Mon. Not. R. astr. Soc. (1989) **236,** 505–544

Why red giants are giant

A. P. Whitworth *Physics Department, PO Box 913, University of Wales College of Cardiff, CF1 3TH*

Accepted 1988 August 19. Received 1988 August 19; in original form 1987 April 13

Summary. There are four primary causes of giantness: [1] the opacity law in the lower envelope (Kramers or Thomson or a mixture); [2] the abrupt molecular weight change across the shell; [3] convection (following gravothermal instability), or electron degeneracy, in hydrogen-exhausted cores above a certain mass, and [4] approximate thermostatic control of the core by simultaneous hydrogen and helium burning. None of these four primary causes can be promoted above the others in the hierarchy of cause and effect underlying giantness. [1] furnishes highly extended envelope solutions, but these solutions would be inaccessible but for [2] and [3], and would be accessible only transiently but for [4]. A simple explanation of giantness simply does not exist.

The details of nuclear energy release are of secondary importance in determining the overall structure of giants. The high luminosities, low surface temperatures and outer convection zones of giants are consequences, not causes, of giantness.

The strong gravitational field and convective stability in the shell and lower envelope are key instruments in making giants giant but they are not primary causes.

To construct a realistic model of a giant from polytropic segments requires at least five segments. Extended giant envelopes are fundamentally different from soft polytropes. Giant cores are fundamentally different from hard polytropes.

Non-quasistatic core contraction is triggered by gravothermal instability, which always pre-empts the Schönberg–Chandrasekhar Limit. Spasmodic mass loss from giants can be triggered by small adjustments in the core.

1 Introduction

When a star evolves from the main sequence to the red giant phase, the core of the star (the region interior to the hydrogen-burning shell) becomes denser, for physical reasons which are well understood. What is at first sight rather unexpected, is that the envelope (the region exterior to the shell) simultaneously swells up until there is an enormous contrast between the mean core density and the mean envelope density, typically by a factor $\gtrsim 10^6$. The problem – and it is a problem which has exercised several authors over the years without any consensus being reached (e.g., Höppner & Weigert 1973; Eggleton & Faulkner 1981; Iben & Renzini

1984; Yahil and van den Horn 1985; Applegate 1988) – is to establish a clear hierarchy of cause and effect in the physics which underlies the swelling of the envelope. There are really two questions here:

(I) Why do giant envelope configurations exist?

(II) Why can giant envelope configurations only sit on cores above a certain mass?

Two points seem particularly important in this context. First the giant phase is not a final or transient instability. Although giants are predicted to make complicated and wide-ranging excursions on the Herzsprung–Russell Diagram, the giant phase is none the less stable, in the sense that the amount of hydrogen which a star burns as a giant is comparable with the amount it burns on the main sequence. In other words, the giant lifetime is ultimately a nuclear time-scale, and not a thermal or dynamical one. One can therefore expect to obtain useful insights from the time-independent equations of quasistatic stellar structure.

Secondly, most of the density contrast between the core and the envelope of a giant is due to a precipitous density drop, $|d\ln[\rho]/d\ln[M]| \gtrsim 33$, across and immediately above the shell (e.g., fig. 28.3 in Schwarzschild, 1958). This precipitous density drop is the key to understanding why giants are giant.

2 Plan of paper

In Section 3 we introduce the basic variables, and in Section 4 we describe a simple model of a giant.

In Sections 5–7 the approach is analytic; the critical features of giant structure are uncovered deductively. Section 5 deals with giant envelopes and the *almost-singular* solutions which describe them, showing how critically dependent on the opacity law these solutions are, the physical reason for their existence, and what conditions are required for them to be set up. Section 6 deals with the hydrogen-burning shell and the role of the associated molecular weight and luminosity changes. Section 7 deals with the core, the role of gravothermal instability in its evolution, the existence of a critical core structure which can support the *almost-singular* envelope solutions, the equivalence of electron degeneracy and convection in producing this critical structure, and the thermostatic control on core structure exerted by hydrogen and helium burning.

In Sections 8 and 9 the approach is synthetic. Section 8 shows how the different key parts of a giant fit together. Section 9 traces the hierarchy of physical cause and effect underlying this structure, and thereby identifies the primary causes at the head of the hierarchy. Section 10 summarizes the main conclusions.

To avoid diverting the reader too far from the main line of reasoning, certain aspects of the problem are treated in appendices. Appendix A deals with conditions at the base of the envelope in the *almost-singular* solutions. Appendix B maps out the quasistatic evolution of isothermal cores up to the Schönberg–Chandrasekhar Limit (hereafter SCL), and demonstrates that the SCL is always pre-empted by gravothermal instability. Appendix C demonstrates that our explanation of why giants are giant is confirmed in every regard by the detailed numerical models tabulated in Novotny (1973). Appendix D reviews previous attempts to explain why giants are giant and explains why they are either incomplete or ill-considered.

3 Variables

In view of the extreme inhomogeneity of giants and the associated steep gradients, it is advantageous to work with logarithmic variables and their derivatives. The following, whilst not unique, are useful since each one embodies a different aspect of stellar structure (energy

transport, hydrostatic balance and mass conservation), and this facilitates the disentangling of different physical effects.

$$A \equiv \frac{d\ln[T]}{d\ln[P]} = \text{Minimum}\left\{\frac{3P\bar{K}L}{64\pi G\sigma T^4 M}, \frac{1}{\Gamma_1}\frac{d\ln[T]}{d\ln[\rho]}\right\}, \tag{3.1}$$

$$B \equiv -\frac{d\ln[P]}{d\ln[R]} = \frac{GM\rho}{RP} \equiv \frac{R}{h_P}, \tag{3.2}$$

$$C \equiv \frac{d\ln[R]}{d\ln[M]} = \frac{M}{4\pi R^3\rho} \equiv \frac{\bar{\rho}}{3\rho}. \tag{3.3}$$

h_P is the pressure scale-height; all other variables have their usual meanings. We use $\ln[M]$ as the independent variable. Except in the core, A is preferred to the more seasoned polytropic exponent $\eta \equiv d\ln[P]/d\ln[\rho]$, because A is better behaved than η in the detailed numerical models tabulated by Novotny (1973); this is because T and P are continuous across a molecular weight discontinuity, whilst ρ is not. The polytropic index $n = (\eta - 1)^{-1}$ is a hapless variable in giants, since it tends to switch through $\pm\infty$ twice at each molecular weight change (i.e. four times in helium-burning giants); its mathematical advantages are redundant in the age of the computer.

4 Model

We start with a simple quasistatic model of a giant comprising a core, a shell, and an envelope. All chemical inhomogeneity and all luminosity generation are confined to the shell, which has negligible thickness. Inside the shell is the core, and outside is the envelope. We shall eventually need to relax some of the assumptions of this model, in particular to allow luminosity generation in the core. However, we first develop the simple model as far as it will go; and we retain throughout the assumption of zero luminosity generation in the envelope so that equations (3.1) to (3.3) are the full set of differential equations for the envelope structure. Further details of the simple model are summarized in Fig. 1.

5 Envelope

To understand why giant envelope configurations exist, we must simply understand why certain solutions of equations (3.1) to (3.3) involve a precipitous density drop at the base of the envelope. A precipitous density drop requires C to become large and to remain large over a significant part of the solution, i.e. large C and small $|C'| \equiv |d\ln[C]/d\ln[M]|$.

5.1 EQUILIBRIUM POINTS AND SINGULAR SOLUTIONS

We adopt an ideal gas equation of state $P = \rho kT/\bar{m}$ and an opacity law $\bar{K} = K_0\rho^\alpha T^{-\beta}/\bar{m}_e$, where $\bar{m}$ and $\bar{m}_e$ are the mean masses per gas particle and per electron, respectively. We also assume radiative energy transport. Then for the envelope, with L, $\bar{m}$ and $\bar{m}_e$ all uniform, equations (3.1) to (3.3) give

$$A' \equiv d\ln[A]/d\ln[M] = (\gamma A - \delta)BC - 1, \tag{5.1}$$

$$B' \equiv d\ln[B]/d\ln[M] = ABC - C + 1, \tag{5.2}$$

$$C' \equiv d\ln[C]/d\ln[M] = (1 - A)BC - 3C + 1, \tag{5.3}$$

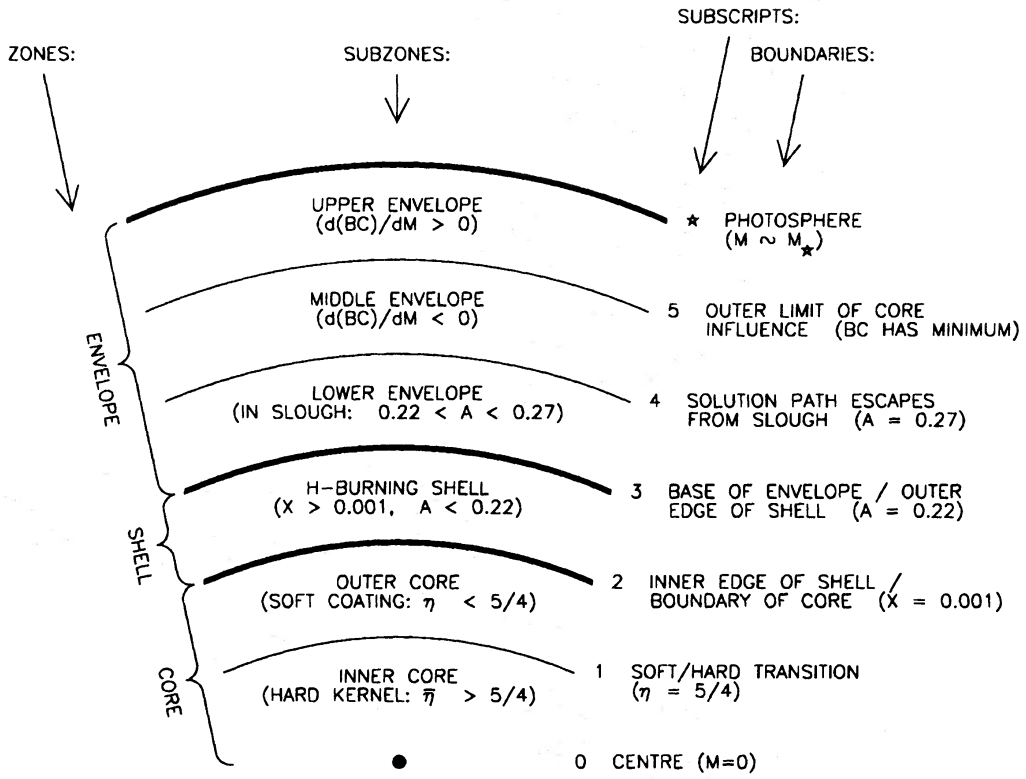


Figure 1. Schematic diagram of a giant showing (i) the different zones (lefthand column); (ii) the different subzones and their definitions (central column), and (iii) the boundaries between subzones, their definitions and the subscripts used to distinguish them (righthand column).

where $\gamma=\alpha+\beta+4$ and $\delta=\alpha+1$. Primes will be used throughout this paper to denote the operation $d\ln[\ ]/d\ln[M]$.

These equations have an equilibrium point $(A'=B'=C'=0)$ at

$$A_{\rm eq}=\frac{(2\delta+1)}{2(\gamma+2)},\qquad B_{\rm eq}=\frac{2(\gamma+2)}{(\gamma+1-2\delta)},\qquad C_{\rm eq}=\frac{(\gamma+1-2\delta)}{(\gamma-4\delta)}, \tag{5.4}$$

corresponding to a singular solution with $R\propto M^{C_{\rm eq}}$, $P\propto M^{-B_{\rm eq}C_{\rm eq}}$, $T\propto M^{-A_{\rm eq}B_{\rm eq}C_{\rm eq}}$, $\rho\propto M^{-(1-A_{\rm eq})B_{\rm eq}C_{\rm eq}}$, etc. These singular solutions are analogous to those for polytropes discussed by Chandrasekhar (1939). However, here we include the equation of radiative energy transport, so that there are three differential equations (rather than two) and we solve for the polytropic exponent $\eta=(1-A)^{-1}$ (rather than assuming it *ab initio*).

5.2 ALMOST-SINGULAR SOLUTIONS AND THE *SLOUGH*

Since in reality the opacity does not obey a perfect power law, the equilibrium point of equation (5.4) is blurred. However, there is still an extended *slough* around the equilibrium point, i.e. a finite volume of (A, B, C)-space in which solution paths advance very slowly with changing path parameter $\ln[M]$. This *slough* corresponds to a sheaf of *almost-singular* solutions, i.e. solutions which over a significant part of their range, $\Delta\ln[M]_{slough}$, diverge only slowly from the singular solution.

5.3 KRAMERS AND THOMSON OPACITY

For the two opacity laws operating near the base of a giant envelope, i.e. Kramers ($\alpha \simeq 1$, $\beta \simeq 7/2$, therefore $\gamma \simeq 17/2$ and $\delta \simeq 2$) and Thomson ($\alpha \simeq 0$, $\beta \simeq 0$, therefore $\gamma \simeq 4$ and $\delta \simeq 1$), the equilibrium points and the *sloughs* surrounding them have the following important properties, which are elaborated in Appendix A:

(i) Both equilibrium points, and hence also both *sloughs*, embrace exceptionally large C-values:

$$(A_{\rm eq}, B_{\rm eq}, C_{\rm eq}) \simeq \begin{cases} (5/21, 42/11, 11) & \text{(Kramers)}, \\ (1/4, 4, \infty) & \text{(Thomson)}. \end{cases} \tag{5.5}$$

Consequently the *almost-singular* solutions involve very steep density gradients, $|\rho'| = (1-A)\,BC \gtrsim 33$.

(ii) Both *sloughs* are very extended in the C-direction (see Figs 2 and A1).

(iii) The two *sloughs* are almost exactly co-extensive (again see Fig. A1). Consequently our arguments are still valid for a mixture of Kramers and Thomson opacity. This is important because conditions near the base of a giant envelope may run very close to the locus in (ρ, T)-space where Kramers and Thomson make equal contributions to the opacity.

(iv) In both *sloughs* the *almost-singular* solution paths move away from the associated equilibrium point, and hence out of the *slough*, with increasing path parameter $\ln[M]$ (i.e. the equilibrium points are unstable). Consequently, *almost-singular* envelope solutions must start near the singular solution at the base of the envelope and diverge from it further out.

(v) *Almost-singular* solution paths do not in general make multiple turns about the associated equilibrium point. They tend to move parallel to the elongation of the *slough* so that they expend a lot of mass before escaping from the *slough*, i.e. $\Delta \ln[M]_{slough}$ is large. This means that the resulting density drop,

$$\Delta \ln[\rho]_{slough} \simeq -(1-A_{\rm eq})\,B_{\rm eq}\,C_{\rm eq}\,\Delta \ln[M]_{slough}, \tag{5.6}$$

is both precipitous and large.

(vi) The condition for $C' > 0$ is $B > (3C-1)/(1-A)\,C$, which in the vicinity of the *slough* ($A \simeq 1/4$ and $C \gg 1$) is roughly equivalent to $B > 4$.

Thus, for Kramers opacity, Thomson opacity, or a mixture of the two, there is a very thick sheaf of *almost-singular* envelope solutions, in which the base of the envelope (subscript 3) is bogged down in the *slough* with $0.22 \lesssim A_3 \lesssim 0.27$ (typically $A_3 \simeq 1/4$), $3.4 \lesssim B_3 \lesssim 4.6$ (typically $B_3 \simeq 4$), $11 \lesssim C_3 \lesssim \infty$ (typically $C_3 \simeq 14$). This means that the density falls precipitously through several orders of magnitude near the base of the envelope, before bottoming out. All the giant models tabulated by Novotny (1973) have envelopes described by *almost-singular* solutions of precisely this form.

Here then we have the mathematical answer to question (I): giant envelope configurations exist because the opacity law (Kramers, Thomson, or a mixture) furnishes the equations of envelope structure with *almost-singular* solutions which involve a precipitous density drop at the base of the envelope. Moreover, the answer to question (II) (why can giant envelope configurations sit only on cores above a certain mass?) now reduces to identifying the circumstances which will deposit the base of the envelope in the *slough*. We reiterate that envelope solution paths based on equations (5.1) to (5.3) necessarily climb up out of the *slough* as they proceed outwards through the star. Therefore we have to look to additional terms in the structure equations for the shell, over and above those in equations (5.1) to (5.3), to transport the solution path into the *slough*.

5.4 INFLUENCE OF THE OPACITY LAW ON THE *ALMOST-SINGULAR* SOLUTIONS

At this juncture it is worth considering how critical the opacity law is. The condition for $11 \lesssim C_{eq} \lesssim \infty$ is

$$3.0\alpha \lesssim \beta \lesssim 3.2\alpha + 0.3. \tag{5.7}$$

We stress that this should not be interpreted as a discontinuous constraint on α and β. The inequalities are chosen simply to demonstrate how tight the limits on α and β can be made and yet still embrace both Kramers and Thomson opacity. The point is that if one were to relax these limits, the additional combinations of α and β thereby admitted would still have equilibrium points, *sloughs*, and *almost-singular* solutions, but they would be less effective in promoting giantness on two counts.

Firstly, the equilibrium points, and hence also the *sloughs*, would be at smaller C-values for $\beta > 3.2\alpha + 0.3$, whilst for $\beta < 3.0\alpha$ the *sloughs* would be shallower and therefore easier to escape from, giving smaller $\Delta \ln[M]_{slough}$. Either way, it is clear from equation (5.6) that the density drop at the base of the envelope would tend to be smaller.

Secondly, the *slough* would be less extended in (A, B, C)-space, and would therefore correspond to a thinner sheaf of *almost-singular* solutions (i.e. a narrower range of giant envelope configurations). This would mean that a more specific disposition of the core boundary is required for the base of the envelope to be deposited in the *slough* (i.e. for a giant envelope configuration to be set up). And this would tend to make gianthood a less sustained, less universal, phase of stellar evolution.

In other words, nature has contrived two distinct opacity laws, both of which are optimized to make a wide range of stars develop and sustain a giant phase. (This is demonstrated graphically in Fig. A1.)

5.5 THE PHYSICAL BASIS OF *ALMOST-SINGULAR* SOLUTIONS

Having shown mathematically how critical the opacity law is, it is appropriate to determine the physical basis of the *almost-singular* solutions which describe giant envelope configurations.

$A_3 \simeq 1/4$ means that the gas is stable against convection (i.e. $A < 2/5$) at the base of the envelope and so energy transport is radiative. This convective stability is very fundamental to giant structure (*cf.* Eggleton & Faulkner 1981) and is due to two factors. First shell hydrogen burning displaces luminosity generation from the centre of the star, thereby relieving the flux bottleneck. Hence, despite the high luminosity, a relatively small temperature gradient $|T'|$ is required for energy transport. Secondly, because the gravitational field is very strong at the boundary of a sufficiently massive and condensed core, a very large pressure gradient $|P'|$ is required for hydrostatic balance.

Now, if $|P'|$ is very large at the base of the envelope (typically $|P'_3| = B_3 C_3 \simeq 56$) and $|T'|$ is relatively small (typically $|T'_3| = A_3 B_3 C_3 \simeq 14$), then $|\rho'|$ is very large because in a chemically homogeneous ideal-gas region $\rho' = P' - T'$ (typically $|\rho'_3| = (1 - A_3) B_3 C_3 \simeq 42$).

On the one hand this means that as we move outwards from the base of the envelope the effective radiative conductivity,

$$\lambda_{rad} = \frac{16\sigma T^3}{3\rho\bar{K}} = \begin{cases} (16\sigma\bar{m}_e/3K_0^{Kr})\, T^{13/2}\rho^{-2} & \text{(Kramers)}, \\ (16\sigma\bar{m}_e/3K_0^{Th})\, T^3\rho^{-1} & \text{(Thomson)}, \end{cases} \tag{5.8}$$

stays high, and so $|T'|$ stays relatively small. Hence as long as $|P'|$ stays large, $|\rho'|$ also stays large. On the other hand, if $|\rho'|$ is too large, M hardly increases at all with increasing R and so g falls off roughly as R^{-2}. Then $|P'| = g^2/4\pi GP$ decreases, and with it $|\rho'|$.

The *almost-singular* solutions arise when an approximate balance is struck between these opposing influences on $|\rho'|$, so that a large $|\rho'|$ can be sustained over a large $\Delta \ln[M]$, thereby producing a large and precipitous density drop.

We note that for equilibrium points with large C_{eq} it is inevitable that $A_{eq} \simeq 1/4$ and $B_{eq} \simeq 4$. This is because large C_{eq} means that $|\rho'|_{eq}$ is large, and so M increases only slowly with increasing R and g falls off roughly as R^{-2}. Consequently, $P = g^2/4\pi G|P'|_{eq} \propto g^2$ falls off roughly as R^{-4}, i.e. $B_{eq} \simeq 4$. But then $dP/dR = -GM\rho/R^2 \propto \rho R^{-2} \propto R^{-5}$, i.e. $\rho \propto R^{-3} \propto P^{3/4}$, giving $T \propto P/\rho \propto P^{1/4}$, i.e. $A_{eq} \simeq 1/4$, and $dM/dR \propto R^2\rho \propto R^{-1}$, i.e. $M \simeq$ constant $+ 0(\ln[R])$.

5.6 CONDITIONS AT THE BASE OF THE ENVELOPE

We now consider question (II): why can giant envelope configurations only sit on cores above a certain mass? Or equivalently: what circumstances will so dispose the core boundary that the base of the envelope falls in the *slough*? At first sight it appears that the answer to this question is rather simple: only a massive core can (a) supply a strong gravitational field at its boundary, and (b) substantially relieve the flux bottleneck by displacing hydrogen burning well away from the centre of the star.

However, although a relatively low flux F_3 and a strong gravitational field g_3 are necessary conditions at the base of the envelope for setting up an *almost-singular* envelope solution, they are not sufficient. As we have just discussed, the *almost-singular* solutions involve a delicate balance between strong opposing influences on $|\rho'|$. This balance is embodied in the requirement that the base of the envelope falls in the *slough* with $A_3 \simeq 1/4$ (± 15 per cent), $B_3 \simeq 4$ (± 15 per cent), and $C_3 \simeq 14$ (± 30 per cent say, although strictly there is no upper limit on C_3). This is equivalent to requiring that the flux, gravitational field, and radius, all adopt rather specific values (in terms of local functions of state) at the base of the envelope, namely:

$$F_3 = 4\pi A_3(2B_3C_3/3)^{1/2}[\bar{l}F_{BB}/at_{FF}] \simeq 19.2[\bar{l}F_{BB}/at_{FF}] \pm 40 \text{ per cent}, \tag{5.9}$$

$$|g_3| = (\pi/2)(3B_3C_3/2)^{1/2}[a/t_{FF}] \simeq 14.4[a/t_{FF}] \pm 22 \text{ per cent} \tag{5.10}$$

and

$$R_3 = (2/\pi)(2B_3/3C_3)^{1/2}[at_{FF}] \simeq 0.278[at_{FF}] \pm 22 \text{ per cent}. \tag{5.11}$$

Here $\bar{l} = (\bar{K}\rho)^{-1}$ is the mean-free-path for a representative energy-transporting photon, $a = (P/\rho)^{1/2}$ is the isothermal sound speed, $t_{FF} = (3\pi/32G\rho)^{1/2}$ is the freefall time, and $F_{BB} = \sigma T^4$ is the equivalent blackbody flux. We reiterate that all the quantities on the righthand side of equations (5.9) to (5.11) (i.e. $\bar{l}$, a, t_{FF}, F_{BB}) are local functions of state, whilst those on the lefthand side are not. Equations (5.9) to (5.11) are a quantitative and testable prediction of this paper.

5.7 RÉSUMÉ

Giant envelopes are extended because they have a precipitous density drop at their base. Giant envelopes are described by a family of *almost-singular* solutions. This family of *almost-singular* solutions is critically dependent on the opacity law in the lower envelope. Both the extent of the family of *almost-singular* solutions (i.e. the range of available giant envelope configurations), and the extent of the density drop which they entail are maximized by a constraint which embraces both Kramers and Thomson opacity, and a mixture of the two. To set up a giant envelope configuration, the base of the envelope must fall in a particular region of (A, B, C)-space which we call the *slough*.

6 Shell

6.1 THE ROLE OF MOLECULAR WEIGHT AND LUMINOSITY GRADIENTS

To see how massive cores fulfil the rather precise requirements of equations (5.9) to (5.11), we must first consider the hydrogen-burning shell and identify the critical roles played here by luminosity and molecular weight changes. Since the *almost-singular* solution paths based on equations (5.1) to (5.3) lead out of the *slough* with increasing $\ln[M]$, it follows that the base of the envelope cannot be deposited in the *slough* unless equations (5.1) to (5.3) do not apply in the shell. However, as long as ideal gas presssure dominates in the shell, the only modifications to equations (5.1) to (5.3) are the additional terms due to changing luminosity and composition, namely:

$$A' = \begin{cases} (17A/2-2)BC-1+L'+\bar{m}'-\bar{m}'_{\rm e} & \text{(Kramers)}, \\ (4A-1)BC-1+L'-\bar{m}'_{\rm e} & \text{(Thomson)}, \end{cases} \tag{6.1}$$

$$B' = ABC-C+1+\bar{m}', \tag{6.2}$$

$$C' = (1-A)BC-3C+1-\bar{m}'. \tag{6.3}$$

Here $L' \equiv d\ln[L]/d\ln[M] = \varepsilon M/L$, $\bar{m}' \equiv d\ln[\bar{m}]/d\ln[M]$, and $\bar{m}'_{\rm e} \equiv d\ln[\bar{m}_{\rm e}]/d\ln[M]$. These are the new terms which must act to transport the solution path into the *slough* against the collective opposition of all the other terms.

We shall show later (in Section 7 and Appendix B) that to deposit the base of the envelope in the *slough*, starting from the core boundary, necessarily requires a change of B and/or C, and a large increase in A. From a comparison of equations (6.2) and (6.3) with (5.2) and (5.3) it follows that $\bar{m}'$ is the only term which can produce a change in B or C, and L' is the only term which can produce a large increase in A. Therefore $\bar{m}'$ and L' both play critical roles in making giants giant.

L' and $\bar{m}'$ must be large if they are to overpower the other terms in equations (6.1) to (6.3). Large L' and $\bar{m}'$ means that the shell must be thin and convectively stable. We have already explained how a massive core promotes convective stability at its boundary. Massive cores also constrict the thickness of the shell, again basically by supplying a large gravitational field which in turn induces steep gradients in the density, the temperature, and hence the specific nuclear energy release rate. Typically in the shell $A \simeq 1/8$ and $BC \simeq 56$, so that $|\rho'| \simeq 49 + |\bar{m}'|$ and $|T'| \simeq 7$.

6.2 JUMP CONDITIONS ACROSS A THIN SHELL

In a sufficiently thin shell the terms (L', $\bar{m}'$, $\bar{m}'_{\rm e}$) dominate equations (6.1) to (6.3). We can then neglect the thickness of the shell and apply jump conditions between the boundary of the core (subscript 2) and the base of the envelope (subscript 3):

$$\Delta\ln[A] = \begin{cases} \Delta\ln[L]+\Delta\ln[\bar{m}]-\Delta\ln[\bar{m}_{\rm e}], & \text{(Kramers)}, \\ \Delta\ln[L]-\Delta\ln[\bar{m}_{\rm e}], & \text{(Thomson)}, \end{cases} \tag{6.4}$$

$$\Delta\ln[B] = \Delta\ln[\bar{m}] = -\Delta\ln[C]. \tag{6.5}$$

6.3 THE *IMAGE*

Most of a giant's luminosity is generated in the shell, so $\Delta\ln[L]$ is large; and in crossing the shell (outwards) the molecular weight decreases by a factor $f \simeq 2$. Therefore, if the base of the

envelope is to fall in the *slough* with $A_3 \simeq 1/4$, $B_3 \simeq 4$, $C_3 \gtrsim 11$, the boundary of the core must fall in an *image* of the *slough* with $A_2 \ll 1/4$, $B_2 \simeq 8$, $C_2 \gtrsim 5$ (see Fig. 2).

6.4 RÉSUMÉ

To set up a giant envelope configuration, the base of the envelope must fall in the *slough*, but the core boundary cannot reach the *slough*. To deposit the base of the envelope in the *slough* starting from the core boundary necessarily requires both a change in B and/or C (which can only be effected by the molecular weight gradient $\bar{m}'$ across the shell), and a large increase in A (which can only be effected by the luminosity gradient L' across the shell). The shell must be thin and convectively stable so that $\bar{m}'$ and L' are large enough to overpower the other terms in the equations, since these other terms act collectively to lead the solution path out of the *slough*. Given the roughly twofold decrease in the molecular weight, and the large increase in the luminosity across the shell, the base of the envelope will only be deposited in the *slough* if the core boundary falls in a particular region of (A, B, C)-space which we call the *image*.

We must distinguish two strands in this argument. The fact that the equilibrium points based on equations (5.1) to (5.3) are unstable (see Appendix A) proves that solution paths based on these equations lead out of the *slough* with increasing path parameter $\ln[M]$. This in turn proves that the solution path can only be transported into the *slough* by the additional terms which operate in the shell, i.e. the molecular weight gradient which changes B and C, and the luminosity gradient which changes A. However, this on its own does not prove that the molecular weight gradient is essential, because it might be that only a change in A is needed between the core boundary and the base of the envelope. Therefore the final step in proving that the molecular weight gradient is essential is to demonstrate that a change in B and/or C is required between the core boundary and the base of the envelope, i.e. that the core boundary cannot reach $B \simeq 4$ and $C \gtrsim 11$ with any value of A. We prove this in the next section.

7 Core

What we must do to answer question (II) (why can giant envelope configurations only sit on cores above a certain mass?) is to investigate the different types of structure the core might have, starting with the simplest, and to determine for each:

(i) whether the core boundary is suitably disposed to deposit the base of the envelope in the *slough*, and

(ii) whether the changes in luminosity and molecular weight across the shell are vital in this regard (i.e. whether the core boundary is in the *slough* or the *image*).

7.1 GRAVOTHERMAL INSTABILITY AND THE SCHÖNBERG-CHANDRASEKHAR LIMIT

In conducting this investigation of different core structures, we need to have regard, not simply for the existence of core solutions, but also for which core solutions actually arise in the natural course of stellar evolution. In this context it is important to appreciate the role of gravothermal instability. As a criterion for gravothermal instability we use the inequality $\mathcal{G} \equiv \partial P_2/\partial R_2)_{M_2,T_2} > \mathcal{G}_c = 0$, where (P_2, R_2, T_2) are the pressure, radius and temperature at the core boundary, and M_2 is the core mass. Since the pressure exerted on the core by the overlying layers is likely to increase if the core shrinks, $\mathcal{G} > \mathcal{G}_c = 0$ is a rather conservative instability criterion. In other words, the gravothermal stability limit (hereafter GSL) probably occurs at $\mathcal{G}_c < 0$. This would make the core even more prone to instability, which would simply strengthen our conclusions.

The most important consequence of gravothermal instability is that for stars with mass $M_* \gtrsim M_\odot$; it pre-empts the Schönberg–Chandrasekhar Limit (SCL), and marks a significant change in the pattern and pace of core evolution. The early stages of post-main sequence evolution are essentially quasistatic. The rate of release of gravitational potential energy in the core is small, and so the core approximates to isothermality at the hydrogen-burning temperature T_H. The SCL represents the maximum core mass for which there exist isothermal equilibrium states, stable or unstable; and it has frequently been inferred that the core only starts to contract rapidly enough to heat up once it has passed the SCL. However, in reality the core encounters the GSL before the SCL, and it is on passing the GSL that contraction becomes rapid enough for the core to heat up.

There is a 'Catch 22' here. Even if gravothermal instability does not develop (i.e. the instability growth time-scale is too long) and the evolution tries to continue quasistatically through unstable equilibrium states beyond the GSL and up to the SCL, it cannot do so. For beyond the GSL the rate of release of gravitational energy in the core, $(-d\Omega_2/dt)_{eq} = (-d\Omega_2/dM_2)_{eq}(dM_2/dt)$, becomes too large to neglect, and so the evolution can no longer be considered quasistatic. Recalling that for a hydrogen-burning shell $(dM_2/dt) \simeq L_*/0.005c^2$, and that during these early stages of post-main sequence evolution L_* is roughly constant, the enormous increase in $(-d\Omega_2/dt)$ on passing the GSL is essentially an enormous increase in $(-d\Omega_2/dM_2)_{eq}$, where

$$-\frac{d\Omega_2}{dM_2} = \frac{d}{dM_2}\int_0^{M_2} \frac{GM\,dM}{R(M)} - P_2 4\pi R_2^2 \frac{dR_2}{dM_2}. \tag{7.1}$$

The second term on the right represents work done on the core by the weight of the overlying layers. The enormous increase in $(-d\Omega_2/dM_2)$ arises because beyond the GSL small increases in core mass start to involve very large changes in the equilibrium core structure (see Appendix B for details).

The non-quasistatic contraction which ensues at the onset of gravothermal instability causes the core to heat up, particularly towards the centre, so that there is a temperature gradient to drive out some of the gravitational energy being released. We shall see later that, in the absence of electron degeneracy, this temperature gradient is a vital feature of the core of a giant.

7.2 STRATEGY FOR INVESTIGATING POSSIBLE CORE STRUCTURES

We now start our investigation of different types of core structure, concentrating first on those which are quasistatic. What we shall find is that most types of core structure must be rejected because they are unable to deliver the base of the envelope into the *slough*. In fact it will turn out that there is only one type of core structure which can deliver the base of the envelope into the *slough*. This critical core structure involves the core having a hard kernel (i.e. $\eta > 5/4$ in the centre, typically $\eta \simeq 5/3$) and a soft coating (i.e. $\eta < 5/4$ towards the boundary); and it puts the core boundary in the *image*, so that the molecular weight and luminosity changes across the shell are vital to deliver the base of the envelope into the *slough*.

In discussing core solutions we shall make extensive use of the (B, C)-plane, but we shall abandon A in favour of η because by using η we can emphasize the similarity between isothermal inner cores supported by electron degeneracy pressure ($\eta \simeq 5/3$, $A \simeq 0$) and convective ideal-gas inner cores ($\eta \simeq 5/3$, $A \simeq 2/5$). All core solutions start from $B_0 = 0$, $C_0 = 1/3$, at the centre.

We shall refer to the projection of the *slough* on the (B, C)-plane simply as the *slough*; and similarly to the projection of the *image* on the (B, C)-plane simply as the *image*. This will

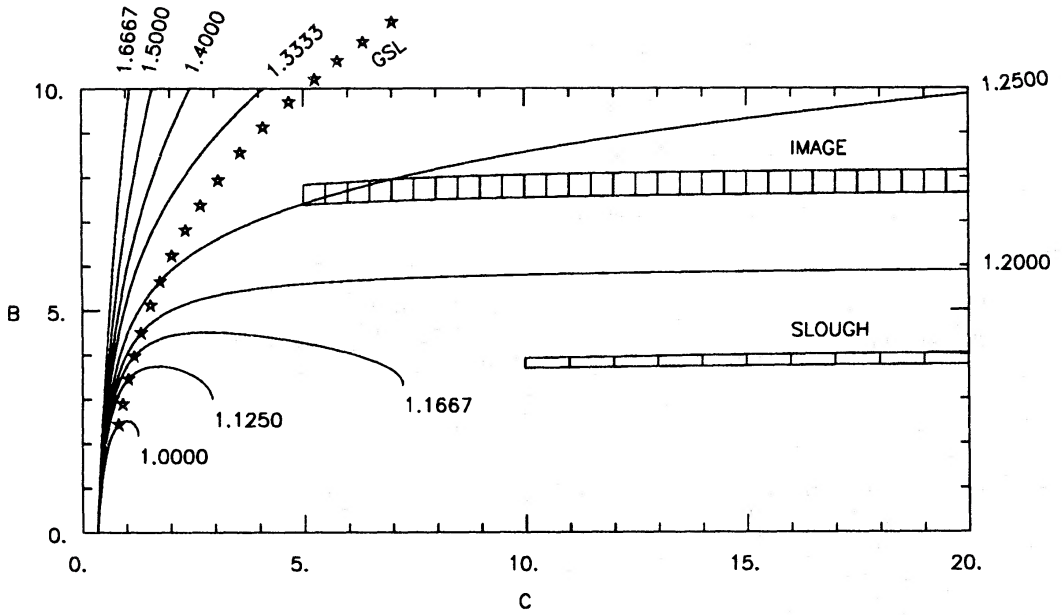


Figure 2. Solution paths on the (B, C)-plane for pure polytropic cores. Each path is labelled with the value of the polytropic exponent η. Asterisks mark points of gravothermal instability. The hatched regions represent the *slough* and its *image*. The full three-dimensional *slough* in (A, B, C)-space is defined, analysed and illustrated in Appendix A. What is shown here is the projection of one contour of the *slough* on the (B, C)-plane. The *image* is obtained by transforming the *slough* according to $B \rightarrow 2B$ and $C \rightarrow C/2$.

enable us to focus our attention on the key question of whether changes in B and/or C are needed between the core boundary and the base of the envelope, and hence whether the molecular weight change across the shell is critical. (There is no problem with increasing A because the existence of a large luminosity gradient in the shell is not in dispute.) At the same time it must be borne in mind that for the core boundary to fall in the *slough* with $A_2 \simeq 1/4$ corresponds to $\eta_2 \simeq 4/3$. Similarly for the core boundary to fall in the *image* with $A_2 \ll 1/4$ corresponds to $\eta_2 \simeq 1$.

7.3 NON-DEGENERATE ISOTHERMAL QUASISTATIC CORES

The simplest credible core structure is the isothermal ideal-gas sphere. This is a reasonable model for the core of a young post-main sequence star. The corresponding solution path is plotted on the (B, C)-plane in Fig. 2 and labelled '1.0000'. Any point on this solution path, up to the GSL at $B_{\rm GSL}=2.43$, $C_{\rm GSL}=0.822$, can represent the core boundary in a quasistatic solution. The solution path goes nowhere near the *slough* or its *image* (even if we neglect the GSL). Therefore isothermal ideal-gas cores cannot support giant envelopes (whether there is a molecular weight change across the shell or not).

7.4 PURE POLYTROPIC QUASISTATIC CORES

Now we consider pure polytropic cores, i.e. cores with uniform polytropic exponent η. For a chemically homogeneous ideal gas, $\eta=(1-A)^{-1}$; the isothermal ideal gas considered in the previous subsection is just a special case of this type with $A=0$ and $\eta=1$. Solution paths for polytropic cores are given by

$$B'=(1-\eta^{-1})BC-C+1, \qquad B_0=0, \tag{7.2}$$

$$C'=\eta^{-1}BC-3C+1, \qquad C_0=1/3. \tag{7.3}$$

Some paths for representative values of η are plotted in Fig. 2. For $\eta \geq 4/3$ the path extends to $(B, C)=(\infty, \infty)$ and is gravothermally stable all the way. For $0<\eta<4/3$ the path ends on an equilibrium point at $B_{eq}=2\eta/(2-\eta)$ and $C_{eq}=(2-\eta)/(4-3\eta)$, but becomes gravothermally unstable before reaching the equilibrium point. Only points on the solution path up to the GSL can represent the core boundary in a quasistatic solution. From Fig. 2, we see that for gravothermally stable, pure polytropic cores (of whatever exponent) the core boundary never reaches the *slough* or its *image*. Therefore quasistatic giants cannot have pure polytropic cores.

7.5 TWO-ZONE POLYTROPIC QUASISTATIC CORES

Now we consider two-zone polytropic cores, i.e. cores with an inner zone having uniform $\eta=\eta_i$ and an outer zone having uniform $\eta=\eta_o\neq\eta_i$. Again we ask whether the core boundary can reach the *slough* or its *image* without encountering gravothermal instability.

We first note from equations (7.2) and (7.3) that at any point on the $(B>0, C>0)$-plane the slope of the solution path is a monotonically increasing function of η for $\eta>0$. Consequently there are two distinct possibilities for a two-zone polytropic core with its boundary in the *image*. In the first possibility the solution path lies below that for a pure polytropic core with $\eta\simeq 5/4$, so that $\eta_i<5/4<\eta_o$ (see Fig. 3). This possibility is always gravothermally unstable and can therefore be rejected. In the second possibility the solution path lies above that for a pure polytropic core with $\eta\simeq 5/4$, so that $\eta_i>5/4>\eta_o$ (again see Fig. 3). In this case the core boundary can indeed reach the *image* without encountering gravothermal instability, although $\eta_i>5/4>\eta_o$ is only a necessary condition. Similarly $\eta_i>13/11>\eta_o$ is a necessary condition for the core boundary to reach the *slough* without encountering gravothermal instability.

We therefore have a general theorem: 'for two-zone polytropic cores, only the hard-kernel/soft-coating ones with $\eta_i>5/4>\eta_o$ (or $\eta_i>13/11>\eta_o$) afford the possibility of the core boun-

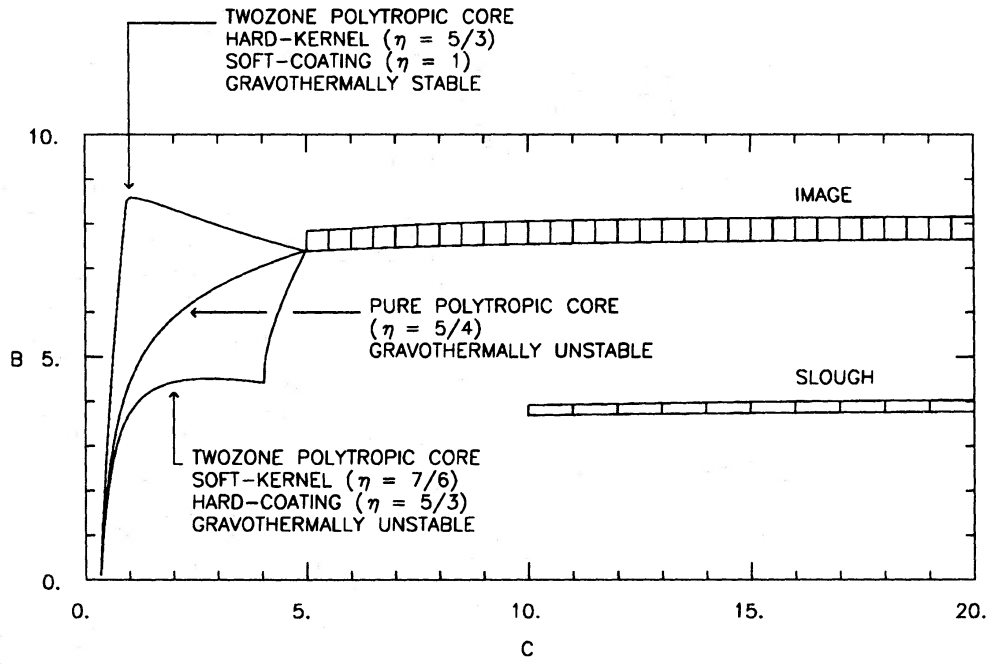


Figure 3. Solution paths on the (B, C)-plane for three representative polytropic cores which just deliver the core boundary into the *image*. One is a pure polytropic core with $\eta=5/4$; this core is gravothermally unstable. The other two are hybrid two-zone polytropic cores. Of these, one has a soft kernel and a hard coating and is also gravothermally unstable, whilst the other has a hard kernel and a soft coating and is gravothermally stable.

dary reaching the *image* (or the *slough*) without encountering gravothermal instability.' The proof of this theorem is straightforward but tedious and so it is not reproduced here. It consists simply of choosing representative points in the *image* (or the *slough*) and then conducting sequences of numerical integrations to determine for each such point which combinations of η_i and η_o allow the solution path to reach that point without encountering gravothermal instability.

7.6 PARTIALLY DEGENERATE ISOTHERMAL QUASISTATIC CORES

Once an isothermal core becomes electron degenerate in its central regions, it approximates to a hard-kernel/soft-coating two-zone polytropic structure with $\eta_i \simeq 5/3$ and $\eta_o \simeq 1$. The properties of partially degenerate isothermal cores are discussed in Appendix B. Some solution paths are plotted on Fig. 4. Each curve is marked with the central density y_0, where y is a dimensionless density defined so that for $y > 1$ degeneracy pressure dominates with $\eta \simeq 5/3$, and for $y < 1$ ideal gas pressure dominates with $\eta \simeq 1$. The line of asterisks marks the GSL.

From Fig. 4 it appears that for $y_0 \gtrsim 5$ the core boundary can reach the *slough* without encountering gravothermal instability, but only just. We should treat this result with reserve in case, as we suspect, the gravothermal instability condition $\mathcal{G} > \mathcal{G}_c = 0$ is too conservative (i.e. in case $\mathcal{G}_c < 0$). In fact we show in Appendix B that these solutions cannot occur during the natural course of stellar evolution because they are pre-empted by gravothermal instability (i.e. to get to them the core would have to pass through gravothermally unstable states.) It follows that B and/or C must change between the core boundary and the base of the envelope, and hence that the steep molecular weight gradient across the shell is essential. We must therefore concentrate on solutions with the core boundary in the *image*.

From Fig. 4 it is clear that for $y_0 \gtrsim 10$ the core boundary can reach the *image* without going anywhere near gravothermal instability. Thus we have at last found a family of gravothermally stable core solutions with a realistic equation of state which (given a twofold molecular weight decrease across the shell) can support giant envelope configurations. Indeed the $M_\odot$ giant model tabulated by Novotny (1973) involves a core of precisely this type. However, there are two complications.

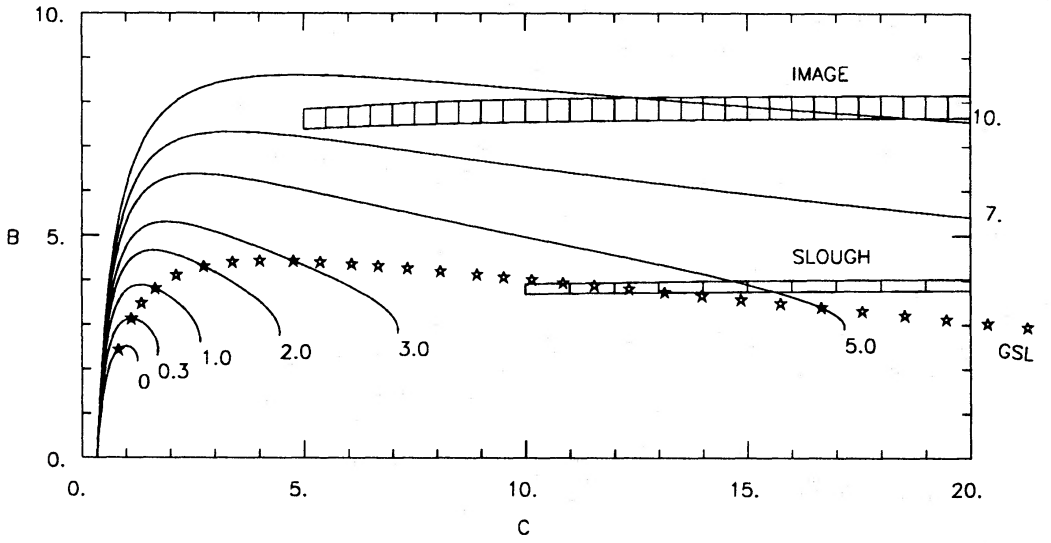


Figure 4. Solution paths on the (B, C)-plane for isothermal cores. Each plot is labelled with the dimensionless central density y_0. Asterisks mark points of marginal gravothermal instability.

The first complication is that during the course of quasistatic post-main sequence evolution for stars with $M_* \lesssim M_\odot$ partially degenerate core solutions with their boundary in the image arise only as a rather transient phase (see Appendix B). As the evolution proceeds, B_2 and C_2 increases monotonically so that the core boundary enters the *image* from below on the (B, C)-plane (see Fig. B3) and then re-emerges above the *image*. As the core boundary enters the *image* the star swells up to become a giant. However, if the core boundary re-emerges above the *image*, the base of the envelope will no longer be deposited in the *slough*, but above it where $C' > 0$, and so the solution path in the lower envelope will head off to enormous C-values. It seems likely that this will cause the star to lose mass; it is possible that mass loss will occur in a self-regulating manner so that the core boundary stays close to the *image* and the star remains a giant, but the evolution will clearly cease to be quasistatic.

The second complication is that the cores of more massive stars do not become degenerate until long after they have become giant. Therefore partially degenerate isothermal cores cannot be the whole solution to our problem.

7.7 THE CRITICAL CORE STRUCTURE

However, we are now very close to the solution to our problem. For it follows that a non-degenerate non-isothermal core could also have its boundary in the *image* without succumbing to gravothermal instability if it were to adopt a hard-kernel/soft-coating structure similar to that identified in Section 7.5. This would require a temperature gradient in the core, and so could only arise if there were quasistatic luminosity generation in the core, i.e. helium burning. Ideally it would involve the kernel being convective with $\eta \simeq 5/3$, and the coating being radiative with η not much greater than unity at the boundary. But this is precisely what occurs in all the $5M_\odot$ giant models tabulated by Novotny (1973). Thus the hard-kernel/soft-coating structure is a fundamental property of giant cores.

7.8 NON-QUASISTATIC CORES

We can now understand what happens to stars during the transient non-quasistatic evolutionary phase between the onset of gravothermal instability and the ignition of helium, as the core contracts and heats up, and the envelope swells to gianthood. This phase is non-quasistatic because the specific energy generation rate ε is not a local function of state (in the sense of the Vogt–Russell Theorem), and so the structure is not governed exclusively by time-independent equilibrium equations.

At the onset of gravothermal instability, the core is approximately isothermal. The values of (B, C) at the core boundary fall near the line of asterisks on Fig. 4, i.e. $B_2 > 2$, $C_2 \gtrsim 1$. The ensuing rapid contraction and heating leads to the development of a temperature gradient in the core, thereby increasing the polytropic exponent and displacing the core solution path upwards on the (B, C)-plane. The core starts to develop a hard-kernel/soft-coating structure of the type defined in Section 7.5.

From equations (7.2) and (7.3) we see that it is the hard kernel which carries the solution path to larger B-values; whilst the soft coating carries the solution path to larger C-values, particularly once $B > 4$. The simultaneous increase of B_2 and C_2 transports the core boundary ever deeper into the *image* and so the base of the envelope falls ever deeper into the *slough*. The result is a rapid growth in the precipitous density drop at the base of the envelope (see equation (5.6)), causing a sudden swelling of the envelope to giant proportions.

7.9 THERMOSTATIC CONTROL OF THE CORE STRUCTURE DUE TO HYDROGEN AND HELIUM BURNING

The non-quasistatic contraction of the core, and the associated upward displacement of the solution path on the (B, C)-plane, halt once the central temperature becomes high enough to ignite helium. The reason why this occurs rather precisely at the stage when the core boundary is near the middle of the *image* (with $B_2 \simeq 8$), and why the core boundary then remains roughly locked in the middle of the *image*, has to do with the ratio between the helium- and hydrogen-burning temperatures. Core helium burning requires $100 \lesssim T_{\rm He}/10^6\,{\rm K} \lesssim 160$, and shell hydrogen burning requires $25 \lesssim T_{\rm H}/10^6\,{\rm K} \lesssim 40$, so that $T_{\rm He}/T_{\rm H} \simeq 4$ (very approximately). Now for a complete $\eta \simeq 5/3$ polytrope supported by ideal gas pressure the central temperature is given by $T_0 \simeq GM_*\bar{m}/2kR_*$. If we apply this result to the core (i.e. if we neglect the small competing corrections due to the soft coating and the finite pressure at the core boundary) and put $T_0 \simeq GM_2\bar{m}/2kR_2$, it follows that

$$B_2 = GM_2\bar{m}/R_2kT_2 \simeq 2T_0/T_2. \tag{7.4}$$

But $T_2 = T_{\rm H}$, and T_0 increases until $T_0 = T_{\rm He}$. Thus B_2 increases until $B_2 \simeq 2T_{\rm He}/T_{\rm H} \simeq 8$ and then becomes stuck at this value, which corresponds rather precisely to the *image*.

We note that if helium burnt at a much higher temperature, $T_{\rm He} \gg 4T_{\rm H}$, non-quasistatic core contraction and heating would have to proceed much further. The solution path would then be displaced even further upwards on the (B, C)-plane. The core boundary would fall above the *image* and the base of the envelope would fall above the *slough*. Since the solution paths here all involve rapidly increasing C (i.e. large positive C'), the star would presumably experience mass loss.

The reason why giants shrink somewhat when helium ignites, is that the core starts to make a larger contribution to the luminosity. The growth of the luminosity in the core increases the extent of its hard kernel at the expense of its soft coating, so that the solution path does not reach such large C-values. This reduces the density drop at the base of the envelope and so the envelope shrinks (whilst still remaining giant).

7.10 RÉSUMÉ

To set up a giant envelope configuration the base of the envelope must fall in the *slough*. This in turn necessitates a particular disposition of the core boundary, which depends on whether there is a molecular weight change across the shell or not. If there is no molecular weight change across the shell, the core boundary also has to fall in the *slough*; but the boundary of a quasistatic core cannot reach the *slough* because it cannot reach such large C-values ($C \gtrsim 11$) at such small B-values ($B \simeq 4$) without encountering the GSL. If there is a molecular weight change across the shell, the core boundary has to fall in the *image*; and the boundary of the core can reach the *image* because it can reach more modest C-values ($C \gtrsim 5$) at larger B-values ($B \simeq 8$) without encountering the GSL. Hence the molecular weight change across the shell plays a critical role in setting up giant envelope configurations.

In order for the core boundary to reach the *image*, the core must adopt a critical structure involving a hard kernel which increases B to ~ 8, and a soft coating which increases C to $\gtrsim 5$. The soft coating ensures that the core boundary is truly in the *image* with $\eta_2 \lesssim 1$ and hence $A_2 \ll 1/4$. In low-mass stars the critical core structure occurs initially because of electron degeneracy. Otherwise it occurs because of convection, following non-quasistatic contraction and heating. In the latter case, the ratio of the helium- and hydrogen-burning temperature acts to lock the core boundary into the *image*.

If the core encounters the GSL, the evolution becomes non-quasistatic and the solution path is displaced to larger B-values, so that the boundary of a non-quasistatic core is even less able to reach the *slough*.

8 Anatomy of a giant

Fig. 5 illustrates schematically the characteristic form of giant solution paths on the (B, C)-plane. Fig. 6 illustrates schematically the characteristic run of $\log_{10}[M/M_*]$, η, A, B and C against $\log_{10}[R/R_*]$ for a giant. Fig. 1 indicates how a giant can be divided into six subzones. Each of the following subsections describes one of these subzones. The subsection heading gives the name of the subzone plus the two subscript numbers used to identify its inner and outer boundaries. The centre of the giant has subscript 0.

8.1 INNER CORE (SUBSCRIPTS 0 TO 1)

The inner core has $\bar{\eta} > 5/4$, due to convection or electron degeneracy. This is the hard kernel. The density gradient is small (a general property of the inner regions of hard polytropes). B increases rapidly, whilst C increases slowly. The outer boundary (subscript 1) is where, as we scan outwards, η falls below 5/4 for the last time interior to the shell. Typically $B_1 \simeq 7$, $C_1 \simeq 1$.

8.2 OUTER CORE (SUBSCRIPTS 1 TO 2)

In the outer core the electrons are non-degenerate and the gas is robustly stable against convection, so that $\eta < 5/4$. This is the soft coating. The density gradient steepens dramatically here, as if the surface of the star were approaching. B is roughly constant and C increases rapidly. The outer boundary (subscript 2) is the point where $X = 0.001$. It falls in the *image* with $A_2 \ll 1/4$, $B_2 \simeq 8$, and $C_2 \gtrsim 5$.

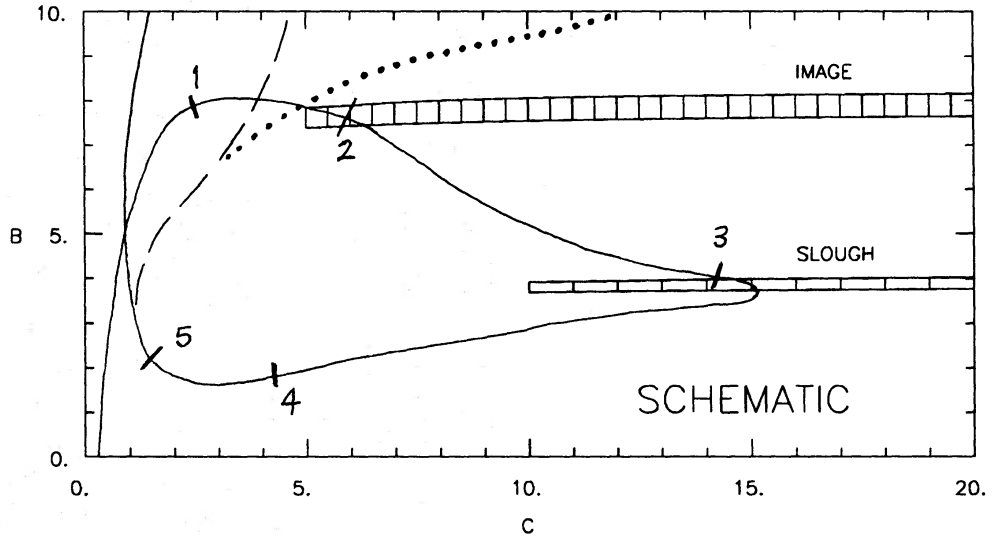


Figure 5. Schematic solution paths on the (B, C)-plane. The full curve is for a giant in which helium is about to ignite and there is a deep outer convection zone. The changes which result when the outer convection zone is either shallow or non-existent are represented by dashes and dots, respectively. [When helium ignites there is a small glitch around $(B, C) \simeq (2, 1)$ (see, for example, Models 30 and 31 on Fig. C1), but we have omitted this variant to avoid confusion.] Numbered tick marks show the location of the subzone boundaries as defined in Fig. 1.

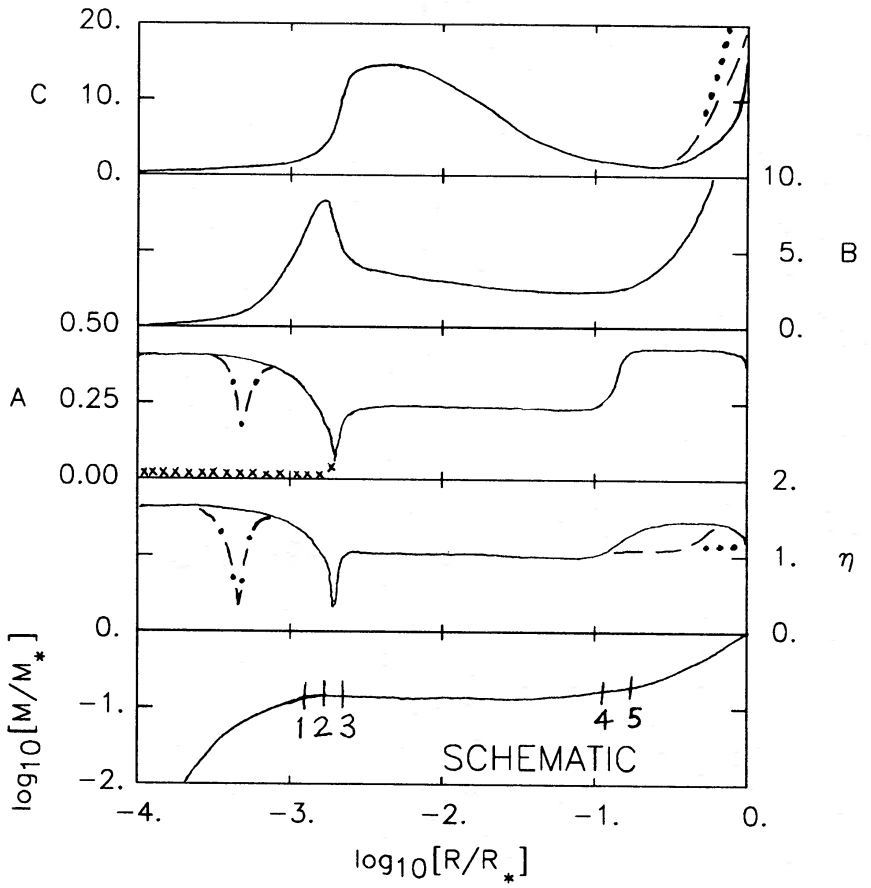


Figure 6. Schematic profiles. The abscissa is $\log_{10}[R/R_*]$. Reading from the top, the ordinate is C, then B, A, η, and lastly $\log_{10}[M/M_*]$. Notation is as in Fig. 5. In addition, the dot-dashes indicate how A and η are suppressed where there is helium burning; and the crosses indicate that A is approximately zero in an isothermal core.

8.3 SHELL (SUBSCRIPTS 2 TO 3)

The hydrogen-burning shell is thin and convectively stable. Consequently the molecular weight gradient $\bar{m}'$ in the shell is steep, which in turn makes the density gradient, $\rho' = \bar{m}' - (1-A)BC$, steep, with $\eta < 1$. However, because the shell is thin (small $\Delta \ln[M]_{\text{shell}}$), the fractional density decrease across the shell, $|\Delta \ln[\rho]_{\text{shell}}| = |\bar{\rho}'| \Delta \ln[M]_{\text{shell}}$, is quite modest.

The main influence of the shell is the changes it induces in (A, B, C), for it is these changes which transport the solution path from the *image* into the *slough*. The almost discontinuous decrease in the molecular weight across the shell (by a factor $f \simeq 2$) causes B to decrease by a factor f, and C to increase by a factor f. The enormous increase in the luminosity across the shell produces a proportional increase in A from its very low value at the core boundary. The rapid increase in A only stops when the solution path becomes trapped in the *slough* with $A \simeq 1/4$.

The outer boundary of the shell (subscript 3) is where A increases through 0.22, i.e. $A_3 = 0.22$, $A_3' > 0$. Typically $B_3 \simeq 4$ and $C_3 \simeq 14$, so that the outer boundary of the shell, i.e. the base of the envelope, falls in the *slough*.

8.4 LOWER ENVELOPE (SUBSCRIPTS 3 TO 4)

This is the critical part of a giant's envelope, where the precipitous density drop occurs, i.e. a very steep density gradient ρ' is maintained over a large mass range $\Delta \ln[M]_{slough}$. By definition,

$0.22 \leq A \leq 0.27$ (equivalently $1.28 \leq \eta \leq 1.37$) so the gas is convectively stable. Typically $2 < B < 4$ and $3 < C < 20$. B and C slowly decrease as the solution path crawls out of the *slough*. The outer boundary (subscript 4) is the point where A rises above 0.27, i.e. $A_4 = 0.27$, $A'_4 > 0$.

8.5 MIDDLE ENVELOPE (SUBSCRIPTS 4 TO 5)

The density gradient flattens out in the middle envelope. The outer boundary (subscript 5) is where the product BC has its last minimum, i.e. where the envelope has put enough distance between itself and the core and accumulated enough mass of its own to be shielded from the gravitational influence of the core.

8.6 UPPER ENVELOPE (SUBSCRIPTS 5 TO *)

The density gradient steepens again in the upper envelope as the star runs out of mass. The outer boundary (subscript *) is the star's photosphere.

8.7 VARIATIONS

We note the following three variations on the above scheme.

(i) If the core is non-isothermal and the inner core is convective, A decreases from $A \simeq 0.4$ at the centre to $A \lesssim 0.2$ at the core boundary. Conversely, if the core is isothermal and the inner core is electron degenerate, it has $A \simeq 0$ throughout.

(ii) If helium is burning, there is a second molecular weight change in the interior of the core, where the composition changes from mainly helium to mainly carbon. This causes a sharp local minimum in both η and A, and a small loop on the (B, C)-plane, but it has no significant influence on the overall structure.

(iii) The upper envelope is not necessarily convective. We include plots for the cases of a deep upper envelope convection zone, a shallow one, and none at all.

9 The hierarchy of cause and effect which makes and keeps giants giant

This section is a commentary on Fig. 7. Fig. 7 demonstrates formally the hierarchy of cause and effect which makes and keeps giants giant, and thereby identifies the four primary causes to which giantness is ultimately attributable. The subsection numbers correspond to the numbers in the righthand margin of Fig. 7.

A note of qualification is in order here. Factors which are equally relevant to the structure of giant and pre-giant stars (e.g. the law of gravitation, local thermodynamic equilibrium, the ideal gas equation of state, and the diffusion approximation for radiative energy transport) are not perceived as primary causes of giantness, although the hierarchy of cause and effect is dependent on their validity.

9.1 THE CRITICAL STRUCTURE OF CORES ABOVE A CERTAIN MASS; THE COMPLEMENTARY ROLES OF CONVECTION AND ELECTRON DEGENERACY

Hydrogen-exhausted cores above a certain mass develop a critical structure, in which the polytropic exponent is high ($\bar{\eta}_i > 5/4$) in the inner core (the hard kernel), and low ($\eta_o < 5/4$) in the outer core (the soft coating). This is critical because all other structures which might deliver the core boundary into the *image* ($A_2 \ll 1/4$, $B_2 \simeq 8$, $C_2 \gtrsim 5$) are gravothermally unstable and

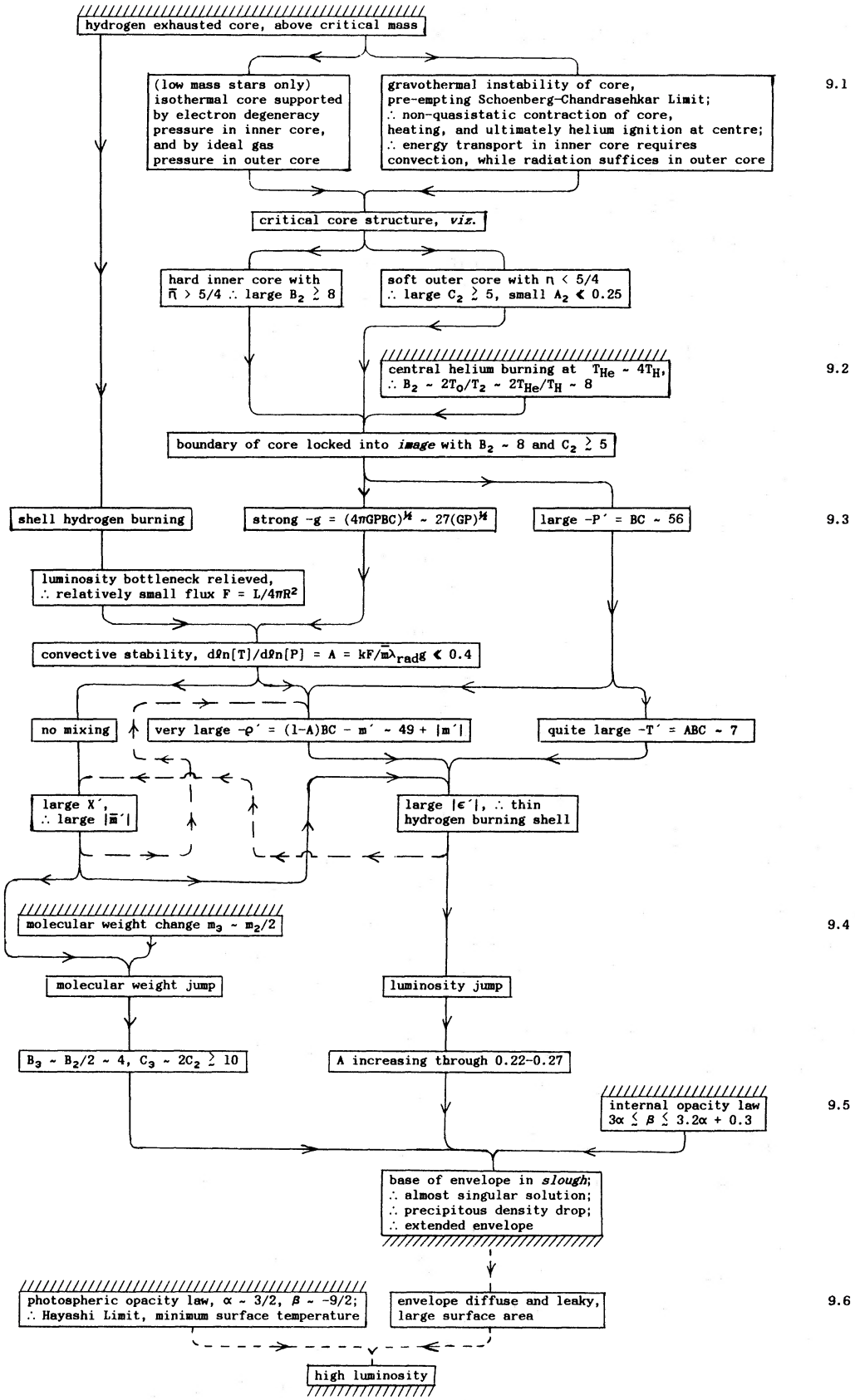


Figure 7. Flow diagram demonstrating the hierarchy of cause and effect which makes giants giant (and luminous). Primary causes are highlighted with hatching above, and final effects with hatching below. There are various levels of the hierarchy and these are distinguished by numbers in the righthand margin, which refer to the subsections in which they are discussed. Positive feedback loops are indicated by long dashes (level 9.3 only). The causes of high luminosity (as distinct from giantness) are connected by short dashes (level 9.6).

cannot therefore be sustained. In low-mass stars the critical structure arises, even while the core is still isothermal, due to the development of electron degeneracy in the inner core, so that $\eta_i \simeq 5/3$ and $\eta_o \simeq 1$. Otherwise the critical structure arises only after the core has become gravothermally unstable, undergone non-quasistatic contraction and heated up, so that the inner part is convective with $\eta_i \simeq 5/3$, and the outer part is radiative with $\eta_o < 5/4$. We stress that the gravothermal instability which drives non-quasistatic contraction and heating of the core is a transient phase; global gravothermal stability returns, at least in a relative sense, once helium ignites.

9.2 LOCKING THE CORE BOUNDARY INTO THE *IMAGE*; THE HYDROGEN- AND HELIUM-BURNING THERMOSTAT

The critical structure is only a necessary condition for the core boundary to be in the *image*. As the core heats up due to non-quasistatic contraction, the core boundary advances to ever increasing values of B_2. It is the ignition of helium at a temperature $T_{He} \simeq 4T_H$ which halts the steady increase of B_2 and holds it near the value $B_2 \simeq 2T_0/T_2 \simeq 2T_{He}/T_H \simeq 8$. This locks the core boundary into the *image*, thereby ensuring that the star remains giant for a while.

9.3 CONVECTIVE STABILITY AND THINNESS OF THE SHELL

Both the convective stability of the shell and its thinness are consequences of the critical core structure, so they are not primary causes. However, since they play key roles in marrying the critical core structure to the *almost-singular* envelope solutions, it is important to understand how they arise.

Convective stability in the shell is promoted by two circumstances. First, the gravitational field is large, typically $|g| = (4\pi GPBC)^{1/2} \simeq 27(GP)^{1/2}$, and so the pressure gradient is also large, typically $|P'| = BC \simeq 56$. Secondly, the main site of luminosity generation has been displaced from the centre of the star, giving a relatively small flux, $F = L/4\pi R^2$, and hence a relatively low temperature gradient, typically $|T'| = ABC \simeq 7$. The upshot is that $d\ln[T]/d\ln[P] \equiv A = kF/\bar{m}\lambda_{rad}|g| \simeq 1/8$, which ensures robust convective stability ($A < 2/5$).

The thinness of the shell is due to several effects and includes two positive feedback loops. Because (as discussed in the preceding paragraph) the pressure gradient is large, the density and/or temperature gradients must also be large. Typically $A \simeq 1/8$, so that $|\rho'| = |\bar{m}'| + (1-A)BC \simeq |\bar{m}'| + 49$ and $|T'| = ABC \simeq 7$. Shell hydrogen burning normally proceeds by the CN-cycle, so at a representative temperature $T \simeq 3 \times 10^7$ K the energy generation rate gradient is $\varepsilon' = \rho' + 15T' + X' \simeq -154 + X'$. This implies that the shell is very sharply bounded on the outer side where $X \simeq 0.7$ (constant) and $\varepsilon' < 0$.

Since convective stability means no mixing, the composition and molecular weight changes associated with hydrogen burning are confined to the shell. Thus, if the shell is thin, the composition gradient is steep (i.e. X' is large), and this makes the shell thin by ensuring that it is sharply bounded on its inner side where $\varepsilon' > 0$: this is the first feedback loop. Furthermore, if the shell is thin, the molecular weight gradient is steep (i.e. $|\bar{m}'|$ is large), and this makes the shell thin by increasing the density gradient $|\rho'| = |\bar{m}'| + (1-A)BC$: this is the second feedback loop.

9.4 JUMP CONDITIONS ACROSS THE SHELL DUE TO THE MOLECULAR WEIGHT AND LUMINOSITY CHANGES

Because the shell is thin and convectively stable, the molecular weight and luminosity gradients in the shell are large and dominate the structure equations, which can therefore be approxi-

mated by jump conditions. The changes in B and C are mainly due to the molecular weight change giving $B_3 \simeq B_2/2$ and $C_3 \simeq 2C_2$. This transports the solution path from the *image* into the *slough* on the (B, C)-plane. The change in A is mainly due to the luminosity change, giving $A_3 \gg A_2$. This brings the solution path up from the *image* at $A_2 \ll 1/4$ to $A_3 = 0.22$ where it becomes stuck in the *slough*. The dominance of the molecular weight and luminosity gradients in the shell structure equations is crucial (and hence the thinness and convective stability of the shell are crucial) because all the other terms in the shell structure equations tend to lead the solution path out of the *slough*.

9.5 THE PRECIPITOUS DENSITY DROP IN THE LOWER ENVELOPE; THE ROLE OF THE OPACITY LAW IN THE *ALMOST-SINGULAR* SOLUTIONS

With the base of the envelope in the *slough*, the envelope is inescapably hooked up to an *almost-singular* solution with a large density drop in the lower envelope. The enormous size of this density drop is due to the opacity law, $\bar{K} = K_0 \rho^\alpha T^{-\beta}/\bar{m}_e$. If α and β were to fall outside the narrow range $3.0\alpha \lesssim \beta \lesssim 3.2\alpha + 0.3$, the *slough* would be smaller so it would be harder to land in (i.e. there would be a narrower range of *almost-singular* solutions so it would be harder to set one up), and/or the *slough* would be much shallower (i.e. the *almost-singular* solutions would on average involve smaller density drops).

9.6 MIDDLE AND UPPER ENVELOPE; HIGH LUMINOSITY

Because the density in the middle and upper envelope is so low, the envelope is both very extended so that the star has a large surface area from which to radiate, and very leaky so that its internal energy diffuses out at a high rate. Because of the photospheric opacity law ($\alpha \simeq 3/2$, $\beta \simeq -9/2$) the surface temperature cannot fall below about 3000 K (the Hayashi Limit). The combination of these factors makes for a very high luminosity. Because the CN-cycle is very temperature sensitive, quite small internal adjustments will normally suffice to maintain nuclear generation of this high luminosity. Thus the high luminosities of giants are a consequence, and not a cause, of their being giant.

10 Conclusions

10.1 PRIMARY CAUSES OF GIANTNESS

We have identified the four primary causes underlying giantness:

[1] 'The opacity law in the lower envelope (Kramers or Thomson or a mixture).' If one adopts an opacity law of the form $\bar{K} = K_0 \rho^\alpha T^{-\beta}/\bar{m}_e$, then for certain α and β the envelope structure equations possess a family of *almost-singular* solutions in which the density falls precipitously through several orders of magnitude and then bottoms out. These are the solutions governing giant envelopes. They are due to a subtle balance between opposing influences on the density gradient. Both the extent of the family of *almost-singular* solutions, and the extent of the density drop which they entail, are maximized by the constraint $3.0\alpha \lesssim \beta \lesssim 3.2\alpha + 0.3$, which is satisfied by both Kramers and Thomson opacity, and by a mixture of the two. However, these *almost-singular* solutions require a very particular (and at first sight rather unlikely) disposition at the base of the envelope. This disposition involves the base of the envelope falling in the *slough* ($A_3 \simeq 1/4$, $B_3 \simeq 4$, $C_3 \gtrsim 11$). The other three primary causes are critical because without them the base of the envelope cannot fall in the *slough*.

[2] 'The abrupt molecular weight change across the shell.' The core boundary cannot reach the *slough*, but it can – given the critical core structure, see [3] – reach the *image* ($A_2 \ll 1/4$, $B_2 \simeq 8$, $C_2 > 5$). The roughly-twofold decrease in molecular weight across the shell is unique in being able to transport the solution path in (B, C)-space from the *image* to the *slough*. (Simultaneously, the large luminosity increase across the shell transports the solution path in A-space from the *image* to the *slough*.) Without the molecular weight decrease across the shell it would be impossible to deposit the base of the envelope in the *slough* and so the *almost-singular* envelope solutions would be inaccessible.

[3] 'Convection (following gravothermal instability), or electron degeneracy, in hydrogen-exhausted cores above a certain mass.' In order for a quasistatic core to have its boundary in the *image*, it must adopt a critical structure which comprises a hard ($\eta > 5/4$) kernel and a soft ($\eta < 5/4$) coating. This critical structure arises naturally in cores above a certain mass as a consequence of convection (following a transient episode of gravothermal instability which normally ends in helium ignition) or electron degeneracy (low-mass stars only). Both effects engender a polytropic exponent $\eta \simeq 5/3$ in the inner core, and η then falls towards unity in the outer core.

One cannot make a precise distinction between giants whose cores have the critical structure due to electron degeneracy and giants whose cores have the critical structure due to convection. For instance, an $M_\odot$ star is already well on the way to gianthood with a partially degenerate essentially quasistatic core, before the rate of contraction becomes sufficiently non-quasistatic to heat up the core and render it convective.

[4] 'Thermostatic control of the core by simultaneous hydrogen and helium burning.' Very approximately, the B-value at the core boundary is given by $B_2 \simeq 2T_0/T_2$, where T_0 is the central temperature and T_2 is the boundary temperature. This means that as long as helium burns in the core ($T_0 \simeq T_{\rm He}$) and hydrogen burns in the shell ($T_2 \simeq T_{\rm H}$), the core boundary has $B_2 \simeq 2T_{\rm He}/T_{\rm H}$. But the core helium-burning temperature is roughly four times the shell hydrogen-burning temperature, so that $B_2 \simeq 8$. Thus the core boundary is locked thermostatically into the *image*.

10.2 CAUTIONS

We stress that the explanation of giantness cannot be made more simple. A complex interplay of many different physical effects is involved, and none of the four primary causes we have identified should be elevated above the others. Previous explanations of giantness are all at best highly incomplete, as we discuss further in Appendix D.

It is also appropriate to note that one cannot construct a realistic model of a giant by simply appending a pure polytropic envelope to a pure polytropic core with a molecular weight discontinuity and a fixed ideal gas temperature at the interface. Such models can be made to mimic post-main sequence evolution in the sense that as the core mass grows, the core shrinks and the envelope expands. However, all such models are unacceptable as representations of real giants on at least two of the following fundamental counts:

(i) They are gravothermally unstable and therefore unsustainable.
(ii) The giant phase is too transient.
(iii) There is no precipitous density drop at the base of the envelope.
(iv) The implied opacity law is unreal.

Since we have explained why giants are giant without explicitly including the physics of luminosity generation, it follows that a giant's high luminosity is more a consequence than a

cause of its structure. A giant is luminous because it is extended and diffuse, being extended it has a large surface area from which to radiate, and being diffuse it is leaky and so has a rapid turnover of thermal internal energy. If the growth of the giant pushes it against the Hayashi Limit, then there is a back reaction on the luminosity through the development of a convection zone in the upper envelope, but this is a secondary effect.

10.3 PREDICTIONS AND SPECULATIONS

We can make a number of testable predictions and speculations. Firm predictions can be stated thus: giants should subscribe broadly to the structure we have delineated in Sections 8 and 9 and Figs 1, 5 and 6; conditions at the base of the envelope should satisfy equations (5.9)–(5.11). These predictions can be tested against existing detailed models of giants; I have only had access to the Iben models tabulated by Novotny (1973).

In addition, existing numerical codes could be used to test how giant evolution and structure responds to (artificial) changes in the basic physics, and hence to test how critical the real physics is. This might appeal to aficionados of the Anthropic Principle, since giants play key roles in the origin of both the carbon vital for life, and the grains vital for planet formation. With the caveat that, given the highly non-linear nature of giant structure and evolution they are tentative speculations rather than firm predictions, we offer the following.

(i) If the opacity law is modified to fall well outside the range identified in inequality (5.7), we expect that on average giants will have a smaller density drop at the base of the envelope; the Giant Branch (hereafter GB) will be less extended on the HR Diagram. If $\beta < 3.0\alpha$, evolution to gianthood should be delayed until the core mass is a larger fraction of the total. If $\beta > 3.2\alpha + 0.3$, evolution to gianthood should be advanced.

(ii) If the molecular weight change $f \simeq 2$ is reduced significantly, a star should swell up more slowly as it leaves the main sequence; the GB should be less extended and the star should ascend to the tip of the GB and then start losing mass (or possibly come some way back down the GB) before helium ignites. If f is increased, helium should ignite quite low down on the GB.

(iii) If the helium-burning temperature $T_{\rm He} \simeq 100\text{–}160 \times 10^6$ K is increased significantly, the star should ascend to the tip of the GB and then start losing mass (or possibly come some way back down the GB) before helium ignites. If $T_{\rm He}$ is reduced, helium should ignite before the star has ascended very far up the GB.

(iv) The critical core structure becomes even more critical if $\beta < 3.0\alpha$ and/or $T_{\rm He}$ is reduced and/or f is reduced.

(v) To make giants even larger one should leave the opacity law alone and increase f and $T_{\rm He}$, roughly in proportion.

In conclusion, we note that the overall extent of a giant is very sensitive to conditions in the *slough* and hence very dependent on local functions of state and their derivatives near the base of the envelope. Once the base of the envelope starts to fall in the *slough*, the envelope becomes a long tapering tail, and the core becomes a capricious dog wagging this tail. A small increase in B_3 at the base of the envelope can mean that the solution path suddenly finds itself above the $C' = 0$ null-surface so that it has to make an excursion to very large C-values and then loop back in order to escape from the *slough* (there is some evidence that this has just happened in Model 29; see Fig. C1). This will result in a rapid increase in the extent of the envelope. If B_3 increases still further, the solution path may run off to such large C-values that the star is driven to throw off some of its outer envelope. It is therefore of great importance that numerical codes should be as accurate as possible in the convectively stable regions at the base of a giant's envelope.

Acknowledgments

I thank Mike Disney for bringing this problem to my attention, and Mike Edmunds for directing me to several important references.

References

Applegate, J. H., 1988. *Astrophys. J.*, **329,** 803.
Chandrasekhar, S., 1939. *An Introduction to the Study of Stellar Structure,* Dover, New York.
Eggleton, P. P. & Faulkner, J., 1981. In: *Physical Processes in Red Giants,* p. 179, eds Iben, I., Jnr. & Renzini, A., Reidel, Dordrecht.
Höppner, W. & Weigert, A., 1973. *Astr. Astrophys.*, **25,** 99.
Iben, I., Jnr., 1965. *Astrophys. J.*, **141,** 993.
Iben, I., Jnr., 1966. *Astrophys. J.*, **143,** 483.
Iben, I., Jnr., 1967. *Astrophys. J.*, **147,** 624.
Iben, I., Jnr. & Renzini, A., 1984. *Physics Reports,* **105,** 329.
Novotny, E., 1973. *An Introduction to Stellar Atmospheres and Interiors,* Oxford.
Schönberg, M. & Chandrasekhar, S., 1942. *Astrophys. J.*, **96,** 161.
Schwarzschild, M., 1958. *Structure and Evolution of the Stars,* Dover, New York.
Weiss, A., 1983. *Astr. Astrophys.*, **127,** 411.
Yahil, A. and van den Horn, L., 1985. *Astrophys. J.*, **296,** 554.

Appendix A: The *slough*

The *slough* is the region around an equilibrium point where solution paths advance slowly with increasing path parameter. Because in nature opacity does not obey a strict power law, the equilibrium points of equation (5.5) are blurred, and it is more appropriate to discuss solution paths with reference to the *slough*. We emphasize that the *slough* does not have a precise boundary. The solution path can be more or less deeply ensnared in the *slough*, depending on its position in (A, B, C)-space.

Since we are interested in *sloughs* at large C, we delineate the *slough* with contours of constant

$$C[A'^2+B'^2+C'^2]^{-1/2}=D, \tag{A1}$$

so that contours of larger D correspond to greater depth into the *slough*. Fig. A1 illustrates contours for $D=2$, 5, 11, and for various opacity laws. We see that for both Kramers and Thomson opacity, the *slough* has almost exactly the same position, extent and shape, being deepest for $C\gtrsim 11$ and elongated parallel to the C-axis around the line $A\simeq 1/4$, $B\simeq 4$.

Inspection of the null-surfaces,

$$A'=(\gamma A-\delta)BC-1=0 \tag{A2}$$

$$B'=ABC-C+1=0, \tag{A3}$$

$$C'=(1-A)BC-3C+1=0, \tag{A4}$$

reveals why this is so. For $(\gamma, \delta)=(17/2, 2;$ Kramers) or $(4, 1;$ Thomson), and at large C-values, the intersection of any pair of null-surfaces lies very close to the line $A=1/4$, $B=4$. Consequently the two *sloughs* are elongated along this line and nearly co-extensive. This is important for the following reason.

The base of a giant envelope is often close to the locus where Kramers and Thomson make equal contributions to the opacity. This locus has the slope $d\ln[T]/d\ln[\rho]\simeq 2/7\simeq 0.286$. The *slough* and its associated *almost-singular* solutions will exist only for a mixed opacity law (i.e.

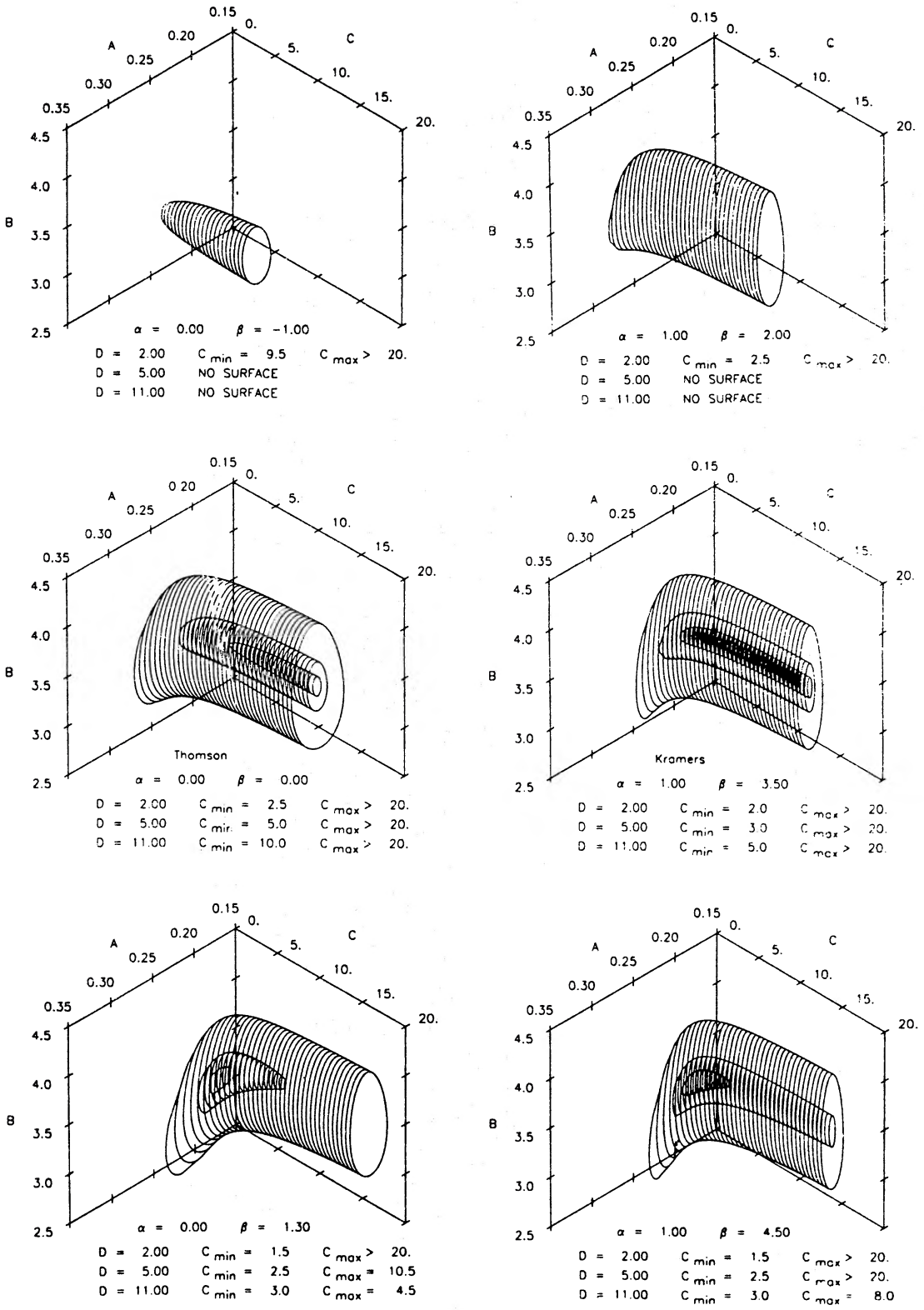


Figure A1. Hidden-line illustration of the *sloughs* for different opacity laws. Each plot is labelled with the values of α and β. The lefthand plots have $\alpha = 0$ and, reading downwards, $\beta = -1$, $\beta = 0$ (Thomson), and $\beta = 1.3$. The righthand plots have $\alpha = 1$ and, again reading downwards, $\beta = 2$, $\beta = 3.5$ (Kramers), and $\beta = 4.5$. The surfaces are defined by equation (A1) with $D = 2$, 5 and 11. The lines delineating the surfaces are lines of constant integral C, plus a silhouette line. Surfaces which extend beyond $C = 20$ are terminated at $C = 20$. For each D-value there is a tabulation below the plot giving the extent of the surface in the C-direction.

Kramers and Thomson in comparable amounts) if the *sloughs* for the two individual opacity laws are extensively overlapping and the proportions in the mixture vary slowly with changing path parameter. Since $d\ln[T]/d\ln[\rho]_{\rm eq} = A_{\rm eq}/(1-A_{\rm eq}) \simeq 0.313$ (Kramers) and 0.333 (Thomson), both conditions are satisfied. The balance tends to switch from Thomson to Kramers, moving radially outwards through the lower envelope.

The topology of solution paths in the *slough* is established by making equations (5.1) to (5.3) linear in the vicinity of the equilibrium point. The first step is to construct the 3×3 matrix

$$\begin{pmatrix} \partial A'/\partial A & \partial A'/\partial B & \partial A'/\partial C \\ \partial B'/\partial A & \partial B'/\partial B & \partial B'/\partial C \\ \partial C'/\partial A & \partial C'/\partial B & \partial C'/\partial C \end{pmatrix},$$

where the derivatives are all evaluated at the equilibrium point. The eigenvalues of this matrix are plotted against α (assuming $\beta = 3.5\alpha$, i.e. along the line connecting Kramers and Thomson opacity) on Fig. A2. For $\alpha < 0.653$ there are three positive real eigenvalues, so the equilibrium

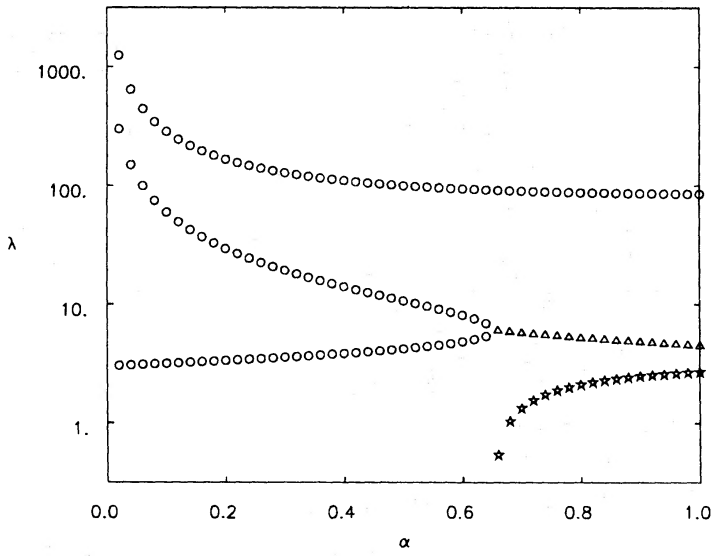


Figure A2. Eigenvalues λ of the linearized equations of stellar structure in the vicinity of the equilibrium point, as a function of α, assuming $\beta = 3.5\alpha$ (i.e. along the line connecting Thomson opacity on the left to Kramers opacity on the right). The circles represent positive real eigenvalues. The triangles and asterisks represent (respectively) the positive real and imaginary parts of complex conjugate eigenvalue pairs.

point is an unstable node. For $\alpha > 0.653$ there is one large positive real eigenvalue and a complex conjugate pair with a real part which is positive and at least a factor of two larger than the imaginary part. Thus these equilibrium points are unstable helices in which the solution path only makes multiple turns around the equilibrium point in its immediate vicinity. Since this region is blurred, the topology of solution paths is actually more like an inflected node.

The hatched region labelled 'SLOUGH' on the (B, C)-plane in Figs 2–5, B2, and C1 is the projection of the common portion of the $D = 11$ *sloughs* for Kramers and Thomson opacity. It therefore represents the deepest recesses of the *slough*. The hatched region labelled 'IMAGE'is obtained by translating 'SLOUGH' according to $B \rightarrow 2B$ and $C \rightarrow C/2$.

Appendix B: Isothermal quasistatic cores and evolution to the SCL

B1 CORE STRUCTURE

The early stages of post-main sequence evolution are essentially quasistatic, with the core approximately isothermal at the hydrogen-burning temperature T_H. As the core grows in mass, the central density increases and the electrons may eventually become degenerate. To investigate how this affects the structure of the core, we adopt the expression for the pressure of a completely degenerate non-relativistic electron gas, $(3/\pi)^{2/3}(h^2/20m_e)(\rho/\bar{m}_e)^{5/3}$, and simply add this to the ideal gas pressure, $\rho kT_H/\bar{m}$. The two are equal at a characteristic density ρ_c, which enables us to introduce dimensionless variables:

$$R = xR_c, \qquad R_c = (kT_H/4\pi G\rho_c\bar{m})^{1/2}, \tag{B1}$$

$$\rho = y\rho_c, \qquad \rho_c = (\pi/3)(20m_e kT_H/\bar{m}h^2)^{3/2}\bar{m}_e^{5/2}, \tag{B2}$$

$$P = [y + y^{5/3}]\,P_c, \qquad P_c = \rho_c kT_H/\bar{m}, \tag{B3}$$

$$M = zM_c, \qquad M_c = 4\pi R_c^3\rho_c. \tag{B4}$$

The above prescription for the equation of state, although inexact, is adequate for our purposes because it possesses the essential feature that, as the density increases above ρ_c, the effective polytropic exponent changes from $\eta \simeq 1$ to $\eta \simeq 5/3$.

The structure equations,

$$\frac{dy}{dx} = -\frac{yz}{x^2(1 + 5y^{2/3}/3)}, \qquad y(x=0) = y_0, \tag{B5}$$

$$\frac{dz}{dx} = x^2 y, \qquad z(x=0) = 0, \tag{B6}$$

have been integrated numerically for different values of y_0. Some solutions are shown in Fig. 4, where $B = z/x(1 + y^{2/3})$, $C = z/yx^3$, and the line of asterisks marks the GSL.

Fig. B1 shows the variation of z_{GSL} with y_0, where z_{GSL} is the dimensionless mass at the GSL. The asymptote $z_{GSL} = 15.7y_0^{-1/2}$ $(y_0 \ll 1)$, corresponds to the standard isothermal ideal-gas sphere truncated at the GSL (i.e. in the notation of Chandrasekhar (1939), at $\xi \simeq 6.45$ where $\xi^4(\psi')^2 e^{-\psi}$ has its first maximum). The other asymptote $z_{GSL} = 10.7y_0^{1/2}$ $(y_0 \gg 1)$ corresponds to the $\eta = 5/3$ polytrope untruncated (i.e. again in the notation of Chandrasekhar (1939), extending to $\xi = 3.65$ where $\theta = 0$).

B2 EARLY POST-MAIN SEQUENCE EVOLUTION

Consider how conditions at the core boundary change as a star evolves off the main sequence. Since $\bar{m}$ and $\bar{m}_e$ are constant in the core, and T_H is also roughly constant, we can treat (ρ_c, R_c, P_c, M_c) as roughly constant [see equations (B1) to (B4)]. Thus, as the core mass ($M_2 = z_2 M_c$) and the central density ($\rho_0 = y_0\rho_c$) increase, the point (y_0, z_2) moves upwards and to the right on the (y, z)-plane, until eventually – at least for stars with total mass $M_* \gtrsim M_\odot$ – it encounters the GSL (see Fig. B1).

We wish to establish that the GSL is always encountered before the SCL, and that the rate of release of gravitational potential energy in the core is increasing so rapidly at the GSL that the core evolution can no longer be treated as quasistatic. Our treatment will be much cruder than

that of Schönberg & Chandrasekhar (1942), but it will include both the possibility of electron degeneracy in the core (thereby enabling a distinction to be made between stars of different mass), and the possibility of a radiative or a convective envelope.

Consider first an uncontained sphere of gas in hydrostatic balance. The Virial theorem requires

$$-\Omega = 3\int P\,dV = 3\int (P/\rho)\,dM = 3M_*(\overline{P/\rho}) \tag{B7}$$

where Ω is the self-gravitational potential energy, and $(\overline{P/\rho})$ is the mass-weighted mean-squared isothermal sound speed. Putting $-\Omega = \omega GM_*^2/R_*$, equation (B7) reduces to

$$\bar{B} \equiv \frac{GM_*}{R_*(P/\rho)} = \frac{3}{\omega}, \tag{B8}$$

where $\bar{B}$ is an average B-value for the whole sphere. For a polytropic gas with $\eta = 4/3$, $\omega_{4/3} = 3/2$ and $\bar{B}_{4/3} = 2$; whilst for $\eta = 5/3$, $\omega_{5/3} = 6/7$ and $\bar{B} = 7/2$.

Suppose that the envelope of the star is a pure polytrope with polytropic constant K_e and polytropic exponent η_e, so that the mean-squared isothermal sound speed and density in the envelope are related to those at the base (subscript 3) thus:

$$(\overline{P/\rho})_e \bar{\rho}_e^{(1-\eta_e)} = K_e = (P/\rho)_3 \rho_3^{(1-\eta_e)}. \tag{B9}$$

We can substitute for the mean-squared isothermal sound speed in the envelope $(\overline{P/\rho})_e = GM_*/R_*\bar{B}_e$, and for the mean density in the envelope $\bar{\rho}_e = 3(M_* - M_2)/4\pi R_*^3$; we neglect the volume of the core since $R_2 \ll R_*$. The values of $\bar{B}_\eta$ obtained in the preceding paragraph are not applicable here because we are only assuming that the envelope (not the whole star) is a pure polytrope, but they are representative of the trend in going from $\eta = 4/3$ to $\eta = 5/3$. The envelope on its own will have a larger value of $\bar{B}$, and so we put $\bar{B}_e = 2\bar{B}_{\eta_e}$, i.e. $\bar{B}_e = 4$ for $\eta_e = 4/3$, and $\bar{B}_e = 7$ for $\eta_e = 5/3$. This prescription for $\bar{B}_e$ is guided by the fact that it brings $\bar{B}_e$ into line with the values obtaining in the envelopes of detailed models.

If the shell is vanishingly thin, we can substitute for the squared isothermal sound speed and density at the base of the envelope in terms of the same quantities at the core boundary: $(P/\rho)_3 = f(P/\rho)_2 = fGM_2/R_2B_2$ and $\rho_3 = \rho_2/f = M_2/4\pi R_2^3 c_2 f$. $f \simeq 2$ is the factor by which the molecular weight increases in crossing the shell from the base of the envelope to the boundary of the core.

With the above substitutions, equation (B9) reduces to

$$\left[\frac{R_*}{R_2}\right]^{(3\eta_e-4)} = \left[\frac{\bar{B}_e}{B_2}\right]\left[\frac{M_2}{M_*}\right]\left[\frac{M_2}{3C_2(M_*-M_2)}\right]^{(1-\eta_e)} f^{\eta_e} \tag{B10}$$

Finally, we express the pressure at the base of the envelope, $P_3 = P_2 = GM_2^2/4\pi R_2^4 B_2 C_2$, in terms of the weight of the envelope, $W \simeq G(M_* - M_2)\,M_2/(R_*/2)^2$:

$$P_3 = \frac{GM_2^2}{4\pi R_2^4 B_2 C_2} = \frac{W}{4\pi R_2^2} \simeq \frac{G(M_*-M_2)\,M_2}{\pi R_*^2 R_2^2},$$

$$\therefore \left[\frac{R_*}{R_2}\right]^2 = \frac{4(M_*-M_2)\,B_2C_2}{M_2}. \tag{B11}$$

Eliminating $[R_*/R_2]$ between equations (B10) and (B11), and substituting for $(\eta_e, \bar{B}_e, f)$, gives

$$\left[\frac{M_2}{M_*}\right]\left[\frac{M_*-M_2}{M_*}\right]^{1/2} \simeq 0.21\left[\frac{B_2}{4}\right]^{3/2}[3C_2]^{-1/2}, \quad \text{for } \eta_e = 4/3 \tag{B12}$$

and

$$\left[\frac{M_2}{M_*}\right]\left[\frac{M_*-M_2}{M_*}\right]^{1/5} \simeq 0.80\left[\frac{B_2}{7}\right]^{9/5}[3C_2]^{-1/5}, \quad \text{for } \eta_e = 5/3. \tag{B13}$$

All other things $(M_2, R_2, \rho_2, P_2, T_2)$ being equal, a radiative envelope with $\eta_e \simeq 4/3$ is more extended (than a convective envelope with $\eta_e \simeq 5/3$) and therefore more massive so that it has the same weight. For instance, a non-degenerate isothermal core has $B_2 \lesssim 2.43$ and $C_2 \lesssim 0.822$. These limits give $(M_2/M_*) \lesssim 0.066$ for $\eta_e = 4/3$, and $(M_2/M_*) \lesssim 0.103$ for $\eta_e = 5/3$. The close agreement with the result of Schönberg & Chandrasekhar (1942) for the case of a convective envelope on a non-degenerate isothermal core is somewhat fortuitous. However, our main concern here is not with absolute values but with the overall evolutionary pattern as the GSL and the SCL are approached. For this purpose the analysis developed above is adequate.

B3 QUASISTATIC EVOLUTIONARY SEQUENCES

For any y_0, we obtain the run of (z, B, C) against x by integrating equations (B5) and (B6). Any point on this integration path up to the GSL is a prospective core boundary, i.e. $x \to x_2$, $z \to z_2$, $B \to B_2$, $C \to C_2$. From equation (B12) or (B13) we obtain corresponding values of (M_2/M_*); and from equation (B11) corresponding values of (R_*/R_2). Finally, given T_H, $\bar{m}$, and $\bar{m}_e$, equations (B1)–(B4) enable us to compute $R_2 = x_2 R_c$, $M_2 = z_2 M_c$, and hence R_* and M_*. By this means we can isolate solutions corresponding to a particular value of M_*.

We consider stars with $M_* = M_\odot$ and $M_* = 5M_\odot$, for which we adopt $T_H = 2.3 \pm 0.2 \times 10^7$ K and $3.3 \pm 0.2 \times 10^7$ K, respectively. We put $\bar{m} \simeq 1.34 m_P \simeq 2.22 \times 10^{-24}$ gm and $\bar{m}_e \simeq 2.00 m_P \simeq 3.33 \times 10^{-24}$ gm, corresponding to the composition $X = 0$ and $Z = 0.02$.

By treating many different values of y_0, we map out evolutionary paths corresponding to the two different total masses ($M_\odot$ and $5M_\odot$), the two different envelope polytropic exponents (4/3 and 5/3), and the two extremes of T_H (i.e. 2.1×10^7 K and 2.5×10^7 K for $M_\odot$; 3.1×10^7 K and 3.5×10^7 K for $5M_\odot$). Without treating in detail the generation and transport of luminosity we cannot distinguish between a radiative envelope with $\eta_e \simeq 4/3$ and a convective envelope with $\eta_e \simeq 5/3$, but the real situation is unlikely to lie outside these limits.

The resulting evolutionary paths are plotted on Figs B1–B3. Fig. B1 gives the run of the dimensionless core mass z_2 against the dimensionless central density y_0. Fig. B3 gives the run of B_2 against C_2. Fig. B2 gives the run of $(\bar{m}/kT_H)(-d\Omega_2/dM_2)$ against core mass M_2, where $(-d\Omega_2/dM_2)$ is the rate of release of gravitational energy in the core, per unit increase in the mass of the core, as defined in equation (7.1), and the coefficient is a convenient way of rendering it dimensionless. The paths are terminated at the SCL, and the GSL is marked by a line of asterisks. From these figures we draw the following conclusions.

Consider Fig. B1. The quasistatic post-main sequence evolution of an $M_\odot$ star is influenced by electron degeneracy; the quasistatic post-main sequence evolution of a $5M_\odot$ star is not. As a consequence of electron degeneracy, a low-mass quasistatic post-main sequence star with a radiative envelope need not encounter either the GSL or the SCL; here 'low mass' appears to mean $M_* \lesssim M_\odot$, but in view of the crudity of our analysis and the conservativeness of our gravothermal instability condition, this is very approximate.

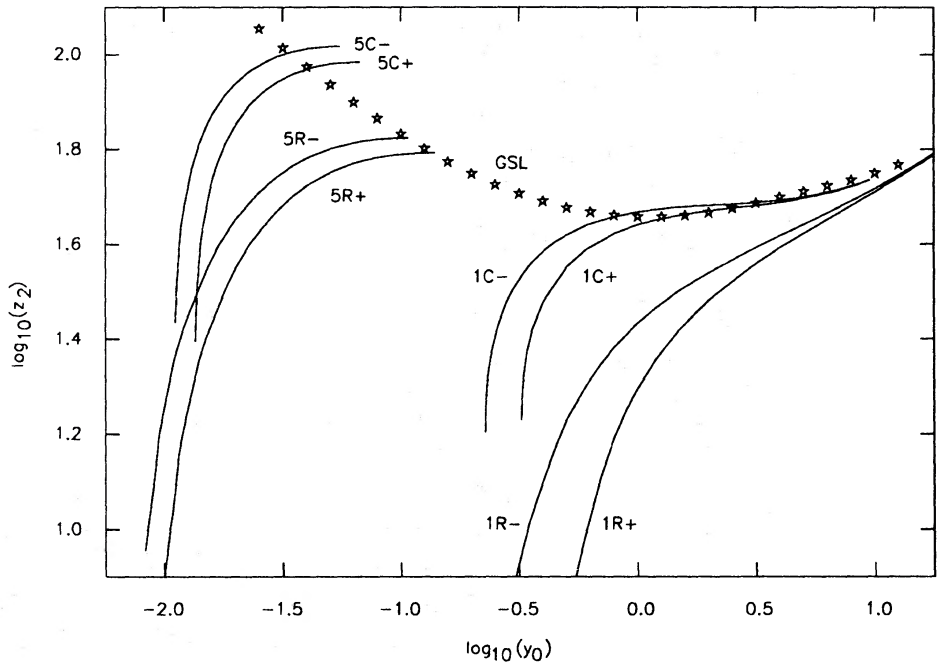


Figure B1. The (z_2, y_0)-plane, where z_2 is the dimensionless core mass and y_0 is the dimensionless central density. The string of asterisks shows the GSL for isothermal cores. The full curves are evolutionary tracks for giants with isothermal cores, showing how the central density increases with increasing core mass and where the GSL is encountered. The tracks are terminated at the SCL. Tracks for giants with $M_* = M_\odot$ and $5M_\odot$ are labelled '1' and '5', respectively. Those with convective envelopes are labelled 'C', and those with radiative envelopes 'R'. Those with a high estimated hydrogen-burning temperature are labelled '+', and those with a low one '−'. Both axes are logarithmic.

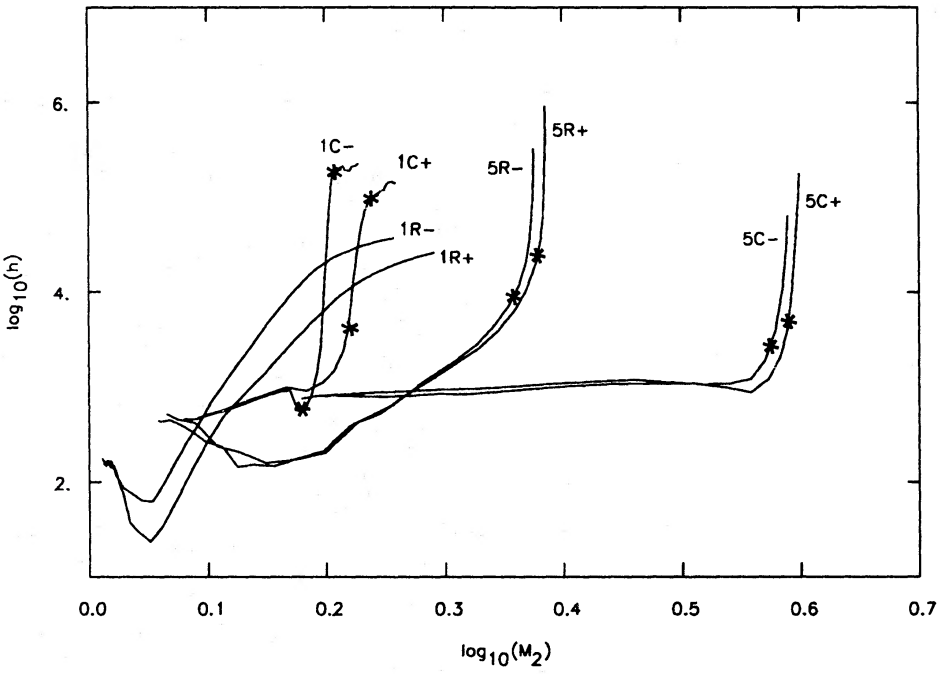


Figure B2. Evolutionary tracks on the (h, M_2)-plane, where h is the dimensionless rate of release of gravitational potential energy in the core per unit increase in core mass, and M_2 is the core mass. Notation is as in Fig. B1. The h-axis is logarithmic.

All quasistatically evolving stars reaching the SCL have already passed the GSL (although for $5M_\odot$ stars with radiative envelopes this cannot be resolved on Fig. B1). Obviously any tightening of the gravothermal instability condition would strengthen this conclusion.

Consider Fig. B2. As a quasistatically evolving star passes the GSL, the rate of release of gravitational potential energy increases by several orders of magnitude. Consequently, evolution beyond the GSL cannot be quasistatic. Note that only the $5M_\odot$ stars pass the GSL once and for all. Quasistatically evolving $M_\odot$ stars with convective envelopes pass into gravothermally unstable states and then out again; only in the gravothermally unstable states is the rate of release of gravitational energy a rapidly rising function of the core mass. Quasistatically evolving $M_\odot$ stars with radiative envelopes do not encounter the GSL.

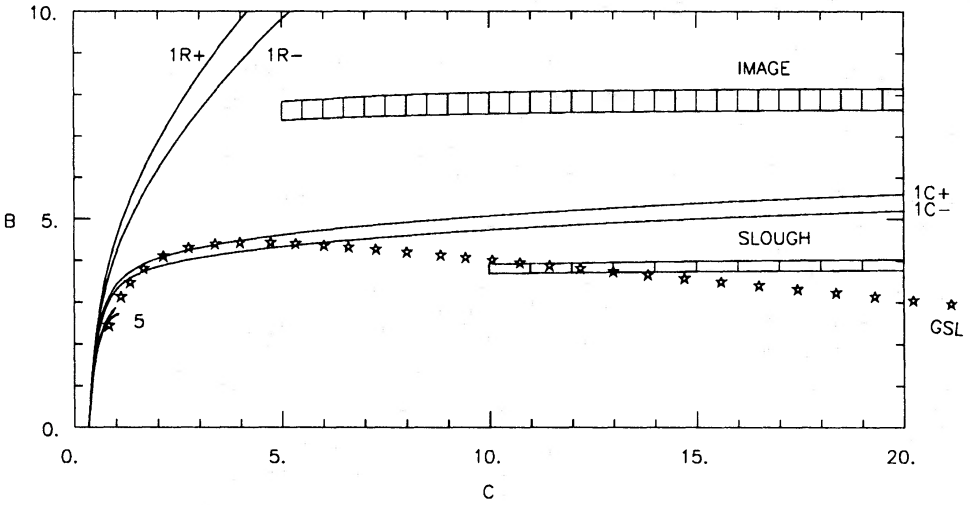


Figure B3. The GSL and evolutionary tracks on the (B, C)-plane for giants with isothermal cores. Notation is as in Fig. B1. Note that for $5M_\odot$ giants the different tracks cannot be resolved because they are all effectively non-degenerate.

Consider Fig. B3. A quasistatically evolving $5M_\odot$ star cannot have its core boundary anywhere near the *slough* or its *image* (even if the evolution is continued beyond the GSL right up to the SCL).

A quasistatically evolving $M_\odot$ star with a predominantly convective envelope might have its core boundary close to the *slough*. However, long before the star reaches this stage it passes the GSL; the core will then start to contract non-quasistatically and heat up, so that the solution path is displaced away from the *slough* towards larger B-values. In other words, although quasistatic solutions exist for which the core boundary is close to the *slough* (and therefore a giant envelope configuration could be supported without a molecular weight change), these solutions do not arise in the natural course of stellar evolution.

A quasistatically evolving $M_\odot$ star with a predominantly radiative envelope might have its boundary near the *image*. However, since B_2 and C_2 are steadily increasing functions of M_2, this can only be a transient phase. Once the core boundary falls well above the *image* on the (B, C)-plane, the base of the envelope falls well above the *slough*. We speculate in the main text that under these circumstances the solution path will extend to such large C-values that the star loses mass.

In summary, when evolutionary considerations are taken into account (in particular gravothermal instability), the boundary of an isothermal quasistatic core (a) cannot fall in the *slough*, (b) can only fall in the *image* temporarily and for low-mass stars.

Appendix C: Iben's detailed models

In Figs C1 and C2 we plot detailed stellar models due to Iben (1965, 1966, 1967), as tabulated in Novotny (1973). Each model is referred to by a number N, where 7-N is the designation of the corresponding table in Novotny. Fig. C1 gives solution paths on the (B, C)-plane. We present a chronological sequence of Iben's models for $M_\odot$ (Models 12–18) and $5M_\odot$ (Models 23–31) in order to emphasize the differences between giants and less-evolved stars. Fig. C2 gives the run of $\log_{10}[M/M_*]$, η, A, B, and C against $\log_{10}[R/R_*]$, for the giant models only (Models 18 and 28 –31). Global parameters of the models are assembled in Table C. Further details of the models, including their evolutionary status, can be obtained from Novotny and the original Iben papers.

Figs C1 and C2 must be compared with Figs 5 and 6, respectively. On the giant models note the positions of the tick marks (1–5), which indicate the locations of the subzone boundaries according to the objective criteria given in Fig. 1. Also keep in mind the three variations listed in Section 8.7, namely:

(i) The core always has a hard kernel with $\eta \simeq 5/3$, but this may be due to convection, in which case $A \simeq 2/5$ (e.g. Models 28–31), or electron degeneracy in an isothermal core in which case $A \simeq 0$ (e.g. Model 18).

(ii) If helium is burning, the inner core contains a molecular weight change which is manifest as a zig-zag on the (B, C)-plane near $(B, C) \simeq (2, 1)$, and as a sharp local minimum in the η- and A-profiles. This has no effect on the overall structure of the star. It is noteworthy because it demonstrates that a molecular weight change on its own is not necessarily effective in producing a large local density drop.

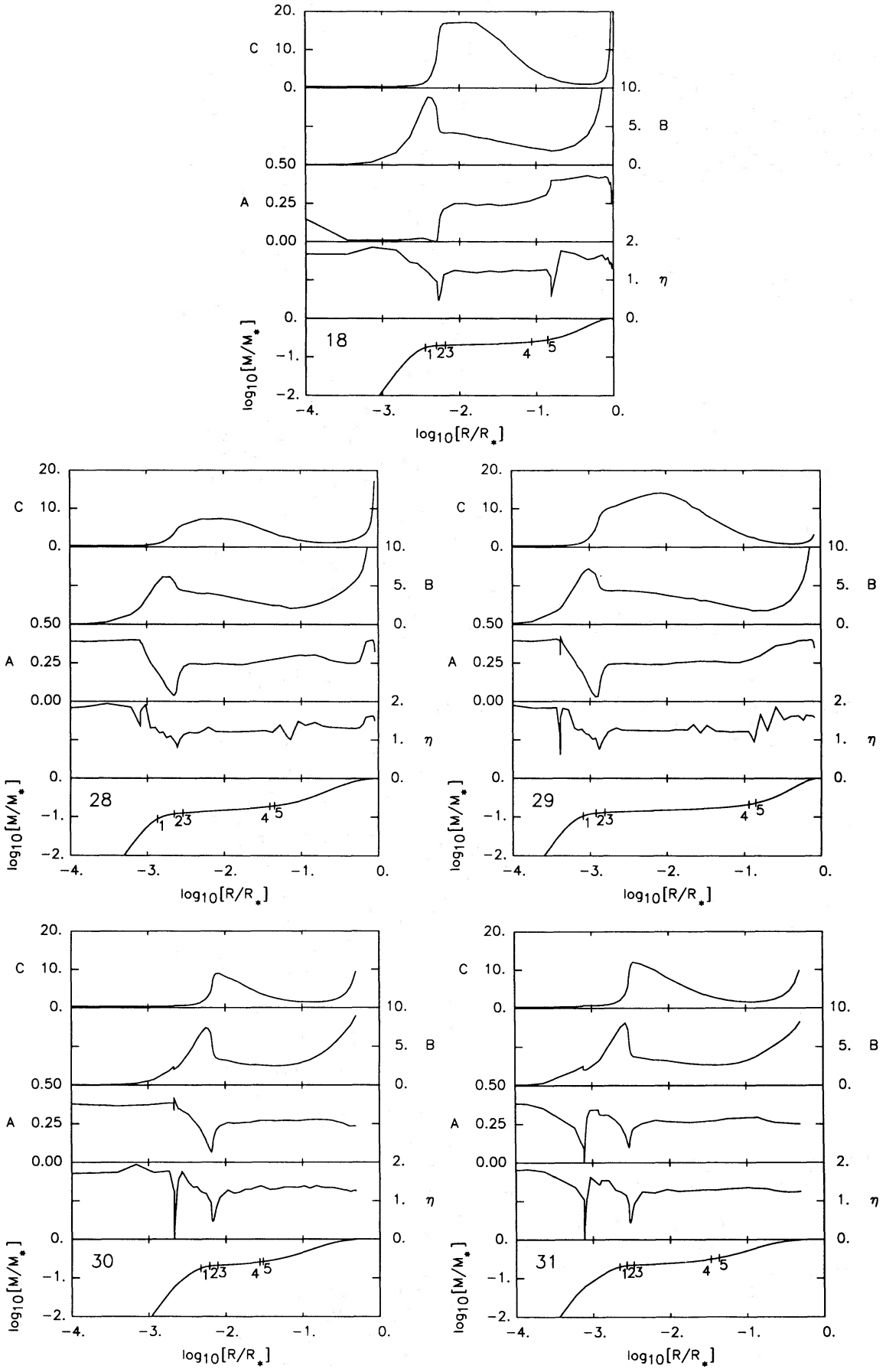
(iii) The giant envelope may have a deep outer convection zone (e.g. Models 18 and 29), a shallow one (e.g. Model 28), or none at all (e.g. Models 30 and 31), as revealed by the η- and A-profiles.

Comparing Figs 5 and C1, we see that the critical part of the solution path, i.e. the loop taking in very large C-values, is common to all giant models, and unaffected by the variations listed above. In all cases the molecular weight change across the shell (between tick marks 2 and 3) is instrumental in transporting the solution path to large C-values and thereby depositing the base of the envelope in the *slough*.

Comparing Figs 6 and C2, we again see that the following features in the profiles are common to all giant models: the core has a hard-kernel/soft-coating structure ($\bar{\eta} > 5/4$ between the centre and tick mark 1, $\eta < 5/4$ between tick marks 1 and 2). B has a maximum in the soft coating (between tick marks 1 and 2). η and A have minima in the shell (between tick marks 2 and 3; this is the Eggleton-Faulkner effect). C has a maximum $C_{\max} \gtrsim 11$, and $A \simeq 1/4$, $B \simeq 4$ at the base of the envelope (near tick mark 3). The lower envelope (between tick marks 3 and 4) is bogged down in the *slough* and makes the dominant contribution to stretching the star (as evidenced by a very small $d\ln[M]/d\ln[R]$). B has a minimum in the middle envelope (between tick marks 4 and 5). C has a minimum in the upper envelope (beyond tick mark 5). A is much better behaved than η.

From the entries in Table C, we see that for giants both the overall extent of the star R_* and the stretching parameter $\Sigma \equiv \log_{10}[\rho_0/\bar{\rho}]$ (where ρ_0 is the central density and $\bar{\rho} = 3M_*/4\pi R_*^3$ is

Figure C2. Profiles for some detailed models due to Iben. The abscissa is $\log_{10}[R/R_*]$ and, reading from the top, the ordinate is C, then B, A, η, and lastly $\log_{10}[M/M_*]$. Each plot is labelled with the model number in the bottom lefthand corner; only giant models are represented in this figure. Numbered tick marks indicate the subzone boundaries.



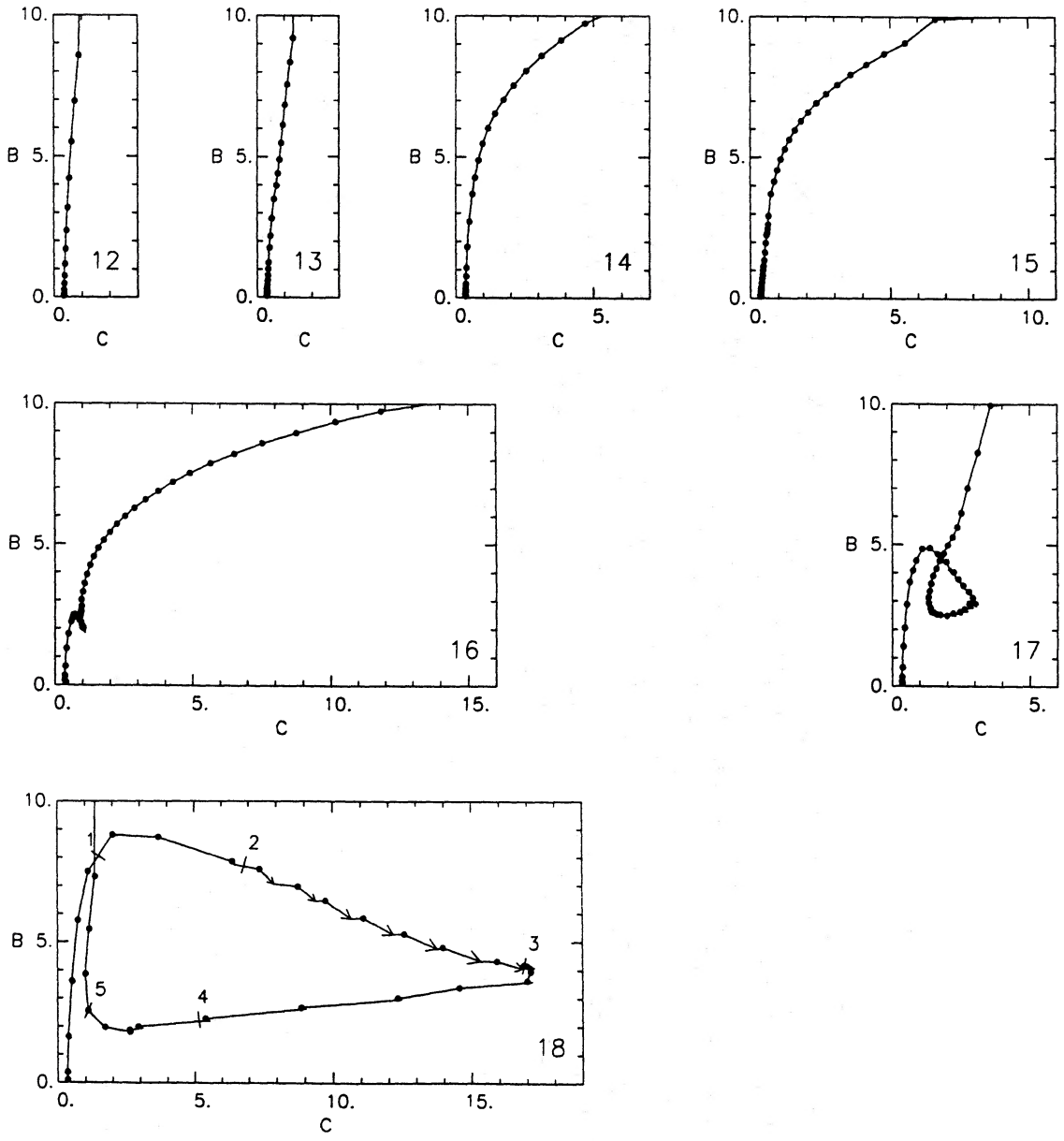


Figure C1. Solution paths on the (B, C)-plane for some detailed models due to Iben. Each plot is labelled with the model number in the bottom righthand corner. Models 12–18 (on the lefthand side) are for $M_* = M_\odot$ and Models 23–31 (righthand side) are for $M_* = 5M_\odot$. Black dots represent the tabulated points. For the true giants (Models 18 and 28–31), numbered tick marks indicate the subzone boundaries as defined in Fig. 1. Arrows demonstrate what part of the displacement between successive tabulated points is attributable to the changing molecular weight. These arrows are concentrated between tick marks 2 and 3, and are most easily resolved for Models 18 and 31.

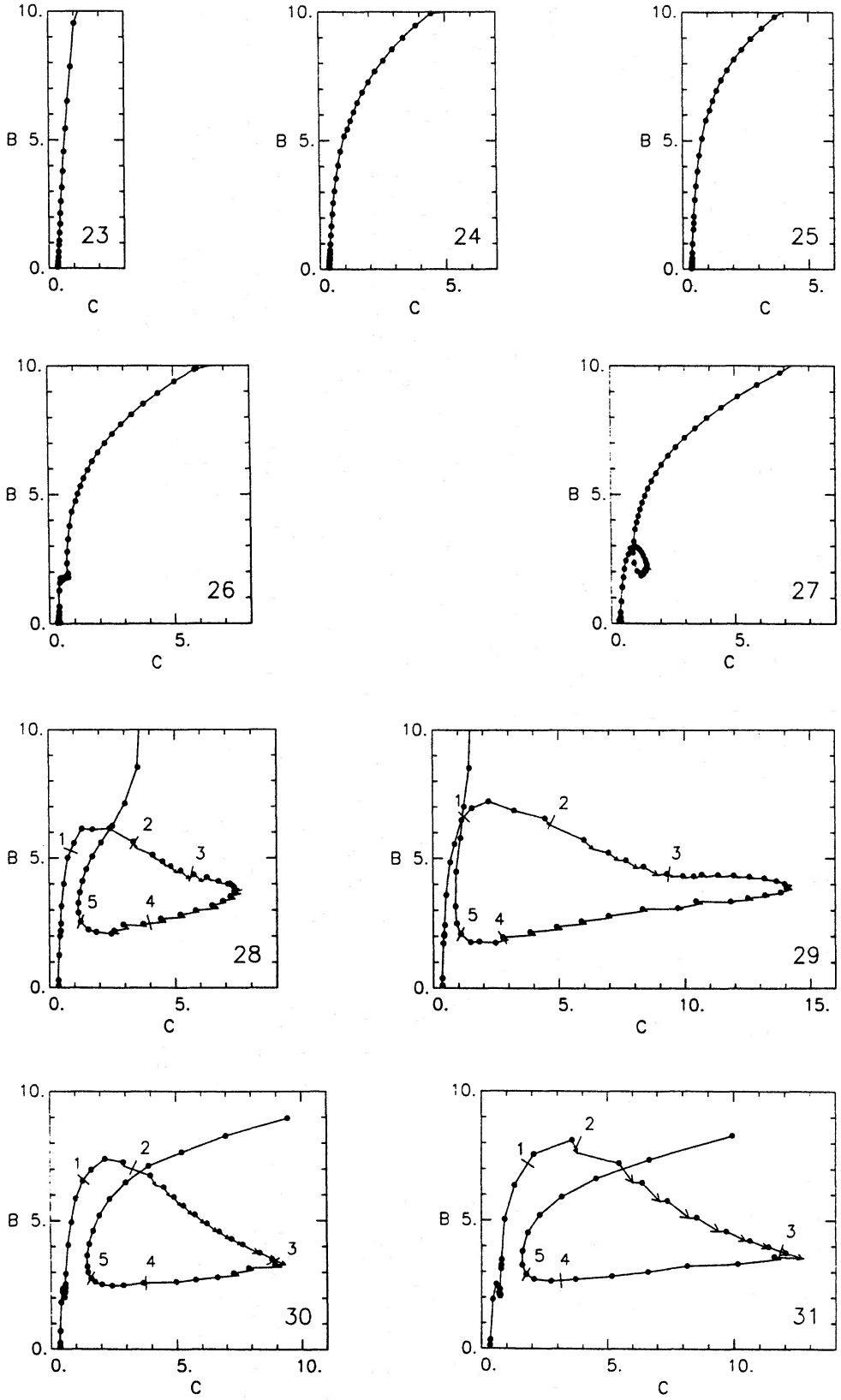


Table c. Parameters of detailed models due to Iben (1965, 1966, 1967), as tabulated in Novotny (1973). Column 1 gives $M_*/M_\odot$, where M_* is the star's total mass. Column 2 gives the model number N, where $7-N$ is the designation of the corresponding table in Novotny. Columns 3–12 give (respectively) $\tau \equiv \log_{10}[t_*/\mathrm{yr}]$, where t_* is the star's age; M_2/M_*, where M_2 is the core mass; $R_*/R_\odot$, where R_* is the star's radius; T_*/K, where T_* is the effective surface temperature; $L_*/L_\odot$, where L_* is the star's total luminosity; L_G/L_*, where L_G is the rate of release of gravitational potential energy; $\phi \equiv \log_{10}[\rho_0/\mathrm{g\ cm^{-3}}]$, where ρ_0 is the central density; the stretching parameter $\Sigma \equiv \log_{10}[4\pi R_*^3 \rho_0/3M_*]$; the mass-weighted mean polytropic exponent $\bar{\eta}$; and the maximum C-value near the base of the envelope $C_{\max}$ (for giants only).

$M_*/M_\odot$	N	τ	M_2/M_*	$R_*/R_\odot$	T_*/K	$L_*/L_\odot$	L_G/L_*	ϕ	Σ	$\bar{\eta}$	$C_{\max}$
1	12	3.630	——	8.095	3967	14.51	1.0000	-1.736	0.840	1.678	——
1	13	6.950	——	1.261	4297	0.485	0.9983	0.959	1.111	1.443	——
1	14	7.699	——	0.865	5740	0.727	0.0143	1.930	1.600	1.383	——
1	15	9.831	0.003	1.076	6034	1.373	-0.0007	2.477	2.428	1.228	——
1	16	9.996	0.100	1.626	5783	2.644	-0.0024	3.443	3.931	1.223	1.011
1	17	10.017	0.135	2.350	4868	2.771	-0.0088	4.296	5.260	1.322	2.931
1	18	10.037	0.199	6.178	4276	11.41	0.0026	4.960	7.185	1.536	17.13
5	23	2.933	——	35.23	4261	365.9	1.0000	-2.101	0.893	1.597	——
5	24	5.301	——	6.450	11190	583.1	0.9998	0.359	1.940	1.318	——
5	25	6.394	——	2.422	18588	626.0	-0.0035	1.319	1.622	1.449	——
5	26	7.823	0.101	4.331	15664	1009.4	-0.0010	1.489	2.550	1.303	0.743
5	27	7.843	0.116	5.819	14395	1299.3	-0.0350	3.116	4.562	1.282	1.427
5	28	7.847	0.123	43.61	4502	698.5	-0.3635	3.907	7.977	1.359	7.320
5	29	7.850	0.128	73.80	3988	1231.5	-0.0040	4.103	8.860	1.542	14.18
5	30	7.931	0.206	20.94	8294	1854.6	0.0004	3.823	6.938	1.353	8.917
5	31	7.944	0.219	44.14	5775	1936.7	0.0679	4.334	8.420	1.302	12.02

the mean density) are well correlated with one another and with $C_{\max}$. Since pure polytropes are only infinite for $\eta \leq 6/5$, one might anticipate a correlation between R_* or Σ and the (mass-weighted) mean polytropic exponent $\bar{\eta}$ for the star as a whole, with the more extended configurations having lower $\bar{\eta}$. However, the correlation, such as it is, is in the opposite sense. For instance, the two most extended giants (Models 18 and 29) have the largest $\bar{\eta}$ ($\simeq 1.5$). In part this is due to the deep convection zones in their outer envelopes, but even if the outer convection zone is excluded from the mean, these models still have large $\bar{\eta}$ ($\simeq 1.4$). Indeed, even the lower envelopes of giants where most of the stretching actually occurs have quite large $\bar{\eta}$ ($\simeq 1.32$).

Appendix D: Previous work

In this appendix we consider previous explanations of giantness due to Höpper & Weigert (1973; hereafter HW), Eggleton & Faulkner (1981; EF), Iben & Renzini (1984; IR), Yahil & van den Horn (1985; YvdH), and Applegate (1988; A). We explain why they are all either incomplete or ill-considered.

D1 THE ROLE OF A STRONG GRAVITATIONAL FIELD IN THE HYDROGEN-BURNING REGION

HW conclude that the primary cause of giantness is the strong gravitational field in the hydrogen-burning region. They base this conclusion on models of chemically homogeneous stars (hereafter HW models), in which hydrogen burns at the centre and the gravitational field is doctored so as to be finite at the centre:

$$g(R) = -\frac{GM(R)}{(R+R_0)^2}, \qquad M(R) = M_0 + \int_{r=0}^{r=R} \rho(r)\, 4\pi r^2\, dr.$$

HW choose M_0 and R_0 so that $g(0)$ and '$\phi(0)$' $= -GM_0/R_0$ match the corresponding quantities at the inner edge of the hydrogen-burning shell in undoctored models of giants. What HW find is that as they increase M_0 their models become much more extended and so evolve away

from the main sequence to become quasi-giants. Note that the detailed evolution of the HW models on the HR Diagram is not very significant in the present context. Once a star swells up to giant proportions (and unless its luminosity is very large), its evolution on the HR Diagram is controlled by its encountering the Hayashi Limit and the consequent adjustments in its upper envelope. All HW have shown is that HW models swell up with increasing M_0. Weiss (1983) has recently performed similar calculations with much the same conclusions.

What HW have done is to find a prescription for doctoring a central hydrogen-burning model so that it can support a giant envelope configuration. It remains the case that giant envelope configurations are a consequence of the opacity law in the lower envelope, and are dependent on convective stability in the lower envelope. In undoctored models this convective stability is engendered not only by the strong gravitational field due to the dense core, but also by the displacement of luminosity generation from the centre of the star. In the HW models, central luminosity generation aggravates the flux bottleneck problem, but this is conveniently compensated by the fact that the luminosity is underestimated, typically by a factor of 3 or 4.

It also remains the case that in undoctored models the critical core structure based on convection or electron degeneracy, and the abrupt molecular weight change across the shell, both play crucial roles in setting up giant envelope configurations. The strong gravitational field supplied by the critical core structure is instrumental in this regard, but the requirements of the critical core structure go beyond just supplying a strong gravitational field in the lower envelope, as we have emphasized in Section 5.6 [i.e. equations (5.9)–(5.11)].

D2 THE ROLE OF REGIONS OF LOW POLYTROPIC EXPONENT

EF propose that the reason for giantness may have to do with the behaviour of polytropes, and in particular with the suppression of η below the critical value 6/5 at the inner edge of the shell. We refer to this suppression of η below 6/5 as the EF effect, and to the region in which it occurs as the EF region. The EF effect is manifest on the η-profiles of all the giants in Fig. C2 as a sharp minimum near tick mark 2. [In helium-burning giants (e.g. Models 30 and 31) there is a second similar sharp minimum associated with the boundary of the helium-burning region.] However, from Fig. C2 it is evident that the region where a giant is most stretched is its lower envelope, i.e. the extensive region above the shell where the solution path is stuck in the *slough* with $0.22 < A < 0.27$ and $1.28 < \eta < 1.37$ (tick marks 3 and 4), rather than the much smaller EF region where $A < 1/6$ and $\eta < 6/5$. In Model 29, for instance, $|\Delta \log_{10}[\rho]| \simeq 1.5$ and $\Delta \log_{10}[R] \simeq 0.3$ in the EF region, whereas $|\Delta \log_{10}[\rho]| \simeq 5$ and $\Delta \log_{10}[R] \simeq 1.8$ in the lower envelope. Giantness cannot be explained simply in terms of lowering the polytropic exponent.

The EF effect actually plays a much more crucial role than the one which EF identify. The same circumstances which suppress η at the inner edge of the shell also suppress A, and thereby ensure that as A subsequently rises through the shell it inevitably runs into the range $0.22 < A < 0.27$ corresponding to the *slough*.

Taking EF's lettered points one by one, we make the following observations:

(a) The development of a convection zone in the outer envelope is an indication that the star has swollen so much that it has run into the Hayashi Limit, i.e. it is a consequence (rather than a cause) of the star's swelling. On the basis of the arguments presented following equation (B13), one might argue that the development of convection in the upper envelope promotes giant-hood by setting an upper limit on η; i.e. that stars pressed against the Hayashi Limit would tend to be less extended if the envelope gas had a higher adiabatic exponent. However, it is difficult to conceive of any physical effect which would increase the adiabatic exponent above 5/3.

(b) Electron degeneracy in the cores of low-mass stars does help to make them giant, by increasing η towards 5/3 at their centres and hence producing the critical core structure.

(c) The Virial Theorem is obeyed by all stars in hydrostatic balance, so it cannot be a cause of giantness.

(d) Giantness is not due to the Schönberg–Chandrasekhar Limit but rather to the GSL. High-mass stars only become giants when their cores have passed the GSL, undergone non-quasistatic contraction, heated up and become convective or close to convective with $\bar{\eta} > 5/4$ in their inner parts, thereby adopting the critical structure.

(e) The molecular weight change plays a vital role in transporting the solution path from the *image* to the *slough*; without a molecular weight change the *slough* would be inaccessible and *almost-singular* solutions could not be sustained (i.e. giant envelope configurations could not be set up). The molecular weight change becomes increasingly effective as the giant evolves and C_2 increases, because it has a multiplicative effect on C (i.e. $C_3 \simeq 2C_2$) and so the larger C_2 is, the larger the increase in C.

(f) The change from central- to shell-burning is important, first because it makes the molecular weight gradient $\bar{m}'$ steeper by reducing $\Delta \ln[M]$ across the burning region, secondly because it relieves the flux bottleneck, thereby promoting convective stability in the lower envelope, and thirdly because as the shell eats its way out through the star and the core grows in mass, the core boundary advances towards and ever deeper into the *image* and the base of the envelope therefore falls ever deeper into the *slough*.

(f′) In pure helium stars the molecular weight change and hence the changes in B and C are small, $f \simeq 1.28$, so that the base of the envelope would only fall in the *slough* if the disposition at the boundary of the helium-exhausted core were $B_2 \simeq 5$ and $C_2 \gtrsim 8$. This disposition is extremely unlikely in view of the arguments presented in Section 7; and it would not be thermostatically locked because the ratio of helium- and carbon-burning temperatures gives $B_2 \simeq 2T_c/T_{He} \simeq 8$ (see Section 7.9).

D3 THE ROLE OF LUMINOSITY IN INFLATING GIANTS

IR propose that the expansion of a giant's envelope is due to the increase in the luminosity generated in the shell. They argue that given the temperature and density dependence of the opacity law, the envelope is obliged to swell in order to remain in radiative equilibrium. However, this argument presupposes without any explicit justification that when the core contracts the shell luminosity increases. The luminosity of the shell depends on its density, temperature, composition, and mass, and it cannot be presumed *a priori* that these quantities will vary in such a way as to increase the luminosity. In effect, IR have put the cart before the horse. What they have actually demonstrated is that as a giant swells the lower envelope becomes more leaky, and this in turn tends to push up the luminosity (although there are other factors influencing the luminosity).

This is not a subtle distinction; it is fundamental. The essence of the problem is to identify in the known physics of giants a clear hierarchy of cause and effect. In this hierarchy giants are giant for the reasons outlined in the main text. Giants have high luminosities because being extended they have a large external radiating surface and are very diffuse and leaky; because the opacity law in the photosphere places a stringent lower limit on the surface temperature; and because being leaky they have a rapid turnover of thermal energy. By rather modest adjustments, particularly in its temperature, the shell can modify its luminosity to meet the turnover of internal thermal energy.

We can justify these assertions from two angles. First, from a purely theoretical point of view, we have been able to explain why giants are giant almost without mentioning the detailed

physics of nuclear energy release. Mathematically this is because there is an extreme imbalance in the interdependence between on the one hand the variables (A, B, C) which we have investigated in detail, and on the other hand L' which we have largely ignored (except to mention its role in increasing A between the core boundary in the *image* and the base of the envelope in the *slough*). We have not demonstrated this imbalance, but it is trivial to do so. Ultimately it derives from the extreme temperature sensitivity of the CN-cycle hydrogen-burning rate.

Secondly, and this is perhaps the more telling argument, evolutionary sequences computed by Iben and others predict that when a star first starts to swell its luminosity increases very little, and may even fall. The luminosity only increases significantly when the star runs into the Hayashi Limit.

A's explanation of giantness is also based on the notion that the expansion is driven by the increasing luminosity, and so it too must be discarded. Moreover, whilst A recognizes the importance of the singular solutions for Kramers opacity, he identifies their role incorrectly. These solutions do not describe the bulk of the swelling envelope (as A suggests) but only the growing density drop immediately above the base of the envelope. This is an important point: a giant envelope is not extended because it goes on and on like a soft polytrope, but because it has a precipitous density drop at its base. Furthermore, A gives no indication of what happens when (as is often the case) Thomson opacity dominates in the lower envelope or there is a mixture of Kramers and Thomson opacity. Finally, he fails to recognize that the singular solutions can be sustained only if a very particular disposition obtains at the base of the envelope, and hence that there are other critical physical effects at work in achieving this disposition.

D4 THE RELEVANCE OF BIFURCATING M-SOLUTIONS

YvdH point out that the stretched region of a giant manifests itself as a loop on the (U, V)-plane, and they relate this to the loops made by polytropic M-solutions. In particular they draw attention to the cut on the (U, V)-plane which (for a given polytropic exponent $\eta < 4/3$) separates solutions which head straight off for the asymptote $(U, V) \rightarrow (0, \infty)$ from solutions which first make another loop around the equilibrium point. They suggest that what distinguishes giants from stars close to the main sequence is whether following the molecular weight change across the shell, the solution is above (main sequence) or below (giant) this cut on the (U, V)-plane. However, there are three inconsistencies in this suggestion.

First, there is no evidence for such a bifurcation. Rather, the loop appears soon after the star leaves the main sequence and as the evolution proceeds simply becomes larger and takes in smaller values of U (i.e. larger values of $C \equiv U^{-1}$). Moreover, during the transition from the main sequence to the Hayashi Limit, the V-value at the outer edge of the shell actually increases monotonically, contrary to what YvdH predict.

Secondly, if the molecular weight change is to be treated as discontinuous and invoked to produce a finite jump from one polytropic solution to another, then the polytropic exponent $\eta_{\rm eff}$ controlling the solutions on either side of this jump must be defined so as not to include the effects of the molecular weight gradient, i.e. $\eta_{\rm eff} = (1 - A)^{-1}$. Now the value of $\eta_{\rm eff}$ at the base of the envelope is typically $\simeq 1.32$ and so the cut on the (U, V)-plane is a very short one connecting a bifurcation point at $(U, V) \simeq (0, 3.9)$ with an equilibrium point at $(U, V) \simeq (0.1, 3.8)$. In early post-main sequence stars which are not giant, the molecular weight change does not deposit the base of the envelope above this cut as YvdH would require, but well below and to the right of it, i.e. $V_3 < 3.8$ and $U_3 > 0.1$, and in giants, although the molecular weight change does deposit the base of the envelope just below the cut, the significant thing about this is that being just below the cut the base of the envelope is right in the *slough*.

Thirdly, the loops in Iben's giant models contain 65–75 per cent of the total stellar mass, and of the remainder most is in the core. Thus if a post-main sequence star were to make a sudden switch from a loopless solution (i.e. one which following the molecular weight change was above the cut and so had no loop) to a looped solution (i.e. one which following the molecular weight change was below the cut etc.), then it would be the looped solution which would have the larger fraction of its mass outside the molecular weight change (i.e. outside the hydrogen-burning shell), and this would imply that it was less evolved.

Finally it should be remarked that YvdH offer little by way of physical justification for the polytropic segments from which their model is assembled.