Arch. Hist. Exact Sci. (2010) 64:203–267
DOI 10.1007/s00407-009-0055-0

Solving the Giant Stars Problem: Theories of Stellar Evolution from The 1930s to The 1950s

Davide Cenadelli

Received: 1 October 2009 / Published online: 25 December 2009

Abstract Historiography has pointed out that the time between the mid 1910s and the early 1930s can be considered a pivotal period in the history of stellar astrophysics. In those years, scholars like Saha and Eddington first applied atomic physics to astrophysics. Theoretical astrophysics was born. This led to the development of the first physically sound models for stellar interiors and atmospheres. These landmark achievements spurred scholars to elaborate theories for stellar evolutions, and in the following decades several astrophysicists focused on this problem. The evolutionary role of red giants turned out to be the main issue. Those stars were initially assumed to be young ones going through the formation stage, but astrophysicists gradually realized that they were rather to be considered old, evolved stars. The solution of the giant stars issue required a couple of decades: it was not until the mid 1950s that a satisfactory explanation was obtained. This provides a detailed picture of the theories of stellar evolution from the 1930s to the 1950s and of the solution to the red giants problem, with special emphasis on how such a solution was made possible by a series of subsequent steps: the identification of changing chemical composition as a main evolutionary feature of a star, the inclusion of nuclear physics within the theoretical framework of stellar astrophysics, the recognition of the importance of inhomogeneities that settle within stars as nuclear processes go on.

Communicated by Tilman Sauer.

D. Cenadelli (✉)
Dipartimento di Fisica, Università degli Studi di Milano, Via Brera 28, 20121 Milano, Italy
e-mail: davide.cenadelli@unimi.it

1 Introduction

The history of stellar astrophysics in the twentieth century is widely dealt with in historiography. Many works on earlier periods up to the 1920s and the 1930s exist in the literature, but it appears that the subsequent decades have been less frequently accounted for. A reason for this is the well-established fact that the time between the mid 1910s and the early 1930s is a pivotal period in the development of stellar astrophysics.[1] In those years, scholars like Saha, Russell, Fowler, Milne, Jeans and Eddington first applied atomic physics to the description of both the inner and the outer regions of the stars. The so-called Quantitative Era of astrophysics[2] began, with the description of the stars' internal structure, the theoretical explanation of stellar spectra via the quantum theory, and the determination of the chemical composition of the stars as its first outcome. In other words, scholars developed the first sound models for stellar structure. Although these achievements were later subject to further discussion, revision and improvement, they are still the pillars current astrophysics is built upon.

These landmark achievements spurred scholars to elaborate theories for stellar evolution: the description of the different features of dwarf and giant stars demanded an evolutionary interpretation. Several astrophysicists focused on this problem. In particular, the evolutionary role of red giants turned out to be a thorny issue. Whereas in the twenties these stars were generally thought to be young bodies in the stage of formation, in the following decades such a picture was put into question and it became increasingly clear that giants are rather old, evolved stars. Consequently, the giants issue appears to be a well-defined problem emerging from the meaningful discontinuity astrophysics experienced in the twenties and the thirties.

[1] Historiography has stressed this issue quite extensively. It is commonly recognized that at the time astrophysics experienced a discontinuity that would lead to a new definition of the discipline itself—with the onset of theoretical astrophysics—and to the identification of new perspectives and problems. Whether the term "discontinuity" should be replaced with "revolution" is questionable. I lean towards the former, as I think that period cannot be labelled as a scientific revolution in the Kuhnian sense (i.e. a whole redefinition of a disciplinary field that gives rise to a completely different perspective). Actually, the main change astrophysics underwent was the attainment of a sound theoretical basis through its newly created link with quantum physics. Of course, this was conducive to the rise of new important issues, like the investigation of both the inner and the outer regions of the stars. By the way, it should be noticed that perspective changes if the onset of observational astrophysics in the mid-nineteenth century is also taken into account. In fact, astrophysics as a whole has a revolutionary character indeed. Meadows (Meadows 1984) suggests that astrophysics can be considered a "revolution" in the sense that it shifted astronomers' attention elsewhere. That this field of research opened up new horizons and dramatically changed modern astronomy is hardly questionable. Among the most interesting works on this subject, see DeVorkin and Kenat (1983a), DeVorkin and Kenat (1983b), Dingle (1963), Goldberg (1988), Hearnshaw (1986, pp. 208–254), Herrmann (1984, pp. 115–120), Hoskin (1999, pp. 252–267), Meadows (1984), Menzel (1972), Sitterly (1970), Strömgren (1972), Tassoul and Tassoul (2004, pp. 96–100).

[2] A division of the history of stellar astrophysics into an earlier "Qualitative" and a later "Quantitative Era" is suggested by DeVorkin and Kenat (1983a). The authors claim they took this suggestion from a previous paper by Menzel (Menzel 1972). Saha's first application of atomic physics to stellar spectroscopy in 1920 acts as a watershed between the two eras.

This article carries out an in-depth investigation of the giants affair from the post-Eddington period to the mid-1950s, when scholars devised the first sound model for giants.

Although such a description might be a bit too schematic, the main steps that drove scholars from the Eddington era to the first reliable giant stars models can be summarized as follows.[3] The first step was the identification of chemical composition as a major issue in evolution. This followed directly from Eddington's and Strömgren's work about hydrogen abundance in the early thirties. Strömgren himself played a key role in this task by identifying the ever-changing composition throughout stellar life as a feature to be accounted for. However, Strömgren only focused on general changes that affected the whole stellar mass, i.e. he thought that stars could change their composition while remaining homogeneous at each stage.

Later on, scholars realized they had to describe these changes quantitatively and, to do that, they needed to turn to nuclear physics, which was growing impressively at the time.[4] This is a major reason why stellar evolution theories from the 1920s and 1930s were to be totally discarded soon afterwards. They did incorporate modern features such as the perfect-degenerate gas duality, but they had no links to nuclear processes theories that were still in their infancy at the time. As we shall see, an understanding of the evolutionary patterns of stars required a detailed description of nuclear processes, which would then have to be connected to the laws of stellar structure. During the twenties, scholars relied solely on atomic physics to penetrate the mysteries of stellar matter. When nuclear physicists like Bethe and Gamow began to turn their attention to astrophysics, new perspectives opened and, as a major outcome, the haze that wrapped up the theories of stellar evolution began to fade.[5] It is no surprise that, then, scholars were still a long way from regarding giants as evolved stars at the time.

A further step forward was made when Chandrasekhar and other scholars argued that chemical inhomogeneities, rather than a mere changing composition, could occur within a star as hydrogen burning went on. This led them to investigate 'composite' models in which the stellar core and envelope were differently described and suitably fitted to one another. This turned out to be a good assumption giants could be built upon.

Finally, between the late 1940s and the mid 1950s, Hoyle, Sandage, Schwarzschild and collaborators developed models that included a detailed description of chemical inhomogeneities, as well as other key features like gas degeneracy and core collapse, and they succeeded in developing the first sound theories of giants.

[3] Reference texts that clearly highlight these points are Arny (1990), Tassoul and Tassoul (2004, pp. 133–56), Hufbauer (2006). More generally, this paper was inspired by these works.

[4] Although there exists a tight relation between the giants problem and the details of nuclear processes, I will not probe the latter topic. I will just mention the main results that must be exploited when developing giant stars models.

[5] This consideration by Hufbauer (Hufbauer 2006, p. 203) emphasizes the "breakthrough" role played by nuclear physics in astrophysics. Such a term is justified by the circumstance that the introduction of nuclear physics into astrophysics has similarities with the previous, ground-breaking exploitation of the atomic models. However, such an analogy should not be pushed too far. This issue will be further discusses in this paper.

The main contents of this paper will be preceded by two introductory sections: Sect. 2 lays down a theoretical background of stellar models, whereas Sect. 3 provides an overview of the theories devised by scholars in a former period. Sections 4–9 are devoted to the historical issues associated to the giants problem. Finally Sect. 10 draws some conclusions about the main features of the scientific development process in the period considered.

2 Theoretical background: basic stellar models

2.1 The basic equations

This section lays down some basic concepts about stellar structure that will be useful later on in this paper. Those concepts were developed by a number of scholars between the mid-nineteenth century and the first decades of the twentieth century.

A basic assumption when building a stellar model is that spherical symmetry holds, i.e. that the star's features are solely dependent upon the distance r from the centre. This makes it possible to introduce r as an independent variable such as $0 \leq r \leq R$, where R is the stellar radius. Pulsations, rotation and magnetic fields are not considered. A stellar model aims at describing the internal structure of a star via the calculation of the values of the state parameters $T(r)$ (temperature), $P(r)$ (pressure), $\rho(r)$ (density) and of the quantities $M(r)$ (mass within radius r) and $L(r)$ (luminosity generated within radius r), for any value of r. In the following, L and M will indicate the total luminosity and mass of the star.

If the star is supposed to be mechanically stable, then the hydrostatic equilibrium equation holds:

$$\frac{\mathrm{d}P(r)}{\mathrm{d}r} = -G\frac{M(r)}{r^2}\rho(r) \tag{1}$$

It should be noticed that Eq. 1 is valid as far as an equilibrium between gravity and pressure exists. If this is not strictly the case, Eq. 1 can still be assumed valid if a star passes through subsequent states of quasi-equilibrium, i.e. provided it does not undergo quick and abrupt stages of contraction or expansion. During stellar evolution Eq. 1 is usually satisfied.[6]

Moreover, general considerations allow to introduce a continuity equation for the mass:

$$\frac{\mathrm{d}M(r)}{\mathrm{d}r} = 4\pi r^2\rho(r) \tag{2}$$

Consequently, an equation of state for the gas is required. If gas is in perfect-gas conditions, we have:

$$P(r) = P_{\mathrm{gas}}(r) + P_{\mathrm{rad}}(r) = KN(r)T(r) + \frac{1}{3}aT(r)^4 \tag{3}$$

[6] Prialnik (2000, p. 72).

where P_{gas} is the pressure due to the gas, P_{rad} is the radiation pressure, $N(r)$ is the number of particles per unit volume, K is the Boltzmann's constant (here and henceforth, the uppercase "K" will indicate Boltzmann's constant, the lowercase "k" any generic constant—see for example Eqs. $3'$ and $3''$, and the Greek "κ" the absorption coefficient —see below in this section).

If the gas is degenerate, then the equation of state is:

$$P(r) = k_1 \rho(r)^{5/3} \tag{3'}$$

in the non-relativistic case, or:

$$P(r) = k_2 \rho(r)^{4/3} \tag{3''}$$

in the relativistic case.

So far, we have three equations in the four quantities P, ρ, T, M, so another one is required. Hence the need to introduce an energy-generation equation:

$$\frac{\mathrm{d}L(r)}{\mathrm{d}r} = 4\pi r^2 \rho(r) \varepsilon(r) \tag{4}$$

where $\varepsilon(r)$ is the energy generated per unit time and mass. Equation 4 features two more unknowns: $L(r)$ and $\varepsilon(r)$ itself. The latter ceases to be such if nuclear physics supplies relations capable to estimate it as a function of temperature, density and chemical composition.[7] Once $\varepsilon(r)$ is known, a further relation is needed to close the system of equations. This can be an equation tying $T(r)$ with $L(r)$, i.e. an equation for energy transport, which is caused by convection or radiation depending on the conditions. Alternatively, the system of equations can be closed even avoiding the introduction of Eq. 4, by simply making the useful assumption that there exists a further relation among the variables P, ρ, T, M. A power-law relation between P and ρ like the one shown in Eq. 5 is, in some cases, a tenable choice

$$P(r) = k\rho(r)^{\gamma}. \tag{5}$$

Equation 5 is called polytropic relation, $n = (\gamma - 1)^{-1}$ is called polytropic index and a stellar model built following this method is called a polytrope.[8]

[7] It is important to point out that the term ε was introduced many years before any actual knowledge about nuclear processes could provide information about it. This is the case for Eddington's standard model. Models like Eddington's are perfectly reliable notwithstanding this lack of knowledge, provided some assumptions are made about the probable nature and distribution of the energy-generating processes within the stars. Eddington and other scholars made such assumptions and worked out stellar models. This issue will be further discussed later in this section. As we shall see, it is precisely through the term $\varepsilon(r)$ that nuclear physics enters the equations of stellar structure.

[8] For a full description of polytropes, see for example Collins (1989, pp. 42–53), Kippenhahn and Weigert (1990, pp. 174–190) or Prialnik (2000, pp. 74–79).

It should be noticed that the constant k in (5) is usually a free constant. However, k may also be fixed, as in the case of a degenerate gas (Eqs. $3'$ and $3''$).[9] A polytrope is usually integrated by eliminating $P(r)$ and $M(r)$ from the set of equations to obtain a differential, second-order equation featuring two dimensionless variables tied to r and ρ (this is the well-known Lane-Emden equation.)[10]

The theory of polytropes was first put forth by the American engineer Jonathan H. Lane and by the German structural engineer August Ritter in the years around 1870–1880, well ahead of the well-known work carried out by the Swiss physicist Robert Emden at the beginning of the twentieth century. Lane developed the first model of a gaseous, self-gravitating sphere in convective equilibrium. As long as convection is supposed to take place in such a way that the gas cells undergo quasi-adiabatical changes, Lane's sphere is well described by a polytrope (see Eq. 20 in Sect. 5). Ritter in his turn solved the soon-to-be-called Lane–Emden equation for different values of n (Lane, in turn, had solved it numerically for the values $n = 3/2$ and $n = 5/2$). After expanding and streamlining the work of his predecessors, in 1907 Emden published the milestone book *Gaskugeln*, where he numerically integrated the equation named after him for various values of the index n. This book would grow to become the reference work about polytropes for many years.[11]

It should be noticed that no issues associated to energy production and transport are found in the polytropic case. This has sometimes led to allegations that it might not be a likely description of any actual star, but this is not the case. For example, the aforementioned perfect-gas sphere, where convection is dominant and a quasi-adiabatic equilibrium sets up, happens to be a polytrope. Another important case is the one for which $n = \infty$. This turns out to be an isothermal sphere—a likely occurrence, as we shall see, in certain stages of stellar evolution.

2.2 Eddington's "standard model"

The famous British astrophysicist Arthur S. Eddington developed his "standard model" of stellar structure based on a star for which the perfect-gas law held and energy transport via radiation was dominant.[12] By resorting to the usual expression of gas and radiation pressure, and by introducing the quantity β (defined as the gas-to-total-pressure ratio), we can write:[13]

[9] The degenerate gas state equation $3'$ or $3''$ are in fact polytropic relations, with indexes $n = 3/2$ and $n = 3$, respectively.

[10] Collins (1989, p. 45).

[11] About these pioneering contributions, see Tassoul and Tassoul (2004, pp. 73–79 and 250–251). As the authors suggest, the Lane–Emden equation could also be called the Lane–Ritter–Emden equation: "It is unfortunate that the value of Ritter's astronomical papers has never been adequately recognized." (p. 79)

[12] Eddington worked out his "standard model" in a number of papers published in the years 1916–1932, the most important of them being Eddington (1916, 1917, 1924a,b, 1932).

[13] Masani (1984, p. 67).

$$P(r) = \left[\left(\frac{K}{\mu \mathrm{H}} \right)^4 \frac{3}{a} \frac{1-\beta}{\beta^4} \right]^{1/3} \rho(r)^{4/3} \tag{6}$$

where H is the hydrogen mass and μ the mean molecular weight (i.e. the mean mass per particle in units of H). It is clear that, generally speaking, both μ and β depend upon r. For the moment, we can neglect any r-dependence for μ.[14] If we further suppose β to be constant, based on the previous equation the star turns out to be a $n = 3$ polytrope. In order to assume β as constant, Eddington was forced to make a questionable assumption. He first introduced a parameter η defined as follows:

$$\frac{L(r)}{M(r)} = \eta \cdot \frac{L}{M} \tag{7}$$

In Eddington's words "[...]η varies as the energy liberated per unit mass, averaged through the part of the star interior to the point considered".[15] It can be demonstrated that $\beta = \mathrm{const} \Leftrightarrow \eta\kappa = \mathrm{const}$,[16] where κ is the absorption coefficient. This was defined by Eddington as "the absorption by a cylinder of unit mass and unit cross-section".[17]

Therefore, Eddington supposed that η times κ must be constant anywhere in the star. Although this assumption was criticized, Eddington felt confident that it would always be approximately true. In fact, he reckoned that the absorption coefficient was likely to decrease towards the innermost regions, whereas η was likely to increase as the energy sources were supposed to be more powerful around the hotter centre of the star.[18]

Strictly speaking, the luminosity of a star does not appear in a polytropic model. However, it is a natural component of any radiative model. Eddington succeeded in introducing it and showed that it is tied to the absorption coefficient κ through the expression:[19]

$$L = \frac{4\pi c G(1-\beta)M}{\kappa\eta} \tag{8}$$

[14] In other words, we consider a star to be a homogeneous body and we suppose that high (complete) ionization occurs almost everywhere inside it. The fact that an actual star may be not homogeneous is a focal point of this paper and will be discussed at length in the following sections.

[15] Eddington (1917, p. 599).

[16] Prialnik (2000, p. 83).

[17] Eddington (1916, p. 18). An equivalent definition for κ by (Smith 1995, p. 247) is "the absorption cross-section per unit mass". In other words, κ is given by $1/\rho l$ where l is the mean free path of a photon. The dimensions of κ are $\mathrm{cm}^2\ \mathrm{g}^{-1}$.

[18] This is an example of the problems that the lack of knowledge about the energy-supplying processes could raise, and of the tentative ways that were sometimes used to bypass them. Moreover, Eddington had to decide which values to adopt for the product $\eta\kappa$. After long speculations, he finally went for $\eta\kappa = \alpha\kappa_c$, where κ_c is the value of the absorption coefficient at the centre of star and α is a coefficient whose value depends upon the distribution of the energy sources. Eddington often supposed $\alpha = 2.5$. For further information about the assumptions Eddington made and their acceptability, see Mestel (2004).

[19] Eddington (1924a, p. 109).

From Eq. 8 Eddington derived its famous "quartic" equation:[20]

$$1-\beta = \text{const.} \times M^2\mu^4\beta^4 \tag{9}$$

and from (8) and (9) he worked out the mass–luminosity relation.[21] To attain this famous relation, Eddington had to estimate the value of κ and he used Kramers' photoionization law:[22]

$$\kappa(r) = \kappa_0\rho(r)T(r)^{-7/2} \tag{10}$$

where κ_0 is a constant that depends upon the constitution of stellar matter. Eddington actually worked out its value from observational data and this gave rise to the well-known "opacity discrepancy".[23]

As we shall see, the relation expressing the ρ and T-dependence of the absorption coefficient would become a major concern for scholars in the following years.

2.3 Cowling's "point-source" model

Polytropes are a tentative, relatively simple technique to deal with stellar structure, and in some cases they are reliable models. However, simple stellar models can also be constructed by incorporating luminosity, Eq. 4, without going through Eqs. 7 and 8, but rather by making other likely assumptions about the distribution of energy sources within the star, i.e. about the function $L(r)$. In general, from the definition of radiation pressure and absorption coefficient it is possible to write:

$$\frac{\mathrm{d}P_{\mathrm{rad}}(r)}{\mathrm{d}r} = -\frac{\kappa(r)\rho(r)L(r)}{4\pi r^2 c} \tag{11}$$

No solution can be derived from Eqs. 1, 2, 3 and 11, as another relation is needed. This is found in the assumption about energy generation. In 1930 Thomas G. Cowling developed the so-called point-source model,[24] in which the whole of the stellar energy is generated in the central point: $L(r) = L$ for any value of r from 0 to R. It follows that $L(r)$ is constant and the system can be closed by choosing an opacity law. Cowling assumed that κ and μ were constant anywhere in the star[25] and introduced a dimensionless variable u given by $u = \text{const.} \times r^{-1}$. He then worked out a conclusive

[20] Eddington (1924a, p. 109).

[21] Eddington (1924b, p. 310); the mass–luminosity relation is a major topic in literature; see for example Mestel (2004) and Hufbauer (2006).

[22] Eddington (1924b, p. 310).

[23] On the opacity discrepancy see DeVorkin and Kenat (1983b) and Cenadelli (2008).

[24] Cowling (1930); for further information about this model see also Prialnik (2000, pp. 86–89).

[25] Of course, these are very questionable assumptions, and Cowling was perfectly aware of this. He remarked: "It is not suggested that the model represents physical conditions accurately: but a full discussion of this model should give insight into the nature of stellar structure." (Cowling 1930, p. 92).

equation that plays the same role as the Lane–Emden equation in the polytropic case and contains only a function $T(u)$:[26]

$$\frac{1}{T^3}\frac{\mathrm{d}}{\mathrm{d}T}\left\{\frac{1}{T^3}\frac{\mathrm{d}}{\mathrm{d}T}\left(T^4\frac{\mathrm{d}T}{\mathrm{d}u}\right)\right\} = -\frac{1}{u^4} \tag{12}$$

Finally, he then integrated this equation numerically.

2.4 Composite models

In the early thirties, Eddington's work on stellar models fostered further investigation by other scholars (besides Cowling), notably Edward A. Milne and Bengt Strömgren. Milne worked out models[27] with a constant value for κ. Such models turned out to be centrally condensed, i.e. with T_c and $\rho_\mathrm{c} \to \infty$ (subscript 'c' standing for central). This led Milne to realize that the perfect-gas law could not hold, and he suggested that such stars be studied as constituted by two different regions: an outer $n = 3$ polytrope—as suggested by the standard model—and an inner $n = 3/2$ polytrope corresponding to a state equation like ($3'$). The idea of the existence of a "stellar nucleus" was rising. Milne's paper inspired Strömgren. The Danish astrophysicist postulated that there exists a stellar nucleus where energy is supplied, surrounded by an inert outer region.[28] This model had some similarities with Cowling's, but Strömgren removed a very questionable point. i.e. that κ is constant, and rather assumed Eq. 10 to hold. For the outer region, he made use of Eqs. 1, 2 and 3, and Eq. 11 that he wrote in the equivalent form:

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{1}{3}aT^4(r)\right) = -\frac{1}{c}\frac{\kappa(r)\rho(r)L(r)}{4\pi r^2} \Rightarrow \frac{\mathrm{d}T(r)}{\mathrm{d}r} = -\frac{3}{4ac}\frac{\kappa(r)\,\rho\,(r)L(r)}{4\pi r^2 T^3(r)} \tag{13}$$

that I report here, as it is often written thus. This is the equation of energy transport for radiation.

Strömgren took reasonable values for κ_0 and μ, the latter being equal to 2.2[29] throughout the whole star: no chemical inhomogeneities were considered. It is a common feature of all the early models that possible chemical composition variations *within* a star are not accounted for. Then he started the numerical integration from the surface, moving inward with the condition $L(r) = L$ as long as he was in the inert region. He accounted for the fact that in the innermost region $\rho T^{-3/2}$ could become so large as to switch to the degenerate-gas case, and thus he had to replace Eq. 3 with $3'$. Then the main issue was to fit two polytropes together, i.e. to use the values of radius, mass and density at the deepest point of the $n = 3$ model as boundary conditions

26 Cowling (1930, p. 94).

27 Milne (1930).

28 Strömgren (1931b).

29 At the time scientists had not yet come to the conclusion that hydrogen is very abundant in stars. See also note 49. On the other hand, Strömgren carefully analysed possible variations of κ_0 within the star.

for the $n = 3/2$ model.[30] The main result Strömgren arrived at was that, if Eq. 10 holds, similar conclusions (as in the κ = const. case) can be drawn about the possible existence of centrally condensed configurations.

Apart from further details about the method used, it is interesting to notice that Milne's and Strömgren's models highlight some noteworthy issues: the importance of a full understanding of the nature of the energy-supplying sources as a key point to construct a physically realistic stellar model, the possible existence of centrally condensed models, and the idea that a star could be described via different models in the different zones, to be fitted together at the interface between them.

The fitting of a $n = 3$ to an $n = 3/2$ polytrope came back in the limelight a few years later, when Cowling built up a model with a convective core and a radiative envelope.[31] He was investigating the convective and vibrational stability of a point-source model where energy generation has a strong dependence upon ρ and T. Cowling found that, with Kramer's opacity law, convection was supposed to appear in the central region of a star if the dependence on temperature was higher that the eighth power. On the other hand, the envelope kept fully radiative. In fact, Cowling's point-source model does not—strictly speaking—describe an actual physical situation due to the infinities it implies. Nevertheless, as we shall see, it is a likely model for nuclear reactions that possess a strong dependence upon temperature, like the CNO cycle.[32] As Cowling recalls: "Since the pioneer work of Atkinson had already suggested a very rapid increase of energy-generation with temperature, this model was useful in indicating the properties of real stars".[33]

2.5 The problem of fitting polytropes together

As we have seen, even in those early days the idea was emerging that the inner and the outer regions of a star could be described by different models to be fitted to one another. It is useful for the fitting process to define the two following quantities U and V, that are homology-invariant, i.e. two homologous spheres have the same values for U and V:[34]

$$\begin{aligned} U &= \frac{\mathrm{dlog}M(r)}{\mathrm{dlog}r} = \frac{r}{M(r)}\frac{\mathrm{dM}(r)}{\mathrm{d}r} = \frac{4\pi r^3\rho(r)}{M(r)} \\ V &= -\frac{\mathrm{dlog}P(r)}{\mathrm{dlog}r} = -\frac{r}{P(r)}\frac{\mathrm{d}P(r)}{\mathrm{d}r} = \frac{G\rho(r)M(r)}{rP(r)} \end{aligned} \tag{14}$$

[30] For completeness, it should be pointed out that, when using this method, Strömgren also relied on previous works by himself (Strömgren 1931a) and Cowling (Cowling 1931).

[31] Cowling (1934, 1935a); see also Cowling (1935b, 1966, pp. 131–132).

[32] In the following, I will always speak of "CNO cycle", whereas in the original papers this nuclear process is also labelled as "CN cycle", "carbon cycle" etc.

[33] Cowling (1966, p. 131).

[34] Kippenhahn and Weigert (1990, p. 200); it should be pointed out that "the derivatives of the logarithm of any one of the variables with respect to any other logarithm can also be used." (Cowling 1966, p. 130). It was Milne who first introduced these homology-independent variables.

Such quantities can be used any time homologous models are constructed. This is the case for polytropes: two polytropes having the same index n are homologous. This is also the case for the point-source model if κ depends upon ρ and T via a power-law, and provided radiation pressure is negligible.[35]

As r varies from 0 to R, U and V vary accordingly and draw a curve in the U–V plane. It may be useful to discuss a stellar structure in that plane as all homologous models (e.g. all polytropes of a same index) are described by the same curve. Most interesting for us are the aforementioned cases in which we have polytropes with indexes 3/2, 3 and ∞.[36] All the corresponding curves have the stellar centre in $U = 3$ and $V = 0$, but the $n = 3/2$ and $n = 3$ polytropes move towards $U \to 0$ and $V \to \infty$ as long as the star's surface is approached, whereas the curve spirals around the point $U = 1$ and $V = 2$ in the isothermal sphere case ($n = \infty$).

As long as polytropes of different indexes are used to describe the inner and the outer regions of a star, it is convenient to investigate the possibility to fit different structures on the U–V plane. This, as we shall see, is a key point that scholars will develop to work out the structure of giant stars. In particular, the polytropic approximation was widely used before the computer age, when the hand-made numerical integration of the full stellar structure equations was excessively long.

Consider then a star whose core and envelope are described by different models. At the interface between core and envelope the two solutions must be fitted together. Here T and P change continuously but this may not be the case for ρ. If subscripts "c" and "e" stand for "relative to the core solution" or "to the envelope solution", respectively, for the perfect-gas law (if radiation pressure is negligible) we have that at the interface:

$$\frac{\rho_{\mathrm{e}}}{\rho_{\mathrm{c}}} = \frac{\mu_{\mathrm{e}}}{\mu_{\mathrm{c}}} \tag{15}$$

If there is no discontinuity in chemical composition, then we can suppose that, besides T and P, ρ is continuous at the interface, too. The fit between the two curves will then occur continuously in the U–V plane. But if a chemical discontinuity is assumed, then for Eq. 15 ρ passes through a discontinuity, too, and the U–V plane will display a sudden jump of the kind $(U_{\mathrm{c}}, V_{\mathrm{c}}) \to (U_{\mathrm{e}}, V_{\mathrm{e}})$, where the U- and V-values are calculated at the core-envelope interface. In other words, the values of U and V "just inside" and "just outside of" the interface will be different. As both U and V are directly proportional to ρ, this jump will occur along a straight line through the origin. Consequently two polytropes can be fitted together by starting at the point $U = 3$ and $V = 0$ and moving along the curve that represents the index n of the nucleus until the nucleus' mass is reached, and by subsequently moving (or possibly jumping) to a curve that has the index n of the envelope and by following it until the whole stellar mass is reached.[37]

35 Chandrasekhar (1939, pp. 234–239).

36 The curves corresponding to these cases can be found in Collins (1989, p. 49) and Kippenhahn and Weigert (1990, p. 201).

37 See for example Fig. 2.3 in Collins (1989, p. 52). Such a process clearly requires fixed boundary conditions, such as the stellar and the core's mass, or other quantities that determine these ones.

3 Historical background: a brief synopsis of stellar evolution theories up to the early 1930s

3.1 Early theories

Stellar evolution theories are as old as the first spectral classifications schemes.[38] As astrophysicists began to identify differences in stellar spectra and to work out relations between spectra and temperatures, they started to wonder whether, and in what ways, stars could evolve from one kind to another, possibly by heating up or cooling down in the process. Towards the end of the nineteenth century, tentative theories interpreted the temperature sequence as an evolutionary path. Basically, stars were thought to gain energy in some unknown way and then to cool down, thereby changing colour. The idea was substantiated by the observation that hot blue stars were typically located in the neighbourhood of nebular regions, which in turn were suspected to be involved in stellar formation.

Alternative theories speculated about a possible stage of contraction that heated up the star's gas, followed by a cooling phase when contraction stopped. Scholars consequently assumed that stars might overstep the spectral sequence twice—first in one direction, then in the opposite one. They guessed that the failure of the perfect gas law was the phenomenon responsible for the arrested contraction. Such a picture was quite useful, in that it explained where the star's thermal energy came from: a gravitational contraction along the Helmholtz–Kelvin timescale.

3.2 Russell and Eddington on stellar evolution

Further advances were not possible until accurate, quantitative measures of stellar temperatures and luminosities became available, i.e. until the H–R diagram was worked out and scholars realized the existence of giant and dwarf stars. That the heating-and-cooling theory turned out to be a very suitable interpretation key for the diagram itself is indeed noteworthy. Henry N. Russell suggested that the giant branch could correspond to the perfect-gas contraction stage, and the main sequence to the cooling one.[39] His theory also accounted for such facts as the different masses of different stars. Russell reckoned that stars of different mass must follow different paths—corresponding to different luminosities— in the giant region:

> A large mass of gas will therefore arrive at a higher maximum temperature, upon reaching its critical density, than a small one. The highest temperatures will be attained only by the most massive bodies, and all through their career these will reach any given temperature at a lower density, on the ascent, and return to it at a higher density, on the descending scale, than a less massive body. They will

[38] In this section I will just briefly sketch the evolutionary theories to the 1920s, without dwelling too much on the details. For further elaboration see for example Sitterly (1970, pp. 357–362), Tassoul and Tassoul (2004, pp. 84–88), Leverington (1995, pp. 128–129), Celnikier (2006, pp. 69–73).

[39] Russell presented this theory in the very same paper in which he published his first diagram in graphical form (Russell 1914).

> therefore be of much greater luminosity, for the same temperature, than bodies of small mass if both are rising towards their maximum temperature. On the descending side the difference will be less conspicuous. Bodies of very small mass will reach only a low temperature at maximum, which may not be sufficient to enable them to shine at all.[40]

Russell's theory was a pretty natural step to take at the time. Nevertheless, it would be totally discarded within a few years. Reasons for this lie in Russell's lack of knowledge—not so much about the observational features of the stars (masses, luminosities, temperatures), but rather about their internal structure, chemical composition and the phenomena involved in their energy supply. Generally speaking, what Russell really lacked was a theory of matter at the atomic and nuclear level. As time went by and progress was made in these fields, it became increasingly clear that new theories were needed.

When Eddington developed the mass–luminosity law, Russell's theory could not but be abandoned. In fact, Eddington realized that both the giant and the dwarf stars are in a perfect gas condition, so no failure of this law actually occurred. Moreover, the only way to explain the huge difference in luminosity between a giant and a low main sequence star was to suppose that a star burned out almost its whole mass during evolution. Unfortunately, no known process was able to account for this. Eddington mulled over the problem and devised a new possible evolutionary path for a star:[41] he guessed that the main sequence could be identified as a sequence of quasi-equilibrium structures of different masses, and that any evolutionary path would move almost horizontally through the diagram to account for the little (at most) changes in the star's mass. Whatever the yet unknown source of energy capable to fuel a star may be, such source must be activated when the star reaches the main sequence, be kept active for as long as the star remains there (allegedly for most of stellar life), to finally become exhausted and boost the star's departure from the main sequence. In this respect, the giant stage could be passed through both before and after the main sequence period.

In 1925, Russell came back to the stellar evolution topic.[42] He could avail himself of the major breakthroughs in atomic physics and stellar structure theory achieved in the meantime. However, no clear evidence about the nature of the energy-supplying processes had emerged yet. Russell was reluctant to abandon the idea of evolution along the main sequence, so he—driven by Einstein's mass-energy equivalence principle—guessed that some kind of matter annihilation must be at work. Although he was not able to specify anything more about this, he kept thinking that mass loss caused a star to descend the main sequence during its life. Moreover, the increasing density and consequently increasing opacity of the stellar gases lowered the surface temperature, thus enhancing the process.

Russell realized that this did not explain why the main sequence was where it actually is in the diagram, i.e. why certain luminosities—and not others—correspond to the different temperatures. Thus, relying on Eddington's stellar structure model,

---

40 Russell (1914, p. 285).

41 Eddington (1924b, pp. 325–328).

42 Russell (1925). See also Hufbauer (2006, pp. 207–208).

he calculated that an inclined line could be obtained in the position of the main sequence on the diagram by assuming that all stars of different masses had a same central temperature of about 30 million degrees. If—Russell argued—energy supply is enhanced at these temperatures for some unknown reason, this should explain the existence of the main sequence. In turn, giant stars corresponded to lower central temperatures.

This theory was, to some extent, a mix between Russell's former one and Eddington's: it built on the major achievements of the British astronomer in describing stellar structure, but it still retained the idea that giant stars are younger (cooler inside) and that stars descend the main sequence as they burn out their mass.

To explain the relative stability of giants, Russell guessed that they could go through a stage where a different energy-supplying process settled in. He spoke of "dwarf stuff" and "giant stuff", by which he meant the two kinds of matter undergoing energy-generating processes in the two cases. As we shall see, this idea of two different kinds of fuels would become very popular again in the following decades.

Russell's 1925 theory displays a number of modern features. It explains the main sequence as a sequence of different masses and similar central temperature, and it argues that the energy output is due to a cause other than gravitational processes. However, it was to be abandoned when scholars realized that the timescale of evolution would have to be incredibly long for that to be the case—for example, it would take the Sun some $M_{\odot}c^2/L_{\odot} \approx 10^{11}$ years to burn out most of its mass. This became untenable when, around 1930, the first estimates of the lifetime of the Universe yielded a value around 10^{10} years.[43] It followed that no actual star could have evolved much and consequently low-mass dwarfs should not exist at all.

3.3 Further developments up to the Thirties

Meanwhile, other astronomers were getting involved in the evolution issue. In the years around 1920, Harlow Shapley and Robert J. Trumpler carried on important works on stellar clusters. In 1925 an analysis by Trumpler about stellar types in clusters appeared.[44] Trumpler analysed about 50 open clusters and outlined the differences in their $H–R$ diagrams, noticing how in some cases the whole main sequence (but no giants) appeared, while in others giants could be found and the main sequence was cut off above a well-defined brightness. To explain this variety Trumpler guessed that differences in the initial mass of the clusters could play a role, thus implying that it was not possible to figure out a definite evolutionary line among the different kinds. But, more importantly for us, he pointed out how "we are led to the conclusion that the open clusters are already of considerable age; otherwise we would not find the dwarf branch so well formed in all cases, nor could we explain the general scarcity or total absence of yellow and red giant stars".[45] This conclusion is essentially the opposite of what we know today. The idea to rely on clusters as ideal laboratories was retrieved a decade later by Gerard P. Kuiper, who speculated that clusters could be

43 Hufbauer (2006, pp. 210–211).

44 Trumpler (1925). On Trumpler's contributions, see also Tassoul and Tassoul (2004, p. 112).

45 Trumpler (1925, pp. 317–318).

groups of stars of different hydrogen abundance.[46] Kuiper came to this conclusion as he observed that the curves representing clusters in the $H-R$ diagram were similar to curves of different hydrogen contents Strömgren had worked out in 1933.[47] As for the evolutionary role of giants, Trumpler and Kuiper remained faithful to the most popular ideas of their times. Strömgren, however, would make a real step forward and would play a key role in the evolutionary theories of the thirties. His contribution is described in detail in the next section.

4 The influence of chemical composition upon evolution

4.1 Strömgren on the influence of changing chemical composition upon evolution

If we consider the scenario of stellar evolution theories around 1930, we notice how, all in all, the idea of a varying chemical composition in stars was barely taken into account. This was natural to some extent, as only in the years 1929–1932 did scholars become aware of the great abundance of hydrogen. In 1929, Russell reckoned that stars must be mainly made of hydrogen,[48] and in 1932 both Strömgren and Eddington figured out that stellar models matched real stars only if a great hydrogen abundance was taken into account.[49] As they focused on it, astrophysicists began to wonder whether chemical composition had to play a role in evolution. The well-known Vogt–Russell theorem fostered this idea. This theorem states that under proper, reasonable assumptions[50] the structure of a star is entirely determined by its mass and composition to such an extent that stars of different masses but with the same chemical composition must lie along a continuous curve in the $H-R$ diagram. This seemed to be a promising way to explain the main sequence and, possibly, giants as stars of different composition.

It was Strömgren who tackled this issue.[51] His starting points were the Vogt–Russell theorem on one side, and the ascertained influence of the hydrogen content upon the equilibrium structure of a star on the other. Instead of starting from a theoretical estimate of the hydrogen content to work out observable stellar features, Strömgren reversed the method. He began from known radii, masses and luminosities of stars to figure out their hydrogen content. He adopted a stellar composition made up of

46 Kuiper (1937).

47 Strömgren (1933).

48 Russell (1929).

49 Strömgren (1932) and Eddington (1932). While Russell assumed that hydrogen must be much more abundant than any other element, Strömgren's and Eddington's calculations pointed out that percentages around 30 and 99 were both suitable. They opted for the lower value just because they deemed it to be less extreme. In the following decades scholars realized that the higher values was indeed much closer to truth. The recognition of the influence of hydrogen content upon the stellar structure is a major step in the history of astrophysics. See for example DeVorkin and Kenat (1983b) and Cenadelli (2008).

50 Namely, that "the compressibility, the opacity and the energy generation of an arbitrary volume element of the star depend only on its temperature, its density and its chemical constitution." (words by Strömgren, see Strömgren (1933), p. 239).

51 Strömgren (1933).

hydrogen plus a "Russell mixture"[52] of other elements, and attempted to work out the abundances of these two components. Both Russell, and Strömgren after him, considered helium absent.

The mean molecular weight, i.e. the parameter through which chemical composition appears in stellar structure equations, strongly depends upon the hydrogen abundance but is almost insensitive to the exact proportion of other elements in the mixture. This means that, at least as a first estimate, specifying how much hydrogen and how much other stuff is present is enough. Strömgren used the "standard model" as a basis, took into consideration both photon absorption and scattering as sources of opacity, and came to the significant conclusion that stars with the same mass but different radius must differ in hydrogen content. He stated:

> Comparing stars of equal mass but of appreciably unequal radius, we see that generally the hydrogen contents differ appreciably. Let us consider for instance Algol A and Capella A. The masses are about equal. Algol A is a B8-star with comparatively small radius, while Capella is a G0-star of comparatively large radius. Algol A according to our calculations contains 53% hydrogen, while Capella A contains only 30%. This difference in hydrogen content implies a difference in the structure which may well influence the subatomic energy-sources in such a way that the equilibrium configurations have quite different radii. [...] Quite generally we see that for stars of the same mass the hydrogen content diminishes with increasing radius.[53]

In fact, Strömgren's hydrogen content estimates are too low for both stars, but they are still noteworthy for that time. Strömgren worked out the hydrogen content of stars of different masses and radii. Secondly, he calculated their luminosities from the stellar structure equations, and finally their surface temperature via the Stefan–Boltzmann law. This allowed him to depict the curves corresponding to different hydrogen contents on a H–R diagram, with hydrogen-depleted stars lying high up and to the right (Fig. 1).[54]

Strömgren discussed the evolutionary implications of his work. Starting from the idea that a star could evolve following changes to both its hydrogen content and mass, he noticed that two different time scales could be devised for those changes—a shorter one that he called "intermediate time-scale" for hydrogen content, and a longer one that he called "long time-scale" for mass. Strömgren gave a satisfactory answer to the problem of the very long evolutionary time of stars that had bothered Russell, when he observed:

[52] This was the name given to the mixture of elements other than hydrogen, with the abundances that Russell had calculated in (1929).

[53] Strömgren (1933, pp. 239–240).

[54] As pointed out by Tassoul and Tassoul (2004, pp. 112–113), a basic reason for this can be found in Eddington's mass–luminosity relation, that is actually a M– L–μ–T_{eff} relation. At fixed M, it is possible to trace curves in a T_{eff}–L plane for several μ-values. A decrease in hydrogen content triggers a boost in μ and, due to the positive (and large) exponent of the term μ in the aforementioned relation, this causes—for a fixed T_{eff}—a strong increase in L, i.e. a moderate increase in radius.

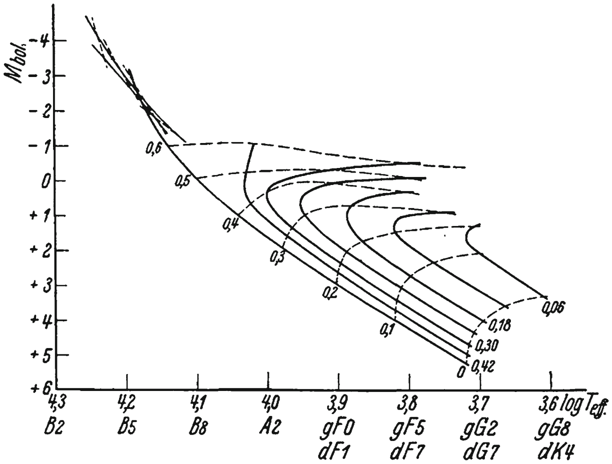

Fig. 1 The *continuous lines* in this *H–R* diagram correspond to equal hydrogen content (values are indicated as fractions of the total mass); the *dotted lines* correspond to equal mass and varying hydrogen content and are possible evolutionary trends. An evolutionary trend towards larger radii emerges, apart in the region of *B* stars where superpositions between different curves occur (from Strömgren 1933, p. 244). Reprinted with kind permission of the Springer Science and Business Media

> The hydrogen content may change during times of the order of the intermediate time-scale. Times of the order of the long time-scale must elapse before the mass changes appreciably. One would therefore expect that the tracks of evolution in the H.-R.-diagram are the lines of constant mass. One would further expect that the hydrogen content decreases, so that the stars expand.
> Perhaps the simplest hypothesis that can be made is that the stars start as pure-hydrogen stars in which, in the course of time, hydrogen is transformed to complex elements, the energy radiated away in the successive equilibrium configurations being equal to the energy set free by the transformation. On this hypothesis the rate at which hydrogen is used up is given by the luminosity, and the course of evolution can be followed.[55]

4.2 The helium issue

In those same years, Strömgren had also dealt with the helium issue.[56] Helium was not a major problem at first for astrophysicists, as its spectral lines are only visible in the hottest stars. Russell disregarded this element in (1929), when determining the "Russell mixture" Strömgren referred to. On the other hand, its presence in stellar interiors does not significantly affect mean molecular weight and thus it was not a

55 Strömgren (1933, p. 247).

56 Strömgren (1938). For further information about this, and more generally about Strömgren's scientific personality and the tight relationship between his interests in stellar structure and spectroscopy, see also Rebsdorf (2007).

great concern for Eddington or Strömgren himself when working on the inner composition of stars. Nevertheless, as we shall see in the following sections, by the end of the 1930s knowledge about stellar energy-generating processes had improved considerably, and the role of helium had been clearly pointed out. This fostered interest in the astrophysical role of this element.

Strömgren made explicit reference to this when tackling the helium problem. He noticed that it was necessary to account for helium in stellar interiors and to possibly work out its abundance through independent methods, to compare it to the theoretical predictions based on nuclear processes. Strömgren decided to carry on an analysis based upon the laws of stellar structure, in the same way he had done for hydrogen.

Let us call the abundance of hydrogen (fraction of total mass) X, the abundance of helium Y, and the abundance of other (heavier) elements Z. If we consider a gas solely made of hydrogen plus other stuff, i.e. if no helium is present ($Y = 0$), Strömgren observed that we have:

$$\begin{aligned} X + Z &= 1 \\ \mu^{-1} &= 2X + \frac{1}{2}Z \end{aligned} \tag{16}$$

while in the actual case $Y \neq 0$ we have:

$$\begin{aligned} X + Y + Z &= 1 \\ \mu^{-1} &= 2X + \frac{3}{4}Y + \frac{1}{2}Z \end{aligned} \tag{17}$$

Both Eqs. 16 and 17 hold if we suppose that a very high (total) degree of ionization is present, as does the second equation in (17) if Z is supposed to be $\ll 1$. In the former case there is but one degree of freedom, namely the parameter μ that is derived from the mass–luminosity law applied to stars of known mass, luminosity and radius. In the latter case one further assumption is needed to solve the system. By exploiting the mass–luminosity relation and the hypotheses that $Z \ll X$ and $Z \ll Y$,[57] Strömgren succeeded in determining X, Y and Z for several stars suitably distributed in the H–R diagram.

He commented that the hydrogen abundance—which turned out to be between 0.79 and 0.19 for the different stars—was not very much affected by the hypothesis of helium presence. This is natural, as Y and Z have a similar coefficient in the second equation of (17), i.e. helium contributes to the mean molecular weight in a way that is similar to heavy elements and much more different from hydrogen. Helium abundance turned out to be between 0.19 and 0.74—larger where hydrogen's was smaller, and vice versa. Heavy elements were always around a few percentage points. Although the helium abundance Strömgren deduced is admittedly tentative, his attempts show

[57] The hypothesis that helium is much more abundant than the heavy elements was deduced by Strömgren from considerations about nuclear reactions (see Strömgren 1938, pp. 520–521).

that hydrogen and helium are supposedly to be considered the main components of stars.[58]

4.3 A focus shift from mass to chemical composition

Strömgren's work is an early attempt to incorporate chemical composition changes into evolutionary models. Even if they were to be abandoned later on, his conclusions about hydrogen percentage had the merit to shift attention from mass to chemical composition changes when studying the evolutionary path of a star. Moreover, this was also the first time when the giant stage was reckoned to be a possible later stage in star evolution. This turned out to be a seminal issue in the following years, when scholars began to inquire into the role of changing chemical composition. On the other hand, at that time Strömgren could not but rely on the theoretical background constituted by Eddington's "standard model" and coeval theories (including his own). This means that his results were subject to some degree of speculation about the possible distribution of energy sources within a star. When faced with the need to make a decision, Strömgren adopted the "standard model" as a basis to elaborate upon.[59] To some extent, Strömgren pushed the speculations of the pre-nuclear era to their limit, but his achievements—remarkable as they are—cannot go beyond these limits and rather point out how a precise, quantitative description of such terms as ε and η was growing necessary.[60]

However, it is important to underline a meaningful fact: speaking of a changing chemical composition in a star as a whole is one thing, and speaking of changes occurring within a star (i.e. chemical inhomogeneities) is another. As we shall see in the sections below, this turned out to be a crucial point in the following decades. Strömgren focused only on the former case. According to him, stars evolved as long as their overall composition changed.

5 The inclusion of nuclear reactions into stellar evolution theories

5.1 A major advance: nuclear physics and the stars

During the 1930s, focus shifted to the topic of energy generation in stars. This issue was beginning to get sound answers, but in stellar models the new achievements about

[58] It should be noticed, though, that the same conclusion was reached by Cecilia Payne in 1925, although it was called into question and thought to be spurious at the time. About this latter issue see Hearnshaw (1986, pp. 229–231) and DeVorkin and Kenat (1983a, pp. 124–127).

[59] Following Eddington, he opted for a model in which the concentration of the sources towards the centre corresponded to $\alpha = 2.5$. See also note 18.

[60] As the main flaws of the Danish astronomer, Sitterly points out (Sitterly 1970, pp. 364–365) the fact that he relied on the standard model and did not account for other models (such as Cowling's, that had become available in the meantime); also, that he did not develop a real theory of evolution, although he gave important indications. Those underlined by Sitterly are real "flaws", as they were fully within reach in those years. On the other hand, the limits I previously pointed out as typically found in Strömgren's theories are— quite literally—natural "limits" for the time, i.e. they cannot be attributed to Strömgren but rather to the current lack of knowledge about crucial topics (e.g. nuclear reactions).

nuclear reactions had not found any application yet. Stellar astrophysics was due to slip into a dead-end street if no details about the energy-supplying processes could be provided, as no further understanding of stellar structure and evolution would be possible in that case. In the same way as the impressive growth of stellar astrophysics in the twenties was made possible by atomic physics, nuclear physics could yield major breakthroughs as well. As for the evolutionary role of giant stars, nuclear reactions provided the reason why—and the ways through which—stars' composition changed over time.

The problem was first tackled by George Gamow, who focused on the details of nuclear processes when describing stellar structures. It is no mere coincidence that this scientist was skilled in both nuclear physics and astrophysics. Gamow worked out new stellar models that—although based upon simplified assumptions—gave remarkable results and paved the way for subsequent, more reliable theories. In this sense, Gamow's work really throws a bridge from nuclear physics to astrophysics—in a way somehow similar to what Saha had done a couple of decades before, when he linked atomic physics to stellar spectroscopy. Nuclear physics was at the forefront of theoretical physics research in the thirties, just like atomic physics had been in the previous 20 years. This analogy cannot be pushed too far, though. Saha's theory had been awaited for a long time and really constituted a discontinuity, as long as it pointed out that there exists a tight connection between atoms and stars. This achievement was epitomized better than elsewhere by Eddington's famous statement: "The road to a knowledge of the stars leads through the atom, and important knowledge of the atom has been reached through the stars".[61] Eddington's words have a meaning that goes beyond any distinction between atomic and nuclear physics. They rather stress the way a new field of investigation—theoretical astrophysics—was born. On the other hand, we can rather figure out Gamow's contributions as major steps in a process that was still going on.

5.2 Gamow's simplified stellar models

Gamow retained the hypothesis that stars remain in an overall homogeneous state at each evolutionary stage, although their composition changes in time. His assumptions may be summarized as follows:

(1) Stars evolve as hydrogen is burnt into helium; all the hydrogen in a star will undergo this process;
(2) As evolution goes by, any stars goes through different, slowly varying equilibrium conditions that are homologous.

As long as details about nuclear reactions became available, scholars could decide whether to lean towards one stellar model or the other more consciously than ever before. For example, a "point-source" model, i.e. a model in which stellar energy is produced in a very small region at the centre of the star (this is Cowling's case), is a suitable choice for any energy-generation process that is strongly dependent upon

[61] Eddington (1927, p. 10).

temperature. On the other hand, if that dependence is not so steep, a model with energy sources distributed throughout a wider zone is more suitable.

Gamow considered both a "point-source" and a "shell-source" model, the latter being a model in which energy is generated in a shell surrounding an inert core. The former case is a good description if barrier penetration is at work—an exponential dependence is expected in this case—whereas the latter applies if there happens to be some nuclear resonance. If this is the case, once the resonance temperature is reached the nuclear reactions start and, as time goes by, they migrate towards the outer regions as the fuel in the inner is exhausted. Over time, the resonance temperature is reached at larger and larger radii. This is a suitable model to describe an isothermal, exhausted core surrounded by a thin shell of burning hydrogen.[62]

In other words, Gamow pointed out how the exact location of the energy-supplying processes within a star had a strong influence upon its evolutionary pattern.

Gamow started from the standard equations:[63]

$$\begin{aligned}
\frac{\mathrm{dP}(r)}{\mathrm{d}r} &= -\frac{GM(r)\rho(r)}{r^2} \\
\frac{\mathrm{d}M(r)}{dr} &= 4\pi r^2 \rho(r) \\
P(r) &= \frac{K\rho(r)T(r)}{\beta\mu\mathrm{H}} \\
P_{\mathrm{rad}}(r) &= P(r)(1-\beta) \\
\frac{\mathrm{d}L(r)}{\mathrm{d}r} &= 4\pi r^2 \rho(r)\varepsilon(r)
\end{aligned} \tag{18}$$

(where all symbols have their usual meaning).

Moreover, Gamow adopted Kramer's Law for opacity (Eq. 10). To these equations, we must add an equation for energy transport. If radiative equilibrium settles in, it is Eq. 13. However, if convective equilibrium arises, Eq. 13 must be replaced by:

$$\frac{\mathrm{d}T(r)}{\mathrm{d}r} = -\frac{\gamma-1}{\gamma}\frac{\mu\mathrm{H}G}{K}\frac{M(r)}{r^2} \tag{19}$$

where γ is the ratio of the specific heats in the gas c_p/c_v. Equation 19 holds as long as the convective cells are supposed to undergo quasi-adiabatical changes. Coupling it with the perfect gas law, it follows that:

[62] This is a likely occurrence in certain stages of stellar evolution. In fact, once the core-burning processes stop, if the core does not undergo abrupt stages of contraction and keeps an overall hydrostatic equilibrium, we have $\varepsilon \approx 0$ and, as $\mathrm{d}L/\mathrm{d}r \approx \varepsilon$, L = const. As the central luminosity is zero, we have $L \approx 0$ throughout the core. This prevents any major energy flow outside the core—at least in first approximation—so that hydrostatic equilibrium can be maintained without any need for a contraction. Moreover, as $\mathrm{d}T/\mathrm{d}r \approx L$, we have that T = const., i.e. the core is isothermal. Obviously enough, perfect isothermality must be seen as an idealization, because a slow gravitational contraction can settle it. But the energy production rate can be low enough not to remove the condition $\varepsilon \approx 0$ in any significant way. See Beech (1988, p. 219).

[63] Gamow (1938a).

$$P \propto \rho^{\gamma} \propto T^{\frac{\gamma}{\gamma-1}} \tag{20}$$

Gamow first considered a point-source model, for which the last of Eq. 18 must be replaced by $L(r) = L = F(\rho_0, T_0)$ for any value of r (ρ_0 and T_0 are the central values of ρ and T and F is an appropriate function of such values). In this case, from Eq. 13 it follows that $\mathrm{d}T/\mathrm{d}r$ increases indefinitely as r approaches 0, as long as other r-dependent quantities are expected to remain finite and $\neq 0$. Consequently, as r diminishes and we approach the innermost region of the star, at a certain radius R^* a superadiabatic gradient is reached and convection settles in. Thus, energy transport occurs by convection within this radius and by radiation outside of it, in such a way that Eq. 19 holds for $r < R^*$ and Eq. 13 for $r > R^*$.

These equations allow a calculation of the stellar model, once the constant κ_0 and the function $F(\rho_0, T_0)$ are known, and a value for M is chosen. This means that for a given mass and chemical constitution everything is fixed, as chemical constitution determines both κ_0 and F. Of course, the way in which it determines F should be explained: this is the main issue in Gamow's paper, as thermonuclear reactions actually enter the model through F. Incidentally, we may notice that the idea to start from mass and composition and to work out R and L is a direct consequence of the possibility to describe the energy-generating processes. Before that became possible, it was rather a matter of taking L and R from observation, hence, introducing surface conditions in stellar models.[64]

Gamow threw into the fray the newly discovered properties of nuclear reactions.[65] As far as it was known, at the temperatures typical of stellar interiors only reactions between very light nuclei were expected to occur, with hydrogen fusion being the most important of them. Therefore, Gamow considered the possible ways in which hydrogen could undergo fusion and concluded that, in the case of a "Russell mixture" composition, "the disappearance of 9 units of mass of hydrogen will lead to the formation of 8 units of mass of helium and 1 unit of mass of heavier elements".[66] Then Gamow focused on how hydrogen could be transformed into helium and on the resulting dependence of ε upon T, ρ and chemical composition. He accounted both for the case in which the reaction was made possible by barrier penetration, and the case in which there was some resonance at stellar thermal energies, and he worked out two possible expressions for ε.

In the former case (point-source), nuclear physics suggested how a changing chemical composition could affect the values of F and κ_0. All luminosity is produced in a small volume ω around the centre and it can be expressed as follows:

$$L \approx \varepsilon \rho_0 \omega \approx X(1 - X)\rho_0^2 T_0^n \tag{21}$$

where X is as usual the hydrogen abundance and n has a suitable value, rather approximately estimated to be $\sim$7.

64 About this, see for example Hoyle and Lyttleton (1942a, p. 177).

65 Gamow relied (see Gamow 1938a, p. 598) on Atkinson and Houtermans (1929).

66 Gamow (1938a, p. 599).

From this starting point, Gamow engaged in the calculations of stellar models for different values of X and M. Fixing a value of M, he calculated the expected variation in L and T_{eff} caused by a given variation in X, i.e. in μ. Then, the role of X and M was reversed, and Gamow investigated the behaviour of stars of different masses when X was fixed. What is most important, he exploited the hypothesis that *homology relation should hold*, i.e he worked out the expected variations in L and T_{eff} triggered by changes in X (or M) through homology relations.

As for the case of variable X,[67] once the functions $L(X)$ and $T_{\mathrm{eff}}(X)$ are devised, it is possible to work out a track on the H–R diagram corresponding to a changing X value. Gamow found:

$$\frac{\mathrm{dLog}L(X)}{\mathrm{dLog}T(X)} \approx 5 \tag{22}$$

except for small values of X, where the value changes to 0.8. The value 5 is the approximate value along the main sequence, so that stars turned out to *evolve along the main sequence* as they burned out their hydrogen.

On the other hand, in the case of a constant X[68] and a variable mass, Gamow worked out an expression for $\mathrm{dlog}L/\mathrm{dlog}M$. The expression showed that stars of growing masses are placed in the top-left portion of the H–R diagram, as it happens along the main sequence.

Such results were plotted in Fig. 2

As for the evolutionary tracks at changing values of X, the conclusions are, in Gamow's own words, "extremely unsatisfactory".[69] They are, of course. Gamow concluded that: "It seems that the real behaviour of stars cannot be interpreted in terms of ordinary thermonuclear sources of energy",[70] and that this was possibly an indication that resonance temperature effects, rather than barrier penetration, were at work.

A posteriori, it is clear what is wrong in Gamow' model—not the barrier penetration, but rather the homology hypothesis. Gamow did not realize that he had put forward the idea that main sequence and giant stars are *not* homologous. As it was discovered later, a red giant's overall structure is very different from that of a main sequence star.

On the other hand, the conclusion Gamow arrived at in the $X = \mathrm{const.}$ case is a satisfactory explanation of the main sequence. More specifically, for main sequence stars of different masses the homology conditions apply much more closely than in the dwarfs-giants case.

In an effort to find a better explanation for evolution at decreasing values of X, Gamow was forced to assume that a resonance temperature did indeed exist, to make a

[67] Gamow observed that, once a value for a stellar mass is chosen, any mass variations during stellar life are to be considered negligible (as nuclear reaction theories of the time discarded any former belief about matter annihilation within stars), and that the actual mass defect is almost null.

[68] As for the value of X, Gamow kept faithful to the commonly adopted value of 0.35. See also note 49.

[69] Gamow (1938a, p. 603). See also Gamow (1938b) on this.

[70] Gamow (1938a, p. 603).

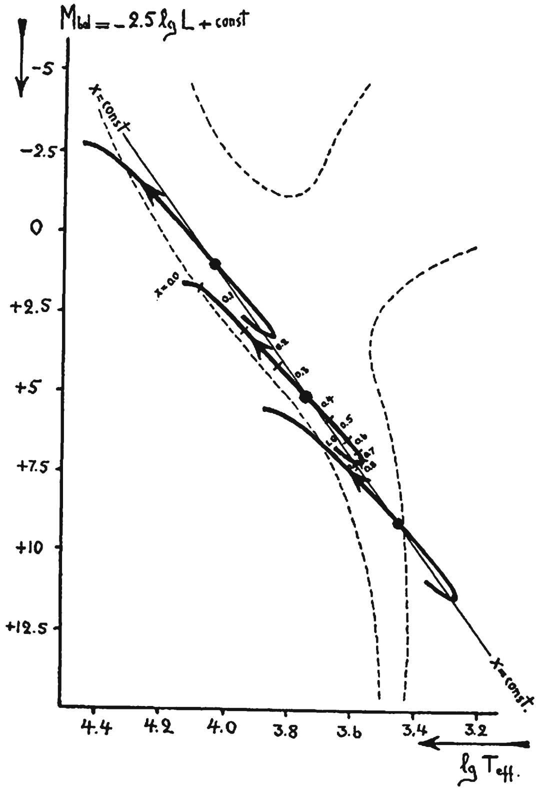

Fig. 2 The *heavy lines* are evolutionary tracks for stars of masses 0.5, 1 and 2 in respect to the Sun (adjusted so that the curve for mass 1 matches the position of the Sun at $X = 0.35$); the *full line* that crosses the diagram from bottom right to top left (labelled by $X =$ const.) is the line corresponding to different M values at the fixed value $X = 0.35$: it represents the main sequence (from Gamow 1938a, p. 603). Reprinted with kind permission of the American Physical Society

shell-source model tenable, as we previously hinted.[71] In such a model, an isothermal core is surrounded by a shell where hydrogen burning goes on, then by a convective region and finally by a radiative envelope. Gamow did not tackle the full calculations in this case, but he performed a preliminary analysis that led him to conclude that in this case a slow increase in L and a slow decrease of R (accompanied by an increase in T_{eff}) were likely to take place. So the following picture of stellar evolution could be figured out: a star begins its life as a large contracting body whose central temperature is far below the resonance value. As times goes by, the star approaches the main sequence from the right side and can be described via a point-source model. When the resonance temperature is reached, the hydrogen-burning shell settles in. The star remains on the main sequence and its luminosity barely changes at all. When the hydrogen is exhausted, the star moves towards the white dwarf stage. At the end of the day, giant stars are young ones again.

71 Gamow (1938a, pp. 603–604), see also Gamow (1938b).

5.3 Further progress in nuclear physics and its incorporation into stellar models

Later on, Gamow learnt about a recent discovery by Hans A. Bethe and Charles L. Critchfield—namely that the energy in main sequence stars was very likely supplied by p–p collisions, a process that does not actually show any resonance temperature. Although yielding a better estimate for the exponent $n(n \approx 3.5)$ in formula (21),[72] however, this implied only minor changes in the evolutionary path.

In the meantime, von Weizsäcker and Bethe described the CNO cycle. Gamow immediately applied this achievement to his description of stars.[73] Both Bethe and Gamow himself[74] had suggested that in stars as massive as the Sun, whose central temperature was assessed to be $\approx 2 \times 10^7$ °C, the CNO cycle could be overriding, while in low-mass stars the p–p process must be dominant. Utilizing an improved version of Kramers' formula for opacity (namely $\kappa = \kappa_0\, \rho^{0.5}\, T^{-2.75}$) and leaving n undetermined to account for the two possible processes going on, Gamow again worked out evolutionary relations via homology transformations. The resulting evolutionary path was thus depicted in a R–L plane (see Fig. 3). When X becomes very small, both μ and κ_0 become independent from it and the evolutionary relations assume the simple form:

$$\left.\begin{array}{l} L \propto X^{-\frac{1.25}{n+1.75}} \\ R \propto X^{\frac{1}{n+1.75}} \end{array}\right\} \Rightarrow L \propto R^{-5/4} \tag{23}$$

If a shell-source model is considered, a further analysis by Gamow together with Critchfield pointed out that no striking changes from this pattern could be expected.[75]

Still, red giants are a problem. Main sequence stars do not evolve towards their region. In other words, Gamow's model does not explain why a giant star should be brighter—if a strong temperature dependence is assumed for nuclear processes—as the idea that a main sequence and a giant star should be homologous implies that giants should be *colder* inside:

> In fact, because of comparatively low temperatures and densities in the central regions of these stars ($T_0 \approx 1 \times 10^6$ °C, $\rho_0 \approx 0.001$) the rate of the carbon-nitrogen reaction will be negligibly small, and cannot account for the observed high luminosities.[76]

Hence, Gamow was forced to speculate whether other nuclear processes, possibly due to light nuclei like Li, Be or B, could settle in before the main sequence was reached.[77]

---

72 Gamow (1938c). Reference to Bethe's and Critchfield's work is at p. 907.

73 Gamow (1939b). Reference to Bethe is provided at p. 720. On this work by Bethe, see also Gingerich and Lang (1979, pp. 321–337).

74 Gamow (1939a).

75 Critchfield and Gamow (1939).

76 Gamow (1939b, p. 724).

77 See also Gamow and Teller (1939) on this. As we can see, Russell's old idea of a "dwarf stuff" and a "giant stuff" is here retrieved.

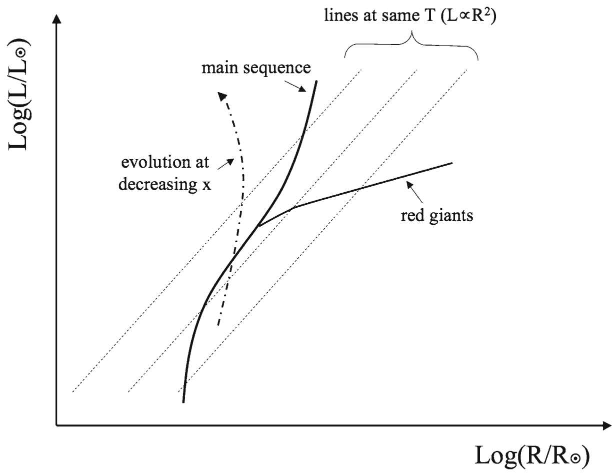

Fig. 3 Evolutionary path on the *R*–*L* plane. The *dotted line* matches the Sun's position when crossing the main sequence. In its *upper part*, where *X* becomes small enough, it follows Eq. 23 (figure rehashed from Gamow 1939b, p. 721). Reprinted with kind permission of the American Physical Society

A few years later, Schönberg and Chandrasekhar were to point out the main flaw of Gamow's approach:

> The assumption that the successive equilibrium configurations are homologous cannot be expected to be rigorously valid; for the nuclear reaction reduces the hydrogen content in the neighbourhood of the centre of the star, and therefore the molecular weight in this region becomes increasingly larger than that of the rest of the stellar material [...][78]

If we want to summarize Gamow's contributions till the end of the thirties, we can observe that his achievements, together with the work by other scholars like von Weizsäcker and Bethe, led to a sound description of the H → He process, and to a satisfactory understanding of main sequence stars. Gamow pointed out how the main sequence could be explained like a collection of stars of different mass and equal chemical composition, in turn fuelled by the CNO cycle or the *p*–*p* process.[79] But when the evolution following hydrogen depletion was investigated, a number of problems arose: a satisfactory understanding of giants was still lacking.

This clarifies the pros and cons of Gamow's theory. It has the merit to search the evolutionary patterns for stars in a detailed description of nuclear processes, rather than adding an undefined parameter ε as was the case before. At first, Gamow just claimed a process of hydrogen fusion must take place, but did not give any details about it. Later, as details became available, he carefully included them in a theory of

78 Schönberg and Chandrasekhar (1942, p. 161).

79 Gamow (1939c); Gamow incorrectly thought the Sun to be mainly powered by the CNO cycle, but this is not relevant to our discussion. On the other hand, he correctly recognized that the *p*–*p* process must be overriding in lighter stars.

stellar evolution. After Gamow's work, it became clear that no evolutionary theory could disregard a full description of nuclear processes.

As for the cons, Gamow focused only on a changing chemical composition that affected the star as a whole, rather than investigating the possible role played by chemical inhomogeneities that settle in within the star. This caused his attention to shift towards homologous states that are not correct in this case. However, Gamow's investigations paved the way for further steps towards the first reliable pictures of giant stars.

6 Early attempts to embed chemical inhomogeneities into stellar models

6.1 Öpik's pioneering work

It sometimes happens that remarkable work is published in inaccessible places and there it lies, almost unknown to coeval scholars. It is well known that this was the case for the first announcement of the relationship between star colour and brightness, made by Hertzsprung in 1905. The same happened with a work by Öpik on stellar evolution, that appeared in the *Publications de l'Observatoire Astronomique de l'Universite' de Tartu*,[80] in Estonia in 1939 and proved to be a crucial contribution to the solution of the giants issue.

Öpik admitted that "our discussion is but qualitative; a more definite picture can be obtained only with the aid of laborious calculations".[81] Nevertheless, his conclusions proved to have striking insight, in that he observed that a star for which mixing is very effective (i.e. a fully convective star), is expected to remain in the main sequence stage for its whole lifetime. However, if such mixing does not occur, hydrogen burning and exhaustion will cause the star to develop a composite structure that will drive it to a completely different stage. The nucleus will then begin to contract in the Helmholtz–Kelvin timescale. As for the envelope:

> Outside the nucleus the material is not exhausted; with the progress of the central condensation the temperature of the shell adjacent to the nucleus rises, and subatomic energy is released in an intermediate shell; the rapid increase of energy generation with increasing temperature and density in the intermediate shell prevents it and the rest of the star from being drawn into the overdense nucleus; on the contrary, if the outmost shell is in radiative equilibrium [...] it is forced to expand, and a giant star is formed.[82]

The main argument in favour of expansion was that a radiative envelope was unable to carry all of the energy flow from the star's interior. Consequently, structural changes were to be expected.

In other words, Öpik explained giants by developing a model consisting of three different zones, a decade before models of this kind became customary. In the inner-

80 Öpik (1939); On Öpik's contribution, see also Gingerich and Lang (1979, pp. 342–348).

81 Öpik (1939, p. 70).

82 Öpik (1939, p. 64).

most part we find a superdense, hydrogen-depleted, convective core collapsing under gravity. Then we have an intermediate, convective shell surrounding the core, where hydrogen is converted into helium. This nuclear process takes place at an accelerated rate due to the very high temperature triggered by the core collapse. Finally, the enclosing, radiative envelope is inert, hydrogen-rich and very extended, since it has to radiate all the energy produced by the star.

Öpik pointed out that Russell's idea of a "giant stuff" could be discarded: no exotic processes were at work in the build-up of a giant star. Such a star is fuelled by gravitational collapse and hydrogen burning in a shell outside the nucleus. The relative paucity of red giants can be explained by admitting that they burn up all available energy sources very quickly, as their huge brightness testifies.

The importance of Öpik's work has been emphasized in the literature.[83] On the one side, he made tentative speculations and proved to possess deep physical insight, more than the ability to make detailed calculations. However, his conclusions relied on sound, basic physical consideration. In particular, he underlined the role played by a discontinuity of μ, as well as the fact that evolution was not expected to go across homologous states.

6.2 Hoyle and Lyttleton: a first attempt to recover Öpik's work

At the beginning of the 1940s, it was Fred Hoyle's and Raymond A. Lyttleton's turn to give interesting contributions. First, they stressed the importance of details about nuclear processes in a paper entitled "On the internal constitution of the stars", published in the *Monthly Notices of the Royal Astronomical Society*.[84] Hoyle and Lyttleton showed once more that if a strong temperature dependence for ε is advocated, then a structure composed of a convective core and a radiative envelope should be expected. Holding on to the hypothesis that μ = const. across the star, they focused on main sequence stars. They exploited the homology hypothesis and achieved satisfactory relations for mass, luminosity, radius and composition, thereby contributing to "put the structure of main-sequence stars on a secure physical basis".[85]

In the same issue of the *Monthly Notices*, a further paper by the same two authors appeared. Its title was "On the nature of red giant stars".[86] In it, Hoyle and Lyttleton introduced the idea that μ could vary within a star, and namely that it assumed two possible values in the inner and the outer regions. They supposed the μ value to change discontinuously; although, "a continuous change in a layer of small depth could equally be discussed, but it can be verified that this case leads to no essential

83 According to Arny (Arny 1990, p. 221) "E. Öpik had successfully built red giants models"; Longair (Longair 2007, p. 61) states that "the solution to the red giant problem was discovered in 1938 by the Estonian astrophysicist Ernst Öpik"; Tassoul and Tassoul (2004, p. 148) reckon that "the first decisive step towards the resolution of the red-giant problem was made by Öpik".

84 Hoyle and Lyttleton (1942a).

85 Longair (2007, p. 188).

86 Hoyle and Lyttleton (1942b).

difference from that of a sudden change."[87] In doing this, they went back to Öpik's work.

The two scholars correctly identified the changes of μ occurring in a star as Öpik's main contribution. An approximate numerical integration gave evidence that "a decrease in μ in the outer part of the star will lead to a considerable increase in the actual radius".[88] It was a remarkable achievement—the first time, after Öpik, the role of chemical inhomogeneities was highlighted. However, Hoyle and Lyttleton's calculations showed that, to obtain large radii, "the regions of lower μ [must] contain only a minor part of the mass, say less than 25%."[89] This seemed to rule out an explanation *à la* Öpik, that the discontinuity in μ could be due to the H $\rightarrow$ He transmutation occurring in the core. It rather drove Hoyle and Lyttleton to focus on processes of hydrogen accretion in the stellar atmospheres, possibly from interstellar matter hitting the star's surface. The idea of focusing on the changing μ value *in the envelope* was quite original, and to some extent it can be claimed to be a precursor to current studies about surface accretion (e.g. blue horizontal branch stars). However, this was not how the red giants problem could be solved.[90] Rather, the μ value at the core–envelope interface would have to be investigated.

7 Inhomogeneous models and the Schönberg–Chandrasekhar limit

7.1 Henrich and Chandrasekhar on homogeneous stars

In the same period another major astrophysicist, Chandrasekhar, was focusing on that very problem and published two important papers on stellar evolution in the early 1940s.

Chandrasekhar's first paper was written together with Louis R. Henrich and appeared in 1941.[91] Henrich and Chandrasekhar focused on stellar models in which the core and the envelope were supposed to have different constitutions, i.e. they had to be described via different models to be further fitted for the values of the state parameters to coincide at the boundary between them. Henrich and Chandrasekhar considered two models. Both were endowed with an isothermal core that, as Gamow had shown, could indeed be there if resonance hydrogen burning or processes involving light nuclei occurred.[92]

The first case they discussed had a polytropic envelope with $n = 3$. The second featured a radiative point-source envelope with opacity given by (10) holding true anywhere in the star, although κ_0 was supposed to change as a function of chem-

[87] Hoyle and Lyttleton (1942b, p. 219).

[88] Hoyle and Lyttleton (1942b, p. 223).

[89] Hoyle and Lyttleton (1942b, pp. 223–224).

[90] Arny states that "it was not at all easy to see why there should be a sharp change in the molecular weight within the envelope" and that Hoyle and Lyttleton's model was "a dead end for models of red giants." (1990, p. 223).

[91] Henrich and Chandrasekhar (1941); a reference paper on Chandrasekhar's contributions is Srinivasan (1996).

[92] Gamow (1938a,b).

ical composition. Henrich and Chandrasekhar described the envelope via standard theories—the aforementioned cases are just the models by Eddington and Cowling—Strömgren. What is most important to us, μ was supposed to be the same in the core and the envelope. In fact, this work was published before Hoyle and Lyttleton's 1942 paper, in which the μ variation played a key role.

In the first case the authors considered, $T = T_0 (T_0$ = central temperature) everywhere in the isothermal core (a $n = \infty$ polytrope). It follows that:

$$P(r) = \frac{K}{\mu \mathrm{H}} T_0 \rho(r) + \frac{1}{3} a T_0^4 \tag{24}$$

In the $n = 3$ polytropic envelope, the two scientists employed Eq. 6.[93] These equations can be numerically integrated, the only constraint being that $P(r)$, $\rho(r)$, and $M(r)$ must be identical at the interface between the core and the envelope.

In short, Henrich and Chandrasekhar had to fit a $n = \infty$ to a $n = 3$ polytrope together. They decided to exploit the homology invariant quantities (14), and to write the fitting conditions as follows:

$$\begin{aligned} U_\infty(r_\mathrm{i}) &= U_3(r_\mathrm{i}) \\ V_\infty(r_\mathrm{i}) &= V_3(r_\mathrm{i}) \end{aligned} \tag{25}$$

where
r_i = radius of the interface;
U_∞ and $V_\infty = U$ and V calculated in the core ($n = \infty$ polytrope);
U_3 and $V_3 = U$ and V calculated in the envelope ($n = 3$ polytrope).

In other words, Henrich and Chandrasekhar opted for a continuous link in the U–V plane.

The second investigated case was a point-source envelope, which the authors assumed to be physically sounder as more akin to "the physical circumstances under which we might expect isothermal cores"[94]—those circumstances being, as we saw, an exhausted core surrounded by a very thin, hydrogen-burning shell. In this case, Eq. 24 still held for the core, whereas (6) was replaced by standard hydrostatic-equilibrium and energy-transport equations that were supposed to hold true:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{K}{\mu \mathrm{H}} \rho(r) T(r) + \frac{1}{3} a T(r)^4\right) &= -\frac{G M(r)}{r^2} \rho \\ \frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{1}{3} a T(r)^4\right) &= -\frac{\kappa_0 L}{4\pi c r^2} \frac{\rho(r)^2}{T(r)^{3.5}} \end{aligned} \tag{26}$$

In the second of Eqs. 26, L is a constant as $L(r) = L$ at any value of r in the envelope, and κ is supposed to obey formula (10). The fitting conditions were similar

[93] As is customary when solving polytropes, Henrich and Chandrasekhar replaced ρ and r with dimensionless quantities and used these variables for their equations. From a conceptual viewpoint, this makes no difference whatsoever.

[94] Henrich and Chandrasekhar (1941, p. 529).

to (25) but, since in this case the envelope wasn't described by a polytrope having a well-defined index, Eqs. 25 were written as follows:

$$
\begin{aligned}
U_{\infty}(r_{\mathrm{i}}) &= 4\pi \left(\frac{r^3 \rho(r)}{M(r)} \right)_{r=r_{\mathrm{i}}} \\
V_{\infty}(r_{\mathrm{i}}) &= \frac{\mu \mathrm{H}}{K} G \left(\frac{M(r)}{r T(r)} \right)_{r=r_{\mathrm{i}}}
\end{aligned}
\tag{27}
$$

It should be noted that such equations have a clearly defined meaning only if the core and the envelope are described by homologous models. This is surely the case for polytropes, but it can also be the case for point-source if κ is made to depend upon ρ and T via a power-law, and provided radiation pressure is negligible. Consequently, Henrich and Chandrasekhar restricted their investigation to this case.

Please notice that it is not the star as a whole to be treated as a structure undergoing homologous transformation, but rather the two different parts it consists of. This was a major step forward compared to the simple models Gamow developed a few years before.

Henrich and Chandrasekhar could exploit Strömgren's integrations for the point-source envelope.[95] These calculations had been performed only for a few possible values of L, M, R and κ_0, with $\mu = 1$, and much more general integrations were certainly desirable. Nevertheless, this was not a priority for the two scientists, who were rather interested in developing a new method to describe a star as consisting of two different parts to be treated separately and then fitted together by means of the homology-invariants U and V. Integration was merely a calculus problem that could be tackled separately.

Henrich and Chandrasekhar reached two important conclusions:

(1) There exists a maximum core mass expressed as a fraction of the whole stellar mass. Its value in the two models considered is around 0.38 or 0.35, respectively;
(2) Hydrogen burning is accompanied by a core contraction and an envelope expansion. Although brightness and radius only increase by a few units, it is clear that this is "a promising way to create giants".[96]

However, the two scholars stuck to the unrealistic hypothesis that μ remained constant throughout the star—an approach that would be discarded a year later in a further paper by Schönberg and Chandrasekhar.

95 Strömgren (1931b); Henrich and Chandrasekhar also relied on integrations performed by Nielsen under Strömgren's supervision (see Henrich and Chandrasekhar (1941, p. 531)).

96 Arny (1990, p. 222).

7.2 Schönberg and Chandrasekhar on inhomogeneous stars: the Schönberg–Chandrasekhar limit

Published in 1942,[97] the paper by Schönberg and Chandrasekhar is widely recognized as a milestone in the development of stellar astrophysics. The starting point for the two scientists was the overcoming of Gamow's homogeneous and homologous models. Schönberg and Chandrasekhar developed a model in which an inhomogeneity settles in between the core (where $\mu = \mu_c$) and the envelope (where $\mu = \mu_{\mathrm{e}} \neq \mu_{\mathrm{c}}$). The idea to treat the core and the envelope as two separate parts is fully applied here. The two astrophysicists considered two cases, namely a convective (Cowling) core and an isothermal one, both topped by a radiative point-source envelope. This decision was consistent with the standard evolutionary picture of the time, according to which a main sequence star has a convective, hydrogen-rich core that over time gets replaced with an isothermal, helium-rich one as hydrogen is burnt and fusion moves outwards in a thin shell.

Radiation pressure was neglected in both cases. Furthermore, $P(r)$, $T(r)$ and $M(r)$ were assumed to have the same interface values both in the core and in the envelope solution but, from the new assumption $\mu_{\mathrm{e}} \neq \mu_{\mathrm{c}}$, it followed that $\rho(r)$ passed through a discontinuity:

$$\begin{aligned} P_{\mathrm{c}}(r_{\mathrm{i}}) &= P_{\mathrm{e}}(r_{\mathrm{i}}) \\ T_{\mathrm{c}}(r_{\mathrm{i}}) &= T_{\mathrm{e}}(r_{\mathrm{i}}) \\ M_{\mathrm{c}}(r_{\mathrm{i}}) &= M_{\mathrm{e}}(r_{\mathrm{i}}) \\ \rho_{\mathrm{c}}(r_{\mathrm{i}}) &\neq \rho_{\mathrm{e}}(r_{\mathrm{i}}) \end{aligned} \tag{28}$$

In Eqs. 28 the subscript "c" stands for "calculated via the core equations" and "e" for "calculated via the envelope equations". The subsequent fitting equations take the following form:[98]

$$\left.\begin{aligned} U_{3/2}(r_{\mathrm{i}}) &= 4\pi \frac{r_{\mathrm{i}}^3 \rho_{\mathrm{e}}(r_{\mathrm{i}})}{M_{\mathrm{e}}(r_{\mathrm{i}})} \frac{\mu_{\mathrm{c}}}{\mu_{\mathrm{e}}} \\ V_{3/2}(r_{\mathrm{i}}) &= \frac{2}{5} \frac{G M_{\mathrm{e}}(r_{\mathrm{i}}) \rho_{\mathrm{e}}(r_{\mathrm{i}})}{r_{\mathrm{i}} P_{\mathrm{e}}(r_{\mathrm{i}})} \frac{\mu_c}{\mu_e} \end{aligned}\right\} \text{convective}$$
$$\left.\begin{aligned} U_{\infty}(r_{\mathrm{i}}) &= 4\pi \frac{r_{\mathrm{i}}^3 \rho_{\mathrm{e}}(r_{\mathrm{i}})}{M_{\mathrm{e}}(r_{\mathrm{i}})} \frac{\mu_c}{\mu_e} \\ V_{\infty}(r_{\mathrm{i}}) &= \frac{\mu_{\mathrm{e}} \mathrm{H}}{K} \frac{G M_{\mathrm{e}}(r_{\mathrm{i}})}{r_{\mathrm{i}} T_{\mathrm{e}}(r_{\mathrm{i}})} \frac{\mu_c}{\mu_e} \end{aligned}\right\} \text{isothermal} \tag{29}$$

In turn, the term $\mu_{\mathrm{c}}/\mu_{\mathrm{e}}$ is a function of time, as its initial value is 1 and it increases as nuclear processes take place. The key idea here was to get an understanding of

[97] Schönberg and Chandrasekhar (1942).

[98] Instead of using the authors' notation, I labelled the core quantities with the polytropic index (3/2 for a convective core and ∞ for an isothermal core) to preserve the previously used notation. Note that a non-relativistic degenerate core also fits the condition $n = 3/2$.

how the size and mass of the core, as well as the star's radius and luminosity, change as a function of μ_c/μ_e, as this ratio increases from 1 to 2 (which corresponds to a transmutation from a hydrogen-rich to a helium-rich mixture).[99]

In the convective core case, the relation:

$$3/2 = \left(\frac{\mathrm{dlog}\,P}{\mathrm{dlog}\,T}\right)_{\mathrm{c}} = \left(\left.\frac{\mathrm{dlog}\,P}{\mathrm{dlog}\,T}\right|_{r=r_i}\right)_{\mathrm{e}} \tag{30}$$

was assumed to hold. In other words, the effective polytropic index of the envelope at the interface was supposed to equal the core's.[100] Schönberg and Chandrasekhar found that both the radius and the mass of the core decreased over time, whereas the star's radius and luminosity increased by a factor of approximately 1.5 when central temperature remained constant.

However, in the isothermal case the hydrogen-burning shell moves outwards and the core's radius grows, with its mass first increasing and then oscillating around a maximum value. As a consequence, the radius and luminosity of the star increase by a factor of about 2 and then start oscillating, too. Actual temperature keeps constant for a while, then starts decreasing. Fig. 4 depicts the evolution on a L–T_{eff} plane.

In the isothermal core case, the authors came to a crucial conclusion. They found that for any value of μ_c/μ_e there exists a definite core mass fraction above which no connection with the envelope is possible. The core mass uppermost value is a decreasing function of μ_c/μ_e and reaches the maximum value for a homogeneous star, i.e. $\mu_c/\mu_e = 1$. When μ_c/μ_e increases to 2, the value gets as low as 10% ca. of the stellar mass. Today the maximum mass (expressed as the core mass/total stellar mass ratio) that an isothermal, non-fusing, non-degenerate core can have to sustain the weight of the overlying layers is known as the *Schönberg-Chandrasekhar limit*.

This limit can be expressed as:[101]

---

[99] In case of pure hydrogen being converted into helium, μ should change from 1/2 to 4/3, and consequently the core-to-envelope ratio should increase from 1 to 8/3. However, the presence of a certain amount of helium in the initial mixture, as well as a slight amount of heavier elements, smoothes out this difference and makes a transition from 1 to 2 an acceptable choice.

[100] About the definition of the "effective polytropic index" see Eddington (1938), see also Hoyle and Lyttleton (1946, p. 525). In this latter paper, Hoyle and Lyttleton questioned that such an index was continuous at the interface.

[101] Collins (1989, p. 131); the existence of such a limit can be intuitively understood on the basis of plausibility arguments. In fact, the aforementioned core configuration is described by a $n = \infty$ polytrope, for which P depends upon the first power of ρ, i.e. the *minimum* possible power. This means that an increase in the gravitational pressure by the envelope yields a *maximum* increase in ρ, which in turn entails a *maximum* increase in gravity. Consequently, one can argue that there must be a limit beyond which an increase in gravitational pressure is no longer sustainable by the core. See Smith (1995, p. 261). In more recent times it was recognized that stars with a mass between 1, 5 and 6 solar masses approximately develop an isothermal core at the end of the main sequence stage. Below 1, 5 $M_{\odot}$ degeneracy settles in, whereas above 6 $M_{\odot}$ the mass is so high that even a slight contraction turns into a large release of energy and isothermality does not actually hold. See Beech (1988, p. 219). This means that for stars within this range, the isothermal core mass grows while hydrogen burns in the shell, but when the Schönberg–Chandrasekhar limit is reached, the core can no longer sustain the pressure of the upper layers and begins to contract towards the helium ignition phase. The Schönberg— Chandrasekhar limit is a major feature of stellar evolutionary models. It

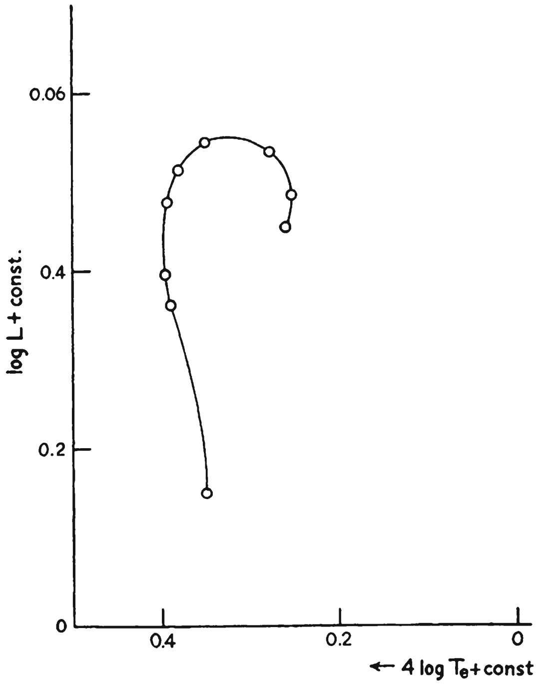

Fig. 4 Evolution on a L–$T_{\rm eff}$ plane in the isothermal core case. Despite some evolutionary inconsistencies, this pattern begins to show some similarity with an evolution upwards and to the right. Notice the wrong labelling of a point on the vertical axis (0.06 instead of 0.6) (from Schönberg and Chandrasekhar 1942, p. 171). Reprinted with kind permission of the American Astronomical Society

$$q_{S-C} = \frac{M_{\rm core}}{M_{\rm star}} = \frac{M(r_{\rm i})}{M} = 0.37 \left(\frac{\mu_e}{\mu_c}\right)^2 \tag{31}$$

However, if gas degeneracy happens to play a major role, the Schönberg-Chandrasekhar limit disappears.[102]

Schönberg and Chandrasekhar once again found a "promising way to create giants". Similar arguments were put forth by Marjorie Hall Harrison a couple of years later.[103]

To summarize, Schönberg's and Chandrasekhar's key achievement was the idea that the core and envelope must be treated differently due to chemical inhomogeneities. This allows a dynamical uncoupling between the two. Furthermore, Schönberg and Chandrasekhar (together with Henrich) introduced the idea of transferring the problem to the U–V plane to investigate it for different stellar configurations—something that would become a standard choice in the following years. They also estimated the maximum mass of the isothermal core by introducing the mass limit that bears their

Footnote 101 continued
would be beyond the historical perspective of the present paper to conduct a far-reaching analysis of this topic. I just cross-refer to Collins (1989, pp. 130–132), Kippenhahn and Weigert (1990, pp. 285–290) and Srinivasan (1996, pp. 64–68).

102 Kippenhahn and Weigert (1990, p. 289). If degeneracy settles is, no Schönberg–Chandrasekhar limit exists anymore: degenerate electron pressure contributes greatly to support gravitational pressure. This issue will be discussed later on in the present paper.

103 Hall Harrison (1944). Plots of the variation of R and L as functions of $\mu_{\rm c}/\mu_{\rm e}$ can be found at p. 345.

name. In so doing, the two scientists were entering uncharted territory and changing the background of the whole issue. In Arny's words: "From that point on, inhomogeneity became a basic ingredient in stellar models".[104]

On the other hand, their evolved configurations were still very small compared to actual giants. This is not surprising, as they considered the possible evolution of stars at the limit of the main sequence stage, but didn't fully account for further evolution along collapsing core stages. Schönberg and Chandrasekhar were perfectly aware of this:

> So far we have discussed the evolution of a star during the relatively early stages of the exhaustion of hydrogen in its central regions. The question now arises as to what can be said concerning the evolution during the later stages, i.e., after the isothermal core has grown to include the maximum possible mass. When this stage has been reached, the liberation of energy from the carbon cycle must cease, and we should expect the star to adjust itself to a contractive model [...] and evolve according to the Helmholtz-Kelvin time scale.[105]

It was becoming clear that the life of a star on the main sequence is limited to the time required to burn out about 10% of its hydrogen, after which it must evolve through a collapsing core stage—an issue that would be retrieved and further developed by Allan R. Sandage and Martin Schwarzschild a decade later (see Sect. 9.2).

Furthermore, Schönberg and Chandrasekhar did not fully account for other crucial issues such as any mixing of material within a star[106] or the settling in of degeneracies within the core. Finally, the choice of an abrupt discontinuity in the value of μ could be further refined into a steep (but continuous) change of its value.

8 Giant stars models with partially degenerate cores

8.1 Gamow's discontent with the Schönberg–Chandrasekhar limit

As we discussed in the previous section, some of the main ingredients required to build giants became available in the early 1940s. Others were still lacking, though. Among these was the possible degeneracy in the stellar core. Once again, it was Gamow who tackled this issue. Gamow felt uncomfortable with the Schönberg–Chandrasekhar limit. He commented upon it in two papers published on the *Physical Review* in 1944 and 1945.[107]

[104] Arny (1990, p. 223).

[105] Schönberg and Chandrasekhar (1942, p. 172). The two authors further speculated that the gravitational contraction and temperature rise could trigger a re-ignition of the carbon cycle at the border of the core, which would cause the core to become isothermal once again. But since the core was now expected to exceed the limiting mass fraction, the star was assumed to evolve through non-equilibrium configurations.

[106] For further details about this issue see Tassoul and Tassoul (2004, p. 150).

[107] Gamow (1944, 1945).

Gamow remarked that:

> [...] the exact mathematical treatment of the isothermal core model[108] seemed to lead to the conclusion that no such model is possible if more than ten percent of total stellar mass is to be contained in the core.
> It must be indicated here that the above paradoxical result is entirely owing to the arbitrary assumption that the isothermal core of the star must be governed by the ideal gas law, and disappears at once if we take into account the possibility of degeneracy near the centre of the core. Indeed the impossibility of fitting in a massive isothermal core of ideal gas (polytrope $n = \infty$) rests on the fact that above a certain density limit, the gas pressure in the centre of the core is unable to support the weight of the core thus leading to an unlimited contraction. Physically, however, such a contraction will be rapidly stopped because of the decreased compressibility of gas in the degenerated region. The stable state of a star with a core exceeding ten percent of the total mass, will thus consist of three different regions: (1) degenerate nucleus, (2) isothermal layer of ideal gas, and (3) radiative envelope.[109]

Gamow investigated this argument further with G. Keller in 1945. The two scientists made an attempt to insert degeneracy into the core of an inhomogeneous star.[110] The scenario they depicted starts from the assumption that main sequence stars burn out hydrogen via the CNO cycle. This is described through a Cowling's point source model, where convection takes place solely in the innermost part of the stars and the envelope is fully radiative. As time goes by, nuclear fusion increases the value of μ_c/μ_e from 1 to 2 and the core mass fraction grows, as Schönberg and Chandrasekhar (and Harrison after them) had shown. Gamow and Keller pointed out that stuff mixing within the core should be efficient to let it evolve as a whole. When hydrogen begins to be in short supply, the core starts shrinking and increasing in temperature. This in turn ignites nuclear fusion in a thin shell outside the core, and the core tends to become isothermal. The transition to a shell source model is complete by now. So a star should experience both cases Schönberg and Chandrasekhar had dealt with, in subsequent stages of evolution. When the energy-producing shell is established, it cannot but move outwards towards the surface. However, in so doing it is bound to exceed the Schönberg–Chandrasekhar limit that is expected to hold in the isothermal case. Schönberg and Chandrasekhar claimed that the star must enter a non-equilibrium state and eventually evolve towards explosion, but, as stated before, this outcome can be avoided by bringing into the equation another feature, namely gas degeneracy in the core.

Gamow and Keller declared their aim to be the investigation of:

> [...] the fitting of partially degenerate cores of fixed temperature $T^* = 2 \times 10^7\,{}^\circ K$ (corresponding to the C–N cycle in the energy producing shell) and molecular

108 In the original text, reference is made to Henrich and Chandrasekhar (1941) and Schönberg and Chandrasekhar (1942).

109 Gamow (1945, p. 120).

110 Gamow and Keller (1945).

weight $\mu_{\text{core}} = 2$ to radiative envelopes of molecular weight $\mu_{\text{env}} = 1$ and [a proper value for κ_0]. The value of the molecular weight chosen for the envelope corresponds to a hydrogen content of 35 percent. […] The fitting conditions are that the gas pressure and temperature must be continuous at the interface between the isothermal and radiative parts.[111]

As for the radiative envelope, the two scholars exploited custom equations and relied on the ever-present Strömgren's and Cowling's integrations, as well as on others they developed themselves. But their real focus point was the core. By calling the core radius and mass r^* and M^* (with ρ^* being the core density at the boundary), in the isothermal, perfect-gas case they found results in agreement with the findings of Chandrasekhar and colleagues (once again, radiation pressure was neglected). Gamow and Keller developed a $\log r^* - \log \rho^*$ relation for a given mass value $M^* = 0.1\, M_\odot$. The resulting curve did not extend into the small radii region:

Physically this means that it is impossible to construct an isothermal core of an ideal gas with a radius smaller than a certain value determined by its mass. This fact underlies the above mentioned result of Chandrasekhar and his collaborators concerning the impossibility of building a stellar model with a core containing more than 10 percent of the mass. In fact […] the fitting curves form the envelopes […] cease to intersect the core curves when the mass of the envelope becomes smaller than 9 times the mass of the core.[112]

But when Gamow and Keller considered a partially degenerate core, the curve did extend towards small radii and a fitting to the envelope proved to be possible.[113] Here is another way to put it. Let us consider the given value $M^* = 0.1\, M_\odot$ and let us assume the core is in perfect gas state. If curves corresponding to envelopes of different stellar (core + envelope) masses—e.g. 0.4 and 1.6 $M_\odot$—are drawn in the logr*–logρ^* plane, it can be said that the core's curve fits the second one, but not the first one. In fact, in the latter case the core's mass exceeds the Schönberg–Chandrasekhar limit. However, if a partial degeneration of the core is taken into account, it turns out that the fitting becomes possible without limitations (Fig. 5).

More generally, Gamow and Keller depicted curves corresponding to different core and envelope masses in the logr*–logρ^* plane. For a given stellar mass (core + envelope) they investigated the values of logr* and log ρ^* at which the curves met for different relative values of the core's and envelope's mass (e.g. for a core containing 12.5, 25 and 50% of the whole stellar mass). This allowed them to determine how r* and ρ^* changed as the core mass increased, i.e. with advancing evolution.

As a last step in drawing the evolutionary path on the H–R diagram, the values of r^* and ρ^* had to be linked to L and T_{eff}, or L and R. The authors did this in an approximate manner, by exploiting homology relations that, strictly speaking, did not

111 Gamow and Keller (1945, p. 126).

112 Gamow and Keller (1945, p. 128).

113 Gamow and Keller exploited actual physical formulae based upon the Fermi-Dirac functions as well as the numerical integrations already performed by Gordon W. Wares (1944).

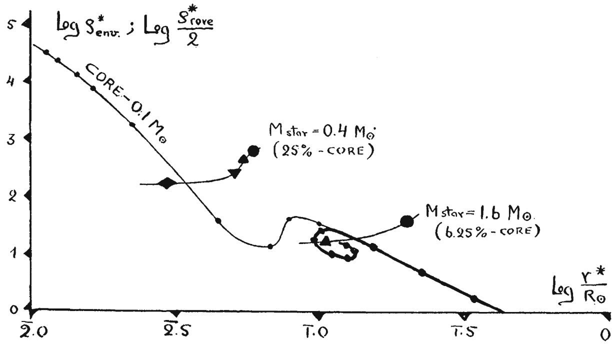

Fig. 5 Core–envelope fitting in the logr^*–log ρ^* plane. The *thick line* corresponds to a perfect-gas core and it does not intersect the 0.4 $M_\odot$ line because of the Schönberg–Chandrasekhar limit. If degeneracy is brought into the equations, it unwinds the core line (*thin line*) and the intersection is possible (from Gamow and Keller 1945, p. 127). Reprinted with kind permission of the American Physical Society

really hold. It is interesting to notice that they justified this by claiming that integrating the equilibrium equations into every single case would require a huge amount of work.

This is a typical feature of many coeval works. The integration problem was but a mere calculus issue, and scholars tried to keep it as simple as possible. All in all, scientist focused on physical issues and did not much care about integration processes. They relied on existing calculations or, if they did new ones, they basically considered that a boring—albeit necessary—task. They tried to keep equations as simple as possible to avoid time-consuming manual calculations. A much sharper focus was placed on the issue of laying down physical giant stars models. Things would change later on, when basic giant models became available and the life-changing development of computers would remove most calculus-related barriers.

8.2 Evolution on the H–R diagram

Gamow and Keller drew the evolutionary path for stars of different masses on a R–L diagram (that is equivalent to an H–R diagram as long as stars are considered to be black bodies) (see Fig. 6). The two scientists figured out that, in the case of a stellar mass $M = 0, 1 M_\odot$, fitting solutions did exist even for a core mass = 0.5 M or more, and that as the core mass increased its radius became steadily larger, too. This implied that the hydrogen-burning shell would actually approach the surface during later stages of evolution. This caused an increase—rather than a decrease—in surface temperature. If M is assumed to be $0, 4\, M_\odot$, then fitting solutions exist only for a core mass of about 1/3 and as the core mass increases, its radius first becomes larger, then begins to shrink. Again, this seemed to be a promising way to create giants, since both stellar radius and luminosity increased by a factor 40 and 10 ca., respectively (see Fig. 6). In the 4 $M_\odot$ case calculations were performed, too, although they were reckoned to be rather uncertain.

Finally, Gamow and Keller remarked that, based on their explanation, red giants are evolved stars and the previous idea of young, contracting stars could be entirely

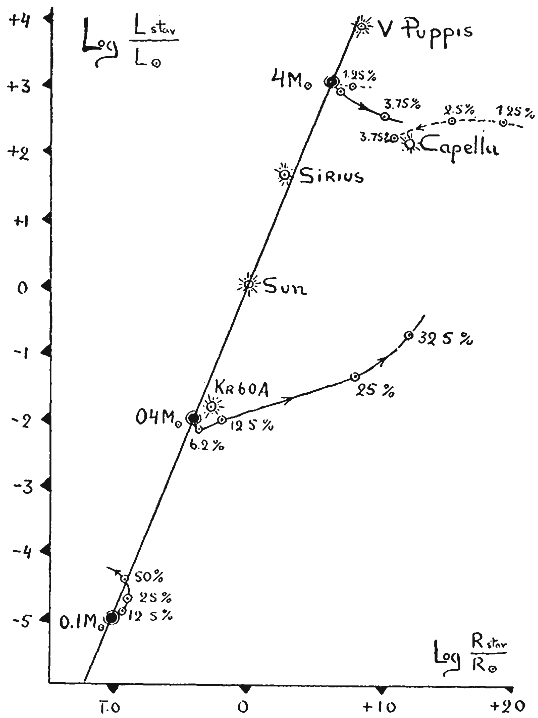

Fig. 6 Evolution of stars of different mass on a *R*–*L* diagram. The percentage of the core's mass is indicated along the evolutionary paths (from Gamow and Keller 1945, p. 135). Reprinted with kind permission of the American Physical Society

dismissed due to the very short timescale it implies. Moreover, they discarded the former idea by Teller and Gamow itself, that light element (Li, Be, B) fusion was a possible energy source for giants, as it yielded different bands that ran parallel to the main sequence in the *H*–*R* diagram.[114]

Although Gamow's and Keller's paper undoubtedly established some important facts, Harrison found an error by a factor of eight in the degenerate gas pressure formula soon afterwards.[115] More generally, she described a model similar to those of Gamow and Keller, but she failed to enlarge the radius enough to produce actual giants. One noteworthy point she made was that the Gamow– Keller star enlarged its nucleus so much as to exceed Chandrasekhar's mass for a degenerate autogravitating structure, and this led Gamow and Keller to wrong conclusions. White dwarfs—she argued—have no links to giant stars. After Harrison's contribution, the core of evolving stars seemed to be constrained between Chandrasekhar's mass and the Schönberg–Chandrasekhar limit, so that its mass could not grow too much either in the perfect or in the degenerate gas case.

To recap, up to that time scholars had often come up with crucial insights into the problem, but no accurate, reliable giant star models were yet available. However, times were ripe for that at least.

114 Gamow and Teller (1939).

115 Hall Harrison (1946, p. 195), see also Arny (1990, p. 223).

9 The first reliable giant star models

The scenario finally changed in the years around 1950. A lot of ingredients to explain giant stars were available, and that growing knowledge eventually came together to form a satisfactory picture— "satisfactory" in the sense that scientists succeeded in laying a sound physical basis that they used to build evolutionary models in which stars did evolve towards the red giant stage after the main sequence age, and the calculated features of the giants actually matched the observational features of these stars. These models cannot be considered conclusive in the light of today's knowledge (further progress would be made by scholars in the following decades). However, they constituted the valuable basis subsequent models would be built upon. Papers from this period are still mentioned today as key steps towards the understanding of giant stars.

Obviously enough, those results were made possible by the significant contributions of the previous decades. As discussed above, the great names of twentieth century astrophysics (namely Strömgren, Gamow, Chandrasekhar) played a major role. Unsurprisingly, the theory of red giants finally came to solid ground when other great names, notably Hoyle, Sandage and Schwarzschild, focused on it.

9.1 Inhomogeneous models in the years around 1950

We already saw Hoyle hard at work, trying to introduce inhomogeneity into stellar models together with Lyttleton at the beginning of the fourties. The two scholars gave another noteworthy contribution in 1949.[116]

They developed a stellar model by identifying a hydrogen-poor inner region and a hydrogen-rich outer one, as usual. However, they made a step ahead in that the hydrogen-poor region was further divided into a convective central zone that supplied the whole star with energy via the CNO cycle, and a surrounding radiative one (see Fig. 7).

Hoyle's and Littleton's model consisted of an H-poor core, an H-poor intermediate zone and an H-rich envelope. Degeneracy was not considered and radiation pressure was supposed to be negligible, as "the parameter β is always near unity for the models of chief interest."[117] Discontinuity in chemical composition was described by a sudden change:

$$\begin{aligned} &(X_{\mathrm{i}}, Y_{\mathrm{i}}, Z_{\mathrm{i}}) \rightarrow (X_{\mathrm{e}}, Y_{\mathrm{e}}, Z_{\mathrm{e}}) \\ &\mu_i \rightarrow \mu_e = 0,5\mu_i \end{aligned} \tag{32}$$

where "i" stands for the inner zone (core + intermediate zone) and "e" stands for the envelope. As for the absorption coefficient, Hoyle and Lyttleton assumed electron scattering to be the dominant source of opacity in the inner region. In the envelope they considered two possible cases, namely electron scattering and a photoelectric model.

116 Hoyle and Lyttleton (1949).

117 Hoyle and Lyttleton (1949, p. 616). Hoyle and Lyttleton restricted their investigation to stars of a few solar masses.

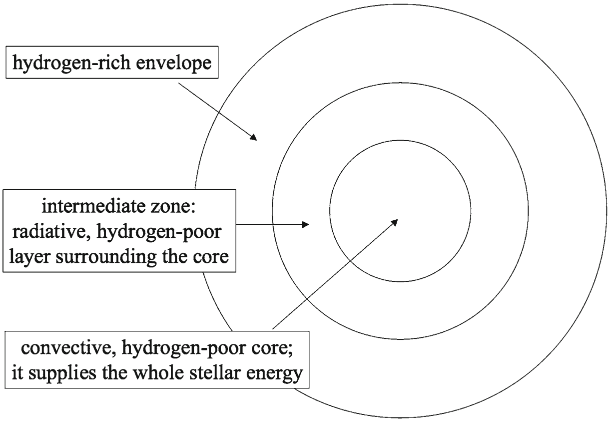

Fig. 7 Structure of an evolved star according to Hoyle and Lyttleton

In the latter case, opacity was supposed to change discontinuously, just like chemical composition.

Hoyle and Lyttleton integrated stellar structure equations into those basic assumptions. For both possible absorption coefficients, they found that an increase of the μ value in the inner region boosted a large growth of both radius and brightness. In the H–R diagram (see Fig. 8), homogeneous stars having a ubiquitously uniform value of $\mu = 1/2$ make up a sequence that lies close to the left of the main sequence. In the photoelectric model, as inhomogeneities settle in and μ rises to 2/3 in the inner region, stars move to a sequence that lies close to the right of the main sequence, only to move again to the far right side as μ reaches 4/3. Meanwhile, brightness increases significantly. A similar overall behaviour is still valid in the electron scattering model, despite somehow different values.

These results were thus commented by the authors:

> It is a satisfactory result of the theory that the main sequence is well-contained between the points corresponding to $\mu = 1/2$ and those for $\mu = 2/3$. As the time required for μ to change, owing to the conversion of hydrogen to helium by nuclear reactions, from 1/2 to 2/3 is considerably greater than the time for an increase from 2/3 to 4/3, it follows that the position of a star in the diagram must be in the main sequence throughout most of its lifetime.[118]

The values found are in fair agreement with observational values. When plotting them, a distinction was made between population-I and -II giant stars.[119]

Hoyle and Lyttleton investigated physical processes that could cause inner chemical inhomogeneities. Clearly, they advocated nuclear processes as a major reason, provided no strong mixing occurred in the star's interior. If that was not the case and

118 Hoyle and Lyttleton (1949, p. 628).

119 As is well-known, in 1944 Baade introduced the idea of the existence of two populations of stars after observing the Andromeda Galaxy and its two elliptical companions M32 and NGC 205 (see Baade (1994)). This is a collateral topic to the main argument of this paper and I shall not elaborate upon it.

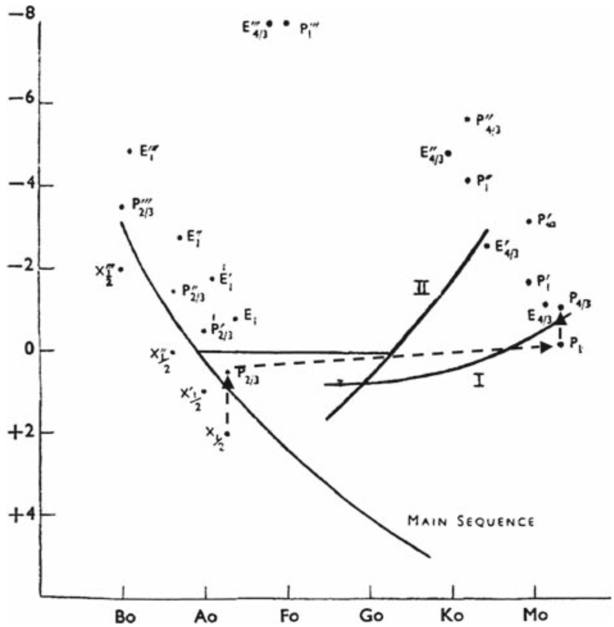

Fig. 8 Hoyle and Lyttleton's results depicted on the *H–R* diagram. The *full lines* represent the main sequence and the giant branches for population-I and -II stars, as observed by Baade in (1994, p. 143). The points $X_{1/2}$, $X'_{1/2}$, $X''_{1/2}$ and $X'''_{1/2}$ stand for stars of different masses with $\mu = 1/2$ everywhere. The masses considered range between 2 and 5 $M_{\odot}$. The unevolved stars lie close to the main sequence, lower and to the left. The points *P* stand for the evolved stars in the photoelectric model, while the points *E* stand for the evolved stars in the electron scattering model. *P* and *E* have the same mass as *X*, P' and E' as X', P'' and E'' as X'', P''' and E''' as X'''. The subscripts of the *P* and *E* points are the μ values in the inner region. Hence, a star of a given mass evolves across the points with growing μ values. The *dashed line* in the figure is an example of evolution (it is by the author of this paper and is not found in the original figure) (from Hoyle and Lyttleton 1949, p. 627). Reprinted with kind permission of Wiley-Blackwell

mixing did occur, then an accretion of hydrogen at the surface was a possible explanation. This is the same idea they advocated in their 1942 paper. However, the 1949 paper goes well beyond that, in that the role of inner changes due to nuclear processes is much more emphasized, and the radii of the calculated giant models are much larger than those previously obtained in 1942.

A paper by Li Hen and Schwarzschild was published in the very same issue of the *Monthly Notices of the Royal Astronomical Society*.[120] It was the first in a series of papers written by Schwarzschild together with several colleagues— notably Sandage and Hoyle himself—that greatly contributed to a satisfactory explanation of the whole giants issue.

Hen and Schwarzschild summarized previous achievements by pointing out how inhomogeneity was a key factor:

120 Hen and Schwarzschild (1949).

> When these [= inhomogeneous] models were applied to stars with a given mass and central temperature, the resulting radii were generally found to be larger than those obtained for chemically homogeneous stars. Thus it appeared possible that these models should represent giants of intermediate spectral type. In all the cases which had been computed through in detail, however, the increase in radius was less than a factor four and hence insufficient to represent late-type giants.[121]

In their paper, they succeeded in elaborating models with "stellar radii sufficiently large for moderate red giants".[122]

Similarly to Hoyle and Lyttleton, they adopted a tripartite model like the one depicted in Fig. 7: a convective, hydrogen-poor core; a radiative, hydrogen-poor intermediate region; and an radiative, hydrogen-rich enclosing envelope. Discontinuity in chemical composition is expressed by a jump $(X_\mathrm{i}, Y_\mathrm{i}, Z_\mathrm{i}) \rightarrow (X_\mathrm{e}, Y_\mathrm{e}, Z_\mathrm{e})$. Again, radiation pressure and degeneracy were neglected. The convective core was supposed to be a polytrope of index $n = 3/2$. The discontinuity in chemical composition was described as a sudden change $\mu_\mathrm{i} \rightarrow \mu_\mathrm{e}$.

As for the absorption coefficient, it was assumed to be:

$$\begin{aligned} \kappa &= \kappa_0 \rho^{0.75} T^{-3.5} \\ \kappa_0 &= 10^{25} Z (1 + X)^{0.75} \end{aligned} \tag{33}$$

The variables of differential stellar structure equations were replaced by the dimensionless quantities p, t, q, x, l, j:

$$\begin{aligned} P(r) &= p(r) \frac{GM^2}{4\pi R^4} \\ T(r) &= t(r) \mu_\mathrm{e} \frac{\mathrm{H}}{K} \frac{GM}{R} \\ M(r) &= q(r) M \\ r &= x(r) R \\ l(r) &= \frac{\mu(r)}{\mu_\mathrm{e}} \\ j(r) &= \frac{Z(r)}{Z_\mathrm{e}} \left\{ \frac{[1 + X(r)]\, \mu(r)}{[1 + X_\mathrm{e}]\, \mu_\mathrm{e}} \right\}^{0.75} \end{aligned} \tag{34}$$

[121] Hen and Schwarzschild (1949, p. 631); interestingly enough, the authors added a reference to Öpik in a note, pointing out how his models, although capable of accounting for larger giants, could actually be neglected, as they assumed too large a difference in chemical composition between the core and the envelope. In other words, Hen and Schwarzschild realized the pioneering nature of Öpik's work, but they also pointed out that so significant a result was premature for his time. Moreover, Hen and Schwarzschild did not appear to be aware of the aforementioned results by Hoyle and Lyttleton, whose paper is not mentioned and does not appear in the reference list. This is hardly surprising as that paper had been submitted only a few months earlier.

[122] Hen and Schwarzschild (1949, p. 631).

In Eqs. 34 it has been stressed that μ, Z, X, and consequently l and j, depend on r, although their dependence is rather simple as they only assume two different values in the inner and the outer part of the star. In particular, in the envelope $l_\mathrm{e} = j_\mathrm{e} = 1$. As usual, the stellar mass M and radius R are to be considered constant for integration purposes, but from a physical viewpoint they are rather free parameters describing different stellar configurations. The standard equations for the radiative part (hydrostatic equilibrium, mass continuity, radiative transfer) take the following form:

$$\begin{aligned} \frac{\mathrm{d}p}{\mathrm{d}x} &= -l\frac{pq}{tx^2} \\ \frac{\mathrm{d}q}{\mathrm{d}x} &= l\frac{px^2}{t} \\ \frac{\mathrm{d}t}{\mathrm{d}x} &= -ljC\frac{p^{1.75}}{t^{8.25}x^2} \end{aligned} \tag{35}$$

where parameter C is constant throughout the star and is expressed as follows:

$$C = \frac{3\kappa_{0e}}{4ac}\left(\frac{\mathrm{K}}{\mu_\mathrm{e}\mathrm{HG}}\right)^{7.5}\left(\frac{1}{4\pi}\right)^{2.75}\frac{LR^{1.25}}{M^{5.75}} \tag{36}$$

The standard quantities U, V are written as follows:

$$\begin{aligned} U &= \frac{\mathrm{dlog}M(r)}{\mathrm{dlog}r} = \frac{r}{M(r)}\frac{\mathrm{d}M(r)}{\mathrm{d}r} = l\frac{px^3}{tq} \\ V &= -\frac{\mathrm{dlog}P(r)}{\mathrm{dlog}r} = -\frac{r}{P(r)}\frac{\mathrm{d}P(r)}{\mathrm{d}r} = l\frac{q}{tx} \end{aligned} \tag{37}$$

It should be noticed that, due to the supposed inhomogeneity, U and V go through a discontinuity at the interface between the intermediate zone and the envelope. As a matter of fact, all quantities in their expressions are continuous except l. On the other hand, no discontinuity occurs at the interface between the core and the intermediate zone.

Moreover, the polytropic index is $n = 1.5$ in the convective core, whereas in the intermediate zone and envelope the effective index was written by Hen and Schwarzschild as being:

$$n + 1 = \frac{T(r)}{P(r)}\frac{\mathrm{d}P(r)}{\mathrm{d}T(r)} = \frac{1}{j}\frac{qt^{8.25}}{Cp^{1.75}} \tag{38}$$

The introduction of the "intermediate zone", first by Hoyle and Lyttleton and later by Hen and Schwarzschild, was a real novelty. Then, it is not surprising that they could rely on former integrations for the other two regions[123] but had to integrate the

[123] Solutions of Eqs. 35 with the condition $l_\mathrm{e} = j_\mathrm{e} = 1$ were available for the radiative envelope, as Schwarzschild himself together with Richardson had already numerically integrated them for different values of C (Richardson and Schwarzschild 1948).

equations in that zone themselves. The authors had the integration start at an arbitrary point ξ_1, that is the point where the core ends and the intermediate zone begins.[124] So they found a single-parameter family of solutions.

Then those solutions were fitted to the envelope. If we call the values of U "just inside" and "just outside" the interface U_i and U_e, respectively, and do the same for V_i, V_e and n_i, n_e, fitting at the intermediate zone/envelope interface requires that:

$$\left.\begin{array}{l} \frac{U_\mathrm{i}}{U_\mathrm{e}} = \frac{l_\mathrm{i}}{l_\mathrm{e}} = l_\mathrm{i} \\ \frac{V_\mathrm{i}}{V_\mathrm{e}} = \frac{l_\mathrm{i}}{l_\mathrm{e}} = l_\mathrm{i} \end{array}\right\} \Rightarrow \frac{U_\mathrm{i}}{V_\mathrm{i}} = \frac{U_\mathrm{e}}{V_\mathrm{e}}$$
$$\frac{n_\mathrm{e}+1}{n_\mathrm{i}+1} = \frac{j_\mathrm{i}}{j_\mathrm{e}} = j_\mathrm{i} \tag{39}$$

Then, as a consequence of the discontinuity of U and V at the interface, it follows that the point representing stellar status at the interface must "jump" between $(U_\mathrm{i}, V_\mathrm{i})$ and $(U_\mathrm{e}, V_\mathrm{e})$ along a straight line passing through the origin.

Figure 9 shows the results of the fitting process.

Based on Fig. 9, a stellar model can be constructed as follows. Start from the centre ($U = 3$, $V = 0$), then follow the curve corresponding to the core. At an arbitrary point continuously switch to the intermediate zone curve, and follow it all the way to another arbitrary point, where you will need to jump to the corresponding point of an envelope curve where C = arbitrary. Finally follow this curve to the surface. Therefore three arbitrary choices must be pursued. From a physical viewpoint, this means that the amount of change in composition, the point at which such change occurs and the amount of change in the absorption coefficient can be set at will. Once a model is chosen, the dimensionless quantities defined in (34) and consequently the state parameters can be computed for any value of r. If the stellar mass M, central temperature T_0 and chemical composition of the core and envelope are fixed, then a clearly defined model can be calculated. Hen and Schwarzschild chose $M = 10M_\odot$ and set $T_0 = 3 \times 10^7\,\mathrm{K}$ in order for the CNO cycle to be sustainable. They then calculated the stellar models and especially the resulting radii for 15 different values of $(X_\mathrm{i}, Y_\mathrm{i}, Z_\mathrm{i})$ and $(X_\mathrm{e}, Y_\mathrm{e}, Z_\mathrm{e})$, and consequently of μ_i and μ_e. The resulting radii turned out to be approximately between 10 and 70 solar radii, which led the authors to conclude that "for the majority of the 15 cases, the computed radii are smaller than the radii of heavy red giants", although in the best-case scenarios "the radii approach the right size".[125] If we take a close look at Hen's and Schwarzschild's results[126] we can see that the largest radii (43.2 and $66.0R_\odot$) correspond to low hydrogen and high helium abundance in the innermost part, and vice versa in the outer.

124 ξ is the adimensional variable tied to r that is usually exploited when integrating polytropes (it appears as an independent variable in the Lane–Emden equation). Its use is justified by the fact that the core is a $n = 3/2$ polytrope.

125 Hen and Schwarzschild (1949, p. 643).

126 Hen and Schwarzschild (1949, pp. 644–645).

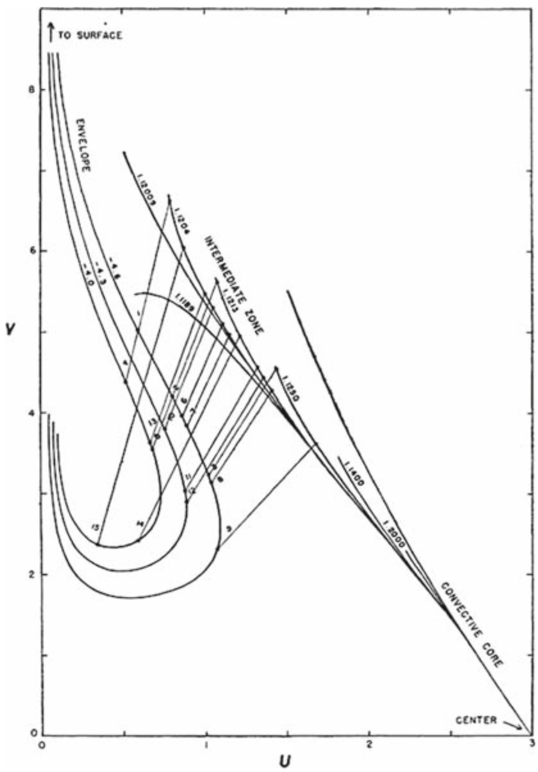

Fig. 9 Fitting in the U–V plane. Different values of log C are shown for the envelope. Similarly, several values of ξ_1 are shown for the intermediate zone. Only one curve represents the core, namely the one corresponding to a polytrope of index 1.5. The *straight lines* indicate some possible "jumps" connecting the intermediate zone to the envelope (from Hen and Schwarzschild 1949, p. 639). Reprinted with kind permission of Wiley-Blackwell

Table 1 Physical structure of giant stars of different masses (from Hen and Schwarzschild 1949, p. 645)

Total mass	$10M_\odot$	$5M_\odot$	$3M_\odot$
Mass of convective core	$0.3M_\odot$	$0.2M_\odot$	$0.1M_\odot$
Mass inside chemical discontinuity	$2.0M_\odot$	$1.0M_\odot$	$0.6M_\odot$
Total radius	$66R_\odot$	$33R_\odot$	$20R_\odot$
Radius of convective core	$0.3R_\odot$	$0.2R_\odot$	$0.1R_\odot$
Radius of chemical discontinuity	$1.0R_\odot$	$0.5R_\odot$	$0.3R_\odot$
Bolometric magnitude	−5.1	−1.8	+0.8
Effective temperature	6500	4300	3100
Spectral type	F5	gK0	gM2

Reprinted with kind permission of Wiley-Blackwell

The two scientists further tried to vary the stellar mass for a fixed composition consisting of 0% hydrogen–90% helium inside and 90% hydrogen–0% helium outside. Their results are summarized in Table 1.

It is obvious that all three models in Table 1 have a rather small and dense core and an extended, thin envelope. Nevertheless, the resulting radii are still not large enough. In conclusion, the method developed by Hen and Schwarzschild was promising in that it showed it was indeed possible to regard red giants as evolved stars. Inhomogeneity with a strong hydrogen difference was proved to be a key factor, and radius enlargement was proved to follow hydrogen depletion in the core. However, the method had to be perfected. Schwarzschild and his associates engaged in this task in the following years. They dealt with changes in central temperature, degree of inhomogeneity and absorption coefficient, and also tried to account for possible degeneracy effects and a gravitational collapse of the core.

In 1952, the absorption coefficient issue got perfected in a paper by John B. Oke and Schwarzschild.[127] The two scholars constructed models much in the same way as Hen and Schwarzschild had done, but they introduced a more precise description of the absorption coefficient. They investigated two different cases for the stellar interior: a dependence $\kappa \propto \rho^{1-\alpha}$ with $\alpha = 0.25$ and $\alpha = 0$. As for the envelope, they only considered $\alpha = 0$. The $\alpha = 0.25$ case had already been developed by Hen and Schwarzschild, while the $\alpha = 0$ case went back to the Kramers' Law and still needed to be integrated. Thus, in the $\alpha = 0$ case Oke and Schwarzschild had to focus on integration both for the envelope—where further values of C were considered—and for the core. The authors finally obtained two series of solutions corresponding to the two possible values of α that they depicted in the U–V plane (Fig. 10, left).

Oke and Schwarzschild found that, following changes in ξ_1, the mass fraction lying in the core varied within a large range of values (from 14% to 70% approximately). On the other hand, the core radius hardly exceeded a few percents, and this implied that the constructed models were much denser in the centre, just like the one shown in Table 1.

By choosing stellar mass, chemical composition and central temperature, it is possible to derive the value of R from the second of Eq. 34 if applied to the centre of the star, and the value of L from (36).[128] Consequently, any stellar model can be depicted in the H–R diagram (Fig. 10, right).

In the diagram, possible variations of ξ_1 and M are both accounted for. The heavy lines represent models of equal mass, while the dashed ones represent models of equal ξ_1 values (for the $\alpha = 0.25$ case). More specifically, these values decrease monotonically as the roman numerals increase, and such a decrease is faster at first and slows down afterwards. As the mass value of the hydrogen-poor region depends on ξ_1, and increases as ξ_1 decreases, following the heavy lines from left to right means following the evolution of a star as its core burns out hydrogen and shrinks, quickly at first and more slowly afterwards. After making comparisons with the position of actual giants, the authors concluded that "the models here considered amply cover the territory in the Hertzsprung-Russell diagram occupied by the red giants. Indeed, the most extreme models [...] give effective temperatures probably lower than that

[127] Oke and Schwarzschild (1952).

[128] It should be pointed out that the equations exploited by Oke and Schwarzschild in (1952) differ slightly from (34) and (36), which Hen and Schwarzschild used in 1949. This is due to the differences assumed for the absorption coefficient. However, from a conceptual viewpoint they can be regarded as equivalent. More generally, the same consideration also applies to the subsequent papers by Schwarzschild and associates that are discussed later in this section. The method used is always based upon equations similar to (34), (35), (36) and (37). However such equations take different forms according to the underlying physical models.

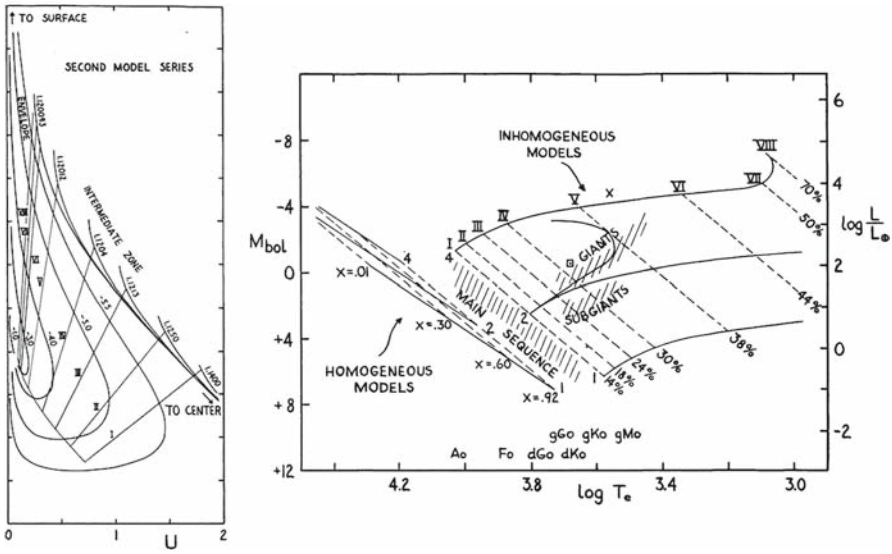

Fig. 10 *Left*—fitting in the U–V plane, similar to the one shown in Fig. 9, for the $\alpha = 0.25$ case (a similar figure is supplied in the original paper for the $\alpha = 0$ case). As for the envelope, different values for log C are shown. Similarly, several values of ξ_1 are shown for the intermediate zone. The curve representing the core is outside the right bottom corner. The *straight lines* indicate some possible "jumps" connecting the intermediate zone to the envelope (from Oke and Schwarzschild 1952, p. 320). *Right*—corresponding evolution in the H–R diagram. The *full lines* stand for fixed stellar mass values ($1\,M_\odot$, $2M_\odot$ and $4M_\odot$ from bottom to top, respectively) and a variable percentage of the total mass in the hydrogen-poor zone. The *shaded areas* indicate the actual position of main sequence, subgiant and giant stars in the diagram (from Oke and Schwarzschild 1952, p. 327). Reprinted with kind permission of the American Astronomical Society

of any observed star".[129] This final remark is hardly surprising, as convection in the envelope was not taken into consideration at that time.

The line corresponding to $2M_\odot$, which turns back once temperature drops to approximately 3,500 K, represents the evolution of a star in the $\alpha = 0$ case. This highlights the great influence the inner absorption coefficients has on the structure of an evolved star. Furthermore, the diagram indicates the possible evolution of homogeneous models. It appears evident how the evolution of a star where strong mixing occurs is expected to be very different from the evolution of a star where mixing is not so effective and inhomogeneities settle in.

However, Oke and Schwarzschild felt that their models were still work in progress. They stated so at the end of their paper:

> Stellar models like the present ones must still be considered as exploratory, owing to the arbitrariness with which the character of the chemical inhomogeneity is assumed. If, indeed, a chemical inhomogeneity is an essential feature in the internal structure of the red giants, as now appears likely, it will be possible to

[129] Oke and Schwarzschild (1952, p. 326).

> derive definite models for red giants only after the physical mechanisms which govern the degree and character of the internal mixing have been determined.[130]

9.2 Sandage and Schwarzschild: inhomogeneous stars with a collapsing core

A work by Sandage and Schwarzschild[131] appeared in the same *Astrophysical Journal* issue that contained the paper by Oke and Schwarzschild. Sandage and Schwarzschild's work turned out to be a milestone on the way to solving the giants issue. It was the first paper ever to lay down a sound model in which an inhomogeneous structure was investigated after reaching the Schönberg-Chandrasekhar limit and consequently experiencing a core collapse.[132]

Sandage's interest in this kind of theoretical speculations was fostered by his own work under Baade on the observational features of globular clusters. In particular, he studied the H–R diagram for the two globular clusters M3 and M92[133] and observed that the giant branch was tied to the mid main sequence through a continuous distribution of stars. This pointed towards the need to explain the evolutionary stages of a star that moves from the main sequence to the giant branch.

The paper by Oke and Schwarzschild had described a point-source, convective core, under the assumption that a certain reservoir of hydrogen, big enough as to sustain nuclear reactions, was still present. Sandage and Schwarzschild described the core as exhausted, isothermal, in a radiative quasi-equilibrium state. In other words, the core was considered to be exhausted and to have reached the Schönberg–Chandrasekhar limit (hence the gravitational collapse), with an extremely (infinitely) thin shell at the temperature of 3×10^7 K where the CNO cycle occurs, and a radiative envelope:

> This paper considers a model whose inhomogeneity arises from the following evolutionary process. An initially homogeneous star with a convective core and radiative envelope (Cowling model) which experiences no mixing between the core and envelope starts exhausting its hydrogen supply in the core. The subsequent early stages of the evolution follow those computed by Schönberg and Chandrasekhar[134] (reference is given to Schönberg and Chandrasekhar 1942), with the core finally exhausted of hydrogen and therefore isothermal. The nuclear-energy production is then confined to a shell between the exhausted core and the radiative envelope. The assumption of no mixing creates a chemical discontinuity between the core and the shell. When the shell has burned outward until it reaches the Schönberg – Chandrasekhar limit for an isothermal core, a new evolutionary process must take place, which is most likely a gravita-

130 Oke and Schwarzschild (1952, p. 330).

131 Sandage and Schwarzschild (1952).

132 The importance of the 1952 paper by Sandage and Schwarzschild is emphasized in literature. See Arny (1990, p. 224); Longair (2007, p. 188); Tassoul and Tassoul (2004, p. 150). Moreover, Gingerich and Lang, in their selected bibliography of astronomy in the years 1900–1975, chose this paper as representative of the solution of the giants problem (1979, pp. 353–363).

133 Arp et al. (1952). A further work on the same subject is on 1953.

134 In the original paper reference is here given to Schönberg and Chandrasekhar (1942).

> tional contraction of the core. This paper is concerned with the quasi-equilibrium states through which an unmixed model passes after reaching the Schönberg – Chandrasekhar limit.[135]

Sandage and Schwarzschild assumed that the energy ε supplied by gravitational contraction was equally distributed throughout the core (an admittedly unrealistic assumption) and that degeneracy and radiation pressure were negligible. They further made a reasonable assumption we will not discuss about the shift from Kramer's to the electron-scattering regime for the absorption coefficient. Finally, they relied on the hypothesis that the inhomogeneities could be described by a sudden jump from the core to the envelope, such as:

$$\begin{aligned} (X_\mathrm{i}, Y_\mathrm{i}, Z_\mathrm{i}) = (0, 0.98, 0.02) \rightarrow (X_\mathrm{e}, Y_\mathrm{e}, Z_\mathrm{e}) = (0.596, 0.384, 0.02) \\ \mu_\mathrm{i} \rightarrow \mu_\mathrm{e} = 0.5\mu_\mathrm{i} \end{aligned} \tag{40}$$

The two scientists followed the U–V fitting method and plotted a graph similar to the one in Fig. 10—left. In particular, the luminosity L_g due to gravitational contraction was estimated and included in the integration process. This term obviously depends on time. Consequently, the evolution of a star of a given mass had a further time-dependence, beyond the one implicitly contained in the variation of the μ value.

Once the fitting process was done, the physical features of three model stars of $1M_\odot$, $2M_\odot$ and $4M_\odot$ could be determined in a way similar to the one used in the paper by Oke and Schwarzschild (via the second of Eqs. 34 and Eq. 36). As a result, the core contracts and the envelope expands dramatically:

> The star must pump energy into the expansion of the outer layers so as to overcome the gravitational potential. This energy is supplied by the outgoing flux produced by the nuclear and gravitational sources closer in.[136]

The authors also observed that central temperature rose up from 30 to 170 million degrees, and they correctly pointed out that this was due to the fact that only a portion of the gravitational energy was radiated away—a noteworthy feature, since in the point-source model Oke and Schwarzschild had previously developed, central temperature was assumed to be constantly equal to 3×10^7 K as hydrogen was supposed to be continuously burning in the star's interior. Sandage and Schwarzschild's paper gives a more reliable description of a giant star, in that it describes a further evolutionary stage in which the core begins to shrink and to heat up. The associated evolutionary tracks are depicted in the H–R diagram (see Fig. 11).

The first part of the evolutionary track shows a moderate brightness increase at an almost constant surface temperature, as Schönberg and Chandrasekhar had already pointed out. This corresponds to a transition from the Cowling configuration on the main sequence to the isothermal core, that is still close by in the diagram. When it sets in, the core collapse boosts the envelope expansion and a quick shift to the red giant zone.

[135] Sandage and Schwarzschild (1952, p. 463).

[136] Sandage and Schwarzschild (1952, p. 471).

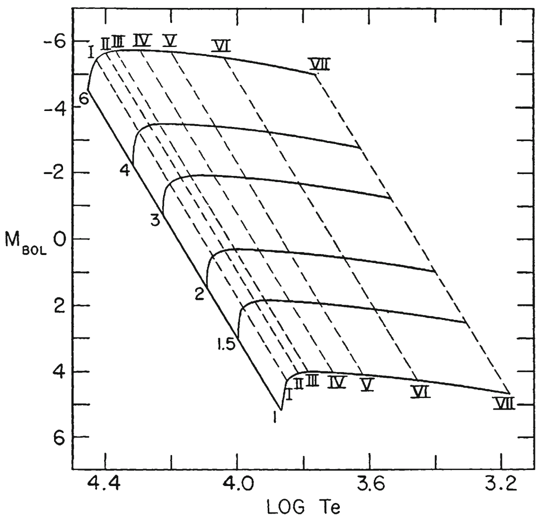

Fig. 11 Evolutionary tracks in the *H–R* diagram. The main sequence (based on Cowling's model) is shown on the *left*. The *full* and *dashed lines* have the same meaning as in Fig. 10—*right*. No comparison with observational data is present (from Sandage and Schwarzschild 1952, p. 473). Reprinted with kind permission of the American Astronomical Society

This conclusion proves to be a powerful tool for interpreting the *H–R* diagram of globular clusters. The fainter stars have not reached the Schönberg-Chandrasekhar limit yet, but the brighter ones have, moving to the right in the process. Some stars caught in their expansion stage link the main sequence to the giant branch.

The work by Sandage and Schwarzschild is a major achievement. However, it is not flawless. The evolutionary tracks move towards large radii but they fail to account for the increase in luminosity that is observed in actual giants. Sandage and Schwarzschild commented:

> A [...] difficulty arises when the extent of the envelope expansion is considered. Under the present assumptions there is no reason why the envelope expansion should stop at or before [the point labelled as] VII, while the observed H-R diagrams of globular clusters seem to indicate that the expansion should essentially stop about at [the point labelled as] V, and then mainly a brightening [...] and only little further expansion [...] should occur. One may speculate that around [the point labelled as] V a physical process not included in the present computations should start to play an essential role.[137]

The two authors considered two tentative explanations for the "physical process not included in the present computations": (1) helium ignition and burning in the central regions and (2) a certain amount of mixing near the shell due to stellar rotation, which might reduce the degree of inhomogeneity.

As for helium ignition, the central temperature Sandage and Schwarzschild came up with was too low if compared to Salpeter's coeval estimates.[138] Nevertheless, it was of the same order of magnitude, which implied that helium fusion could not

137 Sandage and Schwarzschild (1952, p. 475).

138 Salpeter (1952).

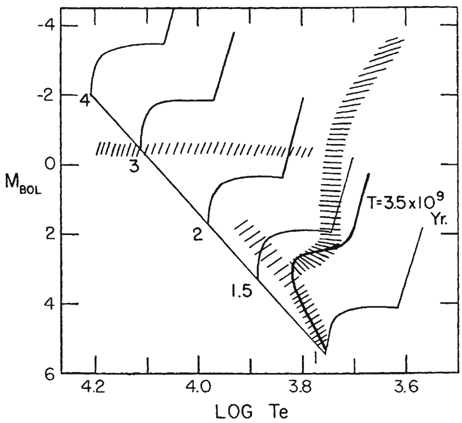

Fig. 12 Evolutionary tracks (which the authors defined "speculative") in the *H–R* diagram for stars of different masses, according to the helium-burning model. The *heavy line* is the expected distribution of a globular cluster 3.5 billion years in age (from Sandage and Schwarzschild 1952, p. 475). Reprinted with kind permission of the American Astronomical Society

be thoroughly ruled out. Assuming helium actually underwent the fusion process, preliminary estimates were made for stars constituted by "a hydrogen-rich envelope, a shell in which hydrogen burns, a helium-rich intermediate zone, a shell in which helium burns, and an inert core of heavy elements."[139] These led to an *H–R* diagram (see Fig. 12) more closely matching the diagram of globular clusters.

Noticeably, the idea that convection could settle in the envelope was not considered as a possible explanation of why the tracks in Fig. 11 move far to the right without bending upwards.

More generally, we can identify three critical assumptions in Sandage and Schwarzschild's models: (1) a lack of gas degeneracy, (2) an abrupt discontinuity in the μ value and (3) obviously enough, the absence of convection in the envelope.

All the three assumptions were further investigated by Schwarzschild and associates in the following years, and the whole issue of the giant stars was finally solved.

9.3 Gas degeneracy and continuous mixing

The degeneracy issue was addressed in 1953 by Schwarzschild, together with I. Rabinowitz and R. Härm.[140] Core degeneracy was not considered in the previous models, and Sandage and Schwarzschild had actually discussed whether such an assumption was tenable. It turned out that in fact it was, although some degeneration could be present in the innermost zones of less massive stars.

Then Schwarzschild, Rabinowitz and Härm focused on a more detailed investigation of possible, partial degeneracy in the stellar core, following in Gamow and Keller's footsteps. They considered a shell-source model with an isothermal core surrounded by an inert envelope. Composition was again thought to change abruptly at the envelope–core interface (i.e. in the infinitesimally thin shell). The basic equations remained similar to (35) and (37) for the envelope, but were changed for the core.

139 Sandage and Schwarzschild (1952, p. 475).

140 Schwarzschild et al. (1953).

As had been the case with Gamow and Keller, Wares' solutions[141] were adopted for integration purposes. The resulting stellar models happened to fit the red giant region of the H–R diagram well:

> [...] the present models cover well the area occupied by the red giants. Hence the present detailed computations appear to substantiate the earlier tentative conclusions by Gamow and Keller[142] that models with partially degenerate isothermal cores may under certain circumstances have very large radii.[143]

Once again, however, some models turned out to have huge (larger than expected) radii and very low (around 10^3 K) surface temperatures, as no convection in the envelope was considered.

This issue was further investigated in a paper by Härm and Schwarzschild,[144] who assumed a continuously changing chemical composition, with μ varying by a factor 2.5 from the core to the surface. Their models were constructed as follows: $\mu = \mu_{\mathrm{i}}$ (constant) in the core and $\mu = \mu_{\mathrm{e}} = 0.4\mu_{\mathrm{i}}$ (constant) in the envelope, μ(r) continuously varying from μ_{c} to μ_{e} in an intermediate zone. Degeneracy and (as in all previous papers) radiation pressure were neglected. Both convective-core (point-source) and isothermal-core (shell-source) models were investigated. Equations 39 were replaced to meet the requirement that U and V be continuous throughout the star. The resulting numerical integrations showed that, for stars having the same mass, central temperature and overall chemical compositions, no great differences in brightness were to be expected when moving from the discontinuous to the continuous case. As for the radii, Härm and Schwarzschild observed that red giants were well described and that, once more, "for the more extended stars, however, the discontinuous models appear to exaggerate the radius."[145]

The issue of the continuous variation of μ was also tackled by Roger J. Tayler in 1954.[146] The English astrophysicist pointed out that, as the core retreats during evolution, a region of continuously varying composition must exist. Consequently, he calculated some models for a massive star of 9.9 $M_{\odot}$—in which electron opacity dominates and radiation pressure is taken into account—with a diminishing hydrogen content in the interior and a shrinking core. Following hydrogen depletion in the core, the star's radius increases and its surface temperature decreases. Although the evolutionary track for this star is somewhat different from the ones developed by Sandage and Schwarzschild for the stage preceding the collapse of the core, overall evolution was not much different and the resulting H–R diagram for a globular cluster of a given age proved to be quite similar.

---

141 Wares (1944).

142 In the original paper reference is here given to Gamow and Keller (1945).

143 Schwarzschild et al. (1953, p. 333).

144 Härm and Schwarzschild (1955).

145 Härm and Schwarzschild (1955, p. 453).

146 Tayler (1954).

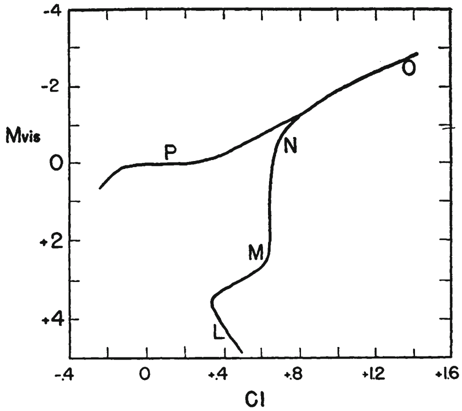

Fig. 13 *H–R* diagram (colour index versus visual magnitude) for a globular cluster (from Hoyle and Schwarzschild 1955, p. 2). Reprinted with kind permission of the American Astronomical Society

9.4 Hoyle, Schwarzschild and the first reliable giant stars models

In 1955, Hoyle and Schwarzschild published a landmark paper entitled "On the evolution of Type II stars".[147] Hoyle and Schwarzschild were awarded the International Balzan Prize for Astrophysics in 1994 for, as the formal citation stated, "their pioneering contributions to the theory of stellar evolution, upon which the modern development of the field is founded."[148] The prize was actually giving credit to their 1955 paper. It is always a little risky to identify a single work as the one that brought a long-debated scientific issue to an end. However, the 1955 paper by Hoyle and Schwarzschild definitely stands out among coeval works, both for the results it yielded and for the physical insight it showed, and can be identified as the one that finally solved the giants issue, at least for low-mass stars.

Hoyle and Schwarzschild's aim is to study stars approaching the turning-point in the *H–R* diagram of globular clusters, i.e. with masses around 1.1–1.2 $M_\odot$ and belonging to population II. Such a choice was fostered by Arp, Baum and Sandage's recent work upon globular clusters M3 and M92 mentioned above, as well as by another paper by Sandage published in 1954.[149]

Hoyle and Schwarzschild started by drawing an *H–R* diagram for a globular cluster (Fig. 13).

They observed that:

> The latter line [i.e. the line occupied by the stars in the diagram] is the locus for stars of various masses at one time, while an evolutionary track is the locus for stars of the same mass at various times. This difference, however, is in practice rather small [because] all the stars in globular clusters which are now observed between points M and P [of fig. 13] have started their evolution within a short

[147] Hoyle and Schwarzschild (1955).

[148] See http://www.balzan.it/en\Cprizewinners\FredHoyleeMartinSchwarzschild.aspx (as of October 2009).

[149] Sandage (1954).

section of the main sequence, have therefore nearly the same masses, and are evolving along a tight family of tracks which closely hug the observed sequence in the colour-magnitude diagram.[150]

Hence, it was a matter explaining the different sections of the path for a star of 1.1–1.2 $M_{\odot}$.

First, the authors investigated the path from L to M. They considered energy supply as mainly due to the CNO cycle.[151] As the core gets exhausted, it must become isothermal due to the electron conduction that occurs as degeneracy settles in. The whole issue then boils down to investigating what happens to an isothermal, partially degenerate core surrounded by a layer of negligible thickness where hydrogen transmutation takes place.[152] Hoyle and Schwarzschild's also assumed no relativistic degeneracy, a negligible radiation pressure, an abrupt change in chemical composition between the core and the envelope. This means that initially $\mu_c = 0.533$ (mainly hydrogen) anywhere in the star. Then a discontinuity is established: μ_c changes to the value 4/3 (pure helium), while μ_e remains equal to 0.533.

Furthermore, the envelope was supposed to be in radiative equilibrium, but "this assumption applies only to the evolution from L to M [and it will] be dropped when we come to consider the later stages of the evolution."[153] One of the paper's key points was clear right from the outset, namely the introduction of convection in the envelope in later stages of evolution. Finally, the envelope was split into two parts as a consequence of two opacity regimes: free–free in the outer part (obeying Kramers' formula (10)) and electron scattering in the inner one—opacity always being continuous, but with an abrupt change from one formula to the other.

The method of the homology invariants U and V had become a standard technique by then, and Hoyle and Schwarzschild introduced them as set out in Eq. 14. They also used as a further homology invariant the effective polytropic index:

$$n + 1 = \frac{\mathrm{dlog}P(r)}{\mathrm{dlog}T(r)} = \frac{T(r)}{P(r)}\frac{\mathrm{d}P(r)}{\mathrm{d}T(r)} = \frac{16\pi acGT(r)^4M(r)}{3\kappa(r)P(r)L(r)} \rightarrow \text{(envelope)} \qquad (41)$$
$$n = \infty \rightarrow \text{(core)}$$

(the expression of $n + 1$ in (41) follows directly from Eqs. 1 and 13; Eq. 13 holds as long as the envelope is supposed to be in radiative equilibrium).

They estimated an abundance of C and N around 0.0005, that seemed appropriate for type-II stars.

The results of the integration process are illustrated in Fig. 14, in which a star is placed in the H–R diagram as a function of the increasing core mass.

[150] Hoyle and Schwarzschild (1955, p. 2).

[151] Although the authors reckoned that such an assumption "is not correct for the early part of the evolution near L", they nevertheless adopted it, as "the present work has been undertaken by way of a preliminary reconnaissance of the problem" (Hoyle and Schwarzschild 1955, p. 3 and 1).

[152] The authors explicitly referred to Schwarzschild et al. (1953), claiming they would essentially follow the same procedure.

[153] Hoyle and Schwarzschild (1955, p. 4).

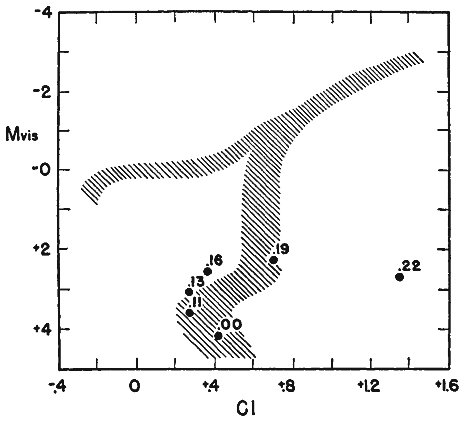

Fig. 14 A comparison between the observational *H–R* diagram for a globular cluster (*shaded area*) and the calculated models with exhausted, partially degenerate, isothermal cores and radiative envelopes (*filled circles*). The numbers indicate the mass fraction of the core, which grows over time (from Hoyle and Schwarzschild 1955, p. 10). Reprinted with kind permission of the American Astronomical Society

The tendency of the calculated points to lie on the left side of the shaded area can be easily explained in the light of the approximations made. However, this is not true for the 0.22 point, and even less so for a 0.25 point that lies so far to the right as to fall outside the diagram. This evolution to the right at an almost constant luminosity is exactly the same result Hoyle and Lyttleton got in 1949 (see Fig. 8), with Sandage and Schwarzschild following suit in 1952 (see Fig. 11). As we saw above, Sandage and Schwarzschild investigated possible ways to get rid of this discrepancy, but they did not consider the possibility that convection might arise in the envelope. Hoyle and Schwarzschild realized that and commented that "radiative-envelope solutions integrated to the mathematical boundary condition of zero pressure and temperature at the surface of the star cease to give a satisfactory expression of the actual physical boundary condition."[154] Consequently, at the surface T must approach T_{eff} and ρ must approach a value consistent with the fact that light escapes from the star. In their late models, Hoyle and Schwarzschild found that density fell off too soon, i.e. reached a value that was acceptable for the photosphere before T decreased to T_{eff}.

Other scholars had already suggested that red stars must possess a convective zone in the outer envelope. In 1951, J. G. Gardiner attempted to build a red giant model moving inwards from the surface of the star.[155] Although he stuck to the assumption of a fully radiative envelope in his calculations, he found that "the outer regions of the atmosphere, as calculated for this model, will probably possess convectively unstable zones".[156] In 1953, following detailed calculations, Donald E. Osterbrock suggested that red dwarf stars might have an outer convective zone extending inwards for about 1/3 of the stellar radius.[157]

Convection in the envelope could well be the solution to the discrepancies Hoyle and Schwarzschild had found. This played a crucial role in explaining the increase in luminosity that occurred in the MN track of Fig. 13.

154 Hoyle and Schwarzschild (1955, p. 12).

155 Gardiner (1951).

156 Gardiner (1951, p. 102).

157 Osterbrock (1953).

Consequently, the two scholars introduced convection in the outer envelope, replacing Eq. 13 with $P = kT^{2.5}$ (Eq. 20 with $\gamma = 5/3$ and k to be determined by surface conditions). The equations for the core remained unchanged. It was then a matter of integrating the different zones: core, radiative envelope and convective envelope.

Let us begin with the core. Its uniform temperature, that had been shown by previous calculation to rise from about 13–20 million degrees along the LM path, was now supposed to keep constant at 20 $\times 10^6$ K throughout the evolution from M to N, so as to sustain the CNO cycle in the burning shell. However, Hoyle and Schwarzschild found that such an assumption could not be extended to the NO path—if it had been, the growing core mass would have triggered too large a decrease in the density of the burning shell. In order for the correct rate of energy production to be secured, this density decrease had to be balanced by a temperature boost. This implies a stage when the core warms up:

> Stellar models for the portion N-O of the evolutionary track may be defined as the portion of the evolution where the core begins to warm up. Along the portion L-M there was some degree of heating of the core, from about 13 $\times$ $10^{6\circ}$K to about 20 $\times 10^{6\circ}$K. This rise was occasioned mainly by the proton-proton chain's being the main source of energy production for stellar models near L. Once the carbon cycle becomes the main source of energy production, as it is near M, the value of T_1 [= temperature in the core] stays nearly constant at 20 $\times$ $10^{6\circ}$K until the neighbourhood at N is reached. Then a further stage of heating sets in.[158]

The computations for the inner zone of the star showed that, along with a core mass increase from approximately 0.3 to 0.6, temperature also grew from about 20 to 50 $\times$ 10^6 K, and the star's luminosity underwent a huge boost from a few tens to about ten thousand solar luminosities. As for the radius, it could not be deduced from the configurations of the core as it depended upon the boundary conditions of the star. More specifically, luminosity turned out to be determined by the core's mass during evolution and to be scarcely dependent upon surface temperature and radius, whereas surface temperature and (for a given luminosity) the radius itself resulted scarcely dependent upon the mass content of the core but strongly dependent upon the boundary conditions.

As for surface conditions, with special emphasis on the newly introduced convective zone, Hoyle and Schwarzschild carefully described the structure of the stellar atmosphere by dividing it into three regions:

> (1) an upper photospheric region with isothermal equilibrium; (2) a lower photospheric region in which energy flow is predominantly by radiation (convection may occur in this region but is unable to convey the whole energy flux); and (3) a convection zone in which the energy flow is predominantly by convection. The convection is divided into two parts, a thin outer shell in which the ionization of hydrogen is incomplete and the deeper zone that runs into the interior solution.[159]

158 Hoyle and Schwarzschild (1955, p. 19).

159 Hoyle and Schwarzschild (1955, p. 21).

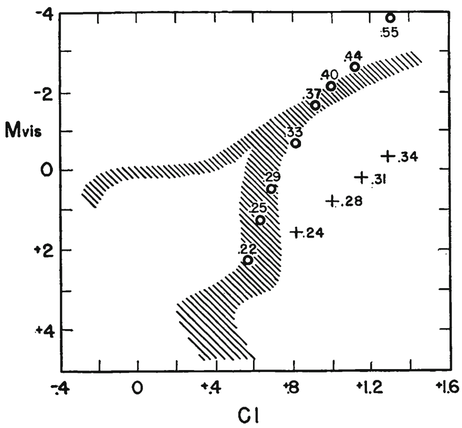

Fig. 15 Comparison between the observational *H–R* diagram for a globular cluster (*shaded area*) and the calculated models overtopped by a convective envelope. The numbers indicate the mass fraction of the core (which grows over time). Circles refer to type-II giants, crosses to type-I giants (from Hoyle and Schwarzschild 1955, p. 27). Reprinted with kind permission of the American Astronomical Society

An elaborate analysis of this multi-layered atmosphere and of the way it fitted the interior led to estimate the value of $R/R_\odot$ as determined by $L/L_\odot$. It turned out that, as luminosity increased up to values of the order of $7000L_\odot$, radius enlargement reached values of the order of $160R_\odot$. The representative points were depicted in the *H–R* diagram (Fig. 15).

A comparison between Figs. 14 and 15 shows that a core mass ratio around 0.20 acts as the "turning point" in the steepness of the giant branch. "For earlier models (i.e. with the mass ratio below 0.2)—the authors commented—it seems highly likely that convection in the envelope plays only a minor role".[160]

From Fig. 15, there appears to be a fair agreement between calculated models and observations, except for the models with a larger core/mass ratio, e.g. 0.55. According to Hoyle and Schwarzschild, this could be due to an assumption they made while discussing the atmosphere, namely that the supply of free electrons was mainly due to hydrogen. Clearly, at decreasing $T_{\rm eff}$ the contribution by metals is expected to exceed the contribution by hydrogen at some point, even if metals are supposed to be present in a very small amount. According to the calculations, the hydrogen–metal ratio required to overcome the discrepancy and to shift the 0.55 point downwards to meet the giant branch must be of the order of $\approx 10^5$, in fair agreement with coeval spectroscopic data.

Moreover, this observation fostered Hoyle and Schwarzschild's interest in metal-richer stars, like Type-I giants. Assuming a hydrogen-to-metals ratio of the order of $\approx 10^4$, the resulting models for type-I giant turned out to be less luminous than for type II, in good agreement with the observations: "in the photospheric regions […] the metal content […] greatly affects the extent of the tenuous envelope, i.e. the radius of the star." [161]

One final issue still remains open: what happens after the star has reached the topmost point O of its evolutionary curve? Why does not brightness get any larger?

[160] Hoyle and Schwarzschild (1955, p. 27).

[161] Hoyle and Schwarzschild (1955, p. 30).

All models constructed so far rested upon the isothermal core assumption. The steady boost in its temperature along the *NO* track goes hand in hand with the increase in energy generation and in brightness. But at some point the isothermal condition begins to fade. As more and more material is added to the core, gravitational energy is released. This generates a temperature gradient capable of triggering a flux of energy out of the core. Such en energy flux is proportional to the rate at which helium is added to the core, i.e. to the star's brightness.[162] As a consequence of that, the isothermal condition becomes increasingly questionable as the star climbs the NO track.

But what is the effect of a rising temperature in the innermost part of the core? As long as the inner part of the core maintains an essentially degenerate structure, the temperature increase has little effect on pressure and consequently on core structure: "it therefore appears that a really marked deviation from the models discussed is contingent on the core's assuming a non-degenerate structure, even in its inner part."[163] It remained to be decided what process could remove the core's degeneracy.

That process was probably—Hoyle and Schwarzschild argued—helium ignition, as suggested by Salpeter in 1952 and by Hoyle himself in 1954.[164] Helium fusion can cause problems when it takes place in a degenerate material. As is well known, a degenerate gas cannot balance itself by expanding if the energy production ratio is too high. Provided the star does not explode altogether, degeneracy must be removed at some point. According to the two scholars, this was the reason why the star stopped its ascent along the giant branch. Hoyle and Schwarzschild performed a quantitative analysis to tackle that issue. The helium ignition temperature was assumed to lie at 1.2×10^8 K. This led to estimate that the luminosity at which such a temperature was achieved in the centre must be around 5000 $L_\odot$—the same order of magnitude of point O in the diagram. This conclusion was the same for both type-I and -II giants.[165]

These considerations made an analysis of the OP track of Fig. 13 finally possible. The authors considered these stars to have a non-degenerate core, a double nuclear energy source resulting from helium burning in the core and hydrogen burning in the shell, and an enclosing radiative envelope (convection in the envelope was no longer considered). Due to the strong dependence of helium fusion on temperature, a convective innermost zone could be expected. Calculations were carried out in a way similar to previous investigations, and results (shown in Fig. 16) were reckoned to be "not entirely discouraging".[166]

At first, these stars began to slip down the giant branch, then moved mainly to the left, parallel to the OP path, although, roughly one order of magnitude above it. Hoyle

162 Provided L continues to be mainly sustained by nuclear reactions in the shell, i.e. provided the contribution of gravitational energy to L is minor. Hoyle and Schwarzschild reckoned this assumption to be valid, as they thought that the release of gravitational energy had the main effect to heat up the core.

163 Hoyle and Schwarzschild (1955, p.31).

164 Salpeter (1952) and Hoyle (1954).

165 Due to the fact that their work is, in the authors' words, "a preliminary reconnaissance of the problem" (see also note 151), small discrepancies between the estimated value of 5000 $L_\odot$ and the actual position of the point O should not be overrated. A greater problem seems to arise from the fact that the points associated to type-I giants lie much lower down. Hoyle and Schwarzschild ascribed this to different bolometric corrections for the two star types.

166 Hoyle and Schwarzschild (1955, p. 37).

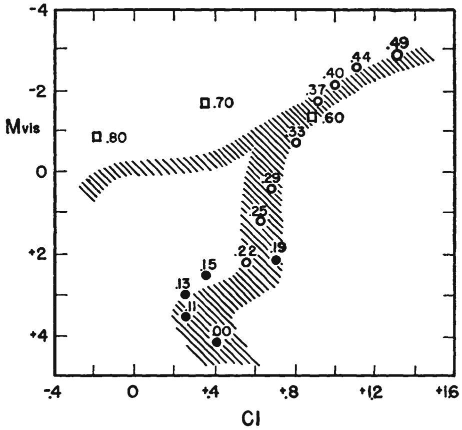

Fig. 16 *H–R* diagram with the whole evolution of type-II stars. The *squares* stand for the helium-burning calculated models (with core to total mass ratios equal to 0.60, 0.70 and 0.80); *filled and open circles* have the same meaning as in Figs. 14 and 15 respectively (from Hoyle and Schwarzschild 1955, p. 36). Reprinted with kind permission of the American Astronomical Society

and Schwarzschild regarded this as possibly due to the fact that all helium had to burn up before the cores attained a mass ratio of 0.7. This led to an even more complicated model, with an inner (perhaps partially degenerate and isothermal) carbon core and a helium-burning shell. But the integration of such a model was judged by the authors to be beyond the scope of their paper.

The paper by Hoyle and Schwarzschild is rightfully considered to be a pivotal work in the history of twentieth-century stellar astrophysics, especially when it comes to the giants issue. As we said, it provided the first sound model of red giant stars. Moreover, it depicted a grand scenario that brought together contributions from different fields: the concepts developed by Hoyle and Schwarzschild themselves and their colleagues regarding giant stars, as well as suggestions from the theory of stellar atmospheres and recent insights from nuclear physics about the helium ignition stage.

Moreover, Hoyle and Schwarzschild's paper plays a symbolic role, in that it "marks the end of the older era of stellar models, when models were generated one at a time by incredibly tedious work using hand-operated office desk calculators. This allowed snap-shot views of a star's history, but conveyed only a rough sense of evolution".[167] Here and there this difficulty surfaces in the words of the scientists involved as a vindication for making reasonable approximations. For example Hoyle and Schwarzschild stated that "the arithmetical work was found to be so heavy that it seemed desirable, always remembering that the present investigation is only a preliminary reconnaissance, to introduce an approximation into the calculations."[168] The two scientists also remarked, towards the end of their 1955 paper, that the huge difficulties they met in carrying on the calculations possibly were "an indication that an entirely new line of attack is desirable. In this case, we feel that the new line may turn out to be a fully automatic representation, using large electrical machines".[169]

167 Arny (1990, p. 225).

168 Hoyle and Schwarzschild (1955, p. 15).

169 Hoyle and Schwarzschild (1955, pp. 39–40).

1955 was also the year when the first computer-based stellar models appeared.[170] This opened up an era of much easier calculations and the subsequent possibility to develop more complex stellar models. The research fields involved in stellar evolution that benefited from it include convection in stellar envelopes, nuclear reactions beyond hydrogen fusion (helium ignition, helium flash, further reactions up to iron formation), post-giant evolution, mass loss.[171] Clearly, some of them were already included in Hoyle' and Schwarzschild's work.

10 Conclusions

All things considered, the road that led to the recognition of giants as evolved stars and to a satisfactory description of their structure was long but not excessively winding. The discontinuity of the twenties was followed by a period of "normal science"[172] that lasted at least until the 1950s, during which a number of scholars focused on explaining the evolutionary role played by giant stars.

What was the interplay of theory and observation in solving the giants issue, and how did progress in both fields influence the search for a solution?

Major advances were made in observational astronomy during the decades considered. In fields of astrophysics such as cosmology, ground-breaking achievements opened up new scenarios for the structure of the universe. On the other hand, if we restrict our analysis to the field we are dealing with, namely stellar astrophysics, we can easily spot a number of significant results like the study of the H–R diagram of globular clusters and the discovery of the existence of two populations of stars. Obviously enough, these achievements had an impact on the giants issue in the final part of the timeframe investigated in this paper. However, in our opinion they cannot be considered ground-breaking observations: from an observational viewpoint, the giants issue was already well established in the 1910s. The new observational data gathered afterwards provided additional insights that had to be incorporated in the theories, but did not have an impact on the basic features of the issue as a whole.

The issue was mainly a theoretical one. The final breakthroughs that led to a satisfactory explanation of the giants' structure came in steps. The ideas put forth by Strömgren, Gamow, etc. laid the foundations for further speculations. Of course, one should not forget that scientists went down a number of blind alleys as well. Some of them unveiled new possible paths to be pursued, others did not. Furthermore, the scientists involved depicted evolutionary paths on the H–R diagram that were to be discarded afterwards. But this was fully to be expected. These considerations do by no means affect the conclusion that, all in all, the solution to the giants issue should be regarded as "normal science".

170 Henyey et al. (1955a,b).

171 For a first-level elaboration about these topics see Arny (1990, pp. 225–227), Tassoul and Tassoul (2004, pp. 152–158).

172 Such a term is to be put in quotes, as it is not meant in a fully Kuhnian sense. For the same reason, the birth of theoretical astrophysics can more accurately be described as a discontinuity rather than as a revolution. See also note 1.

The issue was mainly a theoretical one, as I said. At a closer inspection, although the onset of nuclear physics obviously played a crucial role, the description of degenerate gas behaviour and the rejection of electron–proton annihilation (in turn an outcome of the mass-energy equivalence principle) were also important. However, although the inclusion of nuclear physics theories into astrophysics was a necessary and huge step, it was rather different in character from the discontinuity that the ground-breaking inclusion of atomic physics had brought about in a previous period. It was only natural that the theories about the nucleus should enter the existing field of theoretical astrophysics and provide scientists with new tools to look deeper into the mysteries of stellar matter, building on a picture of stellar structure that was already available. From nuclear physics scholars took what they needed to integrate stellar structure equations.

Having reached the end of the long road that led to a satisfactory understanding of red-giants structure, it might be appropriate to hint at an issue that has basically remained in the background so far. What does *physically* cause red giants to be formed? Answering this question is by no means a straightforward task. As we saw, the description of the structure of red giants resulted from a huge number of time-consuming numerical integrations. Clearly, there is no other way to try and figure out the steps of stellar evolution. But telling the technical aspects of integration apart from the essential underlying physics is not easy. If there is but one message to take home from the quest for an explanation of giants, it is that a complex interplay of several primary causes is at work here.

The difficulty of developing a simple physical explanation of red giants becomes apparent when reading books about their nature. Entry-level books point to the core collapse and core–envelope uncoupling as main factors that cause giants to puff up. Which they are, of course. But the other factors are seldom mentioned. Higher-level books are obviously very accurate, but usually much more focused on integration. I think it fair to conclude by highlighting this issue, as well as the existence of considerable literature about it.[173] This might be an interesting topic for further investigation work.

References

Applegate, J.H. 1988. Why stars become red giants. *Astrophysical Journal* 329:803–807.

Arny, T. 1990. The star makers: a history of the theories of stellar structure and evolution. *Vistas in Astronomy* 33:211–233.

Arp, H.C., Baum, W.A., and Sandage, A.S. 1952. The HR diagrams for the globular clusters M 92 and M 3. *Astronomical Journal* 57:4–5.

Arp, H.C., Baum, W.A., and Sandage, A.S. 1953. The color-magnitude diagram of the globular cluster M 92. *Astronomical Journal* 58:4–9.

Atkinson, R.d'E., and Houtermans, G. 1929. Zur Frage der Aufbaumöglichkeit der Elemente in Sterne. *Zeitschrift für Physik* 54:656–665.

Baade, W. 1944. The resolution of Messier 32, NGC 205, and the central region of the Andromeda Nebula. *Astrophysical Journal* 100:137–146.

[173] This topic is discussed, for example, in Yahil and van der Horn (1985), Applegate (1988), Celnikier (1990), Renzini et al. (1992), Faulkner (1997), Eggleton et al. (1998). Whitworth stated that "a simple explanation of giantness simply does not exist." (Whitworth 1989, p. 505). This list of references is just a tentative suggestion for further elaboration and should not be regarded as exhaustive.

Beech, M. 1988. The Schoenberg-Chandrasekhar limit: a polytropic approximation. *Astrophysics and Space Science* 147:219–227.
Celnikier, L.M. 1990. A simple way to assess the structure of red giants. *American Journal of Physics* 58:169–177.
Celnikier, L.M. 2006. *Find a hotter place! A history of nuclear astrophysics*. Singapore: World Scientific Publishing.
Cenadelli, D. 2008. The hydrogen abundance in stars: a first major step for quantitative astrophysics. *Journal of Astronomical History and Heritage* 11:134–145.
Chandrasekhar, S. 1939. *An introduction to the study of stellar structure*. Chicago: University of Chicago Press.
Collins, G.W. II. 1989. *The fundamentals of stellar astrophysics*. New York: W H Freeman & Co.
Cowling, T.G. 1930. On a point-source model of a star. *Monthly Notices of the Royal Astronomical Society* 91:92–108.
Cowling, T.G. 1931. Note on the fitting of polytropic models in the theory of stellar structure. *Monthly Notices of the Royal Astronomical Society* 91:472–478.
Cowling, T.G. 1934. The stability of gaseous stars. *Monthly Notices of the Royal Astronomical Society* 94:768–782.
Cowling, T.G. 1935a. The stability of gaseous stars (Second paper). *Monthly Notices of the Royal Astronomical Society* 96:42–60.
Cowling, T.G. 1935b. Convection in stars. *Observatory* 58:243–247.
Cowling, T.G. 1966. The development of the theory of stellar structure. *Quarterly Journal of the Royal Astronomical Society* 7:121–137.
Critchfiled, C.L., and Gamow, G. 1939. The shell-source stellar model. *Astrophysical Journal* 89:244–254.
DeVorkin, D.H., and Kenat, R. 1983a. Quantum physics and the stars (I): the establishment of a stellar temperature scale. *Journal for the History of Astronomy* 14:102–132.
DeVorkin, D.H., and Kenat, R. 1983b. Quantum physics and the stars (II): Henry Norris Russell and the abundances of the elements in the atmospheres of the Sun and stars. *Journal for the History of Astronomy* 14:180–222.
Dingle, H. 1963. A hundred years of spectroscopy. *The British Journal for the History of Science* 1:199–216.
Eddington, A.S. 1916. On the radiative equilibrium of the stars. *Monthly Notices of the Royal Astronomical Society* 77:16–35.
Eddington, A.S. 1917. Further notes on the radiative equilibrium of the stars. *Monthly Notices of the Royal Astronomical Society* 77:596–612.
Eddington, A.S. 1924a. The absorption of radiation inside a star. *Monthly Notices of the Royal Astronomical Society* 84:104–123.
Eddington, A.S. 1924b. On the relation between the masses and luminosities of the stars. *Monthly Notices of the Royal Astronomical Society* 84:308–332.
Eddington, A.S. 1927. *Stars and atoms*. Oxford: Oxford University Press.
Eddington, A.S. 1932. The hydrogen content of the stars. *Monthly Notices of the Royal Astronomical Society* 92:471–481.
Eddington, A.S. 1938. Star models with variable polytropic index. *Monthly Notices of the Royal Astronomical Society* 99:4–13.
Eggleton, P.P., Faulkner, J., and Cannon, R.C. 1998. A small contribution to the giant problem. *Monthly Notices of the Royal Astronomical Society* 298:831–834.
Faulkner, J. 1997. Why stars don't become red giants—and why they DO! *Advances in Stellar Evolution, Proceedings of the Workshop Stellar Ecology, held in Marciana Marina, Elba, Italy, 23–29 June 1996*, 9–10.
Gamow, G. 1938a. Nuclear energy sources and stellar evolution. *Physical Review* 53:595–604.
Gamow, G. 1938b. A star model with selective thermo-nuclear source. *Astrophysical Journal* 87:206–208.
Gamow, G. 1938c. Tracks of stellar evolution. *Physical Review* 53:907–908.
Gamow, G. 1939a. The energy-producing reaction in the Sun. *Astrophysical Journal* 89:130–133.
Gamow, G. 1939b. Physical possibilities of stellar evolution. *Physical Review* 55:718–725.
Gamow, G. 1939c. Nuclear reactions in stellar evolution. *Nature* 144:575–577.
Gamow, G. 1944. The evolution of contracting stars. *Physical Review* 65:20–32.
Gamow, G. 1945. The red-giant stage of stellar evolution. *Physical Review* 67:120–121.
Gamow, G., and Keller, G. 1945. A shell source model for red giant stars. *Reviews of Modern Physics* 17:125–137.

Gamow, G., and Teller, E. 1939. Energy production in red giants. *Physical Review* 55:791.
Gardiner, J.G. 1951. A model of a red giant star. *Monthly Notices of the Royal Astronomical Society* 111:102–110
Gingerich, O., and Lang, K. (eds.). 1979. *A source book in astronomy and astrophysics, 1900–1975*. Cambridge: Harvard University Press.
Goldberg, L. 1988. Atomic spectroscopy and astrophysics. *Physics Today* 41:38–45.
Hall Harrison, M. 1944. The generalized Cowling Model. *Astrophysical Journal* 100:343–346.
Hall Harrison, M. 1946. Stellar models with partially degenerate isothermal cores and point-source envelopes. *Astrophysical Journal* 103:193–206.
Härm, R., and Schwarzschild, M. 1955. Inhomogeneous stellar models. IV. Models with continuously varying chemical composition. *Astrophysical Journal* 121:445–453.
Hearnshaw, J.B. 1986. *The analysis of starlight: one hundred and fifty years of astronomical spectroscopy*. Cambridge: Cambridge University Press.
Hen, L., and Schwarzschild, M. 1949. Red-giant models with chemical inhomogeneities. *Monthly Notices of the Royal Astronomical Society* 109:631–646.
Henrich, L.R., and Chandrasekhar, S. 1941. Stellar models with isothermal cores. *Astrophysical Journal* 94:525–536.
Henyey, L.G., Lelevier, R., and Levée, R.D. 1955a. The early phases of stellar evolution. *Publications of the Astronomical Society of the Pacific* 67:154–160.
Henyey, L.G., Lelevier, R., and Levée, R.D. 1955b. Evolution of Sirius. *Publications of the Astronomical Society of the Pacific* 67:341–342.
Herrmann, D.B. 1984. *The history of astronomy from Herschel to Hertzsprung*. Cambridge: Cambridge University Press.
Hoskin, M. (ed.).1999. *The Cambridge concise history of astronomy*. Cambridge: Cambridge University Press.
Hoyle, F. 1954. On nuclear reactions occuring in very hot stars. I. The synthesis of elements from carbon to nickel. *Astrophysical Journal Supplement* 1:121–146.
Hoyle, F., and Lyttleton, R.A. 1942a. On the internal constitution of the stars. *Monthly Notices of the Royal Astronomical Society* 102:177–193.
Hoyle, F., and Lyttleton, R.A. 1942b. On the nature of red giant stars. *Monthly Notices of the Royal Astronomical Society* 102:218–225.
Hoyle, F., and Lyttleton, R.A. 1946. Note on stellar structure. *Monthly Notices of the Royal Astronomical Society* 106:525–530.
Hoyle, F., and Lyttleton, R.A. 1949. The structure of stars of non-uniform composition. *Monthly Notices of the Royal Astronomical Society* 109:614–630.
Hoyle, F., and Schwarzschild, M. 1955. On the evolution of type II stars. *Astrophysical Journal Supplement* 2:1–40.
Hufbauer, K. 2006. Stellar structure and evolution, 1924–39. *Journal for the History of Astronomy* 37:203–227.
Kippenhahn, R., and Weigert, A. 1990. *Stellar structure and evolution*. Berlin-Heidelberg: Springer.
Kuiper, G.P. 1937. On the hydrogen content of clusters. *Astrophysical Journal* 86:176–197.
Leverington, D. 1995. *A history of astronomy from 1890 to the present*. London: Springer.
Longair, M.S. 2007. *The cosmic century: a history of astrophysics and cosmology*. Cambridge: Cambridge University Press.
Masani, A. 1984. *Astrofisica*. Roma: Editori Riuniti.
Meadows, J. 1984. The origins of astrophysics. *American Scientist* 72:269–274.
Menzel, D.H. 1972. The history of astronomical spectroscopy I: qualitative chemical analysis and radial velocities and The history of astronomical spectroscopy II: quantitative chemical analysis and the structure of the solar atmosphere. *Annals of the New York Academy of Sciences* 198:225–244.
Mestel, L. 2004. Arthur Stanley Eddington: pioneer of stellar structure theory. *Journal of Astronomical History and Heritage* 7:65–73.
Milne, E.A. 1930. The analysis of stellar structure. *Monthly Notices of the Royal Astronomical Society* 91:4–55.
Oke, J.B., and Schwarzschild, M. 1952. Inhomogeneous stellar models. I. Models with a convective core and a discontinuity in the chemical composition. *Astrophysical Journal* 116:317–330.
Öpik, E. 1939. Stellar structure, source of energy, and evolution. *Publications de l'Observatoire Astronomique de l'Universite' de Tartu* 30:1–115.

Osterbrock, D.E. 1953. The internal structure of red dwarf stars. *Astrophysical Journal* 118:529–546.
Payne, C.H. 1925. Stellar atmospheres; a contribution to the observational study of high temperature in the reversing layers of stars. Harvard Observatory Monographs, no. 1. Cambridge.
Prialnik, D. 2000. An introduction to the theory of stellar structure and evolution. Cambridge: Cambridge Univerity Press.
Rebsdorf, S.O. 2007. Bengt Strömgren: interstellar glow, helium content, and solar life supply, 1932–1940. *Centaurus* 49:56–79.
Renzini, A., Greggio, L., Ritossa, C., and Ferrario, L. 1992. Why stars inflate to and deflate from red giant dimensions. *Astrophysical Journal* 400:280–303.
Richardson, R.S., and Schwarzschild, M. 1948. A stellar model for red giants of high central temperature. *Astrophysical Journal* 108:373–387.
Russell, H.N. 1914. Relations between the spectra and other characteristics of the stars. *Nature* 93, 227–231, 252–258 and 281–286.
Russell, H.N. 1925. The problem of stellar evolution. *Nature* 116:209–212.
Russell, H.N. 1929. On the composition of the Sun's atmosphere. *Astrophysical Journal* 70:11–82.
Salpeter, E.E. 1952. Nuclear reactions in stars without hydrogen. *Astrophysical Journal* 115:326–328.
Sandage, A.R. 1954. A survey of present knowledge of globular clusters and its significance for stellar evolution. *Les Processus Nucléaires dans les Astres, Communications présentées au cinquième Colloque International d'Astrophysique tenu à Liège les 10–12 Septembre, 1953*, 254–274.
Sandage, A.R., and Schwarzschild, M. 1952. Inhomogeneous stellar models. II. Models with exhausted cores in gravitational contraction. *Astrophysical Journal* 116:463–476.
Schönberg, M., and Chandrasekhar, S. 1942. On the evolution of the main-sequence star. *Astrophysical Journal* 96:161–172.
Schwarzschild, M., Rabinowitz, I., and Härm, R. 1953. Inhomogeneous stellar models. III. Models with partially degenerate isothermal cores. *Astrophysical Journal* 118:326–334.
Sitterly, B.W. 1970. Changing interpretations of the Hertzsprung-Russell Diagram, 1910–1940: a historical note. *Vistas in Astronomy* 12:357–366.
Smith, R.C. 1995. *Observational astrophysics*. Cambridge: Cambridge University Press.
Srinivasan, G. 1996. Stars: their structure and evolution. *Journal of Astrophysics and Astronomy* 17:53–76.
Strömgren, B. 1931a. The possible solutions of the "equations of fit" on the Standard Model. *Monthly Notices of the Royal Astronomical Society* 91:466–472.
Strömgren, B. 1931b. The point-source model with coefficient of opacity $\mathrm{k} = \mathrm{k}_1 \rho \mathrm{T}^{-3.5}$. *Zeitschrift für Astrophysik* 2:345–369.
Strömgren, B. 1932. The opacity of stellar matter and the hydrogen content of the stars. *Zeitschrift für Astrophysik* 4:118–152.
Strömgren, B. 1933. On the interpretation of the Hertzsprung-Russell-Diagram. *Zeitschrift für Astrophysik* 7:222–248.
Strömgren, B. 1938. On the helium and hydrogen content of the interior of the stars. *Astrophysical Journal* 87:520–534.
Strömgren, B. 1972. The rise of astrophysics. *Annals of the New York Academy of Sciences* 198:245–254.
Tassoul, J.-L., and Tassoul, M. 2004. *A concise history of solar and stellar physics*. Princeton: Princeton University Press.
Tayler, R.J. 1954. Evolution of massive stars. *Astrophysical Journal* 120:332–341.
Trumpler, R.J. 1925. Spectral types in open clusters. *Publications of the Astronomical Society of the Pacific* 37:307–318.
Wares, G.W. 1944. Partially degenerate stellar models. *Astrophysical Journal* 100:158–175.
Whitworth, A.P. 1989. Why red giants are giant. *Monthly Notices of the Royal Astronomical Society* 236:505–544.
Yahil, A., and van der Horn, L. 1985. Why do giants puff up? *Astrophysical Journal* 296:554–564.