UNDERSTANDING YOUNG STARS: A HISTORY*

STEVEN W. STAHLER

Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

ABSTRACT

The history of pre-main-sequence theory is briefly reviewed. The paper of Henyey et al. (1955) is seen as an important transitional work, one which abandoned previous simplifying assumptions yet failed to incorporate newer insights into the surface structure of late-type stars. The subsequent work of Hayashi and his contemporaries is outlined, with an emphasis on the underlying physical principles. Finally, the recent impact of protostar theory is discussed, and speculations are offered on future developments.

Key words: stellar evolution--star formation

I. Introduction

The paper by Henyey, LeLevier, and Levee (1955, hereafter HLL), reprinted in this issue, represents a clear turning point in the development of pre-main-sequence theory. Louis Henyey and his collaborators were the first to study early stellar evolution by integrating numerically, using a computer, the four full differential equations of stellar structure. Abandoning earlier restrictive assumptions about the nature of pre-main-sequence contraction, the authors achieved unprecedented accuracy. Moreover, the matrix technique for solving the equations first used in this paper, the famous Henyey method, is still the working heart of most stellar-evolution calculations. On the other hand, Henyey was a transitional figure and was himself guilty of a simplifying assumption that undermined his results. By ignoring the possibility of surface convection in his stars, he was following the canonical wisdom of his day. It is appropriate that Henyey, with one foot in the past and one in the future, turned out to be half right. We now know that his results describe only the final approach of stars to the main sequence. At earlier epochs, stars are wholly convective and follow, in the H-R diagram, vertical paths which are roughly orthogonal to Henyey's (see Fig. 1).

In this paper I offer a brief historical sketch of our changing conception of young stars. I tell of the progression of ideas that led to Henyey's work, of the correction of his crucial oversight by Hayashi, and of the more recent refinements and modifications of the theory. Throughout, my focus is on theoretical models of pre-main-sequence stars, those objects which are too young and cold to burn hydrogen but which are past the earlier, protostar phase of dynamical accretion from a parent interstellar cloud. I begin with the idea of surface convection, the ingredient, missing in Henyey's work, that plays a key role in our current picture of early stellar evolution.

II. The Rise and Fall of Surface Convection

The notion that the interior transport of heat in stars occurs by turbulent convection has a venerable tradition. Astrophysicists of the late nineteenth century were convinced that the Sun and other stars had contracted to their present sizes from much larger configurations and that

![Diagram of theoretical pre-main-sequence evolutionary tracks](image)

FIG. 1—Theoretical pre-main-sequence evolutionary tracks (schematic). A star of a given mass is first convective and descends a nearly vertical path, eventually turning onto the more horizontal radiative track discovered by HLL. The vertical path and its downward extension (dotted line) form the border of Hayashi's forbidden region (shaded) for stars of that mass.

---

*One of a series of invited reviews celebrating the centenary of the Astronomical Society of the Pacific.
the time scale for this gravitationally-driven process was set by the rate of energy loss from the stellar surface. This cooling luminosity was thought to drive interior convection in both the present-day stars and their larger antecedents. Whatever the physical motivation for this preference of convective to radiative transport, it was, in any case, a highly convenient one. Given the undeveloped state of radiation theory and stellar opacities, physicists would have been hard-pressed to calculate the structure of radiative stars, even if they had believed in them. Fortunately, convective stars could be accurately described using polytrope theory, the elegant mathematical formalism developed by Kelvin, Lane, Emden, and Ritter. I remind the reader that a polytropic gas is one whose pressure $P$ and density $\rho$ are related by

$$ P = K_0 \rho^{1+1/n} ,$$

where $n$, the polytropic index, and $K_0$ are constants. Stars with fully ionized, convective interiors have nearly uniform specific entropy and are well modeled as polytropes with $n = 1.5$.

At the turn of the century, a key step was taken by Karl Schwarzschild (1906). Schwarzschild found, by a simple argument, the quantitative criterion for convective instability. He showed that, in any local region of the star, the inequality

$$ \nabla > \nabla_s$$

has to be satisfied for instability. Here $\nabla_s$ is defined as $\nabla_s = (\partial \ln T / \partial \ln P)_s$, the logarithmic temperature gradient at constant specific entropy, while $\nabla$ is the actual logarithmic gradient in the region. Equation (2) is equivalent to the statement that the specific entropy must decrease outward in the star. Schwarzschild also knew that, once initiated, convection is such an efficient mechanism for heat transport that the actual temperature gradient in a convective star is close to the adiabatic one, i.e., that the specific entropy is nearly constant spatially.

Unfortunately for the development of stellar-evolution theory, Schwarzschild went on to offer a proof that convection is impossible in stellar atmospheres! Schwarzschild’s argument, as recounted in Woolley and Stibbs (1953), began by assuming the familiar Eddington approximation for the run of temperature with optical depth in the atmosphere:

$$ T^4 = \frac{T_{eff}^4}{2} [1 + (3/2)\tau] .$$

Here, the effective temperature $T_{eff}$ is related to the stellar luminosity $L$ by the blackbody law

$$ L = 4\pi R^2 \sigma T_{eff}^4 ,$$

where $R$ is the radius of the photosphere. Schwarzschild next gave the condition for hydrostatic equilibrium as

$$ \frac{dP}{d\tau} = g ,$$

where $g$ is the acceleration of gravity and $\kappa$ is a frequency-averaged opacity. From equation (3), Schwarzschild derived the relation

$$ \frac{1}{T} \frac{dT}{d\tau} = \frac{3}{8+12\tau} ,$$

while he obtained, from equation (5), assuming constant opacity,

$$ \frac{1}{P} \frac{dP}{d\tau} = \frac{1}{\tau} .$$

Combining these last two equations, he obtained

$$ \frac{d \ln T}{d \ln P} = \frac{3 \tau}{8+12 \tau} .$$

Now, for any perfect gas, the polytropic law in equation (1) is equivalent to the relation

$$ P = K_0 T^{n+1} ,$$

where $K_0$ is another constant. Assuming that a gas undergoing an adiabatic transformation can indeed be modeled as a polytrope, Schwarzschild used equations (6) and (7) to show that his criterion for convective instability, equation (2), becomes

$$ \frac{1}{n+1} \leq \frac{3 \tau}{8+12 \tau} .$$

But this last inequality never holds at any $\tau$ if $n$ is less than 3. For a fully ionized gas with $n$ equal to 1.5, the criterion fails.

Although Schwarzschild’s argument seems naive today for its assumption of constant opacity, it apparently impressed his contemporaries, for convection quickly faded as the mechanism for stellar energy transport. Over the next 25 years only radiatively stable models were seriously considered. In his influential book, The Internal Constitution of Stars, Eddington (1926) devoted a scant few paragraphs to the description of a star undergoing convection. He closed the section with the following remark.

We shall not enter further into the historic problem of convective equilibrium since modern researches show that the process is untenable.

Within a few years, however, Eddington’s fellow theorists were finding the untenable process to be tenable once again. Unsöld (1930) pointed out that in late-type stars the region of hydrogen ionization below the photosphere would have a very shallow temperature gradient. Thus, from equation (7), $n$ would be large, and the instability criterion on equation (8) would indeed be satisfied, at least in that one region of the star. About the same time, Jeffreys (1931) noted that nuclear-energy generation in any star could plausibly create a central convection zone. Jeffreys did not know the precise nuclear reactions pow-
tering stars, but he saw that if these reactions were sufficiently temperature sensitive, they would be confined to a small central core. Thus, the temperature gradient needed to carry the nuclear-generated luminosity would become superadiabatic as one approached the center, and Schwarzschild’s original criterion for instability (eq. (2)) would be met.

Unsöld’s “hydrogen convection zone” was quickly seen to provide an important link to the energetic surface phenomena, such as granulation, long known to solar physicists, and it was implicated in the turbulent broadening in spectral lines of distant stars as well. However, the argument used by Unsöld showed that this convection was confined to the region where hydrogen was partially ionized. Such a convective layer, perhaps a few hundred kilometers thick (Unsöld 1938), would have negligible effect on the star as a whole. It was perfectly reasonable, therefore, that theorists working on stellar models took very seriously the suggestion by Jeffrey’s but ignored the phenomenon described by Unsöld. Chandrasekhar (1939), in his classic treatise on stellar structure, treats the case of stars with convective cores in great detail but never mentions the possibility of surface convection. Throughout the 1930s and 1940s the respectable models for stars of all masses and ages were ones with convective cores and radiative envelopes. Chandrasekhar (1939) modeled such stars as composite polytropes and matched inner and outer solutions of the Lane-Emden equation (with different values of n) at the convective-radiative interface. More-realistic models involved the numerical solution of the stellar-structure equations for radiative equilibrium in the outer region.

At this point a brief technical digression is necessary. We will see shortly that the error made by Henyey and corrected by Hayashi was essentially an inadequate treatment of boundary conditions, so it is important to note how previous workers dealt with this issue. The major difficulty in solving the stellar-structure equations is the fact that two of the four boundary conditions must be specified at the center of the star, while the other two must be given at the surface. This unfortunate fact means that any straightforward solution of the equations entails both inward and outward integrations, along with a procedure to match variables at some internal fitting point (see, e.g., Schwarzschild 1958). The central conditions are the values of any two thermodynamic variables (usually pressure and temperature), while the surface conditions are the blackbody law, equation (4), and the value of $P_{\text{phot}}$, the photospheric pressure. This quantity is obtained by integrating equation (5) from $\tau = 0$ to $\tau = 2/3$:

$$P_{\text{phot}} = \frac{2g}{3k}.$$  

To simplify their task, theorists generally replaced these exact outer boundary conditions by the approximate ones

that the pressure and temperature fall to zero simultaneously at the stellar surface. It can be shown that these “zero boundary conditions” cause no appreciable error if the interior opacity rises sufficiently slowly with increasing temperature (see eq. (6–9b) of Clayton 1968). Such a situation is indeed encountered in early-type stars, where the opacity generally falls with rising temperature. In any case, the simplified boundary conditions were the ones universally adopted by theorists of the 1930s and 1940s in their models with convective cores and radiative envelopes.

III. Red Dwarfs and Giants

By the early 1950s it was becoming apparent that the standard core-envelope models were inadequate to describe stars of all masses and ages. The first real difficulties were encountered not with young stars, for which observations were scarce, but with late-type main-sequence stars. Here theorists had to confront reliable data on stellar masses, radii, and luminosities. In the case of the Sun there were early signs of trouble in the fact that the models needed implausible chemical abundances in order to fit the observations (Schwarzschild 1946). An additional source of confusion in the solar models was the question of whether the CNO cycle or the PP chains were the relevant nuclear reactions. For less massive main-sequence stars, however, it was clear that only the PP reactions, which do not require as high a central temperature, were operative.

The PP reactions are markedly less temperature sensitive than those in the CNO cycle, so the region of nuclear burning in low-mass stars is not strongly concentrated at the center. Thus, the original motivation for convective cores no longer applied. But now theorists studying red dwarfs reached an impasse. It was found that all models employing radiatively stable interiors were greatly overluminous with respect to observed stars of known mass and radius (e.g., Aller 1950).

It was Osterbrock (1953) who saw the root of the problem and quickly supplied its solution. Following a suggestion by Strömgren (1952), he tried red-dwarf models that consisted of radiative interiors and deep outer convection zones. In effect, he widened the original hydrogen convection zone of Unsöld several hundredfold, until it encompassed a substantial portion of the stellar mass and radius. To be sure, neither Osterbrock nor Strömgren had been the first to suggest that stars could have large convective envelopes. Biermann (1938) replaced Unsöld’s original argument by a much more careful study indicating that the extent of surface convection had been greatly underestimated. However, it is clear that it was not Biermann’s analysis, but the urgent need to reconcile theory and observation, that motivated the new picture of late-type stars.

Figure 2 shows how Osterbrock’s introduction of deep
convective envelopes resolved the luminosity problem in red dwarfs. The figure depicts, in a schematic fashion, three possible distributions of the specific entropy, $s$, in a star of fixed mass and radius. Note that the figure does not attempt to depict the change in entropy in the thin subphotospheric layers; this surface variation, as we shall see, is important in late-type stars. Curve 1 represents the flat entropy distribution found in a fully convective star, while curve 3 shows the rising entropy encountered in a fully radiative interior. Now the radius of a star is a sensitive, increasing function of the total entropy content of the star. Hence, in a star of fixed radius, the star in radiative equilibrium must have a lower specific entropy at the center than in the fully convective case. If we now add to this radiative model, as Osterbrock did, a convective envelope (curve 2), the central entropy must be higher in order to compensate for the lowered entropy in the envelope. Raising the central entropy decreases the central density and temperature, so that Osterbrock’s models burned hydrogen less vigorously than their fully radiative counterparts. In the end the stellar luminosity was lowered, as was needed to match the observations.

Osterbrock was well aware that his new models implied a strong connection between the interior structure of the star and the properties of its photosphere. He supplied, in fact, a detailed analysis of the outer layers of his stars. Figure 3 depicts, again schematically, the subsurface entropy distribution, as a function of pressure, in late-type stars. The region shown occupies a negligible fraction of the mass and radius of the star. Beginning at the photospheric pressure given in equation (9), the specific entropy falls in a thin, radiatively stable zone. In stars with surface temperatures less than 5000 K, the opacity in this region is provided chiefly by H$^-$ ions. At the time of Osterbrock’s paper a detailed calculation of this opacity source had recently become available (Chandrasekhar and Münch 1946). The H$^-$ ion receives its electrons from ionizable metals, and this supply of electrons increases copiously with rising temperature, resulting in a sharp rise in the opacity itself. Moving inward from the surface, the increasing opacity, now also due to photoionization of hydrogen, induces a rise in the temperature gradient needed to push out the stellar luminosity. Eventually, at the pressure marked $P_{\text{cr}}$, the gradient becomes superadiabatic and convective instability sets in. Due to the increasing number of free electrons in the hydrogen ionization zone, the entropy interior to $P_{\text{cr}}$ first rises sharply, then levels off to the value marked $s_{\text{int}}$. Because of the high efficiency of convective heat transport, $s_{\text{int}}$ remains the value of the entropy for the entire depth of the convection zone. Thus, in stars with deep convective envelopes, the photosphere and the entropy structure in the associated subphotospheric layers play a decisive role.

Since the H$^-$ opacity is a rapidly increasing function of temperature, Osterbrock knew that an inward integration using zero boundary conditions would result in gross inaccuracies. Moreover, he felt that the mixing-length theory of convection was too unreliable to permit an inward integration from the photospheric boundary conditions. The method he chose for his actual model calculations was to ignore the surface structure, revert to the zero boundary conditions, but to treat $s_{\text{int}}$ (or, equivalently, the depth of the uniform-entropy convection zone) as a free parameter, adjusted to match the observed stellar luminosity for a given stellar mass and radius. In this way he found that convection zones covering about 30% of the stellar radius could match the observations of his sample stars, with no need for exotic chemical compositions.

At the same time as the theory of red dwarfs was undergoing radical revision, a closely parallel situation was developing in the analysis of red giants. For stars which have exhausted their hydrogen fuel, there could hardly be the possibility of a central convective core. It was realized that an isothermal core was more appropriate, and models incorporating a radiative envelope with the zero boundary condition successfully reproduced the

---

1Actually, Osterbrock ignored the entropy rise in the ionization zone, citing the high efficiency of convective transport and the near equality of $\nabla$ and $\nabla_s$. However, these two gradients can be nearly equal in a region of rapidly rising entropy, as explained in Stahler, Palla, and Salpeter (1986). Fortunately, Osterbrock’s error did not vitiate his quantitative results, nor those of Hayashi and Hoshi (1961) who also neglected the ionization entropy rise.
initial departure of stars from the main sequence seen in old clusters. However, these same evolutionary tracks failed to turn up the red-giant branch, continuing instead on a nearly horizontal path (Sandage and Schwarzschild 1952). In a masterful analysis of Population II stars, Hoyle and Schwarzschild (1955) produced the upturn by introducing deepening convection zones into their evolving models. Unlike Osterbrock, they used the exact photospheric boundary conditions, employing mixing-length theory to determine the flat convective entropy $s_{\text{int}}$ and joining this outer integration to one from the center.

IV. Henyey, Hayashi, and the Forbidden Region

It is time to return, after this account of convection in mature and aging stars, to our main topic of pre-main-sequence contraction. We have seen that the earliest models of contracting stars assumed them to be in a state of convective equilibrium, with spatially constant entropy. The contraction of such stars is said to be "homologous". Homology transformations, which have historically played an important role in stellar-structure theory, are those in which all physical variables in a given fluid element in one model are related by a spatially-independent scale factor to the variables in the corresponding fluid element in another model. In homologously contracting stars, the factors depend on time, so that successive models have distributions of pressure, temperature, etc., that look the same at any time except for an overall change in scale.

The mathematical simplicity of the homologous contraction of fully convective stars is due to a corresponding simplicity in their thermal evolution. Convective eddies within the star homogenize the entropy on a much faster time scale than that for gravitational contraction. The latter, known as the Kelvin-Helmholtz time scale, $t_{\text{KH}}$, is set by the rate at which the stellar surface can radiate away a significant fraction of the gravitational binding energy of the star:

$$t_{\text{KH}} = \frac{G \mathcal{M}^2}{RL}.$$  

(10)

Here $\mathcal{M}$, $R$, and $L$ are the stellar mass, radius, and luminosity, respectively. Over this time scale, the specific entropy declines, while the central temperature
of the star, which varies inversely as the radius, increases. A central assumption of pre-main-sequence theory is that \( t_{\text{KH}} \) is much longer than the sound-crossing time in the star. If so, the star can effectively be considered to be in hydrostatic equilibrium at each instant, despite its slow, or quasi-static, contraction.

If we now consider stars which are not fully convective, their contraction still takes place over the time scale \( t_{\text{KH}} \), but this contraction cannot, in general, be homologous. Indeed, the specific entropy in a radiative star changes on the time scale for photon diffusion, a time which, along with the opacity, varies strongly from one region of the star to another. Nevertheless, the early studies of pre-main-sequence evolution considered only homologous contraction, even after fully convective stars had passed out of fashion. Thomas (1931) offered a detailed analysis of the homologous contraction of stars in radiative equilibrium. He showed that the homology condition required stringent initial and boundary conditions that were hardly likely to be met in practice. Nevertheless, researchers had little hope of following the local variations of entropy before the advent of the computer.

The paper by HLL was published shortly after an earlier study of the same problem by Levee (1953). Levee assumed that solar-type stars would be radiatively stable in their pre-main-sequence phase, and he used the computer to integrate his equations. However, he did not solve the full heat equation:

\[
\frac{dL}{dM} = \epsilon - T \frac{ds}{dt},
\]

where \( L \) and \( M \) are the internal luminosity and mass, and \( \epsilon \) is the nuclear-energy generation rate. Instead, Levee replaced the right-hand side of this equation by the gravitational energy release in a homologously contracting sphere with the stellar mass and radius. It was the virtue of HLL to solve equation (11) without approximation. They did so by dividing the star into concentric shells and writing the difference equations relating variables at adjacent shells in matrix form. Starting with an initial guess for the structure, a matrix inversion was performed to arrive at the next distribution of all variables; this inversion process was iterated until successive values of the variable at each shell differed negligibly. The great advantage of this method over the older two-way integrations was that it could incorporate both the central and surface boundary conditions automatically, bypassing the need for tedious multiple integrations and cumbersome matching procedures. The method, first described by Henyey et al. (1959), has since been widely adopted for stellar structure and evolution calculations. It is an inappropriate technique, however, for stars which have widely disparate thermal time scales in different regions, as is the case in low-mass accreting protostars (Stahler, Shu, and Taam 1980).

Henyey either ignored or was unaware of the recent and contemporary work on convection in red dwarfs and giants. In any case, HLL adopted the usual assumption of a convective core and radiative envelope, using the zero boundary conditions. Actually, the code only established a convective core if Schwarzschild's criterion was met, which it never was. The main result of HLL, that stars follow nearly horizontal tracks during pre-main-sequence contraction, can be reproduced by a simple argument. Any star in radiative equilibrium, if its interior opacity follows Kramer's Law, has a luminosity given by

\[
L_{\text{rad}} = C \Phi^{0.5}/R^{0.5},
\]

where \( C \) is a constant. This relation follows simply from a dimensional analysis of the radiation diffusion equation and application of the virial theorem (see, e.g., Cox and Giuli 1968, Chap. 22). The quantity \( C \) depends weakly on the precise entropy distribution within the star. For each radiative star of a fixed mass, equation (12) shows that \( L_{\text{rad}} \) rises slowly, as \( R^{-0.5} \), in the course of contraction. Interestingly, a similar derivation of these tracks had earlier been given by Salpeter (1954). Salpeter assumed provisionally that stars would contract homologously, with fully radiative interiors, to their final main-sequence positions. His tracks look similar to those in HLL, but he missed the downward turn next to the main sequence. As explained in HLL, this final drop in luminosity is caused by the halting of contraction due to incipient hydrogen burning.

The rest of the 1950s saw no new theoretical breakthroughs, but major observational advances, in the study of young stars. Equation (10) shows that the duration of the pre-main-sequence phase for solar-type stars is of order 10^7 years, perhaps 1% of their main-sequence lifetimes. Thus, these stars are relatively rare, and it is not surprising that their identification did not come quickly. It was largely through the work of George Herbig that the recently discovered T Tauri stars became recognized as low-mass pre-main-sequence stars at this time. The construction of the first observational H–R diagrams of young clusters containing such stars (Walker 1956) is another major development that can only be mentioned in passing here. The next advance in theory was the seminal paper of Hayashi (1961), which significantly altered the results of HLL and essentially formed our modern conception of pre-main-sequence stars.

At the time of his discovery, Hayashi and his colleagues had been studying the evolution of red giants. Building on the previous work of Hoyle and Schwarzschild, Hayashi and Hoshi (1961) presented again the surface structure shown in Figure 3 and showed how this structure, in the case of giants with deep convective envelopes, is of central importance. Their analysis was an approximate one. Rather than integrate the stellar-structure equations for the entire star, they contented themselves with determi-
nation of the interior convective entropy, $s_{int}$. For chosen values of stellar mass, radius, and luminosity, they first calculated conditions at the photosphere, then numerically integrated inward through the radiative and hydrogen ionization zones, using mixing-length theory in the superadiabatic region. When the entropy leveled off, they obtained, not its value directly, but the asymptotic value of the ratio $P/\rho T^{5/2}$, i.e., the quantity $K_3$ in equation (7). They then used $K_3$ to evaluate a nondimensional quantity, denoted $\tilde{E}$, which is formed from $K_3$ and the stellar mass and radius. To motivate the use of $E$, which figures prominently in Hayashi’s subsequent work, consider first a star with spatially constant specific entropy. Let us define a nondimensional pressure, $p$, and a nondimensional temperature, $t$, by the relations

$$P = \frac{p}{\mathcal{M}_{\text{grav}}^2}, \quad t = \frac{\mu G \mathcal{M}}{\mathcal{M}_{\text{grav}}},$$

where $\mathcal{M}$ is the gas constant and $\mu$ the interior molecular weight. Since the star is isentropic, we have

$$p = E t^{5/2},$$

where $E$ is a kind of nondimensional entropy. Using equation (7) in conjunction with equations (13a) and (13b), we readily evaluate $E$ as

$$E = 4\pi \mathcal{M}^{-5/2} G^{3/2} \mu^{-5/2} \mathcal{M}^{1/2} R^{-3} K_3.$$  

We may also evaluate $E$ directly from polytrope theory. We begin with the expression for stellar radius in terms of the constant $K_1$ from equation (1) and the central density, $\rho_c$:

$$R = \left[ \frac{5K_1}{8\pi G} \right]^{1/2} \xi \rho_c^{-1/6},$$

as found, e.g., in equation (23.18) of Cox and Giulii (1968). Here $\xi$ is the first zero of the Lane-Emden function, $\Phi(\xi)$, for $n$ equal to 1.5. Now $K_1$ can be related to $K_3$ through the perfect gas law

$$K_1 = K_3^{-2/3} \mathcal{M}_{\text{grav}}^{5/3} \mu^{-5/3}. $$

Equation (17), together with the polytropic equation for the stellar mass (eq.23.30 of Cox and Giulii 1968), can be substituted into equation (16) and the result solved for $K_3$:

$$K_3 = \left( \frac{125}{128\pi^2} \left( \frac{\mathcal{M}^3}{G^3 \mu^3} \right) \left( \frac{1}{\mathcal{M} R^3} \right) \left( -\frac{\xi}{\xi} \frac{d\Phi}{d\xi} \right) \right)^{1/2}. $$

Finally, we may substitute this expression for $K_3$ into equation (15) to obtain

$$E = \left( \frac{-125}{8} \xi \frac{d\Phi}{d\xi} \right)^{1/2} \xi \left( \left( \frac{1}{\xi} \right) \right)_{\xi_1} = 45.48. $$

Thus, we see that in a fully convective star, the quantity $E$, defined through equation (15), has a unique numerical value. It may be shown, using the Lane-Emden equation, that stars in which $E$ exceeds this critical value have the property that the mass contained in an interior sphere approaches a nonzero value as the sphere shrinks to zero radius. Conversely, in stars with subcritical values of $E$, the interior mass vanishes for some nonzero radius of the sphere. This latter problem can be remedied by switching, at some radius where the mass is still finite, to an interior structure in which the entropy declines toward the center. In other words, stars for which equation (15) gives $E$ less than 45.48 have radiative interiors and convective envelopes.

Using equation (15) Hayashi and Hoshi could obtain $E$ for stars with selected $\mathcal{M}$, $R$, and $L$, where $L$ was needed to evaluate $K_3$ by inward integration from the photosphere. Turning this process around, they actually fixed $\mathcal{M}$ and $E$ and used equation (15) to plot evolutionary tracks in the H-R diagram for stars of Population I and Population II composition. They found the results that (a) different values of $E$ gave parallel, almost vertical, tracks with higher $E$ values corresponding to higher effective temperatures, and (b) for the “reasonable” value of $E$ equal to 20, the tracks were in good agreement with the observed diagrams of old clusters.

This analysis by Hayashi and Hoshi would have remained an interesting postscript to the work of Hoyle and Schwarzschild had not Hayashi seen the significance of his results for pre-main-sequence evolution. In a paper submitted the same day as the work on red giants, Hayashi (1961) noted that an evolutionary track corresponding to a fixed mass and to the critical $E$ value of 45.48 forms the border of a “forbidden region” in the H-R diagram (see Fig. 1). Any star of that mass which finds itself within the region has too low an effective temperature to be in hydrostatic equilibrium. Hayashi claimed that the star would quickly move in the diagram toward the border of the forbidden region. Once there, the star, now fully convective, would descend the border until, intersecting the appropriate horizontal track computed by HLL, the star would become radiatively stable and follow the HLL track to the main sequence. Although Hayashi’s description of how stars initially inside the forbidden region reach its border is no longer considered valid, his account of their subsequent evolution forms the basis of modern pre-main-sequence theory. The convective portions of the first “Hayashi tracks” were obtained by setting $E$ equal to 45.48 in equation (15) and plotting $L$ versus $T_{\text{eff}}$ for fixed values of $\mathcal{M}$.

Hayashi and Hoshi recognized that, in a red giant of fixed mass, the value of $E$ must increase as the star ages and the convection zone deepens. However, the position of the tracks was only weakly dependent on $E$, and they fixed the value at 20 to provide a rough comparison with the observations.
In his original derivation, Hayashi did not offer a physical interpretation of his far-reaching results. Why do stars of a given mass and radius have a minimum surface temperature? Why should stars of sufficiently large radius be fully convective? Why is the actual surface temperature in such stars the minimum possible value? And, finally, why is this minimum temperature insensitive to stellar radius, so that the convective portions of the Hayashi tracks are nearly vertical? Our previous analysis of the surface conditions in late-type stars provides the answers to all these questions.

Stars have a minimum surface temperature because of the decline of opacity with falling temperature in their outer layers. For too low a surface temperature, these layers become optically thin, violating the condition that the photosphere be located at an optical depth of order unity from the surface. Mathematically, the photospheric pressure in equation (9) increases as $T^4$ and $\kappa$ both fall. The photospheric specific entropy, $s_{\text{phot}}$, which is proportional to $\ln(T^4/P)$, correspondingly declines. From Figure 3, a lower $s_{\text{phot}}$ corresponds to a lower value of $s_{\text{int}}$, the entropy in the outer regions of the deep interior. However, Figure 2 shows that a star of fixed mass and radius has a minimum value of $s_{\text{int}}$, that corresponding to a fully convective interior. Therefore, this star also has a minimum value of $s_{\text{phot}}$ and $T_{\text{eff}}$.

Once we recognize that stars have a minimum surface temperature, it is readily seen why they must be fully convective when their radii are large. A large radius corresponds, from equation (4), to a large surface luminosity. If this luminosity exceeds that which can be carried radiatively ($L_{\text{rad}}$ in eq. (12)), the star must be convectively unstable. In such a state the surface temperature is the minimum value, i.e., the star descends along the border of the forbidden region. As the contraction continues, the surface luminosity falls and $L_{\text{rad}}$ rises until the two come into equality. It is at this point that the star becomes radiatively stable and follows a horizontal track to its main-sequence position. For stars of sufficiently low mass, $L_{\text{rad}}$ is so small that the stars remain fully convective to the main sequence (see Fig. 4).

The fact that the convective portion of the Hayashi track is so nearly vertical follows from the extreme temperature sensitivity of the photospheric opacity. As noted by Hayashi and Hoshi (1961), the opacity varies as $T^6$, with $\beta$ between 7 and 13 for the temperature regime of interest. Only a minor change of $T_{\text{eff}}$, therefore, is needed for a substantial change in $P_{\text{phot}}$ and $s_{\text{phot}}$. On the other hand, the interior entropy $s_{\text{int}}$ is relatively insensitive to the stellar radius, dropping slowly as the radius contracts. Thus, for the entropy moving in from the photosphere to match that in the deep interior, $T_{\text{eff}}$ must not change substantially. For solar-type stars this temperature is about 4000 K.

Following their discovery by Hayashi, the new evolutionary models were computed with ever-increasing accuracy by subsequent researchers. Hayashi, Hoshi, and Sugimoto (1962) abandoned the assumption that $E$ remains strictly constant above the radiative track; they followed, instead, the variation in this quantity as a radiative core grows in the star toward the bottom of its vertical track. Ezer and Cameron (1963) and Weymann and Moore (1963) repeated the calculations with improved low-temperature opacities but assumed homologous contraction in computing the gravitational energy release. Finally, Iben (1965) and Ezer and Cameron (1965) incorporated the correct heat equation and, using the Henyey method, constructed detailed evolutionary sequences for a number of different masses. Also included in their models were a number of preliminary thermonuclear processes that precede the main phase of hydrogen burning. When, a few years later, Grossman and Graboske (1971) provided a study of comparable detail for stars of very low mass, a physically self-consistent framework for the pre-main-sequence phase had been established.

### V. The Impact of Protostar Theory

Even before the detailed models of Iben (1965) and Ezer and Cameron (1965) were presented, it was realized that the underlying theory was manifestly incomplete as a description of early stellar evolution. The best observational candidates for pre-main-sequence stars, those of the T Tauri class, displayed irregular variability and anomalous spectroscopic features which the stellar models did not begin to address (Herbig 1962). The discovery of vigorous mass loss (Kuhi 1964), of continuum infrared excess (Mendoza 1966), and of eruptive increases in bolo- metric luminosity (Herbig 1977) further underscored the need for refinements of the basic theory. In the last two decades, these refinements have not followed any single, systematic program, but have sprung up in response to the observations. As one example, consider the ongoing efforts to understand pre-main-sequence angular-momentum loss in terms of magnetized winds (e.g., Kawaler 1988); this program is motivated by the observations of extraordinarily low rotation speeds in T Tauri stars (Vogel and Kuhi 1981). As another example, the infrared excess problem has stimulated a great deal of interest in the theoretical modeling of circumstellar disks (e.g., Adams and Shu 1986). We cannot attempt, in this brief review, to follow each of these diverse strands of inquiry. Suffice it to say that no single, modified pre-main-sequence theory has arisen that successfully matches all the observations.

Another issue faced by theorists has been the origin of the pre-main-sequence phase itself. Although it was originally hoped that the model of a star undergoing slow gravitational contraction would suffice to describe the earliest stages of evolution, it became apparent, by the early 1960s, that such was not the case. Cameron (1962) and Gaustad (1963) pointed out that, under the diffuse
Fig. 4—Observational H-R diagram of the Taurus-Auriga molecular cloud complex. Open circles represent the T Tauri observations of Cohen and Kuhi (1979). The light solid lines are the theoretical pre-main-sequence tracks of Iben (1965) and Grossman and Graboske (1971), with the appropriate masses (in solar units) labeled. The heavy solid line is the birthline of Stahler (1983).

conditions typical of observed interstellar clouds, the cooling luminosity of any stellar object would be so large that its Kelvin-Helmholtz time (see eq. (10)) would be significantly less than the sound-crossing time. Thermal pressure would, therefore, be incapable of providing mechanical support, and the object would undergo, not quasi-static contraction, but free-fall collapse.

For a time it was believed that detailed calculation of this early collapse, or protostar, phase was unnecessary for establishing the initial location of stars on their Hayashi tracks. Cameron (1962) argued that protostellar collapse was so rapid that negligible energy would be radiated away prior to the formation of the pre-main-sequence star. Thus, all of the lost gravitational potential energy of the parent cloud would appear as thermal and ionization energy in the newly formed star. Using the virial theorem to relate the gravitational and thermal energies in the star, Cameron obtained

$$R = 50 R_\odot \left(\frac{M}{M_\odot}\right)$$

(20)

for the initial radii of pre-main-sequence stars (see also Hayashi 1966). As for the initial thermal state of these stars, Bodenheimer (1966) and Von Sengbusch (1968) showed that solar-type stars with radii given by equation
provide a reference point for the many poorly-understood properties of these objects.

Protopstar theory does more than set the initial loci of pre-main-sequence stars on their Hayashi tracks. For stars of higher mass, the theory actually predicts that they should first appear optically far removed from their classical tracks. To see how this situation arises, we must first understand that, in the absence of nuclear burning, an accreting protostar is always radiatively stable. The reason is that the specific entropy at each mass shell in the hydrostatic core is essentially that which resulted from passage of that layer through the accretion shock. As the mass of the protostar grows, the strength of the shock increases, resulting in progressively higher entropies. In other words, the specific entropy in the hydrostatic core increases outward. This increase is slight, but it is enough to violate Schwarzschild's criterion for convective instability. Despite this tendency for radiative stability, the ignition of deuterium in low-mass protostars turns the star convectively unstable. However, for intermediate-mass cores ($\mathcal{M} \approx 2 \mathcal{M}_{\odot}$), $L_{\text{rad}}$ in equation (12) becomes adequate to carry away the deuterium-generated luminosity, and the star reverts to a state of radiative stability.

Consider now the fate of a newly revealed pre-main-sequence star of intermediate mass. Since its internal entropy, $s_{\text{int}}$, obtained through accretion, is rather low, its surface thermal structure resembles that in Figure 3. Therefore, the surface temperature of this star is the typical Hayashi value of 4000 K to 5000 K, and the radius, $R_{\text{core}}$, is also of modest size (e.g., 5 $R_{\odot}$). From equation (4), therefore, the surface luminosity is also relatively low, perhaps 20 $L_{\odot}$. In fact, if $R_{\text{core}}$ is less than $R_{\text{mag}}$, the smallest radius of a convective star of that mass, the values of $L$ and $T_{\text{eff}}$ place the star directly below the vertical portion of its classical Hayashi track (see Fig. 5). Now, the interior luminosity of this radiative object is governed by equation (12) and can be extraordinarily high, depending on the stellar mass. The internal distribution of luminosity resembles that in Figure 6, with $L_{\nu}$ peaking at some off-center mass shell and falling to the surface value in a thin outer convection zone. The star is in a state of thermal imbalance and will contract in a highly nonhomologous fashion. The region between the peak luminosity and the surface will heat up until the rising value of $T_{\text{eff}}$, forces the surface and internal luminosities to come into equality. In the meantime, the star will move in the H-R diagram to a position on the radiative portion of the classical Hayashi track (Fig. 5), where it will attain a luminosity close to the original peak value. A detailed calculation of this thermal relaxation process, including the advance of the luminosity peak, will be necessary to see if the process can account for any of the eruptive luminosity increases that appear to be endemic among young stars.

At still higher masses ($\mathcal{M} \approx 5 \mathcal{M}_{\odot}$), protostar theory indicates that the pre-main-sequence phase is skipped.
entirely. Massive protostars, as we have seen, are radiatively stable, and $L_{\text{rad}}$ increases so rapidly with mass that the thermal relaxation process just described occurs even before the end of accretion. More quantitatively, the Kelvin-Helmholtz time associated with the internal luminosity of the star has become smaller than the accretion time scale of the protostar. The latter is given by $\frac{\dot{M}}{\dot{M}_{\text{ac}}}$, where $\dot{M}$ is the mass accretion rate onto the hydrostatic core. For the values of $\dot{M}$ found in numerical and analytical cloud-collapse studies, the two time scales become equal at a stellar mass of about 5 $M_\odot$. If the hydrostatic core grows beyond this mass, it will contract internally at a faster rate than it is accreting mass from the parent cloud. For a time it will be in a curious hybrid state, where it is invisible optically because of the dust in the surrounding, infalling envelope, but where it is deriving virtually all of its luminosity from gravitational contraction rather than accretion. Eventually, the central ignition of hydrogen via the CNO cycle must halt the internal contraction. The star will have arrived at the main sequence without passing through a normal, optically visible pre-main-sequence phase. It will be interesting to see if future studies of such massive stars can elucidate the properties of the most luminous infrared sources now seen embedded in interstellar gas.

VI. Conclusion

The study of early stellar evolution is today one of the most active fields in astronomy. Dramatic advances in radio and infrared instrumentation have enabled observers to pinpoint regions of star formation and to discern properties of the nascent objects. As yet, there is no single theoretical model that can accommodate this flux of exciting and surprising observational results. We may profitably ask, however, which portion of the historical advances outlined here are likely to survive in this future theoretical framework.

First, it is clear that the observed stars which are found near interstellar gas and are located in the H-R diagram above the main sequence are indeed the pre-main-sequence objects studied for decades by theorists. The observed stars include both the T Tauri class mentioned previously and, at higher masses, the less well-studied Herbig Ae and Be stars (Herbig 1960). Further, there is every reason to believe that these young stars are undergoing quasi-static contraction, first as convective, then as radiative, objects to their final main-sequence configurations. However, it is also clear that the theoretical models perfected by Hayashi and his contemporaries represent idealized, smoothly-evolving stars that give no hint of the erratic, puzzling behavior actually seen. To explain this behavior will require substantial modification, but not radical revision, of the underlying theory. For example, it is a matter of current debate whether the UV excess exhibited by many T Tauri stars is a manifestation of an active chromosphere or of a viscous dissipation process in a star-disk boundary layer (e.g., Bertout, Basri, and Bouvier 1988). In either case, the underlying star is still a
quasi-statically contracting object with substantial outer convection.

It is at the interface between the protostar and pre-main-sequence phases that the boldest theoretical advances will be necessary. Observations currently indicate that the end of the protostellar accretion phase and the disruption of the parent cloud occur by the action of an energetic stellar wind. This wind often manifests itself as a "bipolar flow" in the surrounding material of molecular gas (Lada 1985). The origin of this wind is unknown, but it is tempting to relate it to another puzzling observational fact. The increasing evidence for disks surrounding many pre-main-sequence stars supports the long-held belief that a typical interstellar cloud fragment collapsing to a star contains more angular momentum than can be incorporated in the star itself. However, if the disks are the repositories of this excess angular momentum, then the surfaces of the newly-revealed young stars should be rotating nearly at breakup speed, corresponding to about one stellar revolution per hour. In fact, observations of the youngest pre-main-sequence stars, those close to the birthline in the H-R diagram, yield rotation speeds that are a minor fraction of this value (Hartmann et al. 1986).

The obvious implication is that the large kinetic energy of rotation that "should" be contained in the stars was somehow lost in the form of energetic winds. How such a conversion process could occur is a formidable theoretical challenge, one at the forefront of our understanding of young stars.

REFERENCES


———. 1938, Physik der Sternatmosphären (Berlin: Springer).