

GENERAL QUALIFYING EXAM SOLUTIONS: STELLAR ASTROPHYSICS

Jessica Campbell, Dunlap Institute for Astronomy & Astrophysics (UofT)

Contents

1 Stellar Astrophysics	2
1.1 Question 1	2
1.2 Question 2	11
1.3 Question 3	13
1.4 Question 4	18
1.5 Question 5	20
1.6 Question 6	26
1.7 Question 7	29
1.8 Question 8	33
1.9 Question 9	43
1.10 Question 10	46
1.11 Question 11	49
1.12 Question 12	50
1.13 Question 13	54
1.14 Question 14	61
1.15 Question 15	65
1.16 Question 16	66
1.17 Question 17	67
1.18 Question 18	68
1.19 Question 19	74
1.20 Question 20	77
1.21 Resources	79

1 Stellar Astrophysics

1.1 Question 1

Sketch out a Hertsprung-Russell diagram. Indicate where on the main sequence different spectral classes lie. Draw and describe the post main-sequence tracks of both low- and high-mass stars.

1.1.1 Short answer

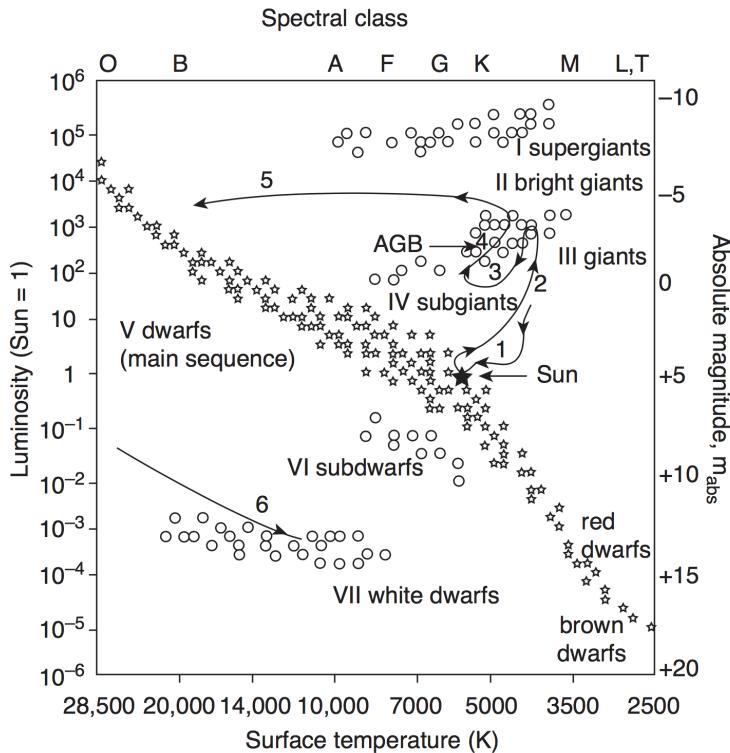


Figure 1: The HR diagram. (left abscissa) Luminosity relative to that of the Sun; (right abscissa) absolute magnitude. Main sequence stars are represented by the symbol \star , and other stars by \circ . The Roman numbers I to VII refer to classification according to luminosity. The contorted solid black line indicates the evolution of a Sun-like star. Track 1: birth line, including the T-Tauri interim state; track 2: expansion to a red giant by H fusion in the stars shell; track 3: recontraction after He fusion in the core has ceased; track 4: AGB, the second red giant state by He fusion in the shell; track 5: mass loss (PN); track 6: final stage (development toward a white dwarf). Figure taken from Rehder (2010).

1.1.2 Additional context

Evolution of a $9 M_\odot$ star: Figure 2 shows the calculated evolutionary path on the HRD of a $9 M_\odot$ star of solar composition. Letters mark critical points in the course of evolution. Specifically, the various points mark the following events:

- A: Beginning of steady hydrogen burning, ZAMS.
- C-C0: Exhaustion of hydrogen in the core, and ignition of hydrogen burning in a shell surrounding the hydrogen-exhausted core.
- E: Arrival on the Hayashi line, that is, the envelope is (almost) fully convective.
- F: Ignition of helium burning in the center.
- J: Exhaustion of helium in the core.
- JK: Ignition of helium burning in a shell surrounding the helium-exhausted core.
- L: Back to the Hayashi line (fully convective envelope).
- LM: Early asymptotic giant branch phase (E-AGB).

Before disclosing what happens at points BDGHI and K (not mentioned in the previous list) it is necessary to introduce a few concepts in the language of stellar model makers. The **core** is the innermost part of the star, where nuclear reactions have greatly altered the original composition. By **envelope** one means the outer part of the star, over (most of) which nuclear burning is negligible and whose composition may still be close to the original one.

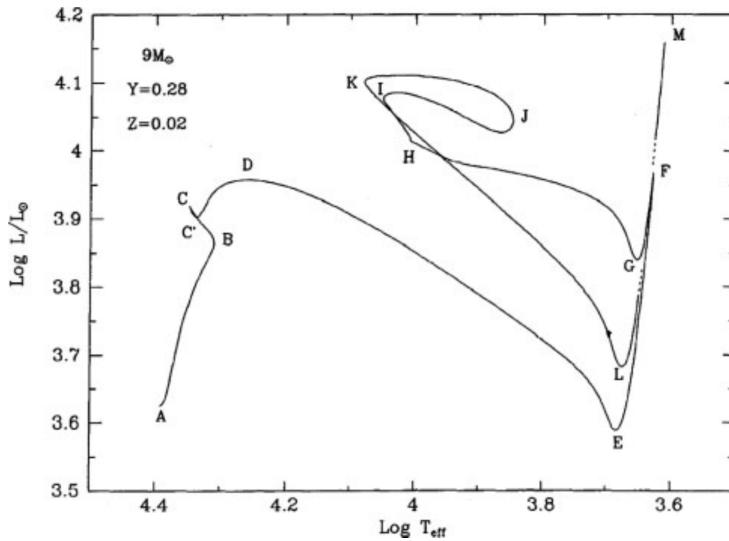


Figure 2: The evolutionary track in the HRD diagram of a $9 M_{\odot}$ star with solar composition from the ZAMS all the way to the AGB phase (hydrogen and helium burning in two separate shells). What happens at the labeled points. Source: Renzini et al. (1992, *Astrophys. J.*, 400, 280). Figure taken from Greggio & Renzini (2011).

Next is the concept of **thermal equilibrium**. A star is said to be in thermal equilibrium (TE) when the total rate of nuclear energy generation (L_N) is almost perfectly equal to its surface luminosity (L_S). TE means that the envelope is able to transfer outside and radiate into space quite precisely the same amount of energy which per unit time is produced by nuclear reactions in the core. If $L_S \simeq L_N$ the star is in TE and its evolution proceeds on a nuclear timescale. Conversely, if $L_S \neq L_N$ the star is out of TE, and its evolution proceeds on a thermal timescale, which usually is much shorter than the nuclear timescale. TE is often broken when the core runs out of fuel or when a new fuel is ignited. A dramatic example of breaking TE is the helium ignition in a degenerate core, called a **helium flash**. At flash peak $L_N \simeq 10^{10} L_{\odot}$ while $L_S \simeq 2000 L_{\odot}$ (and decreases!). The energy that does not escape the star is used to expand the helium core, hence relieving its degeneracy.

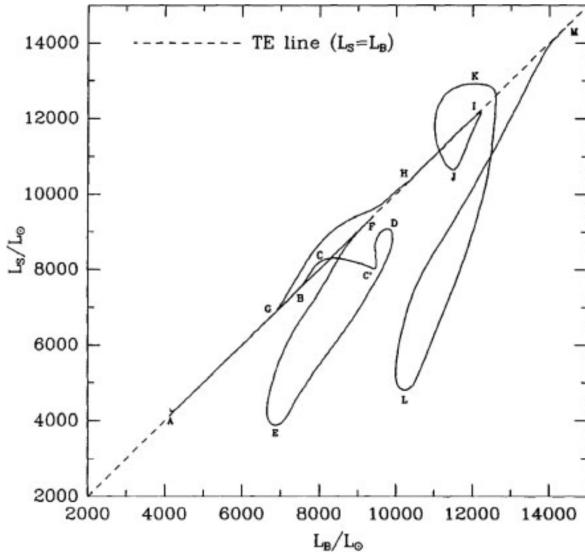


Figure 3: The luminosity radiated from the star's surface as a function of the luminosity released by the stellar core and entering the envelope from its base, for the same track shown in Figure 2. Labels of relevant points along the sequence are the same as in Figure 2. Source: Renzini et al. (1992, *Astrophys. J.*, 400, 280). Figure taken from Greggio & Renzini (2011).

Evolutionary phases in TE and those out of TE can be easily identified by plotting L_S versus L_N , or, equivalently, versus the luminosity impinging at the base of the envelope (L_B), as shown in Figure 3 relative to the evolution of the same $9 M_{\odot}$ star model whose HRD is shown in Figure 2. Letter labels in this figure mark the same events as in Figure 2, so that one can identify those phases that are in TE and those which are out of it.

Thus, phase AB (which is most of the main sequence phase) proceeds in strict TE, but as hydrogen approaches exhaustion in the core, nuclear burning starts to fall short of keeping in pace with the rate at which energy is being radiated away, the star starts departing from TE and begins to contract (point B). At point B, as the core is running out of fuel the envelope starts losing more energy than it receives, and contracts until at point C hydrogen is effectively exhausted over the central regions, and nuclear energy generation quickly shifts to a shell surrounding a hydrogen-exhausted helium core. The quick readjustment from core to shell burning leaves the star at point C', somewhat out of TE, but then the

structure tends to approach again TE, until at point D this tendency is reverted and a major excursion away from TE begins. In stars a little less massive than the one considered here, TE is actually well restored shortly after point C', and yet (as here at point D) TE is broken and stars undergo an extensive loop in the $L_S - L_B$ diagram, as shown in Figure 3. The journey across the HRD from point D to point E takes place on a thermal timescale, and the star expands to red giant dimensions. Clearly, a thermal instability erupts at point D. This is indeed a quite severe thermal instability, suffice to say from Figure 3 that at the peak of the instability the core releases $\sim 7000 L_\odot$ but the envelope radiates away only $\sim 4000 L_\odot$, and $\sim 3000 L_\odot$ are absorbed for its expansion.

The physical origin of this thermal instability is actually quite easy to understand. During phase C'D the luminosity provided by the hydrogen burning shell steadily increases as the shell moves out in the mass coordinate thanks to its own burning, and sits on a progressively more massive helium core. In response to the increasing luminosity impinging on its base (L_B) the envelope slowly expands. By expanding the envelope cools, and by cooling heavy metal ions begin to recombine. Besides being scattered by free electrons, photons now begin to be absorbed by such heavy ions via bound-bound and bound-free transitions: *radiative opacity increases*. This opacity increase is the key factor that determines the onset of the thermal instability. At any point within the star the luminosity transmitted outwards by the radiation field is:

$$L_r = 4\pi r^2 \frac{4acT^3}{3\kappa\rho} \frac{dT}{dr} = 4\pi r^2 F_r \text{ [erg s}^{-1}\text{]},$$

where r is the distance from the center, T the temperature, ρ the density, κ the opacity, and F_r the radiative energy flux. During phase C'D the envelope slowly expands, hence r^2 increases while the flux F_r decreases, but their product still increases and the star is approaching TE. However, as this trend continues the increase in opacity accelerates and eventually the flux drops faster than r^2 increases, and their product L_r starts to decrease. The decrease happens first near the stellar surface, and then (very quickly) through the whole envelope. At this point the envelope is transferring outwards and radiating away less energy than it receives from the stellar core, that is, $\Delta L = L_B - L_S > 0$. But as the envelope expands more ions recombine, opacity increases further, the flux drops even more, the thermal imbalance $L_B - L_S$ increases and expansion accelerates: the stellar envelope is in a thermal runaway, as it becomes more and more unable to radiate away the energy it receives from the stellar interior.

At point E the surface luminosity starts to rise again, $L_B - L_S$ begins to drop, and TE is rapidly restored. What relieves the instability and saves the star from literally falling apart is convection. As the expanding envelope cools, and opacity increases, eventually the radiative gradient ∇_{rad} ¹ exceeds the adiabatic gradient ∇_{ad} , first near the photosphere, where hydrogen is only partly ionized, and then rapidly through the whole envelope. Thus, convection replaces radiative transfer in carrying out energy through the envelope, and the thermal instability is quenched since it was intimately related to the radiative mode of energy transfer. Most of the energy flux in the envelope being now carried by convective motions, the envelope ceases to absorb energy, and the surface luminosity LS starts to increase again: the star now ascends the Hayashi line, until at point F the helium burning reactions ignite in the core.

Following helium ignition, the helium core initiates a very slow expansion, which will last through a major fraction of the helium burning phase. This is because the helium burning core works like a breeder reactor during this stage, that is, it produces more new fuel than it burns. As triple- reactions produce fresh ^{12}C , the $^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$ reactions release an increasing amount of energy, and the inner core is forced to expand slightly. Such a modest expansion of the core is sufficient to cause a decrease of temperature and density in the surrounding hydrogen burning shell and the stellar luminosity correspondingly starts to decrease. The star now evolves along the **Hayashi line** from F to G, burning helium in the core and hydrogen in the shell, until thermal stability is broken again at point G. As the envelope contracts it heats up gradually, and in particular near its base, heavy ions start losing electrons. Hence opacity decreases along with the radiative gradient. As ∇_{rad} drops below ∇_{ad} a radiative region appears at the base of the envelope and grows outwards in mass. As this growth progresses the envelope becomes more and more transparent to radiation, until it starts radiating away more energy than it receives from the core: $L_B - L_S$ turns increasingly negative, the envelope starts deflating, but the more it contracts the more “transparent” it becomes, and the more energy it loses into space. The thermal instability bringing the star in its envelope deflation from point G to point H is precisely the reverse analog of the thermal instability that causes the runaway expansion from point D to point E. As envelope inflation can be ascribed to a runway recombination of heavy elements in the envelope, envelope deflation is due to a runway ionization of such heavy elements. A comparison of the three figures (Figures 2-4) helps in visualizing the onset of the thermal instability and of its results.

Thus, in this star, the core helium burning phase in thermal equilibrium is spent in two distinct locations

¹Temperature gradients are defined as $\nabla T(\rho, T, X_i) = (\partial \log T / \partial \log P)$.

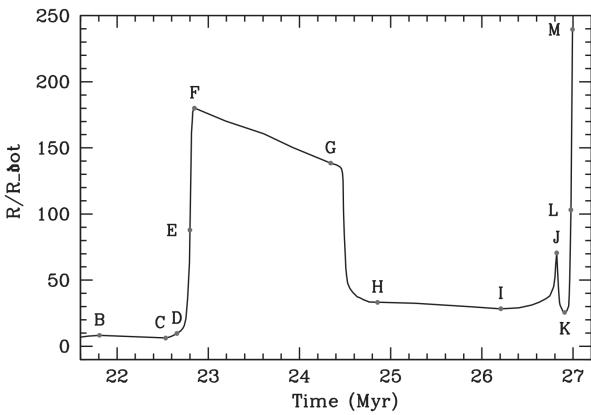


Figure 4: The stellar radius as a function of time for the same track shown in Figure 2. Labels of relevant points along the sequence are the same as in Figure 2. Source: Renzini et al. (1992, *Astrophys. J.*, 400, 280). Figure taken from Greggio & Renzini (2011).

in the HRD: part on the Hayashi line as a red giant (F-G) and part as a blue giant (H-I) separated by a runaway contraction out of TE (G-H). This first blue loop (also called the **Cepheid blue loop**) continues after point I when TE is broken again. Here the envelope is slowly expanding in response to the slowly increasing luminosity from the core and shell burning, when the envelope turns unstable again due to the same physical process already described for the DE phase.

However, the journey towards the Hayashi line is suddenly interrupted and inverted at point J, which marks the start of the so-called **second blue loop**. What happens at point J and after is a rather complex series of interconnected events: helium is exhausted in the core, the core contracts and helium burning rapidly shifts from the helium-exhausted CO core to a shell surrounding it; helium shell ignition is quite violent and causes expansion of the helium buffer above this shell: when the expansion front breaks through the hydrogen burning shell temperature and density in the shell drop; burning in the hydrogen shell (which was still providing most of the stellar luminosity) is effectively shut off completely causing a sudden drop of the luminosity impinging on the base of the envelope: L_B drops below L_S and the envelope stops expanding and contracts. In the meantime the strength of the helium burning shell steadily increases until it leads L_B to exceed L_S again, contraction stops and the star resumes at point K its runaway expansion that was temporarily stopped at J. The rest, from K to L to M is quite similar to the DEF phase, with convection replacing radiative transfer in the envelope and TE being rapidly restored, shortly before point M. Figure 4 clearly illustrates the dramatic effects of the envelope thermal instabilities on the overall structure of the star. Those phases which are in TE are nearly flat in this plot, that is, the stellar radius changes quite slowly with time. With one exception, those phases which are out of TE are instead nearly vertical, that is, the radius varies very rapidly during such runaway inflations or deflations. The exception is phase BC, which is only modestly out of TE, and contraction is relatively slow.

What happens past point M is still an open issue. A $9 M_\odot$ star lies between two domains: massive stars that eventually undergo core collapse and supernova explosion, and intermediate-mass stars, which shedding all their hydrogen-rich envelope die as a white dwarf. One believes that in a $9 M_\odot$ star carbon is ignited in the central core under only mildly degenerate conditions, and the star keeps ascending along the Hayashi line as a super asymptotic giant branch star experiencing a few thermal pulses in its deeper regions, where hydrogen and helium are still burning in two separate shells. Thus, if the envelope is completely lost in a wind the star leaves an ONeMg white dwarf. If instead mass loss is less severe then the core keeps growing in mass thanks to the active burning shells until electron captures in the core trigger a core collapse and we have a supernova explosion. Clearly the fate critically depends on the strength of the mass loss process.

Evolution of a Solar-composition star: Figure 1.4 shows stellar evolutionary tracks of solar composition, covering a wide range of initial masses, from a $0.8 M_\odot$ star, whose MS evolutionary lifetime is longer than one Hubble time, up to a $20 M_\odot$ model. The evolutionary tracks of stars with mass greater than $20 M_\odot$ are heavily affected by mass loss.

Different gray shades in Figure 5 pertain to the different mass ranges as customarily distinguished in stellar evolution: low-mass stars, up to $\lesssim 2.2 M_\odot$, intermediate-mass stars, up to $\lesssim 8 M_\odot$, and high-mass stars. These ranges correspond to different physical behavior during the evolution of the stars, and the mass limits depend on chemical composition. Low-mass stars develop an **electron degenerate helium core** after their MS evolution; intermediate mass stars ignite helium under non-degenerate conditions, but develop a **CO degenerate core** after central helium burning; massive stars experience all successive nuclear burnings up to the production of an **iron core**.

During their MS evolution, stars with mass below $\sim 1 M_\odot$ burn hydrogen through the **p-p chain**, whose

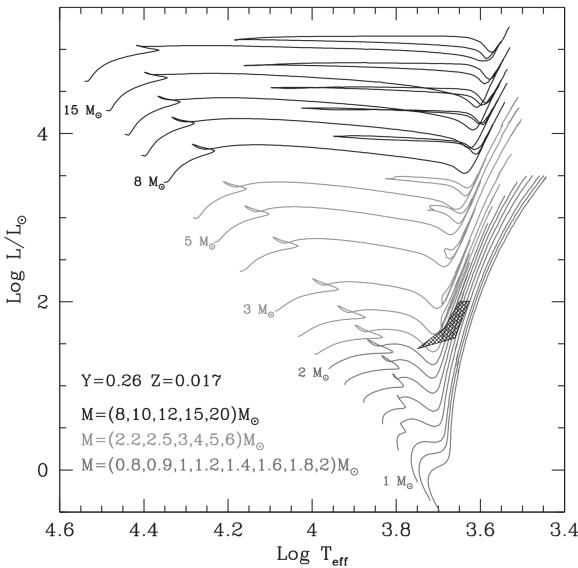


Figure 5: Evolutionary tracks of solar composition star. The shaded area shows the location of low-mass ($0.55 \leq M/M_{\odot} \leq 2$) core helium burning models. Drawn using the YZVAR database (Bertelli, G. et al. 2008, Astron. Astrophys., 484, 815; 2009, Astron. Astrophys., 508, 355). Figure taken from Gregg & Renzini (2011).

reaction rate is not extremely sensitive to temperature, so that these stars possess a radiative core. As evolution proceeds hydrogen is progressively depleted in the inner core, more in the center than in the periphery, leading to a smooth hydrogen profile from a central minimum up to the initial abundance in the outer layers. On the HRD, the stellar track climbs to higher luminosities and temperatures until the central hydrogen is severely depleted; at this point the effective temperature starts decreasing, producing the turnoff (TO), that is, the maximum temperature point on the MS, which is easily recognizable on the CMD of globular clusters. Shortly after the TO, hydrogen is completely exhausted in the center and the hydrogen shell burning phase starts as a natural progression from the previous core hydrogen burning phase. During shell hydrogen burning, the track forms the subgiant branch (SGB), evolving at (almost) constant luminosity and decreasing temperature until the external convection penetrates deeply inside, and the star becomes almost fully convective at the base of the red giant branch (RGB).

The MS stars (with $M \gtrsim 1 M_{\odot}$) have a convective core, since at least part of the hydrogen burning occurs through the CNO cycle, whose energy generation rate is extremely sensitive to temperature. Because of central mixing, the hydrogen profile is characterized by an inner plateau; as evolution proceeds, the extension of the convective core progressively decreases, leaving behind a gradient of hydrogen abundance. In stars with a convective core, the fuel depletion affects a sizable central region, which starts rapid contraction when approaching fuel exhaustion. This happens at the local minimum effective temperature point during the MS evolution of stars with $M > 1 M_{\odot}$ (Figure 5), which signals the beginning of the overall contraction phase. The evolutionary behavior across this phase and the following runaway expansion has been already described in detail in the previous section.

As already mentioned, the helium core of low-mass stars is (electron) degenerate: this implies that central helium ignition is delayed because core contraction does not lead to an efficient increase of the central temperature. The (almost) fully convective star climbs along the RGB, while the hydrogen burning shell progresses outward, thereby increasing the mass of the helium core. When the core reaches a critical limit, a helium flash occurs off-center, since neutrino losses induce a temperature inversion in the innermost layers. This event is not catastrophic for the star because local expansion removes the degeneracy; instead, a sequence of flashes occurring at progressively inner locations totally remove the core degeneracy. During this phase, which lasts 1 Myr, the star moves downward along the Hayashi line to settle on the core helium burning locus, either the red clump, or the horizontal branch (HB). The maximum luminosity reached on the RGB (RGB Tip, or TRGB) is a very important feature in the color-magnitude diagram (CMD) of stellar populations because it is virtually independent of mass, and then of evolutionary lifetime. This is because on the one hand the critical mass for helium ignition under degenerate conditions is almost constant ($0.5 M_{\odot}$), and on the other, along the RGB there exists a core mass-luminosity relation. The evolutionary lifetimes of low-mass stars range from 1 Gyr up to the Hubble time, as seen in Figure 5, and all of them experience the helium flash at almost the same luminosity; therefore, the CMD of a stellar population with stars older than 1 Gyr will show a prominent TRGB feature whose luminosity is known, thus allowing us to determine the distance of the stellar population. The TRGB luminosity depends slightly on metallicity, being higher for higher metal content. However, by a fortunate combination, the absolute magnitude in the I-band of the TRGB does not depend much on metallicity, for metal-poor populations. This is because the effective temperature at the TRGB also depends on metallicity: the higher Z , the cooler the TRGB stars, and the higher the bolometric correction (in absolute value). The

trend with metallicity of the bolometric correction to the I-band largely compensates that of the tip luminosity, so that the I-band absolute magnitude of the TRGB ($M_{I,\text{TRGB}}$) is almost independent of age and weakly dependent on metallicity of the parent stellar population. This makes $M_{I,\text{TRGB}}$ a very effective distance indicator in galaxies.

The luminosity of core helium burning low-mass stars, being fixed by the mass of their hydrogen-exhausted core, is also largely independent of their total mass, thereby providing another distance indicator. However, evolution during the core helium burning phase spreads the red clump stars over 0.6 magnitudes, as can be seen from the vertical size of the hatched region in Figure 5. In addition, the core mass at helium ignition is not a monotonic function of the total mass, and as the latter increases beyond the limit of the low-mass stars regime (M_{Hef}), it decreases, reaches a minimum of about $0.326 M_{\odot}$, and then increases. As a result, stars with mass just above M_{Hef} start their core helium burning evolution at fainter luminosities, have the longest helium burning lifetimes, and cover a wider range of luminosities, compared to stars with $M < M_{\text{Hef}}$. Therefore, the core helium burners of a composite stellar population with a sizable component at ages just below 1 Gyr are spread over a wide magnitude range, which limits the use of this feature for distance determinations.

The effective temperature of low mass core helium burning stars depends on the mass of their envelope, a dependence which is very pronounced when the envelope is thinner than for example $0.2 M_{\odot}$ for $Z \lesssim 0.1 Z_{\odot}$. Below this threshold, the lower the envelope mass, the hotter the core helium burning stars. In a coeval and homogeneous stellar population the core helium burning stars all have virtually the same core mass and then luminosity; a spread of their envelope masses produces a feature on the HRD at about constant (bolometric) magnitude extending over a temperature range. The observational counterpart of this locus is the horizontal branch: a prominent feature in the HRD of globular clusters whose age is old enough to host core helium burning stars with low-mass envelopes. The existence of wide HBs in globular clusters was explained as due to a dispersion of mass lost during the RGB phase, but recently it became apparent that other effects are also at play. Evolutionary tracks during the core helium burning phase exhibit a wide variety of morphologies, depending on their mass, envelope mass, metallicity and helium abundance. At central helium exhaustion, a rapid core contraction leads to shell helium ignition; the model star expands and moves again towards the Hayashi line to start the asymptotic giant branch (AGB) evolutionary stage. However, if the envelope mass is small enough, the shell helium burning phase is entirely spent at high effective temperatures, as an AGB manqué star. The hot HB stars and their AGB manqué progeny are likely responsible for the UV emission from elliptical galaxies which host old stellar populations.

The evolution of intermediate and high-mass stars during the hydrogen and helium burning stages is very similar to what has already been described for the $9 M_{\odot}$ star. But, following helium exhaustion, intermediate-mass stars behave similarly to low-mass stars. We only notice here that for stars with mass in the vicinity of M_{Hef} the core helium burning phase is spent in a red clump, which becomes a wider and wider blue loop as the model mass grows. Thus, the core helium burning phase of intermediate-mass stars is spent part in the blue and part in the red. The very occurrence of the loop, its extension and the fraction of lifetime spent on each side of the loop, are sensitive to a number of parameters describing the input physics of the models, like metallicity, opacity, convection and others. Therefore, the blue-to-red ratio in stellar populations with stars in this mass range is difficult to interpret. The luminosity of intermediate mass core helium burners is instead a more robust prediction of the models, and can be effectively used as an age indicator of stellar populations.

Dependence on chemical composition: The evolutionary tracks of given mass are very sensitive to the (initial) helium content (Y) and metallicity (Z), which control the energy generation rates and the opacity. We briefly illustrate some dependencies relevant to the interpretation of the HR diagram and of the spectral energy distribution of stellar populations.

At a given initial mass, evolutionary lifetimes become shorter as Y increases, because stars with higher molecular weight are more compact, hotter, brighter, hence faster in consuming the hydrogen fuel, of which less is available. Instead, lifetimes become longer as Z increases, because the hydrogen burning models get fainter due to the higher opacity, while the hydrogen fuel reservoir remains virtually unchanged. Table 1.2 lists the values of the coefficients of the following relation:

$$\log t = a \log^2 M_0 + b \log M_0 + c \text{ [yr]},$$

adopted to describe the evolutionary lifetimes (in years) as a function of initial mass (in M_{\odot}). The parabolic fit over the whole considered mass range is a rather drastic approximation; nevertheless these analytic relations can be useful to estimate evolutionary lifetimes and their dependence on composition. As mentioned previously, the values of mass defining the low, intermediate and high-mass range depend somewhat on chemical composition. For example M_{Hef} decreases with Y increasing and with Z decreasing. Therefore, an extended RGB on the HRD of a stellar population traces the presence of evolved stars with

Z, Y	H burning lifetime				Lifetime up to first pulse or C ignition			
	a	b	c	$\left\langle \frac{\Delta t_{TO}}{t_{TO}} \right\rangle$	a	b	c	$\left\langle \frac{\Delta t_{tot}}{t_{tot}} \right\rangle$
0.0001, 0.23	0.8751	-3.240	9.754	0.041	0.8130	-3.194	9.809	0.022
0.0001, 0.40	0.8248	-3.007	9.288	0.049	0.7784	-2.992	9.371	0.023
0.001, 0.26	0.8192	-3.168	9.705	0.040	0.7679	-3.147	9.776	0.031
0.001, 0.40	0.7590	-2.951	9.310	0.049	0.7363	-2.976	9.409	0.032
0.004, 0.26	0.8085	-3.230	9.796	0.040	0.7534	-3.211	9.875	0.042
0.017, 0.26	0.8071	-3.423	10.029	0.038	0.7260	-3.369	10.104	0.053
0.04, 0.26	0.7466	-3.516	10.213	0.049	0.6694	-3.457	10.281	0.063
0.04, 0.40	0.7649	-3.361	9.767	0.073	0.7324	-3.371	9.866	0.063
0.04, 0.46	0.7362	-3.243	9.554	0.082	0.7427	-3.303	9.669	0.068
0.07, 0.40	0.7322	-3.420	9.855	0.079	0.7065	-3.437	9.948	0.066

Table 1: Coefficients of $\log t$ for various chemical compositions, resulting from the least square fit of the hydrogen burning and the total lifetimes as a function of M_0 . The fit covers the range $0.6 \leq M_0/M_\odot \leq 20$; columns 5 and 9 report the average relative accuracy on the evolutionary lifetimes at the TO and at the first thermal pulse or central carbon ignition. The YZVAR database (Bertelli, G. et al. 2008, Astron. Astrophys., 484, 815; 2009, Astron. Astrophys., 508, 355) has been used to derive the coefficients. Table taken from Greggio & Renzini (2011).

mass lower than $2.1 M_\odot$ for solar composition, or with mass lower than $1.5 M_\odot$ if $Z = 0.001, Y = 0.4$. However, evolutionary lifetimes at given mass also depend on composition, and, by and large, an extended and well populated RGB is developed in stellar populations older than 1 Gyr almost irrespective of chemical composition.

At fixed initial mass, zero age main sequence models are hotter and brighter for a higher helium content, and/or a lower metallicity. Indeed, the locus of the corresponding stars on the HRD (ZAMS) is used to infer the composition of the target stellar population, and is virtually the only way to estimate the helium content of these stars, if Z is known from spectroscopy. The puzzling composition of multiple stellar populations in some galactic globular clusters has been derived from ZAMS fitting.

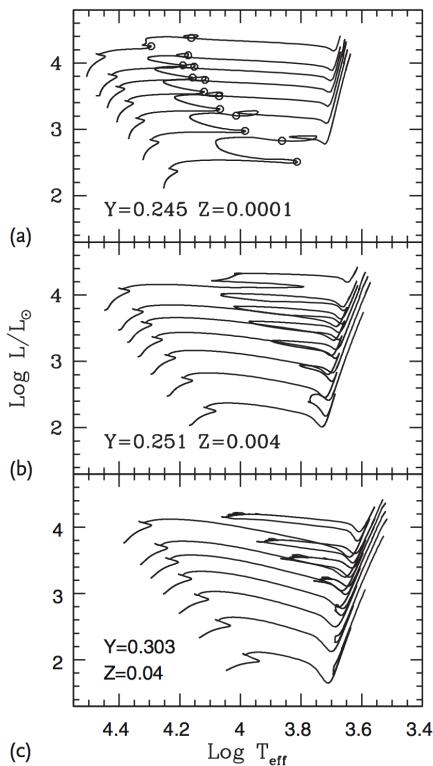


Figure 6: a-c) Evolutionary tracks of intermediate-mass stars ($M/M_\odot = 3, 4, 5, 6, 7, 8, 10$) for different compositions, as labeled. In (a), the open circles mark start and the end of the core helium burning phase. Drawn using the BaSTI database (Pietrinferni, A. et al. 2004, Astrophys. J., 612, 168). the Figure taken from Greggio & Renzini (2011).

Figure 6 shows some evolutionary tracks of intermediate-mass stars for three different initial compositions. The described trend of the ZAMS with metal content is readily visible, together with some other properties already mentioned: at very low metallicity (Figure 6a) the entire core helium burning phase occurs in the blue part of the HRD, and the thermal runaway in the envelope which brings the model star to the Hayashi

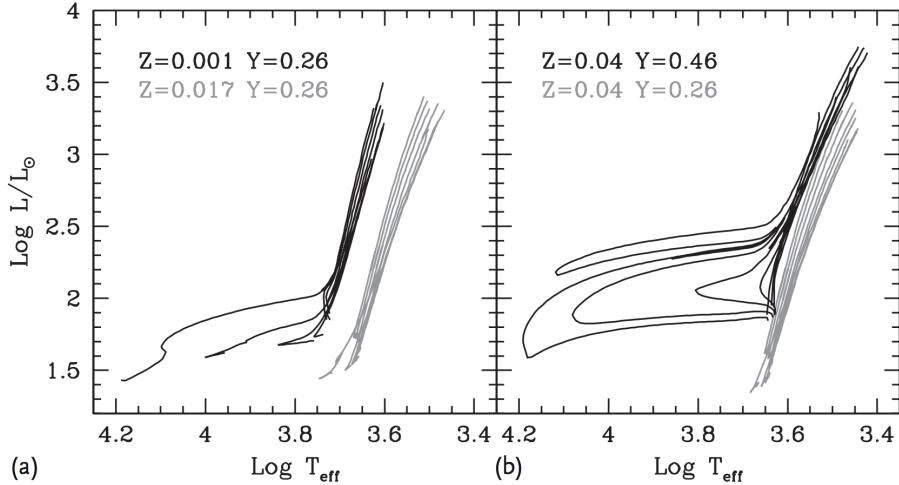


Figure 7: (a,b) Evolutionary tracks of low-mass stars during the core helium burning and early AGB phases. The chemical composition is labeled. Initial model masses are $M/M_{\odot} = 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.9, 1.0, 1.1, 1.2, 1.4, 1.6$ except for the ($Z = 0.04, Y = 0.46$) set, for which the $1.6 M_{\odot}$ track ignites helium under non-degenerate conditions. Drawn using the YZVAR database (Bertelli, G. et al. 2008, Astron. Astrophys., 484, 815; 2009, Astron. Astrophys., 508, 355). Figure taken from Greggio & Renzini (2011).

track is delayed to the very latest stages. As metallicity increases, the luminosity decrease associated with the thermal runaway gets more and more pronounced: this reflects the progressively higher opacity, and then radiative energy trapping in the envelope. At the same time, at higher metallicity it is more difficult to produce extended blue loops: the 3 and $4 M_{\odot}$ tracks at $Z \sim 2 Z_{\odot}$ do not present a blue loop at all, and the loop of the $5 M_{\odot}$ track is just alluded to. The models in Figure 6 are computed adopting classical recipes for the input physics, and even slight modifications of the assumptions lead to dramatic variations of the tracks shape. For example, intermediate-mass models with $Z = 0.04, Y = 0.40$ which adopt a modest overshooting from the convective core lack the loops completely, and the core helium burning phase is totally spent close to the Hayashi line.

Figure 7 illustrates the effect of chemical composition on core helium burners of low-mass. The gray tracks in Figure 7a show that at solar composition this evolutionary phase is completely spent in the red, even for masses as low as $0.55 M_{\odot}$. Conversely, at low metallicity (black lines), low-mass core helium burners are blue, opening the possibility of producing extended HBs in old stellar populations. The tracks in Figure 1.6b show the effect of enhancing the helium abundance: at high metallicity the core helium burning phase is completely spent close to the Hayashi line for a solar helium abundance, but if the helium abundance is high, a blueward excursion occurs during the core helium burning phase which is very wide for low-mass stars. Therefore the production of blue HB stars in old stellar populations can be achieved either assuming heavy mass loss on the RGB, or a high helium content, or both. Actually, the existence of multiple stellar populations with different helium content in NGC 2808 has been first suggested on the basis of its HB stars distribution, and confirmed later from the multiple MSs. Mass loss and helium abundance have indeed an important impact on the HRD of stellar populations, as well as on their spectral energy distribution, since stars in the core helium burning phase provide an important contribution to the total light.

1.1.3 Follow-up Questions

- How does metallicity affect this diagram?
- Elaborate on the life sequence of low-mass stars.
- Elaborate on the life sequence of high-mass stars.
- Which direction does radius increase?
- What's the functional form for how radius increases as a function of temperature and luminosity?
- Why can you use the tip of the red giant branch for distance determination?
- How can you use the H-R diagram to determine ages? Which of the two techniques is more accurate?

1.2 Question 2

Sketch a plot of radius versus mass for various “cold” objects made of normal matter, including planets, brown dwarfs and white dwarfs. Explain the mass-size relationship for rocky and gaseous objects. Why is there an upper mass limit?

1.2.1 Short answer

Using conservation of mass, we can write:

$$\begin{aligned} dM &= 4\pi r^2 \rho dr \\ M &= \frac{4}{3}\pi R^3 \rho \\ \rho &\sim \frac{M}{R^3}. \end{aligned}$$

Then, using hydrostatic equilibrium, we get:

$$\begin{aligned} dP &= -\frac{GM}{r^2} \rho dr \\ P &\sim \frac{M}{R} \rho \\ P &\sim \frac{M^2}{R^4}. \end{aligned}$$

We know that $P \sim \rho^\alpha$, which we can solve using the equation of state for each object:

Planets (rocky):

$$\begin{aligned} \rho &\sim \frac{M}{R^3} \sim \text{const} \\ R &\sim M^{1/3}. \end{aligned}$$

Planets (gas giants):

$$\begin{aligned} P &\sim \rho^2 \\ \frac{M^2}{R^4} &\sim \frac{M^2}{R^3} \\ &\text{independent.} \end{aligned}$$

Brown dwarfs/low-mass white dwarfs:

$$\begin{aligned} P &\sim \rho^{5/3} \\ \frac{M^2}{R^4} &\sim \frac{M^{5/3}}{R^5} \\ R &\sim M^{-1/3}. \end{aligned}$$

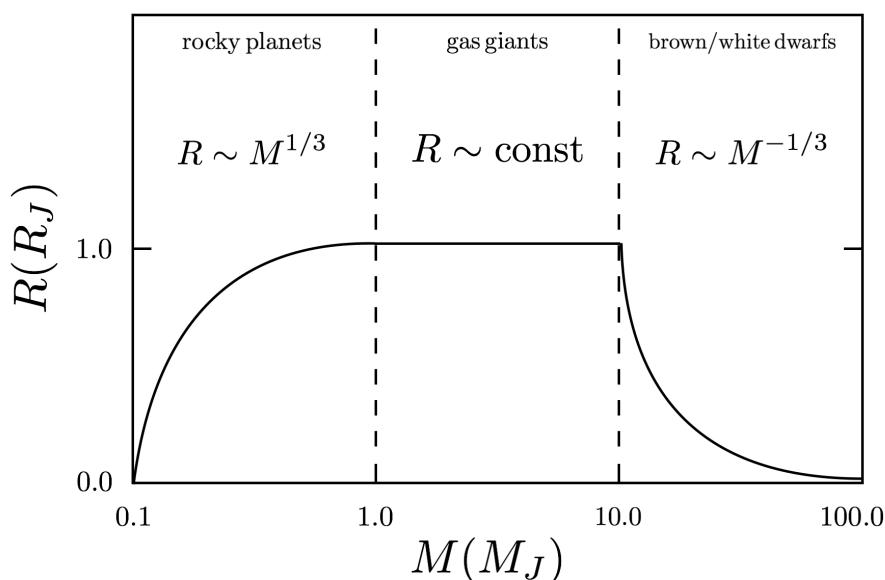


Figure 8: Radius-mass relationship for various cold objects.

1.2.2 Follow-up Questions

- How do you calculate the Chandrasekhar mass limit?
- Why is Saturn smaller than Jupiter? Or, why do we see a range of radii in extrasolar planets (e.g., hot Jupiters)?

1.3 Question 3

Describe the physical conditions that lead to the formation of absorption lines in stars' spectra. What leads to emission lines?

1.3.1 Short answer

We can use Kirchoff's laws of radiation to understand the formation of spectral lines in stellar photospheres. These laws are:

1. Hot dense gas (e.g., stars) emits blackbody radiation following the description of the Planck function.
2. Diffuse gas emits emission lines via electronic transitions from a higher state to a lower state, with the energy of the emitted photon corresponding to the energy difference between electronic states.
3. Cool, diffuse gas in front of a hot dense gas will produce absorption lines on top of the continuum emission. These will occur where the diffuse gas has emission lines, but instead result in the formation of absorption lines instead of emission lines.

So in the example of stellar photospheres, the hot, dense stellar interior produces blackbody emission while the hot, diffuse photosphere produces emission lines. Since there is a temperature gradient in the stellar photosphere, the outer layers of the photosphere will lie at a cooler temperature which allows for the formation of absorption lines.

1.3.2 Additional context

If an absorber X is in a level ℓ and there is radiation present with photons having an energy equal to $E_u - E_\ell$, where E_ℓ and E_u are the energies of levels ℓ (for "lower") and u (for "upper"), the absorber can absorb a photon and undergo an upward transition:

$$\text{absorption : } X_\ell + h\nu \rightarrow X_u, \quad h\nu = E_u - E_\ell.$$

Suppose that we have number density n_ℓ of absorbers X in level ℓ . The rate per volume at which the absorbers absorb photons will obviously be proportional to both the density of photons of the appropriate energy and the number density n_ℓ , so we can write the rate of change of n_ℓ due to photo-absorption by level ℓ as

$$\left(\frac{dn_u}{dt} \right)_{\ell \rightarrow u} = - \left(\frac{dn_\ell}{dt} \right)_{\ell \rightarrow u} = n_\ell B_{\ell u} u_\nu, \quad \nu = \frac{E_u - E_\ell}{h},$$

where u_ν is the **radiation energy density per unit frequency**, and the proportionality constant $B_{\ell u}$ is the **Einstein B coefficient** for the transition $\ell \rightarrow u$.

An absorber X in an excited level u can decay to a lower level ℓ with emission of a photon. There are two ways this can happen:

$$\text{spontaneous emission : } X_u \rightarrow X_\ell + h\nu, \quad \nu = (E_u - E_\ell)/h,$$

$$\text{stimulated emission : } X_u + h\nu \rightarrow X_\ell + 2h\nu, \quad \nu = (E_u - E_\ell)/h.$$

Spontaneous emission is a random process, independent of the presence of a radiation field, with a probability per unit time $A_{u\ell}$ – the **Einstein A coefficient**.

Stimulated emission occurs if photons of the identical frequency, polarization, and direction of propagation are already present, and the rate of stimulated emission is proportional to the density of these photons. Thus the total rate of depopulation of level u due to emission of photons can be written

$$\left(\frac{dn_\ell}{dt} \right)_{u \rightarrow \ell} = - \left(\frac{dn_u}{dt} \right)_{u \rightarrow \ell} = n_u (A_{u\ell} + B_{u\ell} u_\nu), \quad \nu = \frac{E_u - E_\ell}{h},$$

where the coefficient $B_{u\ell}$ is the **Einstein B coefficient** for the downward transition $u \rightarrow \ell$. Thus we now have three coefficients characterizing radiative transitions between levels u and ℓ : $A_{u\ell}$, $B_{u\ell}$ and $B_{\ell u}$. We will now see that they are not independent of one another.

In TE, the radiation field becomes the "blackbody" radiation field, with intensity given by the **blackbody spectrum**

$$B_\nu(T) = \frac{2h\nu^2}{c^2} \frac{1}{\exp(h\nu/k_B T) - 1} [\text{erg s}^{-1} \text{cm}^{-2} \text{Hz}^{-1} \text{sr}^{-1}],$$

with specific energy density

$$(u_\nu)_{\text{LTE}} = \frac{4\pi}{c} B_\nu(T) = \frac{8\pi h\nu^3}{c^3} \frac{1}{e^{h\nu/k_B T} - 1} [\text{erg s}^{-2} \text{cm}^{-1} \text{Hz}^{-1} \text{sr}^{-1}].$$

If we place absorbers X into a blackbody radiation field, then the net rate of change of level u is

$$\begin{aligned} \frac{dn_u}{dt} &= \left(\frac{dn_u}{dt} \right)_{\ell \rightarrow u} + \left(\frac{dn_u}{dt} \right)_{u \rightarrow \ell} \\ &= n_\ell B_{\ell u} \frac{8\pi h\nu^3}{c^3} \frac{1}{\exp(h\nu/k_B T) - 1} - n_u \left(A_{u\ell} + B_{u\ell} \frac{8\pi h\nu^3}{c^3} \frac{1}{\exp(h\nu/k_B T) - 1} \right). \end{aligned}$$

If the absorbers are allowed to come to equilibrium with the radiation field, levels ℓ and u must be populated according to $n_u/n_\ell = (g_u/g_\ell)e^{(E_\ell E_u)/k_B T}$, with $dn_u/dt = 0$. From the equation above, it is easy to show that $B_{u\ell}$ and $B_{\ell u}$ must be related to $A_{u\ell}$ by

$$\begin{aligned} B_{u\ell} &= \frac{c^3}{8\pi h\nu^3} A_{u\ell} [\text{s}^{-1}], \\ B_{\ell u} &= \frac{g_u}{g_\ell} B_{u\ell} = \frac{g_u}{g_\ell} \frac{c^3}{8\pi h\nu^3} A_{u\ell} [\text{s}^{-1}]. \end{aligned}$$

Thus the strength of stimulated emission ($B_{u\ell}$) and absorption ($B_{\ell u}$) are both determined by $A_{u\ell}$ and the ratio g_u/g_ℓ .

Rather than discussing absorption and stimulated emission in terms of the radiation energy density u_ν , it is helpful to characterize the intensity of the radiation field by a dimensionless quantity, the photon occupation number n_γ :

$$\begin{aligned} n_\gamma &\equiv \frac{c^2}{2h\nu^3} I_\nu \text{ [dimensionless]}, \\ \bar{n}_\gamma &\equiv \frac{c^2}{2h\nu^3} \bar{I}_\nu = \frac{c^3}{8\pi h\nu^3} u_\nu \text{ [dimensionless]}, \end{aligned}$$

where the bar denotes averaging over directions. With this definition of n_γ , we can rewrite equations for $(dn_u/dt)_{\ell \rightarrow u}$ and $(dn_\ell/dt)_{u \rightarrow \ell}$ as simply

$$\begin{aligned} \left(\frac{dn_\ell}{dt} \right)_{u \rightarrow \ell} &= n_u A_{u\ell} (1 + \bar{n}_\gamma), \\ \left(\frac{dn_u}{dt} \right)_{\ell \rightarrow u} &= n_\ell \frac{g_u}{g_\ell} A_{u\ell} \bar{n}_\gamma. \end{aligned}$$

If the radiation field depends on frequency in the vicinity of the transition frequency $\nu_{u\ell}$, then n_γ needs to be averaged over the emission and absorption profiles in the above two equations.

From the first of the above two equations, we immediately see that the photon occupation number n_γ determines the relative importance of stimulated and spontaneous emission: stimulated emission is unimportant when $\bar{n}_\gamma \ll 1$, but should otherwise be included in analyses of level excitation.

Absorption cross section: Having determined the rate at which photons are absorbed by an absorber exposed to electromagnetic radiation, it is useful to recast this in terms of an absorption cross section. The photon density per unit frequency is just $u_\nu/h\nu$. Let $\sigma_{\ell u}(\nu)$ be the cross section for absorption of photons of frequency ν with resulting $\ell \rightarrow u$ transition. The absorption rate is then

$$\left(\frac{dn_u}{dt} \right)_{\ell \rightarrow u} = n_\ell \int \sigma_{\ell u}(\nu) c \frac{u_\nu}{h\nu} d\nu \approx n_\ell u_\nu \frac{c}{h\nu} \int \sigma_{\ell u}(\nu) d\nu [\text{s}^{-1}],$$

where we have assumed that u_ν (and $h\nu$) do not vary appreciably over the line profile of $\sigma_{\ell u}$. Thus

$$B_{\ell u} = \frac{c}{h\nu} \int \sigma_{\ell u}(\nu) d\nu [\text{s}^{-1}],$$

and, using our equation that relates $B_{\ell u}$ to $A_{u\ell}$, we obtain the integral over the absorption cross section:

$$\int \sigma_{\ell u}(\nu) d\nu = \frac{g_u}{g_\ell} \frac{c^2}{8\pi \nu_{\ell u}^2} A_{u\ell} [\text{Hz cm}^2].$$

Thus we may relate the monochromatic absorption cross section $\sigma_{\ell u}(\nu)$ to a normalized line profile ϕ_ν :

$$\sigma_{\ell u}(\nu) = \frac{g_u}{g_\ell} \frac{c^2}{8\pi \nu_{\ell u}^2} A_{u\ell} \phi_\nu [\text{cm}^2], \quad \int \phi_\nu d\nu = 1.$$

Intrinsic line profile: The intrinsic line profile is characterized by a normalized profile function ϕ_ν^{intr} :

$$\sigma_{\nu}^{\text{intr}} = \frac{\pi e^2}{m_e c^2} f_{\ell u} \phi_{\nu}^{\text{intr}} [\text{cm}^2], \quad \int \phi_{\nu}^{\text{intr}} d\nu = 1.$$

The intrinsic line profile of an absorption line is normally described by the **Lorentz line profile** function:

$$\phi_{\nu}^{\text{intr}} = \frac{4\gamma_{u\ell}}{16\pi^2(\nu - \nu_{u\ell})^2 + \gamma_{u\ell}^2} [\text{erg s}^{-1} \text{cm}^{-2} \text{sr}^{-1}],$$

where $\nu_{u\ell} \equiv (E_u - E_{\ell})/h$. The Lorentz profile provides an accurate (but not exact)² approximation to the actual line profile. The Lorentz line profile has a **full width at half maximum** (FWHM) of

$$(\Delta\nu)_{\text{FWHM}}^{\text{intr}} = \frac{\gamma_{u\ell}}{2\pi} [\text{Hz}].$$

The intrinsic width of the absorption line reflects the uncertainty in the energies of levels u and ℓ due to the finite lifetimes of these levels³ against transitions to all other levels, including both radiative and collisional transitions. If the primary process for depopulating levels u and ℓ is spontaneous decay (as is often the case in the ISM), then

$$\gamma_{u\ell} \equiv \gamma_{\ell u} = \sum_{E_j < E_u} A_{uj} + \sum_{E_j < E_{\ell}} A_{\ell j}.$$

In the case of a **resonance line**, where ℓ is the ground state, the second sum vanishes.

Doppler broadening; The voigt line profile: Atoms and ions are generally in motion, and the velocity distribution is often approximated by a Gaussian, this being of course the correct form if the velocities are entirely due to thermal motions:

$$p_v = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_v} e^{-(v-v_0)^2/2\sigma_v^2} = \frac{1}{\sqrt{\pi}} \frac{1}{b} e^{-(v-v_0)^2/b^2},$$

where $p_v dv$ is the probability of the velocity along the line of sight being in the interval $[v, v + dv]$, σ_v is the one-dimensional velocity dispersion, and b is the **broadening parameter**, $b \equiv \sqrt{2}\sigma_v$.

The width of the velocity distribution is also sometimes specified in terms of the FWHM; for a Gaussian distribution of velocities, this is just

$$(\Delta v)_{\text{FWHM}} = \sqrt{8 \ln 2} \sigma_v = 2\sqrt{\ln 2} b [\text{m s}^{-1}].$$

If the velocity dispersion is entirely due to thermal motion with kinetic temperature $T = 10^4 T_4$ K, then

$$\begin{aligned} \sigma_v &= \left(\frac{k_B T}{M} \right)^{1/2} = 9.12 \left(\frac{T_4}{M/\text{amu}} \right)^{1/2} [\text{km s}^{-1}] \\ b &= \left(\frac{2k_B T}{M} \right)^{1/2} = 12.90 \left(\frac{T_4}{M/\text{amu}} \right)^{1/2} [\text{km s}^{-1}] \\ (\Delta v)_{\text{FWHM}}^{\text{therm}} &= \left[\frac{(8 \ln 2) k_B T}{M} \right]^{1/2} = 21.47 \left(\frac{T_4}{M/\text{amu}} \right)^{1/2} [\text{km s}^{-1}]. \end{aligned}$$

The intrinsic absorption line profile ϕ_{ν}^{intr} must be convolved with the velocity distribution of the absorbers to obtain the line profile

$$\phi_{\nu} = \int p_v(v) \frac{4\gamma_{u\ell}}{16\pi^2[\nu - (1 - v/c)\nu_{u\ell}]^2 + \gamma_{u\ell}^2} dv [\text{erg s}^{-1} \text{cm}^{-2} \text{sr}^{-1}],$$

where $p_v dv$ is the probability of the absorber having radial velocity in the interval $[v, v + dv]$. If the absorbers have a **Maxwellian** (i.e., Gaussian) one-dimensional velocity distribution p_v , then the absorption line will have a so-called **Voigt line profile**:

$$\phi_{\nu}^{\text{Voigt}} \equiv \frac{1}{\sqrt{2\pi}} \int \frac{e^{-v^2/2\sigma_v^2}}{\sigma_v} p_v(v) \frac{4\gamma_{u\ell}}{16\pi^2[\nu - (1 - v/c)\nu_{u\ell}]^2 + \gamma_{u\ell}^2} dv [\text{erg s}^{-1} \text{cm}^{-2} \text{sr}^{-1}].$$

Unfortunately, the Voigt line profile cannot be obtained analytically except for limiting cases.⁵ However, if, as is generally the case, the one-dimensional velocity dispersion $\sigma_v \gg (\Delta v)_{\text{FWHM}}^{\text{intr}}$, the central core of the line profile is well-approximated by treating the intrinsic line profile as a δ -function, so that the central core of the line has a Maxwellian profile:

²The line profile is more accurately given by the **Kramers-Heisenberg formula**; Lee (2003) discusses application of this formula to the Lyman α line.

³The **Heisenberg uncertainty principle** $\Delta E \Delta t \geq \hbar$ implies that an energy level u has a width $\Delta E_u \approx \hbar/\tau_u$, where τ_u is the level lifetime.

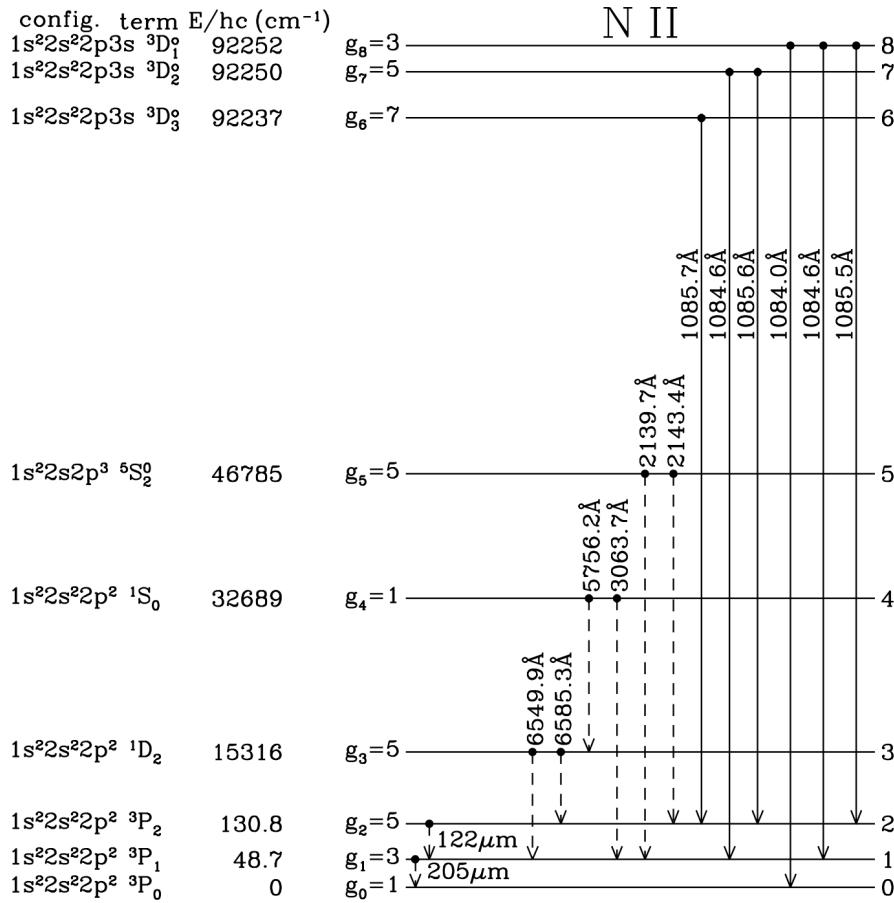


Figure 9: First nine energy levels of NII. Forbidden transitions are indicated by broken lines, and allowed transitions by solid lines; forbidden decays are not shown from levels that have permitted decay channels. Fine-structure splitting is not to scale. Hyperfine splitting is not shown. Figure taken from Draine (2015).

$$\phi_\nu \approx \frac{1}{\pi} \frac{1}{\nu_{ul}} \frac{c}{b} e^{-v^2/b^2} [\text{erg s}^{-1} \text{cm}^{-2} \text{sr}^{-1}], \quad b \equiv \sqrt{2}\sigma_v.$$

Selection rules for radiative transitions: Some energy levels are connected by strong radiative transitions; in other cases, radiative transitions between the levels may be extremely slow. The strong transitions always satisfy what are referred to as the **selection rules for electric dipole transitions**. Here, we summarize the selection rules for the strong electric dipole transitions, and we also give the selection rules for **inter-system** and **forbidden transitions** that do not satisfy the electric dipole selection rules but nevertheless are strong enough to be astrophysically important. We will use the ion NII as an example; the first nine energy levels of NII are shown in Figure 9.

Allowed: Electric dipole transitions: The strongest transitions are electric dipole transitions. These are transitions satisfying the following selection rules:

1. Parity must change.
2. $\Delta L = 0, \pm 1$.
3. $\Delta J = 0, \pm 1$, but $\Delta J = 0 \rightarrow 0$ is forbidden.
4. Only one single-electron wave function $n\ell$ changes, with $\Delta\ell = \pm 1$.
5. $\Delta S = 0$: Spin does not change.

An allowed transition is denoted *without* square brackets, for example,



This is a transition between the $\ell = 1s^2 2s^2 2p^2 \ ^3P_0$ and $u = 1s^2 2s^2 2p3s \ ^3D_1^o$ levels of NII, with a wavelength $\lambda_{ul} = 1084.0 \text{\AA}$. The transition has $A_{ul} = 2.18 \times 10^8 \text{s}^{-1}$. This decay is very fast – the lifetime of the $^3D_1^o$ level against this decay is only $1/A_{ul} = 4.6 \text{ ns}$!

Spin-Forbidden or inter-system transitions: These are transitions that fulfill the electric dipole selection rules 1 to 4 but have $\Delta S \neq 0$. These transitions are considerably weaker than allowed transitions. Such transitions are sometimes referred to as **semi-forbidden**, or **inter-combination**, or **inter-system transitions**; the latter is the terminology that we will use. An inter-system transition is denoted with a single right bracket – for example,

$$[\text{NII}] 2143.4\text{\AA} \ ^3P_2 - ^5S_2^o.$$

This is a transition between the $\ell = 1s^2 2s^2 2p^2 \ 3P_2$ and $u = 1s^2 2s^2 2p^3 \ ^5S_2^o$ levels of NII, with a wavelength $\lambda_{ul} = 2143.4 \text{\AA}$ and $A_{ul} = 1.27 \times 10^2 \text{ s}^{-1}$.

Forbidden transitions: Forbidden transitions are those that fail to fulfill at least one of the selection rules 1 to 4. The transition probabilities vary widely, depending on the values of the electric quadrupole or magnetic dipole matrix elements between the upper and lower states. A forbidden transition is denoted with two square brackets – for example,

$$[\text{NII}] 6549.9\text{\AA} \ ^3P_1 - ^1D_2^o.$$

This is a transition between the $\ell = 1s^2 2s^2 2p^2 \ 3P_1$ and $u = 1s^2 2s^2 2p^2 \ ^1D_2^o$ levels of NII, with a wavelength $\lambda_{ul} = 6549.9 \text{\AA}$ and $A_{ul} = 9.20 \times 10^{-4} \text{ s}^{-1}$. This fails rule 1 (parity is unchanged) and it fails rule 4 (single electron wave functions are unchanged). This is an example of a **magnetic dipole transition**.

Another example of a forbidden transition is the **electric quadrupole transition**

$$[\text{NII}] 5756.2\text{\AA} \ ^1D_2 - ^1S_0^o.$$

This is a transition between the $\ell = 1s^2 2s^2 2p^2 \ 1D_2$ and $u = 1s^2 2s^2 2p^2 \ ^1S_0^o$ levels of NII, with a wavelength $\lambda_{ul} = 5756.2 \text{\AA}$ and $A_{ul} = 1.17 \text{ s}^{-1}$. This fails rules 1 (parity is unchanged) and 4 (single electron wave functions are unchanged) and it fails rules 2 and 3 ($\Delta L = -2$ and $\Delta J = -2$), yet its transition probability is three orders of magnitude larger than the magnetic dipole transition [NII] 6549.9\AA!

We see then that there is a hierarchy in the transition probabilities: very roughly speaking, inter-system lines are $\sim 10^6$ times weaker than permitted transitions, and forbidden lines are $\sim 10^2 - 10^6$ times weaker than inter-system transitions.

Despite being very “weak,” forbidden transitions are important in astrophysics for the simple reason that every atom and ion has excited states that can only decay via forbidden transitions. At high densities, such excited states would be depopulated by collisions, but at the very low densities of interstellar space, collisions are sufficiently infrequent that there is time for forbidden radiative transitions to take place.

1.3.3 Follow-up Questions

- Why aren’t emission and absorption lines delta functions?
- How does this relate to population levels and excitation temperatures?
- Are there emission lines in the Sun? Why is there emission from the Calcium doublet?
- Write down the heat transfer equation. What do solutions look like?

1.4 Question 4

Describe these important sources of stellar opacity: electron scattering, free-free, bound-free, and the H⁻ ion.

1.4.1 Short answer

- **Electron scattering:** Also known as Thomson scattering. When an EM wave passes by an electron, the radiation field makes the electron oscillate and emit dipole radiation in all directions. This process has a small interaction cross section, so it dominates at high electron densities when the temperature is high (e.g., stellar interiors).
- **Free-free:** A free electron in the vicinity of an ion (which is necessary to conserve energy and momentum) will absorb a photon which results in an increased energy and momentum of the electron. This process is a source of continuum absorption.
- **Bound-free:** Also known as photoionization. This occurs when an atom absorbs a photon that exceeds its ionization energy and becomes ionized, resulting in a free electron that takes away any excess energy. This process is a source of continuum absorption.
- **H⁻ ion:** A second electron can become loosely bound to atomic hydrogen resulting in a negatively charged ion.

1.4.2 Additional context

Electron scattering: If an EM wave passes an electron, the electric field makes the electron oscillate. The oscillating electron represents a classical dipole that radiates in other directions, (i.e., the electron scatters part of the energy of the incoming wave). The weakening of the original radiation due to scattering is equivalent to that by absorption, and we can describe it by way of a cross section at frequency ν per unit mass (κ_ν). This can be calculated classically giving the result

$$\kappa_\nu = \frac{8\pi}{3} \frac{e_e^2}{\mu_e m_u} = 0.20(1 + X) [\text{cm}^2 \text{ g}^{-1}],$$

where r_e is the classical electron radius, X the mass fraction of hydrogen, and the constant is in $\text{cm}^2 \text{ g}^{-1}$. The term $\mu_e m_u$ arises because κ_ν is taken per unit mass; and μ_e is replaced by (4.30). Since κ_ν does not depend on the frequency, we immediately obtain the **Rosseland mean for electron scattering**:

$$\kappa_{\text{sc}} = 0.20(1 + X) [\text{cm}^2 \text{ g}^{-1}].$$

The Thomson scattering just described neglects the exchange of momentum between electron and radiation. If this becomes important, then κ_ν will be reduced compared to the value given above, though this effect plays a role only at temperatures sufficiently high for the scattered photons to be very energetic. In fact, during the scattering process the electron must obtain such a large momentum that its velocity is comparable to c , say $v \gtrsim 0.1c$ for the equation above to become a bad approximation. The momentum of the photon is $h\nu/c$, which after scattering is partly transferred to the electron, $m_e v \sim h\nu/c$. Therefore relativistic corrections (**Compton scattering**) become important if the average energy of the photons is $h\nu \gtrsim 0.1 m_e c^2$. For $h\nu$ we take the frequency at which the Planck function has a maximum; then according to Wien's law this is at $h\nu = 4.965 k_B T$, and the full Compton scattering cross section has to be taken into account if $T > 0.1 m_e c^2/(4.965 \text{ K})$, or roughly $T > 10^8 \text{ K}$. In fact even at $T = 10^8 \text{ K}$ Compton scattering reduces the opacity by only 20% of that given by the equation above.

Free-free: If during its thermal motion a free electron passes an ion, the two charged particles form a system which can absorb and emit radiation. This mechanism is only effective as long as the electron and ion are sufficiently close. Now, the mean thermal velocity of the electrons is $v \sim \sqrt{T}$, and the time during which they form a system able to absorb or emit is proportional to $1/v \sim 1/\sqrt{T}$; therefore, if in a mass element the numbers of electrons and ions are fixed, the number of systems temporarily able to absorb is proportional to $1/\sqrt{T}$.

The absorption properties of such a system have been derived classically; the absorption coefficient per system is proportional to $Z^2 \nu^{-3}$, where Z is the charge number of the ion. We therefore expect the absorption coefficient κ_ν of a given mixture of (fully ionized) matter to be

$$\kappa_\nu \sim Z^2 \rho T^{-1/2} \nu^{-3} [\text{cm}^2 \text{ g}^{-1}].$$

Here the factor ρ appears because for a given mass element the probability that two particles are accidentally close together is proportional to the density.

For the determination of the Rosseland mean κ of this absorption coefficient we make use of a simple theorem: a factor ν^α contained in κ_ν gives a factor T^α in κ . With this and the above equation for κ_ν we find

$$\kappa_{ff} \sim \rho T^{-7/2} [\text{cm}^2 \text{g}^{-1}].$$

All opacities of this form are called **Kramers opacities** and give only a classical approximation. One normally multiplies the Kramers formula by a correction factor g , the so-called **Gaunt factor**, in order to take care of the quantum-mechanical correction. In the equation for κ_{ff} , we have still omitted the factor Z^2 which appears in κ_ν . In general, one has a mixture of different ions, and therefore one has to add the contributions of the different chemical species. The (weighted) sum over the values of Z^2 is taken into the constant of proportionality in κ_{ff} , which then depends on the chemical composition. For a fully ionized mixture a good approximation is given by

$$\kappa_{ff} = 3.8 \times 10^{22} (1 + X)[(X + Y) + B] \rho T^{-7/2} [\text{cm}^2 \text{g}^{-1}],$$

with the numerical constant in cgs. The mass fractions of H and He are X and Y , respectively. Here the factor $(1 + X)$ arises, since ν_{ff} must be proportional to the electron density which is proportional to $(1 + X)\rho$. The term $(X + Y)$ in the brackets can be understood in the following way: there are X/m_u hydrogen ions and $Y/(4m_u)$ helium ions. The former have the charge number 1, the latter the charge number 2. But since $\kappa_\nu \sim Z^2$, when adding the contributions of H and He to the total absorption coefficient, we obtain the factor $X/m_u + 4Y/(4m_u) = (XY)m_u$. Correspondingly the term B gives the contribution of the heavier elements:

$$B = \sum_i \frac{X_i Z_i^2}{A_i} [\text{dimensionless}],$$

where the summation extends over all elements higher than helium and A_i is the atomic mass number. **Bound-free:** We first consider a (neutral) hydrogen atom in its ground state, with an ionization energy of χ_0 (i.e., a photon of energy $h\nu > X_0$ can ionize the atom). Energy conservation then demands that

$$h\nu = \chi_0 + \frac{1}{2} m_e v_e^2 [\text{eV}],$$

where v_e is the velocity of the electron released (relative to the ion, which is assumed to be at rest before and after ionization).

If we define an absorption coefficient a_ν per ion ($a_\nu = \kappa_\nu \rho / n_{\text{ion}}$), we expect $a_\nu = 0$ for $\nu < \chi_0/h$ and $a_\nu > 0$ for $\nu \geq \chi_0/h$. Classical considerations similar to those which lead to the Kramers dependence of κ_ν for freefree transitions give a $a_\nu \sim \nu^{-3}$ for $\nu \geq \chi_0/h$. Quantum-mechanical corrections can again be taken into account by a Gaunt factor. But if we have neutral hydrogen atoms in different stages of excitation, the situation is different: an atom in the first excited stage has an absorption coefficient $a_\nu = 0$ for $h\nu < \chi_1$, where χ_1 is the energy necessary to ionize a hydrogen atom from the first excited state, while $a_\nu \sim \nu^{-3}$ for $h\nu \geq \chi_1$. The absorption coefficient κ_ν for a mixture of hydrogen atoms in different states of excitation is a superposition of the a_ν for different stages of excitation. The resulting κ_ν is a saw-tooth function. In order to obtain κ_ν for a certain value of the temperature T , one has to determine the relative numbers of atoms in the different stages of excitation by the Boltzmann formula; then their absorption coefficients a_ν , weighted with their relative abundances, are to be summed.

If there are ions of different chemical species with different degrees of ionization, one has to sum the functions a_ν for all species in all stages of excitation and all degrees of ionization before carrying out the Rosseland integration. An important source of opacity are bound-free transitions of neutral hydrogen atoms, in which case the opacity must be proportional to the number of neutral hydrogen atoms and κ can be written in the form

$$\kappa_{bf} = X(1 - x)\tilde{\kappa}(T) [\text{cm}^2 \text{g}^{-1}].$$

Here $\tilde{\kappa}(T)$ is obtained by Rosseland integration over (weighted) sums of functions a_ν for the different stages of excitation, while x is the degree of ionization.

H⁻ ion: Hydrogen atom has a bound state for a second electron in the field of the proton, though it has a very low ionization potential, $E_{H^-} = 0.75 \text{ eV}$. The number density of negative hydrogen ions will be proportional to the electron density, which, in all but the most metal-poor stars, will be set by ionization of the metals (which have much lower ionization potentials than hydrogen and helium). Thus, the H⁻ opacity will scale as $\kappa_{H^-} \propto \rho X Z$ at low temperatures; H⁻ is of course easily ionized at higher temperatures, and at very low temperatures even metals will not be ionized, so there will be no electrons to form H⁻ by combining with H.

1.5 Question 5

Describe the processes that can cause pulsations in a star's luminosity, and provide at least one example of a class of stellar pulsation.

1.5.1 Short answer

Stellar pulsations can occur when a star's surface is expanding and contracting which causes its surface area, and thus luminosity, to vary: $L = 4\pi R^2 \sigma T^4$. Radial oscillations of a pulsating star, for example, are the result of sound waves resonating in the star's interior. For such radial oscillations, displacement of stellar material occurs primarily on the surface. A star has an infinite number of modes of radial pulsation. The simplest is called the fundamental mode. In this mode, all parts of the star expand together and contract together in unison. The next simplest mode is the first overtone. In this mode, there is a nodal sphere in the star where the material remains at rest. When the part of the star outside this sphere is expanding, the part inside is contracting, and vice versa. In the second overtone mode, there are two nodal spheres, where the material remains at rest. The large-amplitude pulsating stars (Cepheids, RR Lyrae variables, and Mira variables) pulsate primarily in radial modes.

1.5.2 Additional context

Variable stars are stars which change in brightness. The change may be as small as a few parts in a million, or it may be a factor of a thousand or more. It may occur in a second or less, or it may take years, decades, or centuries. These are extremes, but astronomers have developed an array of techniques for discovering, measuring, and analyzing the full range of possible variable stars. Why? Because the variations provide important and often unique information about the nature and evolution of the stars. This information can be used to deduce even more fundamental knowledge about our universe in general. For a star to begin and continue to pulsate, there must be some mechanism which converts radiant energy into energy of motion in a timely way, overcoming friction and building up the motion. Such a mechanism can exist in a star: it is the thermodynamic effect of an ionization zone in a star. In a star of moderate surface temperature, the layers below the surface have temperatures up to a few tens of thousands of degrees. The helium tends to be singly-ionized; it has lost one of its two electrons but not the second. If the star contracts, it heats slightly which causes the helium to lose its second electron. This requires energy which comes from the radiation flowing outward through the gas; it is absorbed by the opacity of the gas. The energy then becomes stored in the form of ionization energy. When the star subsequently expands and cools, the electrons recombine with the helium atoms, and release the stored-up ionization energy. This is the 'push' which maintains the pulsation in all of the pulsating stars in the Cepheid instability strip.

Classification of variable stars: The earliest classification scheme dates back nearly two centuries: Pigott divided variables according to the nature of their light curve into novae, long-period variables, and short-period variables. A century later, Pickering devised a more detailed scheme: (Ia) normal novae, primarily nearby ones in our own galaxy; (Ib) novae in nebulae: now known to be primarily SNe in distant galaxies; (IIa) long-period variables, cool, large-amplitude pulsating variables; (IIb) U Geminorum stars, dwarf novae; (IIc) R Coronae Borealis stars, stars which suddenly and unpredictably decline in brightness; (III) irregular variables, a motley collection; (IVa) short-period variables, such as Cepheids and later including the cluster-type or RR Lyrae stars; (IVb) Beta Lyrae type eclipsing variables; and (V) Algol type eclipsing variables.

Pickering's classification scheme contains some hints as to the nature and cause of the variability. Classification and explanation go hand in hand (in principle), and in the late nineteenth and early twentieth centuries, progress was made in understanding the physical nature and the physical processes in variable stars, and in stars in general.

Pulsating variable stars: Pulsation is the astronomer's word for *vibration or oscillation*. Every physical object has natural patterns or modes of vibration, each with a corresponding period – the time required for one vibration. The radial oscillations of a pulsating star are the result of sound waves resonating in the star's interior. A rough estimate of the pulsation period may be obtained by considering how long it would take a sound wave to cross the diameter of a model star of radius R and constant density ρ . The adiabatic sound speed is given by

$$v_s = \sqrt{\frac{\gamma P}{\rho}} [\text{m s}^{-1}]$$

The pressure may be found using hydrostatic equilibrium along with (again) the assumption of constant density:

$$\frac{dP}{dr} = -\frac{GM_r\rho}{r^2} = -\frac{G(\frac{4}{3}\pi r^3 \rho)\rho}{r^2} = -\frac{4}{3}\pi G\rho^2 r \text{ [Pa m}^{-1}\text{]}$$

This is readily integrated using the boundary condition that $P = 0$ at the surface to obtain the pressure P as a function of radius r :

$$P(r) = \frac{2}{3}\pi G\rho^2(R^2 - r^2) \text{ [Pa].}$$

Thus, the pulsation period is roughly,

$$T \approx 2 \int_0^R \frac{dr}{v_s} \approx 2 \int_0^R \frac{dr}{\sqrt{\frac{2}{3}\gamma G\rho(R^2 - r^2)}} \text{ [yr]},$$

or,

$$T \approx \sqrt{\frac{3\pi}{2\gamma G\rho}} \text{ [yr].}$$

Quantitatively, this shows that the pulsation period of a star is inversely proportional to the square root of its mean density. This **period-mean density relation** explains why the pulsation period decreases as we move down the instability strip from the tenuous supergiants to the very dense WDs.⁴ The tight period-luminosity relation discovered by Leavitt exists because the instability strip is roughly parallel to the luminosity axis of the HR diagram, the finite width of the instability strip being reflected in the ± 0.5 mag uncertainty in the period-luminosity relation.

Pulsation modes: Stars are, to a first approximation, spherical. The simplest form of pulsation is radial pulsation – a simple, spherically symmetric in-and-out expansion and contraction. A star has an infinite number of modes of radial pulsation. The simplest is called the **fundamental mode**. In this mode, all parts of the star expand together and contract together, in unison. The next simplest mode is the **first overtone**. In this mode, there is a nodal sphere in the star, where the material remains at rest. When the part of the star outside this sphere is expanding, the part inside is contracting, and vice versa. In the **second overtone mode**, there are two nodal spheres, where the material remains at rest. Figure 10 shows the fractional displacement, $\delta r/R$, of the stellar material from its equilibrium position for several radial modes of a $12 M_\odot$ main-sequence star. Note that $\delta r/R$ has been arbitrarily scaled to unity at the stellar surface. The large-amplitude pulsating stars – **Cepheids**, **RR Lyrae variables**, and **Mira variables** – pulsate primarily in radial modes. Radial pulsation produces substantial changes in luminosity (and therefore brightness), temperature (and therefore colour), and radial velocity; these are observed, and provide direct evidence for the radial pulsation. Through techniques such as the **Baade-Wesselink method**, they can be used to determine the average radius of the star.

⁴Pulsating WDs exhibit non-radial oscillations, and their periods are longer than predicted by the period-mean density relation.

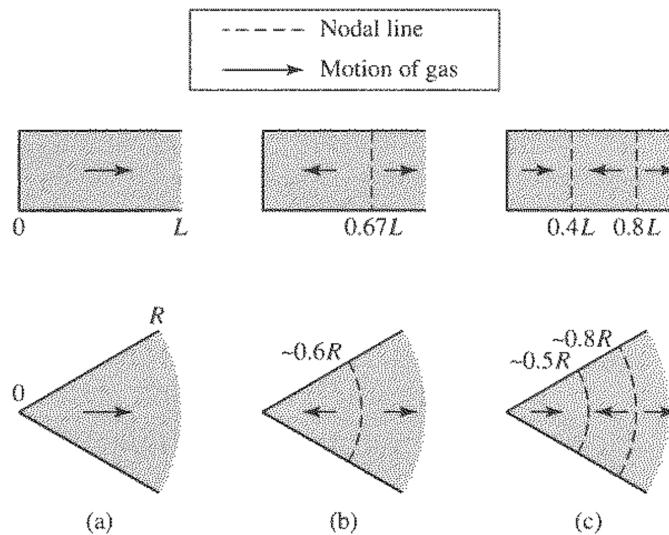


Figure 10: Standing sound waves in an organ pipe and in a star for (a) the fundamental mode, (b) the first overtone, and (c) the second overtone. Figure taken from Carroll & Ostlie (2007).

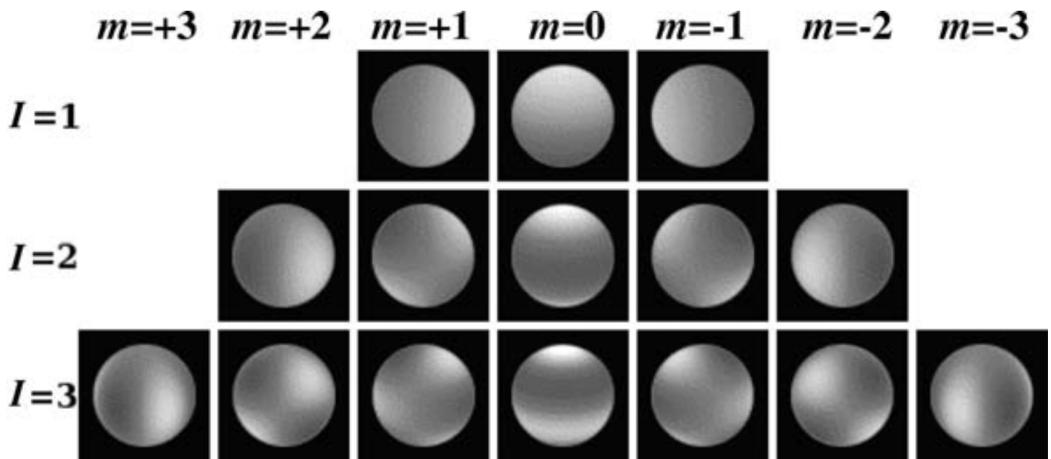


Figure 11: Illustration of non-radial pulsation modes. When the light-coloured regions are moving downward, the dark-coloured regions are moving upward. In radial pulsation, all parts of the star move in and out in unison. (Source: Whole Earth Telescope). Figure taken from Percy (2007).

It is less easy to tell which radial mode a star is pulsating in. Most, but not all, large-amplitude variables pulsate in the fundamental mode. If the radius and mass of the star can be estimated, the observed period can be compared with the theoretical or expected period for each mode. If the star pulsates in two or more radial modes, then the period ratio can be used to deduce the modes.

In **non-radial pulsations**, the star changes shape, not volume (Figure 11). There is a triply-infinite set of modes, corresponding to the three different co-ordinate axes on the star. These are chosen to be: the distance from the centre, the angular distance ('latitude') above or below the star's equator, and the angular distance ('longitude') around the star's equator. Non-radial modes can be divided into (i) **p (pressure) modes**, in which the motion is primarily radial and the restoring force is pressure (as it is in radial modes), and (ii) **g (gravity) modes**, in which the motion is primarily horizontal, and the restoring force is buoyancy or gravity (as it is in water waves). A wide assortment of stars pulsate non-radially. Non-radial pulsation tends to produce smaller amplitudes than radial modes in the brightness and colour variation, but, if they can be observed, their relative amplitudes and phases (along with the periods themselves) can be used to identify the precise non-radial mode. Non-radial pulsation also produces characteristic absorption line profile variations. With modern high S/R spectrographs, these can and have been observed and studied over time. Also, there are often very close non-radial periods, whose differences are a clue to the mode identification. In non-rotating stars, these close periods are actually identical or degenerate, but in a rotating star, they become different; the difference is proportional to the amount of rotation. The situation is complicated by the fact that many of these non-radially pulsating stars rotate rapidly, and the theory of rapidly rotating pulsating stars is not well understood.

Modelling stellar pulsation: The theory of stellar pulsation begins with a model of the star which gives the physical properties of the star as a function of distance from the centre. The first models of stars were primitive (and so were the deductions from stellar pulsation theory). Now, with improved understanding of the physics of stars, and powerful computers to implement this knowledge, the models are much more detailed and accurate.

The simplest level of modelling is the **linear adiabatic theory** (LAT) of radial pulsation of a spherically symmetric, non-rotating, non-magnetic star. This was formulated by Arthur Eddington, early in the twentieth century. It assumes that the pulsations are infinitesimally small, and that there is no transfer of energy between parts of the star. It provides reasonable estimates of the periods of the radial modes and their dependence on the physical properties of the star. For instance, it shows that the period of any mode is approximately inversely proportional to the square root of the mean density of the star. Alternatively, the period times the square root of the density is approximately constant; we call it the **Q-value, or pulsation constant**. It is this relationship which leads to the period-luminosity relation; since both period and luminosity are related to the radius of the star, and since the temperatures of pulsating stars are about the same, period and luminosity will be related to each other. The LAT also shows that the relative amplitude of pulsation tends to be largest in the outer layers, where the density is lowest. Corresponding calculations can be done for non-radial pulsation. The LAT assumes that the pulsation takes the form of a standing wave in the star (i.e., a wave which is perfectly reflected from the centre and the surface).

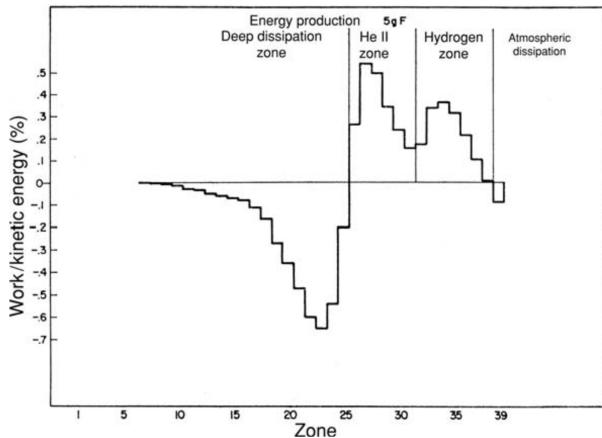


Figure 12: Diagram showing the zones in an unstable star which convert radiant energy into kinetic energy, and vice versa, in a model of a pulsating RR Lyrae star, expressed as the fractional energy production (positive) or dissipation (negative) per period. The centre of the star is at the left, the atmosphere at the right. The two energizing zones (the hydrogen zone and the ionized helium (He II) zone) are shown, along with the principal dissipation zone, deeper in the star. (From Christy, 1966.) Figure taken from Percy (2007).

The next level of complexity is the **linear non-adiabatic theory** (LNAT). This also assumes that the pulsations are infinitesimally small, but it allows for the possibility that any layer of the star may gain or lose energy. In particular, the amplitude of the infinitesimal pulsations can increase or decrease, and we can tell what modes may be unstable, and grow. We cannot tell what the eventual amplitudes may be (i.e., which mode(s) may grow to become observable). The LNAT gives periods which are slightly more realistic than the LAT, though they are about the same in most cases.

The next level of complexity is the **non-linear, non-adiabatic theory**, which amounts to a full, time-dependent, hydrodynamical model of the star. This includes the equation which relates the motion of a layer in the star to the forces acting on it. The calculation begins with the static, ‘equilibrium’ model of the star. Infinitesimal random motions are imposed on the model (such as from the round-off errors in the calculations), to see if they grow or decay. After several hundred cycles, the model (usually) reaches some constant, stable amplitude in one or more modes. The non-linear theory accurately predicts the amplitudes of Cepheids and RR Lyrae stars, which pulsate in one or two radial modes. It has not been able to explain stars like Delta Scuti stars, which can and do pulsate in many radial and non-radial modes. An ongoing problem in stellar pulsation is: what determines what mode(s) a star will pulsate in, and with what amplitude?

These models are non-rotating and non-magnetic. Non-linear analysis of rotating magnetic models is at the frontier of present research. But we know that all stars rotate (some of them very rapidly) and that many of them have significant magnetic fields. In reality, there is no such thing as a truly spherically symmetric star; for most of them, however, that is a reasonable approximation.

Non-linear effects: The pulsation waves in a star are not reflected perfectly from the surface, but move outward in an atmosphere of decreasing density. In many variables, such as Mira stars, this creates **shock waves** as the atmospheric layers pile up against each other. In Mira stars, this creates layers of higher density where gases condense into dust. The dust absorbs the star’s radiation and is pushed outward, carrying gas along with it. The result is mass loss which has a profound effect on the star and the space around it. In the largest, most luminous supergiants, the outward radiation pressure almost balances or perhaps even exceeds gravity; the star borders on instability; and non-linear effects may be significant.

The instability strip(s): Normally, stars are stable. The inward pull of gravity is balanced by the outward pressure of the hot gas in the star. If the star begins to contract, its internal pressure and temperature increase and reverse the contraction. If the star begins to expand, its internal pressure and temperature decrease and gravity restores the initial balance. Furthermore, if a star expands or contracts, the energy of motion is dissipated by friction. Eventually the star returns to rest.

For a star to begin and continue to pulsate, there must be some mechanism which converts radiant energy into energy of motion in a timely way, overcoming friction and building up the motion. Such a mechanism can exist in a star: it is the thermodynamic effect of an ionization zone in a star.

In a star of moderate surface temperature, the layers below the surface have temperatures up to a few tens of thousands of degrees. The helium tends to be singly-ionized; it has lost one of its two electrons but not the second. If the star contracts, it heats slightly which causes the helium to lose its second electron. This requires energy which comes from the radiation flowing outward through the gas; it is absorbed by the opacity of the gas. The energy then becomes stored in the form of ionization energy. When the star subsequently expands and cools, the electrons recombine with the helium atoms, and release the stored-up ionization energy. This is the ‘push’ which maintains the pulsation in all of the pulsating stars in the Cepheid instability strip (Figure 12). If there is sufficient mass in the ionization zone to store up significant radiant energy during the pulsation cycle, then the star will pulsate. This mechanism is

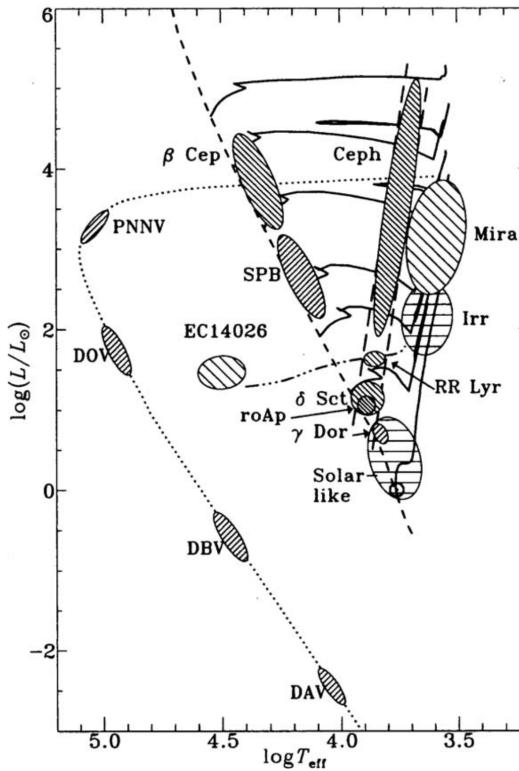


Figure 13: The location of various types of pulsating star on the HR diagram. The main instability regions are the Cepheid strip, including its intersection with the main sequence (dashed line), the region of the coolest stars (Miras and their relatives), and the region of the Beta Cephei and SPB stars. There are also instability regions along the white dwarf cooling track (dotted line) at the bottom of the diagram, including the nuclei of planetary nebulae (PNNV) and the DOV, DBV, and DAV pulsating white dwarfs. (From J. Christensen-Dalsgaard, private communication.) Figure taken from Percy (2007).

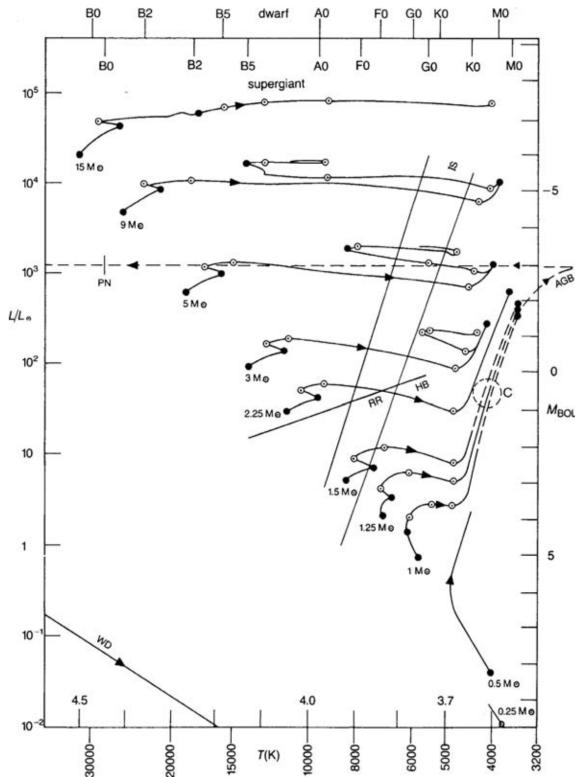


Figure 14: Evolution tracks on the HR diagram, showing the changes which take place in a star as it evolves. The tracks are marked by the mass of the star in solar units. The Cepheid pulsation instability strip is marked IS, and the horizontal branch where RR Lyrae (RR) variables are found is marked HB. The time scales for evolution depend on the mass: the more massive the star, the faster the evolution. (From Kaler, 1997.) Figure taken from Percy (2007).

referred to as **self-exciting**, in contrast to the pulsations of the Sun and other Sun-like stars, which are excited by the turbulence in their convection zones.

But what about the other types of pulsating stars? The Beta Cephei stars were a real mystery, because they are much hotter than the Cepheid instability strip. Only in the 1980s was it realized that the ionization of iron, at temperatures near 150,000 K deep within the star had sufficient effect to maintain the pulsation. This realization came about as a result of physicists' recalculation of the opacity of iron at these temperatures. This solved the long-standing mystery of the Beta Cephei stars, and some other nagging problems in pulsation theory as well.

Figure 13 shows the location of various types of pulsating variable stars in the HR diagram. There is a misconception often included in astronomy textbooks that pulsation is something that happens only in evolved stars such as giants and supergiants. In fact, the Cepheid instability strip extends from the top of the HR diagram to the bottom, at approximately constant temperature. That is because the instability mechanism occurs at a particular temperature – the temperature at which there is a helium ionization zone in the outer layers of the star. So supergiant Cepheids pulsate, main-sequence Delta Scuti stars pulsate, and white dwarfs pulsate – all as a result of the same mechanism.

The Cepheid instability strip is not exactly vertical: its temperature is cooler at its high-luminosity end. That is because more luminous stars have outer regions of lower density; at lower density, ionization can occur at a lower temperature because there is lots of ‘room’ for the electrons that are released.

Stars can evolve into the instability strip in different ways Figure 14. Massive young stars can evolve quickly from left to right in the HR diagram; old low-mass stars can evolve slowly to the same position. So there may be Population I and II stars in the same region of the instability strip.

Stars can evolve into the instability strip in different ways (Figure 14). Massive young stars can evolve quickly from left to right in the HR diagram; old low-mass stars can evolve slowly to the same position. So there may be Population I and II stars in the same region of the instability strip.

1.5.3 Follow-up Questions

- What about the instability strip? RR Lyrae?
- What is the period-luminosity relation?
- What is the form of the period-luminosity relation?
- How would you derive the time scale of pressure waves in a star?
- How would you order-of-magnitude estimate the period for a pulsation?

1.6 Question 6

Briefly describe the sources of thermal energy for stars and planets.

1.6.1 Short answer

The sources of thermal energy of stars include:

- **gravitational energy:** As a star contracts, gravitational potential energy can be converted into thermal energy.
- **rotational energy:** The gas cloud from which stars form acquire angular momentum from their environment which is conserved as the collapse, thus increasing their rotation rate.
- **nuclear energy:** The primary source of energy for main sequence stars. Most of the energy to be gained from nuclear fusion occurs by the conversion of hydrogen to helium and less than one half of that energy can be obtained by all other fusion processes that carry helium to iron.

The sources of thermal energy of planets include:

- **radioactive decay:** The exothermic decay of radioactive isotopes within planet cores.
- **differentiation:** The separation of heavier material towards the core of a molten planet from lighter materials which float towards the surface during planet formation which converts gravitational potential to thermal energy.
- **solar radiation:** The incident radiation on a planet's surface from its host star can heat the surface either via absorption or the greenhouse effect.
- **accretion/impact:** The conversion of kinetic energy to thermal energy upon asteroids etc. hitting the surface of planets; primarily during planet formation (i.e., bombardment).
- **tidal heating:** Tidal heating occurs through the tidal friction processes: orbital energy is dissipated as heat in either the surface ocean or interior of a planet or satellite. When an object is in an elliptical orbit, the tidal forces acting on it are stronger near periapsis than near apoapsis.

1.6.2 Additional context

Stars:

One of the great mysteries of the late nineteenth and early twentieth centuries was the source of the energy required to sustain the luminosity of the sun. By then, the defining Solar parameters of mass, radius, and luminosity were known with sufficient precision to attempt to relate them. For instance, it was clear that if the Sun derived its energy from chemical processes typically yielding less than $10^{12} \text{ erg g}^{-1}$, it could shine no longer than about 10,000 years at its current luminosity. It is said that Lord Kelvin, in noting that the liberation of gravitational energy could only keep the sun shining for about 10 million years, found it necessary to reject Charles Darwin's theory of evolution because there would have been insufficient time for natural selection to provide the observed diversity of species.

Gravitational energy: It is generally conceded that the Sun has shone at roughly its present luminosity for at least the past 2 billion years and has been in existence for nearly 5 billion years. With this in mind, let us begin our study of the sources of stellar energy with an inventory of the stores of energy available to the sun. Perhaps the most obvious source of energy is that suggested by Lord Kelvin, namely gravitation:

$$E_p \leq -\frac{3}{5} \frac{GM^2}{R} [\text{J}].$$

The right-hand side of the inequality is the gravitational potential energy for a uniform density sphere, which provides a sensible upper limit for the energy. Remember that the gravitational energy is considered negative by convention; a rather larger magnitude of energy may be available for a star that is more centrally concentrated than a uniform-density sphere. We may acquire a better estimate of the gravitational potential energy by using the results for a polytrope. Chandrasekhar obtains the following result for the gravitational potential energy of a polytrope:

$$E_p = -\frac{3}{5-n} \frac{GM^2}{R} [\text{J}].$$

For a star in convective equilibrium (that is, $n = 3/2$) the factor multiplying GM^2/R becomes $6/7$ or nearly unity. Note that for a polytrope of index 5, $U_g \rightarrow -\infty$ implying an infinite central concentration of material. This is also one of the polytropes for which there exists an analytic solution. Thus, one has the picture of a mass point surrounded by a massless envelope of infinite extent. This equation also tells us that as the polytropic index increases, so does the central concentration.

It is not at all obvious that the total gravitational energy would be available to permit the star to shine. Some energy must be provided in the form of heat, to provide the pressure which supports the star. We may use the Virial theorem [equation (1.2.35)] to estimate how much of the gravitational energy can be utilized by the luminosity. Consider a star with no mass motions, so that the macroscopic kinetic energy T in equation (1.2.35) is zero. Let us also assume that the equilibrium state is good enough that we can replace the time averages by the instantaneous values. Then the **Virial theorem becomes**

$$2E_k + E_p = 0 \text{ [J].}$$

Recall that E_k is the total internal kinetic energy of the gas which includes all motions of the particles making up the gas. Now we know from thermodynamics that not all the internal kinetic energy is available to do work, and it is therefore not counted in the internal energy of the gas.

For simplicity, let us assume that γ is constant throughout the star. Remembering that the total energy E is the sum of the internal energy and the gravitational energy, we can express the Virial theorem in the following ways:

$$\begin{aligned} E_k &= -\frac{E_p}{3(\gamma - 1)} \text{ [J]} \\ E_{\text{tot}} &= -(3\gamma - 4)E_{\text{th}} \text{ [J]} \\ E_{\text{tot}} &= \frac{3\gamma - 4}{3(\gamma - 1)}E_p \text{ [J].} \end{aligned}$$

It is clear that for $\gamma > 4/3$ (that is, $n < 3$), the total energy of the star will be negative. This simply says that the star is gravitationally bound and can be in equilibrium. So we can look for the physically reasonable polytropes to have indices less than or equal to 3. The case of $n = 3$ is an interesting one for it represents radiation dominated gas. In the limit of complete radiation dominance, the total energy of the configuration will be zero.

Rotational energy: Stars do rotate, and we should not forget to count the rotational energy in the inventory of energies. We may place a reasonable upper limit on the magnitude of the rotational energy that we can expect by noting that (1) the moment of inertia of the star will always be less than that of a sphere of uniform density and (2) there is a limit to the angular velocity ω_c at which the star can rotate. Thus, for a centrally condensed star

$$\begin{aligned} \omega^2 &= \frac{8GM}{27R^3} \text{ [km}^2 \text{s}^{-2} \text{kpc}^{-2}\text]} \\ I_z &< \frac{2}{5}MR^2 \text{ [kg km}^2\text]}, \end{aligned}$$

which implies that the rotational energy must be bounded by

$$E_{\text{rot}} = \frac{1}{2}I_z\omega^2 < \frac{8}{135}\frac{GM^2}{R} \text{ [J].}$$

One may quibble that we have used the angular velocity limit for a centrally condensed star and the moment of inertia for a uniform-density star, but the fact remains that it is extremely difficult for a star to have more than about 10 percent of the magnitude of its gravitational energy stored in the form of rotational energy.

Nuclear energy: Of course, the ultimate upper limit for stored energy is the energy associated with the rest mass itself. It is also the common way of estimating the energy available from nuclear sources. Indeed, that fraction of the rest mass which becomes energy when four hydrogen atoms are converted to one helium atom provides the energy to sustain the solar luminosity. Below in Table 2 is a short table giving the mass loss for a few common elements involved in nuclear fusion processes.

Clearly most of the energy to be gained from nuclear fusion occurs by the conversion of hydrogen to helium and less than one-half of that energy can be obtained by all other fusion processes that carry helium to iron. Nevertheless, 0.7% of mc^2 is a formidable supply of energy. Table 2 is a summary of the energy that one could consider as being available to the Sun. All these entries are generous upper limits. For example, the sun rotates at less than 0.5% of its critical velocity, it was never composed of 100% hydrogen and will begin to change significantly when a fraction of the core hydrogen is consumed, and not all the gravitational energy could ever be converted to energy for release. In any event, only nuclear processes hold the promise of providing the solar luminosity for the time required to bring about

Form of Energy	Amount, erg	$\%M_{\odot}$	Lifetime for constant L_{\odot} , years
Chemical (10^{12} erg/g)	1.98×10^{45}	10^{-7}	14,000
Rotational ($8 \Omega/81$)	2.21×10^{47}	10^{-5}	1.4 million
Gravitational [$3GM^2/(5R)$]	2.24×10^{48}	10^{-4}	14 million
Nuclear ($0.0106M_{\odot}c^2$)	1.89×10^{52}	1	140 billion
Rest mass energy ($M_{\odot}c^2$)	1.78×10^{54}	100	1.4×10^{13}

Table 2: Possible sources of Solar energy. Table taken from Kippenhahn, Weigert & Weiss (2012).

agreement with the age of the solar system as derived from rocks and meteorites. However, the time scales of Table 2 are interesting because they provide an estimate of how long the various energy sources could be expected to maintain some sort of equilibrium configuration.

Planets:

1.6.3 Follow-up Questions

- Why do different nuclear reaction pathways have different temperature sensitivities?
- If I assume a constant core temperature on the main sequence, how does stellar radius depend on mass?
- What are some other thermal sources, like say for neutron stars?

1.7 Question 7

Describe the process by which supernovae produce light. Why are Type Ia supernovae generally brighter than Type II events?

1.7.1 Short answer

Type I SNe were identified first as those SNe that do not exhibit any hydrogen lines in their spectra. Given that hydrogen is the most abundance element in the Universe, this fact alone suggests something unusual about these objects. Conversely, the spectra of Type II SNe contain strong hydrogen lines. Type I SNe can be further subdivided according to their spectra. Those Type I spectra that show a strong Si II line at 615 nm are called **Type Ia**. The others are designated **Type Ib** or **Type Ic**, based on the presence (Ib) or absence (Ic) of strong helium lines. Type II SNe can also be further sub-divided into either **Type II-P** (plateau) or **Type II-L** (linear). Type Ia SNe are most likely the explosions of WDs while Type II SNe and SNe Ib,c are the final stages in the evolution of massive ($\gtrsim 8 M_{\odot}$) stars. The following describes the main sources of light production in SNe:

- **Type Ia SNe:** Two general scenarios have been proposed to explain the mechanism(s) that trigger these explosions: the **double degenerate model** in which two WDs exist in a binary orbit and a **single degenerate** model in which an evolving star orbits a WD. One version of the single degenerate models suggests that as the material from the secondary falls onto the primary the helium in the gas will settle on top of the CO WD, becoming degenerate. When enough helium has accumulated, a helium flash will occur. Not only will this cause the helium to burn to C and O, but it will also send a shock wave down into the degenerate CO WD, causing ignition of the degenerate C and O. A second version of the single degenerate models doesn't invoke degenerate helium burning on the surface but simply has C and O igniting in the interior of the WD as the star nears the Chandrasekhar mass, at this point the degenerate gas is no longer able to support the mass of the star. As the star approaches the fatal limit, two- and three-dimensional simulations suggests that multiple, independent ignition points may occur deep within the core, resulting in non-spherical events.
- **Type II (core-collapse) SNe:** A Type II-P SN is the most common type of core-collapse SN, occurring about ten times as often as Type II-L SNe. The source of the plateau in Type II-P light curves is due largely to the energy deposited by the shock into the hydrogen-rich envelope. The gas, which was ionized by the shock, enters a stage of prolonged recombination, releasing the energy at a nearly constant temperature of about 5000 K. The plateau may be supported further by the energy deposited in the envelope by the radioactive decay of ^{56}Ni that was produced by the shock front (the half life being $\tau_{1/2} = 6.1$ days). It is expected that the explosive nucleosynthesis of the SN shock should have produced significant amounts of other radioactive elements as well, such as ^{57}Co ($\tau_{1/2} = 271$ days), ^{22}Na ($\tau_{1/2} = 2.6$ years), and ^{44}Ti ($\tau_{1/2} = 247$ years). If these isotopes are present in significant quantities, they may also contribute to the light curve, causing the slope of the curve to change.

1.7.2 Additional context

Classification of supernovae: Based on their spectral properties, SNe are divided into several classes. SNe of Type I do not show any Balmer lines of hydrogen in their spectrum, in contrast to those of Type II. The Type I SNe are further subdivided: SNe Ia show strong emission of SiII $\lambda 6150\text{\AA}$ whereas no SiII at all is visible in spectra of Type Ib,c. Our current understanding of the supernova phenomenon differs from this spectral classification.⁵ Following various observational results and also theoretical analyses, we are confident today that SNe Ia are a phenomenon which is intrinsically different from the other supernova types. For this interpretation, it is of particular importance that SNe Ia are found in all types of galaxies, whereas we observe SNe II and SNe Ib,c only in spiral and irregular galaxies, and here only in those regions in which blue stars predominate. The stellar population in elliptical galaxies consists almost exclusively of old stars, while spirals also contain young stars. From this observational fact it is concluded that the phenomenon of SNe II and SNe Ib,c is linked to a young stellar population, whereas SNe Ia occur also in older stellar populations.

⁵This notation scheme (Type Ia, Type II, and so on) is characteristic for phenomena that one wishes to classify upon discovery, but for which no physical interpretation is available at that time. Other examples are the spectral classes of stars which are not named in alphabetical order nor according to their mass on the main sequence; or the division of Seyfert galaxies into Type 1 and Type 2. Once such a notation is established, it often becomes permanent even if a later physical understanding of the phenomenon suggests a more meaningful classification.

Core-collapse supernovae: SNe II and SNe Ib,c are the final stages in the evolution of massive ($\gtrsim 8 M_{\odot}$) stars. Inside these stars, ever heavier elements are generated by nuclear fusion: once all the hydrogen in the inner region is used up, helium will be burned, then carbon, oxygen, etc. This chain comes to an end when the iron nucleus is reached, the atomic nucleus with the highest binding energy per nucleon. After this no more energy can be gained from fusion to heavier elements so that the pressure, which is normally balancing the gravitational force in the star, can no longer be maintained. The star then collapses under its own gravity. This gravitational collapse proceeds until the innermost region reaches a density about three times the density of an atomic nucleus. At this point the so-called **rebounce** occurs: a shock wave runs towards the surface, thereby heating the infalling material, and the star explodes. In the center, a compact object probably remains – a **neutron star** or, possibly, depending on the mass of the iron core, a **black hole**. Such neutron stars are visible as pulsars at the location of some historically observed SNe, the most famous of which is the Crab pulsar which has been identified with a supernovae explosion seen by Chinese astronomers in 1054. Presumably all neutron stars have been formed in such core-collapse supernovae.

The major fraction of the binding energy released in the formation of the compact object is emitted in the form of neutrinos: about 3×10^{53} erg. Underground neutrino detectors were able to trace about 10 neutrinos originating from SN 1987A in the LMC. Due to the high density inside the star after the collapse, even neutrinos, despite their very small cross section, are absorbed and scattered, so that part of their outward-directed momentum contributes to the explosion of the stellar envelope. This shell expands at $v \sim 10000 \text{ km s}^{-1}$, corresponding to a kinetic energy of 10^{51} erg. Of this, only about 10^{49} erg is converted into photons in the hot envelope and then emitted – the energy of a SN that is visible in photons is thus only a small fraction of the total energy produced.

Owing to the various stages of nuclear fusion in the progenitor star, the chemical elements are arranged in shells: the light elements (H, He) in the outer shells, and the heavier elements (C, O, Ne, Mg, Si, Ar, Ca, Fe, Ni) in the inner ones – see Figure 15. The explosion ejects them into the ISM which is thus chemically enriched. It is important to note that mainly nuclei with an even number of protons and neutrons are formed. This is a consequence of the nuclear reaction chains involved, where successive nuclei in this chain are obtained by adding an α -particle (or ${}^4\text{He}$ nucleus; i.e., two protons and two neutrons). Such elements are therefore called α -elements. The dominance of α -elements in the chemical abundance of the ISM, as well as in the Solar System (see Figure 16), is thus a clear indication of nuclear fusion occurring in the He-rich zones of stars where the hydrogen has been burnt.

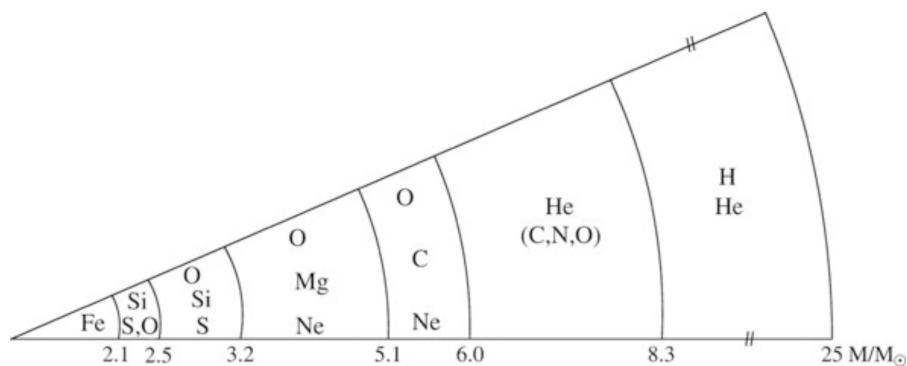


Figure 15: Chemical shell structure of a massive star at the end of its life with the axis labeled by the mass within a given radius. The elements that have been formed in the various stages of the nuclear burning are ordered in a structure resembling that of an onion, with heavier elements being located closer to the center. This is the initial condition for a supernova explosion. Adapted from A. Unsöld & B. Baschek, *The New Cosmos*, Springer-Verlag. Figure taken from Schneider (2006).

Type Ia supernovae: SNe Ia are most likely the explosions of WDs. These compact stars which form the final evolutionary stages of less massive stars no longer maintain their internal pressure by nuclear fusion. Rather, they are stabilized by the degeneracy pressure of the electrons – a quantum mechanical phenomenon related to the Fermi exclusion principle. Such a white dwarf can be stable only if its mass does not exceed a limiting mass, the **Chandrasekhar mass**; it has a value of $M_C \approx 1.44 M_{\odot}$. For $M > M_C$, the degeneracy pressure can no longer balance the gravitational force.

A WD can become unstable if its mass approaches the Chandrasekhar mass limit. There are two different scenarios with which this is possible: If the white dwarf is part of a close binary system, matter from the companion star may flow onto the WD; this is called the '**single-degenerate**' model. In this process, its mass will slowly increase and approach the limiting mass. At about $M \sim 1.3 M_{\odot}$, carbon burning will ignite in its interior, transforming about half of the star into iron-group elements (i.e., iron, cobalt, and

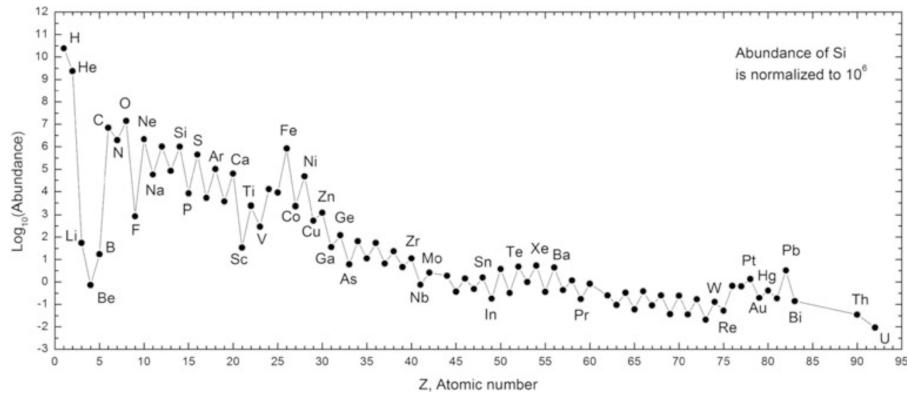


Figure 16: The relative abundance of chemical elements in the Solar System, normalized such that silicon attains the value 10^6 . As a general trend, the abundances decrease with increasing atomic number, except for the light elements lithium (Li), beryllium (Be), and boron (B), which are generated in stars, but also easily destroyed due to their low binding energy. Superposed on this decrease, the abundances show an oscillating behavior: nuclei with an even number of protons are more abundant than those with an odd atomic number – this phenomenon is due to the production of alpha elements in core-collapse supernovae. Furthermore, iron (Fe), cobalt (Co) and nickel (Ni) stick out in their relatively high abundance, given their atomic number, which is due to their abundant production mainly in Type Ia SNe. Source: Wikipedia, numerical data from: Katharina Lodders. Figure taken from Schneider (2006).

nickel). The resulting explosion of the star will enrich the ISM with $0.6 M_{\odot}$ of Fe, while the WD itself will be torn apart completely, leaving no remnant star. A second scenario (so-called ‘double-degenerate’) for the origin of SNe Ia is that of the merger of two WDs for which the sum of their masses exceeds the Chandrasekhar mass. Of course, these two scenarios are not mutually exclusive, and both routes may be realized in nature.

This interpretation of the different types of SNe explains why one finds core-collapse SNe only in galaxies in which star formation occurs. They are the final stages of massive (i.e., young) stars which have a lifetime of not more than 2×10^7 yr. By contrast, SNe Ia can occur in all types of galaxies, since their progenitors are members of an old stellar population.

Since the initial conditions are probably very homogeneous for the class of SNe Ia in the single-degenerate scenario (defined by the limiting mass prior to the trigger of the explosion), they are good candidates for standard candles: all SNe Ia have approximately the same luminosity. This is not entirely the case, but nevertheless SNe Ia play a very important role in the cosmological distance determination, and thus in the determination of cosmological parameters. On the other hand, in the double-degenerate scenario, the class of SNe Ia is not expected to be very homogeneous, as the mass prior to the explosion no longer attains a universal value. In fact, there are some SNe Ia which are clearly different from the majority of this class, by being far more luminous. It may be that such events are triggered by the merging of two white dwarfs, whereas the majority of the explosions is caused by the single-degenerate formation process.

Since Type Ia supernovae are supposed to be the result of explosion processes of white dwarfs which cross a critical mass threshold by accretion of additional matter, this threshold should be identical for all SNe Ia making it at least plausible that they all have the same luminosity. If this were the case, they would be ideal for standard candles: owing to their high luminosity, they can be detected and examined even at very large distances.

However, it turns out that SNe Ia are not really standard candles, since their maximum luminosity varies from object to object with a dispersion of about 0.4 mag in the blue band light. This is visible in the top panel of Figure 17. If SNe Ia were standard candles, the data points would all be located on a straight line, as described by the Hubble law. Clearly, deviations from the Hubble law can be seen, which are significantly larger than the photometric measurement errors. It turns out that there is a strong correlation between the luminosity and the shape of the light curve of SNe Ia. Those of higher maximum luminosity show a slower decline in the light curve, as measured from its maximum. Furthermore, the observed flux is possibly affected by extinction in the host galaxy, in addition to the extinction in the MW. With the resulting reddening of the spectral distribution, this effect can be derived from the observed colors of the SN. The combined analysis of these effects provides a possibility for deducing an empirical correction to the maximum luminosity from the observed light curves in several filters, accounting both for the relation of the width of the curve to the observed luminosity and for the extinction. This correction was calibrated on a sample of SNe Ia for which the distance to the host galaxies is very accurately known. With this correction applied, the SNe Ia follow the Hubble law much more closely, as can be seen in the bottom panel of Figure 17. A scatter of only 0.15 mag around the Hubble relation remains. Figure 18

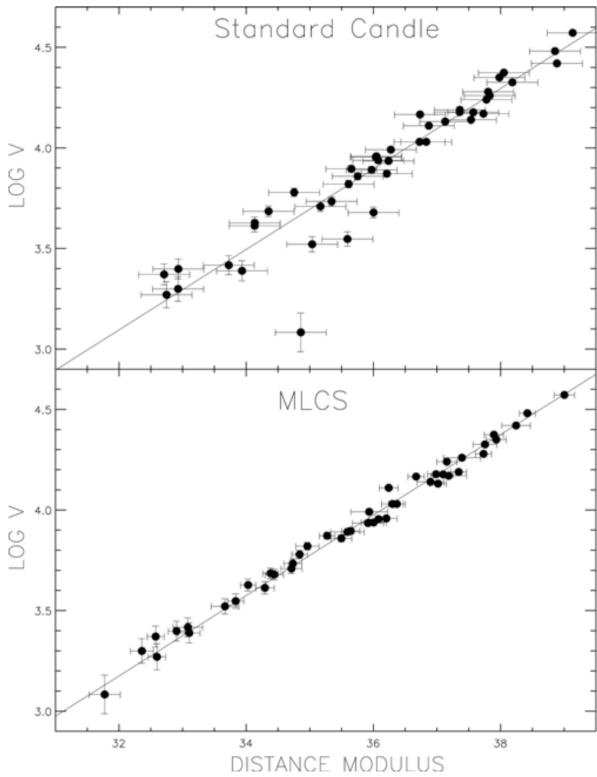


Figure 17:
The Hubble diagram for relatively nearby SNe Ia. Plotted is the measured expansion velocity cz as a function of the distance modulus for the individual supernovae. In the top panel, it is assumed that all sources have the same luminosity. In the bottom panel, the luminosities have been corrected by means of the so-called MLCS method in which the shape of the light curve and the colors of the SN are used to ‘standardize’ the luminosity. By this the deviations from the Hubble law become dramatically smaller (the dispersion is reduced from 0.42 to 0.15 mag). Source: A.V. Filippenko & A.G. Riess 2000, Evidence from Type Ia Supernovae for an Accelerating Universe, astro-ph/0008057, p. 5, Fig. 1. Figure taken from Schneider (2006).

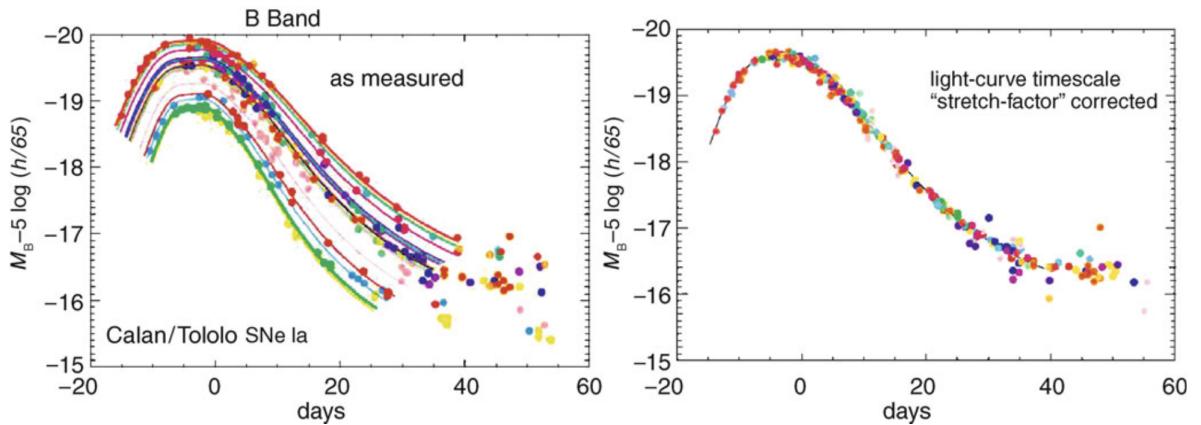


Figure 18: **Left panel:** B-band light curves of different SNe Ia. One sees that the shape of the light curves and the maximum luminosity of the SNe Ia differ substantially among the sample. A transformation was found empirically with a single parameter described by the width of the light curve. By means of this transformation, the different light curves can all be made congruent, as displayed in the **right panel**. Credit: M. Hamuy, S. Perlmutter, Supernova Cosmology Project. Figure taken from Schneider (2006).

demonstrates the effect of this correction on the light curves of several SNe Ia which initially appear to have very different maximum luminosities and widths. After correction they become nearly identical. The left panel of Figure 18 suggests that the light curves of SN Ia can basically be described by a one-parameter family of functions, and that this parameter can be deduced from the shape, in particular the width, of the light curves.

1.7.3 Follow-up Questions

- When a star goes supernova, how much of the luminous energy generated at the rebound is available for heating the gas? As in, where does the heat come from?
- Are there cases where the rebound shock wave can't blow up the star? Why, and what happens then?

1.8 Question 8

Describe the condition for a star's envelope to become convective. Why are low mass stars convective in their outer envelopes while high mass stars are convective in their inner cores?

1.8.1 Short answer

In general, convection will occur when

1. the stellar opacity is large, implying that an unachievable steep temperature gradient would be necessary for radiative transport;
2. a region exists where ionization is occurring, causing a large specific heat and a low adiabatic temperature gradient; and
3. the temperature dependence of the nuclear energy generation rate is large, causing a steep radiative flux gradient and a large temperature gradient.

In the atmosphere of many stars, the first two conditions can occur simultaneously, whereas the third condition would occur only deep in stellar interiors. In particular, the third condition can occur when the highly temperature-dependent CNO cycle or triple alpha processes are occurring.

The Sun is purely radiative below $r/D_\odot = 0.714$ and becomes convective above that point. Physically this occurs because the opacity in the outer portion of the Sun becomes large enough to inhibit the transport of energy by radiation. Stars only slightly more massive than the Sun are convective in their centers because of the stronger temperature dependence of the CNO cycle as compared to the pp chain.

1.8.2 Additional context

Three different energy transport mechanisms operate in stellar interiors. **Radiation** allows the energy produced by nuclear reactions and gravitation to be carried to the surface via photons, the photons being absorbed and re-emitted in nearly random directions as they encounter matter. This suggests that the opacity of the material must play an important role, as one would expect. **Convection** can be a very efficient transport mechanism in many regions of a star, with hot, buoyant mass elements carrying excess energy outward while cool elements fall inward. Finally, **conduction** transports heat via collisions between particles. Although conduction can play an important role in some stellar environments, it is generally insignificant in most stars throughout the majority of their lifetimes.

Radiative transport of energy: basic estimates Rough estimates show important features of the radiative transfer in stellar interiors and justify an enormous simplification of the formalism. Let us first estimate the mean free path ℓ_{mfp} of a photon at an “average” point inside a star like the Sun:

$$\ell_{\text{mfp}} = \frac{1}{\kappa\rho} [\text{m}].$$

where κ is a mean absorption coefficient (i.e., a radiative cross section per unit mass averaged over frequency). Typical values for stellar material are of order $\kappa \approx 1 \text{ cm}^2 \text{ g}^{-1}$; for the ionized hydrogen in stellar interiors, a lower limit is certainly the value for electron scattering, $\kappa \approx 0.4 \text{ cm}^2 \text{ g}^{-1}$. Using this and the mean density of matter in the Sun, $\bar{\rho}_\odot = 1.4 \text{ g cm}^{-3}$, we obtain a mean free path of only

$$\ell_{\text{mfp}} = \frac{1}{\kappa\rho} = \frac{1}{1 \text{ cm}^2 \text{ g}^{-1} \cdot 0.4 \text{ g cm}^{-3}} \approx 2 [\text{cm}],$$

i.e., stellar matter is very opaque.

The typical temperature gradient in the star can be roughly estimated by averaging between centre ($T_c \approx 10^7 \text{ K}$) and surface ($T_0 \approx 10^4 \text{ K}$):

$$\frac{\Delta T}{\Delta r} \approx \frac{T_c - T_0}{R_\odot} \approx 1.4 \times 10^{-4} [\text{K cm}^{-1}].$$

The radiation field at a given point is emitted from a small, nearly isothermal surrounding, the differences of temperature being only of order $\Delta T = \ell_{\text{mfp}}(dT/dr) \approx 3 \times 10^{-4} \text{ K}$. Since the energy density of radiation is $u \sim T^4$, the relative anisotropy of the radiation at a point with $T = 10^7 \text{ K}$ is $\Delta T/T \sim 10^{-10}$. The situation in stellar interiors must obviously be very close to TE, and the radiation very close to that of a blackbody. Nevertheless, the small remaining anisotropy can easily be the carrier of the stars' huge luminosity: this fraction of 10^{-10} of the flux emitted from 1 cm^2 of a blackbody of $T = 10^7 \text{ K}$ is still 10^3 times larger than the flux at the solar surface ($6 \times 10^{10} \text{ erg cm}^{-2} \text{ s}^{-1}$). Radiative transport of energy occurs via the non-vanishing net flux (i.e., via the surplus of the outwards-going radiation emitted from somewhat hotter material below over the inwards-going radiation emitted from less-hot material above).

Diffusion of radiative energy: The above estimates have shown that for radiative transport in stars, the mean free path ℓ_{mfp} of the “transporting particles” (i.e., photons) is very small compared to the characteristic length R (stellar radius) over which the transport extends: $\ell_{\text{mfp},\odot}/R_\odot \approx 3 \times 10^{-11}$. In this case, the transport can be treated as a diffusion process, which yields an enormous simplification of the formalism. We derive the corresponding equation by analogy to those for particle diffusion.

The diffusive flux j of particles (per unit area and time) between places of different particle density n is given by

$$\vec{j} = -D\nabla n,$$

where D is the coefficient of diffusion,

$$D = \frac{1}{3}v\ell_{\text{mfp}} [\text{s}^{-1}],$$

determined by the average values of mean velocity v and mean free path ℓ_{mfp} of the particles.

In order to obtain the corresponding diffusive flux of radiative energy \vec{F} ; we replace n by the energy density of radiation U ;

$$u = aT^4 = \frac{4\sigma}{c}T^4 [\text{erg cm}^{-3}].$$

Here, $a = 7.57 \times 10^{-15} \text{ erg cm}^3 \text{ K}^{-4}$ is the **radiation density constant**. Owing to the spherical symmetry of the problem, \vec{F} has only a radial component $F_r = |\vec{F}| = F$ and ∇U reduces to the derivative in the radial direction

$$\frac{\partial u}{\partial r} = 4aT^3 \frac{\partial T}{\partial r} [\text{erg cm}^{-4}].$$

Using this result along with the equation for \vec{j} , this immediately gives us

$$F = -\frac{4ac}{3} \frac{T^3}{\kappa\rho} \frac{\partial T}{\partial r} [\text{erg s}^{-1} \text{ cm}^{-2}].$$

Note that this can be interpreted formally as an **equation for heat conduction** by writing

$$\vec{F} = -k_{\text{rad}} \nabla T [\text{erg s}^{-1} \text{ cm}^{-2}],$$

where

$$k_{\text{rad}} = \frac{4ac}{3} \frac{T^3}{\kappa\rho} [\text{erg cm}^{-1} \text{ K}^{-1} \text{ s}^{-1}]$$

represents the **coefficient of conduction** for this radiative transport.

We solve F for the gradient of the temperature and replace it by the usual local luminosity $L = 4\pi r^2 F$; then

$$\frac{\partial T}{\partial r} = -\frac{3}{16\pi ac} \frac{\kappa\rho L}{r^2 T^3} [\text{K cm}^{-1}]$$

Of course, this neat and simple equation becomes invalid when one approaches the surface of the star. Because of the decreasing density, the mean free path of the photons will there become comparable with (and finally larger than) the remaining distance to the surface; hence the whole diffusion approximation breaks down, and one has to solve the far more complicated full set of transport equations for radiation in the stellar atmosphere (these equations indeed yield our simple diffusion approximation as the proper limiting case for large optical depths).

The Rosseland mean for κ_ν : The above equations are independent of the frequency ν ; F and L are quantities integrated over all frequencies, so that the quantity must represent a “proper mean” over ν . We shall now prescribe a method for this averaging.

In general the absorption coefficient depends on the frequency ν . Let us denote this by adding a subscript ν to all quantities that thus become frequency dependent: κ_ν , ℓ_ν , D_ν , u_ν etc.

For the diffusive energy flux \vec{F}_ν of radiation in the interval $[\nu, \nu + d\nu]$, we write now

$$\begin{aligned}\vec{F}_\nu &= -D_\nu \nabla u_\nu [\text{erg s}^{-1} \text{ cm}^{-2}], \\ D_\nu &= \frac{1}{3}c\ell_\nu = \frac{c}{3\kappa_\nu\rho} [\text{m}^2 \text{ s}^{-1}],\end{aligned}$$

while the energy density in the same interval is given by

$$u_\nu = \frac{4\pi}{c} B(\nu, T) = \frac{8\pi h}{c^3} \frac{\nu^3}{e^{h\nu/k_B T} - 1} [\text{erg cm}^{-3}].$$

$B(\nu, T)$ denotes here the **Planck function** for the intensity of blackbody radiation (differing from the usual formula for the energy density simply by the factor $4\pi/c$). From this equation for u_ν , we have

$$\nabla u_\nu = \frac{4\pi}{c} \frac{\partial B}{\partial T} \nabla T,$$

which together with D_ν is inserted into F_ν , the latter then being integrated over all frequencies to obtain the total flux F :

$$\vec{F} = - \left[\frac{4\pi}{3\rho} \int_0^\infty \frac{1}{\kappa_\nu} \frac{\partial B}{\partial T} d\nu \right] \nabla T \text{ [m}^2 \text{s}^{-1}\text{].}$$

We have thus regained $\vec{F} = -k_{\text{rad}} \nabla T$ with

$$k_{\text{rad}} = \frac{4\pi}{3\rho} \int_0^\infty \frac{1}{\kappa_\nu} \frac{\partial B}{\partial T} d\nu \text{ [erg cm}^{-1} \text{K}^{-1} \text{s}^{-1}\text{].}$$

Equating this expression for k_{rad} with that in the averaged form, we have immediately the proper formula for averaging the absorption coefficient:

$$\frac{1}{\kappa} = \frac{\pi}{acT^3} \int_0^\infty \frac{\partial B}{\partial T} d\nu \text{ [cm}^2 \text{g}^{-1}\text{].}$$

This is the so-called **Rosseland mean**. Since

$$\int_0^\infty \frac{\partial B}{\partial T} d\nu = \frac{acT^3}{\pi},$$

the Rosseland mean is formally the harmonic mean of with the weighting function $\partial B/\partial T$, and it can simply be calculated, once the function κ_ν is known from atomic physics.

In order to see the physical interpretation of the Rosseland mean, we rewrite $\vec{F}_\nu = -D_\nu \nabla u_\nu$ as

$$\vec{F}_\nu = - \left(\frac{1}{\kappa_\nu} \frac{\partial B(\nu, T)}{\partial T} \right) \frac{4\pi}{3\rho} \nabla T \text{ [m}^2 \text{s}^{-1}\text{].}$$

This shows that, for a given point in the star (ρ and ∇T given), the integrand in $1/\kappa$ is at all frequencies proportional to the net flux \vec{F}_ν of energy. The Rosseland mean therefore favours the frequency ranges of maximum energy flux. One could say that an average *transparency* is evaluated rather than an *opacity* – which is plausible, since it is to be used in an equation describing the transfer of energy rather than its blocking.

One can also easily evaluate the frequency where the weighting function $\partial B/\partial T$ has its maximum. One finds that, for given a temperature, $\partial B/\partial T \sim x^4 e^x (e^x - 1)^{-2}$ with $x = h\nu/k_B T$. Differentiation with respect to x shows that the maximum of $\partial B/\partial T$ is close to $x = 4$.

The way we have defined the Rosseland mean κ , which is a kind of weighted harmonic mean value, has the uncomfortable consequence that the opacity κ of a mixture of two gases having the opacities κ_1, κ_2 is not the sum of the opacities: $\kappa \neq \kappa_1 \kappa_2$. Therefore, in order to find κ for a mixture containing the weight fractions X of hydrogen and Y of helium, the mean opacities of the two single gases are of no use. Rather one has to add the frequency-dependent opacities $\kappa_\nu = X_{\kappa_\nu, \text{H}} + Y_{\kappa_\nu, \text{He}}$ before calculating the Rosseland mean. For any new abundance ratio X/Y the averaging over the frequency has to be carried out separately.

In the above we have characterized the energy flux due to the diffusion of photons by F_ν . Since in the following we shall encounter other mechanisms for energy transport, from now on we shall specify this radiative flux by the vector \vec{F}_{rad} . Correspondingly we shall use κ_{rad} instead of κ , etc.

Conductive transport of energy: In heat conduction, energy transfer occurs via collisions during the random thermal motion of the particles (electrons and nuclei in completely ionized matter, otherwise atoms or molecules). A basic estimate shows that in “ordinary” stellar matter (i.e., in a non-degenerate gas), conduction has no chance of taking over an appreciable part of the total energy transport. Although the collisional cross sections of these charged particles are rather small at the high temperatures in stellar interiors ($10^{-18} - 10^{-20} \text{ cm}^2$ per particle), the large density ($\rho = 1.4 \text{ g cm}^{-3}$ in the Sun) results in mean free paths several orders of magnitude less than those for photons; and the velocity of the particles is only a few per cent of c . Therefore the coefficient of diffusion is much smaller than that for photons.

The situation becomes quite different, however, for the cores of evolved stars where the electron gas is highly degenerate. The density can be as large as 10^6 g cm^{-3} . But degeneracy makes the electrons

much faster, since they are pushed up close to the Fermi energy; and degeneracy increases the mean free path considerably, since the quantum cells of phase space are filled up such that collisions in which the momentum is changed become rather improbable. Then the coefficient of diffusion (which is proportional to the product of mean free path and particle velocity) is large, and heat conduction can become so efficient that it short-circuits the radiative transfer.

The energy flux \vec{F}_{cd} due to heat conduction may be written as

$$\vec{F}_{cd} = -k_{cd}\nabla T \text{ [m}^2\text{s}^{-1}\text{].}$$

The sum of the conductive flux \vec{F}_{cd} and the radiative flux \vec{F}_{rad} is

$$\vec{F} = \vec{F}_{rad} + \vec{F}_{cd} = -(k_{rad} + k_{cd})\nabla T \text{ [m}^2\text{s}^{-1}\text{],}$$

On the other hand, we can just as well write the conductive coefficient k_{cd} formally in analogy to k_{rad} as

$$k_{cd} = \frac{4ac}{3} \frac{T^3}{\kappa_{cd}\rho} \text{ [cm}^2\text{g}^{-1}\text{],}$$

hence defining the **conductive opacity** κ_{cd} . Then the flux becomes

$$\vec{F} = -\frac{4ac}{3} \frac{T^3}{\rho} \left(\frac{1}{\kappa_{rad}} + \frac{1}{\kappa_{cd}} \right) \nabla T,$$

which shows that we arrive formally at the same type of equation as in the pure radiative case if we replace $1/\kappa$ by $1/\kappa_{rad} + 1/\kappa_{cd}$. Again the result is plausible, since the mechanism of transport that provides the largest flux will dominate the sum (i.e., the mechanism for which the stellar matter has the highest “transparency”).

The basic equation for energy transport which, if we define properly, holds for radiative and conductive energy transport, can be rewritten in a convenient form. Starting with

$$\frac{\partial T}{\partial r} = -\frac{3}{16\pi ac} \frac{\kappa\rho L}{r^2 T^3} \text{ [K cm}^{-1}\text{],}$$

we can transform to the independent variable mass m using the fact that $\partial r / \partial m = (4\pi r^2 \rho)^{-1}$:

$$\frac{\partial T}{\partial m} = -\frac{3}{64\pi^2 ac} \frac{\kappa L}{r^4 T^3} \text{ [K g}^{-1}\text{].}$$

Assuming hydrostatic equilibrium, we can divide this by $\partial P / \partial r = -(Gm\rho)/r^2$ and obtain

$$\frac{(\partial T / \partial m)}{(\partial P / \partial m)} = \frac{3}{16\pi ac G} \frac{\kappa L}{m T^3} \text{ [K P g}^{-2}\text{].}$$

We call the ratio of the derivatives on the left $(dT/dP)_{rad}$, and we mean by this the variation of T in the star with depth, where the depth is expressed by the pressure, which increases monotonically inwards. In this sense, in a star which is in hydrostatic equilibrium and transports the energy by radiation (and conduction), $(dT/dP)_{rad}$ is a gradient describing the temperature variation with depth. If we use the customary abbreviation

$$\nabla_{rad} \equiv \left(\frac{d \ln T}{d \ln P} \right)_{rad} \text{ [K P}^{-1}\text{],}$$

which can be written in the form

$$\nabla_{rad} = \frac{3}{16\pi ac G} \frac{\kappa L P}{m T^4} \text{ [K P}^{-1}\text{],}$$

in which conduction effects are now included. ∇_{rad} means a spatial derivative (connecting P and T in two neighbouring mass shells), while ∇_{ad} describes the thermal variation of one and the same mass element during its adiabatic compression. Only in special cases $(d \ln T / d \ln P)$ and ∇_{ad} will have the same value, and we then speak of an **adiabatic stratification**.

Transport of energy by convection: Convective transport of energy means an exchange of energy between hotter and cooler layers in a dynamically unstable region through the exchange of macroscopic mass elements (“blobs”, “bubbles”, “convective elements”), the hotter of which move upwards while the cooler ones descend. The moving mass elements will finally dissolve in their new surroundings and thereby deliver their excess (or deficiency) of heat. Owing to the high density in stellar interiors, convective transport can be very efficient. However, this energy transfer can operate only if it finds a sufficient driving mechanism in the form of the buoyancy forces.

A thorough theoretical treatment of convective motions and transport of energy is extremely difficult. It is the prototype of the many astrophysical problems in which the bottleneck preventing decisive progress

is the difficulty involved in solving the well-known hydrodynamic equations. For simplifying assumptions, solutions are available that may even give reasonable approximations for certain convective flows in the laboratory. Unfortunately, convection in stars proceeds under rather malicious conditions: turbulent motion transports enormous fluxes of energy in a very compressible gas, which is stratified in density, pressure, temperature, and gravity over many powers of ten. Nevertheless, large efforts have been made over many years to solve this notorious problem, and they have partly arrived at promising results. None of the so-called Reynolds stress models, however, have reached a stage where it could provide a procedure easy enough to be handled in everyday stellar-structure calculations, and at the same time would describe the full properties of convection accurately enough. On the other hand, full two- and three-dimensional hydrodynamical simulations have also made large progress, thanks to the impressive advances in supercomputer technology and efficient numerical algorithms. They give valuable hints to the true nature of convection and often serve as numerical experiments to test the dynamical methods. Nevertheless, these numerical simulations are still limited in their size and thus can follow convection in most cases only for a limit time and only for thin convection zones. But even if these restrictions can be foreseen to get relaxed with time, such full hydrodynamical simulations will never be used in full stellar evolution models, as they would unnecessarily follow the star's evolution on a dynamical timescale, which is so much shorter than the dominant nuclear one. Therefore, we limit ourselves exclusively to the description of the old so-called **mixing-length** theory. The reason for this is not that we believe it to be sufficient, but it does provide at least a simple method for treating convection locally, at any given point of a star. Moreover, empirical tests of the resulting stellar models show a surprisingly good agreement with observations. And, finally, even this poor approximation shows without any doubt that in the very deep interior of a star, a detailed theory is normally not necessary.

Note that in the following we are dealing only with convection in stars that are in hydrostatic equilibrium. We furthermore assume that the convection is time independent, which means that it is fully adjusted to the present state of the star. Otherwise, a convection theory for rapidly changing regions (time-dependent convection) has to be developed.

The gradient ∇_{rad} is what would be maintained in a star if the whole luminosity L had to be transported outwards by radiation only. If convection contributes to the energy transport, the actual gradient ∇ will be different (namely smaller). In the following, we will estimate ∇ in the case of convection.

The basic picture: The mixing-length theory goes back to Ludwig Prandtl, who in 1925 modelled a simple picture of convection in complete analogy to molecular heat transfer: the transporting "particles" are macroscopic mass elements ("blobs") instead of molecules; their mean free path is the **mixing length** after which the blobs dissolve in their new surroundings. Prandtl's theory was adapted for stars afterwards.

The radiation pressure gradient is given by

$$\frac{dP_{\text{rad}}}{dr} = -\frac{\kappa\rho}{c} F_{\text{rad}} \quad [\text{P m}],$$

where F_{rad} is the outward radiative flux. The radiation pressure may also be expressed as

$$\frac{dP_{\text{rad}}}{dr} = \frac{4}{3} a T^3 \frac{dT}{dr} \quad [\text{P m}].$$

Equating the two, we have

$$\frac{dT}{dr} = -\frac{3}{4ac} \frac{\kappa\rho}{T^3} F_{\text{rad}} \quad [\text{K m}^{-1}].$$

If we use the expression for the radiative flux written in terms of the local radiative luminosity of the star at radius r ,

$$F_{\text{rad}} = \frac{L_r}{4\pi r^2} \quad [\text{erg s}^{-1} \text{ cm}^{-2}],$$

the temperature gradient for radiative transport becomes

$$\frac{dT}{dr} = -\frac{3}{4ac} \frac{\kappa\rho}{T^3} \frac{L_r}{4\pi r^2} \quad [\text{K m}^{-1}].$$

As either the flux or the opacity increases, the temperature gradient must become steeper (i.e., more negative) if radiation transport is to carry all of the required luminosity outward. The same situation holds as the density increases or the temperature increases.

We now consider a situation where a hot convective bubble of gas rises and expands **adiabatically**, meaning that the bubble does not exchange heat with its surroundings. After it has travelled some distance, it finally **thermalizes**, giving up any excess heat as it loses its identity and dissolves into the surrounding gas. Differentiating the **ideal gas law**

$$P = \frac{\rho k_B T}{\mu m_H} [\text{P}]$$

yields an expression involving the bubble's **temperature gradient**

$$\frac{dP}{dr} = -\frac{P}{\mu} \frac{d\mu}{dr} + \frac{P}{\rho} \frac{d\rho}{dr} + \frac{P}{T} \frac{dT}{dr}.$$

Using the adiabatic relationship between pressure and density

$$PV^\gamma = K [\text{P m}^3]$$

and recalling that $V \equiv 1/\rho$ is the **specific volume**, we have that

$$P = K\rho^\gamma [\text{P}].$$

Differentiating and re-writing, we obtain

$$\frac{dP}{dr} = \gamma \frac{P}{\rho} \frac{d\rho}{dr} [\text{P m}^{-1}].$$

If we assume for simplicity that the mean molecular weight μ is constant, we can combine equations for dP/dr to give the **adiabatic temperature gradient**

$$\left(\frac{dT}{dr} \right)_{\text{ad}} = \left(1 - \frac{1}{\gamma} \right) \frac{T}{P} \frac{dP}{dr} [\text{K m}^{-1}].$$

Using the equation of **hydrostatic equilibrium**

$$\frac{dP}{dr} = -\frac{GM_r \rho}{r^2} = -\rho g [\text{P m}^{-1}], \quad g \equiv \frac{GM_r}{r^2} [\text{m s}^{-2}]$$

and the ideal gas law, we finally obtain

$$\left(\frac{dT}{dr} \right)_{\text{ad}} = \left(1 - \frac{1}{\gamma} \right) \frac{\mu m_H}{k} \frac{GM_r}{r^2} [\text{K m}^{-1}].$$

If the star's actual temperature gradient is *steeper* than the adiabatic temperature gradient,

$$\left| \frac{dT}{dr} \right|_{\text{act}} > \left| \frac{dT}{dr} \right|_{\text{ad}} [\text{K m}^{-1}],$$

the temperature gradient is said to be **super-adiabatic** (recall that $dT/dr < 0$). In the deep interior of a star, if the actual temperature gradient is just *slightly* larger than the adiabatic temperature gradient, this may be sufficient to carry nearly all of the luminosity by convection. Consequently, it is often the case that either radiation or convection dominates the energy transport in the deep interior of stars, while the other energy transport mechanism contributes very little to the total energy outflow. The particular mechanism in operation is determined by the temperature gradient. However, near the surface of the star, the situation is much more complicated: both radiation and convection can carry significant amounts of energy simultaneously.

This condition for convection can be used to find another useful relation. Since $dT/dr < 0$ and $1/\gamma - 1 < 0$ (recall that $\gamma > 1$),

$$\frac{T}{P} \left(\frac{dT}{dr} \right)^{-1} \frac{dP}{dr} < -\frac{1}{\gamma^{-1} - 1} [\text{dimensionless}],$$

which may be simplified to give

$$\frac{T}{P} \frac{dP}{dT} < \frac{\gamma}{\gamma - 1} [\text{dimensionless}],$$

or, for convection to occur,

$$\frac{d \ln P}{d \ln T} < \frac{\gamma}{\gamma - 1} [\text{P K}^{-1}].$$

For an ideal monatomic gas, $\gamma = 5/3$ and convection will occur in some region of a star when $d \ln P/d \ln T < 2.5$. In that case, the temperature gradient (dT/dr) is given approximately by

$$\left(\frac{dT}{dr} \right)_{\text{ad}} = -\left(1 - \frac{1}{\gamma} \right) \frac{\mu m_H}{k} \frac{GM_r}{r^2} [\text{K m}^{-1}].$$

By comparing this with the temperature gradient for radiative transport

$$\frac{dT}{dr} = -\frac{3}{4ac} \frac{\bar{\kappa}\rho}{T^3} \frac{L_r}{4\pi r^2} [\text{K m}^{-1}]$$

together with the condition for convection written in terms of the temperature gradient,

$$\left| \frac{dT}{dr} \right|_{\text{act}} > \left| \frac{dT}{dr} \right|_{\text{ad}} [\text{K m}^{-1}],$$

it is possible to develop some understanding of which conditions are likely to lead to convection over radiation. In general, convection will occur when

1. the stellar opacity is large, implying that an unachievable steep temperature gradient would be necessary for radiative transport;
2. a region exists where ionization is occurring, causing a large specific heat and a low adiabatic temperature gradient; and
3. the temperature dependence of the nuclear energy generation rate is large, causing a steep radiative flux gradient and a large temperature gradient.

In the atmosphere of many stars, the first two conditions can occur simultaneously, whereas the third condition would occur only deep in stellar interiors. In particular, the third condition can occur when the highly temperature-dependent CNO cycle or triple alpha processes are occurring.

Convective regions: Knowledge of the extension of convective regions is very important in view of their influence on the ensuing chemical evolution. A rough overview can be obtained from Fig. 22.7, where m/M and $\log M/M_\odot$ are ordinate and abscissa. For any given stellar mass M along a line parallel to the ordinate it is indicated what conditions we would encounter when drilling a radial borehole from the surface to the centre. In particular, one can see whether the corresponding mass elements are convective or radiative. Aside from the stars of smallest mass ($M < 0.25 M_\odot$), we can roughly distinguish between two types of models:

$$\begin{aligned} &\text{radiative core + convective envelope (lower MS)} \\ &\text{convective core + radiative envelope (upper MS).} \end{aligned}$$

The transition from one type to the other occurs near $M = 1 M_\odot$.

The distinction between convective and radiative regions is made here by using the **Schwarzschild criterion**, which predicts convection if the radiative gradient of temperature ∇_{rad} exceeds the adiabatic gradient ∇_{ad} . The variation of these gradients (together with that of the actual gradient ∇) throughout the star is plotted in Fig. 22.8 for $M = 1 M_\odot$ and $10 M_\odot$. For the abscissa, $\log T$ is chosen, since this conveniently stretches the scale in the complicated outer layers.

Let us start with the simpler situation concerning the convective core. When comparing Figure 20a, b, we see that the convective core in the more massive models is caused by a steep increase of ∇_{rad} towards the centre. The reason for this is that the dominating CNO cycle, with its extreme temperature sensitivity, concentrates the energy production very much towards the centre (cf. the curve l=L D 0:5 in Fig. 22.7, and Fig. 22.4e). Therefore we find in these stars very high fluxes of energy ($L = 4\pi r^2$) at small r , which produce large values of ∇_{rad} . Figure 19 shows a remarkable increase in the extent of the convective core for increasing M . The core covers as much as 65% of the stellar mass in a star of $50 M_\odot$, an increase caused by the increasing radiation pressure which depresses the value of ∇_{ad} well below its standard value of 0.4 for an ideal monatomic gas. In the centre of the $50 M_\odot$ model, roughly $1/3$ of P is radiation pressure, and $\nabla_{\text{ad}} \approx 0.27$. From Figure 20b it is clear that a depression of ∇_{ad} in the central region will shift the intersection with ∇_{rad} (i.e., the border of the convective core) outwards to smaller T . When we increase M to much larger values still, the top of the convective core will finally approach the surface such that we should obtain fully convective stars. We then approach models of the so-called **supermassive stars**. In less massive stars, the pp chain with its smaller temperature sensitivity dominates. This distributes the energy production over a much larger area, so that the flux ($L/4\pi r^2$) and ∇_{rad} are much smaller in the central region, which thus remains radiative.

Outer convective envelopes can generally be expected to occur in stars of low effective temperature. When studying the different gradients in the outer layers of cool stars (Figure 20a), one finds a variety of complicated details. The variation of ∇_{ad} clearly shows depressions in those regions where the most abundant elements, hydrogen ($T \gtrsim 10^4 \text{ K}$) and helium ($T \approx 10^5 \text{ K}$), are partially ionized. The most striking feature is that ∇_{rad} reaches enormous values (more than 10^5). This is due to the large opacity κ , which here increases by several powers of 10. Therefore the Schwarzschild criterion indicates convective instability: the models have an outer convective zone. In the largest part of it, the density is so high that convection is very effective and the actual gradient ∇ is close to ∇_{ad} . Convective transport becomes

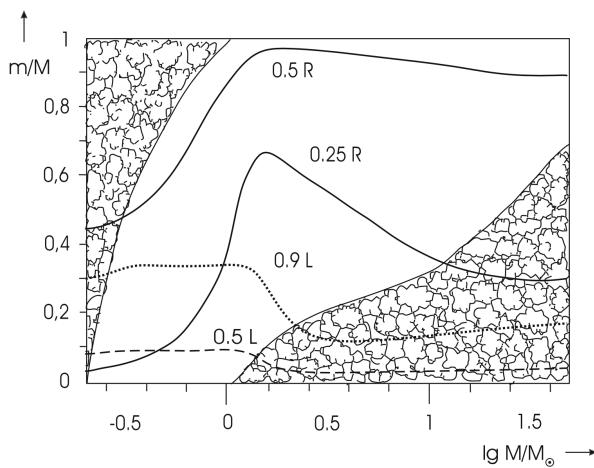


Figure 19:
The mass values m from centre to surface are plotted against the stellar mass M for the same ZAMS models as in Fig. 22.1. “Cloudy” areas indicate the extension of convective zones inside the models. Two solid lines give the m values at which r is $1/4$ and $1/2$ the total radius R : The dashed and dotted lines show the mass elements inside which 50% and 90% of the total luminosity L are produced. Figure taken from Kippenhahn, Weigert & Weiss (2012).

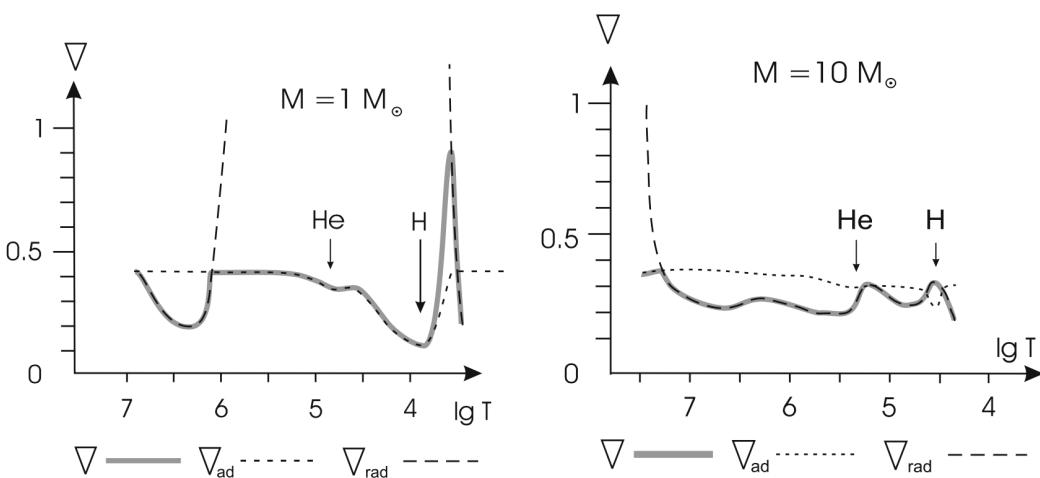


Figure 20:
The grey solid lines show the actual temperature gradient $\nabla = d \ln T / d \ln P$ over the temperature T (in K) inside two ZAMS models of $1 M_\odot$ (left panel) and $10 M_\odot$ (right panel). The corresponding adiabatic gradients ∇_{ad} (dotted lines) and radiative gradients ∇_{rad} (dashed lines) are also plotted, and the location of the ionization zones of hydrogen and helium are indicated (arrows). The chemical composition of the models is the same as for those of Figure 19. Figure taken from Kippenhahn, Weigert & Weiss (2012).

ineffective only in the outermost, super-adiabatic part, where ∇ is clearly above ∇_{ad} . Scarcely anything of all these features appears in the hot envelope of the $10 M_\odot$ star (Figure 20b). ∇_{rad} remains nearly at the same level; even the photosphere is too hot for hydrogen to be neutral, and only the small dip from the second He ionization is seen immediately below the photosphere. This causes such a shallow zone with convective instability that only for special cases, depending on the detailed chemical composition, convective motions set it.

The outer convection zone gradually penetrates deeper into the star with decreasing T_{eff} . Its lower border finally reaches the centre at $M \lesssim 0.25 M_\odot$ (left end of Figure 19), such that the main-sequence stars of even smaller masses are fully convective.

The present-day interior structure of the Sun: Consistent with the current age of the Sun, a **solar model** may be constructed for the present-day Sun. Table 3 gives the values of the central temperature, pressure, density, and composition for one such solar model, and a schematic diagram of the model is shown in Figure 21. According to the evolutionary sequence leading to this model, during its lifetime the mass fraction of hydrogen (X) in the Sun’s center has decreased from its initial value of 0.71 to 0.34, while the central mass fraction of helium (Y) has increased from 0.27 to 0.64. In addition, due to diffusive settling of elements heavier than hydrogen, the mass fraction of hydrogen near the surface has increased by approximately 0.03, while the mass fraction of helium has decreased by 0.03.

Because of the Sun’s past evolution, its composition is no longer homogeneous but instead shows the influence of ongoing nucleosynthesis, surface convection, and elemental diffusion (i.e., settling of heavier elements). The composition structure of the Sun is shown in Figure 21 for 1H , 3He , and 4He . Since the Sun’s primary energy production mechanism is the pp chain, 3He is produced and then destroyed again. At the top of the hydrogen burning region where the temperature is lower, 3He is relatively

Temperature	$1.57 \times 10^7 \text{ K}$
Pressure	$2.342 \times 10^{16} \text{ N m}^{-2}$
Density	$1.527 \times 10^5 \text{ kg m}^{-3}$
X	0.3397
Y	0.6405

Table 3: Central conditions in the Sun. (Data from Bahcall, Pinsonneault, and Basu, Ap. J., 555, 990, 2001.) Table taken from Carroll & Ostlie (2007).

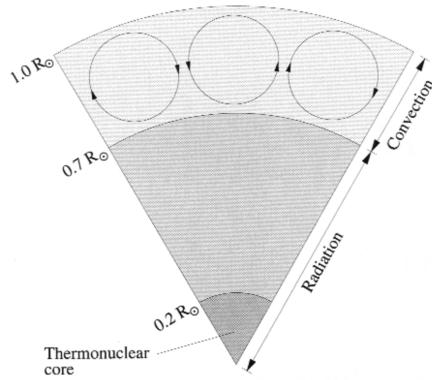


Figure 21:
Figure taken from
<https://web.njit.edu/gary/321/Lecture9.html>.

more abundant because it is produced more easily than it is destroyed. At greater depths, the higher temperatures allow the $^3\text{He}-^3\text{He}$ interactions to proceed more rapidly, and the ^3He abundance again decreases (the temperature profile of the Sun is shown in Figure 22). The slight ramp in the ^1H and ^4He curves near $0.7 R_\odot$ reflects evolutionary changes in the position of the base of the surface convection zone, combined with the effects of elemental diffusion. Within the convection zone, turbulence results in essentially complete mixing and a homogenous composition. The base of the present-day convection zone is at $0.714 R_\odot$.

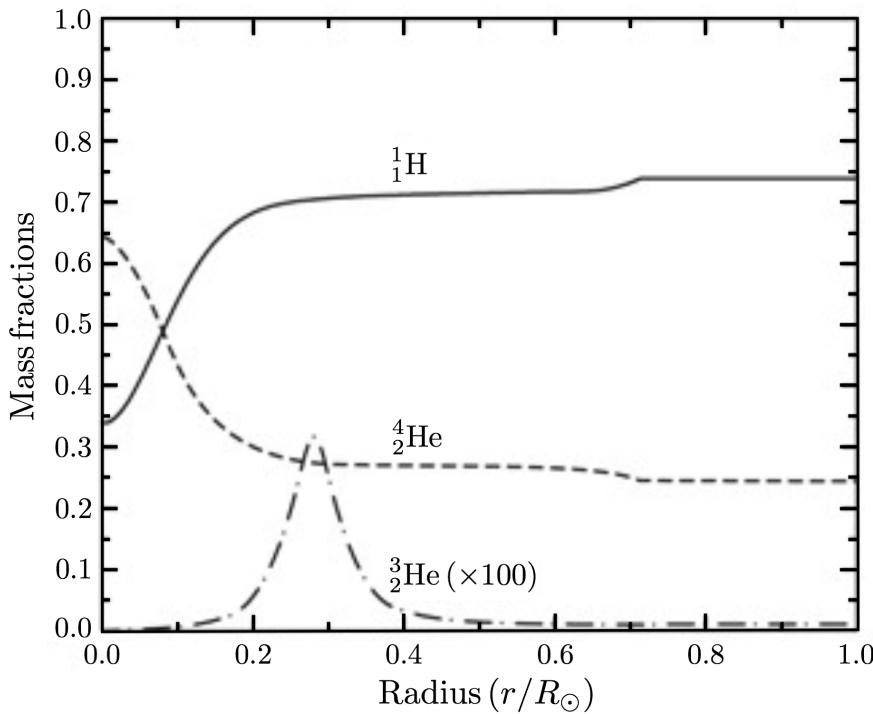


Figure 22:
The abundance of ^1H , ^3He , and ^4He as a function of radius for the Sun. Note that the abundance of ^3He is multiplied by a factor of 100. (Data from Bahcall, Pinsonneault, & Basu, Ap. J., 555, 990, 2001.) Figure adapted from Carroll & Ostlie (2007).

Figure 23 shows $d \ln P / d \ln T$ versus r/R_\odot . As can be seen, the Sun is purely radiative below $r/D_\odot = 0.714$ and becomes convective above that point. Physically this occurs because the opacity in the outer portion of the Sun becomes large enough to inhibit the transport of energy by radiation; recall that the radiative temperature gradient is proportional to the opacity:

$$\frac{dT}{dr} = -\frac{3}{4ac} \frac{\bar{\kappa}\rho}{T^3} \frac{L_r}{4\pi r^2} [\text{K m}^{-1}].$$

When the temperature gradient becomes too large, convection becomes the more efficient means of energy transport. Throughout most of the region of convective energy transport, $d \ln P / d \ln T \approx 2.5$, which is

characteristic of the nearly adiabatic temperature gradient of most convection zones. The rapid rise in $d \ln P / d \ln T$ above $0.95 R_\odot$ is due to the significant departure of the actual temperature gradient from the adiabatic one. In this case convection must be described by a more detailed treatment, such as **mixing length theory**.

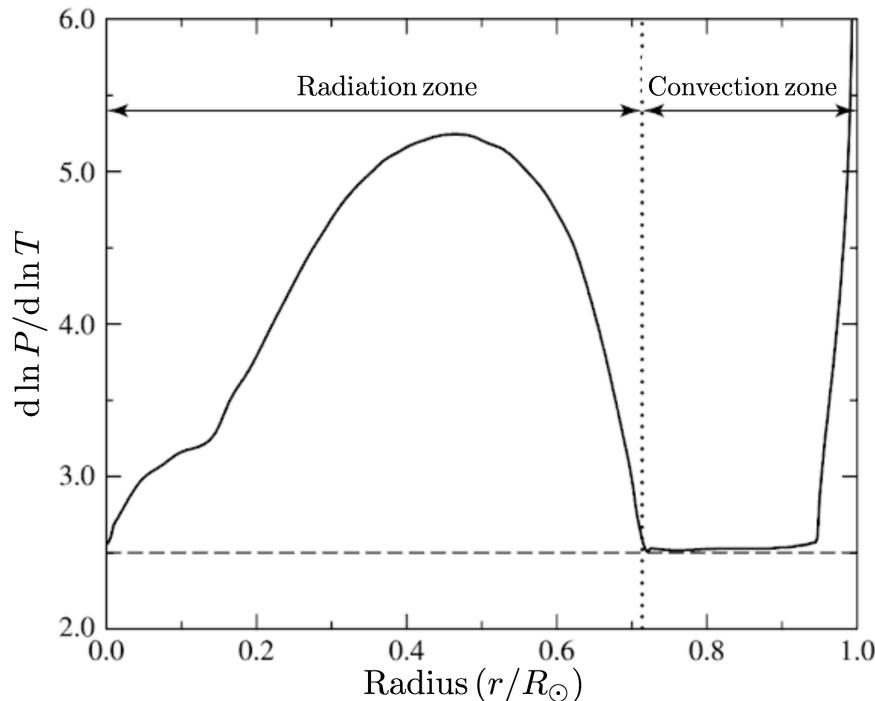


Figure 23:
The convection condition $d \ln P / d \ln T$ plotted versus r / R_\odot . The dashed horizontal line represents the boundary between adiabatic convection and radiation for an ideal monatomic gas. The onset of convection does not exactly agree with the ideal adiabatic case because of the incorporation of a sophisticated equation of state and a more detailed treatment of convection physics. The rapid rise in $d \ln P / d \ln T$ near the surface is associated with a highly super-adiabatic nature of convection in that region (i.e., the adiabatic approximation that convection occurs when $d \ln P / d \ln T < 2.5$ is invalid near the surface of the Sun.) [$d \ln P / d \ln T$ was computed using data from Bahcall, Pinsonneault, and Basu, Ap. J., 555, 990, 2001. The data for the zones above $0.95 R_\odot$ are from Cox, Arthur N. (editor), Allen's Astrophysical Quantities, Fourth Edition, AIP Press, New York, 2000.] Figure adapted from Carroll & Ostlie (2007).

Notice that $d \ln P / d \ln T$ also decreases to almost 2.5 at the center of the Sun. Although the Sun remains purely radiative at the center, the large amounts of energy that must be transported outward pushes the temperature gradient in the direction of becoming super-adiabatic. Stars only slightly more massive than the Sun are convective in their centers because of the stronger temperature dependence of the CNO cycle as compared to the pp chain.

1.8.3 Follow-up Questions

- How do we know that the Sun's outer envelope is convective?
- How far into the surface of the Sun does the convective zone permeate? How can we measure this?

1.9 Question 9

What is Eddington's luminosity limit? Explain why this limit is important for the properties and lifetimes of massive stars.

1.9.1 Short answer

The Eddington luminosity limit is the luminosity limit where the radiative force is balanced by the gravitational force. This limit is important for the properties and lifetimes of massive stars because stars massive enough to reach such luminosities can have their properties and lifetimes affected; at such high luminosities, stellar material can become unbound and ejected from the stellar surface via stellar winds.

1.9.2 Additional context

Sir Arthur Stanley Eddington observed that radiation and gravitation both obey inverse-square laws and so there would be instances when the two forces could be in balance irrespective of distance. Thus there should exist a maximum luminosity for a star of a given mass, where the force of radiation on the surface material would exactly balance the force of gravity.

There exists a gradient of the radiation pressure

$$\frac{dP_{\text{rad}}}{dr} = \frac{a}{3} T^3 \frac{dT}{dr} [\text{Pa m}^{-1}],$$

which exerts, just like the gas pressure gradient, an outward acceleration ($dP_{\text{rad}}/dr < 0$)

$$g_{\text{rad}} = -\frac{1}{\rho} \frac{dP_{\text{rad}}}{dr} [\text{m s}^{-2}]$$

Using

$$F = -\frac{4ac}{3} \frac{T^3}{\kappa\rho} \frac{\partial T}{\partial r} [\text{erg s}^{-1} \text{cm}^{-2}]$$

we see that we can rewrite this as

$$g_{\text{rad}} = \frac{\kappa F_{\text{rad}}}{c} = \frac{\kappa L_r}{4\pi r^2 c} [\text{m s}^{-2}].$$

In case that radiation pressure completely dominates over gas pressure, a star can no longer be in hydrostatic equilibrium if $g_{\text{rad}} > g$. The sum of both accelerations can be written as

$$g + g_{\text{rad}} = -\frac{Gm}{r^2} \left[1 - \frac{\kappa L_r}{4\pi c G m} \right] = -\frac{Gm}{r^2} [1 - \Gamma_r] [\text{m s}^{-2}],$$

where Γ_r can be understood as the ratio of the luminosity relative to the critical luminosity at which the bracket changes sign, and thus the star becomes unbound. For $m = M$ this critical luminosity is called the **Eddington luminosity** and is

$$L_E = \frac{4\pi c GM}{\kappa} [\text{J s}^{-1}].$$

Expressed in Solar units it is

$$\frac{L_E}{L_\odot} = 1.3 \times 10^4 \frac{1}{\kappa} \frac{M}{M_\odot}$$

and grows linearly with stellar mass. Since $L \propto M^3$, stars obviously reach a limit, where radiation pressure is able to drive a strong stellar wind, and which depends on the opacity.

For hot, massive stars electron scattering is the dominating opacity source, which can be approximated by

$$\begin{aligned} \kappa_\nu &= \frac{8\pi}{3} \frac{e_e^2}{\mu_e m_u} \\ &= 0.20(1 + X) [\text{cm}^2 \text{g}^{-1}], \end{aligned}$$

and is $\kappa_{\text{sc}} = 0.20(1 + X)$. For a mass fraction of hydrogen of 0.70 this simplifies to

$$\frac{L_E}{L_\odot} = 3.824 \times 10^4 \frac{M}{M_\odot}.$$

For $M \approx 200 M_{\odot}$ the luminosity of massive main-sequence stars reach the Eddington limit and disperse. This is a rough estimate for an upper limit. In reality the instability of this mechanism occurs at lower mass. However, the Eddington limit can become quite important in other situations.

Any object that has a luminosity greater than L_E will be forced into instability by its own radiation pressure. This effectively provides a mass-luminosity relationship for super-massive stars since these radiation-dominated configurations will radiate near their limit. It is worth noting that **magnetars**, the source for **soft gamma repeaters**, are class of gamma-ray bursts that *exceed the Eddington luminosity by far*, but are characterized by a comparably soft gamma spectrum. In a magnetar, the decaying magnetic field is the source of free energy (rather than rotation, as in pulsars).

Limiting masses for super-massive stars: Any star with a mass less than about half a million solar masses can come to equilibrium burning hydrogen via the CNO cycle, albeit with a short lifetime. More massive stars are destined to continue to contract. Of course, more massive stars will produce nuclear energy at an ever-increasing rate as their central temperatures rise. However, the rate of energy production cannot increase without bound. This is suggested by the declining exponents of the temperature dependence shown in Table 4. The nuclear reactions that involve β decay set a limit on how fast the CNO cycle can run, and β decay is independent of temperature. So there is a maximum rate at which energy can be produced by the CNO cycle operating in these stars.

Proton-Proton			CNO Cycle			Triple- α Process		
T_6	ϵ_0 (cgs)	v	ϵ_0 (cgs)	v	T_8	ϵ_0 (cgs)	v	
10	7×10^{-2}	4.60	3×10^{-4}	22.9	0.8	2×10^{-12}	49	
20	1	3.54	4.5×10^2	18	1.0	4×10^{-8}	41	
40	9	2.72	3×10^7	14.1	2.0	15	19	
80	43	2.08	2×10^{11}	11.1	3.0	6×10^3	12	
100	—	—	2×10^{12}	10.2	4.0	10^5	7.9	

Table 4: Table taken from Collins (2003).

If the nuclear energy produced is sufficient to bring the total energy above the binding energy curve, the star will explode. However, should the energy not be produced at a rate sufficient to catch the binding energy that is rising due to the relativistic collapse, the star will continue an unrestrained collapse to the Schwarzschild radius and become a black hole. Which scenario is played out will depend on the star's mass. For these stars, the temperature gradient will be above the adiabatic gradient, so convection will exist. However, the only energy transportable by convection is the kinetic energy of the gas, which is an insignificant fraction of the internal energy. Therefore, unlike normal main sequence stars, although it is present, convection will be a very inefficient vehicle for the transport of energy. This is why the star remains with a structure of a polytrope of index $n = 3$ in the presence of convection. The pressure support that determines the density distribution comes entirely from radiation and is not governed by the mode of energy transport.

The star will radiate at the Eddington luminosity, and that will set the time scale for collapse. The total energy of these stars is small compared to the gravitational energy. So most of the energy derived from gravitational contraction must go into supporting the star, and very little is available to supply the Eddington luminosity. This can be seen from the relativistic Virial theorem which indicates that any change in the gravitational energy is taken up by the kinetic energy. Relativistic particles (in this case, photons) are much more difficult to bind by gravitation than ordinary matter; thus little of the gravitational energy resulting from collapse will be available to let the star shine. The collapse will proceed very quickly on a time scale that is much nearer to the dynamical time scale than the Kelvin-Helmholtz time scale. The onset of nuclear reactions will slow the collapse, but will not stop it for the massive stars.

A dynamical analysis shows that stars more massive than about $7.5 \times 10^5 M_{\odot}$ will undergo collapse to a black hole. Here the collapse proceeds so quickly and the gravity is so powerful that the nuclear reactions, being limited by β decay at the resulting high temperatures, do not have the time to produce sufficient energy to arrest the collapse. For less massive stars, this is not the case. Stars in the narrow range of $5 \times 10^5 M \leq M \leq 7.5 \times 10^5 M_{\odot}$ will undergo explosive nuclear energy generation resulting in the probable destruction of the star.

Nothing has been said about the role of chemical composition in the evolution of these stars. Clearly, if there is no carbon present, the CNO cycle is not available for the stabilization of the star. Model

calculations show that the triple-alpha process cannot stop the collapse. For stars with low metal abundance, only the pp-cycle is available as an energy source. This has the effect of lowering the value of the maximum stable mass. Surprisingly, there is no range at which an explosion occurs. If the star cannot stabilize before reaching R_m , it will continue in a state of unrestrained gravitational collapse to a black hole. Thus, it seems unlikely that stars more massive than about a half million solar masses could exist. In addition, it seems unlikely that black holes exist with masses greater than a few solar masses and less than half a million solar masses. If they do, they must form by accretion and not as a single entity.

1.9.3 Follow-up Questions

- Draw a force diagram of what's happening.
- What particles experience gravity the most?
- What particles experience photon pressure the most?

1.10 Question 10

Explain why we know what the Sun's central temperature ought to be, and how we know what it actually is.

1.10.1 Short answer

The hydrostatic condition

$$\frac{\partial P}{\partial r} = -\frac{Gm\rho}{r^2} \text{ [P m}^{-1}\text{]}$$

together with an equation of state for a perfect gas

$$P = \frac{\rho k_B T}{\mu m_H} \text{ [P]}$$

enables us to estimate the pressure and the temperature in the interior of a star of given mass and radius. Let us replace the left-hand side of the equation for hydrostatic equilibrium by an average pressure gradient $(P_0 - P_c)/r$ where P_0 and P_c are the pressures at the surface and at the centre, respectively. On the right-hand side, we can replace m and r by rough mean values $M/2$ and $R/2$, and obtain

$$P_c \approx \frac{2GM^2}{\pi R^4} \text{ [P].}$$

From the equation of state for a perfect gas (our assumption of a perfect gas turns out to be fully justified for these values of P and T), and with the mean density

$$\bar{\rho} = \frac{3M}{4\pi R^3} \text{ [g cm}^{-3}\text{]},$$

we find with P_c that

$$\begin{aligned} T_c &= \frac{P_c \mu m_H}{\bar{\rho} k_B} \text{ [K]} \\ &= \frac{\left(\frac{2GM^2}{\pi R^4}\right) \mu m_H}{\left(\frac{3M}{4\pi R^3}\right) k_B} \text{ [K]} \\ &= \left(\frac{2GM^2}{\pi R^4} \mu m_H\right) \cdot \left(\frac{3M}{4\pi R^3} k_B\right)^{-1} \text{ [K]} \\ &= \left(\frac{2GM^2 \mu m_H}{\pi R^4}\right) \cdot \left(\frac{4\pi R^3}{3M k_B}\right) \text{ [K]} \\ &= \frac{8G\mu m_H}{3k_B} \left(\frac{M}{R}\right) \text{ [K].} \end{aligned}$$

Therefore,

$$T_c \propto \frac{M}{R} \text{ [K].}$$

With the mass and the radius of the Sun ($M_\odot = 1.989 \times 10^{33}$ g, $R_\odot = 6.96 \times 10^{10}$ cm) and with $\mu = 0.5$, we find that

$$\begin{aligned} T_c &= \frac{8G\mu m_H}{3k_B} \left(\frac{1.989 \times 10^{33} \text{ g}}{6.96 \times 10^{10} \text{ cm}}\right) \text{ [K]} \\ &= 3.1 \times 10^7 \text{ [K].} \end{aligned}$$

Modern numerical solutions give $T_c = 1.6 \times 10^7$ K.

1.10.2 Additional context

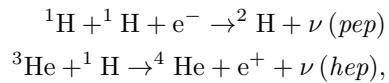
Although the Sun is a very ordinary star of average mass and in a quiet state of MS hydrogen burning, it is a unique object for stellar evolution theorists. For no other star so many quantities are known with comparable accuracy obtained by so many different and independent methods. From Kepler's laws and known distances within the solar system we can derive its mass and radius as well as the total luminosity. This yields the effective temperature by application of the Stefan-Boltzmann law. Neutrino experiments on Earth allow the determination of conditions in the innermost energy producing core. And the art

Quantity	Value	Method
Mass	$(1.9891 \pm 0.0004) \times 10^{33} \text{ g}$	Kepler's third law
Radius	$695,508 \pm 26 \text{ km}$	Angular diameter plus distance
Luminosity	$(3.846 \pm 0.01) \times 10^{33} \text{ erg s}^{-1}$	Solar constant
Effective temp.	$5,779 \pm 2 \text{ K}$	Stefan-Boltzmann law
Z/X	0.0245 ± 0.001 0.0165	Meteorites and solar spectrum (new determination)
Age	$4.57 \pm 0.02 \text{ Gyr}$	Radioactive decay in meteorites
Depth of conv. env.	$0.713 \pm 0.001 R_\odot$	Helioseismology
Env. helium content	0.246 ± 0.002	Helioseismology

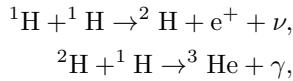
Table 5: Solar quantities and how they are derived. Table taken from Kippenhahn, Weigert & Weiss (2012).

of (helio-)seismology has returned with high accuracy the run of the sound speed throughout most of the solar interior, the helium content of the outer convective envelope, and its depth. These quantities restrict the modelling of the present Sun and allow a comparison with stellar evolution theory at a degree of precision which is almost unique in astrophysics. Table 5 summarizes the fundamental solar parameters and the method to derive them. Note that the rather large uncertainty in the solar mass is the result of the uncertainty in Newton's constant of gravity G. Kepler's third law returns their combination, GM_\odot , with a precision of 10^{-7} !

Solar neutrinos: Some of the nuclear reactions of the pp chain, as well as of the CNO cycle, produce neutrinos. In addition, there are also neutrinos due to the very rare *pep* and *hep* reactions



the latter one being the trivial way to produce ${}^4\text{He}$ after the **pp-chain**,



but it is occurring in only 10^{-8} of all cases. However, the energy of the emitted neutrino is close to 10 MeV, and it is therefore necessary to consider this reaction. The neutrinos leave the star practically without interacting with the stellar matter. The energy spectrum of neutrinos from β decay is continuous, since the electrons can take part of the energy away, while neutrinos released after an inverse β decay are essentially monochromatic. Therefore most reactions of the pp chain have a continuous spectrum, while the pep-reaction and the electron capture on ${}^7\text{Be}$ have a line spectrum. Since ${}^7\text{Be}$ can decay into ${}^7\text{Li}$ either in the ground state or in an excited state, this reaction gives two spectral lines. The neutrino spectrum of the Sun as predicted from the reactions of the pp chain, computed from our standard solar model, is given in Figure 24.

Since the solar neutrinos can leave the Sun almost unimpeded they can in principle be measured in terrestrial laboratories and thus be used to learn directly about conditions in the innermost solar core. This difficult task indeed has been undertaken since 1964, when John Bahcall and Raymond Davies began to plan for an underground neutrino detector in a mine in Homestead, North Dakota. Forty years later the experiments finally have confirmed the standard solar model, and R. Davies received the Nobel Prize for his work. The time in between, however, was characterized by the **solar neutrino problem**.

The solar neutrino problem consisted in the fact that since the first results from the so-called chlorine experiment by Davies there was a lack of neutrinos compared to solar model predictions. The chlorine experiment is sensitive to neutrinos with energies above 0.814 MeV and therefore, as can be seen in Figure 24 mainly to the ${}^8\text{B}$ neutrinos, with some contribution from pep, hep, and ${}^7\text{Be}$ neutrinos. The experiment is based on the reaction ${}^{37}\text{Cl} + \nu \rightarrow {}^{37}\text{Ar}$, where the decays of radioactive argon nuclei are counted. The rate of neutrino captures is commonly measured in solar neutrino units (SNU). One SNU corresponds to 10^{-36} captures per second and per target nucleus. The predicted counts amount to 7.5 SNU for the chlorine experiment, the measurements averaged over several decades to only 2.5 ± 0.2 SNU. The deficit could indicate that the solar centre is cooler than in the models.

To improve the experimental evidence, additional experiments were started. First, another kind of radiochemical detector using gallium in the detector fluid measured, due to a much lower energy threshold, the majority of neutrinos, including those from the pp-reaction. Later, electron-scattering detectors were developed, which are sensitive to the highest energies only, but which provide directional information about the neutrino source (for these detectors the *hep*-neutrinos have to be taken into account.). All

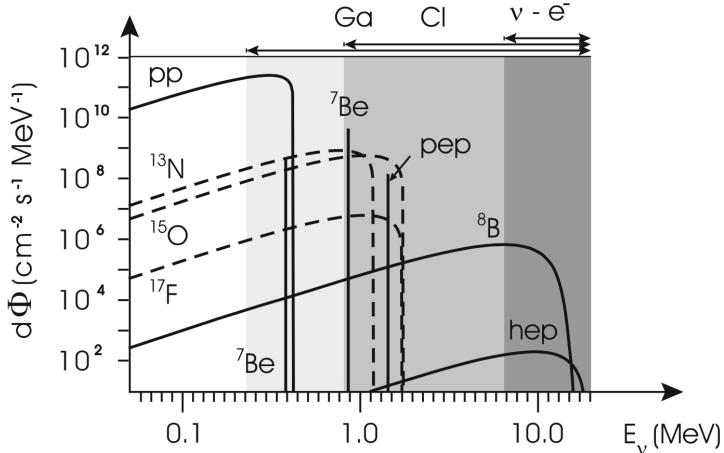


Figure 24: The neutrino spectrum of the Sun as predicted from the theoretical standard solar model. The solid lines belong to reactions of the pp chain while the broken lines are due to reactions of the CNO cycle. The neutrinos from most of the reactions have continuous spectra, while mono-energetic neutrinos come from ${}^7\text{Be}$ and from the pep-reaction. The flux for the continuum sources is given in $\text{cm}^2 \text{s}^{-1} \text{MeV}^{-1}$ and for the line sources in $\text{cm}^2 \text{s}^{-1}$. The sensitivity of the three types of neutrino experiments is indicated above the figure and by the shaded regions. Figure taken from Kippenhahn, Weigert & Weiss (2012).

experiments confirmed that the solar neutrino flux was of the right order of magnitude, and therefore that indeed the Sun shines by the nuclear fusion of hydrogen, but they also consistently measured a deficit of neutrinos. This deficit, however, varied between different kinds of detectors.

With more and more experimental data it became evident that even hypothetical changes to the solar centre cannot solve the problem and that the solution is most likely to be found in the properties of neutrinos. All nuclear reactions emit electron neutrinos, and these are the only ones that were measured in terrestrial experiment, with the exception of the electron-scattering Sudbury Neutrino Observatory (SNO) experiment in Canada, where heavy water (with a high percentage of deuterium isotopes) was used as the detector. Here also reactions with the two other types (flavours) of neutrinos, muon and tau neutrinos can be detected. Summing these and the electron neutrinos up, the total number of detections is completely consistent with the solar model prediction, within a few percent. What created the apparent solar neutrino deficit is the fact that neutrinos can change their flavour, both while travelling through vacuum and more efficiently in the presence of electrons in the solar interior. A similar effect was also confirmed for muon neutrinos arising in the Earth's upper atmosphere from high-energy cosmic radiation, when measured before or after they have travelled through the Earth's interior. The modelling of the solar interior, together with sophisticated experiments, has therefore resulted in new knowledge about fundamental properties of neutrinos. In particular, these so-called **neutrino oscillations** are possible only if neutrinos have mass.

1.11 Question 11

Which have higher central pressure, high-mass or low-mass main-sequence stars? Roughly, what is their mass-radius relation? Derive this.

1.11.1 Short answer

Using hydrostatic equilibrium, we have that:

$$\begin{aligned}\frac{dP}{dr} &= -\frac{GM}{r^2}\rho \\ P &\sim \frac{M}{R}\rho \\ P &\sim \frac{M^2}{R^4}.\end{aligned}$$

With the ideal gas law, we get:

$$\begin{aligned}P &= \frac{\rho k_B T}{\mu m_H} \\ P &\sim \rho \\ P &\sim \frac{M}{R^3}.\end{aligned}$$

Putting these two together, we get the mass-radius relation:

$$\begin{aligned}\frac{M^2}{R^4} &= \frac{M}{R^3} \\ M &\sim R.\end{aligned}$$

We can use this to determine the central pressure relation:

$$\begin{aligned}P &\sim \frac{M^2}{R^4} \\ P &\sim \frac{M^2}{M^4} \\ P &\sim \frac{1}{M^2}.\end{aligned}$$

Therefore, lower-mass main sequence stars have higher central pressures.

1.11.2 Follow-up Questions

- How would we actually know the central pressure?
- What properties can we measure to test models of stellar structure?

1.12 Question 12

Sketch the SED of an O, A, G, M, and T star. Give defining spectral characteristics, such as the Balmer lines and Balmer jump and Calcium doublets, and describe physically.

1.12.1 Short answer

Figure 25 shows the spectral sequence from the hottest **early-type stars**, the O-type stars to spectral type G2. Figure 26 shows the spectral sequence from the coolest **late-type stars**, the G2-type stars to spectral type M4.5. Figure 27 shows the spectrum of a typical mid-T dwarf from the red to the mid-infrared. In the red, the SED is shaped by the very extensive wings of the KI resonance line; in the far red and infrared, the most prominent features are due to CH₄ and H₂O. Indeed, T dwarfs are distinguished from the L dwarfs by the presence of CH₄ absorption in their near-infrared and infrared spectra. All T dwarfs are brown dwarfs.

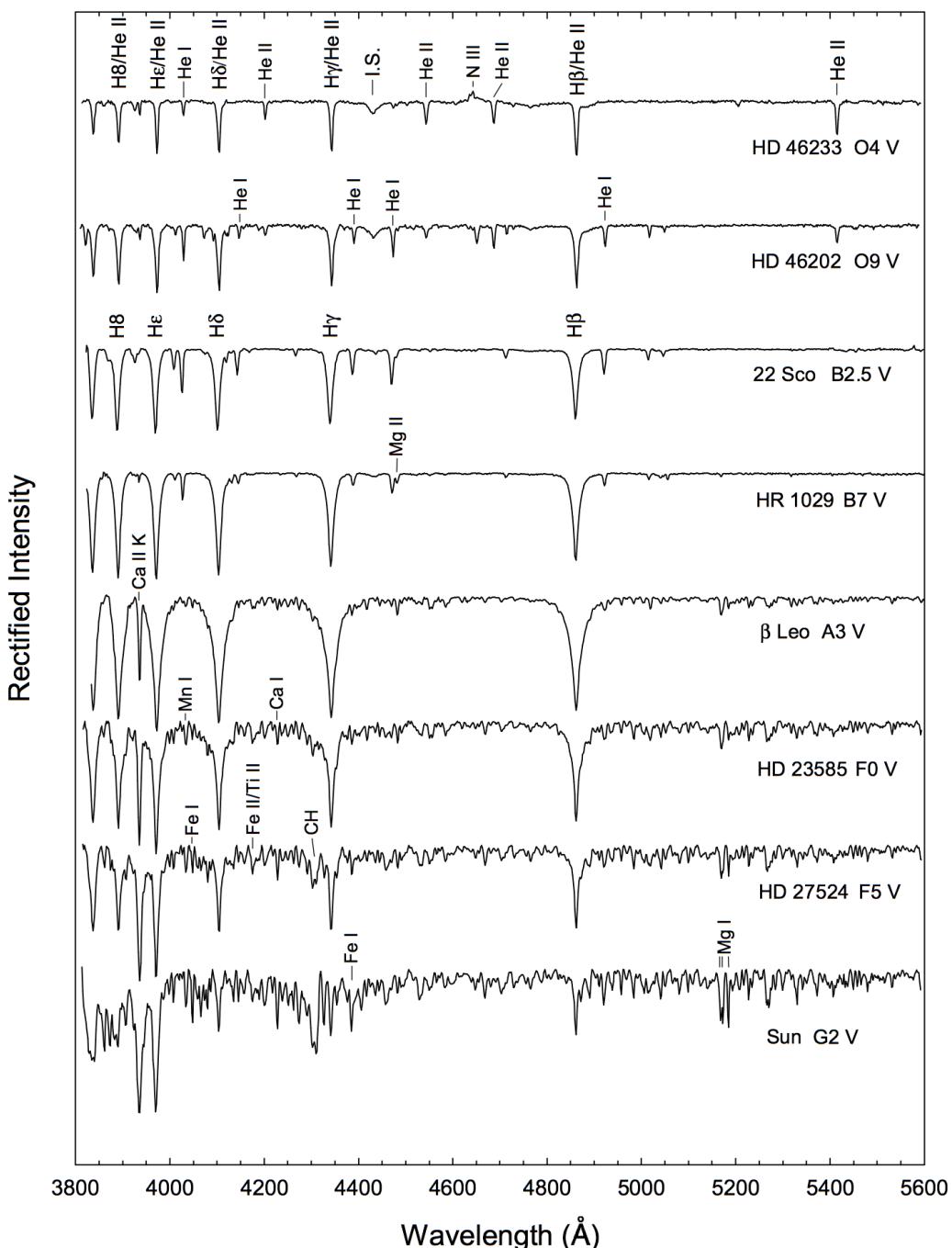


Figure 25: The Main Sequence from O4 to G2, the “early-type” stars. Figure taken from Gray (2009).

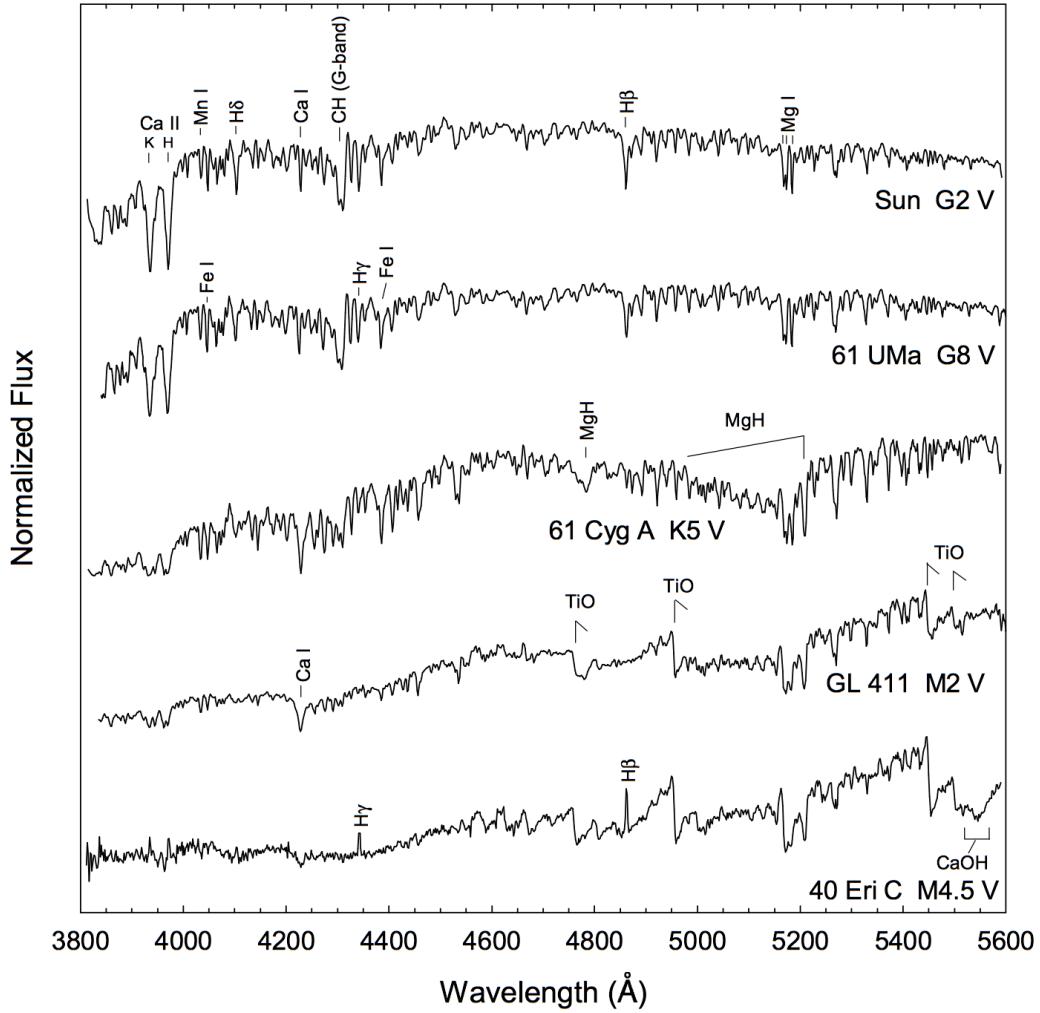


Figure 26: The Main Sequence from G2 to M4.5, the “late-type” stars. Figure taken from Gray (2009).

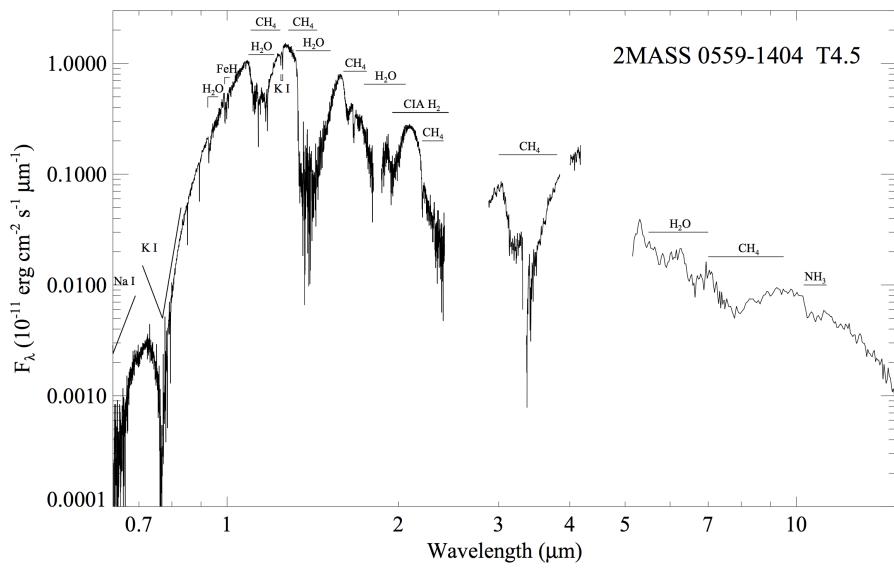


Figure 27: The spectrum of a mid-T dwarf spanning the wavelengths $0.63 - 15 \mu\text{m}$. Prominent atomic and molecular features are labeled. Reproduced from Stellar Spectral Classification, courtesy Adam Burgasser. Figure taken from Gray (2009).

1.12.2 Additional context

The spectra of early-type stars are dominated by the Balmer lines of hydrogen. Note how they increase in strength and reach a maximum at A2, and then fade thereafter. The O- and B-type stars are characterized by lines of helium; the A-, F-, and G-type stars are characterized by lines of metals. The O-type stars constitute the hottest normal stars, and are characterized by moderately weak hydrogen lines, lines of neutral helium (HeI) and by lines of singly ionized helium (HeII). The spectral type can be judged easily by the ratio of the strengths of lines of HeI to HeII; HeI tends to increase in strength with decreasing temperature while HeII decreases in strength. The ratio HeI 4471Å to HeII 4542Å shows this trend clearly. As we move toward later (cooler) types, the helium lines continue to fade until they essentially disappear in spectra of this resolution (1.8Å) at a spectral type of about A0.

The definition of the break between the O-type stars and the B-type stars is the absence of lines of ionized helium (HeII) in the spectra of B-type stars. The lines of HeI pass through a maximum at approximately B2, and then decrease in strength towards later (cooler) types. Were it not for the presence of helium peculiarities in both the early and late B-type stars, the behavior of HeI would be sufficient to accurately estimate the spectral type in the B-type stars. However, because of these peculiarities, the spectral type in the early B-type stars is, instead, estimated on the basis of ratios of lines of silicon ions. For instance, between B0 and B1 (an interval which can be divided into B0, B0.2, B0.5, B0.7 and B1 classes), the SiIV 4089Å/SiIII 4552Å ratio may be used. Between B1 and B3 the SiIII 4552Å/SiII 4128.32Å ratio may be used.

The **Be stars** are B-type stars that are characterized (or have been characterized in the past) by emission in one or more of the Balmer lines of hydrogen, sometimes accompanied by emission in lines of singly ionized metals, most commonly FeII. The Be class excludes B-type supergiants, as well as pre main-sequence stars such as the Herbig Ae/Be stars. Many Be stars may be classified without difficulty on the MK system, although the more extreme present significant difficulties. An extension to the MK system was introduced for the Be stars comprised of an **e index** ($e_1 \rightarrow e_4$, in direction of increasing emission-line strength). Some typical Be spectra are illustrated in Figure 28.

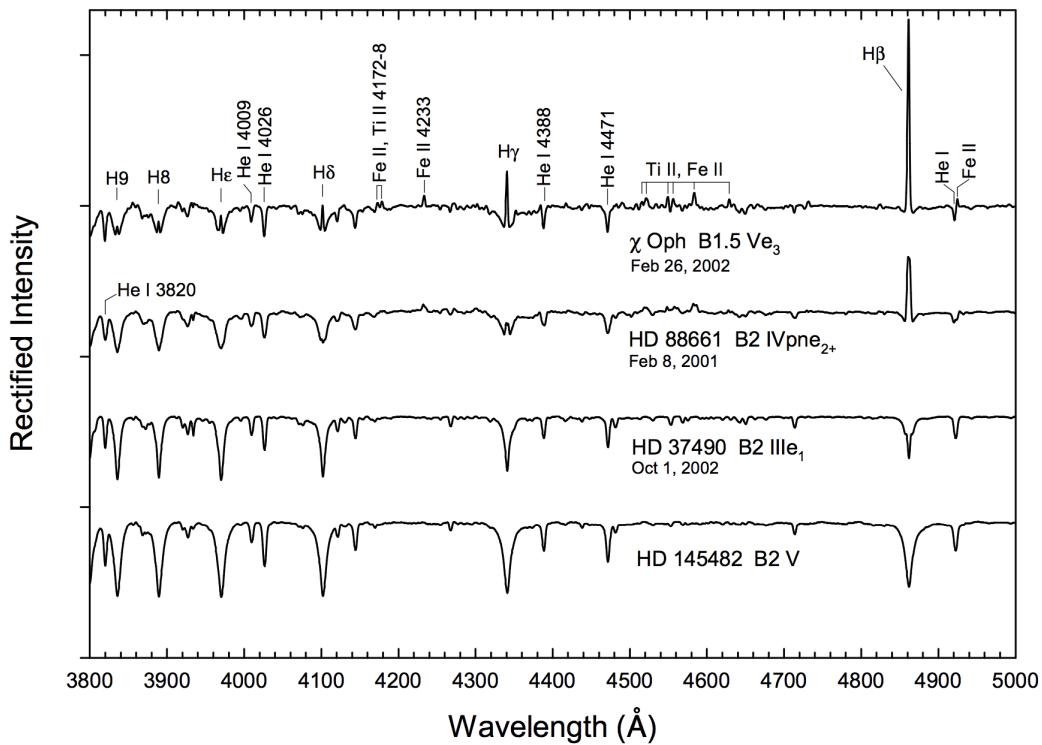


Figure 28: A selection of Be stars compared with a B2V secondary standard, HD 145482. The spectral types are on the Lesh (1968) system. Spectra from the Paranal Observatory spectral library. Figure taken from Gray (2009).

Near a spectral type of A0, the primary luminosity criterion is the progressive widening and strengthening of the hydrogen lines with decreasing luminosity. The hydrogen lines reach a maximum in the A-type stars. The CaII K-line increases significantly in strength through the A-type stars, and its absolute strength, or, more usefully, its ratio with He ϵ or H δ is a sensitive indicator of the temperature type, although one that is dependent on the metallicity. The general metallic-line spectrum also increases in strength through the A-type stars. These three criteria give a consistent temperature type in the normal

A-type stars. Any disagreement between these three criteria indicates the star is peculiar in some way. **Ap stars** are A-type (actually more commonly late B-type) stars that show significant enhancements of certain elements. **Herbig Ae stars** generally show emission in the H α line (outside of the spectral range of most of the spectra used in this atlas), and quite often emission in H β and even H γ . Many Herbig Ae stars are still contracting to the main sequence, and are thus either still surrounded by remnants of their stellar cocoons, or have developed massive stellar winds.

Metallic-line A-type stars, or, for short, **Am stars**, are defined as stars that show a difference between the K-line type and the metallic-line type. Many Am stars show the **anomalous luminosity effect** in which the luminosity criteria in certain spectral regions (in particular, the region from 4077Å–4200Å) may indicate a giant or even a supergiant luminosity, whereas other regions indicate a dwarf luminosity or even lower.

The hydrogen lines continue to weaken through the F-type stars, and the CaII K-line strengthens, although it becomes essentially saturated by the late F-type stars. The general strength of the metallic-line spectrum grows dramatically. Around F2, depending upon the resolution of the spectrum, the G-band makes its first appearance. The G-band is a molecular band, composed of thousands of closely spaced lines due to the diatomic molecule CH. By F0, the hydrogen lines have lost most of their sensitivity to luminosity. Note, however, that they can still be used to distinguish the supergiant classes from lower luminosities. Near F0, the luminosity class is estimated from the strength of lines due to ionized iron and titanium. By F8, the hydrogen lines have lost all sensitivity to luminosity, and it is now necessary to rely solely on lines and blends of ionized species.

Later than G0 along the main sequence, the hydrogen lines continue to fade, while the strength of the general metallic-line spectrum continues to increase. The G-band also increases in strength until the early K-type stars (about K3), and then begins to fade.

Molecular features first appear in the spectra of F-type main sequence stars and grow to dominate the spectra of the late-type stars, especially those of types K and M. The main sequence does not end at M4.5, but continues through the late M-, the L- and the T-type dwarfs, a sequence that spans the transition from hydrogen-burning dwarf stars to brown dwarfs. Visible in a number of stars is a broad, shallow depression, centered near 4430Å. This is an example of a diffuse interstellar band, probably due to molecules in the ISM between us and the star. A number of G and K-type giants show either strong or weak molecular bands involving carbon.

Almost all barium-rich stars are members of binary systems in which the companion is a white dwarf. It is believed that the enhanced abundances of s-process elements, such as barium, were gained from the companion star when it was on the AGB. The mass transfer came about either through **Roche-lobe overflow** or wind accretion. The **s-process elements** are the result of slow neutron capture, which occurs only at advanced stages of nuclear burning.

Notice that H β is in emission in the M4.5 star; many M dwarfs have active chromospheres and exhibit strong flares many times more energetic than solar flares. One of the manifestations of this is emission in the hydrogen lines.

1.12.3 Follow-up Questions

- Are there emission lines?
- What molecular lines are in the Sun?
- What important lines are there are much longer wavelengths than those in the optical?
- What if the A star had a protoplanetary disk?
- What is the significance of a λF_λ (or νF_ν) spectrum?
- How can the relative height of the stellar vs disk bumps change? (i.e., total energy of the system cannot change; dust bump can't be higher than star bump without extinction.)

1.13 Question 13

What can be learned about young stars (T Tauri and pre-main-sequence stars) from an analysis of their spectral features?

1.13.1 Short answer

A range of information can be obtained about T Tauri stars and pre-main-sequence stars using spectral analysis:

- chemical composition
- velocity structure
- information on the existence of inflows and/or outflows
- rate of contraction (i.e., SFR)
- disk properties
- accretion rates
- age determination (via Lithium abundance)
- evolutionary stage (via spectral index)

1.13.2 Additional context

Since we think star formation begins with a core that is purely gas, we expect to begin with a cloud that is cold and lacks a central point source. Once a protostar forms, it will begin gradually heating up the cloud, while the gas in the cloud collapses onto the protostar, reducing the opacity. Eventually enough material accretes from the envelope to render it transparent in the near infrared and eventually the optical, and we begin to be able to see the star directly for the first time. The star is left with an accretion disk, which gradually accretes and is then dispersed. Eventually the star contracts onto the main sequence.

This theoretical cartoon has been formalized into a system of classification of young stars based on observational diagnostics. At one end of this sequence lies purely gaseous sources where there is no evidence at all for the presence of a star, and at the other end lies ordinary main sequence stars. In between, objects are classified based on their emission in the infrared and sub-mm parts of the spectrum. These classifications probably give more of an impression of discrete evolutionary stages than is really warranted, but they nonetheless serve as a useful rough guide to the evolutionary state of a forming star. Consider a core of mass $\sim M$, seen in dust or molecular line emission. When a star first forms at its center, the star will be very low mass and very low luminosity, and will heat up only the dust nearest to it, and only by a very small amount. Thus the total light output will still be dominated by the thermal emission of the dust at its equilibrium temperature. The spectral energy distribution of the source will therefore look just like that which prevailed before the star formed.

However, there might be other indicators that a star has formed. For example, the density distribution might show a very sharp, unresolved peak. Another sign that a star has formed might be the presence of an outflow, which all protostars seem to generate. Outflows coming from the center of a core can be detected in a few ways. Most directly, one can see bipolar, high velocity structures in molecular emission (Figure 29).

One can also detect indirect evidence of an **outflow**, from the presence of highly excited molecular line emission that is produced by shocks at hundreds of km s^{-1} . One example of such a line is SiO($2 \rightarrow 1$) line, which is generally seen in gas moving at several tens of km s^{-1} with temperatures of several hundred K – this is taken to be indication that emission in this line is produced in warm shocks. Since we know of no processes other than formation of a compact object with a $\gtrsim 100 \text{ km s}^{-1}$ escape velocity that can accelerate gas in molecular clouds to such speeds, the presence of such an outflow is taken to indicate that a compact object has formed.

These are the earliest indications of star formation we have available to us. We call objects that show one of these signs, and do not fall into one of the other categories, **class 0 sources**. The dividing line between class 0 and class 1 is that the star begins to produce enough heating that there is non-trivial infrared emission. Before the advent of Spitzer and Herschel, the dividing line between class 0 and 1 was taken to be a non-detection in the IR, but as more sensitive IR telescopes became available, the detection limit went down of course. Today, the dividing line is taken to be a luminosity cut. A source is said to be

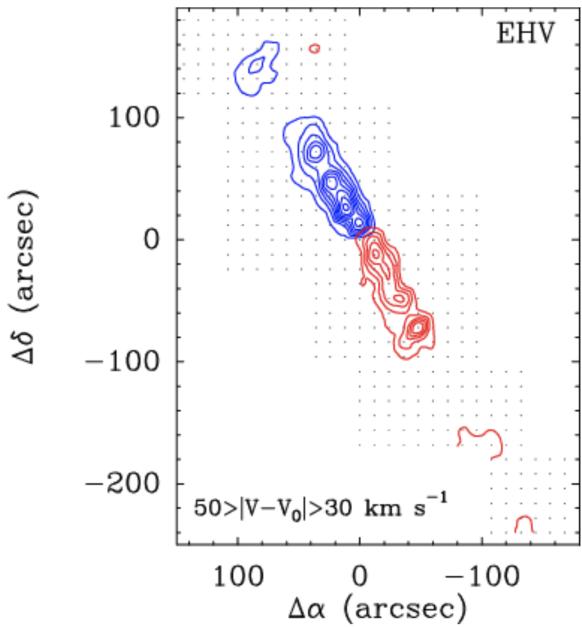


Figure 29:
An integrated intensity map in CO(2 → 1), showing material at velocities between ±30–50 km s⁻¹ (blue and red contours, respectively) relative to the mean (Tafalla et al., 2004). Contours are spaced at intensities of 1 K km s⁻¹. The outflow shown is in the Taurus star-forming region. Figure taken from Krumholz (2015).

class 0 if more than 0.5% of its total bolometric output emerges at wavelengths longer than 350 μm, i.e., if $L_{\text{smm}}/L_{\text{bol}} > 0.5\%$, where L_{smm} is defined as the luminosity considering only wavelengths of 350 μm and longer (see Figure 30).

In practice, measuring L_{smm} can be tricky because it can be hard to get absolute luminosities (as opposed to relative ones) correct in the sub-mm, so it is also common to define the class 0 – 1 divide in terms of another quantity: the bolometric temperature T_{bol} . This is defined as the temperature of a blackbody that has the same flux-weighted mean frequency as the observed SED. That is, if F_ν is the flux as a function of frequency from the observed source, then we define T_{bol} by the implicit equation

$$\frac{\int \nu F_\nu d\nu}{\int F_\nu d\nu} = \frac{\int \nu B_\nu(T_{\text{bol}}) d\nu}{\int B_\nu(T_{\text{bol}}) d\nu}.$$

The class 0 – 1 dividing line is also sometimes taken to be $T_{\text{bol}} = 70$ K. In cases where L_{smm} is accurately measured, T_{bol} is observed to be a reasonably good proxy for $L_{\text{smm}}/L_{\text{bol}}$.

Once protostars reach **class I**, their evolution into further classes is defined in terms of the infrared SED. The motivating cartoon is as follows. At early times, the envelope of dust around the protostar is very optically thick at visible and even near infrared wavelengths. As a result, we don't get to see the stellar photosphere at all. All the radiation is absorbed by the envelope. The dust is in thermal equilibrium, so it re-radiates that energy. Since the radius of the sphere of dust is much larger than that of the star, and the luminosity radiated by the dust must ultimately be equal to that of the star, this emission must be at lower temperature and thus longer wavelengths. Thus as the radiation propagates outward through the dust it is shifted to longer and longer wavelengths. However, dust opacity decreases with wavelength, and thus eventually the radiation is shifted to wavelengths where the remaining dust is optically thin, and it escapes. What we observe is therefore not a stellar photosphere, but a “dust photosphere”.

Given this picture, the greater the column density of the dust around the star, the further it will have to diffuse in wavelength in order to escape. Thus the wavelength at which the emission peaks, or, roughly equivalently, the slope of the spectrum at a fixed wavelength, is a good diagnostic for the amount of circumstellar dust. Objects whose SEDs peak closer to the visible are presumed to be more evolved, because they have lost most of their envelopes.

More formally, this classification scheme was based on fluxes as measured by the IRAS satellite. We define

$$\alpha_{\text{IR}} = \frac{d \log \lambda F_\lambda}{d \log \lambda} \text{ [dimensionless]},$$

as the **IR spectral index**. In practice we measure α_{IR} using two points from the IRAS SED: 2.2 μm and 10 – 25 μm. More positive values of α_{IR} indicate SEDs that peak at longer wavelengths, further into the IR, while more negative values indicate SEDs that peak closer to visible. We define sources with $\alpha_{\text{IR}} \geq 0.0$ (i.e., rising at longer wavelengths from 2 – 25 μm, as class I sources. Alternately, in terms of bolometric temperature, the class I to class II transition is generally taken to be at 650 K.

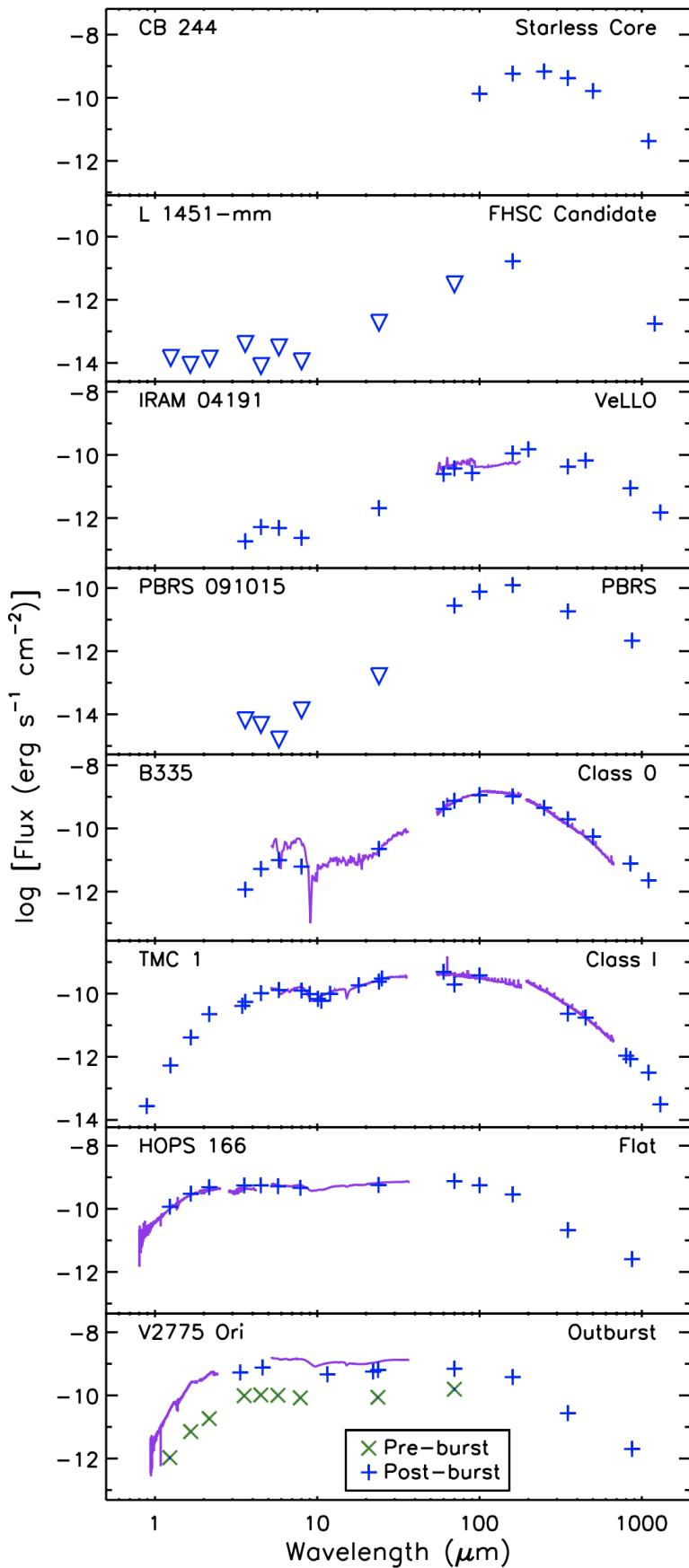


Figure 30: Sample SEDs of protostellar cores, together with the assigned class, as collected by Dunham et al. (2014). Figure taken from Krumholz (2015).

As more of the envelope accretes, it eventually becomes optically thin at the peak emitting wavelengths of the stellar photosphere. In this case we see the stellar blackbody spectrum, but there is also **excess**

infrared emission coming from the disk of warm, dusty gas that till surrounds the star. Thus the SED looks like a stellar blackbody plus some extra emission at near- or mid-IR wavelengths. Stars in this class are also known as **classical T Tauri stars**, named for the first object of the class, although the observational definition of a T Tauri star is somewhat different than the IR classification scheme, so the alignment may not be perfect.

In terms of α_{IR} , these stars have indices in the range $-1.6 < \alpha_{\text{IR}} < 0$. (Depending on the author, the break-point may be placed at -1.5 instead of -1.6 . Some authors also introduce an intermediate classification between 0 and I, which they take to be $-0.3 < \alpha_{\text{IR}} < 0.3$.) A slope of around -1.6 is what we expect for a bare stellar photosphere without any excess infrared emission coming from circumstellar material. Since the class II phase is the last one during which there is a disk of any significant mass, this is also presumably the phase where planet formation must occur.

The final class is **class III**, which in terms of SED have $\alpha_{\text{IR}} < -1.6$. Stars in this class correspond to **weak-line T Tauri stars**. The SEDs of these stars look like bare stellar photospheres in the optical through the mid-infrared. If there is any IR excess at all, it is in the very far IR, indicating that the emitting circumstellar material is cool and located far from the star. The idea here is that the disk around them has begun to dissipate, and is either now optically thin at IR wavelengths or completely dissipated, so there is no strong IR excess.

However, these stars are still not mature MS stars. First of all, their temperatures and luminosities do not correspond to those of MS stars. Instead, they are still puffed up to larger radii, so they tend to have either lower effective temperatures or higher bolometric luminosities (or both) than MS stars of the same mass. Second, they show extremely high levels of magnetic activity compared to MS stars, producing high levels of x-ray emission. Third, they show lithium absorption lines in their atmospheres. This is significant because lithium is easily destroyed by nuclear reactions at high temperatures, and no main sequence stars with convective photospheres show Li absorption. Young stars show it only because there has not yet been time for all the Li to burn.

To summarize,

- **Class 0** protostars are objects that are so heavily obscured that their spectra peak at $\lambda > 100 \mu\text{m}$. For these sources, α_{IR} is not a useful characteristic, because the source may be invisible at $2.2 \mu\text{m}$, and at $10 \mu\text{m}$ there may be deep absorption by cold foreground silicate dust. Inward motions of the gas are sometimes revealed by asymmetric profiles of molecular emission lines. The lifetime of a Class 0 object is short, $\sim 1 - 3 \times 10^4 \text{ yr}$.
- **Class I** protostars have $\alpha_{\text{IR}} > 0$: there is more power being radiated near $10 \mu\text{m}$ than near $2 \mu\text{m}$. Blackbodies with $T < 870 \text{ K}$ have $\alpha_{\text{IR}} > 0$. Class I protostars are thought to have typical ages $\sim 1 - 2 \times 10^5 \text{ yr}$.
- **Class II** YSOs have $-1.5 < \alpha_{\text{IR}} < 0$. Blackbodies with $870 < T < 1540 \text{ K}$ have such IR spectral indices. Class II YSOs correspond to classical T Tauri stars, which are pre-main-sequence stars, still undergoing gravitational contraction, with substantial accretion disks and accretion rates $\sim 10^{-6} M_{\odot} \text{ yr}^{-1}$.
- **Class III** protostars have $\alpha_{\text{IR}} < -1.6$. Blackbodies with $T > 1540 \text{ K}$ have such spectral indices. Class III YSOs correspond to **weak-lined T Tauri stars**, which are pre-main-sequence stars still undergoing gravitational contraction, but where the accretion disk is either weak or perhaps entirely absent.

The observed flux ratios are determined by the temperature of the emitting regions (star and inner disk) and by wavelength-dependent extinction. Classification of a given object as either Class I or Class II may depend on the source orientation. If we are viewing the object face-on it, may be classified as Class II, but an identical disk viewed edge-on could heavily redden the light reaching the observer, so that the object could be classified as Class I or even Class 0.

Infall signatures: The H α is particularly interesting, because it tells us something about the star's immediate environment. For MS stars, the H α line profile is a result of absorption at the stellar photosphere and emission from the chromosphere. At the photosphere there is a population of neutral hydrogen atoms in the $n = 2$ level that absorbs photons at H α frequencies, producing absorption. Above that in the chromosphere is an optically thin, hot gas, which contains atoms in the $n = 3$ level. Some of these emit H α photons, partially filling in the absorption trough, but leaving the line overall in absorption. Producing H α in emission is tricky, however. The emitting material must be above the stellar photosphere, so it can fill in the absorption trough created there. This gas must be at temperatures of $5,000 - 10,000 \text{ K}$ to significantly populate the $n = 3$ level. However, in order to produce enough H α photons to fill in the trough and produce net emission, this gas must also be dense enough for the collision rate to be high enough to force the $n = 3$ level close to LTE.

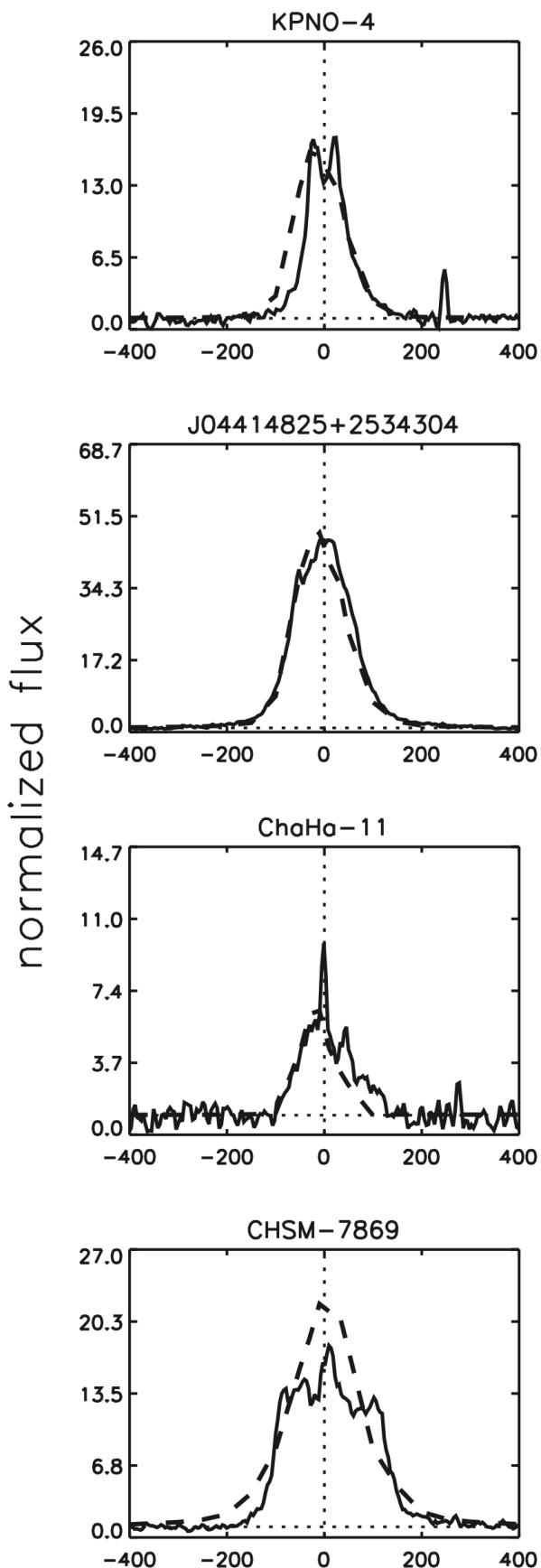


Figure 31: Comparisons between observed (solid) and model (dashed) H α line profiles for a sample of T Tauri stars (Muzerolle et al., 2005). The x axis shows velocity in km s^{-1} . Each model curve is a fit in which the accretion rate is one of the free parameters. Figure taken from Krumholz (2015).

Ordinary stellar chromospheres have densities that are much too low to meet this requirement. This H α emission implies the presence of material around the star at temperatures of 5,000 – 10,000 K, but at densities much higher than found in an ordinary stellar chromosphere. Moreover, the width of the H α emission requires that this material be moving at velocities of hundreds of km s $^{-1}$ relative to the stellar surface (i.e., comparable to the free-fall velocity). This cannot be thermal broadening, because this would require temperatures of $\sim 10^6$ K high enough to completely ionize hydrogen. It must therefore be bulk motion.

The standard inference is that this indicates the presence of gas infalling onto the stellar surface. Such gas would provide the high densities required to produce H α in emission. The infall of this material would provide the requisite bulk motion. Internal shocks and the shocking of this gas against the stellar surface could easily heat gas to the required temperatures. Finally, this hot material would also produce continuum emission, explaining the continuum veiling.

Quantitative radiative transfer calculations that attempt to fit the observed veiling and line emission can be used to infer the densities and velocities of the circumstellar gas, thereby constraining the accretion rate (Figure 31). The inferred accretion rates depend on the strength of the H α emission, and are typically 10^8 M \odot yr $^{-1}$. There is a broad range, however, running from $10^{-11} – 10^{-6}$ M \odot yr $^{-1}$, with a very rough correlation $\dot{M}_* \propto M_*^2$.

These accretion rates are generally low enough so that accretion luminosity does not dominate over stellar surface emission. However, the estimated accretion rates are extremely uncertain, and the models used to make these estimates are very primitive. In general they simply assume that a uniform density slab of material arrives at the free-fall velocity, and covers some fraction of the stellar surface, and the accretion rate is inferred by determining the density of this material required to produce the observed spectral characteristics.

Despite this caveat, though, the H α line and other optical properties do seem to indicate that there must be some dense infalling material even around these stars that lack obvious envelopes. This in turn requires a reservoir of circumstellar material not in the form of an envelope, which is most naturally provided by a disk. Indeed, before the advent of space-based IR observatories, optical indicators like this were the only real evidence we had for disks around T Tauri and Herbig Ae/Be stars.

Disk lifetimes: We now turn to the disks that surround T Tauri and similar stars. Prior to the 2000s, we had very little direct information about such objects, since they are not visible in the optical. That changed dramatically with the launch of space-based IR observatories, and the development of ground-based millimeter interferometers. These new techniques made it possible to observe disks directly for the first time.

One of the most interesting properties of disks for those who are interested in planets is their lifetimes. This sets the limit on how long planets have to form in a disk before it is dispersed. In discussing disk lifetimes, it is important to be clear on how the presence or absence of a disk is to be inferred, since different techniques probe different parts and types of disks. In general what all these techniques have in common is that one uses some technique to survey young star clusters for disks. The clusters can be age-dated using pre-main sequence or main sequence HR diagrams. One then plots the disk fraction against age.

One signature of disks we have already discussed: optical line emission associated with accretion in T Tauri stars, particularly H α . Surveys of nearby groups find that H α line emission usually disappears at times between 1 and 10 Myr (Figure 19.4). This tells us that the inner parts of disks ($\lesssim 1$ AU) which feed stars disappear over this time scale. In contrast, ground-based near IR observations tell us about somewhat more distant parts of the disk, out to a few AU. The timescales implied by these results are very similar to those obtained from the H α : roughly half the systems lose their disks within ~ 3 Myr.

These observations are sensitive primarily to the inner disk, and the IR techniques are generally sensitive only in cases where the dust in these regions is optically thick. Some optically thin material could still be present and would not have been detected. Observations at longer wavelengths, such as Spitzer's 24 μ m band and in the mm regime from ground-based radio telescopes, probe further out in disks, at distances of $\sim 10 – 100$ AU. They are also sensitive to much lower amounts of material. Interestingly, unlike the shorter wavelength observations, these measurements indicate that a small but non-zero fraction of systems retain some disks out to times of $\sim 10^8$ yr. The amounts of mass needed to explain the long wavelength excess are only $\sim 10^{-5}$ M \odot in dust. Thus the older systems are likely looking at an even later evolutionary phase than T Tauri disks, one in which almost all the gas and inner disk material is gone. These are **debris disks**, which are thought to originate from collisions between larger bodies than those made up of dust from interstellar gas.

Transition disks: The observation that accreting disks and inner optically thick disks disappear on a few Myr timescales, but that some fraction leave behind very small amounts of mass in the outer disk is a very interesting one. Observationally, we would like to get some constraints on how the process of a gaseous, accreting T Tauri disk transitioning to a low-mass debris disk occurs. For this purpose, there is

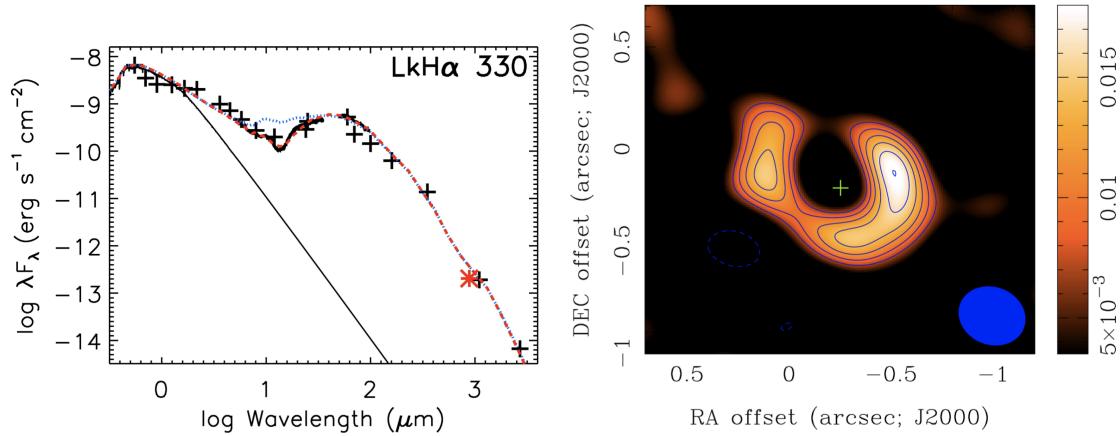


Figure 32: (**Left:**) The SED of the star LkH α 330 (Brown et al., 2008). Plus signs indicate measurements. The black line is a model for a stellar photosphere. The blue line is a model for a star with a disk going all the way to the central star, while the red line is a model for disk with a 40 AU hole in its center. Figure taken from Krumholtz (2015). (**Right:**) Dust continuum image of the disk around LkH α 330, taken at 340 GHz by the SMA (Brown et al., 2008). Colors show the detected signal, and contours show the signal to noise ratio, starting from S/N of 3 and increasing by 1 thereafter. The green plus marks the location of the star. The blue circle is the SMA beam. Figure taken from Krumholtz (2015).

an intriguing class of objects known as **transition disks**. Spectrally, these are defined as objects that have a significant $24\text{ }\mu\text{m}$ excess (or excess at even longer wavelengths), but little or no IR excess (Figure 34, left). This SED suggests a natural physical picture: a disk with a hole in its center. The short wavelength emission normally comes from near the star, and the absence of material there produces the lack of short wavelength excess. Indeed, it is possible to fit the SEDs of some stars with models with holes.

In the last decade it has become possible to confirm the presence of inner holes in transition disks directly, at least cases where the inferred hole is sufficiently large (Figure 34, right). The sizes of the holes inferred by the observations are generally very good matches to the values inferred by modeling the SEDs. The holes are remarkably devoid of dust: upper limits on the masses of small dust grains within the hole are often at the level of $\sim 10^{-6} M_{\odot}$. The sharp edges of the holes indicate that the effect driving them isn't simply the growth of dust grains to larger sizes, which should produce a more gradual transition. Instead, something else is at work. However, in some transition disks gas is still seen within the gap in molecular line emission, which also suggests that whatever mechanism is removing the dust does not necessarily get rid of all the gas as well.

1.13.3 Follow-up Questions

- How does the spectrum change as planets start to form?

1.14 Question 14

Sketch the spectral energy distribution (SED) of a T Tauri star surrounded by a protoplanetary disk. How would the SED change: (a) if the disk develops a large inner hole, (b) if the dust grains in the disk grow in size by agglomeration (with the same total mass)?

1.14.1 Short answer

Figure 33 shows a schematic of what the SED of a T Tauri star surrounded by a protoplanetary disk would look like. It can be seen that the total SED is a composite of the stellar blackbody plus the protoplanetary disk component.

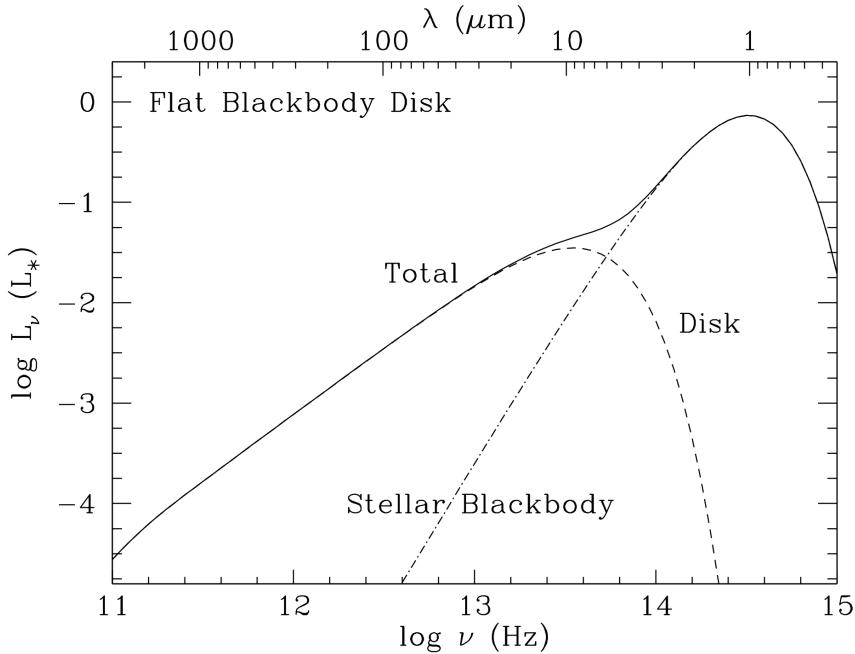


Figure 33:
SED for the flat blackbody disk, with contributions from star and disk identified. Figure taken from Chiang & Goldreich (1997).

Figure 34 (left) shows the SED of a T Tauri star surrounded by a protoplanetary disk with a hole in its center. The short wavelength emission normally comes from near the star, and the absence of material there produces the lack of short wavelength excess. Figure 34 (right) shows a continuum image of such a disk.

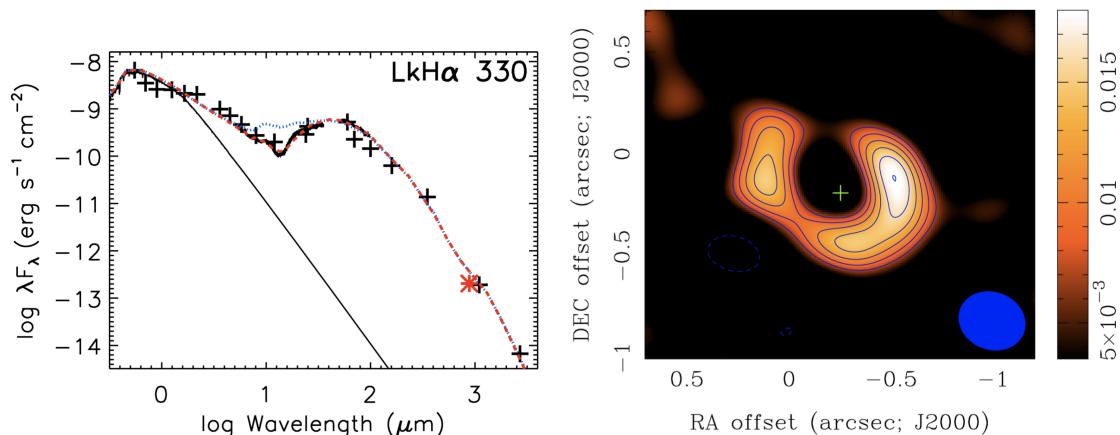


Figure 34: (Left:) The SED of the star LkH α 330 (Brown et al., 2008). Plus signs indicate measurements. The black line is a model for a stellar photosphere. The blue line is a model for a star with a disk going all the way to the central star, while the red line is a model for disk with a 40 AU hole in its center. Figure taken from Krumholtz (2015). (Right:) Dust continuum image of the disk around LkH α 330, taken at 340 GHz by the SMA (Brown et al., 2008). Colors show the detected signal, and contours show the signal to noise ratio, starting from S/N of 3 and increasing by 1 thereafter. The green plus marks the location of the star. The blue circle is the SMA beam. Figure taken from Krumholtz (2015).

1.14.2 Additional context

In the pre-main sequence phase, that is, while developing from a voluminous protostar into a star, the accreting stellar objects are powered by gravitational energy, their central temperature still being too low for hydrogen fusion. This situation of stellar evolution is represented by very young protostars ($10^5 - 10^8$ yr) observed in nebulae where star formation is going on, so-called T Tauri stars (named after their prototype in the constellation of Taurus), which, in many cases, are embedded in protoplanetary nebulae and discs. T-Tauri stars are variable stars; their erratic brightness changes due to instabilities of the accretion disc, violent activity (i.e., stellar winds) in the thin stellar atmosphere (characterized by strong emission lines, mainly H α of the Balmer series, and Ca $^{+}$), and obscuration by parts of the inhomogeneous surrounding cloud. A very characteristic feature of the T Tauri stars is the higher abundance of lithium relative to the main sequence stars detectable by its 670.7 nm line. The lithium isotope ^7Li is formed in the pp (protonproton) chain but subsequently eliminated, at temperatures $> 2.5 \times 10^6$ K by lithium burning during the last highly convective stages when the star enters the MS. The rapid rotation of the T Tauri stars, typically between 1 and 12 days, enforces the transport of lithium into the hot stellar core and eventually lithium depletion.

Excess infrared (IR) emission from T Tauri stars is thought to arise from circumstellar disks. The standard disk is modeled as a flat blackbody. In the simplest scenario, it lacks intrinsic luminosity and passively re-radiates the energy it absorbs from the central star. This blackbody model yields an IR spectral energy distribution (SED) of the form $\nu F_\nu \propto \nu^\alpha$, with spectral index $\alpha = 4/3$. By coincidence, the steady state accretional luminosity from an opaque, thin disk yields an SED with identical spectral index. SEDs measured between 3 and 100 μm are usually well fitted by power laws with $0 \leq \alpha \leq 4/3$. However, in contrast to the prediction of the standard model, most sources exhibit flattish spectra having $n \leq 3/4$.

The failure of the standard disk model spawned a number of alternative proposals. One of these includes a blackbody disk whose surface flares outward with increasing radius as a consequence of vertical hydrostatic equilibrium. Flared disks intercept more stellar radiation than flat ones, especially at large distances from the star. Other models invoke either an “active” disk having a high intrinsic luminosity or a dusty component in addition to the disk.

T Tauri stars have temperatures $T_{\text{TT}} \sim 4000$ K, masses $M_{\text{TT}} \sim 0.5 M_\odot$, and radii $R_{\text{TT}} \sim 2.5 R_\odot$. Dust, which is uniformly mixed with the gas, comprises about 1% of the total mass. Dust grains dominate the continuum opacity from visible through millimeter wavelengths. Dust grains modelled as being spherical have a radius $r \sim 0.1 \mu\text{m}$, mass density $\rho_d \sim 2 \text{ g cm}^{-3}$, and negligible albedo. Their emissivity, which is unity for $\lambda \leq 2\pi r$, decreases as $\epsilon_\lambda \sim (2\pi r/\lambda)^\beta$ for $\lambda > 2\pi r$. We denote by $\epsilon_\lambda \sim (8\pi r k_B T/hc)^\beta \sim (T/T_{\text{TT}})^\beta$ ($\beta \approx 1$) the average dust emissivity at temperature T . The dust opacity at visual wavelengths is $\kappa_V \sim 400 \text{ cm}^2 \text{ g}^{-1}$, which implies an optical depth of $\tau_v \sim 4 \times 10^5 a_{\text{AU}}^{-3/2}$, where a_{AU} is the disk radius measured in AU.

SEDs are computed for disks viewed pole-on. We choose an inner cutoff radius, $a_i \sim 6R_{\text{TT}} \sim 0.07$ AU, to mark the condensation boundary of common silicates. The outer cutoff radius for flat disks, $a_0 \sim 2.3 \times 10^4 R_{\text{TT}} \sim 2.7 \times 10^2$ AU, is fixed to facilitate comparisons among different disk models; a_0 is comparable to the size of the largest disks as seen in silhouette against the Orion Nebula.

To begin, we review standard relations for the temperature and SED of a blackbody disk. The flux of stellar radiation incident upon the disk is $(\alpha/2)(R_{\text{TT}}/a)^2 \sigma T_{\text{TT}}^4$ for $a \gg R_{\text{TT}}$, where α is the grazing angle at which the starlight strikes the disk. Equating the emitted and absorbed fluxes yields the disk temperature

$$T_e \sim \left(\frac{\alpha}{2}\right)^{1/4} \left(\frac{R_{\text{TT}}}{a}\right)^{1/2} T_{\text{TT}} [\text{K}].$$

The SED is computed as

$$L_\nu \equiv 4\pi d^2 \nu F_\nu = 8\pi^2 \nu \int_{a_i}^{a_0} a B_\nu(T_e) da [\text{erg s}^{-1}],$$

where $B_\nu(T)$ is the Planck function, and d is the distance to the source. A scaling relation for L_ν at wavelengths between those that characterize the disk at a_i and a_0 is derived as follows. L_ν is approximated as $8\pi^2 a^2 \nu B_\nu(T_e)$, $\nu B_\nu(T_e)$ is replaced by $\sigma T_e^4 / \pi$, a is related to T_e by $8\pi^2 a^2 \nu B_\nu(T_e)$, and T_e is expressed in terms of ν by $3k_B T_e \sim h\nu$. These yield

$$L_\nu \sim 8\pi a^2 \sigma T_e^4 \sim \alpha L_{\text{TT}} [\text{erg s}^{-1}].$$

In other words, the fraction of the stellar luminosity L_{TT} that is reprocessed to frequencies in an octave centered on ν is approximately equal to the grazing angle α at the location where $3k_B T \sim h\nu$.

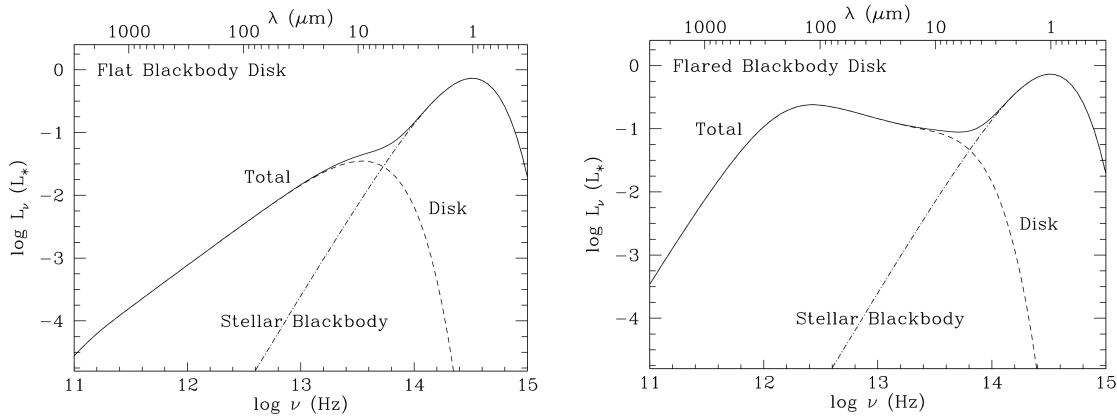


Figure 35: (Left): SED for the flat blackbody disk, with contributions from star and disk identified. The $n = 4/3$ law is evident between $30 \mu\text{m}$ and 1mm . The turnover near 1mm is due to the truncation of the disk at $a_0 = 270 \text{ AU}$. (Right): SED for the flared blackbody disk. At mid-IR wavelengths, $L_\nu \propto \nu^{-2/3}$. At longer wavelengths, $L_\nu \propto \nu^3$. Figure taken from Chiang & Goldreich (1997).

Flat geometry: A flat disk is one whose **aspect ratio** (i.e., opening angle) is independent of α ; let's assume it to be much less than unity. The grazing angle appropriate to this geometry is $\alpha \sim 0.4R_{\text{TT}}/a \ll 1$ for $a/R_{\text{TT}} \ll 1$. Thus, T_e takes the form

$$T_e \sim \left(\frac{2}{3\pi}\right)^{1/4} \left(\frac{R_{\text{TT}}}{a}\right)^{3/4} T_{\text{TT}} [\text{K}].$$

Application of the scaling relation given by $L_\nu \sim 8\pi a^2 \sigma T_e^4 \sim \alpha L_{\text{TT}}$ to the flat disk gives $L_\nu \sim 0.01(\nu/10^{13} \text{ Hz})^{4/3} L_{\text{TT}}$. Figure 1 (left) confirms that L_ν obeys this relation for $30 \lesssim \lambda \lesssim 1000 \mu\text{m}$.

Flared geometry: We retain the assumption that the disk radiates as a blackbody and consider the consequences of vertical hydrostatic equilibrium in a gravitational field $g = \Omega^2 z$.

The disk temperature T_e is still given by

$$T_e \sim \left(\frac{\alpha}{2}\right)^{1/4} \left(\frac{R_{\text{TT}}}{a}\right)^{1/2} T_{\text{TT}} [\text{K}],$$

but now the grazing angle α takes the more general form

$$\alpha \sim \frac{0.4R_{\text{TT}}}{a} + a \frac{d}{da} \left(\frac{H}{a}\right) [\text{deg}],$$

where H is the height of the visible photosphere above the disk midplane.

Taking the gas to be isothermal at temperature T_e yields a Gaussian density profile

$$\frac{n}{n_0} = \exp\left(-\frac{z^2}{2h^2}\right) [\text{dimensionless}],$$

where

$$\frac{h}{a} = \left(\frac{T_e}{T_c}\right)^{1/2} \left(\frac{a}{R_{\text{TT}}}\right)^{1/2} [\text{dimensionless}].$$

The temperature T_c is a measure of the gravitational potential at the surface of the central star;

$$T_c \equiv \frac{GM_{\text{TT}}\mu}{k_B R_{\text{TT}}} \sim 8 \times 10^6 [\text{K}],$$

with μ being the mean molecular weight of the gas. Under the assumption that the dust-to-gas ratio is uniform throughout the disk, we have

$$\frac{H}{h} = \left[2 \ln\left(\frac{n_0}{n_{\text{ph}}}\right)\right]^{1/2} [\text{dimensionless}],$$

where n_{ph} is the number density of the photosphere. In the limit of large radius where α is dominated by the flaring term,

$$\frac{H}{a} \sim 4 \left(\frac{T_{\text{TT}}}{T_c}\right)^{4/7} \left(\frac{a}{R_{\text{TT}}}\right)^{2/7} \sim 0.18a_{\text{AU}}^{2/7} [\text{dimensionless}].$$

It is assumed that $H/h = 4$; in reality, this factor declines from about 5 to 4 between $a_{\text{AU}} = 3$ to $a_{\text{AU}} = 4$. It follows that $\alpha \sim 0.005a_{\text{AU}}^{-1} + 0.05a_{\text{AU}}^{2/7}$. Thus, α is minimal at the transition radius $a_{\text{tr}} \sim 0.4 \text{ AU}$ with $\alpha_{\min} \sim 0.05$. Beyond the transition radius, the disk flares until $H \sim a$ at $a_0 \sim 270 \text{ AU}$. The SED for the flared blackbody disk truncated at a_0 is shown in Figure 35 (right). At wavelengths between 10 and $100 \mu\text{m}$, the disk emission follows the scaling relation $L_\nu \propto \alpha L_{\text{TT}} \sim 0.1(\nu/10^{13} \text{ Hz})^{-2/3} L_{\text{TT}}$. Longward of $300 \mu\text{m}$ in the Rayleigh-Jeans regime, the SED varies as ν^3 .

1.15 Question 15

What are the primary origins of the heat lost to space by infrared luminosity of Jupiter, Earth, and Io?⁶

1.15.1 Short answer

1. **Jupiter:** Found to emit ~ 2.5 times the energy that it absorbs from the Sun, which suggests that it might still be contracting (consistent with a rate of $0.5 \text{ mm century}^{-1}$) or undergoing differentiation of hydrogen and helium in the atmosphere.
2. **Earth:** Main source of IR luminosity is re-radiated Solar flux. The Earth has an albedo of ~ 0.3 , which means it absorbs $\sim 70\%$ of the incident Solar radiation.
3. **Io:** Tidal heating via internal tidal forces not only supplies the IR luminosity of Io, but maintains its strong volcanic activity.

1.15.2 Follow-up Questions

- Saturn is a gas giant like Jupiter; is it also believed to be contracting?
- What supplies the tidal heating of Io (i.e., what is physically happening)?

⁶Answer shamelessly stolen from Miranda.

1.16 Question 16

Explain the observational problem of radius inflation for hot Jupiters and describe two possible solutions.

1.16.1 Short answer

Planet structure and evolution models had predicted there should be a clear maximum size for evolved giant planets. In the absence of any additional energy source, even a completely gaseous planet should never be larger than $\sim 1.2 R_J$ by the time it is several billion years old. Nonetheless, at short orbital periods there are now dozens of highly inflated hot Jupiters, some with radii that are almost $\sim 2 R_J$. Moreover, not all hot Jupiters are inflated to the same degree; instead, we see that even among the highly irradiated gas giants, the most irradiated planets are much more inflated than those that are slightly further out. There are several possible explanations for this observational problem:

- invoking an extra heat source for the interior, perhaps caused by orbital tidal heating, “kinetic heating” in which wind energy is converted into heat, obliquity tides when in a Cassini state, Ohmic dissipation, or penetration of gravity waves into the planetary interior that then dissipate at depth
- enhanced atmospheric opacities which delay the loss of heat and entropy and retain an inflated atmosphere
- a magnetohydrodynamic mechanism with Ohmic heating and attendant magnetic braking of the velocity field at sufficient atmospheric ionization fraction

1.16.2 Additional context

Since the discovery of the first transiting hot Jupiters, models have sought to explain the anomalously large radii of highly irradiated gas giants. We now know that the size of hot Jupiter radius anomalies scales strongly with a planet’s level of irradiation and numerous models like tidal heating, ohmic dissipation, and thermal tides have since been developed to help explain these inflated radii. In general, however, these models can be grouped into two broad categories: models that directly inflate planetary radii by depositing a fraction of the incident irradiation into the interior and models that simply slow a planet’s radiative cooling, allowing it to retain more heat from formation and thereby delay contraction.

1.17 Question 17

Explain the effects of an atmosphere on a planet's surface temperature and the position of the habitable zone. What special considerations must one make for habitability around M-type stars?

1.17.1 Short answer

There are several ways that a planet's atmosphere can affect its surface temperature:

- clouds reflecting heat from the parent star resulting in a higher albedo and lower surface temperature;
- winds transporting heat from the day-side to the night-side, regulating the surface temperature; and
- greenhouse gasses retaining heat and increasing the surface temperature.

Special consideration must be made for M-type stars due to increased stellar activity and intense ultra-violet radiation.

1.17.2 Additional context

The temperatures of planets played a key role in their formation and evolution. During the formation stage, the temperature structure of the Solar nebula influenced whether a planet would become a terrestrial or a giant planet. Temperature also helped to determine the current composition of each planet's atmosphere.

The Stefan-Boltzmann equation is the most significant factor in determining the present-day temperatures of the planets in the Solar System. Under equilibrium conditions, a planet's total energy content must remain constant. Therefore, all of the energy absorbed by the planet must be re-emitted; if this were not so, the planet's temperature would change with time.

To estimate a planet's equilibrium temperature, assume that the planet is a spherical blackbody of radius R_p and temperature T_p in a circular orbit a distance D away from the Sun. For simplicity, we will assume that the planet's temperature is uniform over its surface that that the planet reflects a fraction a of the incoming sunlight (a is known as the planet's *albedo*). From the condition of thermal equilibrium, the sunlight that is not reflected must be absorbed by the planet and subsequently re-emitted as blackbody radiation. Of course, we will also treat the Sun as a spherical blackbody having an effective temperature $T_\odot = T_e$ and radius R_\odot . The temperature of the planet is thus given by

$$T_p = T_\odot(1 - a)^{1/4} \sqrt{\frac{R_\odot}{2D}} [\text{K}].$$

Note that the temperature of the planet is proportional to the effective temperature of the Sun and does not depend on the size of the planet.

What's not included here is the *greenhouse effect*, a significant source of warming (due largely to atmospheric water vapour, carbon dioxide, methane, and chlorofluorocarbons in the case of Earth). According to Wein's law, Earth's blackbody radiation is emitted primarily as infrared wavelengths. This infrared radiation is absorbed then re-emitted by the atmospheric greenhouse gasses, which act as a thermal blanket to warm Earth's surface by about 34°C. Greenhouse warming on Venus has been more dramatic.

1.18 Question 18

Explain the process of nuclear fusion and give two examples of important fusion processes that affect the lives of stars.

1.18.1 Short answer

Most observed stars (including the Sun) live on **thermonuclear fusion**. In such nuclear reactions, induced by the thermal motion, several lighter nuclei fuse to form a heavier one.

1.18.2 Additional context

General properties of the nucleus: The notion that the atom can be viewed as being composed of a nucleus surrounded by a cloud of electrons which are confined to shells led to a very successful theory of atomic spectra. A very similar picture can be postulated for the nucleus itself, namely, that nucleons are arranged in shells within the nucleus and undergo transitions from one excited state (shell) to another subject to the same sort of selection rules that govern atomic transitions. The origin of the shell structure of any nucleus is that nucleons are fermions and therefore must obey the **Pauli Exclusion Principle**, just as the atomic electrons do. Thus, only two protons or two neutrons may occupy a specific cell in phase space (protons and neutrons have the same spin as electrons, so each species can have two of its kind in a quantum state characterized by the spatial quantum numbers).

However, the nucleons are much more tightly bound in the nucleus than the electrons in the atom. Whereas the typical ionization energy of an atom can be measured in tens to thousands of electron volts, the typical binding energy of a nucleon in the nucleus is several million electron volts. This large binding energy and the Pauli Exclusion Principle can be used to explain the stability of the neutron in nuclei. Although free neutrons beta-decay to protons (and an electron and an electron anti-neutrino) with a half-life of about 10 min, neutrons appear to be stable when they are in nuclei. If neutrons did decay, the resulting proton would have to occupy one of the least tightly bound proton shells, which frequently costs more energy than is liberated by the beta decay of the neutron. Thus, unless the neutron decay can provide sufficient energy for the decay products to be ejected from the nucleus, the neutron must remain in the nucleus as a stable entity.

For a nucleus to be stable, its mass must be less than the sum of the masses of any possible combination of its constituents. Thus, ${}^5\text{Li}$ is not stable, whereas ${}^4\text{He}$ is. However, note that the instability of mass-5 nuclei posed one of the greatest barriers of the century to the understanding of the evolution of stars. The nuclear evolution beyond mass 5 was finally solved by Fred Hoyle, who showed that the triple- α process could actually initiate synthesis of all the nuclei heavier than mass 12.

Most observed stars (including the Sun) live on so-called thermonuclear fusion. In such nuclear reactions, induced by the thermal motion, several lighter nuclei fuse to form a heavier one. Before this process, the involved nuclei j have a total mass ($\sum m_j$) different from that of the product nucleus (m_y). This difference is called the **mass defect**:

$$\Delta m = \sum m_j - m_y \text{ [kg].}$$

This mass defect is converted into energy according to Einstein's formula

$$E = \Delta mc^2 \text{ [eV]}$$

and is available (at least partly) for the star's energy balance. Usually for such a reaction to happen, it must be exothermic. That is, the rest energy of the initial constituents of the reaction must exceed that of the products:

$$\Delta mc^2 > 0 \text{ [eV].}$$

An example is the series of reactions called **hydrogen burning**, where four hydrogen nuclei ${}^1\text{H}$ with a total mass $4 \times 1.0079 \text{ amu}$ (atomic mass units, physical scale) are transformed into one ${}^4\text{He}$ nucleus of 4.0026 amu . Atomic masses are given for the neutral atoms (i.e., for the nucleus plus all electrons). However, since the electron mass is only $1/1823 \text{ amu}$, we will assume that masses of nuclei are the same as the atomic masses. Obviously $2.9 \times 10^{-2} \text{ amu}$ per produced ${}^4\text{He}$ nucleus have "disappeared" during the fusion of the four protons, which is roughly 0.7% of the original masses and which corresponds to an energy of about 27.0 MeV . As usual in nuclear physics, as the unit of energy, we take the electron volt eV ($1 \text{ eV} = 1.6018 \times 10^{-12} \text{ erg}$) with the following equivalences:

$$1 \text{ keV} \equiv 1.1606 \text{ K}$$
$$931.49 \text{ MeV} \equiv 1 \text{ amu.}$$

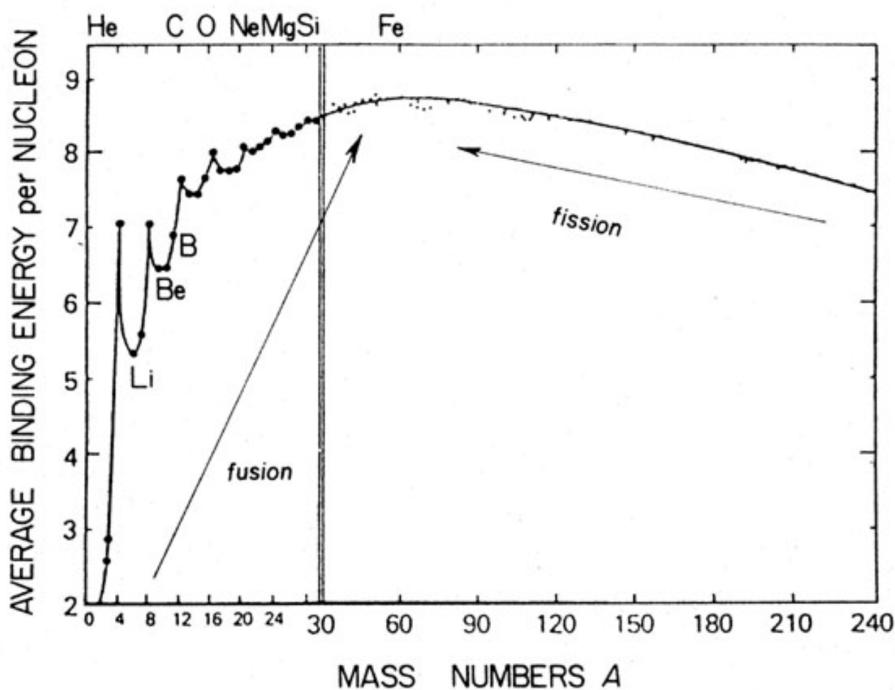


Figure 36: The average binding energy per nucleon f as a function of atomic mass number A . Figure taken from <https://physics.stackexchange.com>.

The Sun's luminosity corresponds to a mass loss rate of $L_{\odot}/c^2 = 4.26 \times 10^{12} \text{ g s}^{-1}$, which appears to be a lot, especially if it is read as "more than four million metric tons per second". If a total of $1 M_{\odot}$ of hydrogen were converted into ${}^4\text{He}$, then the disappearing 0.7% of this mass would be $1.4 \times 10^{31} \text{ g}$, which could balance the Sun's present mass loss by radiation for about $3 \times 10^{18} \text{ s} \sim 10^{11} \text{ yr}$.

The deficiency of mass is just another aspect of the fact that the involved nuclei have different binding energies (E_B). This is the energy required to separate the nucleons (protons and neutrons in the nucleus) against their mutual attraction by the strong, but short-range, nuclear forces. Or else, the binding energy is the energy gained if they are brought together from infinity (which starts here at any distance large compared with, say, 10^{-12} cm , the scale of a nuclear size).

Consider a nucleus of mass M_{nuc} and atomic mass number A (the integer **atomic weight**): it may contain Z protons of mass m_p and $(A - Z)$ neutrons of mass m_n . Its binding energy is then related to these masses by

$$E_B = \Delta mc^2 [\text{MeV}]$$

$$E_B = [(A - Z)m_n + Zm_p - M_{\text{nuc}}]c^2 [\text{MeV}].$$

When comparing different nuclei, it is more instructive to consider the average binding energy per nucleon,

$$f = \frac{E_B}{A} [\text{MeV}]$$

which is also called the **binding fraction**. With the exception of hydrogen, typical values are around 8 MeV, with relatively small differences for nuclei of very different A . This shows that the short-range nuclear forces due to a nucleon mainly affect the nucleons in its immediate neighbourhood only, such that with increasing A , a saturation occurs rather than an increase of f proportional to A . An idealized plot of f against A is shown in Figure 36 (the curve zigzags around this smoothed curve as a consequence of the shell structure of the nucleus and pair effects).

With increasing A , $f(A)$ rises steeply from hydrogen, then flattens out and reaches a maximum of 8.5 MeV at $A = 56({}^{56}\text{Fe})$, after which it drops slowly with increasing A . The increase for $A < 56$ is a surface effect: particles at the surface of the nucleus experience less attraction by nuclear forces than those in the interior, which are completely surrounded by other particles. And in a densely packed nucleus, the surface area increases with radius slower than the volume (i.e. the number A) such that the fraction of surface particles drops. With increasing A , the number Z of protons also increases (the addition of neutrons only would require higher energy states, because the Pauli exclusion principle excludes more than two identical neutrons, and the nuclei would be unstable). The positively charged protons experience a repulsive force which is far-reaching and therefore does not show the saturation of the nuclear forces. This increasing repulsion by the Coulomb forces brings the curve in Figure 36 down again for $A > 56$.

Around the maximum, at ${}^{56}\text{Fe}$, we have the most tightly bound nuclei. In other words, the nucleus of ${}^{56}\text{Fe}$ has the smallest mass per nucleon, so that any nuclear reaction bringing the nucleus closer to this

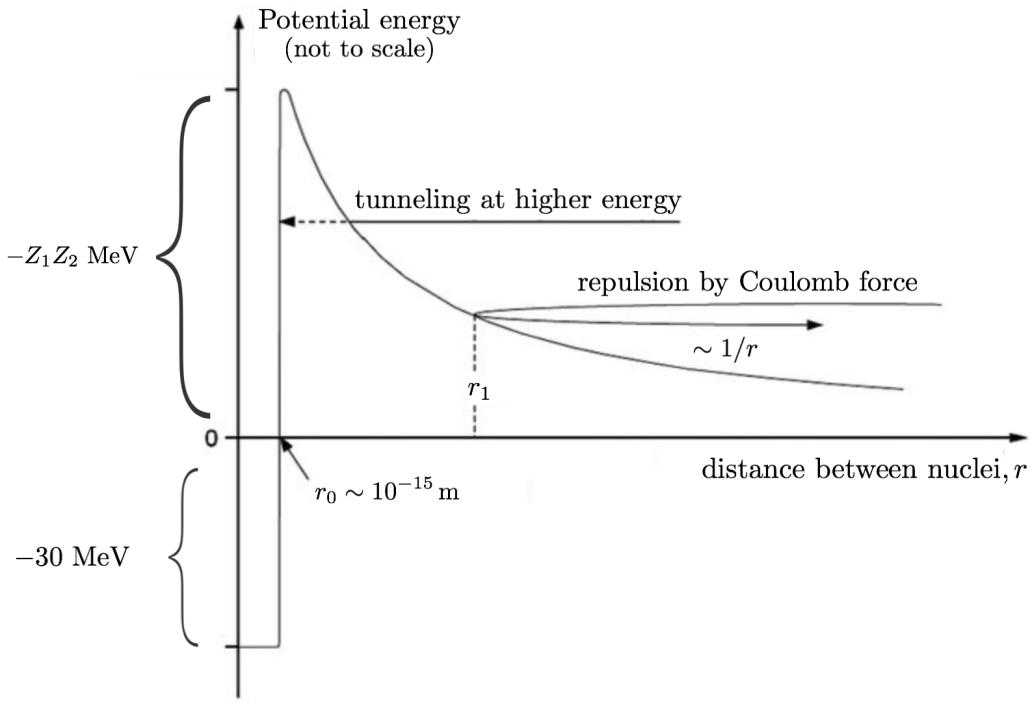


Figure 37: Sketch of the potential over the distance r from the nuclear centre. Nuclear attraction dominates for $r < r_0$ and Coulomb repulsion for $r > r_0$. Figure adapted from Collins (2003) and <https://physics.stackexchange.com>.

maximum will be exothermic (i.e., it will release energy). There are two ways of doing this:

1. By fission of heavy nuclei, which happens, for example, in radioactivity.
2. By fusion of light nuclei, which is the prime energy source of stars (and possibly ours too in the future).

Clearly, both reach an end when one tries to extend them over the maximum of f , which is therefore a natural finishing point for the stellar nuclear engine. So if a star initially consisted of pure hydrogen, it could gain a maximum of about 8.5 MeV per nucleon by fusion to ^{56}Fe , but 6.7 MeV of these are already used up when ^4He is built up in the first step.

In order to obtain a fusion of charged particles, they have to be brought so close to each other that the strong, but very short-ranged, nuclear forces dominate over the weaker, but far-reaching, Coulomb forces. The counteraction of these two forces leads to a sharp potential jump at the interaction radius (Figure 37):

$$r_0 \approx A^{1/3} 1.44 \times 10^{-13} [\text{fm}]$$

(the **nuclear radius** of the order of **femtometer**, $1 \text{ fm} = 10^{-13} \text{ cm}$). For distances less than r_0 , the nuclear attraction dominates and provides a potential drop of roughly 30 MeV, while beyond r_0 , the repulsive Coulomb forces for particles with charges Z_1 and Z_2 yield

$$E_{\text{Coul}} = \frac{Z_1 Z_2 e^2}{r} [\text{MeV}].$$

The height of the Coulomb barrier $E_{\text{Coul}}(r_0)$ is typically of the order

$$E_{\text{Coul}}(r_0) \approx Z_1 Z_2 [\text{MeV}].$$

If, in the stationary reference frame of the nucleus, a particle at infinity has kinetic energy E_1 , it can come classically only to a distance r_1 given by $E_1 = E_{\text{Coul}}(r_1)$, as indicated in Figure 37. Now, the kinetic energy available to particles in stellar interiors is that of their thermal motion, and hence the reactions triggered by this motion are called thermonuclear. Since in normal stars we observe a slow energy release rather than a nuclear explosion, we must certainly expect the average kinetic energy of the thermal motion, E_{th} , to be considerably smaller than $E_{\text{Coul}}(r_0)$. For $T \sim 10^7 \text{ K}$ estimated for the solar centre, $k_B T$ is only 10^3 eV (i.e., E_{th} is smaller than the Coulomb barrier by a factor of roughly 10^3). This is in fact so low that, with classical effects only, we can scarcely expect any reaction at all. In the high-energy

tail of the Maxwell-Boltzmann distribution, the exponential factor drops here to $\exp(-1000) \approx 10^{-434}$, which leaves no chance for the “mere” 10^{57} nucleons in the whole Sun (and even for the $\sim 10^{80}$ nucleons in the whole visible Universe)!

The only possibility for thermonuclear reactions in stars comes from a quantum-mechanical effect found by G. Gamow: there is a small but finite probability of penetrating (**quantum tunnelling**) through the Coulomb barrier, even for particles with $E < E_{\text{Coul}}(r_0)$. This tunnelling probability varies as

$$P = CE^{-1/2} \exp^{-2\pi\eta}, \quad \eta = \sqrt{\frac{m}{2}} \frac{Z_1 Z_2 e^2}{\hbar \sqrt{E}}.$$

Here \hbar is the **reduced Planck constant** $h/2\pi$ and m the **reduced mass**. The factor C is a constant that depends only on the properties of the two colliding nuclei. The exponent $2\pi\eta$ is obtained as the only E -dependent term in an approximate evaluation of the integral over $\hbar^{-1} \sqrt{2m(E_{\text{Coul}} - E)}$, which is extended from r_0 to the distance of closest approach (i.e., where $E = E_{\text{Coul}}$). For $Z_1 Z_2 = 1$ and $T = 10^7$ K, P is of the order of 10^{-20} for particles with average kinetic energy E and steeply increases with E and decreases with $Z_1 Z_2$. Therefore, for temperatures as “low” as 10^7 K, only the lightest nuclei (with smallest $Z_1 Z_2$) have a chance to react. For reactions of heavier particles, with larger $Z_1 Z_2$, the energy (i.e., the temperature) has to be correspondingly larger to provide a comparable quantum tunneling probability. This will result in well-separated phases of different nuclear burning during the star’s evolution.

Nuclear reaction rates: Consider collisions between two different kinds of particles with a number density in phase space of dN_1 and dN_2 . To obtain the number of collisions per second per unit volume, we must integrate over all available velocity space. That is, we must sum over the collisions between particles so that the collision rate r is

$$r = \int \int v \sigma(v) dN_1(\vec{v}_1) dN_2(\vec{v}_2) [\text{s}^{-1} \text{ m}^{-3}].$$

Let us assume that the velocity distributions of both kinds of particles are given by Maxwellian velocity distributions

$$dN = N \left(\frac{m}{2\pi k_B T} \right)^{3/2} e^{-mv^2/(2k_B T)} dv,$$

so that the collision rate equation becomes

$$r = N_1 N_2 \left(\frac{m_1}{2\pi k_B T} \right)^{3/2} \left(\frac{m_2}{2\pi k_B T} \right)^{3/2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{-m_1 v_1^2}{2k_B T} + \frac{-m_2 v_2^2}{2k_B T} \right) v \sigma(v) d\vec{v}_1 d\vec{v}_2 [\text{s}^{-1} \text{ m}^{-3}].$$

If we transform to the center-of-mass coordinate system, assuming the velocity field is isotropic, so that the triple integrals can be written as spherical “velocity volumes”, then we can rewrite this in terms of the center of mass velocity v_0 and the relative velocity v as

$$r = N_1 N_2 \left(\frac{m_1}{2\pi k_B T} \right)^{3/2} \left(\frac{m_2}{2\pi k_B T} \right)^{3/2} \int_0^{\infty} \int_0^{\infty} \exp \left[-\frac{(m_1 + m_2)v_0^2 + \tilde{m}v^2}{2k_B T} \right] v \sigma(v) 4\pi v_0^2 (4\pi v^2) dv_0 dv [\text{s}^{-1} \text{ m}^{-3}],$$

where

$$\begin{aligned} \tilde{m} &= \frac{m_1 m_2}{m_1 + m_2} [\text{kg}] \\ \vec{v} &= \vec{v}_1 - \vec{v}_2 [\text{m s}^{-1}] \\ \vec{v}_0 &= \frac{m_1 \vec{v}_1 + m_2 \vec{v}_2}{m_1 + m_2} [\text{m s}^{-1}]. \end{aligned}$$

The integral over v_0 is analytic and is

$$\int_0^{\infty} \exp \left[-\frac{(m_1 + m_2)v_0^2}{2k_B T} \right] 4\pi v_0^2 dv_0 = \left(\frac{2\pi k_B T}{m_1 + m_2} \right)^{3/2}$$

which reduces r to

$$r = N_1 N_2 \left(\frac{\tilde{m}}{2\pi k_B T} \right)^{3/2} \int_0^{\infty} \exp \left[-\frac{\tilde{m}v^2}{2k_B T} \right] v \sigma(v) (4\pi v^2) dv [\text{s}^{-1} \text{ m}^{-3}].$$

Since the *relative* kinetic energy in the center of mass system is $E = mv^2/2$, we can rewrite this in terms of an average reaction cross section $\langle\sigma(v)v\rangle$, so that

$$r = N_1 N_2 \langle\sigma(v)v\rangle$$

where

$$\langle\sigma(v)v\rangle = \frac{2}{\sqrt{\pi}} \left(\frac{1}{k_B T} \right)^{3/2} \int_0^\infty \exp\left(\frac{-E}{k_B T}\right) \sigma(v) v \sqrt{E} dE = \int_0^\infty n(E) v \sigma(v) dE.$$

Thus $\langle\sigma(v)v\rangle$ is the “relative energy” weighted average of the collision probability of particle 1 with particle 2. When this average cross section is written, the explicit dependence on velocity is usually omitted, so that

$$\langle\sigma v\rangle \equiv \langle\sigma(v)v\rangle.$$

If the collisions involve identical particles, then the number of distinct pairs of particles is $N(N - 1)/2$ so the factor of $N_1 N_2$ is replaced by $N^2/2$.

If we call the **energy produced per reaction** Q , we can write the energy produced per gram of stellar material as

$$\epsilon = \frac{r_{1,2} Q}{\rho} = \frac{N_1 N_2 \langle\sigma v\rangle}{\rho} [\text{eV g}^{-1}]$$

The number densities can be replaced with the more common fractional abundances by mass to get

$$\epsilon = \left[N_A^2 Q \langle\sigma v\rangle \left(\frac{X_1}{m_1} \right) \left(\frac{X_2}{m_2} \right) \right] \rho [\text{eV g}^{-1}]$$

where N_A is **Avogadro's number**. Since $\langle\sigma v\rangle$ is a complicated function of temperature and must be obtained numerically, this is usually approximated numerically as

$$\epsilon \approx \epsilon_0 \rho T^\nu [\text{eV g}^{-1}]$$

where

$$\epsilon_0 = N_A^2 Q \langle\sigma v\rangle \left(\frac{X_1}{m_1} \right) \left(\frac{X_2}{m_2} \right) \frac{1}{T_0^\nu} [\text{eV g}^{-2} \text{cm}^3 \text{K}^{-1}]$$

Here ν itself is very weakly dependent on the temperature. Most of the important energy production mechanisms have this form. The approximation $\epsilon \approx \epsilon_0 \rho T^\nu$ expresses the energy generated for a specific energy generation mechanism in terms of the state variables T and ρ . This is what we were after. Formulas such as these, where ϵ_0 has been determined, will enable us to determine the energy produced throughout the star in terms of the state variables. Let us now consider a few of the specific nuclear reactions for which we have expressions of the type given by this equation of ϵ_0 .

Specific nuclear reactions: The nuclear reactions that provide the energy for MS stars all revolve on the conversion of hydrogen to helium. However, this is accomplished by a variety of ways. We may divide these ways into two groups. The first is known as the **proton-proton cycle** (p-p cycle) and it begins with the conversion of two hydrogen atoms to deuterium. Several possibilities occur on the way to the production of ${}^4\text{He}$. These alternate options are known as **P2-P6 cycles**. In addition to the proton-proton cycle, a series of nuclear reactions involving carbon, nitrogen, and oxygen also can lead to the conversion of hydrogen to helium with no net change in the abundance of C, N, and O. For this reason, it is known as the **CNO cycle**. These reactions and their side chains are given below in Table 6: Besides the steps marked with asterisks which denote reactions that occur by spontaneous decay and do not depend on local values of the state variables, the steps that are the important contributors to the energy supply have their contribution (their Q value) indicated. The energy of the neutrinos has not been included since they play no role in determining the structure of normal stars. When the p-p cycle dominates on the lower MS, most of the energy is produced by means of the **P1 cycle**. *The neutrino produced in the fifth step of the P2 cycle is the high energy neutrino which has been detected, but in unexpectedly low numbers, by the neutrino detection experiments.* In general, the relative importance of the P1 cycle relative to P2 and P3 is determined by the helium abundance, since this governs the branching ratio at step 3 in the p-p cycle. If ${}^4\text{He}$ is absent, it will not be possible to make ${}^7\text{Be}$ by capture on ${}^3\text{He}$.

Virtually all the energy of the CNO cycle is produced by step 6 as the production of ${}^{12}\text{C}$ from ${}^{15}\text{N}$ is strongly favored. However, all the higher chains close with only the net production of ${}^4\text{He}$. The first stage of the P4 cycle is endothermic by 18 keV so unless the density is high enough to produce a Fermi

Step	Q [MeV]	Proton Cycles	Step	Q [MeV]	CNO Cycles
1	1.18	$^1\text{H}(p, \beta, \nu)^2\text{H}$	1	1.94	$^{12}\text{C}(p, \gamma)^{13}\text{N}$
2	5.49	$^2\text{H}(p, \gamma)^3\text{He}$	2*	1.51	$^{13}\text{N} \rightarrow ^{13}\text{C} + \beta^+ + \nu$
3	12.86	$^3\text{He}(^3\text{He}, 2p)^4\text{He}$	3	7.54	$^{13}\text{C}(p, \gamma)^{14}\text{N}$
or			4	7.29	$^{14}\text{N}(p, \gamma)^{15}\text{O}$
3	1.59	$^3\text{He}(\alpha, \gamma)^7\text{Be}$	5*	1.76	$^{15}\text{O} \rightarrow ^{15}\text{N} + \beta^+ + \nu$
4	0.06	$^7\text{Be}(\beta^-, \bar{\nu})^7\text{Li} \rightarrow ^7\text{Li}(p, \alpha)^4\text{He}$	6	4.96	$^{15}\text{N}(p, \alpha)^{12}\text{C}$
or		$Q = 17.35$			$^{15}\text{N}(p, \gamma)^{16}\text{O}$
4	0.13	$^7\text{Be}(p, \gamma)^8\text{B}$	7		$^{16}\text{O}(p, \gamma)^{17}\text{F}$
5*	10.78	$^8\text{B} \rightarrow ^8\text{Be} + \beta^+ + \nu$	8*		$^{17}\text{F} \rightarrow ^{17}\text{O} + \beta^+ + \nu$
6*	0.09	$^8\text{Be} \rightarrow ^2\text{He}$	9		$^{17}\text{O}(p, \alpha)^{14}\text{N}$
or					$^{17}\text{O}(p, \gamma)^{18}\text{F}$
3		$^3\text{He}(\beta^-, \nu^3\text{H})$	10*		$^{18}\text{F} \rightarrow ^{18}\text{O} + \beta^+ + \nu$
		$^3\text{H}(p, \gamma)^4\text{He}$	11		$^{18}\text{O}(p, \alpha)^{15}\text{N}$
4		$^3\text{H}(^3\text{He}, np)^4\text{He}$			$^{18}\text{O}(p, \gamma)^{19}\text{F}$
		$^3\text{H}(^3\text{H}, 2n)^4\text{He}$	12		$^{19}\text{F}(p, \alpha)^{16}\text{O}$
Triple- α process					
1		$2(^4\text{He}) + (\sim 100 \text{ KeV}) \rightarrow ^8\text{Be}^*$			
2		$^8\text{Be}^*(\alpha, ^{12}\text{C}^*)$			
3		$^{12}\text{C}^* \rightarrow ^{12}\text{C} + 2\gamma + 7.656 \text{ MeV}$			

Table 6: Nuclear reactions of the p-p chain and CNO cycle. Asterisk (*) denotes reactions that occur by spontaneous decay and do not depend on the local values of the state variables. Table taken from Collins (2003).

Proton-Proton			CNO Cycle			Triple- α Process		
T_6	ϵ_0	ν	ϵ_0	ν	T_8	ϵ_0	ν	
10	7×10^{-2}	4.60	3×10^{-4}	22.9	0.8	2×10^{-12}	49	
20	1	3.54	4.5×10^2	18.0	1.0	4×10^{-8}	41	
40	9	2.72	3×10^7	14.1	2.0	15.0	19	
80	43	2.08	2×10^{11}	11.1	3.0	6×10^3	12	
100	–	–	2×10^{12}	10.2	4.0	10^5	7.9	

Table 7: Temperature dependence of ν and ϵ_0 .

energy of 18 keV, the reaction does not take place. This requires a density of $\rho > 2 \times 10^4 \text{ g cm}^{-3}$ and so will not be important in MS stars. Once ^3H is produced it can be converted to ^4He by a variety of processes given in step 4. The last two are sometimes denoted **P5** and **P6**, respectively, and are rare.

While the so-called **triple- α process** is not operative in MS stars, it does provide a major source of energy during the red-giant phase of stellar evolution. The extreme temperature dependence of the triple- α process plays a crucial role in the formation of low-mass red giants. The $^8\text{Be}^*$ is unstable and decays in an extremely short time. However, if during its existence it collides with another ^4He nucleus, ^{12}C can form, which is stable. The very short lifetime for $^8\text{Be}^*$ basically accounts for the large temperature dependence since a very high collision frequency is required to make the process productive.

The exponent of the temperature dependence given in $\epsilon \approx \epsilon_0 \rho T^\nu$ and the constant ϵ_0 both vary slowly with temperature. This dependence is shown in Table 7.

The temperature T_6 in Table 7 is given in units of 10^6 . Thus $T_6 = 1$ is 10^6 K . It is a general property of these types of reaction rates that the temperature dependence T^ν “weakens” as the temperature increases, while the efficiency ϵ_0 increases. In general, the efficiency of the nuclear cycles rate is governed by the slowest process taking place. In the case of p-p cycles, this is always the production of deuterium given in step 1. For the CNO cycle, the limiting reaction rate depends on the temperature. At moderate temperatures, the production of ^{15}O (step 4) limits the rate at which the cycle can proceed. However, as the temperature increases, the reaction rates of all the capture processes increase, but the steps involving inverse β decay (particularly step 5), which do not depend on the state variables, do not and therefore limit the reaction rate. So there is an upper limit to the rate at which the CNO cycle can produce energy independent of the conditions which prevail in the star. However, at temperatures approaching a billion degrees, other reaction processes not indicated above will begin to dominate the energy generation and will circumvent even the beta-decay limitation.

1.19 Question 19

What is Fermi's Paradox? Explain its logic and assess the current state of the Paradox in light of modern knowledge.

1.19.1 Short answer

The Fermi paradox, or Fermi's paradox, named after physicist Enrico Fermi, is the apparent contradiction between the lack of evidence and high probability estimates for the existence of extraterrestrial civilizations. The basic points of the argument, made by physicists Enrico Fermi (1901-1954) and Michael H. Hart (born 1932), are:

1. There are billions of stars in the galaxy that are similar to the Sun, and many of these stars are billions of years older than the Solar system.
2. With high probability, some of these stars have Earth-like planets, and if the Earth is typical, some may have developed intelligent life.
3. Some of these civilizations may have developed interstellar travel, a step the Earth is investigating now.
4. Even at the slow pace of currently envisioned interstellar travel, the MW galaxy could be completely traversed in a few million years.

According to this line of reasoning, the Earth should have already been visited by extraterrestrial aliens. In an informal conversation, Fermi noted no convincing evidence of this, leading him to ask, "Where is everybody?" There have been many attempts to explain the Fermi paradox, primarily either suggesting that intelligent extraterrestrial life is extremely rare or proposing reasons that such civilizations have not contacted or visited Earth.

1.19.2 Additional context

Fermi's paradox presents arguably the greatest challenge for any practical SETI activity, as well as one of the least understood of all "grand questions" posed in the history of science. As is well known, the key argument follows a lunchtime remark of the great physicist, Enrico Fermi: "Where is everybody?" First discussed in print by the Russian space-science pioneer Konstantin Eduardovich Tsiolkovsky, and in recent decades elaborated upon in detail by Viewing, Hart, Tipler and others, the argument presents a formidable challenge for any theoretical framework assuming a naturalistic origin of life and intelligence. As such, this should worry not only a small group of SETI enthusiasts, but challenges some of the deepest philosophical and cultural foundations of modern civilization. It is hard to conceive a scientific problem more pregnant in meaning or richer in connections with the other big questions of science throughout the ages.

Tsiolkovsky, Fermi, Viewing, Hart, and their followers argue on the basis of two premises:

1. the absence of extraterrestrials in the Solar System; and
2. the fact that they have had more than enough time in the history of Galaxy to visit, either in person or through their conventional or self-replicating probes.

Characteristic time for colonization of the Galaxy, according to these investigators, is what we shall call the **Fermi-Hart timescale**:

$$t_{\text{FH}} = 10^6 - 10^8 \text{ [yr]},$$

making the fact that the Solar System is (obviously) not colonized hard to explain, if not for the total absence of extraterrestrial cultures. It is enough for our purposes to contend that this timescale is well-defined, albeit not precisely known due to our ignorance regarding the possibilities and modes of interstellar travel. For comparison, the accepted age of the Earth as an object of roughly present-day mass is

$$t_{\oplus} = (4.46 \pm 0.02) \times 10^9 \text{ [yr]}.$$

The drastic difference between these timescales is one of the ways of formulating Fermi's paradox. Even more generally, we need not consider the direct physical contact between an extraterrestrial civilization and Earth or the Solar System (insofar as we do not perceive evidence of extraterrestrial visits in the Solar System; however, this is still an act of faith, considering the volume of space comprising

our planetary system). It is sufficient to consider a weaker requirement: namely, that no extraterrestrial civilizations are detectable by any means from Earth at present. This includes the detectability of astro-engineering or macroengineering projects over interstellar distances. In the words of the great writer and philosopher Stanislaw Lem, who authored some of the deepest thoughts on this topic, Fermi's paradox is equivalent to the “absence of cosmic miracles” or the *Silentium Universi* (cosmic silence). Following the classic review by Brin (1983), we may introduce **contact cross-section** as a measure of the probability of contact (by analogy with introduction of cross-sections in atomic and particle physics) and reformulate Fermi's paradox as the question why this cross-section in the MW at present is so small in comparison to what could be naively expected.

Schematically, Fermi's paradox can be represented as

$$\begin{aligned} & \text{spatiotemporal scales of the Galaxy} + \text{the absence of detected extraterrestrial civilizations} \\ & \quad (+ \text{additional assumptions}) \rightarrow \text{paradoxical conclusion.} \end{aligned}$$

Here, under spatiotemporal scales we include our understanding of the age of the Galaxy, the Solar System and the ages (incompletely known) of other planetary systems in the MW. The additional assumptions can be further explicated as

$$\begin{aligned} \text{additional assumptions} = & \text{“naiverealism”} + \text{naturalism} + \text{Copernicanism} + \text{gradualism} + \\ & \text{non-exclusivity.} \end{aligned}$$

These assumptions are quite heterogeneous. By “naive realism” we denote the working philosophy of most of science (as well as everyday life), implying that there is a material world out there, composed of objects that occupy space and have properties such as size, mass, shape, texture, smell, taste and colour. These properties are usually perceived correctly and obey the laws of physics. In the specific case of Fermi's paradox, the basic premise following from naive realism is that there are, indeed, no traces of extraterrestrial intelligent presence detected either directly or indirectly. We shall discuss below some of the hypotheses for resolving Fermi's paradox which directly violate this realist view; an extreme and ludicrous example (but powerfully present in pop-culture) of such naively anti-realist standpoint is a view that, contrary to scientific consensus, some humans are in contact with extraterrestrial visitors and are conspiring with them. Naive realism and naturalism are methodological assumptions typically in play in any scientific research. Copernicanism and gradualism are somewhat more specific tenets, stemming more from our experiences in the history of physical science than from the general epistemology. **Copernicanism** (often called the **Principle of Mediocrity**) in a narrow sense tells us that there is nothing special about the Earth or the Solar System or our Galaxy within large sets of similar objects throughout the Universe. In a somewhat broader sense, it indicates that there is nothing particularly special about us as observers: our temporal or spatial location, or our location in other abstract spaces of physical, chemical, biological, etc., parameters are typical or close to typical. **Gradualism**, on the other hand, is often expressed as the motto that “the present is key to the past” (with corollary that “the past is key to the future”). This paradigm, emerging from geological science in the 19th century with the work of Charles Lyell (and expanding, through Lyell's most famous pupil, Darwin, into life sciences) has been subject of the fierce criticism in the last quarter of a century or so.

Finally, the role of the non-exclusivity (or “hardness” in some of the literature) assumption needs to be elucidated. Non-exclusivity is simply a principle of causal parsimony applied to the set of hypotheses for resolving Fermi's paradox: we should prefer those hypotheses which involve a smaller number of local causes. Fermi's paradox is eminently not resolved by postulating that a single old civilization self-destructs in a nuclear holocaust. Fermi's paradox is resolved by hypothesizing that all civilizations self-destruct soon after developing nuclear weapons, but the major weakness of such a solution is obvious: it requires many local causes acting independently in uniform to achieve the desired explanatory end. In other words, such a solution is exclusive (or “soft”). As long as we have any choice, we should prefer non-exclusive (or “hard”) solutions (i.e., those which rely on a small number of independent causes). For instance, the hypothesis that a γ -ray burst can cause mass extinction over a large portion of the Galaxy and thus arrest evolution toward advanced technological society, is quite non-exclusive.

Recent developments: Fermi's Paradox has become significantly more serious, even disturbing, of late. This is due to several independent lines of scientific and technological advance occurring during the last two decades:

1. The discovery of more than 350 extrasolar planets so far, on an almost weekly basis⁷. Although most of them are hot Jupiters and not suitable for life as we know it (some of their satellites could still be habitable, however) many other exoworlds are reported to be parts of systems with stable

⁷For regular updates see <http://exoplanet.eu/>

circumstellar habitable zones. It seems that only the selection effects and the capacities of present-day instruments stand between us and the discovery of Earth-like extrasolar planets, envisioned by the new generation of orbital observatories. In addition, this relative wealth of planets decisively disproves old cosmogonic hypotheses regarding the formation of the Solar System as a rare and essentially non-repeatable occurrence, which have been occasionally used to support skepticism on issues of extraterrestrial life and intelligence.

2. Improved understanding of the details of the chemical and dynamical structure of the MW and its Galactic habitable zone. In particular, it's been shown that Earth-like planets began forming more than 9 Gyr ago, and that their median age is $\langle t \rangle = (6.4 \pm 0.7) \times 10^9$ yr – significantly more than the age of the Earth. This means that the age difference $\langle t \rangle - t_{\oplus} = (1.9 \pm 0.7) \times 10^9$ yr is large in comparison with the Fermi-Hart timescale. This also means that not only the oldest ones, but a large majority of habitable planets are much older than Earth. The significance of this result cannot be overstated, since it clearly shows that the naive naturalist, gradualist and Copernican view must be wrong, since it implies that millions of planets in the MW are inhabited by Gyr-old supercivilizations, in clear contrast with observations.
3. Confirmation of the rapid origination of life on early Earth; this rapidity, in turn, offers strong probabilistic support to the idea of many planets in the MW inhabited by at least the simplest lifeforms.
4. Discovery of **extremophiles** and the general resistance of simple lifeforms to much more severe environmental stresses than had been thought possible earlier. These include representatives of all three great domains of terrestrial life (Bacteria, Archaea, and Eukarya), showing that the number and variety of cosmic habitats for life are probably much larger than conventionally imagined.
5. Our improved understanding of molecular biology and biochemistry leading to heightened confidence in the theories of the naturalistic origin of life or **biogenesis**. The same can be said, to a lesser degree, for our understanding of the origin of intelligence and technological civilization.
6. Exponential growth of the technological civilization on Earth, especially manifested through **Moore's Law** and other advances in information technologies. This is closely related to the issue of astro-engineering: the energy limitations will soon cease to constrain human activities, just as memory limitations constrain our computations less than they once did. We have no reason to expect the development of technological civilization elsewhere to avoid this basic trend.
7. Improved understanding of the feasibility of interstellar travel in both the classical sense and in the more efficient form of sending inscribed matter packages over interstellar distances. The latter result is particularly important since it shows that, contrary to the conventional skeptical wisdom, it makes good sense to send (presumably extremely miniaturized) interstellar probes even if only for the sake of communication.
8. Theoretical grounding for various astroengineering/macroengineering projects potentially detectable over interstellar distances. Especially important in this respect is the possible combination of astroengineering and computation projects of advanced civilizations.
9. Our improved understanding of the extragalactic Universe has brought a wealth of information about other galaxies, many of them similar to the MW, while not a single civilization has been found, in spite of the huge volume of space surveyed.

Although admittedly uneven and partially conjectural, this list of advances and developments (entirely unknown at the time of Tsiolkovsky's and Fermi's original remarks and even Viewing's, Hart's and Tipler's later reissues) testifies that Fermi's paradox is not only still with us more than 75 years after Tsiolkovsky and more half a century after Fermi, but that it is more puzzling and disturbing than ever.

1.19.3 Follow-up Questions

- What percentage of stars have planets?

1.20 Question 20

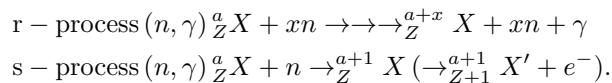
The so-called r- and s- processes are mechanisms that produce elements heavier than iron. Describe these mechanisms and evidence for them from abundance patterns. Where is the r-process thought to act?

1.20.1 Short answer

These are processes by which elements heavier than iron are formed by neutron capture (r- and s-processes). The **r-processes** (r for rapid) take place in core-collapse supernovae. The **s-process** (s for slow) take place in AGB stars, that is, in low- or intermediate-mass stars ($0.8 M_{\odot} < M < 8 M_{\odot}$) in their second red giant phase or close to the final state of this phase.

1.20.2 Additional context

Similar to r- and s-processes, rp-processes form elements heavier than iron by proton capture (rather than neutron capture). “Rapid” and “slow” refer to the timescale of neutron capture as related to that of the β^- decay of the resulting nucleus. The identification of the two types of neutron capture processes goes back to pioneering work carried out by Fred Hoyle and colleagues half a century ago who also proposed the rp-process to explain the presence, and the abundance pattern, of proton-rich heavy nuclei. Related to the r-process is the νp -process, by which a proton combines with an anti-neutrino to form a neutron and a positron, with concomitant incorporation of the neutron into a nucleus. A complementary process is the γ -process, sometimes confusingly also referred to as p-process (a photodisintegration process) by which neutrons, α -particles, or protons are knocked out of a nucleus by γ rays. The following equations provide an overview. Here, X and X' represent elements (**metals**) beyond iron, Z is the atomic number (number of protons, i.e., nuclear charge), and A is the mass number (sum of protons and neutrons).



The r-process (which affords a particularly high neutron density) and the s-process (which goes on at lower neutron densities and intermediate temperatures) contribute about equally to the nucleosynthesis of heavy nuclei, but there are specific nuclei that are (almost) exclusively produced by either the r- or the s-process. Typical r-process elements are Ge, Zr, Te, Xe, Eu, Os, Pt, and Au, and typical s-process elements are Sr and Ba. Nuclei such as ${}^{232}\text{Th}$, ${}^{235}\text{U}$, and ${}^{238}\text{U}$ can only be produced in the r-process. Further, since the s-process occurs later in galactic history than the r-process, elements produced by the s-process are delivered to the ISM at a later stage. So-called **halo stars**, which circle our galaxy in randomly oriented eccentric orbits, and which are among the very oldest stars (evidenced by their very low iron abundance), have a clear signature of elements made in the r-process. Figure 38 is a graphical representation of typical paths for an r- and s-process.

Rapid processes: The r-process, associated with very large neutron densities, requires a few seconds only. The timescale for neutron capture is $\sim 10^{-4}$ s. Up to 10 and more neutrons can be picked up before β^- decay occurs when the neutron flux runs out. For neutron numbers per nucleus exceeding 184, neutron-induced nuclear fission processes also come in, contributing to the distribution of the lighter r-process nuclei. Neutron capture at high temperature produces high-energy γ -rays; this leads to (γ, n) photodisintegration competing with n-capture.

The explosive collapse of a pre-SN leaves behind a neutron core with an initial temperature of several 10^{11} K that produces an extensive flux of neutrinos and anti-neutrinos with energies typically between 10 and 20 MeV. In proton-rich, cooler ejecta of the SNe, protons can be converted into neutrons by the absorption of anti-neutrinos. In principle, this is possible for free protons as well as for protons confined to nuclei. Since protons are sufficiently more abundant in the stellar envelopes than heavy nuclei, anti-neutrino capture occurs predominantly on free protons, leading to a drastic increase of neutron density for a few seconds, and subsequent neutron capture, in particular by nuclei that are comparatively proton-rich; that is, where the number of neutrons is about that of the number of protons. The net result of this process, termed νp -process, compares to that of the r- process. Along with the absorption of neutrons by neutron-deficient nuclei, neutron capture can also be coupled to the disposal of a proton. This (n, p) reaction delivers nuclei with a higher effective cross section for proton capture, producing nuclei such as ${}^{92}\text{Mo}$, ${}^{94}\text{Mo}$, ${}^{96}\text{Ru}$, ${}^{98}\text{Ru}$, and ${}^{102}\text{Pd}$, which are otherwise not easily accessible.

Slow processes: In contrast to the rapid processes and typical of SNe scenarios, a slow neutron capture process is a characteristic event for the formation of trans-iron elements in the AGB phase of a star. In this phase (see track 4 in the HR diagram of Figure 1), an extremely dense core primarily composed of carbon and oxygen is surrounded by a helium burning shell as the main source of energy and, slightly farther out, by a hydrogen burning shell followed by a tenuous hydrogen-rich envelope. When most of

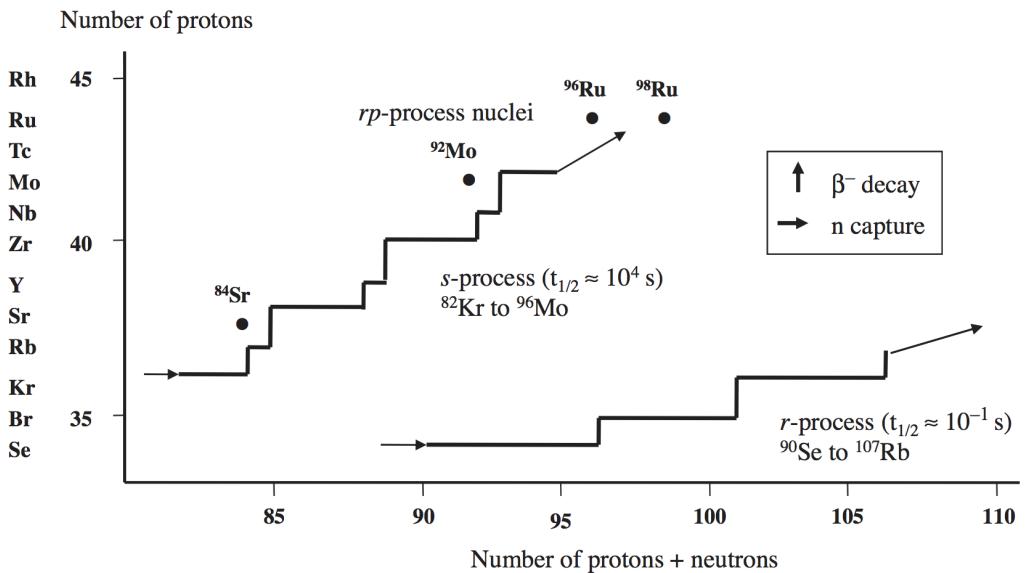


Figure 38: Sections of the tracks for the formation of heavy (trans-iron) elements along the r-process (rapid neutron capture, lower trace) and the s-process (slow neutron capture, upper trace). In both tracks, neutron capture is followed by β^- decay (vertical lines); $t_{1/2}$ = half life. The dots represent nuclei that are formed by rapid proton capture (rp-process). Figure taken from Rehder (2010).

the helium has been consumed, the giant passes through a brief period of flickering (thermal pulses by alternately switching on and off H and He fusion), whereby material from the inner regions is mixed into the outer layers (**dredge-up**), and vice versa, accompanied by mass loss through violent stellar winds. The mass loss gives rise to a circumstellar envelope extending over several light-years.

In the s-process for which the neutron capture time is slow relative to the β^- decay rates of the daughter nucleus, lower neutron densities and temperatures ($\sim 3 \times 10^8$ K) than in the r-process are required.

1.21 Resources

- Stellar Populations; Greggio & Renzini (2011)
- The Fundamentals of Stellar Astrophysics, Collins (2003)
- Stellar Structure and Evolution, Kippenhahn, Weigert & Weiss (2012)
- Evolution of Stars and Stellar Populations, Salaris & Cassisi (2005)
- The Astronomical Reach of Fundamental Physics, Burrows & Ostriker (2014)
- Opacity, Huebner & Barfield (2014)
- Radiative Processes in Astrophysics, Rybicky & Lightman (1979)
- Understanding Variable Stars, Percy (2007)
- A Library of Stellar Spectra; Jacoby & Hunter (1984)
- A Digital Spectral Classification Atlas; Gray (2009)
- Notes on Star Formation; Krumholz (2015)
- Physics of the Interstellar and Intergalactic Medium; Draine (2011)
- Spectral Energy Distributions of T Tauri Stars with Passive Circumstellar Disks; Chiang & Goldreich (1997)
- Fermi's Paradox – The Last Challenge for Copernicanism?; Ćirković (2009)
- Chemistry in Space: From Interstellar Matter to the Origin of Life; Rehder (2010)
- On the Anomalous Radii of the Transiting Extrasolar Planets; Laughlin et al. (2011)
- Possible Solutions to the Radius Anomalies of Transiting Giant Planets; Burrows et al. (2007)
- Re-Inflated Warm Jupiters Around Red Giants