



## ASTROCHALLENGE 2019 SENIOR TEAM ROUND

### SOLUTIONS

Monday, 3<sup>rd</sup> June 2019

#### PLEASE READ THESE INSTRUCTIONS CAREFULLY.

1. This paper consists of 37 printed pages, including this cover page.
2. Do NOT turn over this page until instructed to do so.
3. You have 2 hours to attempt all questions in this paper.
4. At the end of the paper, submit this booklet together with your answer script.
5. Your answer script should clearly indicate your name, school, and team.
6. It is your responsibility to ensure that your answer script has been submitted.
7. The marks for each question are given in brackets in the right margin, like such: [2].
8. The alphabetical parts (i) and (l) have been intentionally skipped, to avoid confusion with the Roman numbering of (i).

# 1 Cosmological Inflation

## Part I: The Universe

(a) In astronomy, what does the term ‘cosmological inflation’ refer to?

[1]

**Solution:**

Cosmological inflation refers to the **exponential expansion** of the **extremely early** Universe, given timescales of  $10^{-36}$  to  $10^{-32}$  seconds after the Big Bang, where the inflation was by a factor of  $10^{22}$ .

(b) Cosmological inflation was hypothesised in order to explain certain observations that could not otherwise be easily explained. Name and describe an example of such a problem.

[2]

**Solution:**

Three problems that are accepted:

- **Horizon problem:** The observed **homogeneity** of distant parts of the Universe implies that there *has* been **information** (i.e. thermal) **transfer** between said parts; however, their distances from us and from each other ( 46 billion light years, and 93 billion light years respectively) suggest that it is impossible for them to have had **causal contact** for such information transfer to occur.
- **Flatness problem:** The current mass-energy density of the Universe is very close to the critical value of  $\Omega = 1$ —in other words, the Universe is likely to be **flat**: it will probably undergo neither the Big Rip nor the Big Crunch, but experience a slow, gently decelerating expansion that would lead to the heat death of the Universe. The mass-energy density should have diverged far from the initial value at an increasing rate as time goes by and the Universe expands; however, as noted earlier, this is not the case.
- **Magnetic monopole problem:** If the Universe was initially extremely hot and dense, a large number of magnetic monopoles (i.e. distributions of *magnetic charge*, with field lines either radiating outwards or falling inwards, akin to electric charge) would have formed, due to the fact that three of the four fundamental forces, electromagnetism, the weak interaction and the strong force, should have been unified. However, there exist **no** such monopoles today.

It is prudent to note that all these problems are *cosmological fine-tuning* problems.

(c) Explain how cosmological inflation resolves the problem you have stated in (b).

[2]

**Solution:**

Three problems that are accepted:

- **Horizon problem:** Given the extremely rapid expansion of the early Universe, the **initially** homogenous early Universe was inflated to modern scales within a matter of  $10^{-32}$  seconds or so, hence ‘locking in’ the homogeneity, despite their current immense distances from us, and

from each other. This also explains the isotropy of the CMBR.

- **Flatness problem:** The argument is similar to the above: the early Universe had a mass-energy density nearly equal to 1, because the rapid, exponential expansion of the early Universe led to a 'locking in' of the mass-energy density, without allowing it to diverge far from the initial value.
- **Magnetic monopole problem:** There exist no magnetic monopoles today, because the rapid expansion of the Universe meant that even though the fundamental forces were unified, they diverged quickly due to the rapidly decreasing temperature and pressure, and there was insufficient time for *many* monopoles to have formed, if at all, given there are *none* today.

(d) What is the Cosmic Microwave Background?

[1]

**Solution:**

The CMBR is remnant electromagnetic radiation left by the Big Bang, Big Bang nucleosynthesis, and Big Bang matter-antimatter annihilation, that has since been redshifted to microwave radiation equal to that emitted by a blackbody of temperature 2.7 K.

(e) What process initially generated the Cosmic Microwave Background? Name and describe this process.

[2]

**Solution:**

Until 380 000 years after the Big Bang, the Universe was opaque to EM radiation, due to the ionised nature of matter (in turn due to the high ambient temperature of the Universe—about 3000 K). At that point, however, **recombination** occurred, where electrons combined with positive nuclei to form neutral atoms, and thus, the Universe became transparent to EM radiation. The radiation seen from this event is what has now been redshifted to microwave radiation, today.

## Part II: Galaxies

Refer to Figure 1 to answer parts (f) and (g).

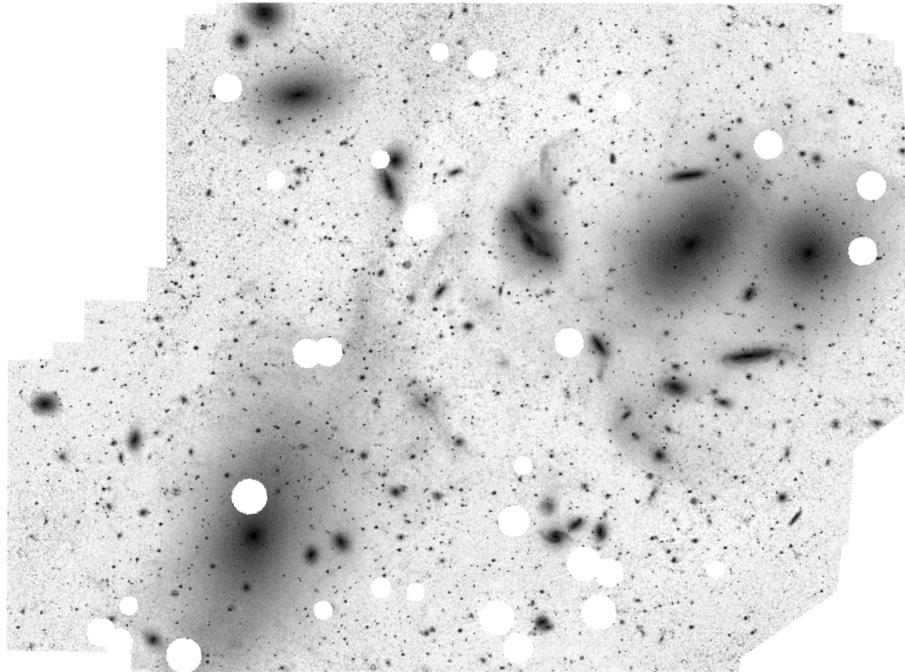


Figure 1: Negative image of the core of the Virgo Cluster; white circles are foreground stars that have been blotted out

- (f) Galaxies tend to group in clusters, rather than exist in complete isolation. Account for this trend. [2]

**Solution:**

Minute **quantum fluctuations** in the distribution of matter in the early Universe eventually led to **clustering of dark matter** around regions of **increased density**. **Over time, ordinary matter coalesced** around these regions of dark matter, leading to clusters of galaxies, that explains the large-scale structure of the Universe, with clusters, superclusters, filaments and voids, rather than a random, uniform distribution of galaxies.

- (g) It is clear from Figure 1 that spiral galaxies are rare within the core of the Virgo Cluster. Furthermore, the few that *do* exist, are highly distorted. Explain this observation. [2]

**Solution:**

The primary reason is **gravity**. The core of the Virgo cluster is a **dense, crowded place** with intergalactic collisions **common**. Most spiral galaxies in the core would have collided and merged with each other, and evolved to form yellow elliptical or lenticular galaxies. Any spiral galaxies *still* left are in the merger process, and hence are highly distorted due to **tidal effects** from neighbouring massive galaxies.

### Part 3: Stars and Star Clusters

- (h) (i) What is parallax? [1]

**Solution:**

Parallax is the change in **apparent position** of an object when viewed from different locations.

In the case of the Earth and distant stars, it is the change in apparent position of a nearer star compared to more distant, apparently fixed background stars, as the Earth moves in its orbit around the Sun.

- (ii) How might one use parallax to determine the distance to a nearby star? [1]

**Solution:**

Answer from the Formula Book: The distance to a star,  $d$  in **parsecs** with a parallax of  $p$  **arc-seconds** (where  $1 \text{ pc} \approx 3.26 \text{ ly}$ ) is given by:

$$d \approx \frac{1}{p}$$

Other accepted answers include a diagrammatic explanation of how parallax works.

- (j) Suppose one wishes to find out the distance to a certain open star cluster within the Milky Way. The star cluster is too far away for accurate parallax measurements.

Name **two** alternative methods to determine the distance to this star cluster, and briefly explain how each method works.

Your answer should explicitly specify what data needs to be collected for each method to work. [6]

**Solution:**

**1. Variable stars, e.g. Cepheids and RR Lyraes:**

**Data needed:** list of stars with variable brightness, and their light curves.

From the above, the periods and absolute magnitude of the variability can be inferred/calculated, and thence be fitted to known variable stars, thus deriving the distance to this star cluster.

**2. Main-sequence fitting:**

**Data needed:** a colour-magnitude diagram (CMD) of the cluster in question, and a (CMD) of a cluster of known distance, so that the two can be compared.

Solutions *not* or partially accepted are **Type Ia supernovae** and **Hubble constant/redshift**.

Other *accepted* answers include standard candles like Vega, etc.

## 2 The Last of the Main Sequence

How do low and intermediate-mass stars leave the main sequence and evolve into red giants? It turns out that this is a fairly complex process! Here, we will try to shed some light into the mechanics of the red giant phase.

### Preliminaries

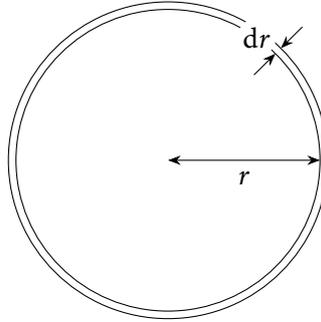
Stars are complex beasts. Several different processes occur within them, and to understand how stars evolve, we need to account for all of them. Table 2 is a (non-exhaustive) list of a few processes within...

Process	Equation	Description
Hydrostatic equilibrium	$\frac{dP}{dM_r} = -\frac{GM_r}{4\pi r^4}$	$M_r$ represents the mass contained within radius $r$ of the star. Note that $r$ is a variable: in other words, this equation must hold true for <i>all layers</i> within a stable star, and not just the outer radius. To avoid confusion, we will use $R$ to denote the outer radius of the star, and <i>not</i> $r$ .
Energy generation equation	$\frac{dL}{dM_r} = \epsilon_{\text{nuc}}$	$\epsilon_{\text{nuc}}$ is the rate of energy release per unit mass of nuclear fuel.
Rate of energy release for the p-p chain	$\epsilon_{\text{pp}} \propto \rho X^2 T^4$	$\rho$ is the density of the gas, while $X$ is the fraction of hydrogen, and $0 \leq X \leq 1$ . You may assume all hydrogen fusion occurs via the p-p chain i.e. $\epsilon_{\text{nuc}} = \epsilon_{\text{pp}}$ .
Energy transport equation	$\frac{dT}{dM_r} = -\left(\frac{1}{16\pi^2 r^4 \rho \lambda}\right) L_r$	$L_r$ is the <i>net</i> energy flowing out of the star's interior from radius $< r$ , while $\lambda$ is a constant that describes the efficiency of conduction and radiation.
Virial Theorem for gravitationally bound systems	$\text{KE} = -\frac{1}{2} \text{GPE}$	If we measure the kinetic and gravitational potential energy of all the particles in a star, we would find that the magnitude of the total kinetic energy would be half that of the magnitude of the total gravitational potential energy. Note that GPE is negative, thus KE is always positive.
Kinetic energy of an average ideal gas particle	$\langle E_k \rangle = \frac{3}{2} kT$	

Figure 2: Equations and descriptions for stellar processes

- (a) The Formula Book has a different variation of the formula for hydrostatic equilibrium. Show that they are equivalent. [1]

**HINT:** Beginning at radius  $r$  of a star and moving outwards by  $dr$ , one crosses a thin shell of volume  $4\pi r^2 dr$ . Multiplying that by the density at that radius yields the resultant change in mass  $dM_r$ . See Figure 3.



**Figure 3:** A cross section of a sphere of radius  $r$  with a shell of thickness  $dr$ , and mass  $dM_r$

**Solution:**

From the hint:

$$dM_r = 4\pi r^2 \rho_r dr$$

Substituting in to the equation,

$$\frac{dP}{4\pi r^2 \rho_r dr} = \frac{GM_r}{4\pi r^2}$$

$$\frac{dP}{dr} = -\rho_r \frac{4\pi r^2 GM_r}{4\pi r^2} = -\rho_r \frac{GM_r}{r^2}$$

From the previous, one can clearly solve for relevant stellar parameters (e.g. luminosity) by expressing them in terms of the enclosed mass  $M_r$ , or in terms of the radius of the star. In practice, however, it is more convenient to work with the enclosed mass, and this convention will be followed in the rest of the question.

[Food for thought: Why is it more convenient?]

**Solution:**

The mass of the star is approximately constant throughout its lifetime, while its radius can change dramatically as it moves from phase to phase. Thus, working with the enclosed mass allows us to avoid solving for the stellar radius *explicitly* when its not needed for our calculations.

## Intermediate-Mass Stars and the Subgiant Stage

- (b) We define low-mass stars as stars that eventually undergo a helium flash. Intermediate-mass stars do **not**, but neither do they undergo supernovae.

What is a helium flash, and how does it occur?

[3]

### Solution:

Due to their high density, the helium cores of low-mass stars become degenerate as they leave the main sequence: the pressure of the core is no longer dependent on temperature [1]. Eventually, conditions in the core become suitable for helium fusion, increasing the core temperature. Yet, due to the degenerate core, the core does not expand in response to the increased energy output [1].

Because of this, the core temperature continues to climb, triggering runaway helium fusion throughout the core that eventually breaks the degeneracy. This is the helium flash [1].

Now, both low mass and intermediate mass stars evolve into red giants via slightly different ways, but the broad mechanics are similar. Let us consider an intermediate mass star that has run out of hydrogen. The core is now completely comprised of inert helium and is surrounded by a shell of fusing hydrogen.

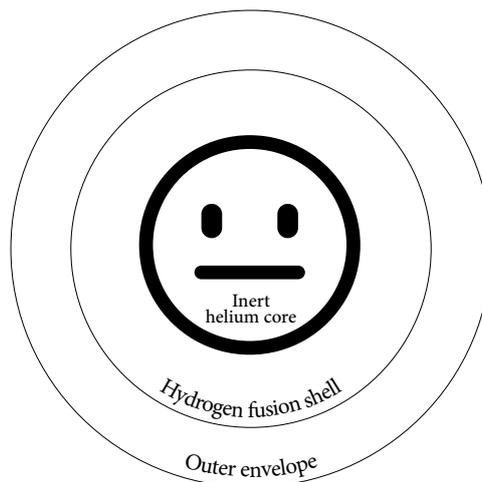


Figure 4: The internal structure of a subgiant star

For an intermediate mass star, this configuration is initially stable: the inert helium core can, at first, support the outer layers without contracting.

- (c) The inert core quickly becomes isothermal with its surroundings, i.e. the same temperature persists *throughout* the core. Explain this phenomenon. [2]

**HINT:** Treat the core as being composed of many unit masses. You may then consider the equations in the Table, or utilise the First Law of Thermodynamics: the change in internal energy of each unit mass is equal to the amount of heat supplied to the unit mass, minus the amount of work done by the unit mass.

**Solution:**

**Table:**

When core nuclear fusion halts, no more energy flows outwards from the core [1]. Formally,  $\frac{dL}{dM_r} = 0$  within the core, and obviously  $L_r = 0$  at the centre, thus  $L_r = 0$  throughout the core.

Applying our energy transport equation, when  $L_r = 0$ ,  $\frac{dT}{dM_r} = 0$  i.e. the temperature does not change within the core as we move outwards. This implies that the core is isothermal [1].

**First Law:**

Since the core is stable, no work is done by the unit mass throughout this process [1]. In other words, the unit mass is not sinking/rising/expanding/contracting, and thus no work is being done against gravity.

Since no fusion is taking place within the core, the only heat flow is that between the unit mass and its surroundings [1]. When each unit mass reaches thermal equilibrium with the base of the hydrogen fusion shell, all further heat exchange halts and the internal energy (i.e. temperature) remains constant. Repeating this argument for each unit mass, we can see that eventually the entire core shares the same temperature as the base of the hydrogen fusion shell and is thus isothermal.

### Things Fall Apart...

The effective surface pressure of an isothermal ideal-gas stellar core,  $P_{\text{core}}$ , is given by the equation:

$$P_{\text{core}} = \frac{3RT_{\text{core}}M_{\text{core}}}{4\pi\mu_{\text{core}}r_{\text{core}}^3} - \frac{\alpha GM_{\text{core}}^2}{4\pi r_{\text{core}}^4}$$

where:

- $R$  is the ideal gas constant;
- $T_{\text{core}}$  is the temperature of the *isothermal* core;
- $M_{\text{core}}$  is the total mass of the core;
- $r_{\text{core}}$  is the radius of the core;
- $\mu_{\text{core}}$  is the *average* molar mass of each particle in the core;
- $\alpha \approx 1$ , a constant that depends on the star's internal structure;
- $G$  is the gravitational constant.

Notice that there are only three *variables* in this equation:  $M_{\text{core}}$ ,  $T_{\text{core}}$  and  $r_{\text{core}}$ . It follows that increasing the core mass eventually reduces the effective surface pressure. Thus, as more helium ash accumulates

[Turn over]

at the core, the effective surface pressure of the core gradually falls. We will investigate if the other two variables change to counteract this effect.

- (d) While  $T_{\text{core}}$  is *technically* a variable, its value does not change much, in reality. This is because the core temperature has equalised with the hydrogen fusion shell, as seen in (c). Explain why the temperature of the hydrogen fusion shell cannot vary dramatically. [1]

**Solution:**

Hydrogen fusion is highly sensitive to temperature (from the table, it varies with  $T^4$ ): small changes in temperature lead to large changes in the fusion rate, which in turn leads to temporary loss of hydrostatic equilibrium for the hydrogen fusion shell. In response the hydrogen fusion shell contracts/expands such as to oppose the change in temperature.

- (e) Let us consider  $r_{\text{core}}$  **only**. Give a *qualitative, non-mathematical* explanation as to why, for a given  $M_{\text{core}}$  and  $T_{\text{core}}$ ,  $P_{\text{core}}$  cannot increase indefinitely. [2]

**HINT:** Treat this massive stellar core like a stress ball (you might need one, anyway, at this point). The core has a fixed mass and temperature. In response to the pressure you exert on it, the core adjusts to a corresponding radius. What happens as you squeeze it harder and harder? Can your stress ball support unlimited amounts of pressure?

**Solution:**

As the core contracts, the self-gravity of the core becomes non-negligible [1]. This counteracts the increased gas pressure from contraction, until at some point it becomes dominant: further contraction causes the pressure to decrease [1].

- (f) Prove, using calculus, that the premise in (e) is indeed true: there exists a maximum surface pressure, provided that *only*  $r_{\text{core}}$  can vary. [4]

**Solution:**

Idea: take the derivative with respect to  $r_{\text{core}}$ , and set  $P'_{\text{core}} = 0$ . Then, show that when  $P' = 0$ ,  $P''_{\text{core}} < 0$ .

Note that the expression is really rather messy; so let's clean it up, by combining all the terms except  $r_{\text{core}}$ :

$$A = \frac{3R}{4\pi} \frac{T_{\text{core}} M_{\text{core}}}{\mu_{\text{core}}}$$

$$B = \frac{\alpha G M_{\text{core}}^2}{4\pi}$$

Observe that  $A, B > 0$ . Therefore, the pressure equation reduces to:

$$P_{\text{core}} = \frac{A}{r_{\text{core}}^3} - \frac{B}{r_{\text{core}}^4}$$

$$\frac{dP_{\text{core}}}{dr_{\text{core}}} = 0 \implies \frac{-3A}{r_{\text{core}}^4} + \frac{4B}{r_{\text{core}}^5} = 0$$

$$4B = 3Ar_{\text{core}}$$

$$r_{\text{core}} = \frac{4B}{3A} = \frac{4}{9} \frac{\alpha GM_{\text{core}} \mu_{\text{core}}}{RT_{\text{core}}}$$

$A, B > 0 \implies r_{\text{core}} > 0$ , and our answer makes sense, and hence a critical point exists. To prove that it is a maximum,

$$\frac{d^2P_{\text{core}}}{dr_{\text{core}}^2} = \frac{12A}{r_{\text{core}}^5} - \frac{20B}{r_{\text{core}}^6}$$

$$= \frac{4}{r_{\text{core}}^5} \left( 3A - \frac{5B}{r_{\text{core}}} \right)$$

Given  $r_{\text{core}} > 0 \implies \frac{4}{r_{\text{core}}} > 0$ . Now, to prove a maximum, we need to prove  $3A - \frac{5B}{r_{\text{core}}} < 0$  when  $r_{\text{core}} = \frac{4B}{3A}$ . Hence:

$$\frac{d^2P_{\text{core}}}{dr_{\text{core}}^2} = 4 \left( \frac{3A}{4B} \right)^5 \left( 3A - 5B \frac{3A}{4B} \right)$$

$$= 4 \left( \frac{3A}{4B} \right)^5 \left( 3A - \frac{5}{4} (3A) \right)$$

$$= 4 \left( \frac{3A}{4B} \right)^5 \cdot -\frac{1}{4} (3A)$$

$$= -3A \left( \frac{3A}{4B} \right)^5 < 0$$

[Turn over]

- (g) We thus informally show that the effective surface pressure can only decrease as the core mass increases. Therefore, as the mass of the core increases, eventually it cannot support the overlying stellar envelope and *also* remain isothermal. With appropriate values, it follows that an isothermal helium core has no more than approximately 10% of the total mass of the star.

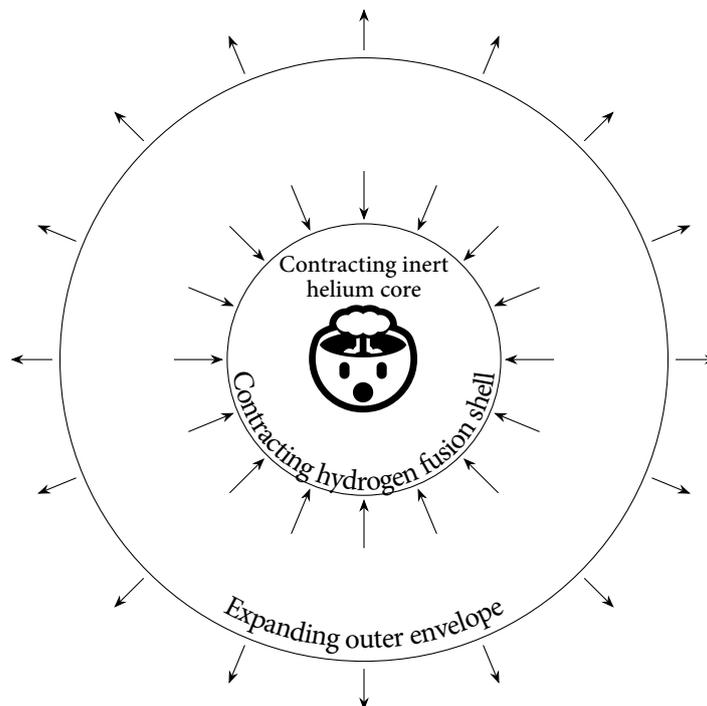
However, things are different for the Sun—simulations suggest that the Sun will develop an isothermal helium core of nearly *half* the mass of the Sun. Briefly explain how the Sun can support such a situation. [1]

**Solution:**

The helium core in the Sun is so dense that degeneracy pressure is significant; therefore, it no longer behaves like an ideal gas and our calculations do not apply.

### The Centre Cannot Hold

What happens as we cross this point? The core starts to contract, converting its gravitational potential energy into thermal energy by the Virial Theorem, and the core heats up. The core thus loses heat energy to the surroundings, which partially arrests core collapse: the core shrinks in a slow and orderly manner. Yet this contraction creates havoc for the hydrogen fusion shell. We illustrate this with the results of a simulation for a 5 solar mass star.



**Figure 5:** Core contraction of an intermediate-mass star

- (h) When core fusion halts, a unit mass at the outer edge of the hydrogen fusion shell lies at 0.4 solar radii, with 0.6 solar masses underlying it. Within a short span of time (approximately 300 000 years) the core and hydrogen fusion shell contracts to their maximum extents such that the same unit mass is now at 0.1 solar radii.

Calculate the change in GPE experienced by the core and hydrogen fusion shell during this process. Assume uniform density. [1]

NOTE: By the Shell Theorem, you should ignore the mass of the outer envelope when performing your GPE calculations, because the *net* gravitational force exerted by the outer envelope upon the core/hydrogen fusion shell is zero.

**Solution:**

$$\begin{aligned}
 \Delta U &= U_{\text{final}} - U_{\text{initial}} \\
 &= -\frac{3}{5} \frac{GM^2}{R_{\text{final}}} + \frac{3}{5} \frac{GM^2}{R_{\text{init}}} \\
 &= \frac{3GM^2}{5} \left[ \frac{1}{R_{\text{init}}} - \frac{1}{R_{\text{final}}} \right] \\
 &= \frac{3G (0.6 \times 1.989 \times 10^{30} \text{ kg})^2}{5} \left[ \frac{1}{0.4 \times 6.963 \times 10^8 \text{ m}} - \frac{1}{6.963 \times 10^7 \text{ m}} \right] \\
 &= -6.1 \times 10^{41} \text{ J}
 \end{aligned}$$

- (j) By the Virial Theorem, this implies that the total energy of the core and hydrogen fusion shell has decreased: energy is being dumped into the outer layers.

Hence, calculate the total amount of energy transferred to the outer layers, **and** the implied average rate of energy transfer per second. Express your answer in terms of joules and solar luminosities ( $L_{\odot}$ ) respectively. [2]

**Solution:**

Since  $E_{\text{total}} = E_{\text{kinetic}} + E_{\text{potential}}$ , and the Virial Theorem tells us that  $E_{\text{kinetic}} = -\frac{1}{2}E_{\text{potential}}$ , this tells us that if GPE changes by  $x$ ,  $E_{\text{kinetic}}$  changes by  $-0.5x$ , and thus  $E_{\text{total}}$  changes by  $0.5x$ .

This implies that the  $E_{\text{total}}$  of the core falls by  $3.05 \times 10^{41}$  J, and this is transferred to the outer layers over 300 000 years [1]. Compute the 'luminosity' for the other mark:

$$L_{\text{core}} = \frac{3.05 \times 10^{41} \text{ J}}{300\,000 \text{ years}} = 3.2 \times 10^{28} \text{ W} = 83.2L_{\odot}$$

- (k) Furthermore, the contraction compresses the hydrogen fusion shell, increasing its temperature and density. This stimulates much higher rates of nuclear fusion.

To prove this, calculate the pressure gradient,  $\frac{dP_{\text{new}}}{dM_r}$  generated by nuclear fusion at this point, in terms of  $\frac{dP_{\text{old}}}{dM_r}$ . [1]

NOTE: Despite the ongoing contraction, the star is *still* approximately in hydrostatic equilibrium.

**Solution:**

Apply the equation for HE and utilise proportionality:

$$\frac{dP_{\text{new}}}{dM_r} \propto \frac{M_r}{r^4} \quad (1)$$

Since the enclosed mass is *constant* in this case, we can drop it. We then have:

$$\begin{aligned} \frac{dP_{\text{new}}}{dM_r} \div \frac{dP_{\text{old}}}{dM_r} &= \frac{1}{r_{\text{new}}^4} \div \frac{1}{r_{\text{old}}^4} \\ \implies \frac{dP_{\text{new}}}{dM_r} &= \frac{r_{\text{old}}^4}{r_{\text{new}}^4} \times \frac{dP_{\text{old}}}{dM_r} \\ &= \left( \frac{0.4r_{\text{sun}}}{0.1r_{\text{sun}}} \right)^4 \times \frac{dP_{\text{old}}}{dM_r} \\ &= 256 \frac{dP_{\text{old}}}{dM_r} \end{aligned}$$

This means that the fusion rate in this part of the hydrogen fusion shell increases by **256 times!** This is the key reason why red giants are so luminous.

Due to these two factors, the luminosity of the star dramatically increases. Some of the emitted energy is absorbed by the envelope: following the Virial Theorem, this causes the envelope to expand and cool. The star has now become a red giant.

### The Mirror Principle and Helium Fusion

We now come to the main takeaway of this question. The observations and mechanisms in the previous sections generally apply to different stellar models, from low- to intermediate mass stars. This leads to a useful empirical observation known as the ‘Mirror Principle’.

When a star has an active hydrogen fusion shell, this shell acts as a mirror between the core and the outer envelope. In other words, core *contraction* leads to *expansion* of the outer envelope, and core *expansion* leads to *contraction* of the outer envelope.

For an example of the Mirror Principle in action, consider a star at the end of the red giant stage. After sufficient core contraction, conditions in the helium core allow helium fusion to begin. In response, the helium core stops contracting and re-expands from its minimum size at the red giant stage.

At this point, the luminosity of the intermediate-mass star decreases from the red giant phase, and the star correspondingly shrinks and becomes hotter. The star thus leaves the red giant branch, following the

predictions of the Mirror Principle.

- (m) Explain why the luminosity of the star decreases *despite* the introduction of a new source of energy, helium fusion. [2]

**Solution:**

Simply reverse the arguments made in (j) and (k):

1. Core expansion requires energy, reducing the energy output by the core
2. The density/temperature of the hydrogen fusion shell is decreased during core expansion, lowering the pressure gradient required to maintain hydrostatic equilibrium—i.e. the amount of nuclear fusion in the hydrogen fusion shell *falls*.

These two factors outweigh the increase in energy production due to helium fusion, and thus overall luminosity falls.

### 3 Flat and *Somewhat* Round

Protoplanetary discs, galactic discs, planets with rings set in concentric discs—the universe is *full* of discs, but we hardly ever stop to wonder why. Take our solar system, for instance. All the planets move with minimal inclination to the ecliptic. This is clearly because the protoplanetary disc is, well, a disc. But why? What is so special about discs? This question is the focus of **Part I: A Universal Slimming Regime**.

**Part II: Expanding Horizontally** is inspired by the hypothetical Planet Nine. Why do we hypothesise that Planet Nine might exist? We will look at a line of reasoning for this hypothesis, using a famous mechanism in orbital theory.

#### Part I: A Universal Slimming Regime

To understand why discs form, one needs to return to the very beginnings of stellar formation. Stars form from the gravitational collapse of interstellar clouds of gas, also known as nebulae.

- (a) Explain briefly why gravitational collapse occurs. Mathematical arguments are **not** required. [2]

##### Solution:

An interstellar cloud of gas remains in hydrostatic equilibrium if the kinetic energy of the gas balances the potential energy of the internal gravitational force. (Expressed in mathematical terms by the Virial theorem, equilibrium is maintained if gravitational potential energy is twice the internal kinetic thermal energy.) Above a certain mass (known as the Jeans mass), gravitational potential energy will overcome the gas pressure, forcing gravitational collapse.

- (b) As collapse occurs, the cloud undergoes fragmentation until the cloud fragments reach stellar mass. These are called protostellar clouds.

- (i) State **four** possible factors which may affect the fragmentation process. [1]

##### Solution:

Turbulence, rotation, magnetic fields, internal cloud density/matter distribution, cloud geometry, internal flows, etc.

- (ii) Hence, or otherwise, comment on the expected geometry and net angular momentum of the protostellar clouds. Justify your answers. [2]

##### Solution:

Non-uniform geometries and non-uniform net angular momenta.

**Also accepted: non-uniform geometry and non-zero net angular momentum.**

The reason is due to the large number of factors affecting the fragmentation process. These factors, in all probability, are present and affect the cloud non-uniformly. It is therefore far more likely to expect the fragmentation process to be non-uniform.

(c) Each protostellar cloud will continue to collapse towards its own centre of mass, as long as gravitational binding energy can be lost—this is done via radiation in the beginning, and then when the cloud becomes opaque to its own radiation, via dust (infrared emissions). Naturally, density and temperature both increase towards the centre of mass.

(i) How does the motion of the gas in each cloud change during this process? Explain your answer. [2]

**Solution:**

The gas in each cloud starts out somewhat random, and averages out towards the direction of the net angular momentum of the cloud.

The reason is because of the increasing rate of gas collisions. As the rate of collision increases due to the collapse, on average there will be a net bias in resultant angular momenta of the particles towards the direction of the net angular momentum of the cloud, since particles interact much more often.

**Also accepted: The velocity and angular velocity of the gas increases. Since the cloud contracts, the radius of orbital motion decreases. By conservation of angular momentum, the angular velocity must increase.**

(ii) Hence, or otherwise, explain why a protoplanetary disc forms. [3]

**Solution:**

By conservation of angular momentum, the rotation increases as the radius of each cloud (now called a solar nebula) decreases. Notice that centripetal acceleration from orbital motion resists gravitational pull only in the radial direction (or in the orbital plane).

That is to say, coupled with the established direction of net angular momentum of the cloud, and the fact that particles in the cloud average out in this direction, most particles resist the gravitational pull in a disk.

Therefore, the cloud collapse is resisted in a radial direction, but is free to collapse in the vertical direction. As this process continues, the disk is established.

### Part II: Expanding Horizontally

Planets, moons, and asteroids (typically) form out of debris in the protoplanetary disc. As such, one expects the orbital planes of planets, asteroids, and other objects about a star to lie within a narrow range of inclinations. An orbit can be described with six parameters, summarised in Figure 6.

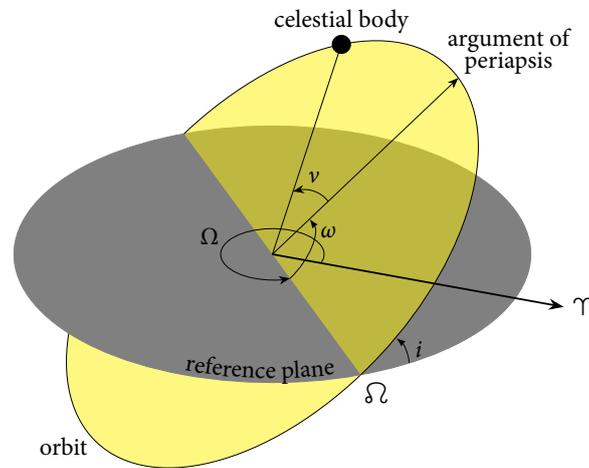


Figure 6: The six parameters of an orbit of a celestial body about another

In this question, we will only focus on **two** parameters: the argument of periaapsis  $\omega$ , which is the angle formed from the plane of reference to the periaapsis, and the inclination,  $i$ , the angle between the plane of reference, and the orbital plane. For the Solar System, the reference plane is the ecliptic.

It is well-known that external gravitational perturbations can alter orbits. Indeed, the discovery of Neptune was due to analyses of perturbations of Uranus’ orbit. Early in the 20<sup>th</sup> century, the search for a ninth planet (Planet X) led to the discovery of Pluto, even if rather inadvertently.

One commonly-known perturbation effect takes place via the Kozai mechanism. This is a three-body system, consisting of two bodies orbiting about a centre of mass (called the *inner binary*) and one outer body also orbiting the centre of mass at an angle, called the *perturber*. The mechanism investigates the effects of the perturber’s gravity on the inner binary.

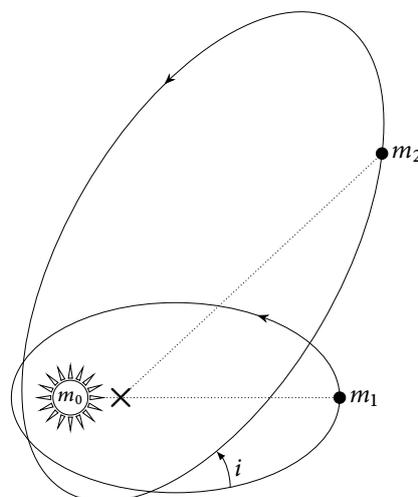


Figure 7: The Kozai mechanism:  $m_0$  and  $m_1$  form the inner binary, and  $m_2$  is the perturber

The Kozai mechanism causes several interesting effects. These include:

- Kozai oscillations, in which the plane(s) of orbit(s) of the inner binary oscillate about the plane of orbit of the perturber. In mathematical terms, this is a periodic oscillation of  $i$ . The period of oscillation is much longer than the orbital periods of the bodies.
- Libration of argument of perihelion/pericentre, i.e. the oscillation of the argument of periapsis of the inner binary orbits about a certain fixed value.

Fundamentally, this is a highly-complicated dynamical problem. We will simplify the situation by making the following assumptions:

- $m_0$  is a highly massive object.
- $m_0 \gg m_2$  (e.g. star-planet systems).
- $m_1 = 0$ .
- All objects are effectively point masses.

Under the above assumptions, we call  $m_0$  the *primary* and  $m_1$  the *secondary*. We need only consider two orbits: the orbit of the secondary  $m_1$ , and the orbit of the perturber  $m_2$ .

For  $m_1$ , there is a conserved quantity  $L_z$ , which we will discuss. If the orbit of  $m_2$  is close to the orbit of  $m_1$ , libration of the argument of periapsis is expected around  $0^\circ$  or  $180^\circ$ . If the orbit of  $m_2$  is distant, libration of the argument of periapsis is expected around  $90^\circ$  or  $270^\circ$ .

The quantity  $L_z$  is dependent on the system and is the component of the secondary's orbital angular momentum parallel to the angular momentum of the perturber. It satisfies

$$L_z = \sqrt{a - be^2} \cos i,$$

where  $a, b$  are integers, which you will be asked to determine,  $i$  is the angle of inclination of the orbit of  $m_1$  with reference to the plane of the perturber, and  $e$  is the orbital eccentricity of the orbit of  $m_1$ .

Table 8 provides an idealised (up to rounding) dataset of six orbits perturbed by  $m_2$  (under the above assumptions).

Eccentricity $\varepsilon$	Inclination $i$
0.225	$39.7^\circ$
0.301	$38.1^\circ$
0.359	$36.5^\circ$
0.413	$34.6^\circ$
0.432	$33.7^\circ$
0.475	$31.5^\circ$

**Table 8:** Ideal data for orbits of massless test particles  $m_1$  about the primary  $m_0$  under perturbation by  $m_2$

- (d) It is given that the ratio,  $a : b$  is in its most simplified form. Plot a suitable **linear** graph to determine  $a$  and  $b$ . Show all appropriate workings, including the linearisation of the given equation. [6]

**Solution:**

- 1 point: linearisation (working required)

Manipulating the equation  $L_z = \sqrt{a - be^2} \cos i$  by squaring and dividing by  $\cos^2 i$ , we obtain

$$\frac{L_z^2}{\cos^2 i} = a - be^2.$$

This is the linearised form for the graph with vertical axis in terms of  $L_z$ .

Alternatively, notice that  $L_z$  is an unknown value. Therefore, the linearised form for the graph using only numbers is

$$\frac{1}{\cos^2 i} = -\frac{b}{L_z^2}e^2 + \frac{a}{L_z^2}.$$

**Note:** some students computed a different linear equation that allowed them to read the **Y-intercept** or the **gradient** directly to obtain the ratio  $a : b$ . This is perfectly fine.

- 1 point: correct coordinate values (working NOT required)

Eccentricity $e$	Inclination $i$	$e^2$	$\frac{1}{\cos^2 i}$
0.225	39.7°	0.051	1.69
0.301	38.1°	0.091	1.61
0.359	36.5°	0.129	1.55
0.413	34.6°	0.171	1.48
0.432	33.7°	0.187	1.44
0.475	31.5°	0.226	1.38

- 1 point: plotted points and axes labelling
- 1 point: line
- 1 point: computation of  $y$ -intercept and gradient of line

Linear regression calculator gives intercept at 1.776 and gradient at  $-1.764$ . The precise value used in creating this question was (of magnitude)  $\frac{16}{9}$

A value of approximately 1.77 or 1.78 is more or less expected. The acceptable range is  $\pm 5\%$ , i.e. 1.68 to 1.87.

- 1 point: values  $a = b = 1$  (minor working or explanation required)

Here, students should realise that they cannot just *multiply* to get astronomically (pun intended) large numbers. Rather, when treating data, often one needs to be aware that the numbers may be slightly off, and account for it. In this case, the  $a : b$  ratio is incredibly close to  $1 : 1$ , so this should give  $a = b = 1$ .

Use the following information to answer part (e).

Recently, analysis of the orbits of trans-Neptunian objects have led to the hypothesis and a renewed search for Planet Nine, thought to be between five to ten Earth masses. Consider the following dataset of Kuiper Belt objects (KBOs). These objects have been used in existing studies debating Planet Nine.

Object	Semi-major axis	Orbital eccentricity	Argument of periapsis	Inclination w.r.t. ecliptic
VP113	266	0.69	292.8°	24.1°
TG422	503	0.93	285.7°	18.6°
GB174	351	0.87	347.8°	21.5°
<b>Sedna</b>	507	0.85	311.5°	11.9°
FT28	295	0.86	40.2°	17.3°
KG163	680	0.95	32.0°	14.0°
SY99	730	0.93	32.4°	4.2°
TG387	1200	0.94	118.2°	11.7°
GT50	310	0.89	129.2°	8.8°
<i>RF98</i>	364	0.90	311.8°	29.6°
SR349	299	0.84	341.4°	18.0°
VN112	327	0.85	327.1°	25.6°
RX245	430	0.89	65.4°	12.2°
FE72	1600	0.98	134.4°	20.6°

**Table 9:** Some properties of orbits of some KBOs; the reference plane is the ecliptic

In Table 9, objects in **boldface** have been studied by Trujillo and Sheppard (2014) in the initial hypothesis of Planet Nine. The *italicised* object was added to the bolded ones in a 2016 study by Brown and Batygin. The most recent study (2019) considers *all* the above objects.

Following are further notes on the above objects. The group of objects studied by Brown and Batygin exclude those with unstable orbits due to close approaches to Neptune or mean-motion resonances. The six objects in **boldface** and *italics* display clustering in longitudes of perihelion, have tilted orbits which are roughly coplanar, and were discovered by six different telescopes on six different surveys (one object each).

[Turn over]

- (e) In Figure 6, for an object in orbit around the Sun, the angle sum of the longitude of the ascending node  $\Omega$  and the argument of periapsis  $\omega$ , is the longitude of perihelion,  $\varpi$ .

Discuss whether the Kozai mechanism could be a valid justification for the presence of Planet Nine. You should provide arguments for and against, weigh them, and decide on a stance. [4]

**Solution:**

First, there are many bodies, so we need to make clear where each body is.  $m_0$  is the Sun.  $m_1$  refers to the eTNOs, and  $m_2$ , the perturber, is the hypothesised Planet Nine.

Possible arguments for:

1. With the exception of TG387, GT50, and FE72, the remainder of the objects have arguments of periapsis clustered around  $0^\circ$ , fitting the expected effect of a perturbation via the Kozai mechanism.
2. The high orbital eccentricities can be explained by a conserved quantity  $L_z$ , which trades off inclination for eccentricity.
3. The clustering of the six objects cannot be reasonably explained by observer bias (it could, but the probability would be really low). The clustering of longitudes have a very low probability of occurrence due to random chance, due to differing rates of precession (due to eccentricities and semi-major axes differing). That is to say, the clustering is very unlikely to be due to an event in the (distant) past, and supports the hypothesis of a gravitational influence.

Possible arguments against:

1. The number of objects studied is small, and the observations of clustering could be due to some astronomical coincidence. Of astronomical probabilities. Astronomically crazy stuff, it is!
2. There exist more than one clustering group, as many as three centred around approximately  $310^\circ$ ,  $40^\circ$ , and  $125^\circ$ . In fact, the study by Brown and Batygin find an argument of periapsis clustering around  $310^\circ$ , which differs from the Kozai prediction.
3. The dataset is largely unreliable due to a number of other effects, such as close approaches to Neptune, or mean-motion resonances.

## 4 Presumptuous Assumptions

### Introduction

When we consider problems regarding a star-planet system, or planet-moon system, assumptions are often made about the system to simplify calculations. For example, one extremely popular assumption is that of ‘circular orbits’ (i.e. eccentricity,  $\varepsilon = 0$ ). Another popular assumption is that of ‘point masses’, when performing gravitational computations.

By the very nature of Astronomy, assumptions abound. Some of these assumptions are perfectly valid—for instance, ‘point masses’ are *usually* valid because the distances between said masses are much greater compared to the objects themselves. Some are invalid: for instance, circular orbits will fail in almost *any* scenario involving comets.

We say that an assumption is *valid* if the deviation from the assumption does *not* affect the theory under the assumption significantly. The assumption *fails* if said deviation affects said theory.

It is clear that some assumptions are simply made for the sake of simplification. In this question, we take a brief look into the world of assumptions.

### Part I: How Big is the Hill?

The Hill sphere of an astronomical body  $X$  is the region in which the gravitational attraction of  $X$  dominates. In other words, if we theoretically consider only gravitational influences, a satellite might be expected to be found in an orbit around  $X$  if and only if the orbit lies within the Hill sphere of  $X$ . The Hill radius is the radius of the Hill sphere. Of course, the word ‘sphere’ is a misnomer—the Hill sphere is not *precisely* spherical.

Consider first the problem of a three-body system, consisting of

- a body  $P$  with mass  $m$ ,
- a body  $Q$  with mass  $M$ , and
- a satellite  $S$  of mass  $\mu$ .

Assume that  $P$  is orbiting about  $Q$  with a circular orbit of radius  $r$ , and  $S$  is orbiting  $P$  at the Hill radius  $r_H$  of  $P$ .

When  $S$  is lying on the line joining  $Q$  and  $P$ , the forces exerted on  $S$  at that point can be expressed as

$$\frac{Gm\mu}{(r_H)^2} + \frac{GM\mu(r - r_H)}{r^3} = \frac{GM\mu}{(r - r_H)^2}.$$

Of course, in doing so, we have made a number of assumptions, one of which being that all the bodies are point masses.

- (a) As the above equation suggests, there are three components to consider in balancing the forces. Name (or explain the origins of) these three components, and match them to their respective terms in the equation. [1.5]

#### Solution:

This is actually a very well-disguised  $F_c = F_g$  (centripetal force equals gravitational force) equation.

- $\frac{Gm\mu}{r_H^2}$ : Gravitational force of  $P$  acting on  $p$ . This is obvious.

- $\frac{GM\mu}{(r-r_H)^2}$ : Gravitational force of  $S$  acting on  $p$ . This is also obvious.
- $\frac{GM\mu(r-r_H)}{r^3}$ : Centripetal force acting on  $p$ , considered as an orbit around  $S$ . This is perhaps not so obvious, but it arises due to fact that  $p$  is being dragged in orbit about  $S$  by the combined gravitational forces. Consider that the centripetal force on  $p$  about  $S$  is given by  $\mu\omega^2(r-r_H)$ , where  $\omega$  is the angular velocity of the planet, i.e.

$$m\omega^2 R = \frac{GMm}{R^2}.$$

Then  $\omega^2 = \frac{GM}{R^3}$ , which is exactly what was needed,

(b) There are **three** major assumptions made in the above scenario. These assumptions are respectively regards to: firstly, the relation between  $M$  and  $m$ ; secondly, the relation between  $\mu$ ,  $Q$ ,  $S$ , and  $P$ ; and thirdly, other forces acting on the system.

State these three assumptions.

[1.5]

**Solution:**

1. The relation between  $M$  and  $m$ :  $M \gg m$ .
2.  $\mu$ ,  $p$ ,  $S$ , and  $P$ :  $\mu$  is sufficiently small such that the gravitational effect of  $p$  on  $S$  and  $P$  is negligible.

**Also accepted:**  $\mu \ll m, M$ .

3. Other forces acting on the system: No other forces, i.e. the system is an isolated system.

(c) Show that the Hill radius  $r_H$  of  $P$  is approximately given by

$$r_H \approx r \sqrt[3]{\frac{m}{3M}}.$$

[3]

NOTE: You may need to use binomial approximation(s). If you do, or if you make any other mathematical assumptions, you are required to justify the validity of the assumption made. Justifications of assumptions need not be rigorous—a short sentence will suffice.

**Solution:**

From

$$\frac{Gm\mu}{r_H^2} + \frac{GM\mu(r-r_H)}{r^3} = \frac{GM\mu}{(r-r_H)^2},$$

by dividing throughout by  $G\mu$ , we obtain

$$\frac{m}{r_H^2} + \frac{M(r - r_H)}{r^3} = \frac{M}{(r - r_H)^2}.$$

Rewriting this equation, we obtain

$$\frac{m}{r_H^2} + \frac{M}{r^2} \left(1 - \frac{r_H}{r}\right) = \frac{M}{r^2} \left(1 - \frac{r_H}{r}\right)^{-2}.$$

Since  $M \gg m$  and  $\mu$  is sufficiently small, therefore  $r_H \ll r$  and  $\frac{r_H}{r}$  is small. (The long version: The binomial approximation ensures  $(1 + x)^m \approx 1 + mx$  if  $x$  is small. Since  $\mu$  is sufficiently small such that the gravitational effect of  $p$  on  $S$  and  $P$  is negligible, and  $M \gg m$ , it is reasonable to conclude that  $r \gg r_H$ . Then  $\frac{r_H}{r}$  is small, so the binomial approximation is a reasonable assumption.) By the binomial approximation,

$$\frac{m}{r_H^2} + \frac{M}{r^2} \left(1 - \frac{r_H}{r}\right) = \frac{M}{r^2} \left(1 - \frac{r_H}{r}\right)^{-2} \approx \frac{M}{r^2} \left(1 + 2\frac{r_H}{r}\right).$$

Multiplying throughout by  $r_H^2 r^2$ , we obtain

$$mr^2 + M \left(r_H^2 - \frac{r_H^3}{r}\right) \approx M \left(r_H^2 + 2\frac{r_H^3}{r}\right).$$

Rearranging yields

$$mr^2 \approx 3\frac{Mr_H^3}{r}.$$

Therefore,

$$r_H \approx r \sqrt[3]{\frac{m}{3M}},$$

as required.

In Astronomy, any theory made must be checked in order to determine its validity. Checking typically consists of three steps:

1. Ensuring it is mathematically consistent.
2. Comparison of the predicted results with real-world data.
3. Ensuring the assumptions made are valid.

If a theory applied to a given scenario holds up to the above three parts, it is a valid theory to describe that scenario.

Tables 10a and 10b list the distance in AU of several of Jupiter's 79 known moons and Saturn's 62 known moons from their respective parent planets (as of 2018). The 'Position' column ranks the distance of the moon from Jupiter in increasing order. That is to say, a moon of Jupiter with position 1 is nearest to Jupiter,

[Turn over]

while a moon of Jupiter with distance 60 has 59 moons closer to Jupiter than itself.

Position	Moon	Semi-major axis (AU)	Position	Moon	Semi-major axis (AU)
1	Metis	$8.61 \times 10^{-3}$	10	Mimas	$1.24 \times 10^{-3}$
8	Callisto	$1.26 \times 10^{-2}$	22	Titan	$8.16 \times 10^{-3}$
27	Orthosie	$1.37 \times 10^{-1}$	48	Hati	$1.32 \times 10^{-1}$
53	Kallichore	$1.54 \times 10^{-1}$	57	Fenrir	$1.47 \times 10^{-1}$
70	Hegemone	$1.58 \times 10^{-1}$	62	Fornjot	$1.64 \times 10^{-1}$
78	Megaclite	$1.65 \times 10^{-1}$	(b) Moons of Saturn		

(a) Moons of Jupiter

Figure 10: Data on moons of Jupiter and Saturn

(d) In (a) to (c), we investigated a simple theory under several assumptions and obtained an expression for the Hill radius. Let us call this Theory H (for Hill).

Based on the given data and the assumptions made, is Theory H valid? Explain your answer clearly. [3]

**Solution:**

It is not valid.

- Assumptions:

- The assumption which fails is that of an isolated system. There are a number of outside factors exerting forces on the Sun-Jupiter and Sun-Saturn system. For example, perturbations by other bodies in the Solar System, radiation pressure, the solar wind, etc.

- Numerical:

- The numerical predictions do not coincide with the observed data.

The Hill radius for Jupiter (use the formula sheet!) is approximately

$$r_H \approx 7.785 \times 10^{11} \times \sqrt[3]{\frac{1.899 \times 10^{27}}{3 \times 1.989 \times 10^{30}}} \approx 5.31514 \text{ m} \approx 0.35530 \text{ AU}.$$

while the Hill radius for Saturn is

$$r_H \approx 1.433 \times 10^{12} \times \sqrt[3]{\frac{5.685 \times 10^{26}}{3 \times 1.989 \times 10^{30}}} \approx 6.54492 \text{ m} \approx 0.43750 \text{ AU}.$$

In both cases, it is clear that the orbits of the moons of both planets lie well within the Hill radius (at about half to one-third the Hill radius). Clearly this contradicts

the predictions made by Theory H, in which we expect moons up to close to the Hill radius as gravitational attraction of the planet dominates.

## Part II: Working in the Age of Assumptions

Assumptions are ultimately used to simplify matters. There are **two** general types of assumptions we make in Astronomy (and everyday life): assumptions used to simplify theories, and assumptions used to rule out possibilities.

Regardless of type, the basic principles of using assumptions remain identical no matter the scenario, and are given in a short two-step process.

1. Make assumptions.
2. Deduce the consequences of holding these assumptions.

The widespread use of Newtonian gravitation in place of general relativity is a classical example of the first sort of assumption. General relativity is mathematically daunting, whereas Newtonian gravitation provides a much simpler alternative, while being valid in most human scales.

In this part, we will focus on the second type. It might *sound* fancy, but this should be nothing new to you. If you have ever solved a question by thinking along the lines of ‘If this is true, then...nope, that can’t be right...’, you have made successful use of the second type of assumptions to rule out a case you were considering. In fact, using this sort of assumption to prove a conjecture, is called *proof by contradiction*. Therein lies the strength of using assumptions.

The following questions can be broken down into cases. Whether you use the method of assumptions is up to you; but, if you do, remember that sometimes, *all* cases have implications, and you need to weigh each case against the others.

- (e) When a satellite is ‘placed at a Lagrangian point’, it is not actually placed *at* the point itself. Rather, the satellite orbits the point. Satellites orbiting these points typically employ *station-keeping*.
- (i) Define station-keeping, and explain why it is needed. [1]

### Solution:

Station-keeping refers to the process of keeping a satellite in a particular orbit via orbital manoeuvres by thruster burns.

It is required because of perturbations of the satellite by external forces. For example, dynamical perturbations of an n-body system. Small perturbations of a satellite from the orbit can occur due to gravitational influences of other massive bodies.

- (ii) At the Sun-Earth  $L_1$  and  $L_2$  points, larger orbits are generally preferred. Give **one** reason why, for each point. [2]

**Solution:**

$L_1$ : There will be less solar interference with satellite-Earth communications.

The idea is that if a smaller orbit is used, there will be more solar interference, as compared to further out. Observational capacity remains approximately the same, therefore a larger orbit is obviously preferred.

$L_2$ : The satellite will spend more time outside Earth's shadow. This allows more time for solar panels to collect sunlight and provide power to the satellite.

Compare this to a tighter energy budget for a smaller orbit. Regardless of the mission, energy is absolutely essential.

- (f) Generally, in literature, astronomers use redshift values to describe the distances of *extremely* distant objects like quasars and active galactic nuclei, but use terms of absolute distance (e.g. light years, kilometres, etc.) to describe distances of nearer objects. Why is this so? [2]

**Part III: Tremulous Theories**

It is unfortunate, however, that sometimes we cannot help but assume certain facts to be true. We cannot freely move about in space-time to confirm every assumption we make! One of the most obvious examples underpins the  $\Lambda$  Cold Dark Matter ( $\Lambda$ CDM) model, the standard model of the universe today.

$\Lambda$ CDM is referred to as the *standard model* of Big Bang cosmology because it is the simplest model which accounts for cosmic microwave background radiation, large-scale structure of galaxy distribution, abundance of hydrogen, helium, and lithium in the universe, and accelerating expansion of the universe.

There are other models based on differing assumptions, but a standard assumption is that of the *cosmological principle*.

- (g) The cosmological principle claims that the universe satisfies *homogeneity* and *isotropy*. Define these two terms. [1]

**Solution:**

**Homogeneity:** Observers at any location in the universe will have the same observational evidence.

**Isotropy:** The same observational evidence is available in any direction of measurement.

- (h) The cosmological principle **cannot** be verified today. Of the two conditions in (g), it is impossible to *truly* assess the validity of one of them. Yet, we can still argue that the cosmological principle is a reasonable assumption.

Explain which condition **can** be assessed for validity today, and explain why the other condition **cannot**. [2]

**HINT:** You may find a diagram helpful.

**Solution:**

Homogeneity is impossible to assess, while isotropy is possible to assess.

The answer for homogeneity is best intuited with a light cone.

As observers from Earth, we only have access to one cosmic location and thus one light cone. It is impossible for us to observe aspects of the universe outside our light cone. Therefore we cannot assess homogeneity.

Isotropy, however, can be assessed, since it relies on *direction* of measurement rather than location. From Earth, we have access to many directions, allowing us to assess isotropy.

- (j) For the assessable condition identified in (h), there are a number of observations which, taken together, strongly imply it. Give **three** such observations. [2]

**NOTE:** While explanations for each observation are *not* required and will *not* be awarded any marks, you may include explanations if you feel that you need to explain your answers to increase the chance of acceptance.

**Solution:**

- Consistent matter/energy distribution (e.g. CMB, GRB, etc.)
- Consistent trend in matter (e.g. galaxies, galaxy clusters) velocity away from us, which is due to universal expansion.
- Consistent gravitational lensing by large-scale objects.
- Consistent measurements of angular diameter distance, i.e. objects of the same absolute size appear smaller with increasing distance. (This one is rather interesting. Beyond approximately  $z = 1.5$ , a further object will in fact appear *larger*.)

- (k) Hence, or otherwise, suggest why the non-assessable condition identified in (h) is a reasonable assumption. [1]

**Solution:**

If we observe consistent isotropy over time, it suggests that the universe in our light cone has

evolved in a matter consistent with that of a homogeneous universe. One may assume this is true at all points of the universe (via the Copernician principle, which is a weaker assumption than the cosmological principle).

A slightly more mathematical answer (which is NOT expected) involves tests on the FLRW metric. In the loosest possible layman's terms, if certain conditions on isotropy hold in a region, then the FLRW metric holds in that region. Since homogeneity implies the FLRW metric, tests for (non-)violation of this metric (e.g. dark energy measurements from Type Ia supernovae) could be construed as an argument for homogeneity together with the Copernician principle.

## 5 Questions of Galactic Proportions

### Introduction

Galaxies—how they so fascinate the imagination! From Star Wars to Doctor Who, science fiction tells tales of aliens and travels throughout various galaxies in the Universe. Yet, in these shows, little consideration is usually given to the *science* of the galaxies themselves. Galaxies are fascinating, and the science is so much more than that.

In this question, we will investigate what we can learn from galaxies.

### Part I: The Hypes and Types of Galaxies

To begin with, there are several different types of galaxies, and we generally categorise them into the following:

- *spiral galaxies*, which have a flat galactic disk with spiral arms and a bulge at the middle, called the galactic centre;
- *elliptical galaxies*, which are ellipsoidal or spheroidal, rather like a rugby ball or an M&M™ chocolate;
- *irregular galaxies*, which have no discernible structure, and hence do not fit in either of the former.

In fact, half to two-thirds of all spiral galaxies have a barred structure, and we call them *barred spiral galaxies*. Our own Milky Way is such a barred spiral.

The visual classification, as it turns out, is not simply for *show* (pardon the pun). Indeed, galaxies of each class tend to display other distinctive properties characteristic of the class. Table 11 lists some properties for spiral and elliptical galaxies.

Characteristic	Spiral	Elliptical
Shape	Flat disk with arms and central bulge	Spheroid/Ellipsoid
Rotation	Rotating disk	Radial orbits about the centre
Interstellar gas	Lots	Little
Stellar ages	Mixed	Old

Table 11: Some characteristics of spiral and elliptical galaxies

- (a) Differentiate, if possible, between spiral and elliptical galaxies, in terms of **star formation rates**, and **colour**. [2]

#### Solution:

Spiral – High/present in significant quantities. Elliptical – low.

Spiral galaxies have mixed stellar ages – evidence of stellar formation. They also have lots of interstellar gas to fuel formation.

Elliptical galaxies have old stellar ages – evidence of little stellar formation. Also, little intersellar gas, so formation will be extremely limited.

Spiral – bluer. Elliptical – redder.

Generally, elliptical galaxies have higher composition of older stars as compared to spirals, and older stars appear redder than newer stars. Consequently, spirals should generally appear bluer, and ellipticals redder.

- (b) It is commonly said that spiral galaxies are dominant throughout the universe, since nearly four-fifths of observed galaxies are spiral. However, deep sky surveys show a larger proportion of elliptical galaxies as compared to spiral galaxies, suggesting elliptical galaxies are more dominant instead. Resolve this apparent paradox. [2]

**Solution:**

The solution is that ellipticals are probably more dominant.

The catch with the spirals making up nearly 80% of observed galaxies is due to observational bias. Spirals are brighter, generally, so they are easier to observe, and so many more are documented as compared to ellipticals. Deep sky surveys are less affected by this bias as it surveys much dimmer objects as well, so they are probably more accurate on this point.

In the morphology of galaxies, there is also the S0 type, also known as lenticular galaxies. They have commonalities with both elliptical and spiral galaxies. Lenticular galaxies are commonly characterised as a disk with a central bulge, but *without* arms. They share stellar formation and age features with elliptical galaxies, but may contain plenty of dust, unlike their elliptical counterparts. One might imagine that if we view a lenticular galaxy edge-on, it can be difficult to distinguish it from an elliptical galaxy. The origins of these galaxies are not yet well-understood. Common theories include fading of spiral arms and galaxy mergers.

Table 12 lists data on approximate compositions of different cluster types by galaxy class:

Cluster Type	Elliptical (E)	S0	Spiral (Sp)
Regular	35%	45%	20%
Intermediate (Spiral-poor)	20%	50%	30%
Irregular (Spiral-rich)	15%	35%	50%
Field	10%	20%	70%

**Figure 12:** Typical galactic content of galaxy clusters ( $r \lesssim 1.5 \times 10^{-1}$  Mpc). The *field* refers to the collection of gravitationally isolated galaxies, i.e. those galaxies not belonging to any cluster.

- (c) Account for the main difference(s) in the composition of galaxy clusters as compared to the field. [2]

**Solution:**

The main differences are that the percentages of elliptical and S0 galaxies is higher than the field, and the percentage of spirals is much lower than the field.

This can be explained since clusters have numerous gravitationally bound galaxies in a small (relatively speaking) space. This increases the rate of galaxy collisions, which facilitate the formation of ellipticals and maybe S0 galaxies. In contrast, the field consists of isolated galaxies, thus the rate of collision is much lower, so lesser E and S0 galaxies form.

- (d) Hence, or otherwise, suggest a measurable quantity (e.g. mass, luminosity, etc.) of clusters correlated with the  $\frac{E+S0}{Sp}$  ratio of clusters. Explain your answer. [1]

**Solution:**

Galaxy density, i.e. galaxies per megaparsec cubed (or similar).

Pretty self-explanatory, actually. Higher galaxy density means higher E and S0, and lower Sp due to higher collision rates, so we expect it to be correlated with the  $\frac{E+S0}{Sp}$  ratio.

**Other well-reasoned answers are also accepted.**

**Part II: Round and Round and Round We Go**

One of the most well-known pieces of evidence of dark matter are the *rotation curves* of galaxies throughout the universe. A rotation curve of an  $N$ -body system orbiting about a given axis is a graph of the radial velocity of the bodies against distance from the axis. For example,

- the plot of radial velocity of the planets ( $N$  bodies) around the Sun (axis) against distance;
- the plot of radial velocities of stars in a galaxy ( $N$  bodies) around the centre of the galaxy (axis) against radial distance from the centre of the galaxy.

It is this second example we are concerned with. Of course, it is virtually impossible to resolve individual stars in an external galaxy, so observations are made by regions instead.

- (e) The HI line, also known as the 21 cm line is a powerful tool used to plot rotation curves of spiral galaxies. Explain **one** advantage and **one** disadvantage of using this method to plot rotation curves. [2]

**Solution:****Advantages:**

- Most of the interstellar media of the spiral galaxies is cold neutral hydrogen gas cloud lying in the galactic disk, giving a good representation of rotation throughout the disk.
- Gas cloud usually extends significantly beyond the visible disk (up to 4 times further, but typically about 2 times further), allowing for measurements of rotational velocities beyond

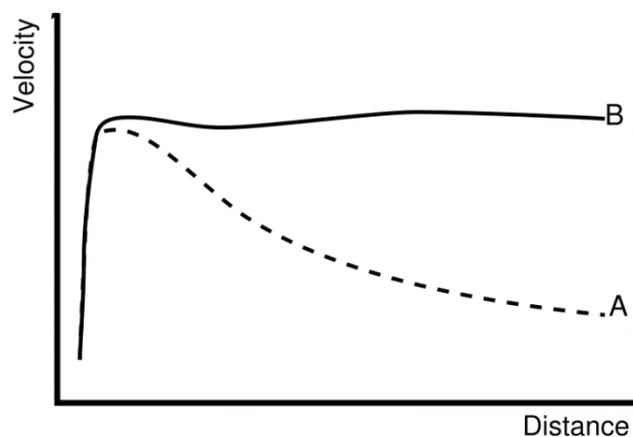
the traditional visible disk.

- Average column density of neutral hydrogen is about the same in all spiral galaxies regardless of surface brightness (postulated to be due to self-shielding), giving a standard base to measure rotation of spiral galaxies with.
- **DO NOT ACCEPT STANDALONE:** The HI line suffers little absorption in spiral galaxies. (Remark: this is true, but has little bearing on the question unless combined with another reason, e.g. the ones given above. Combining these reasons will be looked upon favourably.)

Disadvantages:

- Lower resolution than other gases with shorter emission wavelengths, e.g. CO, so the mapping has poorer spatial resolution as compared to detections in smaller wavelengths. This is especially true nearer to the galactic centre.
- Nearer to the galactic centre, hydrogen gas does not dominate – instead, (hotter) molecular gas forms most of the gas. Thus HI maps of galactic centres are generally poor.
- It is necessary to observe at high resolutions. Rotation curves from HI line observations at low resolutions may have poor correlation with optical velocities (see e.g. Rubin, 1989 on Virgo spirals). There is (quote) excellent correlation at high resolution.

A particularly well-known phenomenon of galactic rotation curves diverging from Keplerian orbital dynamics, is illustrated in Figure 13. The  $x$ - and  $y$ -axes refer to distance from a given galactic nucleus (i.e. galactic orbital radius) and stellar velocity, respectively.



**Figure 13:** Expected Keplerian rotation curve (A) and *actual* rotation curve (B) for spiral galaxies

- (f) There are two main features of the Keplerian graph: the initial rise, and the eventual fall. Explain **both** observations. [4]

**Solution:**

The initial rise follows rigid body rotation and is explained by the close proximity of stars in the

central bulge. Due to the high stellar density, stars interact with each other extremely often and therefore tend to behave more like vibrating atoms rather than a differential gas. Ergo, it rotates like a rigid body, so rigid body rotation it is. For rigid body rotation,  $v \propto r$ , as can be easily proven since the period  $T = \frac{2\pi r}{v}$  is constant, so  $v = \frac{2\pi}{T}r$ , i.e.  $v \propto r$ .

The subsequent drop is due to aforementioned Keplerian rotation. Stars revolve about the galactic centre, now. Equating centripetal force with gravitational force, we have

$$\frac{mv^2}{r} = \frac{GMm}{r^2},$$

i.e.  $v = \sqrt{\frac{GM}{r}}$ , or  $v \propto r^{-\frac{1}{2}}$ . This relationship characterises Keplerian rotation. It is expected to hold since beyond the bulge, stellar density drops sharply, and the effects of neighbouring stars are much lesser. So rotation starts to be expected to mimic those by planets.

This sharp drop in stellar density also explains why the Keplerian effect dominates the increase in mass inside the orbital radius – the increase in (visible) mass is far too low to have significant impact on the Keplerian effect.

The higher actual radial velocity suggests that more mass is present than what is observable. The unobservable component is called *dark matter*.

Spiral galaxies are said to be surrounded by a blob of dark matter, called the *dark matter halo*. The determining equations for the shape of the dark matter halo are rather complicated (for this competition, at least), but suffice to say the blob takes the shape of an oblate spheroid, or a rugby ball.

However, *you* don't have time to investigate an oblate spheroid, so we're going to assume the blob is spherical, just like physicists like to assume everything is a sphere, a point mass, a light string, or a frictionless surface.

The Navarro-Frenk-White (NFW) profile is a density profile commonly used to model dark matter halos. It is valid for isolated halos. Mathematically, the NFW profile for spherical dark matter halos is

$$\rho(r) = \frac{\rho_0}{\frac{r}{R_s} \left(1 + \frac{r}{R_s}\right)^2},$$

where

- $\rho(r)$  is the density of dark matter at radius  $r$  from the centre of the halo, and
- $\rho_0$  and the scale radius  $R_s > 0$  are constants determined by the system in consideration, chosen to fit the profile curve to the observed data. Note that  $R_s$  is *not* the radius of the galaxy.

[Turn over]

In particular,  $\rho_0$  is related to the critical density of the Universe  $\rho_c = \frac{3H_0^2}{8\pi G}$ , where  $H_0$  is the Hubble constant today and  $G$  is the gravitational constant. The constant  $\rho_0$  is given by

$$\rho_0 = \delta_0 \rho_c = \frac{200}{3} \left( \frac{c^3}{\ln(1+c) - \frac{c}{1+c}} \right) \rho_c,$$

where  $c$  is known as the *concentration parameter*, and is determined by the system in consideration.

This model is not without its problems. One of the main problems of this model is the fact that the total mass in the halo is predicted to be infinite, which is physically impossible.

- (g) By considering the mass  $M(R)$  of dark matter contained in a general sphere of radius  $R$ , prove that the model predicts that  $\lim_{R \rightarrow \infty} M(R) = \infty$ , i.e. prove that  $M(R) \rightarrow \infty$  as  $R \rightarrow \infty$ . [4]

**Solution:**

Since  $\rho$  is the density of dark matter, it follows that in a spherical shell at radius  $r$ , the mass of dark matter is

$$\begin{aligned} m(r) &= 4\pi r^2 \frac{\rho_0 R_s^3}{r(R_s + r)^2} \Delta r \\ &= \frac{4\pi r \rho_0 R_s^3}{(R_s + r)^2} \Delta r. \end{aligned}$$

The total mass in a sphere of radius  $R$ , then, is

$$M(R) = 4\pi \rho R_s^3 \int_0^R \frac{r}{(R_s + r)^2} dr.$$

One recognises the obvious partial fraction decomposition

$$\begin{aligned} \frac{r}{(R_s + r)^2} &= \frac{R_s + r}{(R_s + r)^2} - \frac{R_s}{(R_s + r)^2} \\ &= \frac{1}{R_s + r} - \frac{R_s}{(R_s + r)^2}. \end{aligned}$$

This admits the easy integration

$$\begin{aligned} M(R) &= 4\pi \rho R_s^3 \int_0^R \frac{r}{(R_s + r)^2} dr \\ &= 4\pi \rho R_s^3 \int_0^R \left( \frac{1}{R_s + r} - \frac{R_s}{(R_s + r)^2} \right) dr \\ &= 4\pi \rho R_s^3 \left[ \ln(R_s + r) + \frac{R_s}{R_s + r} \right]_0^R \\ &= 4\pi \rho R_s^3 \left( \ln(R_s + R) + \frac{R_s}{R_s + R} - \ln R_s - 1 \right). \end{aligned}$$

Since  $R_s > 0$ , as  $R \rightarrow \infty$  it is clear that  $\ln(R_s + R) \rightarrow \infty$  and  $\frac{R_s}{R_s + R} \rightarrow 0$ . But this means  $M(R) \rightarrow \infty$ .

Thanks to this problem, we need to define the edge of the halo, i.e. a ‘cut-off point’, so to speak. Typically, the virial radius  $R_{200} = cR_s$  is used, i.e. we ‘define’ the halo to have radius  $R_{200}$ . Note that typically,  $4 \leq c \leq 40$ , and the Milky Way itself has  $10 \leq c \leq 15$ .

One other problem is known to arise with the NFW profile (and with many other commonly-used models). Let us call this *Problem X*. The fact is, the rotation curves of most observed dwarf galaxies imply that they have flat central dark matter density profiles. Problem X arises from consideration of this particular fact of dwarf galaxies.

- (h) For dwarf galaxies, we may assume  $R_s \approx 1$  kpc and  $c \approx 7$ . Suggest what Problem X is. Explain your reasoning. [3]

**Solution:**

Problem X, also known as the *core-cusp problem* or the *cuspy halo problem*, is the issue with the behaviour of the NFW profile for low-mass galaxies. Specifically, the profile predicts that density of dark matter sharply increases with small  $r$  as  $r \rightarrow 0$ , in contrast with the observed flat dark matter density profile of dwarf galaxies.

Dwarf galaxies are on the order of 10 kpc diameter. For the range  $0 < r \leq 5$  kpc, the graph of  $\rho$  against  $r$  in the NFW profile is most certainly not flat. Ignoring  $\rho_0$  and computing the graph of

$$y = \frac{1}{r(1+r)^2}$$

yields a noticeable negative gradient starting from about  $r = 3$  and increasing steeply from  $r = 1$ . So for dwarf galaxies, we would expect about 60% of the galaxy to exhibit a noticeably decreasing  $\rho$  as  $r$  increases, which contradicts what we know of dwarf galaxies!

Anyway, the salient point is that it is enough to discuss the shape of the profile as  $r \rightarrow 0$ , with mention of the main graphical feature leading to the contradiction (and hence suggesting what Problem X is), and why indeed it is a contradiction. It is not necessary to perform the blatant substitution of values to compute  $\rho_0$  (since it is a constant).

Also, other well-reasoned answers are certainly welcome.