared to a collapse time) configuration. This would then suggest that the estimates of cloud lifetimes that are in the range of 5–10 Myrs are likely correct. In some sense this shifts the problem of star formation from finding what makes the bulk of the material in a cloud collapse to finding what holds it up (Shu et al. 1987). This is coupled to the first point in that molecular-line and extinction surveys indicate that only a small fraction of the gas in a GMC is at high density. It is in this material that stars form (Lada 1992). The overall efficiency of star formation in dense gas is around 10–20%, considerably higher than the the global star formation
efficiency for GMCs as a whole. Although in an individual core this efficiency can range between 0–30% (e.g., Lada & Lada 2003). As mentioned earlier only about 25% of dense cores in dark clouds contain embedded stars or protostars (Lee & Myers 1999). The highest efficiencies appear to be achieved in the most massive cores which produce embedded clusters (e.g., Lada 1992).

3.2. Embedded clusters in molecular clouds

Given that stars form in dense gas, it is perhaps not surprising that most stars have been found to form in embedded clusters (Lada et al. 1991; Carpenter 2000; Lada & Lada 2003, Porras et al. 2004), since the mass spectrum of dense cores indicates that most of the dense gas is found in the most massive cores. Indeed in the Orion L1630 GMC, which was systematically surveyed for both embedded stars and dense gas, most of the embedded stars were found to confined to three rich clusters occupying only a few % of the area of the cloud and each cluster was in turn located in a massive dense core of similar size. Typically the stars in an embedded cluster are co-extensive with the gas within its associated massive core indicating that the spatial size of the cluster is set by the dimensions of the natal core. Observations of embedded clusters show that in most cases the spatial distribution of the stars within a cluster can be described as being centrally condensed. They are characterized by a dominant central peak in the stellar surface density, as would be expected if gravity dominated the overall formation from the these systems. However, a small but significant number of clusters display a more structured surface density distribution which can sometimes appear hierarchical as if turbulence dominated the formation of these systems (Lada & Lada 2003).

Observations have also shown that embedded clusters span a large range in size, from small groups of 10 or more stars to rich populations containing up to a thousand or more members (Lada & Lada 2003; Porras et al. 2004). Lada & Lada (2003) constructed the mass function of embedded clusters within 2 kpc of the sun and found that it was similar to that of dense cores which perhaps suggests a constant (terminal) star formation efficiency within dense cores. They also found evidence for a truncation at the low mass end indicating that the vast majority of stars form in clusters with 100 members or more. A similar result was found by Porras et al. (2004) who performed a survey for clusters within 1 kpc of the sun that was more complete at the low mass end. The ages of embedded clusters span a relatively narrow range of about 0.5–3 Myr suggesting that the formation time for these objects is quite rapid and that consequently the stars formed in clusters are to a reasonable degree coveal (Hartmann 2001; Moitinho et al. 2001; Preibisch et al. 2001).

The large number of embedded clusters in GMCs is surprising given the relatively small numbers of older open clusters in the field (Lada & Lada 1991). This is illustrated in the frequency distribution of cluster ages for a combined sample of embedded and classical open clusters. Figure 1 shows statistics on cluster ages derived from a combined sample of embedded clusters (Lada & Lada 2003) and classical open clusters (Battinelli & Cappuzzo-Dolcetta 1991) along with the expected age distribution for a constant rate of cluster formation. There is a large and increas-
Fig. 1. Observed frequency distribution of ages for open and embedded clusters within 2 kpc of the Sun (solid line) compared to that (dotted line) predicted for a constant rate of cluster formation adjusted for cluster luminosity evolution. All embedded clusters fall into the first bin. The large discrepancy between the predicted and observed numbers indicates a high infant mortality rate for protoclusters. From Lada & Lada 2003.
formation, new stars will immediately begin to disrupt their surrounding gaseous environments. The birth of high mass stars can be particularly destructive and not only lead to the rapid disruption of a cluster forming core but in addition to the complete dispersal of an entire GMC (e.g., Whitworth 1979). The dynamical response of the stars which are left behind after such explosive gas removal will depend on the star formation efficiency achieved by the core at the moment of gas dispersal. The condition for the cluster to remain bound in the face of rapid gas removal is that the escape speed from the cluster, \( V_{esc} \approx (2GM_{stars}/R)^{0.5} \), is less than \( \sigma \), the instantaneous velocity dispersion of the embedded stars at the time of gas dispersal. Thus a bound group will emerge only if the star formation efficiency is greater than 50% (e.g., Wilking & Lada 1983). Consequently, the fact that the star formation efficiencies of embedded clusters are always observed to be less than about 30%, insures that it is very unlikely for embedded clusters to emerge from molecular clouds as bound entities. The most massive embedded clusters can emerge partially intact and these are likely the precursors of classical open clusters such as the Pleiades (Lada et al. 1984; Kroupa et al. 2001).

Clusters with ages in excess of 100 Myr are very rare. This fact was first emphasized by Oort (1957) and then explained by Spitzer (1958) who demonstrated that tidal encounters with passing interstellar clouds could disrupt all open clusters with mass densities \( \lesssim 1 \text{ M}_\odot \text{pc}^{-3} \) within 200 Myr. In this context, it is interesting to note that both the infant mortality rate and life expectancy of clusters will be a function of distance from the galactic center. This is because both the galactic tidal force and the number of GMCs increases toward the center of the Galaxy, making encounters with GMCs both more frequent and disruptive. Indeed, van den Bergh and McClure (1980) have pointed out that the oldest open clusters are strongly concentrated in the outer galaxy, a fact they attribute to a lower frequency of disruptive encounters with GMCs in the outer regions of the Galaxy. Moreover, Fig. 1 also indicates that the disruption rate for bound clusters between 10–100 Myrs of age is significant, probably due to encounters with GMCs. Many of the observed open clusters in this age range may also not be presently bound (Battinelli & Capuzzo-Dolcetta 1991).

### 3.3. Embedded clusters and the stellar IMF

A fundamental consequence of the theory of stellar structure and evolution is that, once formed, the subsequent life history of a star is essentially predetermined by one parameter, its birth mass. Consequently, the initial distribution of stellar masses at birth (i.e., the IMF) is a fundamental property of all stellar systems. Knowing how this quantity varies through time and space is necessary to predict and understand the evolution of galaxies and clusters. Detailed knowledge of the IMF and its spatial and temporal variations is also particularly important for understanding the process of star formation, since it is the mysterious physics of this process that controls the conversion of interstellar matter into stars. A theory of star formation cannot be considered as complete unless it can predict the form of the stellar IMF. Unfortunately, present day theory is unable to predict the form of the IMF. This quantity must be derived from observations. However, this is not a straightforward exercise, since stellar mass is not itself an observable quantity. Stellar radiant flux or
luminosity is the most readily observed property of a star. Determination of stellar masses therefore requires a transformation of stellar luminosities into stellar masses which in turn requires knowledge of stellar evolutionary states.

Embedded clusters play an important role in IMF studies because they present equidistant, statistically significant and roughly coeval populations of stars of similar chemical composition. In addition, embedded clusters are too young to have lost significant numbers of stars due to stellar evolution or dynamical evaporation, thus their present day mass functions are, to a very good approximation, their initial mass functions. Embedded clusters are also particularly well suited for determining the nature of the IMF for low mass stars and substellar objects. This is because low mass stars in embedded clusters are primarily pre-main sequence stars, and thus are brighter than at any other time in their lives prior to their evolution off the main-sequence. At these young ages, substellar objects or brown dwarfs are also significantly more luminous than at any other time in their subsequent evolution, and moreover have brightnesses comparable to the lowest mass stars. Indeed, infrared observations of modest depth are capable of detecting objects spanning the entire range of stellar mass from 0.01 to 100 $M_\odot$ in clusters within 0.5–1.0 kpc of the sun.

However, a significant disadvantage for embedded clusters is that the member stars are mostly pre-main sequence stars and the timescale for forming them is an appreciable fraction of the cluster age. Consequently, uncertain corrections for pre-main sequence evolution and non-coevality must be applied to the members to derive mass spectra from luminosity functions. To ameliorate this disadvantage numerous researchers have employed various modeling techniques (e.g., Zinnecker et al. 1993; Fletcher and Stahler 1994; Lada & Lada 1995; Muench, Lada & Lada 2000; Muench et al. 2002). Perhaps the best determined IMF is that of the famous Trapezium cluster in Orion. A number of different groups have derived the IMF for this cluster from various modeling its infrared luminosity function (i.e., the frequency distribution of the monochromatic brightness of stars in the cluster). These studies have produced very similar results which are well illustrated by the IMF derived by Muench et al. (2002) who used Monte Carlo modeling techniques to invert the observed luminosity function to the cluster’s mass function with the aid of modern models of pre-main sequence evolution. The resulting IMF is shown in Fig. 2.

The most significant characteristic of this IMF is the broad peak, extending roughly from 0.6 to 0.1 $M_\odot$. This structure clearly demonstrates that there is a characteristic mass produced by the star formation process in Orion. That is, the typical outcome of the star formation process in this cluster is a star with a mass between 0.1 and 0.6 $M_\odot$. The process produces relatively few high mass stars and relatively few substellar objects. Indeed, no more than $\sim 22\%$ of all the objects formed in the cluster are freely floating brown dwarfs. The overall continuity of the IMF from OB stars to low mass stars and across the hydrogen burning limit strongly suggests that the star formation process has no knowledge of the physics of hydrogen burning. Substellar objects are produced naturally as part of the same physical process that produces OB stars (see also Najita, Tiede & Carr 2000; Muench et al. 2001).

In this respect the secondary peak at 0.015 $M_\odot$ is intriguing. The existence of
Fig. 2. The IMF derived for the Trapezium cluster from Monte Carlo modeling of its luminosity function (Muench et al. 2002). This plot displays the binned mass function of the synthetic cluster whose luminosity function was found to best fit the observed KLF of the Trapezium cluster. A vertical dashed line marks the approximate location of the hydrogen burning limit. The derived IMF displays a broad peak between 0.1–0.6 $M_\odot$ and extends deep into the substellar mass regime. The secondary peak is located near 0.015 $M_\odot$ or 15 $M_J$ and may be an artifact of the adopted substellar mass-luminosity relation.

such a peak may imply a secondary formation mechanism for the lower mass brown dwarfs, similar to suggestions recently advanced by Reipurth and Clarke (2000) and thus is potentially very important. However, the significance that should be attached to this feature depends on the accuracy of the adopted mass-luminosity relations for substellar objects used in the modeling. These mass-luminosity relations may be considerably more uncertain than those of pre-main sequence stars and it is quite plausible that the secondary feature in the derived IMF is artificial and does not represent a true feature in the underlying IMF. However, a recent analysis of the IMF of the Trapezium cluster based on analysis of the spectra of its fainter members finds this secondary feature to persist (Slesnick et al. 2004). In addition, the same analysis suggests that the overall number of substellar objects in the IMF has been overestimated in the Muench et al. analysis by their assumption of young ages for all the stars observed in the direction of the Trapezium. The spectroscopic observations of Slesnick et al. (2004) indicate that a significant fraction, perhaps as
much as 1/3, of the faintest objects appear to be hotter than implied by their low luminosities, suggestive of older ages and more stellar-like masses for these objects. Clearly more data, both observational and theoretical is needed to assess the reality and significance of this intriguing feature and precise shape of the IMF at substellar masses.

The derived IMF of the Trapezium cluster spans a significantly greater range of mass than any previous IMF determination whether for field stars or other clusters (e.g., Kroupa 2002). Its statistically meaningful extension to substellar masses and the clear demonstration of a turnover near the hydrogen burning limit represents an important advance in IMF studies. For masses in excess of the hydrogen burning limit the IMF for the Trapezium is in good agreement with the most recent determinations for field stars (Kroupa 2002). This is to some extent surprising since the field star IMF represents an average of the IMF over billions of years of galactic history, assuming a constant star formation rate, and over stars originating from very different locations of galactic space. The Trapezium cluster, on the other hand, was formed within the last million years in a region considerably less than a parsec in extent. Taken at face value this agreement suggests that the IMF and the star formation process that produces it are very robust in the disk of the Galaxy. There is evidence that the IMF may be different for stars and clusters formed in environments that differ significantly from that of the galactic disk. In particular observations of the Arches cluster in the galactic center region show that the IMF in this cluster contains significantly higher numbers of massive stars (i.e., $m_*>10\,M_\odot$) than would be expected for an IMF similar to that of the Trapezium or the field (Stolte et al. 2002). The galactic center region is characterized by significantly higher pressures than the disk and this may suggest that extreme environments can alter the form of the IMF produced in the star formation process.

§4. From cores to stars: The physical process of star formation

4.1. The starless cores and initial conditions

The most critical but least understood aspects of the star formation process are those which concern its initial conditions and earliest stages. In particular, little is known about the initial phases of protostellar development within dense molecular cloud cores and even less is known about the origin of the dense cores themselves. Observations have demonstrated that dense cores are the sites of star formation and protostellar evolution. In active star forming clouds like Orion and Taurus a large fraction of the dense cores are associated with a forming protostar or recently formed pre-main sequence star (Jijina et al. 1999). In clouds with low levels of star formation, such as the Coalsack and Pipe Nebula, most cores are starless (e.g., Kato et al. 1999, Onishi et al. 1999). In a general survey of dark nebulae for dense cores, Lee & Myers (1999) found that 75% of the cores they surveyed were starless. The physical parameters of starless cores likely reflect the initial conditions of star formation and so these objects are prime laboratories for investigating the initial conditions and early phases of the star formation process.
Low mass \((m < 15 \, M_\odot)\) dense cores greatly outnumber high mass \((m > 100 \, M_\odot)\) dense cores. As of yet no high mass core has been found to be starless, although in the L1630 GMC two of its five most massive cores exhibit very low levels of star forming activity (Lada 1991). Most starless cores are of the low mass, isolated variety. Submillimeter measurements of dust emission from a sample of starless cores enabled the radial density structure of the cores to be partially resolved and it was found that the cores exhibited density gradients that were relatively flat in their centers, at least compared to cores which had already formed stars (Ward-Thompson et al. 1994). Such observations are suggestive of equilibrium configurations of self-gravitating gas that are bounded by external pressure. Indeed, using measurements of extinction to thousands of individual stars behind the B 68 dark globule, Alves et al. (2001) were able to construct a highly resolved radial density profile of the core and found it to be exquisitely well fit by the predictions of a so-called Bonnor-Ebert (BE) sphere. The extinction map, radial surface density profile and the fit of the Bonnor-Ebert relation that they obtained are shown Fig. 3.

A BE sphere is a pressure-truncated isothermal ball of gas within which internal pressure everywhere precisely balances the inward push of self-gravity and external surface pressure. The fluid equation that describes such a self-gravitating, isothermal sphere in hydrostatic equilibrium is the following well known variant of the Lane-Emden equation:

\[
\frac{1}{\xi^2} \frac{d}{d\xi} \left( \xi^2 \frac{d\psi}{d\xi} \right) = e^{-\psi},
\]

where \(\xi\) is the dimensionless radius: \(\xi = r/r_c\) and \(r_c\) is the characteristic or scale radius, \(r_c = c_s/(4\pi G \rho_0)^{1/2}\), where \(c_s\) is the sound speed in the cloud and \(\rho_0\) is the density at the origin. Equation (2) is Possion’s equation in dimensionless form where \(\psi(\xi)\) is the dimensionless potential and is set by the requirement of hydrostatic
equilibrium to be \( \psi(\xi) = -\ln(\rho/\rho_0) \). The equation can be solved using the boundary conditions that the function \( \psi \) and its first derivative are zero at the origin. Equation (2) has an infinite family of solutions that are characterized by a single parameter, the dimensionless radius at outer edge (\( R \)) of the sphere:

\[
\xi_{\text{max}} = cR/r_c.
\] (4.2)

Each solution thus corresponds to a truncation of the infinite isothermal sphere at a different outer radius, \( R \). The external pressure at a given \( R \) must then be equal to that which would be produced by the weight of material that otherwise would extend from \( R \) to infinity in an infinite isothermal sphere. The shape of the BE density profile for a pressure truncated isothermal cloud therefore depends on the single parameter \( \xi_{\text{max}} \). As it turns out, the higher the value of \( \xi_{\text{max}} \) the more centrally concentrated the cloud is, that is, the higher the center-to-edge density contrast of the core. The stability of such pressure truncated clouds was investigated by Bonnor (1956) and Ebert (1955) who showed that when \( \xi_{\text{max}} > 6.5 \) the clouds are in a state of unstable equilibrium, susceptible to gravitational collapse.

The curve that fit the B68 density profile was characterized by a \( \xi_{\text{max}} = 6.9 \) indicating the cloud was near the critical stability condition. In addition, the excellent match of theory with observations and the precise determination of \( \xi_{\text{max}} \) for the B68 core allowed the physical conditions describing the globule to be determined from Eq. (4.2). With knowledge of the peak extinction through the cloud from the extinction measurements and an independent determination of the gas temperature from molecular-line measurements the equations are over constrained. Thus, the central density, radial density law, center-to-edge density contrast, the mass and the distance to the cloud could be accurately determined. Although we cannot be certain that B68 will collapse and form a star, these conditions likely reflect the initial conditions for star formation in such cores.

It is interesting that measured linewidths of emitting molecules in the cloud are essentially thermal with very little turbulence present. This indicates that thermal pressure dominates the internal pressure of this object (Lada et al. 2003). This, in turn, is consistent with the notion that cores are formed in turbulent clouds and evolve by dissipating their internal turbulence before forming stars (Myers 1983). Moreover, detailed maps of the cloud’s velocity field were consistent with the pattern expected for a non-radial, low mode, pulsation of the cloud’s outer layers (Lada et al. 2003). The existence of subsonic pulsating modes suggests that the cloud must be close to a perturbed equilibrium state, consistent with the value of \( \xi_{\text{max}} \) derived from the density profile. In a study of the velocity fields of a large sample of starless cores Lee et al. (1999) found evidence for the presence of infall motions in about 25% of the cores they observed. These objects may be beginning the collapse to form stars, but the collapse has not yet produced detectable protostellar objects in their centers.

Lee et al. (1999) also found a tendency for starless cores with collapse signatures to be cores with the largest central column densities. Extinction studies of both starless and protostellar cores suggest that the in the process of forming a stars the structure of a core evolves in such a way that its center-to-edge density contrast...
increases with time. That is, when the density profiles of cores are fitted with BE relations, there is a tendency for the derived value of $\xi_{\text{max}}$ for cores with protostars to be significantly higher than those that are starless (Alves et al. 2001; Harvey et al. 2001; Lada et al. 2004; Teixiera et al. 2005) as qualitatively would be expected from the Bonnor-Ebert stability condition. Of course, it is not at all clear whether dense cores embedded within molecular clouds are true Bonnor-Ebert spheres in the sense of an isothermal-like sphere with a sharp boundary. For example, it is operationally difficult to determine the edges of cores or even whether cores have sharp edges when they are surrounded by more extended molecular material within a larger cloud complex. Nonetheless, such dense cores are stratified and exhibit radial density gradients that appear to be correlated with their evolutionary state. On the whole the fragmentary observational evidence suggests that cores form from turbulent gas and gradually dissipate their inner turbulence as they evolve; once they become near thermal they collapse either because of the loss of turbulent pressure support, or from external perturbations or from the loss of magnetic pressure support due to ambipolar diffusion.

4.2. Protostars, outflows and the formation of a star

If an otherwise stable dense core is sufficiently perturbed, or loses internal support, or if the core initially forms with a mass in excess of the Jeans mass, then it will collapse to form a star. The formation of single, isolated, low mass stars is reasonably understood and this understanding is based on both empirical data and a general theory developed by Shu and his collaborators (e.g., Shu 1977; Shu et al. 1987). This picture can be briefly summarized as follows. Before being incorporated into a star, the molecular material within a collapsing core must increase its density by 20 orders of magnitude and collapse to a size nearly 7 orders of magnitude smaller than the original dimensions of the core. Because the dust in the cloud is optically thin, internal energy gained by its collapse is effectively radiated away. Thus the cloud material remains isothermal and dynamically collapses. The collapse also proceeds in a non-homologous manner, with the inner regions becoming denser and collapsing faster than the outer regions which are left behind. Eventually, the innermost infalling material becomes dense enough to be optically thick to its own radiation resulting in the development of a central quasi static stellar core surrounded by an infalling envelope. Thus is a protostar born.

After the formation of its embryonic core, the protostar enters the accretion phase of protostellar evolution. During this time the central stellar core gradually gains mass via the accretion of material from its infalling envelope. Before being finally incorporated onto the growing stellar core, accreting material must dissipate the gravitational potential energy lost in infall, giving rise to an accretion luminosity:

$$L_{\text{accretion}} = \frac{GM_* \dot{M}}{R_*},$$

(4.3)

where $M_*$, and $R_*$ are the mass and radius of the protostellar core and $\dot{M}$, the mass accretion rate. This accretion luminosity can be a significant component of a protostar’s luminosity. The mass accretion rate is given by: $\dot{M} = m_0 a^3/G$ where
\(a\) is the effective sound speed and \(m_0\) is a constant which is sensitive to the initial conditions in the collapsing core, but which must be of order unity to be consistent with the observed luminosities of protostellar objects.

In order for infalling material to be accreted onto a protostellar core the material must first lose significant amounts of angular momentum. Observations indicate that the specific angular momentum of cloud cores is about six orders of magnitude higher than material on the surface of a typical star. Moreover, the tendency for collapsing material to conserve angular momentum will produce a centrifugal barrier that will prevent infalling material from falling directly onto the protostar in the first place. Instead, the infalling material will form a highly flattened structure or disk surrounding the protostellar core. Observations suggest that this disk can range from a few hundred to a few thousand AU in diameter. Initially the mass of the central embryonic stellar core is very small, approximately \(10^{-2} M_\odot\) and the central temperature of the core is not sufficient to ignite nuclear reactions. In order for this embryonic core to grow into a star it will have to accrete material from its surrounding disk. Material in the disk can only accrete onto the star if the material loses angular momentum. Frictional, viscous accretion can result in the net outward transport of angular momentum in a disk, allowing material to move inward (Lynden-Bell and Pringle 1974). As the material moves inward it also loses energy which is radiated away by the disk. For a Keplerian disk the material that reaches the star still has considerably more angular momentum than can be absorbed by the star. Nature must find a way for this material to accrete onto the star’s surface and solve the angular momentum problem.

Nature’s solution, energetic, bipolar outflows, was completely unanticipated by theory, but discovered by millimeter-wave molecular-line observers in the early 1980s (e.g., Snell et al. 1980; Lada 1985). Bipolar outflows are very energetic flows of cold molecular gas generally consisting of two spatially separate lobes moving diametrically away from an embedded YSO at hypersonic velocities. These outflows are very massive, often containing considerably more mass than the central YSO which drives them. This indicates that the molecular outflows primarily consist of swept-up material and not ejecta from the embedded driving star itself. Bipolar outflows are the manifestation of an underlying primary wind generated by the embryonic protostellar core. Very close to the surface of the protostellar object, the primary wind is most often manifest by a highly collimated, circumstellar jet. These jets contain sufficiently hot and ionized gas to emit at optical as well as centimeter wavelengths. Although such jets appear to originate very close (\(\leq 50\) AU) to the protostellar core, they can also extend to large distances from the central object. Such jets are frequently observed to terminate at Herbig-Haro objects which have the shapes of extended bow shocks.

Bipolar outflows are extremely energetic. The mechanical luminosities of such flows are appreciable fractions (\(\sim 5\%\)) of the total radiant luminosities of the protostellar objects driving them (Bally and Lada 1983; Cabrit and Bertout 1992). Moreover, the mechanical luminosities are nearly linearly correlated with the radiant luminosities of the protostars. Since protostellar luminosities are largely generated by accretion luminosity, which is proportional to the mass accretion rate, the
mechanical energies of outflows must also be more or less proportional to the protostellar mass accretion rate. It is logical to assume therefore that the ultimate source of the energetics of bipolar flows is gravity which is tapped by accretion from deep in the potential well at or near the protostellar surface.

Angular momentum can be carried away from a star by a stellar wind. Consequently, a protostar may be able to gain mass only if it simultaneously loses mass. To allow star formation to proceed once the initial protostellar core is formed, a protostellar wind must be generated that carries away angular momentum and mass at a fraction of the mass accretion rate i.e., $\dot{M}_{\text{wind}} = f \dot{M}_{\text{accretion}}$ where the fraction $f$ is determined by the physics of the wind generating mechanism. The ideal protostellar wind is one that carries away little mass but lots of angular momentum. A number of recent investigations have shown that centrifugally-driven hydromagnetic winds are potentially capable of doing the job and such winds could be driven from the magnetosphere of the central protostar (e.g., Shu et al. 1988; 1994). Thus, the natural consequence of star formation in rotating, magnetic cloud cores may be the formation of protostar-disk systems which generate powerful magnetically driven outflows which remove sufficient angular momentum from accreting gas to allow the central core to add mass and grow.

As the embryonic core begins to grow its luminosity is dominated by accretion. The mass of the protostellar core increases as $M_*(t) = \dot{M}t$. Once a protostellar core reaches a mass of roughly 0.2–0.3 $M_\odot$, its central temperature reaches $10^6$ Kelvins and deuterium burning nuclear reactions are ignited providing the protostar with another source of luminosity. Accretion of new material and a fully convective core enable the protostar to continue burning deuterium as it grows. If placed on the HR diagram the protostellar core would follow a trajectory parallel to and somewhat above the deuterium burning main sequence which itself lies significantly above and roughly parallel to the (hydrogen-burning) main sequence. A protostellar core can continue to gain mass by accretion until its central temperature reaches $10^7$ Kelvins and hydrogen fusion is ignited at which point the embryonic core reaches the main sequence. This occurs when the cores attain masses around 7–8 solar masses. Presumably, the protostar still can continue to grow and increase its mass. However,
for reasons not yet fully understood, most protostars cease growing long before this point. Growth could be terminated by an exhaustion of circumstellar material in the core or perhaps by the outflows generated by the protostars. One consequence of an “accretion driven” outflow is that as a protostar evolves, its outflow disrupts the surrounding material and thus may ultimately limit the mass which can fall onto the central star. As the protostellar envelope disperses, the reservoir of infalling material available to drive disk accretion and mass outflow also diminishes and the YSO evolves to progressively less active states. At this point the post-protostellar phase begins. For low mass stars this is the pre-main sequence phase of stellar evolution.

§ 5. Pre-main sequence evolution and the formation of planetary systems

The initial conditions for pre-main sequence evolution are those which describe the mass, radius and luminosity of an embryonic protostellar core at the the point in time that infall and accretion cease. For this reason the locus of points on the HR diagram which traces the initial starting points of pre-main sequence evolution for all stars is called the birthline (Stahler 1983). The position of a given star on the birthline is a function of the mass it has acquired by the end of its protostellar accretion phase. Presumably, the protostellar evolution of a given star is identical to that of all other stars until the time that the star stops accreting and reaches the birthline. Prior to its appearance near the birthline, a young stellar object is surrounded by an infalling protostellar envelope which renders it invisible. Once the envelope is either mostly accreted onto the protostellar core or removed by an outflow, the protostar becomes a visible pre-main sequence star. Its initial luminosity and surface temperature place it near the birthline which for low mass stars is itself basically coincident with the deuterium burning main sequence on the HR diagram. Because the abundance of deuterium in the young star is relatively low, it is burned up very rapidly. Without accretion to replenish the burned deuterium, nuclear reactions cease and the star slowly contracts to the main sequence. The time scale for this quasistatic contraction to the main sequence is the Kelvin-Helmholtz time. A pre-main sequence star will contract until its central core temperatures reach $10^7$ K, at which point hydrogen burning will commence in its interior increasing the internal pressure to exactly balance gravity so that the star will begin a long-lived phase of overall equilibrium on the main sequence.

During the early parts of the pre-main sequence phase, stars are still surrounded by accretion disks which deplete themselves by continuing to add mass to the central star but at relatively low rates. Perhaps an additional few % of the final mass of the star is added during this phase. It is during this period of evolution that planetary systems are formed within the disk. It generally held that the planet formation process begins with the settling of dust into the midplane of a circumstellar disk due to drag forces resulting from interaction with the gas (e.g., Lissauer 1993; Ruden 1999). In this process the micron sized interstellar dust grains collide and can merge together forming increasingly massive solid particles. In addition cooling processes
in the disk can lead to the production of condensates which will collide and merge with the interstellar dust. These particles can grow to millimeter then centimeter size rather quickly and then when they reach kilometer size they become planetesimals. Planets are constructed from the mutual accretion of such planetesimals. The formation of planetesimal sized bodies can occur fairly rapidly, in 1 Myr or less at 1 AU. The planetesimals then grow quickly to an earth mass or more. Larger planetesimals formed in the outer protoplanetary disk can accrete gaseous atmospheres and form Jovian planets. Such planets are massive enough to create gaps in the circumstellar disk which can generate torques that can result in the inward migration of the planets required to explain the discovery of extra-solar planetary systems with Jovian planets orbiting extremely close to their parent star.

The planetesimal theory of planet formation can account for many of the observations of the solar system but its success in explaining observations of extra solar planetary systems is less clear. Because circumstellar disks are ubiquitous around young stars the theory can, in principle, be tested, by direct observation of nearby protoplanetary systems. Young clusters offer an excellent laboratory for investigating disk evolution. These clusters present statistically significant samples of stars whose mean ages are well determined. Moreover, by combining observations of embedded as well as open clusters one can create a sample that spans a range in age much greater than that which characterizes any individual star forming region. The frequency of disks within a young stellar population is directly related to the physical processes of disk formation and evolution. Knowledge of the cluster disk fraction (CDF) and how it evolves with time has important consequences for understanding the origin of planetary systems. Since most stars probably formed in embedded clusters, the measurement of the CDF in the youngest embedded clusters produces a determination of the initial disk frequency which in turn directly measures the probability of disk formation around newly formed stars. The variation of the CDF with cluster age sets the timescale for disk evolution and thus the duration or lifetime of the circumstellar disk (and planet building) phase of early stellar evolution. This therefore provides a critical constraint for determining the probability of planet formation in circumstellar disks and so therefore also directly bears on the question of the ubiquity of extrasolar planetary systems.

The CDF and its dependence on stellar mass and cluster age, in principle, can be directly measured. This is because stars with circumstellar disks emit excess infrared emission which can be readily detected at any infrared wavelength sufficiently longward of the peak of the underlying stellar energy distribution of the central star. Although, the longer the wavelength the more unambiguous the infrared excess, observations at wavelengths as short as 2 \( \mu \)m (K-band) can detect infrared excesses in the majority of disk bearing stars. Indeed, essentially all stars with circumstellar disks containing \( 10^{-9} M_\odot \) of hot dust or more can be detected using ground-based L-band (3.4 \( \mu \)m) observations (Lada et al. 2000; Wood et al. 2002). Figure 5 shows the variation of the CDF within young clusters as a function of cluster age derived from infrared L-band observations by Haisch et al. (2001). These clusters each contain hundreds of members spanning nearly the entire range of stellar mass and have well determined ages. The youngest clusters exhibit very high excess or
Fig. 5. Disk Fraction as a function of cluster age for a sample of young clusters with consistently determined mean ages. The disk fraction is initially very high, but then rapidly drops with cluster age suggesting maximum disk lifetimes of less than 6 Myrs in young clusters (Haisch et al. 2001).

disk fractions indicating that circumstellar disks are a natural byproduct of the star formation process and that most stars, independent of mass, are therefore born with the ability to form planetary systems. However, the observations show a rapid decline in the CDF with cluster age. Half of the disks in a cluster appear to be lost within only 2–3 Myrs and essentially all the disks are gone in about 5–6 Myrs. Moreover, observations also indicate that disk lifetimes are also functions of stellar mass with disks around higher mass stars evolving more rapidly than disks around low mass stars. Such a rapid timescale for disk evolution places stringent constraints on the timescale for building planets, particularly giant gaseous planets. For example, according to current planetesimal theories it takes more than 10 Myr to form the cores of Jupiter and Saturn and even longer for Uranus and Neptune. These long timescales appear in conflict with the observations. Either the formation of plane-
tary systems is a relatively rare event, or the existing theory needs to be adjusted. For example, an alternative model of (giant) planet formation that has been recently investigated is that of gravitational instability in the disk (Boss 2000; 2003). This model is attractive in that it can form giant planets on timescales consistent with that required by the cluster observations, however more detailed calculations are needed to determine whether it is a viable alternative.

§6. Concluding remarks and future prospects

The story of stellar origins described in the previous section represents the current paradigm for star formation research. However, it is only a small part of the whole story. For example, a major deficiency with the current theoretical paradigm is that it cannot yet predict the form of the IMF. Another issue to be considered is that most stars form in clusters and the physical process of star formation in the clustered environment (or mode) is not yet understood. In addition, the current paradigm has not been extended to massive stars such as those stars ($m_* > 7$–8 $M_\odot$) which begin their post-protostellar evolution as hydrogen burning objects and never experience a pre-main sequence phase of post-protostellar evolution. The details of the formation of such massive stars are much less known. Our lack of understanding of the details of how these heavier stars are constructed is largely due to a combination of the rarity of massive stars, the rapid evolution of such objects and the very destructive nature of their interaction with their surrounding natal material. There are indications that the theory of low mass star formation may be extendable to the formation of high mass stars. For example, very young high mass stars seem to be surrounded by disks and are found to produce bipolar outflows like low mass stars. On the other hand, entirely new mechanisms such as stellar mergers in the central regions of dense clusters may be required to explain the formation of the most massive stars. Developing a theoretical picture of massive star formation and star formation in the clustered environment are two of the most important challenges facing modern star formation research.

Another outstanding problem that deserves attention is the formation and early evolution of GMCs and the dense cores within them. It is clear that the mechanism responsible for creating the mass spectrum of dense cores and, in particular, the most massive dense cores, is ultimately responsible for the fact that most stars form in embedded clusters. Solving the problem of embedded cluster formation requires solving the problem of dense core formation. Yet our understanding of cloud and core formation is meager. Understanding these issues presents another important challenge for star formation studies. What are the critical observations that will reveal how clouds are formed and how dense cores evolve to become star forming factories? It is not clear that new capabilities or technological developments will help answer this question. Our ability to detect and measure the gaseous component of molecular clouds has existed for more than thirty years, yet little progress has been made in determining even such basic facts as the ages of molecular clouds. Here the challenge will be to devise new ways of thinking about the problem and new applications of existing capabilities to address it. Perhaps the most exciting prospects
for major progress lie in the area of understanding planet formation and the early evolution of planetary systems. The development of new observational facilities will have great impact on this problem. Spectroscopic capabilities in the mid-infrared (e.g., Spitzer Space Telescope, SOFIA) will enable detailed studies of disk spectral energy distributions and permit the identification of such features as disk gaps which are expected to accompany planet formation. Facilities such as ALMA may even be able to directly image such disk structures. Such capabilities will also enable the direct observation of the gaseous component of disks, something which is of critical need and has been difficult to obtain with existing instrumentation. In the future large 30 meter class ground-based telescopes operating at the diffraction limit in the near-infrared may be able to resolve and directly image recently formed planetary companions around young pre-main sequence stars. Finally, there is much interest in understanding star formation in the earliest epochs of cosmic history. How were the first stars formed? To what extent can we use what we know about present day star formation to address this problem? Will direct empirical study of this process be possible?

Star formation is a complex and rich phenomenon. Its study will continue to be a major focus of astronomical research into the foreseeable future and will help advance our overall understanding of both the nature and evolution of the astronomical universe.

Acknowledgements

I thank the Organizing Committee for the kind invitation to present this review at the 19th Nishinomiya-Yukawa Memorial Symposium and for the support to enable my participation in the conference. I thank Gus Meunch for assistance in preparing Fig. 3. I also acknowledge support from NASA grant NAG 5-13041 for some of the work reviewed here.

References

17) Hara, A., Tachihara, K., Mizuno, A., Onishi, T., Kawamura, A., Obayashi, A. and Fukui,
33) Lada, C. J., Muench, A. A., Haisch, K. E., Lada, E. A., Alves, J. F., Tollesstrup, E. V. and
52) Myers, P. C., Dame, T. M., Thaddeus, P., Cohen, R. S., Silverberg, R. F., Dwek, E. and
57) Porras, A., Christopher, M., Allen, L., Di Francesco, J., Megeath, S. T. and Myers, P. C.,