NOTES OF
ASTROPHYSICAL PROCESSES

version 1.0

Giacomo Pannocchia

a.y. 2025-2026

For those who come after

PREFACE

Not long ago, it occurred to me how cool it is when someone unexpectedly releases a very detailed and all-comprehensive version of their notes, especially when dealing with a course that has a handful of really different topics often interacting together in unpredictable, yet fascinating, ways—-as it's the case for the Astrophysical Processes class.

These notes will be mainly based on *my* own notes of the lectures by Professor Walter del Pozzo and Professor Marco Crisostomi during the academic year 2025-2026. However, since I take little to no pride in my messy notes, I'll be using more often than not some of the many references you can find on the course catalogue page or in the bibliography of this humble collection.

You can report errors (whatever their nature might be) and suggestions for additions at g.pannocchia3@studenti.unipi.it or through whatever convoluted way (conventional or not) you prefer[1].

Without further ado, we'd better not lose much more time on a preface and get started with it.

There was Eru, the One, who in Arda is called Ilúvatar; and he made first the Ainur [...] *But for a long while they sang only each alone, or but few together, while the rest hearkened; for each comprehended only that part of the mind of Ilúvatar from which he came, and in the understanding of their brethren they grew but slowly.*

Yet ever as they listened they came to deeper understanding, and increased in unison and harmony.

Ainulindalë, "The music of the Ainur",
Silmarillion, J. R. R. Tolkien

1 I'd like, however, not to see my house stormed by homing pidgeons.

CONTENTS

Part I

RADIATIVE TRANSPORT

1 INTERACTION OF RADIATION WITH MATTER

1.1 INTRODUCTION

Most of our knowledge about the Universe is based on the electromagnetic radiation that reaches us from far far away. EM radiation is obviusly not the only way we can probe the Universe we live in but, in respect to neutrinos, cosmic rays or even gravitational waves, it's not a long stretch to claim it is by far the most understood.

It is most important then that an astrophysicist worthy of his (or her) name has a good grasp of the theory of radiative transfer and of its applications.

Apart from a few more key differences, I'll follow the description of radiative transfer of [7] and [20], but I won't fail to emphasize whenever I'll be doing otherwise.

1.2 RELEVANT QUANTITIES FOR RADIATIVE TRANSFER

Although some books often start their description of radiative transfer from the definition of *monochromatic energy* and *monochromatic intensity*, I found that it is most misleading, since, in all but a few cases, what we experimentally measure are fundamentally *fluxes*.

We shall then consider the *monochromatic flux* F_ν ($\mathrm{erg\,s^{-1}\,Hz^{-1}\,cm^{-2}}$) produced by some source passing through a small area $\mathrm{d}A$ located somewhere in space.

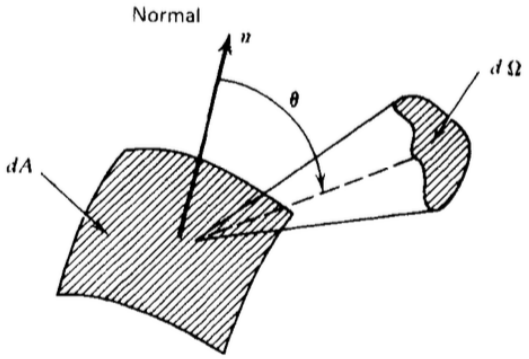

Figure 1: Schematic geometrical representation of the system.
Credits: G. Rybicki, A. Lightman [20].

If we call $\hat{k}$ the propagation direction of the flux and $\hat{n}$ the unit vector emerging from the surface $\mathrm{d}A$, it's easy to get convinced that what is actually passing through the surface is something proportional to $F_\nu(\hat{k}\cdot\hat{n})$.

From the monochromatic flux we can define the *bolometric flux*, which is just the monochromatic flux integrated over all frequencies (or wavelengths)

$$F = \int_0^{+\infty} F_\nu \, \mathrm{d}\nu = \int_0^{+\infty} F_\lambda \, \mathrm{d}\lambda \tag{1}$$

This also tells us how to convert a flux per unit frequency to a flux per unit wavelength

$$F_\nu \, \mathrm{d}\nu = F_\lambda \, \mathrm{d}\lambda$$

By now it should be clear that, despite being experimentally sensible to use the flux, we're losing much information sticking with it, namely directional information.

We consider then the amount of radiation $E_\nu \, \mathrm{d}\nu$ passing through the same area in time $\mathrm{d}t$ and solid angle $\mathrm{d}\Omega$. Hence we can write

$$\mathrm{d}E_\nu \, \mathrm{d}\nu = I_\nu(\mathbf{r}, t, \hat{k})(\hat{k} \cdot \hat{n}) \, \mathrm{d}t \, \mathrm{d}\Omega \, \mathrm{d}A \, \mathrm{d}\nu \tag{2}$$

where the quantity $I_\nu(\mathbf{r}, t, \hat{k})$ is called the *specific monochromatic intensity*. If $I_\nu(\mathbf{r}, t, \hat{k})$ is specified for all directions at every point in a certain region of spacetime, then we'd have a complete prescription of the radiation field we intend on studying.

Capitalizing on the blatant similarities with distribution functions, we can evaluate the moments of the monochromatic intensity.

Definition 1.2.1. *Monochromatic mean intensity J_ν*

$$J_\nu = \frac{1}{4\pi} \int_\Omega I_\nu \, d\Omega = \frac{c}{4\pi} U_\nu$$

with U_ν the total energy density of radiation. Note that J_ν is essentially just an average of the monochromatic intensity over all solid angles.

Definition 1.2.2. *Monochromatic flux $\vec{F}_\nu$*

$$\vec{H}_\nu = \frac{1}{4\pi} \int_\Omega I_\nu(\hat{k}) \hat{k} \, d\Omega \implies \frac{1}{4\pi} F_\nu = \vec{H}_\nu \cdot \hat{n}$$

I haven't explicitly proved the last equality, but it shouldn't be hard for you to convince yourself (or prove it yourself) that it is indeed true.

Definition 1.2.3. *Monochromatic radiation pressure p_ν The monochromatic pressure is defined starting from the different directions correlations of the monochromatic intensity*

$$K_\nu^{ij} = \frac{1}{4\pi} \int_\Omega I_\nu(\hat{k}) n^i n^j \, d\Omega$$

The pressure in particular is usually expressed as

$$P_\nu = \frac{1}{c} \int_\Omega I_\nu(\hat{k}) \cos^2\theta \, d\Omega$$

where $\cos^2\theta = (\hat{k} \cdot \hat{n})^2$.

1.3 BLACKBODY RADIATION

Even at an undergraduate level, we're all fairly familiar with *blackbody radiation*. The easiest way to deduce the expression for the energy density of photons in *thermal equilibrium* (STE) inside a cavity is by the means of statistical mechanics.

Remember the Bose-Einstein distribution ($\mu = 0$)

$$n = \frac{1}{\exp(h\nu/kT) - 1}$$

and the phase space density of states (per unit volume)

$$\rho(\nu)\,\mathrm{d}\nu = \frac{4\pi h g \nu^3}{c^3}\,\mathrm{d}\nu$$

from which deducing the expression from internal energy is straightforward. Remembering $g = 2$ is the quantum degeneracy of photons, a simple moltiplication of the previous expressions yields

$$U_\nu\,\mathrm{d}\nu = \frac{8\pi h}{c^3}\frac{\nu^3}{\exp(h\nu/kT) - 1}\,\mathrm{d}\nu$$

Since blackbody radiation is isotropic (it depends only on the absolute temperature T), the definition of mean monochromatic intensity yields

$$\boxed{B_\nu(T) = \frac{2h\nu^3}{c^2}\frac{1}{\exp(h\nu/kT) - 1}} \tag{3}$$

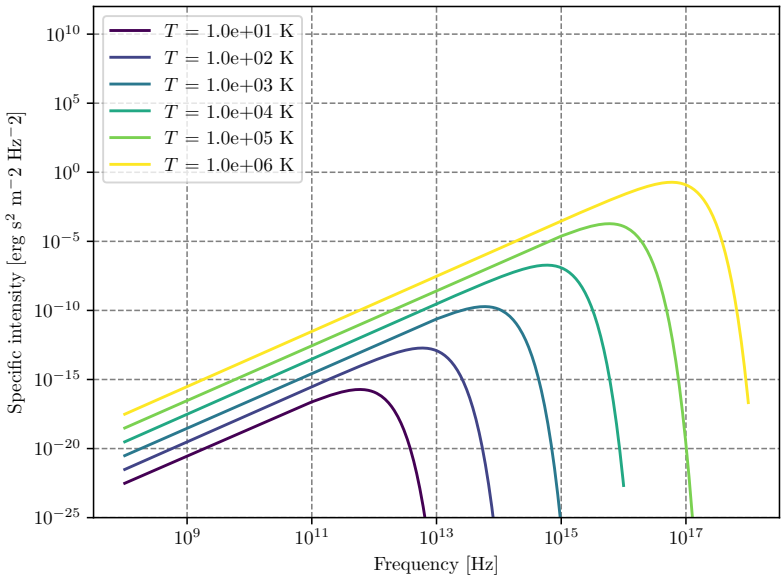

Figure 2: Blackbody frequency spectrum.

It's important to notice that, in principle, such a fundamental result holds only in *strict thermodynamic equilibrium* (STE), but we'll soon see how to generalize this formulation for less "restrictive" environments.

An incredible number of important results descends from (3), and it may be worthwhile to cite at least some of them, starting from Stefan-Boltzmann law. We'll use the following result without proving it

$$\int_0^{+\infty} B_\nu(T)\,\mathrm{d}\nu = \frac{2h}{c^2}\frac{\pi^2}{15}\left(\frac{kT}{h}\right)^4$$

Computing the bolometric flux and the bolometric energy density by integrating over all frequencies using what we've just written down, you find the following

$$U(T) = aT^4 \quad F(T) = \sigma_{SB}T^4$$

Clearly the two constants a and σ_{SB} cannot be independent, and are actually related by the integral we've previously calculated. Using for example[1]

$$F(T) = \pi\int_0^{+\infty} B_\nu(T)\,\mathrm{d}\nu$$

you can easily find out that the *Stefan-Boltzmann constant* is equal to

$$\sigma_{SB} = \frac{2\pi^5k^4}{15c^2h^3}$$

and the relation with a is simply $\sigma_{SB} = ac/4$.

The equation

$$F(T) = \frac{2\pi^5k^4}{15c^2h^3}T^4 \tag{4}$$

is what is usually known as the *Stefan-Boltzmann law*.

Let us now consider two different regimes for eq.3: $h\nu/kT \ll 1$ and $h\nu/kT \gg 1$. The first yields what is commonly known as the Rayleigh-Jeans Law which is, sadly, pretty much relevant only for radioastronomy.

Since

$$\exp\left(\frac{h\nu}{kT}\right) = 1 + \frac{h\nu}{kT} + o\left(\frac{h\nu}{kT}\right)^2$$

the blackbody radiation assumes the much simpler form of

$$B_\nu^{RJ} = \frac{2\nu^2}{c^2}kT \tag{5}$$

Another important results is achieved in the opposite regime, when the exponential term is rather larger than unity

$$B_\nu^{W} = \frac{2h\nu^3}{c^2}\exp\left(-\frac{h\nu}{kT}\right) \tag{6}$$

This expression is known as Wien's Law.

1 The emergent flux from an isotropically emitting surface (such as a blackbody) is $\pi \cdot$ brightness which is none other than the specific intensity.

In the presence of matter, it is not immediately obvious what changes may occur in the specific intensity as we move along a ray path. The aim of this section will be to eviscerate the matter.

Let's consider the following geometric construction

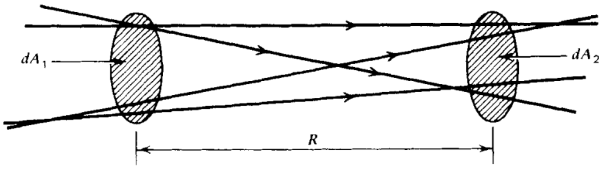

Figure 3: Geometrical construction for ray paths propagating in empty space. Credits: G. Rybicki, A. Lightman.

It won't take a lot of effort to convince yourself that in empty space the monochromatic intensity I_ν is actually conserved. In fact, from simply writing down the definitions and imposing the conservation of energy

$$I_{\nu_2}\,\mathrm{d}A_2\,\mathrm{d}t\,\mathrm{d}\Omega_2\,\mathrm{d}\nu = I_{\nu_1}\,\mathrm{d}A_1\,\mathrm{d}t\,\mathrm{d}\Omega_1\,\mathrm{d}\nu$$

the conclusion follows observing that $\mathrm{d}A_2\,\mathrm{d}\Omega_2 = \mathrm{d}A_1\,\mathrm{d}\Omega_1$.

If we consider an affine parameter of the form $\vec{x} = \vec{x}_0 + \hat{k}s$, we may as well write the previous results in a more familiar fashion

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = 0 \implies (\hat{k}\cdot\nabla)I_\nu = 0 \tag{7}$$

What changes if matter is present along the ray path? Clearly it will no longer be true that $(\hat{k}\cdot\nabla)I_\nu = 0$, but we're not that far off. All that we need is some little work on both terms.

How the right hand side of the equation should change is obvious: It needs to keep track of the "creation" and "destruction" of photons in the considered volume of spacetime.

The left hand side of the equation requires a little more care. Consider infinitesimal time and space displacements along the ray path, respectively $\mathrm{d}t$ and $\mathrm{d}\vec{x}$

$$\Delta E_\nu\,\mathrm{d}\nu = \Big(I_\nu(\vec{x}+\mathrm{d}\vec{x}, t+\mathrm{d}t, \hat{k}) - I_\nu(\vec{x}, t, \hat{k})\Big)\,\mathrm{d}t\,\mathrm{d}\Omega\,\mathrm{d}A\,\mathrm{d}\nu$$

Taking a first order expansion in respect to the affine parameter s along the ray path yields

$$\left(\frac{1}{c}\partial_t I_\nu + \partial_s I_\nu\right)\,\mathrm{d}t\,\mathrm{d}s\,\mathrm{d}\Omega\,\mathrm{d}A\,\mathrm{d}\nu = \text{photon addition} - \text{photon removal}$$

This equation is a generalization of eq.7 for non-stationary radiative transport and in the presence of matter. It's about time we get to know what "lives" in the right hand side of the equation.

1.4.1 Monochromatic emission coefficient

For the moment, we'll define the *spontaneous* monochromatic emission coefficient j_ν as

$$\mathrm{d}E_\nu \,\mathrm{d}\nu = j_\nu \,\mathrm{d}V \,\mathrm{d}t \,\mathrm{d}\Omega \,\mathrm{d}\nu \tag{8}$$

which in general has a non-zero dependence on the emission direction. Sometimes the spontaneous emission coefficient is defined by the *emissivity* ϵ_ν (**please note** that often the two names are used almost interchangeably), which is the energy emitted spontaneously per unit frequency per unit time per unit mass

$$j_\nu = \frac{\epsilon_\nu \rho}{4\pi}$$

where ρ is the mass density of the emitting medium.

If we perform the decomposition $\mathrm{d}V = \mathrm{d}A \,\mathrm{d}s$, the contribution of spontaneous emission to the specific intensity is

$$\mathrm{d}I_\nu = j_\nu \,\mathrm{d}s$$

1.4.2 Absorption coefficient

Similarly, we can consider the energy that is absorbed from the radiation when passing through a medium. There exists various definitions; I'll use the one we gave in class and that is incidentally the one used in [7] and [20] as well.

We define the *absorption coefficient* α_ν through the following expression

$$\mathrm{d}I_\nu = -\alpha_\nu I_\nu \,\mathrm{d}s \tag{9}$$

If we use a microscopic model, then the absorption coefficient can be understood as particles with numeric density n presenting an effective absorbing area, the *cross section* σ_ν. The coefficient α_ν can thus be rewritten in terms of

$$\alpha_\nu = n\sigma_\nu = \rho\kappa_\nu$$

where κ_ν is called the mass absorption coefficient or the *mass-weighed opacity coefficient*.

Note that in eq.9, we consider "absorption" to include both "true absorption" and stimulated emission, because both are proportional to the intensity of the incoming beam. Depending on the entity of the contribution, the α_ν coefficient may be positive or even negative, giving raise to curious phenomena.

Making full use of what we've just defined, we can finally present the celebrated *equation of radiative transfer* (although in the notable absence of scattering)

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = -\alpha_\nu I_\nu + j_\nu \tag{10}$$

which is actually fairly easy to solve when one of the two coefficients vanishes.

Emission only

We set $\alpha_\nu = 0$ and the equation may be solved by direct integration

$$I_\nu(s) = I_\nu(s_0) + \int_{s_0}^{s} j_\nu(s')\,\mathrm{d}s'$$

the result is not that interesting per se.

Absorption only

This time we set $j_\nu = 0$. The equation is easily solved this time as well

$$I_\nu(s) = I_\nu(s_0)\exp\left(-\int_{s_0}^{s}\alpha_\nu(s')\,\mathrm{d}s'\right)$$

In this case, it's rather common to write down the equation in terms of a new variable, the *optical depth* τ_ν

$$\mathrm{d}\tau_\nu = \alpha_\nu\,\mathrm{d}s \tag{11}$$

Given this definition we'll say that if

- $\tau_\nu \gg 1$: the medium is *optically thick or opaque*
- $\tau_\nu \ll 1$: the medium is *optically thin or transparent*

This has some crucial implications we'll be going through in a moment.

In the stationary limit, the equation of radiative transport may be written as

$$(\hat{k}\cdot\nabla)I_\nu(\hat{k},\vec{x}) = j_\nu(\vec{x}) - \alpha_\nu(\vec{x})I_\nu(\hat{k},\vec{x})$$

In terms of the *source function* $S_\nu = j_\nu/\alpha_\nu$ it becomes

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}\tau_\nu} = -I_\nu + S_\nu \tag{12}$$

which can be integrated to yield the formal solution

$$I_\nu(\hat{k},\tau_\nu) = I_\nu(\tau_{\nu,0})\exp(-\tau_\nu) + \int_{\tau_{\nu,0}}^{\tau_\nu} d\tau'_\nu\, S_\nu \exp(-(\tau_\nu - \tau'_\nu))$$

Assume for the moment that the matter through which radiation is passing has constant properties and has no background source. Then the source function S_ν is constant and the formal equation becomes

$$I_\nu = S_\nu(1 - e^{-\tau_\nu})$$

If the medium and is optically thin, then the equation is reduced to

$$I_\nu = S_\nu\tau_\nu = j_\nu L \tag{13}$$

by taking the Taylor expansion of the exponential term and calling L some typical length of the medium.

If, on the other hand, the medium is optically thick, we can neglect the exponential $e^{-\tau_\nu}$ to obtain

$$I_\nu = S_\nu \tag{14}$$

The most notable implication of eq.14 is if we consider the specific intensity coming out of a small hole on a box kept in thermodynamic equilibrium. We know that what's going to come out of there is the blackbody radiation

$$I_\nu = B_\nu(T)$$

but what if we were to put an optically thick object just behind the hole?

If the object is in thermodynamic equilibrium with the surroundings (and it *will* be, given an appropriate amount of time), then the radiation coming out of the hole will still be blackbody radiation. But eq.14 tells us that the source function will tend to be equal to the specific intensity, hence

$$S_\nu = B_\nu(T) \tag{15}$$

which actually puts a constraint on the possible values of the emission coefficient in terms of the absorption coefficient. This is exactly what is expressed in Kirchhoff's law

$$j_\nu = \alpha_\nu B_\nu \tag{16}$$

Let us briefly consider what we have just derived. Matter often tends to emit and absorb at specific frequencies corresponding to what are commonly called *spectral lines*. We would expect then both j_ν and α_ν to have peaks (or depression) around these lines. But Kirchhoff's law forces their ratio to be equal to a smooth blackbody profile.

Thus we can expect to observe two very different scenarios if the medium is optically thin rather than optically thick. In the former, the radiation emerging from the medium is essentially determined by its emission coefficient; since j_ν is expected to present peaks, so will the radiation spectrum, which will be appear in spectral lines, as shown in Fig.4 and Fig.5.

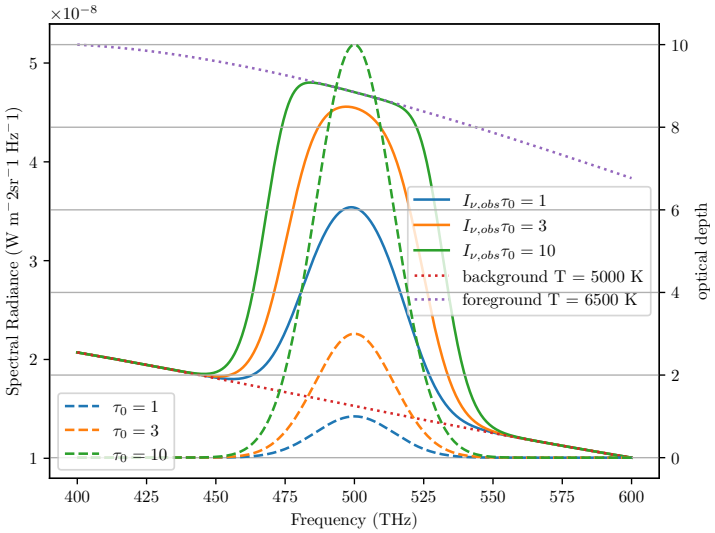

Figure 4: An example of emission features formation for different temperatures and different values of τ. Credits: Prof. Walter del Pozzo.

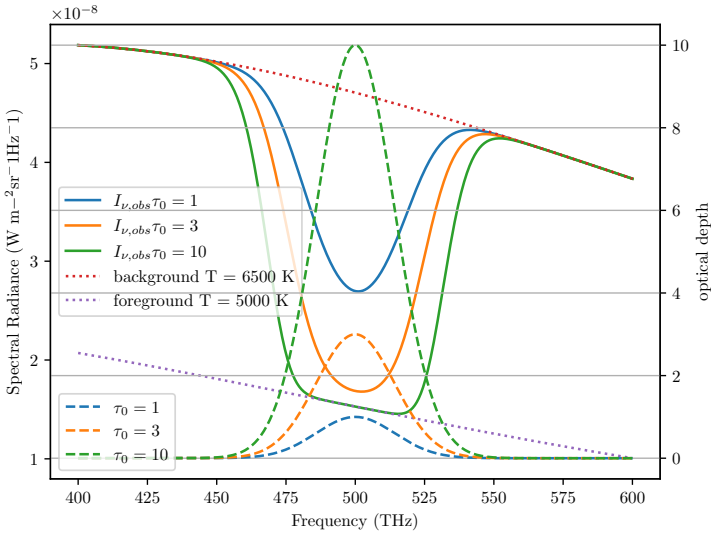

Figure 5: An example of absorption features formation for different temperatures and different values of τ. Credits: Prof. Walter del Pozzo.

On the other hand, the intensity coming out of an optically thick body is its source function, which must be equal to the blackbody function. Hence we expect the medium to emit in a continuum, like a blackbody.

All throughout this description, we've been assuming the medium to have constant properties, which has the perk of being a good approximation for many objects of interest, but still turns out to be a really poor one for many other objects. Stars, for example.

Ingenuously, we may expect stars to emit radiation like blackbodies, but they're not. Actually, stars present absorption lines, many, even, depending on the class of star. What we cannot assume in stars is them having constant properties, starting from temperature.

In fact, we could take a guess and claim that stars are in *strict* thermodynamic equilibrium. It would be a very bad guess indeed.

1.5.1 Local Thermodynamic Equilibrium (LTE)

Let's be honest: In a realistic situation, we *rarely* have strict thermodynamic equilibrium. If a body is in thermodynamic equilibrium, we can assume a number of important physical principles to hold, like the Maxwellian distribution

$$\mathrm{d}n_v = 4\pi n \left(\frac{m}{2\pi kT}\right)^{3/2} v^2 \exp\left(-\frac{mv^2}{2kT}\right) \mathrm{d}v \tag{17}$$

where n is the total number of particles per unit volume and m is the mass of each particle. Similarly, we can expect certain laws to hold, like Boltzmann's law for occupation numbers

$$\frac{n_E}{n_0} = \frac{g_E}{g_0} \exp\left(-\frac{E - E_0}{kT}\right) \tag{18}$$

and Saha's equation

$$\frac{N_{j+1}n_e}{N_j} = 2\frac{Z_{j+1}(T)}{Z_j(T)}\left(\frac{2\pi mkT}{h^2}\right)^{3/2}\exp\left(-\frac{\chi_{j,j+1}}{kT}\right) \tag{19}$$

where n_e is the density of electrons and $\chi_{j,j+1}$ is the ionization potential. Saha's equation in particular is expected to be crucial in interpreting the effect that ionization has on the emission/absorption spectrum.

The proverbial "one-million-dollar-question" then is: When can we expect a system to be in thermodynamic equilibrium and when can we expect the previous principles to hold?

Even if the system initially does not obey the, say, Maxwellian distribution, it will eventually relax to it after undergoing some *collisions*.

Collisions are crucial in establishing thermodynamic equilibrium.

When collisions are frequent, the mean free path of particles will be small, and particles will interact more effectively. When this happens, we can expect the principles aforementioned to hold. Since we're physicists, vague sentences like "*the mean free path of particles will be small*" are destined to elicit a deep sense of unease and distress. How small does the free path have to be? One meter? Two micrometers? Below the Planck lengthscale?

When we've defined the absorption coefficient α_ν, the sharpest among my four readers total may have noticed that α_ν has the dimension of the inverse of a length. It is safe to assume that α_ν^{-1} may define some distance over which a significant fraction of the radiation would get absorbed by matter.

Such a "mean-distance" is defined in a homogeneous medium as

$$< \tau_\nu >= \alpha_\nu l_\nu = 1$$

Thus, if l_ν is sufficiently small such that the temperature can be taken as a constant over such distance, we can safely say that the useful relations we have defined earlier still hold, although only locally.

In such a fortunate scenario, known as *Local Thermodynamic Equilibrium* (LTE), all the important laws requiring thermodynamic equilibrium are expected to hold, provided that we use the local temperature $T(\vec{x})$.

In the interiors of stars, for example, LTE will prove to be a very good approximation, that will get progressively worse as we approach the "surface" of the star.

1.6 PARALLEL PLANE APPROXIMATION

One useful approximation that may be worthwhile to dedicate some of our time to is the *plane parallel atmosphere*, that will allow us to obtain notable results for describing how radiation travels through, say, the inner regions of the stellar atmosphere.

In the following, we're going to neglect the curvature of the stellar atmosphere and assume the various thermodynamic quantities to be constant over horizontal planes.

Using Fig.6 as a reference, we see that

$$\mathrm{d}s = \frac{\mathrm{d}z}{\cos\theta} = \frac{\mathrm{d}z}{\mu}$$

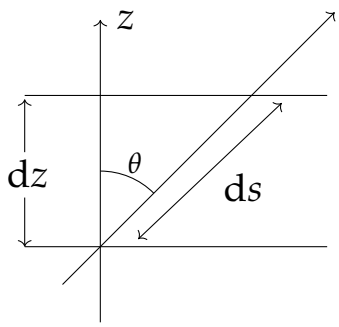

Figure 6: A ray path through a plane parallel atmosphere.

where we used the customary notation in astrophysics $\mu = \cos\theta$.

We shall consider a scattering free, stationary situation for the equation of radiative transport (10). Hence, due to planar symmetry, we expect the specific intensity to depend only on z and μ. For the sake of the current discussion, we perform a slight modification to the definition of optical depth, so that

$$\mathrm{d}\tau_\nu = -\alpha_\nu \,\mathrm{d}z$$

This way the equation of radiative transfer may be cast in the following form

$$\mu \partial_{\tau_\nu} I_\nu(\tau_\nu, \mu) = I_\nu - S_\nu$$

which has a formal solution easily computed

$$I_\nu \exp\left(-\frac{t_\nu}{\mu}\right)|_{\tau_{\nu,0}}^{\tau_\nu} = -\int_{\tau_{\nu,0}}^{\tau_\nu} \frac{S_\nu}{\mu} \exp\left(-\frac{t_\nu}{\mu}\right) \mathrm{d}t_\nu \tag{20}$$

This is customarily solved considering two distinct intervals for μ: (I) $\mu \in [0,1]$ and (II) $\mu \in [-1,0]$. In case (I) we can assume the ray path to begin from a great depth inside the star, so that $\tau_{\nu,0} \to \infty$, while in case (II) we assume the ray to receive contributions beginning from the top of the atmosphere, where $\tau_{\nu,0} \approx 0$. For case (II), we're also assuming no radiation to be coming from *outside the star*[2].

Now we can assume LTE throughout the stellar atmosphere so that eq.16 is verified. The source function at some optical depth shall then be equal to $B_\nu(T(\tau_\nu))$. For the source function at a nearby optical depth we can simply compute a Taylor expansion around the optical depth τ_ν

$$S(t_\nu) = B_\nu(\tau_\nu) - (t_\nu - \tau_\nu)\frac{\mathrm{d}B_\nu}{\mathrm{d}\tau_\nu} + \mathrm{o}(t_\nu^2)$$

We can use this to solve eq.(20), finding for both positive and negative values of μ a very important equation

$$I_\nu(\tau_\nu, \mu) = B_\nu(\tau_\nu) + \mu \frac{\mathrm{d}B_\nu}{\mathrm{d}\tau_\nu} \tag{21}$$

2 Please note that this condition may be not valid at all in close binary systems.

provided the point considered is sufficiently inside the atmosphere so that $\tau_\nu \gg 1$[3]. Using this simple result, we can compute the three momenta of the equation of transport

$$U_\nu = \frac{4\pi}{c} B_\nu(\tau_\nu) \tag{22}$$

$$F_\nu = \frac{4\pi}{3} \frac{\mathrm{d}B_\nu}{\mathrm{d}\tau_\nu} \tag{23}$$

$$P_\nu = \frac{4\pi}{3c} B_\nu(\tau_\nu) \tag{24}$$

1.6.1 The Grey Atmosphere

If we consider the absorption coefficient α_ν constant over all frequencies, then the atmosphere is called a "grey atmosphere". This implies that the value of the optical depth at some physical depth is constant for all frequencies. Under this assumption, we could solve

$$\mu \frac{\partial I}{\partial \tau} = I - S$$

I'll skip the explicit calculation (which is actually fairly easy for once) and present just the final result. Two more assumptions are to be made, however: The first is to assume *radiative equilibrium*, which roughly translates into requiring that there are no sources nor sinks of energy in the atmosphere, thus $\partial_\tau F = 0$; as a further semplification, we assume the *Eddington approximation* to hold everywhere in the atmosphere[4], so that

$$P = \frac{1}{3} U$$

It should be evident that this last equation is verified automatically in presence of an isotropic source of radiation, also in its frequency-dependent form.

We then come to the following conclusion

$$I_{obs}(\tau = 0, \mu) = \frac{3F}{4\pi} \left(\mu + \frac{2}{3} \right) \tag{25}$$

from which we deduce the equation for the *limb darkening*

$$\frac{I(0,\mu)}{I(0,1)} = \frac{3}{5} \left(\mu + \frac{2}{3} \right)$$

which roughly translates into saying that the radiation that we observe at the surface is the one equivalent at a source function S evaluated at $\tau = 2/3$. This is known as the *Eddington-Barbier estimation*.

A somewhat more general way to solve the problem is assuming the following functional relation for the specific intensity

$$I_\nu(\tau, \mu) = a_\nu(\tau_\nu) + b_\nu(\tau_\nu)\mu$$

3 You can see [7], §2.4.1 for more detailed calculations.

4 I find most intriguing that Eddington's approximation essentially assumes that $T^\mu{}_\mu = 0$, with $T^{\mu\nu}$ the electromagnetic energy-momentum tensor if we are to treat electromagnetic radiation as a perfect fluid.

and compute the three momenta proper

$$J_\nu = \frac{1}{2}\int_{-1}^{+1} I_\nu \,\mathrm{d}\mu = a_\nu \tag{26}$$

$$H_\nu = \frac{1}{2}\int_{-1}^{+1} I_\nu \mu \,\mathrm{d}\mu = \frac{b_\nu}{3} \tag{27}$$

$$K_\nu = \frac{1}{2}\int_{-1}^{+1} I_\nu \mu^2 \,\mathrm{d}\mu = \frac{a_\nu}{3} \tag{28}$$

Now we assume a stronger version of the *Eddington approximation* $K_\nu = J_\nu/3$ so that the two following expressions can be written

$$\frac{\partial H_\nu}{\partial \tau_\nu} = J_\nu - S_\nu \tag{29}$$

$$\frac{\partial K_\nu}{\partial \tau_\nu} = H_\nu = \frac{1}{3}\frac{\partial J_\nu}{\partial \tau_\nu} \tag{30}$$

The mixing of the two gives us a second order PDE that it's still (approximately) valid even in the outer regions of the stellar atmosphere.

1.7 RADIATIVE DIFFUSION APPROXIMATION

This section would probably be clearer if you take a brief detour to Chapter 3 to build some groundwork for scattering processes, but it still suits better the main subject of this chapter.

For the sake of coherence (and personal laziness) I'm going to talk about the radiative diffusion approximation right here, also because it makes use of the plane parallel approximation we just went through.

In Chapter 3 we will use random walk arguments to show that S_ν approaches B_ν at large *effective optical depths* in a homogeneous medium. Real media are seldom homogeneous, but often, as in the interiors of stars, there is a high degree of local homogeneity.

The equation of radiative transport in presence of scattering (45) may be cast in a slightly different form

$$I_\nu = S_\nu - \frac{\mu}{\alpha_\nu + \sigma_\nu}\partial_z I_\nu$$

We shall then assume that over a distance l_* (the thermalization length) I_ν is constant and at zero-th order is $I_\nu^{(0)} = S_\nu^{(0)} = B_\nu$. Plugging this in the equation of radiative transfer gives us $I_\nu^{(1)}$ by a simple iterative procedure

$$I_\nu^{(1)} = B_\nu - \frac{\mu}{\alpha_\nu + \sigma_\nu}\partial_z B_\nu \tag{31}$$

With the simple redefining $\mathrm{d}\tau_\nu = -(\alpha_\nu + \sigma_\nu)\,\mathrm{d}z$, we can put eq.(31) in a form functionally equal to (21).

Let us now compute the flux F_ν using the above form for the intensity

$$F_\nu(z) = 2\pi\int_{-1}^{+1} I_\nu^{(1)}\mu \,\mathrm{d}\mu = -\frac{4\pi}{3}\frac{\partial_z B_\nu}{\alpha_\nu + \sigma_\nu} = -\frac{4\pi}{3}\frac{\partial_T B_\nu}{\alpha_\nu + \sigma_\nu}\partial_z T$$

Recalling the result

$$\partial_T \int_0^{+\infty} B_\nu \, \mathrm{d}\nu = \frac{4\sigma_{SB} T^3}{\pi}$$

we can define a *mean absorption coefficient* using the *Rosseland approximation for radiative diffusion*

$$\frac{1}{\alpha_R} := \frac{\int_0^{+\infty} \frac{1}{\alpha_\nu + \sigma_\nu} \partial_T B_\nu \, \mathrm{d}\nu}{\int_0^{+\infty} \partial_T B_\nu \, \mathrm{d}\nu} \tag{32}$$

If we integrate the monochromatic flux over the frequencies and make use of the Rosseland mean, we find a useful expression used in stellar structure models

$$F(z) = -\frac{16\sigma_{SB} T^3}{3\alpha_R} \partial_z T \tag{33}$$

2 THE EINSTEIN COEFFICIENTS

2.1 INTRODUCTION

Kirchhoff's law (eq.16), which relates the (spontaneous) emission and the absorption coefficient, seems to imply some underlying microscopic connection between the two phenomena.

As was first discovered by Einstein, that is exactly the case. Let's consider a two level atom interacting with radiation. As depicted in Fig.7, we'll con-

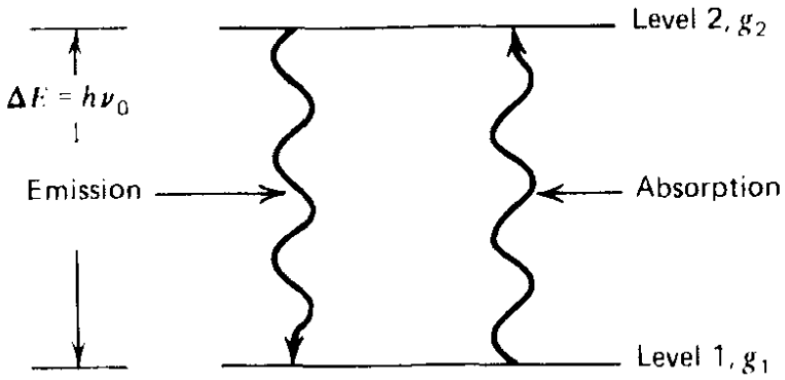

Figure 7: Photon emission and absorption in a two levels atom.
Credits: G. Rybicki, A. Lightman.

sider two discrete energy levels: the lower with energy E and degeneracy g_1, while the upper level has energy equal to $E + h\nu_0$ and degeneracy g_2. Transition between the two levels is possible only through absorption ($1 \rightarrow 2$) or emission ($2 \rightarrow 1$) of photons of energy $h\nu_0$.

Three processes can be thus identified: *spontaneous emission*, *stimulated emission* and *absorption*.

Spontaneous Emission

The process we'll refer to as *spontaneous emission* occurs when the system transitions from the excited state 2 to the lower level 1 through the emission of a photon.

Spontaneous emission can occur even in the **absence of radiation fields** and can be assumed to be **isotropic**. The transition probability for spontaneous emission is defined through the *Einstein A-coefficient* A_{21}, which has units $[A_{21}] = \mathrm{s}^{-1}$.

Absorption

In the presence of a radiation field with the right energy, the system can absorb a photon to transition from state 1 to a higher energy state. Assuming that the radiation field cannot self-interact, we expect the probability per unit time to be proportional to the density of photons (or to the mean intensity J_ν) at frequency ν_0.

However, the energy difference between the two states is not infinitely sharp, but more like a smoother curve we'll call the *line profile function* $\phi(\nu)$,

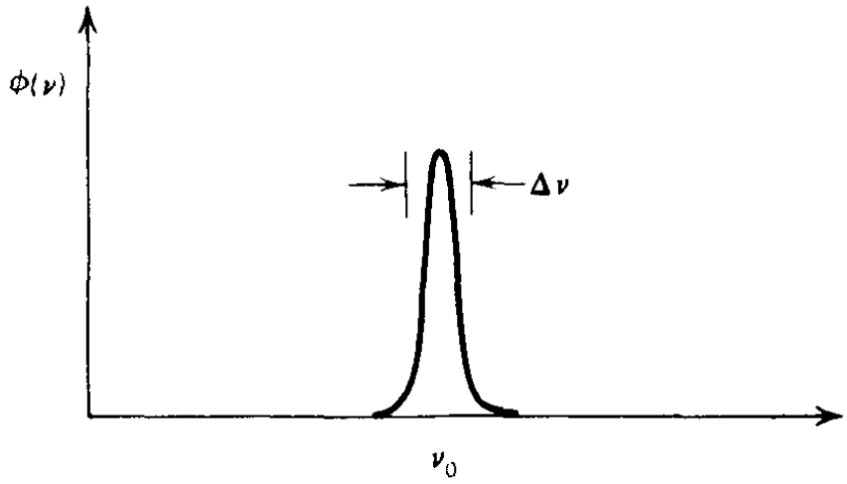

Figure 8: Line profile for a two levels atom.
Credits: G. Rybicki, A. Lightman.

which is conveniently peaked at $\nu = \nu_0$ and is correctly normalized

$$\int_0^{+\infty} \phi(\nu)\, \mathrm{d}\nu = 1$$

Not bothering for the moment with the underlying physical mechanisms that concur in determining the line profile, we'll propose the following definition for the transition probability for absorption: $B_{12}\bar{J}$, where $\bar{J}$ is

$$\bar{J} = \int_0^{+\infty} J_\nu \phi(\nu)\, \mathrm{d}\nu$$

and B_{12} is the *Einstein B-coefficient*.

Stimulated Emission

As anticipated, there exists yet another way for a system to emit a photon, but this time requiring the presence of a radiation field. Taken $\bar{J}$ has the same meaning as before, we'll define the transition probability per unit time for stimulated emission as $B_{21}\bar{J}$.

If we were to assume that the mean intensity J_ν changes slowly over the width $\Delta\nu$ of the line profile, we could in principle approximate the line profile as a δ-function peaked at ν_0.

Some books use the energy density U_ν in the definitions. It's pretty much the same since they carry the same information, but be aware that the two definitions will differ of a $c/4\pi$ factor.

2.2 RELATIONS BETWEEN THE EINSTEIN COEFFICIENTS

It would be most useful to know if there were some kind of relations between the three different coefficients, and if there were some way of relating them to what we've called α_ν and j_ν when dealing the radiative transport.

If this weren't the case, this section would have no reason to exist, so it's safe for you to assume that such relations do in fact exist. To do so, however, we have to invoke the *quantum theorem of detailed balance* (see, for example, [11]), which roughly says

> If states a and b of a system have the same energy, then if P_{ab} is the probability per unit time of a transition from a to b, and P_{ba} from b to a,
>
> $$P_{ab} = P_{ba}$$

For an atom of matter to be in equilibrium with a radiation field, then, the probability of said atom being in the ground state and absorbing a photon of a given frequency must be equal to the probability that it is in the excited state and emits the photon.

If we assume to be in a steady-state condition, the rate of transition from level 1 to level 2 has to be equal to the rate of transition from level 2 to level 1, or, more generally

$$n_i \sum_j R_{ij} - \sum_j n_j R_{ji} = 0$$

If we put in the transition rates we've written down earlier, we get

$$n_1 B_{12} \bar{J} = n_2 A_{21} + n_2 B_{21} \bar{J}$$

which means that the number of transitions per unit time per unit volume out of state 1 must be equal to the number of transitions per unit time per unit volume into state 1.

Solving for $\bar{J}$

$$\bar{J} = \frac{n_2 A_{21}}{n_1 B_{12} - n_2 B_{21}} = \frac{A_{21}/B_{21}}{(n_1/n_2)\cdot(B_{12}/B_{21}) - 1} \tag{34}$$

In thermodynamic equilibrium we can use eq.18 to relate the occupation numbers

$$\frac{n_1}{n_2} = \frac{g_1}{g_2} \frac{\exp(-E/kT)}{\exp[-(E+h\nu_0)/kT]} = \frac{g_1}{g_2} \exp(h\nu_0/kT)$$

but in thermodynamic equilibrium we know that $J_\nu = B_\nu$, so if B_ν varies slowly on the scale of $\Delta\nu$ we can assume $\bar{J} \approx B_\nu(\nu_0)$.

That means the following relations must simultaneously hold

$$g_1 B_{12} = g_2 B_{21} \tag{35}$$

$$A_{21} = \frac{2h\nu^3}{c^2} B_{21} \tag{36}$$

which are known as *Einstein relations*. They connect atomic properties A_{21}, B_{21} and B_{12} and have no reference to the temperature T (unlike Kirchhoff's law).

Although we have derived eq.36 assuming thermodynamic equilibrium, those two relations **must be always valid** whether or not the atoms are in thermodynamic equilibrium.

2.2.1 Absorption and Emission coefficients in terms of Einstein coefficients

To obtain the emission coefficient j_ν we have to make a crucial assumption about the frequency distribution of the emitted radiation: This emission is distributed with the same line profile $\phi(\nu)$ that describes absorption, which is often verified in astrophysics (good for us).

By the definition (8), we already know the amount of energy emitted in volume $\mathrm{d}V$, solid angle $\mathrm{d}\Omega$, frequency $\mathrm{d}\nu$, and time $\mathrm{d}t$. Now, since each atom contributes with $h\nu_0$ to spontaneous emission over a 4π solid angle, we can express the emission coefficient as

$$j_\nu = \frac{h\nu_0}{4\pi} n_2 A_{21} \phi(\nu) \tag{37}$$

Similarly, you can show that the energy absorbed from radiation in frequency range $\mathrm{d}\nu$, solid angle $\mathrm{d}\Omega$, time $\mathrm{d}t$ and volume $\mathrm{d}V$ is

$$\frac{h\nu_0}{4\pi} n_1 B_{12} I_\nu \phi(\nu)\, \mathrm{d}\nu\, \mathrm{d}\Omega\, \mathrm{d}t\, \mathrm{d}V$$

From here follows immediately the *true absorption coefficient*

$$\alpha_\nu = \frac{h\nu_0}{4\pi} n_1 B_{12} \phi(\nu) \tag{38}$$

Note that since stimulated emission is proportional to the specific intensity I_ν and only affects the photons along the given beam, it's not a big stretch to claim that it behaves in a way that is akin to true absorption. We shall then define the *absorption coefficient, corrected for stimulated emission* as

$$\alpha_\nu = \frac{h\nu_0}{4\pi} \phi(\nu)(n_1 B_{12} - n_2 B_{21}) \tag{39}$$

in which we're regarding stimulated emission like a *negative absorption*.

In terms of the newly defined emission and absorption coefficients, the equation of radiative transfer takes a new look

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = -\frac{h\nu_0}{4\pi} \phi(\nu)(n_1 B_{12} - n_2 B_{21}) + \frac{h\nu_0}{4\pi} n_2 A_{21} \phi(\nu) \tag{40}$$

Computing the ratio j_ν / α_ν with the new definitions holds

$$S_\nu = \frac{n_2 A_{21}}{n_1 B_{12} - n_2 B_{21}} = \frac{2h\nu^3}{c^2} \left(\frac{g_2 n_1}{g_1 n_2} - 1 \right)^{-1} \tag{41}$$

where the last equality is obtained just by substitution of the Einstein relations (36). Note that eq.(41) is a generalized Kirchhoff's law. Let's now consider three interesting cases sprouting from this last equality.

Thermal Emission (LTE)

If the matter is in local thermal equilibrium (LTE) with itself (but not necessarily with the radiation), eq.18 must hold locally. In this case, we correctly retrieve

$$S_\nu = B_\nu$$

Nonthermal Emission

We can't assume eq.18 to hold, and particles do not obey the Maxwellian distribution or any of the fancy properties we'd like them to.

Inverted Populations: MASERS

Let's consider the definition of the absorption coefficient corrected for stimulated emission (39). Whether the coefficient is positive or negative depends on the term within parentheses, which can be restated using Einstein's relations

$$\frac{n_1 g_2}{n_2 g_1} - 1 > 0$$

so that when this relation is satisfied, even out of thermal equilibrium, we say that the system has *normal populations*. However, it is possible to put enough atoms in the upper state so that we achieve what is called an *inverted population*

$$\frac{n_1}{g_1} < \frac{n_2}{g_2}$$

which is quite the odd configuration, if you think about it. We're basically saying that a higher energy level is more densely populated than a lower energy one, which, by means of eq.18, corresponds to stating that the temperature T is negative, since E is assumed to be non-negative.

And no, T is not in Celsius degrees.

In such a scenario, the absorption coefficient is negative; rather than being damped to extintion, the intensity of the radiation ray *increases* when passing through matter. Such a system is called MASER (Microwave Amplification by Stimulated Emission of Radiation). For visible light you'd get what we usually call "laser".

The funny thing is that we actually observe this phenomenon occurring spontaneously in Nature, like for example in some molecular clouds formed in the ISM. Some sources in specific molecular lines (like the OH lines) were found to be *abnormally high*. Let me quantify how much "abnormal" we're talking about.

Your typical molecular cloud has a density a bit shy of 10^9 particles per m^3 and temperature in the range of $10 - 30\,K$, so kinda chilly.

If those "abnormal" sources were assumed to be optically thick in the spectral lines and the specific intensity was equated to B_ν, you'd probably roll down you chair reading on your computer that those sources should have temperatures as high as $10^9\,K$.

Not so chilly anymore, huh?

The favored explanation for this weird phenomenon (and most likely the correct one) makes use exacly of what we've just shown: Maser action.

If you want to read something (slightly) more rigorous, you can look at [7], §6.6.3.

For those among you that, like me, were particularly invested in TV shows like *X-Files* and similar, may be interested to know that one of the most likely explanations for the so-called Wow! Signal[1] is in fact attributed to MASER action on a Hydrogen cloud.

2.3 HYDROGENOID ATOMS

Fairly interesting objects to study are *hydrogenoid atoms*, that is atoms with Z protons but that had ended up (no matter how) with only one e^-. For this type of atoms, Quantum Mechanics predicts energy levels distributed as such

$$E_n = -\frac{Z^2\,\mathrm{Ry}}{n^2} \quad \mathrm{Ry} = \frac{m_e e^4}{2\hbar^2} \approx 13.6\,\mathrm{eV}$$

If we were to introduce relativistic corrections through perturbation theory, the expression would get a bit more complicated

$$E(n,\,j) = -\frac{Z^2\,\mathrm{Ry}}{n^2}\left(1+\alpha^2\left[\frac{n}{j+1/2}-3/4\right]\right)$$

introducing *fine splitting* of the energy levels, that now depend on the value of the angular momentum. It is probably worth pointing out that the energy levels are still degenerate in the magnetic quantum number m. Given a problem with spherical symmetry, it would be beneath us to claim that a scalar perturbation could be able to completely solve the degeneracy[2]. Degeneracy is solved completely only if we introduce in the system a preferred direction[3].

Depending on the value of the orbital angular momentum eigenvalue l we can assign a name to distinguish them:

- $l = 0 \rightarrow s$-orbitals
- $l = 1 \rightarrow p$-orbitals
- $l = 2 \rightarrow d$-orbitals
- $l = 3 \rightarrow f$-orbitals

This way, transitions may be represented in a *Grotrian diagram* (Fig.9), where the energy of a given level is plotted in function of the l eigenvalue, through which selection rules are made most evident.

What happens if we are to consider atoms with more than one electron? Will our model still work? Nope, not a chance.

On top of getting stupidly difficult to evaluate even the non relativistic limit, there are *many more ways* to couple interactions, resulting in more and more sources of opacity. It is actually elements with $Z > 1$ and $\#e^- > 1$ that make the greatest contributions to opacity, especially *metals* ($Z \geqslant 3$).

1 This was brought to my attention by a fellow colleague of the Aerospace Engineering department, and I thought it was well-worth sharing it with you. Thanks, Monta.

2 The system is still invariant under rotations.

3 cfr. *Stark-Lo Surdo Effect* and the *Zeeman effect*.

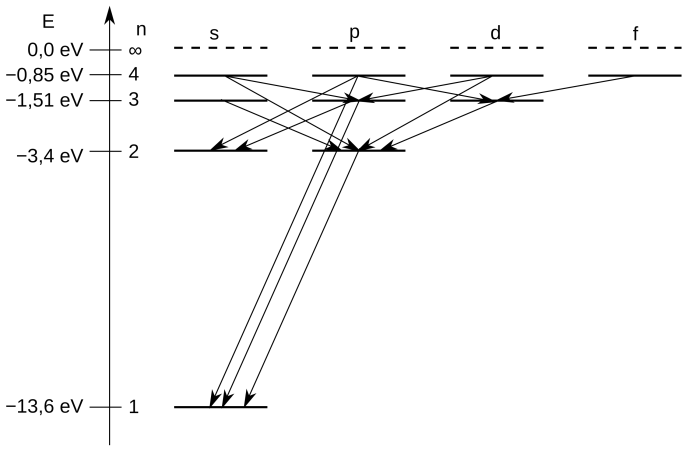

Figure 9: Example of Grotrian Diagram for Hydrogen. Credits: Wikipedia.

Sometimes it is useful to express the electron configuration by means of the *Russel-Saunders notation*. Said S, L, J respectively the spin, the orbital angular momentum and the total angular momentum eigenvalues, we can express a given electron configuration in such a fashion

$$^{2S+1}L_J$$

where L is usually written in spectroscopic notation, that is the symbol associated to the orbital with that L value.

2.3.1 Hyperfine Transition

What we'll be most interested in considering, however, is the hyperfine splitting that arises if we consider the coupling between the spin of the nucleus and the electron.

If we were to consider such interaction, we'd find a hyperfine splitting of the lowest energy state. We'll be interested in the energy of the radiation coming from an electron spin flip: The system will transition from a state with total spin $F = 1$ to a one with $F = 0$, resulting in the emission of radiation (Fig.10). Ignoring higher order corrections to the Hamiltonian, the radiation resulting from the spin flip will have a wavelength

$$\lambda_{hf} = \frac{3\pi\hbar^5 m_p c^3}{g_e g_p m_e^2 e^8} \approx 21.106114054160(30)\,\text{cm} \tag{42}$$

where the expression is in CGS units and g_e, g_p are respectively the electron and the proton spin g-factors (≈ 2 for the electron, ≈ 5.59 for the proton).

Unfortunately, such a transition cannot be observed in lab-experiments. The extimated half-life time of the transition from the Einstein A coefficient[4] is of the order of a few Myrs. Which means that unless you have *a lot* of time to spare to fight against death, you'll probably never record even a single one of these transitions. And that is ignoring the possibility of Hydrogen getting collisionally de-excited, which is unfathomably more probable.

4 Remember that A has the dimensions of the inverse of a time, so $\tau_{HL} \approx A^{-1}$.

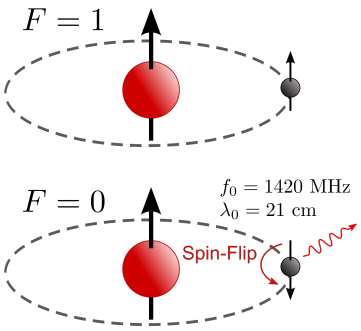

Figure 10: Schematic representation of the spin flip and the resulting emission.
Credits: Wikipedia.

This, however, has a strikingly beautiful consequence: Although "rare", we should be able to detect the 21-cm emission line if we were to look somewhere where the transition may have happened millions of years ago.

If you were to steal a radiotelescope[5] and point it towards the Galactic plane, you'll see a really sharp line in proximity to the 21-cm line ($\sim$ 1420 MHz), as shown in Fig.11.

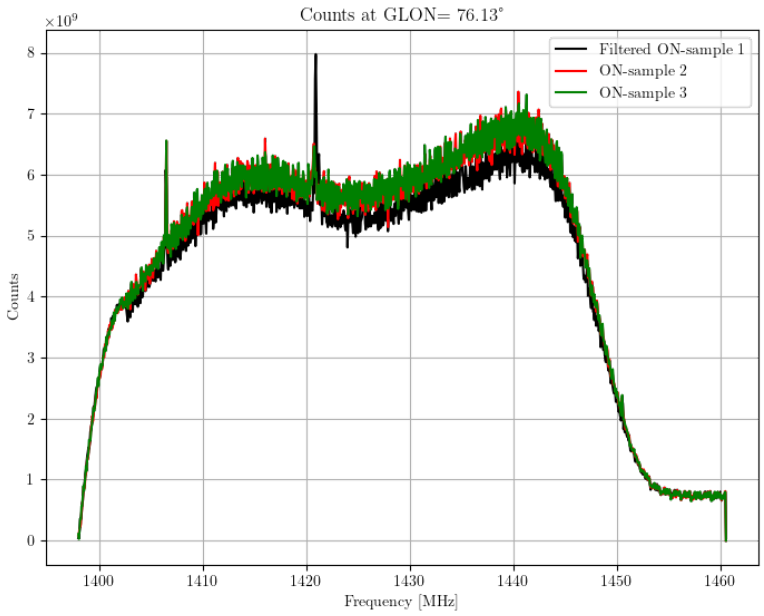

Figure 11: Sampling of the 21-cm emission line with the new Spider500 Radio telescope of the Physics Department of the University of Pisa.
Note how the sharp, black emission line is not centered on 1420 MHz due to Doppler shift (see next section).
Credits: AMLab, Group 8, a.y. 2025/2026.

5 Do not recommend. They get awfully big if you want to see something cool.

2.4 LINE BROADENING

It would be most dumb for us to assume that energy levels, or the lines connecting them, are infinitely sharp. Even a quick look at Heisenberg's uncertainty principle should convince you otherwise.

When we defined the Einstein's coefficients, we had to introduce the line profile $\phi(\nu)$ to account for the non-zero width of the line, and it's about time we consider some of the phenomena that concur in determining the line's actual shape.

2.4.1 Natural Broadening

For this one, you'll have to recall some of the basics results of Quantum Mechanics. The calculation is not particularly difficult, but it's surely long, so I'm going to omit that and take a simpler route.

Note that the spontaneous decay of an atomic state n proceeds at a rate

$$\Gamma = \sum_m A_{nm}$$

where the sum is over all states m of lower energy (which may be quite a few). If radiation is present, we should add the induced rates to this. The coefficient of the wave function of state n, therefore, is of the form $e^{-\Gamma t/2}$ and leads to a decay of the electric field by the same factor. We have an emitted spectrum determined by the decaying sinusoid type of electric field, so the line profile must be a *Lorentz or Natural profile*[6]

$$\phi(\nu) = \frac{\Gamma}{4\pi^2} \frac{1}{(\nu - \nu_0)^2 + \left(\frac{\Gamma}{4\pi}\right)^2}$$

More often than not, however, we're going to assume that the effect of natural broadening is rather peaked around ν_0, similar to a δ-function, since, at least as far as astrophysics is concerned, there are more relevant sources of broadening.

2.4.2 Doppler Broadening

Consider an atom in thermal motion, so that by simple relativistic considerations the frequency of emission or absorption in its own frame corresponds to a different frequency for the observer. Each atom will have its own Doppler shift, so that the net effect is to spread the line out, but not to change its total strength.

Recall the classic Doppler effect

$$\nu - \nu_0 = \nu_0 \frac{v_z}{c} \tag{43}$$

6 Fun fact: The Lorentzian distribution (also known as Cauchy distribution among mathematicians) is a classical example of a pathological distribution. In fact the Cauchy distribution has no mean, variance or higher momenta defined. Its mode and median are well defined and are both equal to ν_0.

Here ν_0 is the rest frequency. Note that this can be extended to *bulk motion* as well. We can decompose the velocity as

$$\vec{v} = \langle \vec{v} \rangle + \delta\vec{v}$$

which goes by the name of *Reynolds' decomposition* in Fluid-dynamics books. Generally, an ensemble of atoms will have different $\delta\vec{v}$, so each atom will absorb photons at different frequencies.

What we have to do is then convolute the Doppler effect over some velocity distribution. Assuming LTE to hold, such distribution will be the Maxwellian distribution. Switching into a reference frame where $\langle \vec{v} \rangle = 0$ (which is always possible), then the small fluctuations $\delta\vec{v}$ shall follow the Maxwellian distribution.

Please note that the one dimensional version of the Maxwellian distribution must be used, which is a Gaussian distribution with $\mu = 0$ and a given variance. The convoluted line profile will then be

$$\phi(\nu) = \int \delta(\nu - \nu_0)\nu_0 \left(1 + \frac{v}{c}\right) \exp\left(-\frac{m_a v^2}{2kT}\right) \mathrm{d}v$$

where, as anticipated, we identified the natural broadening as a δ-function. This is easily computed

$$\phi(\nu) = \frac{1}{\Delta\nu_D\sqrt{\pi}} e^{-(\nu-\nu_0)^2/(\Delta\nu_D)^2} \quad \Delta\nu_D = \frac{\nu_0}{c}\sqrt{\frac{2kT}{m_a}} \tag{44}$$

Here $\Delta\nu_D$ represents the *Doppler width*.

Generally speaking, the minimum broadening we can expect is given by a convolution of Doppler Broadening and Natural Broadening in its proper Lorentzian form. This hasn't a fancy analytical solution, but can be expressed in terms of what is called a *Voigt profile*.

Roughly speaking, the Voigt profile is something that looks like a Gaussian at the center of the line[7] (where Doppler effect dominates) but has the *wings* of the Lorentzian distribution.

$$H(a,u) = \frac{a}{\pi}\int_{-\infty}^{+\infty} \frac{e^{-y^2}}{a^2 + (u-y)^2}\,\mathrm{d}y \quad a = \frac{\Gamma}{4\pi\Delta\nu_D} \quad u = \frac{\nu - \nu_0}{\Delta\nu_D}$$

In terms of this compact definition of the Voigt profile, the overall line profile will be

$$\phi(\nu) = \frac{1}{\Delta\nu_D\sqrt{\pi}} H(a,u)$$

7 Sometimes it's called *kernel* of the distibution.

3 SCATTERING PROCESSES

3.1 TRANSPORT THROUGH SCATTERING

When we've first written down the equation for radiative transport (10) we've neglected the effects of a possibly much relevant source of photons, which is *scattering*, another fairly common emission process. Scattering depends completely on the amount of radiation falling on the spacetime element we're considering.

How do we include it into the picture of radiative transport?

For the present discussion we assume *isotropic* scattering, so that the scattered radiation is emitted equally into equal solid angles. We also assume that the total amount of radiation emitted per unit frequency range is just equal to the total amount absorbed in that same frequency range (*coherent scattering*).

These are some fair requirements if you think about it. Isotropy is most likely assured if the main type of scattering is the (non-relativistic) Thomson scattering on electrons, while coherence of the scattering is just requiring the scattering to be elastic.

Such conditions are not always met, but for the moment we'll turn our heads away from the obvious problem and see where this brings us.

The emission coefficient for coherent, isotropic scattering can be found equating the power absorbed per unit volume and frequency ranges to the corresponding power emitted. Here we define the *scattering coefficient* σ_ν so that

$$j_\nu = \sigma_\nu J_\nu$$

Dividing by the scattering coefficient, we find that the source function for pure scattering is simply equal to the mean intensity within the emitting material

$$S_\nu = J_\nu = \frac{1}{4\pi}\int_\Omega I_\nu \, \mathrm{d}\Omega$$

The equation of radiative transport is thus modified[1]

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = -\sigma_\nu(I_\nu - J_\nu) = -\sigma_\nu(I_\nu - \frac{1}{4\pi}\int_\Omega I_\nu \, \mathrm{d}\Omega) \tag{45}$$

Solving this equation is a bloodbath. The source function is not known a priori and depends on the solution I_ν at all directions through a given point, making this a *integro-differential equation*, or, in short, a "bloody mess".

1 Note that we're assuming for the scattering version of RT a similar form of the purely radiative version (10).

3.2 RANDOM WALKS

A particularly useful way of looking at scattering, which leads to important order-of-magnitude estimates, is by means of random walks. We shall now develope a formalism that interprets the processes of absorption, emission, and propagation in probabilistic terms for a single photon rather than the average behavior of large numbers of photons.

Consider a photon emitted in an infinite, homogeneous scattering region. It travels a displacement $\vec{r}_1$ before being scattered, then travels in a new direction over a displacement $\vec{r}_2$ before being scattered, and so on. After N free paths, the total displacement will look like

$$\vec{R} = \sum_{i=1}^{N} \vec{r}_i$$

The average total displacement will be identically null. Therefore, we must evaluate the mean square of the total displacement, which, in the assumption of independent and isotropic scattering, must be expressed as

$$l_*^2 = \langle \vec{R}^2 \rangle = \sum_{i=1}^{N} \langle \vec{r}_i^2 \rangle$$

The quantity l_* is the root-mean-square net displacement of the photon, while l^2 denotes the mean square of the free path of a photon, which within a factor of order unity, it's simply the mean free path of a photon.

Since the mean square displacements of each scattering iteration have no reason to be different, we'll just write

$$l_* = \sqrt{N} l$$

Said L the linear dimension of the object photons are trying to "escape" from, the number of scatterings events required to actually escape from regions of large optical depth is roughly determined by setting $l_* \approx L$. Then $N \sim L^2/l^2$. But since l is of the order of the mean free path, the optical thickness of the medium τ is of order L/l, our previous results yield

$$\tau^2 \approx N \quad \tau \gg 1$$

For regions of small optical thickness, the mean number of scatterings is small, so that we can approximate $N \approx \tau$. For most order-of-magnitude estimates we could then use

$$N \approx \tau^2 + \tau$$

which has the evident perk of well-behaving in the two limits we've discussed.

3.2.1 Combined scattering and absorption

What happens if we had both scattering and absorption? We'd have two terms on the right hand side of the transfer equation

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = -\alpha_\nu (I_\nu - B_\nu) - \sigma_\nu \left(I_\nu - \frac{1}{4\pi} \int_\Omega I_\nu \, \mathrm{d}\Omega \right) \tag{46}$$

Please note that we're assuming Kirchhoff's law to hold, and the medium to be optically thick, so that $j_\nu = \alpha_\nu B_\nu$. In terms of a suitably defined source function S_ν

$$S_\nu = \frac{\alpha_\nu B_\nu + \sigma_\nu J_\nu}{\alpha_\nu + \sigma_\nu}$$

we can rewrite eq.46 as

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = -(\alpha_\nu + \sigma_\nu)(I_\nu - S_\nu)$$

We may then define an *effective optical depth*[2] τ_*

$$\mathrm{d}\tau_* = (\alpha_\nu + \sigma_\nu)\,\mathrm{d}s$$

so that the mean free path will be just $l_\nu = (\alpha_\nu + \sigma_\nu)^{-1}$.

During the random walk the probability that a free path will end with a true absorption event is

$$\epsilon_\nu = \frac{\alpha_\nu}{\alpha_\nu + \sigma_\nu}$$

while the corresponding probability for scattering being $1 - \epsilon_\nu$, which is known as *single-scattering albedo*. In terms of ϵ_ν, the source function becomes

$$S_\nu = (1 - \epsilon_\nu)J_\nu + \epsilon_\nu B_\nu$$

Let's consider first an infinite homogeneous medium.

A random walk starts with the thermal emission of a photon (creation) and ends, possibly after a number of scatterings, with a true absorption (destruction). Since the walk can be terminated with probability ϵ at the end of each free path, the mean number of free paths is $N = \epsilon^{-1}$. We then have $l_* = l\epsilon^{-1/2}$.

So, if we were to substitute the definition of l_ν into l and use the definition of ϵ_ν, we'd get

$$l_* \approx \left[\alpha_\nu(\alpha_\nu + \sigma_\nu)\right]^{-1/2}$$

This length represents a measure of the net displacement between the points of creation and destruction of a typical photon; it often goes by the name of *diffusion length*, *thermalization length*, or *effective mean path*. Usually it is also frequency dependent.

Essentially, it is the average length over which emitting and absorbing elements are radiatively coupled.

The behavior of a finite medium can be explained with what we've been creating so far. Its properties will depend (strongly) on whether its linear extension L is larger or smaller than the effective free path.

In terms of the effective optical depth $\tau_* = L/l_*$, we can restate its definition as follows

$$\tau_* \approx \left[\tau_a(\tau_a + \tau_s)\right]^{1/2} \quad \tau_a = \alpha_\nu L \quad \tau_s = \sigma_\nu L$$

When the effective free path is large compared to L, we have $\tau_* \ll 1$ and the medium is *effectively thin*. Most photons will then escape by random walking their way out of the medium before being destroyed by a true absorption.

2 Also called *extinction coefficient*.

Conversely, a medium for which $\tau_* \gg 1$ is *effectively thick*. Most photons thermally emitted at depths larger than the effective path length will be destroyed by absorption before they get out.

This new definition allows us to reformulate when for a medium will be possible to be in LTE: Over a distance of order l_*, photons will do many scattering events, unable to leave the medium if $\tau_* \gg 1$, assuring that LTE is established *locally*.

Were this to happen, we could safely claim then that

$$I_\nu \to B_\nu \quad S_\nu \to B_\nu$$

This way is perhaps more clear why l_* is called thermalization length.

4 IONIZATION AND N-LTE PROCESSES

4.1 INTRODUCTION

The subject of interest of this chapter will follow closely the description given by [16] regarding objects called *emission nebulae*.

Emission nebulae are the result of the photoionization of a diffuse gas cloud by ultraviolet photons from a hot, "exciting" star or from a cluster of exciting stars.

At any given point, the ionization equilibrium will be given by the balance of photoionization and recombination processes of electrons with the ions.

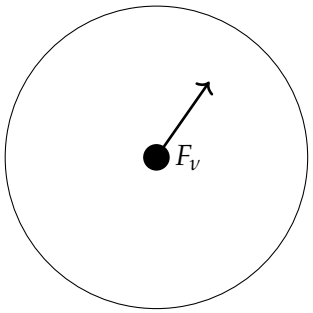

Figure 12: A hot star surrounded by a cloud of Hydrogen. Note that the cloud doesn't have to be spherical, especially if it posseses an inherent angular momentum, as we'll see in Chapter 8.

Now, since Hydrogen's the most abundant element in the Universe, we start considering a pure H-cloud surrounding a very hot star (Fig.12), where "very" is going to be quantified in a moment.

As we've briefly shown in Chapter 2, §2.3, Quantum Mechanics predicts for Hydrogen an energetic spectrum as such

$$E_n = -\frac{\mathrm{Ry}}{n^2}$$

where $\mathrm{Ry} \approx 13.6\,\mathrm{eV}$ is the energy needed to make the single electron of Hydrogen jump to the continuum, or, properly speaking, the energy needed to ionize Hydrogen. It's clear then that our central source of radiation must be able to emit photons of energy at least equal to 13.6 eV; those responding to this characteristic will be, not surprisingly, called *ionizing photons*.

Clearly, a finite source of ultraviolet photons cannot ionize an infinite volume, and therefore, if the star is in a sufficiently large gas cloud, there must be an outer edge to the ionized material. The thickness of this transition zone between ionized and neutral gas, since it is due to absorption, is approximately one mean free path of a ionizing photon.

Let's then try to write down an equation for ionization equilibrium

$$n(\mathrm{H}^0)\int_{\nu_0}^{\infty}\frac{4\pi J_\nu}{h\nu}a_\nu(\mathrm{H}^0)\,\mathrm{d}\nu = n(\mathrm{H}^0)\int_{\nu_0}^{\infty}\phi_\nu a_\nu(\mathrm{H}^0)\,\mathrm{d}\nu = n(\mathrm{H}^0)\Gamma(\mathrm{H}^0) \tag{47}$$

$$= n_e n_p \alpha(\mathrm{H}^0, T)\,[\mathrm{cm}^{-3}\,\mathrm{s}^{-1}] \tag{48}$$

where J_ν is the familiar mean intensity of radiation at the point. Thus, $\phi_\nu = 4\pi J_\nu/h\nu$ is the number of incident photons per unit area, per unit time, per unit frequency interval, and $a_\nu(\mathrm{H}^0)$ is the ionization cross section for Hydrogen by photons with energy $h\nu$; $\Gamma(\mathrm{H}^0)$ therefore represents the number of photoionizations events per Hydrogen atom per unit time.

The neutral atom, electron, and proton densities per unit volume are n, n_e and n_p, while α is a tabulated recombination coefficient.

For a single point-like source, to a first approximation, the mean intensity J_ν is simply

$$4\pi J_\nu = \frac{L_\nu}{4\pi r^2}$$

where L_ν is the luminosity of the star per unit frequency interval.

4.2 PHOTOIONIZATION OF A PURE HYDROGEN NEBULA

As anticipated, only radiation with frequency $\nu \geq \nu_0$ so that $h\nu_0 \approx 13.6\,\mathrm{eV}$ is effective in the photoionization of Hydrogen from its ground state[1]. The equation of radiative transfer takes the form

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = -n(\mathrm{H}^0)a_\nu I_\nu + j_\nu$$

It is convenient to divide the monochromatic intensity of the radiation field into two parts: One strictly due to the star, $I_{\nu,S}$, resulting from the outputted radiation, and one "diffuse" part, which substantially includes the contribution due to the secondary emissions of the ionized gas.

$$I_\nu = I_{\nu,S} + I_{\nu,D} \tag{49}$$

Let's focus on the strictly stellar radiation. Because of geometrical dilution (r^{-2} suppression) and absorption, this contribution to the total intensity decreases outwards[2]

$$4\pi J_{\nu,S} = \pi F_{\nu,S}(r) = \pi F_{\nu,S}(R)\frac{R^2\exp(-\tau_\nu)}{r^2} \tag{50}$$

where $F_{\nu,S}$ is the flux of stellar radiation. The optical depth τ_ν was defined as

$$\tau_\nu(r) = \int_0^r n(\mathrm{H}^0, r')a_\nu\,\mathrm{d}r'$$

On the other hand, for the diffused radiation we have

$$\frac{\mathrm{d}I_{\nu,D}}{\mathrm{d}s} = -n(\mathrm{H}^0)a_\nu I_{\nu,D} + j_\nu$$

1 Note that, were Hydrogen in a higher energy state, less-energetic photons would suffice.

2 Remember that all spontaneous emissions are supposed to occur in the cloud.

If we assume the average kinetic energy per particle to be much smaller than the ionization threshold $kT \ll h\nu_0$, then the only source of ionizing radiation is recombinations of electrons from the continuum to the ground level

$$j_\nu(T) = \frac{2h\nu^3}{c^2}\left(\frac{h^2}{2\pi mkT}\right)^{3/2} a_\nu \exp\left[-h(\nu-\nu_0)/kT\right] n_p n_e$$

which is strongly peaked around the threshold. This means that more often than not the re-emitted photon will be able to ionize again the medium. We can calculate the number of such re-emitted photons

$$4\pi \int_{\nu_0}^{\infty} \frac{j_\nu}{h\nu}\,\mathrm{d}\nu = n_p n_e \alpha_1(\mathrm{H}^0, T) \tag{51}$$

where $\alpha_1(\mathrm{H}^0, T)$ is the ground-state recombination coefficient. In principle we can write down the (total) recombination coefficient as

$$\alpha(\mathrm{H}^0, T) = \sum_{n=1}^{+\infty} \alpha_n \tag{52}$$

with n the principal quantum number. It is clear then that $\alpha_1 < \alpha$ and therefore $J_{\nu,D} < J_{\nu,S}$ on average.

For an optically thin nebula, a good approximation is to take $J_{\nu,D} \approx 0$, while for an optically thick nebula, since no ionizing photons can escape, we assume that every diffuse radiation-field photon generated in such a nebula is absorbed elsewhere in the nebula

$$4\pi \int \frac{j_\nu}{h\nu}\,\mathrm{d}V = 4\pi \int n(\mathrm{H}^0)\frac{a_\nu J_{\nu,D}}{h\nu}\,\mathrm{d}V$$

but we can also give a similar relation assuming it to hold locally (this is sometimes called the "*on the spot*" approximation)

$$J_{\nu,D} = \frac{j_\nu}{n(\mathrm{H}^0)a_\nu} \tag{53}$$

which immediately satisfies the global relation. Generally, this is not a bad approximation, because the diffuse radiation-field photons have $\nu \approx \nu_0$, and therefore have large a_ν[3] and correspondingly small mean free paths before absorption. Actually, eq.53 would be exact if all photons were absorbed very close to the point at which they are generated.

Making use of the on-the-spot approximation and using both (50) and (51), we find that the ionization equation (48) becomes

$$\frac{n(\mathrm{H}^0)R^2}{r^2}\int_{\nu_0}^{\infty} \frac{\pi F_\nu(R)}{h\nu} a_\nu \exp(-\tau_\nu)\,\mathrm{d}\nu = n_p n_e \alpha_B(\mathrm{H}^0, T) \tag{54}$$

where

$$\alpha_B(\mathrm{H}^0, T) = \sum_{n=2}^{\infty} \alpha_n$$

The physical meaning is that in optically thick nebulae, the ionizations caused by (strictly) stellar radiation-field photons are balanced by recombinations to

3 This one is the absorption coefficient, not the recombination coefficient.

excited levels of Hydrogen, while recombinations to the ground level generate ionizing photons that are absorbed elsewhere in the nebula but have no net effect on the overall ionization balance.

Properly solving eq.54 gives us an estimate of the size of the fully ionized region (that for historical reasons is referred to as H_{II}), of which we can get a grasp by looking at Fig.13.

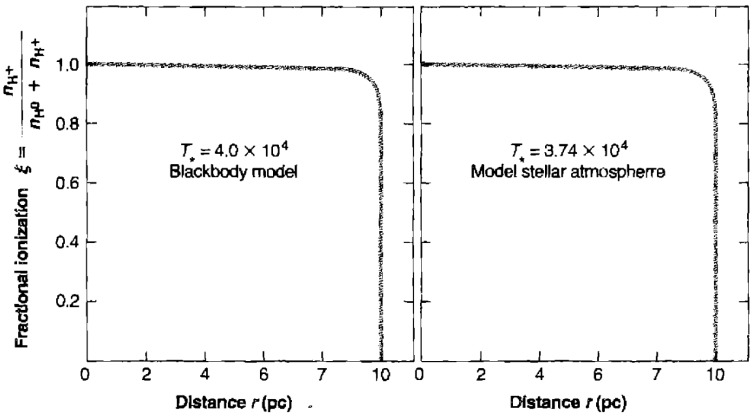

Figure 13: Ionization structure of two homogeneous pure H model H_{II} regions. Credits: Ferland, Osterbrock [16].

Let's assume we're somehow given an input spectrum from the source $\pi F_\nu(R)$, we can find the radius of the H_{II} region by substituting

$$\frac{\mathrm{d}\tau_\nu}{\mathrm{d}r} = n(\mathrm{H}^0)a_\nu$$

back into eq.54 and integrating over r

$$\begin{aligned} R^2 \int_{\nu_0}^{\infty} \frac{\pi F_\nu(R)}{h\nu}\,\mathrm{d}\nu \int_0^{\infty} \mathrm{d}\left[-\exp(-\tau_\nu)\right] &= \int_0^{\infty} n_p n_e \alpha_B(\mathrm{H}^0, T) r^2\,\mathrm{d}r \\ &= R^2 \int_{\nu_0}^{\infty} \frac{\pi F_\nu(R)}{h\nu}\,\mathrm{d}\nu \end{aligned}$$

If we define r_1 as the radius of complete ionization, we have that within this radius $n_p = n_e \approx n(\mathrm{H})$ and nearly zero outside; last equation then becomes

$$4\pi R^2 \int_{\nu_0}^{\infty} \frac{\pi F_\nu(R)}{h\nu} = \int_{\nu_0}^{\infty} \frac{L_\nu}{h\nu} \tag{55}$$

$$= \frac{4\pi}{3} r_1^3 n_{\mathrm{H}}^2 \alpha_B \tag{56}$$

The physical meaning of this is that the total number of ionizing photons emitted by the star balances the total number of recombinations to excited levels within the ionized volume $4\pi r_1^3/3$, which goes by the name of *Strömgren's sphere*. To give an idea about the size of this sphere, O-type stars can fully ionize Hydrogen up to distances of the order of the parsec.

What if we start including more species, for example Helium? Clearly, we should try to understand how the new ionization stages enter the picture,

even more so since Helium has two different ionized variations (He_{II}, He_{III}) with higher ionization thresholds.

For once, photons coming from Helium recombinations can ionize Hydrogen, but the same can't be said if we flip the roles. If then we start including also heavier elements (metals) it gets even crazier. But, alas, their presence makes pressure gradients even more relevant.

One concluding note, of these regions we do see lines in emission and not a continuum as expected. This is because LTE doesn't hold; collisions aren't frequent enough to allow LTE to settle and thus energy levels with long lifetimes become observable due to the lack of de-exciting collisions.

4.3 N-LTE RADIATION TRANSFER

As we may have guessed from last section, the temperature in a static nebula is fixed by the equilibrium between photoionization-heating and cooling by recombination and radiation. Consider for example a photon of energy $h\nu$

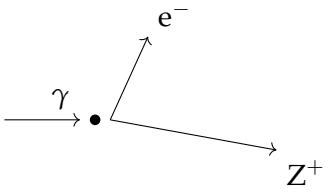

Figure 14: Schematic representation of a photon ionizing a neutral Hydrogen atom.

bumping on a neutral Hydrogen atom as shown in Fig.14. The photoelectron that will be thus produced will have an energy

$$\frac{1}{2}mu^2 = h(\nu - \nu_0)$$

where $h\nu_0$ is the ionization threshold potential. For these electrons we expect to see some distribution for their velocity, not necessarily thermal (17). Only a part of the electrons will be recombining, namely those with low energy; high-energy electrons will instead collide with ions, exciting them. In either of those scenarios, the net result is that for each captured electron $mu^2/2$ disappears from the total energy balance.

It's simplest to begin by considering what happens in a pure Hydrogen nebula. At any given point within the nebula, the photoionization rate is

$$G(\mathrm{H}^0) = \int_{\nu_0}^{\infty} \frac{4\pi J_\nu}{h\nu} n(\mathrm{H}^0) h(\nu - \nu_0) a_\nu \, \mathrm{d}\nu \tag{57}$$

This gives the expected value of the energy over the whole photon distribution. If we assume photoionization-equilibrium, we can use (48) to eliminate $n(\mathrm{H}^0)$

$$\begin{aligned} G(\mathrm{H}^0) &= n_e n_p \alpha(\mathrm{H}^0, T) \frac{\int_{\nu_0}^{\infty} \frac{4\pi J_\nu}{h\nu} h(\nu - \nu_0) a_\nu \, \mathrm{d}\nu}{\int_{\nu_0}^{\infty} \frac{4\pi J_\nu}{h\nu} a_\nu \, \mathrm{d}\nu} \\ &= n_e n_p \alpha(\mathrm{H}^0, T) \frac{3}{2} k T_i \end{aligned}$$

From this equation we can see that the mean energy of a newly created photoelectron depends on the form of the ionizing radiation field, but not on the absolute strength of the radiation. The quantity $3kT_i/2$ represents the initial temperature of the photo-generated electron.

If we assume blackbody behavior $J_\nu \approx B_\nu(T_*)$ with temperature T_*, it's possible to show that $T_i \approx T_*$ as long as $kT_* < h\nu_0$.

We now turn to find a description for the kinetic energy that is lost during recombination processes. It can be given in the following form

$$L_R = n_e n_p kT\beta(\mathrm{H}^0, T) \tag{58}$$

where

$$\beta(\mathrm{H}^0, T) = \sum_{n=1}^{\infty} \beta_n(\mathrm{H}^0, T) = \sum_{n=1}^{\infty} \sum_{L=0}^{n-1} \beta_{nL}(\mathrm{H}^0, T) \tag{59}$$

with

$$\beta_{nL}(\mathrm{H}^0, T) = \frac{1}{kT} \int_0^\infty u\sigma_{nL}(\mathrm{H}^0, T) \frac{1}{2} m u^2 f(u)\, \mathrm{d}u \tag{60}$$

The left-hand side of (60) is thus effectively a kinetic energy averaged recombination coefficient. Note that since the recombination cross sections are approximately proportional to u^{-2}, the electrons of lower kinetic energy are preferentially captured, and the mean energy of the captured electrons is somewhat less than $3kT/2$.

If radiation losses are negligible, then the thermal equilibrium equation would be

$$G(\mathrm{H}^0) = L_R(\mathrm{H}^0)$$

and the solution for the nebular temperature would give $T > T_i$ because of the "heating" due to the preferential capture of the slower electrons.

In presence of radiative losses, the nebula is going to further cool down. As we've already done in the last section, we're going to split J_ν into two contributions, the diffuse radiation and the stellar radiation modified by absorption. Then, using the *on the spot* approximation, we can neglect contributions to the balance coming from $n = 1$ and rewrite the luminosity

$$L_R = n_p n_e \beta_B(\mathrm{H}^0, T) kT$$

with

$$\beta_B(\mathrm{H}^0, T) = \sum_{n=2}^{\infty} \beta_n$$

The photoionization factor becomes

$$\begin{aligned} G(\mathrm{H}^0)_{OTS} &= n(\mathrm{H}^0) \int_{\nu_0}^\infty \frac{4\pi J_{\nu,S}}{h\nu} h(\nu - \nu_0) a_\nu \, \mathrm{d}\nu \\ &= n_e n_p \alpha_B(\mathrm{H}^0, T) \frac{\int_{\nu_0}^\infty \frac{4\pi J_{\nu,S}}{h\nu} h(\nu - \nu_0) a_\nu \, \mathrm{d}\nu}{\int_{\nu_0}^\infty \frac{4\pi J_{\nu,S}}{h\nu} a_\nu \, \mathrm{d}\nu} \end{aligned}$$

For high-energy electrons, a minor contributor to the cooling rate, which nevertheless is important, is free-free (FF) radiation or Bremsstrahlung, in which a continuous spectrum is emitted and does not involve recombinations.

The luminosity by Bremsstrahlung is given by

$$L_{FF}(Z) = 4\pi j_{ff} \tag{61}$$

$$= \frac{2^5 \pi e^6 Z^2}{3^{3/2} hmc^3} \left(\frac{2\pi kT}{m}\right)^{1/2} g_{ff} n_e n_+ \tag{62}$$

where n_+ is the number density of the ions. The numerical factor g_{ff} goes by the name of *Gaunt factor* for free-free emission; it is a slowly varying function of n_e and T. Generally, for nebular conditions is in the range $1.0 < g_{ff} < 1.5$, and usually an average $g_{ff} \approx 1.3$ is adopted.

4.3.1 Loss by Excitational Collisions

A far more important source of radiative cooling is collisional excitation of heavier species. These ions make a significant contribution in spite of their low abundance because they have energy levels with excitation potentials of the order of kT, but all the levels of Hydrogen and Helium have much higer excitation potentials, therefore rendering them less important in this regard.

Let's therefore examine how an ion is excited to a higher level by electron collisions with ions in a lower level. The cross section for this reaction is

$$\sigma_{12}(u) = \frac{\pi\hbar^2}{m^2 u^2} \frac{\Omega(1,2)}{\omega_1} \text{ for } \frac{1}{2} m u^2 > \chi$$

where $\Omega(1,2)$ is the energy-specific collision strength and ω_1 is the statistical weight of the electron. Note that the cross section for excitation is zero below the threshold $\chi = h\nu_{21}$.

There is a relation between the cross section for de-excitation and the cross section for excitation coming from the principle of detailed balance[4]

$$\omega_1 u_1^2 \sigma_{12}(u_1) = \omega_2 u_2^2 \sigma_{21}(u_2)$$

where u_1 and u_2 are related by the conservation of energy

$$\frac{1}{2} m u_1^2 = \frac{1}{2} m u_2^2 + \chi$$

The collisional de-excitation cross section has, not surprisingly, the same functional form as that of the excitation cross section (apart from a permutation $1 \leftrightarrow 2$). The total collisional de-excitation rate per unit volume per unit time is then

$$n_e n_2 q_{21} = n_e n_2 \int_0^\infty u \sigma_{21} f(u)\, \mathrm{d}u \tag{63}$$

$$= n_e n_2 \left(\frac{2\pi}{kT}\right)^{1/2} \frac{\hbar^2}{m^{3/2}} \frac{\Gamma(1,2)}{\omega_2} \tag{64}$$

$$= n_e n_2 \frac{8.629 \cdot 10^{-6}}{T^{1/2}} \frac{\Gamma(1,2)}{\omega_2} \tag{65}$$

where $\Gamma(1,2)$ is the velocity-averaged collision strength

$$\Gamma(1,2) = \int_0^\infty \Omega(1,2;E) \exp(-E/kT)\, \mathrm{d}\left(\frac{E}{kT}\right)$$

4 Check Chapter 2 for the definition.

with $E = mu_2^2/2$. These coefficients must be calculated quantum-mechanically. Similarly, the collisional excitation rate is $n_e n_2 q_{12}$, where

$$q_{12} = \frac{8.629 \cdot 10^{-6} \Gamma(1,2)}{T^{1/2} \omega_1} \exp(-\chi/kT)$$

Although this wasn't discussed in class, I shall point out that, as shown by [3] §3.5, there exists a simple relation for the collision strengths between a term consisting of a single level and a term consisting of various levels, namely

$$\Gamma(SLJ, S'L'J') = \frac{(2J'+1)}{(2S'+1)(2L'+1)} \Gamma(SL, S'L') \tag{66}$$

if either $S = 0$ or $L = 0$. The factors $(2J'+1)$ and $(2S'+1)(2L'+1)$ are the statistical weights of the level and of the term, respectively.

For very low electron densities, every collisional excitation is followed by the emission of a photon, thus the cooling rate is simply

$$L_C = n_e n_1 q_{12} h\nu_{21}$$

For general electron densities we can write

$$\begin{aligned} n_e n_1 q_{12} &= n_e n_2 q_{21} + n_2 A_{21} \\ \implies \frac{n_2}{n_1} &= \frac{n_e q_{12}}{A_{21}} \left[1 + \frac{n_e q_{21}}{A_{21}} \right]^{-1} \end{aligned}$$

so that the total cooling rate becomes

$$L_C = n_2 A_{21} h\nu_{21} = n_e n_1 q_{12} h\nu_{21} \left[1 + \frac{n_e q_{21}}{A_{21}} \right]^{-1} \tag{67}$$

Note that in the limit $n_e \to 0$ we recover the expression we've written earlier, while in the $n_e \to \infty$ limit

$$L_C = n_1 \frac{\omega_2}{\omega_1} \exp(-\chi/kT) A_{21} h\nu_{21}$$

we retrieve the thermodynamic-equilibrium cooling rate.

Some ions have many more levels than just two; for those, our simple formalism doesn't work very well. What we can do, however, is fine-tuning it so to include the fact that collisional and radiative transitions can occur between any of the levels, and excitation and de-excitation cross sections and collision strengths exist between all pairs of the levels.

The equilibrium equations for each of the levels i becomes

$$\sum_{j \neq i} n_j n_e q_{ji} + \sum_{j>i} n_j A_{ji} = \sum_{j \neq i} n_i n_e q_{ij} + \sum_{j<i} n_i A_{ij}$$

This, together with the total number of ions

$$\sum_j n_j = n$$

can be solved for the relative population in each level and then for the cooling rate

$$L_C = \sum_i n_i \sum_{j<i} A_{ij} h\nu_{ij}$$

The temperature at each point in a static nebula is then determined by imposing the equilibrium between the heating and cooling rates we've found

$$G = L_R + L_{FF} + L_C$$

4.3.2 Recombination Lines

Why is it important talking about the lines originating from the occurence of recombination processes?

The radiation emitted by each element of volume in a gaseous nebula depends upon the abundances of the elements, determined by the previous evolutionary history of the gas, and on the local ionization, density, and temperature, determined by the radiation field and the abundances.

Hence, it would be safe to assume that we'd have plenty to gain from studying these lines. Once again, we're going to turn our attention to a nebula whose composition is Hydrogen-dominated.

The equation of statistical equilibrium for any level nL may be written as

$$n_p n_e \alpha_{nL}(T) + \sum_{n'>n}^{\infty} \sum_{L'} n_{n'L'} A_{n'L',nL} = n_{nL} \sum_{n''=1}^{n-1} \sum_{L''} A_{nL,n''L''} \tag{68}$$

Note that, in general, selection rules impose $A_{n'L',n''L''} \neq 0$ only if $L' = L'' \pm 1$.

It is usually convenient to express the population in terms of the dimensionless factors b_{nl} to account for departure from LTE at a given temperature, electron density and proton density. In LTE Saha's equation must hold

$$\frac{n_p n_e}{n_{1s}} = \left(\frac{2\pi mkT}{h^2}\right)^{3/2} \exp(-h\nu_0/kT)$$

as well as the Boltzmann equation for occupation numbers

$$\frac{n_{nL}}{n_{1s}} = (2L+1)\exp(-\chi_n/kT)$$

where the factor $2L+1$ is the ratio of statistical weighs of the nL and $1s$ levels. Then the population in the nL level may be written

$$n_{nL} = b_{nL}(2L+1)\left(\frac{h^2}{2\pi mkT}\right)^{3/2} \exp(X_n/kT) n_p n_e$$

where $b_{nl} = 1$ in LTE and

$$X_n = h\nu_0 - \chi_n = \frac{h\nu_0}{n^2}$$

is the ionization potential of the nL level. We can plug this expression back into the equation for statistical equilibrium (68) and find out that the b_{nL} coefficients are *not* a function of the density as long as recombination and downward-radiative transitions are the only relevant processes

$$\begin{aligned}
&\alpha_{nL} \frac{1}{2L+1}\left(\frac{2\pi mkT}{h^2}\right)^{3/2} \exp(-X_n/kT) \\
&\quad + \sum_{n'>n}^{\infty} \sum_{L''} b_{n'L'} A_{n'L',nL} \left(\frac{2L'+1}{2L+1}\right) \exp\left[(X_{n'} - X_n)/kT\right] \\
&\quad = b_{nL} \sum_{n''=1}^{n-1} \sum_{L''} A_{nL,n''L''}
\end{aligned}$$

Note that the equations above can be solved by working downward in n, for if the b_{nL} are known for all $n \geq n_k$, then the n equations with given L each contain a single unknown b_{nL}. Solution is thus "immediate".

It's convenient to express the solutions as a *cascade matrix* $C(nL, n'L')$ which is the probability to go from level nL to nL' regardless of the route that is taken[5]. Starting from the probability matrix $P(nL, n'L')$

$$P_{nL,n'L'} = \frac{A_{nL,n'L'}}{\sum_{n''=1}^{n-1} \sum_{L''} A_{nL,n''L''}}$$

which is zero unless $L' = L \pm 1$ (recall the usual selection rules from Wigner-Eckart's theorem).

For example, if $n' = n - 1$, the cascade matrix has the form

$$C_{nL,n-1L'} = P_{nL,n-1L'}$$

and so on, so that if we define

$$C_{nL,nL''} = \delta_{LL''}$$

then in general the cascade matrix takes the form

$$C_{nL,n'L'} = \sum_{n''>n'}^{n} \sum_{L''=L'\pm1} C_{nL,n''L''} P_{n''L'',n'L'}$$

Once the populations n_{nL} have been found, we can calculate the emission coefficient in each line

$$j_{nn'} = \frac{h\nu_{nn'}}{4\pi} \sum_{L=0}^{n-1} \sum_{L'=L\pm1} n_{nL} A_{nL,n'L'}$$

All the machination we've built until now works pretty well as an approximation for the diffuse spectrum in optically *thin* nebulae, assuming absorption events are absent.

Unfortunately, for pure Hydrogen all of this works rather poorly.

H-nebulae tends to be optically thick, especially in the Lyman lines[6]. For those, the equation for the central line-absorption cross section is

$$a_0(\mathrm{Ly}_n) = \frac{3\lambda_{n1}^3}{8\pi} \left(\frac{m_\mathrm{H}}{2\pi kT}\right)^{1/2} A_{nP,1s}$$

where λ_{n1} is the wavelength of the line. At a typical temperature for a stellar atmosphere $T \sim 10^4\,\mathrm{K}$, the optical depth in the Lyman-α (Ly_α) is about 10^4 times the optical depth at the Lyman limit $\nu = \nu_0$ of the ionizing continuum. Thus the optical depth will be of order $\tau \sim 1 - 10^4$ and the nebula will be optically thick.

The resulting spectrum will be that of a continuum (due to Bremsstrahlung) superposed with some emission lines.

---

5 In a sense, it is a marginalized probability over the trajectories in the nL space.

6 The Lyman series is a Hydrogen spectral series of transitions and resulting UV emission lines as an electron goes from $n \geq 2$ to $n = 1$.

Part II

FLUID DYNAMICS

5 FUNDAMENTALS OF FLUID DYNAMICS

5.1 PHYSICAL PROPERTIES OF FLUIDS

A "simple" fluid might be defined as a material such that the relative positions of the elements that make it change by an amount which is not small when suitably chosen forces are applied.

Given this definition, we could be tempted to say there's not much difference between gases and liquids, but that wouldn't be a clever guess. The main difference between gases and liquids lays not in density, but in *compressibility*.

To build a coherent and "nice" description of a fluid, we have to be able to attach a definite meaning to the notion of value "at a point" for the various fluid properties. What we'd like is, essentially, to be able to treat a fluid like a continuum.

We know that in the real world an abstract concept like a continuum cannot possibly exists: Matter is discrete and discrete only. However, it may have occurred to you that normally we don't exactly see matter broken down to its fundamental components.

We're able to regard a fluid as a continuum when the measured fluid property is constant for sensitive volumes small on the macroscopic scale but large in respect to the microscopic ones.

When looking at a fluid, we can distinguish two kinds of forces[1] that act on matter in bulk

- long-range forces or *body forces* (like gravity): $\delta\vec{F}_v = \vec{F}(\vec{x}, t)\rho\delta V$
- short-range forces that arise from reactions with matter

The latter can be expressed as: $\delta\vec{F}_s = \vec{\sigma}(\hat{n}, \vec{x}, t)\delta A$, where $\vec{\sigma}$ is the stress exerted by the fluid on the surface element to which $\hat{n}$ points.

The force exerted across the surface element on the fluid on the side to which $\hat{n}$ points is such that

$$-\vec{\sigma}(\hat{n}, \vec{x}, t)\delta A = \vec{\sigma}(-\hat{n}, \vec{x}, t)\delta A$$

The classical procedure to obtain a functional expression for $\vec{\sigma}$ comes from considering all the forces acting instantaneously on the fluid within a δV volume in the shape of a tetrahedron (Fig.15). The calculation is rather tedious but at least straightforward. At the end of it, you'll find out that the i-th component of the stress vector can be written as

$$\sigma^i = \sigma^{ij} n_j$$

where the index placement is definitely overkill, but we'll be clearer in a moment. Here and in the following, we'll be often adopting Einstein convention for summation over repeated indices.

1 Note that the forces proposed are to be intended *per unit mass*.

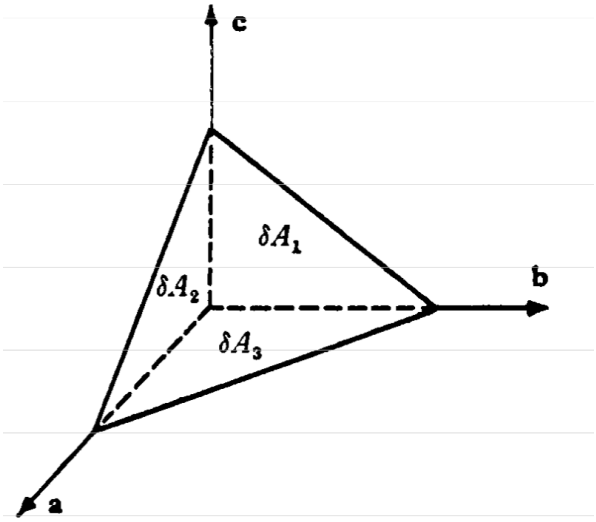

Figure 15: Tetrahedron construction for the stress tensor. Credits: Batchelor [2]

It is customary to bestow a name on σ^{ij}: The *stress tensor*. Curiously enough, the conservation of total angular momentum implies that the the stress tensor must be *symmetric*. This might be ringing a bell in your memory.

If we put ourselves in the fluid's rest frame, where there's no net flux of any of the components of momentum in any of the three orthogonal directions, it's not a far stretch to be claiming that the fluid must, indeed, be isotropic at rest.

Isotropy implies that, at least in this reference frame, the stress tensor must be diagonal. Since the symmetry of the system is essentially spherical, all the spatial components σ^{ii} must therefore be equal. The most general form for the 4-dimensional stress tensor is simply the energy-momentum tensor for a perfect fluid (as far as special relativity is concerned)

$$T^{\mu\nu} = (\rho + p)u^\mu u^\nu + p\eta^{\mu\nu} \tag{69}$$

$$T^{\mu\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}$$

Note that we'll be using the same metric convention of [5], which is $\eta_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$ and we will work in units $c = 1$[2]. The 4-vector u^μ is the 4-velocity $u^\mu = \mathrm{d}x^\mu/\mathrm{d}s$. Please note that requiring the energy-momentum tensor to be locally conserved $\partial_\mu T^{\mu\nu} = 0$ allows us in principle to deduce the three fundamental equations of fluid dynamics: The mass conservation, the energy conservation and Euler's equation. We will, however, get to those equation taking another route, passing briefly from statistical mechanics. For those interested, you can see [5], §1.9.

2 To properly reintroduce c factors, you have to look carefully at the definition of u^μ. Here, and only here, with ρ we're referring to an *energy density* and not to a *matter density*.

The stress tensor we've been discussing so far is the solely spatial component of $T^{\mu\nu}$, which is $\sigma^{ij} = T^{ij} = p\delta^i_j$. Since the stress tensor is isotropic, all its diagonal elements are equal and then we may just take

$$p = -\frac{1}{3}\sigma^i{}_i$$

as a definition of *static fluid pressure*.

5.1.1 Mechanical equilibrium

A necessary condition for equilibrium of a fluid requires that body and surface forces compensate

$$\int_V \rho\vec{F}\,\mathrm{d}V - \int_{\partial V} p\hat{n}\,\mathrm{d}A = 0$$

Applying the divergence theorem and requiring that what we find holds for all possible volumes V, last equation yields

$$\rho\vec{F} = \nabla p \tag{70}$$

If $\vec{F}$ is a conservative force $\vec{F} = -\nabla\phi$, then the condition for mechanical equilibrium may be cast in an equivalent form by taking the curl of both members

$$\nabla\rho \wedge \nabla\phi = 0 \implies \frac{\mathrm{d}p}{\mathrm{d}\phi} = -\rho(\phi)$$

So we find out that equilibrium in a fluid is possible if isochoric and equipotential surfaces are aligned.

For example, for a self-gravitating medium, the following equation holds (in special relativity at least)

$$\nabla^2\phi = 4\pi G\rho \tag{71}$$

which can be suitably rearranged, recalling that $\nabla\phi = -\nabla p/\rho$, into (we assume spherical symmetry)

$$\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{r^2}{\rho}\frac{\mathrm{d}p}{\mathrm{d}r}\right) = -4\pi G\rho \tag{72}$$

which is of crucial importance for stellar structure models. Note that this is a prime example of the *closure problem*: To be properly solved (either numerically or analitically) we need to specify an equation of state of some kind.

Lane-Emden equation

Assume for example a barotropic relation of the form

$$p(\rho) = \kappa\rho^{1+1/n} \tag{73}$$

$$\rho(r) = \rho_c\phi^n(r) \tag{74}$$

with ϕ some adimensional density profile with the following characteristics: $\phi(0) = 1$ and $\partial_r \phi|_{r=0} = 0$.

In terms of an adimensional radius $r = a\eta$, with a constant depending only on n

$$a = \left[\frac{K\rho_c^{1/n-1}(n+1)}{4\pi G}\right]^{1/2}$$

eq.72 takes a much simpler and elegant form

$$\frac{1}{\eta^2}\partial_\eta(\eta^2\partial_\eta\phi) = -\phi^n \tag{75}$$

which is the general *Lane-Emden equation*. Solutions for this equation have been calculated numerically in Fig.16. Albeit not a topic covered by the

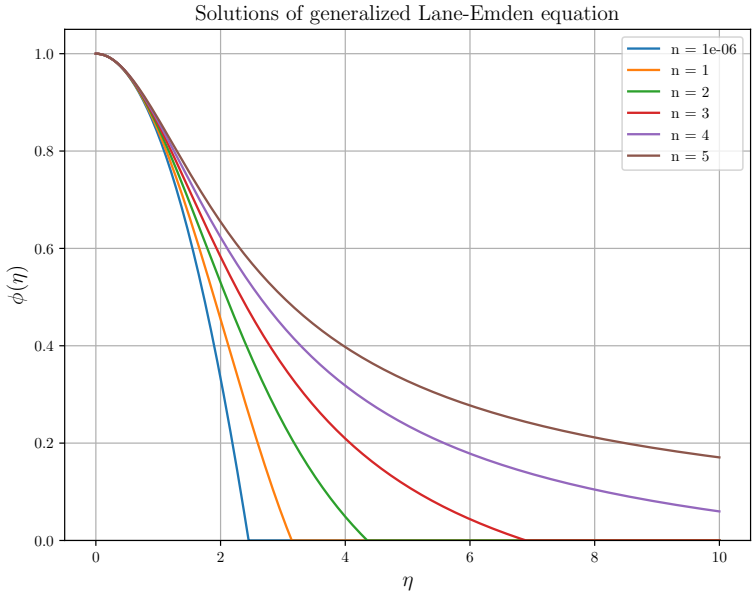

Figure 16: Numerical solutions to the Lane-Emden equation.

course, I'd still like to point out that from eq.75 it's possible to find the fixed value of mass able to satisfy the equation in the ultrarelativistic limit of degenerate electrons ($n = 3$). The mass is given by

$$M = \int_0^R 4\pi r^2 \rho \, \mathrm{d}r = 4\pi a^2 \rho_c \int_0^{\eta_1} \eta^2 \phi^2 \, \mathrm{d}\eta$$

Note that in the equation above, the integral will evaluate to the same result for all stars with a given n. This is the first time we encounter a *self-similar solution*, but we'll have time to get acquaintanted with those in future chapters.

The value of that integral is actually fixed by the original equation itself to $-\eta_1^2 \partial_\eta \phi|_{\eta_1}$.

If we consider now a specific value of n, namely $n = 3$, we'd find out that the mass of the star does not depend in any way from the centeral density of the star. That means that in such a regime, only one fixed value of mass is allowed.

Numerically we find that for $n = 3$, $|\eta_1^2 \partial_\eta \phi|(\eta_1)| \approx 2.018$, so that the limiting mass is of the order of

$$M_{Ch} \approx 1.46 \left(\frac{2}{\mu_e}\right) M_\odot \tag{76}$$

This is the celebrated *Chandrasekhar mass limit* that constrains white dwarves from having larger masses.

Spherically symmetric rotating objects

Another most interesting example is what we observe if we consider a rotating spherical fluid with radius R and angular velocity $\vec{\Omega}$. In the comoving frame, there are obviously fictious forces arising to account for the rotation: *Coriolis' force* $\vec{\Omega} \wedge \vec{v}$ and *Centrifugal acceleration* $\vec{\Omega} \wedge (\vec{\Omega} \wedge \vec{r})$. For the time being, we'll neglect the former.

Centrifugal acceleration can be expressed by the gradient of a scalar potential

$$\vec{\Omega} \wedge (\vec{\Omega} \wedge \vec{r}) = -\frac{1}{2}\nabla(\vec{\Omega}^2 \vec{r}^{\,2}) = -\frac{1}{2}\nabla\phi_C \tag{77}$$

This means that the effective potential is no longer that granting spherical symmetry (say, the usual Newtonian gravitational potential). The equation for mechanical equilibrium will have to be modified into

$$-\rho\nabla(\phi - \phi_C) = \nabla p$$

which in polar coordinates reads something like

$$-\rho\partial_r\left[\phi - \Omega^2 r^2\right] = \partial_r p \tag{78}$$

$$-\rho\partial_z\phi = \partial_z p \tag{79}$$

In conjunction with eq.71 we could, at least theoretically, study the deviation from spherical symmetry, given that we're provided an equation of state, that is. But in general it's safe to assume that rotation **affects the structure of a star**. Since the star is no longer spherically symmetric, we'll no longer

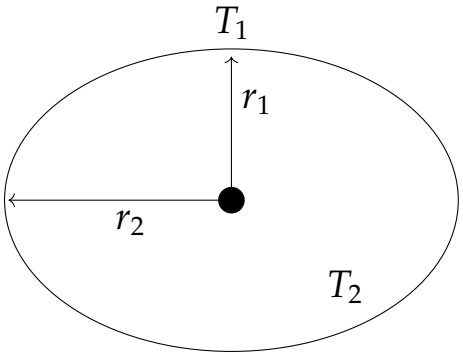

Figure 17: The effect of rotation on a spherically symmetric object. Symmetry is broken and gradients arise.

see the same surface temperature everywhere, but it will change from point to point, giving rise to temperature and pressure gradients.

Now, if we are in presence of a temperature gradient, a sensible call would be remembering of the existence of *thermal conduction*, which is temperature-gradients driven. This implies the presence of blobs of hot gas (or whatever) moving towards the cooler regions (Fig.17). At this point, we'll no longer be able to neglect the Coriolis force, which will further twist the surface of the star.

It's clear that hydrostatic equilibrium is not sustainable in these conditions.

5.2 VLASOV EQUATION

What we'd like to do in the present section is trying to build a formalism that allows to connect the microscopical interpretation of matter through statistical mechanics to the macroscopic space.

To do so, we'll define a distribution function $f = f(\vec{x}, \vec{v}, t)$ so that the number of particles in range $[\vec{x}, \vec{x} + \mathrm{d}\vec{x}]$ and $[\vec{v}, \vec{v} + \mathrm{d}\vec{v}]$ is given by

$$\mathrm{d}n(t) = f(\vec{x}, \vec{v}, t)\, d\vec{x}\, \mathrm{d}\vec{v} \tag{80}$$

To be properly defined, f must meet the usual requirements for distribution functions

$$f \geqslant 0 \quad \int f\, \mathrm{d}\vec{x} < +\infty \quad \int f\, \mathrm{d}\vec{v} < +\infty \tag{81}$$

Thanks to these properties we can easily define the number and mass density

$$n(\vec{x}, t) = \int f(\vec{x}, \vec{v}, t)\, \mathrm{d}\vec{v} \implies \rho(\vec{x}, t) = A m_H n(\vec{x}, t)$$

as well as the concept of *averaged quantities*, for example the average velocity

$$\langle \vec{v} \rangle = \frac{1}{N} \int \vec{u} f(\vec{x}, \vec{u}, t)\, \mathrm{d}\vec{u}$$

This expression is most useful to decompose the velocity as $\vec{v} = \langle \vec{v} \rangle + \delta\vec{v}$, which is known as Reynolds' decomposition. This decomposition implies $\langle \delta\vec{v} \rangle = 0$.

To build the equations of fluid dynamics, we need to ask ourselves what is it of the distribution function f if we let the system evolve in time.

Liouville's theorem tells us that the six dimensional volume in phase space is conserved $\mathrm{d}^3x\mathrm{d}^3v \approx \mathrm{d}^3x_0\mathrm{d}^3v_0$ up to first order[3] in $\mathrm{d}t$. This way we can write the initial infinitesimal number of particles in volume $\mathrm{d}^3x_0\mathrm{d}^3v_0$ and the number of particles in volume $\mathrm{d}^3x\mathrm{d}^3v$

$$\mathrm{d}N_0 = f(\vec{x}_0, \vec{v}_0, t)\, \mathrm{d}^3x_0\mathrm{d}^3v_0$$
$$\mathrm{d}N = f(\vec{x}, \vec{v}, t)\, \mathrm{d}^3x\mathrm{d}^3v$$

Requesting the number of particles to be conserved $\mathrm{d}N_0 = \mathrm{d}N$, we end up with

$$f(\vec{x}_0, \vec{v}_0, t) = f(\vec{x}_0 + \vec{u}_0\, \mathrm{d}t, \vec{v}_0 + \vec{a}_0\, \mathrm{d}t, t) \approx f(\vec{x}_0, \vec{v}_0, t) + \frac{\mathrm{d}f}{\mathrm{d}t}$$

3 Properly speaking, we'd have to consider the transformation $\vec{x} = \vec{x}_0 + \vec{u}_0\, \mathrm{d}t$, $\vec{v} = \vec{v}_0 + \vec{a}_0\, \mathrm{d}t$, which has a Jacobian $\|J\| = 1 + \mathrm{o}(\mathrm{d}t^2)$.

which implies $\mathrm{d}f/\mathrm{d}t = 0$. Writing down what this implies component by component we get

$$\boxed{\partial_t f + v^i \partial_{x^i} f + a^i \partial_{v^i} f = 0} \tag{82}$$

This is called *Collision-less Boltzmann equation* or *Vlasov's equation*. Please, don't bother to ask why of the weird indices placement. Whichever way you put the indices is (here) absolutely irrelevant.

Collisions may be re-introduced by adding a term of the form $\partial_t f_{\text{coll}}$, but in the following we'll assume the detailed balance principle to hold, so that there are as many particles getting kicked out of the volume as those that are pulled in.

5.3 FROM BOLTZMANN TO EULER

Let us consider eq.82 along with Reynolds's decomposition for the velocity $v_i = V_i + u_i$. We have three different terms to evaluate if we integrate over velocities to calculate the momenta of a given quantity $g(\vec{x}, \vec{v}, t)$

$$(i)\ \ \partial_t f \to \int \mathrm{d}^3 v\, \partial_t f\, g(\vec{x}, \vec{v}, t) = \int \mathrm{d}^3 v\, (\partial_t (fg) - f \partial_t g)$$

$$(ii)\ \ v_i \partial_i f \to \int \mathrm{d}^3 v g \vec{v}\, \nabla f = \int \mathrm{d}^3 v\, (\nabla (g \vec{v} f - f \vec{v}\, \nabla g))$$

$$(iii)\ \ a_i \partial_{v_i} f \to - \int \mathrm{d}^3 v f \dot{\vec{v}}\, \nabla_{\vec{v}} g$$

Recalling the definition of the average value of g, the three terms above may be cast in the following form

$$\begin{aligned} &(i) \quad \partial_t (n \langle g \rangle) - n \langle \partial_t g \rangle \\ &(ii) \quad \nabla (n \langle g \vec{v} \rangle) - n \langle \nabla g\, \vec{v} \rangle \\ &(iii) - n \langle \nabla_{\vec{v}} g\, \dot{\vec{v}} \rangle \end{aligned}$$

At this point, consider the following expressions for g

$$g = 1 \implies \text{Mass conservation} \tag{83}$$

$$\partial_t n + \nabla (n \vec{v}) = 0 \tag{84}$$

$$\vec{g} = m \vec{v} \implies \text{Euler's equation} \tag{85}$$

$$\partial_t (\rho \vec{v}) + \vec{v} \cdot \nabla (\rho \vec{v}) = -\nabla p - \rho \vec{F} \tag{86}$$

$$g = \frac{1}{2} m v^2 \implies \text{Energy conservation} \tag{87}$$

$$\partial_t \left(\frac{1}{2} \rho v^2 + \varepsilon \right) + \nabla \left(\rho \vec{v} \left[\frac{v^2}{2} + \varepsilon \right] \right) = \vec{F} \cdot \vec{v} - \nabla (\vec{\phi}_H + p \vec{v}) \tag{88}$$

which are all nice and dandy [4].

Sometimes it is useful to cast the equation of energy conservation in a slightly different form. Called E the total energy, we recall the first law of thermodynamics

$$\frac{\mathrm{d}E}{\mathrm{d}t} = T \frac{\mathrm{d}S}{\mathrm{d}t} + \frac{p}{\rho^2} \frac{\mathrm{d}\rho}{\mathrm{d}t} \tag{89}$$

4 In the last equation, $\vec{\phi}_H$ is the heat flux.

because now we have

$$\partial_t \left(\frac{1}{2}\rho v^2 + \varepsilon \right) + \nabla \left(\rho \vec{v} \left[\frac{v^2}{2} + \varepsilon \right] \right) = \frac{\mathrm{d}E}{\mathrm{d}t} - \rho T \frac{\mathrm{d}S}{\mathrm{d}t} + \Lambda(T, \rho) \tag{90}$$

where Λ includes all energy losses, also the ones not included in Vlasov's equation.

In the following, we'll often use something called the *Lagrangian derivative*, which is just

$$\frac{\mathrm{D}}{\mathrm{D}t} = \partial_t + \vec{v} \cdot \nabla \tag{91}$$

which allows us to rewrite two equations we'll often be using in a more compact form

$$\frac{\mathrm{D}\vec{v}}{\mathrm{D}t} = -\frac{\nabla p}{\rho} - \vec{F} \tag{92}$$

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = -\rho \nabla \cdot \vec{v} \tag{93}$$

From this form of the continuity equation, we can give a definition of *compressibility*: We'll say that a fluid is *incompressible* if $\nabla \cdot \vec{v} = 0$, otherwise we'll say that it's compressible.

So far we've been ignoring the effects of viscosity, which can be actually retrieved if we include collisions in the otherwise collision-less Boltzmann equation. If we were to do that, we would get the celebrated *Navier-Stokes equation*.

We define **streamline** the convolution of the tangents to the velocity field and **streakline** the line traced by all the particles of the fluid that are passing from a specific, fixed point.

Note that the two definitions overlap only if motion is static/stationary.

We'll scarcely make explicit use of these notions but it's worth to write those down, at least for the sake of completeness.

5.3.1 Viscosity and diffusion

So far, we've been blatantly ignoring the collision integral in Vlasov's equation saying that by the detailed equilibrium principle, at equilibrium its contribution is approximately zero. Clearly, this can't be always the case, even more so in fluids.

Notably, fluids display some kind of friction arising from some microscopic molecular effect. We'll call this "friction" *viscosity*[5]. We want to understand what happens when a fluid moves. We have to consider:

- **Dilatation or Compression**: are changes in the volume of the fluid element that can be thought as *isotropic stresses*. Note that in the incompressible fluids they would be absent.
- **Rotation**: we consider a solid body rotation, hence it will cause no net force on the fluid element.
- **Shear**: is the most interacting case, in fact this involves relative motion of different sides of the fluid.

5 For this section, I'm following Dott. Ricciardi's notes [19] and my own notes from the Fluid dynamic class.

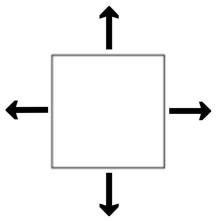

Figure 18: Sketch of dilation or compression.

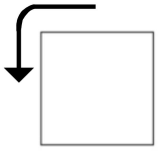

Figure 19: Sketch of rotation.

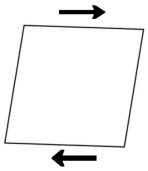

Figure 20: Sketch of shear.

In principle, we could decompose the (purely spatial) stress tensor as follows

$$\sigma_{ij} = -p\delta_{ij} + d_{ij}$$

Here d_{ij} is the *deviatoric stress tensor*, representing the deviation from isotropy. A phenomenological approach may suggest us to consider the deviatoric stress proportional to the local velocity gradient, regarded then as the principal parameter (read as: The more relevant one). To corroborate this approach, we could recall that if a fluid is stationary, velocity gradients vanish, and so would the d_{ij} tensor.

Generally, the deviatoric stress is expressed by means of a fourth rank tensor[6]

$$d_{ij} = A_{ijkl}\partial_l u_k$$

Note that since the stress tensor is symmetric, d_{ij} must be as well. This implies that A_{ijkl} must be symmetric in i, j. After a long and tedious decomposition in symmetric and antisymmetric part, we finally get to the expression for the stress tensor that is customarily used

$$\sigma_{ij} = -p\delta_{ij} + 2\mu\left(\frac{1}{2}\left[\partial_j u_i + \partial_i u_j\right] - \frac{1}{3}\nabla\cdot\vec{u}\delta_{ij}\right) \tag{94}$$

6 It may be worth asking yourself if A is truly a 4-index tensor or just a 4-index symbol.

Here μ is always positive and it's called the viscosity of the fluid. Generally, we'll be interested in the *kinematic viscosity* of the fluid $\nu = \mu/\rho$.

From eq.94 we reckon that if the fluid is uncompressible ($\nabla \cdot \vec{u} = 0$), the Navier-Stokes equation has the form

$$\rho \frac{\mathrm{D}u_i}{\mathrm{D}t} = \rho F_i - \partial_i p + \mu \partial_j^2 u_i \tag{95}$$

hence the viscosity enters the equation with the Laplacian of the velocity, which is often (but definitely not always) negligible. It is often useful to define a quantity called *Reynolds' number*

$$Re = \frac{(\vec{v} \cdot \nabla)\vec{v}}{\nu \nabla^2 \vec{v}} \approx \frac{U \cdot L}{\nu} \tag{96}$$

which quantifies the importance of viscous dissipation in respect to inertial forces. Here U, L are some characteristic scales of the system.

Bernoulli's theorem

If we neglect for the moment dissipative effects in the equation of conservation of energy, we may write

$$\partial_t \left(\frac{1}{2}\rho v^2 + \varepsilon \right) + \nabla \left(\rho \vec{v} \left[\frac{v^2}{2} + \varepsilon \right] \right) - \vec{F} \cdot \vec{v} = 0$$

where ε is some expression for the internal energy of the fluid. If we consider a stationary, incompressible velocity field and only conservative forces

$$\nabla \left(\rho \vec{v} \left[\frac{v^2}{2} + \varepsilon \right] + \rho \phi \right) = 0$$

If we take p/ρ as a measure of the internal energy of the fluid, we obtain *Bernoulli's theorem*

$$\frac{\rho \vec{v}^{\,2}}{2} + p + \rho \phi = \text{const.} \tag{97}$$

which is just a restating of conservation of energy. In the following section we'll try examining one simple application of (97).

5.3.2 The de Laval nozzle

Consider Bernoulli's theorem in the notable absence of an external potential ϕ, and we shall assume that $\vec{v} \cdot \vec{A}$ is constant. Here $\vec{A}$ is the outwards pointing normal to the section of a nozzle of some kind.

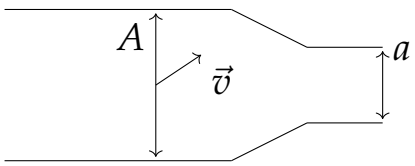

Figure 21: Schematic depiction of the De Lavalle nozzle.

If the fluid is incompressible, Bernoulli's theorem requires that

$$\frac{\rho v^2}{2} + p = \text{const.}$$

which has as a natural consequence that given two sections of areas A and a with velocity $\vec{v}$ at A (see Fig.21), upon reaching the point where the section becomes a, the velocity becomes

$$\vec{V} = \frac{A}{a}\vec{v}$$

You may then be wondering what happens if we relax the condition of the fluid to be incompressible. Consider Euler's equation and the continuity equation (93) in their 1D and stationary form

$$v\partial_x v = -\frac{1}{\rho}\partial_x p \tag{98}$$

$$\partial_x(\rho v) = 0 \tag{99}$$

If we integrate the latter and make use of the divergence theorem, we find out that

$$\int \rho\vec{v}\,\mathrm{d}\vec{A} = \rho\vec{v}\cdot\vec{A} = \text{const.}$$

This can be implicitly differentiated to yield

$$\frac{\mathrm{d}\rho}{\mathrm{d}x}\frac{1}{\rho} + \frac{\mathrm{d}v}{\mathrm{d}x}\frac{1}{v} + \frac{\mathrm{d}A}{\mathrm{d}x}\frac{1}{A} = 0$$

To properly solve the equations, we need to introduce an Equation of State[7] of some sort. Here (and rather often in the following) we'll assume a barotropic relation between pressure and density

$$p(\rho) = c_s^2\rho \tag{100}$$

where c_s is the *speed of sound* in the fluid.

Plugging in what we've just found in the 1D equation of motion

$$\begin{aligned}
\rho v\partial_x v &= -c_s^2\partial_x\rho \\
v\partial_x v &= c_s^2\left(\frac{\mathrm{d}v}{\mathrm{d}x}\frac{1}{v} + \frac{\mathrm{d}A}{\mathrm{d}x}\frac{1}{A}\right) \\
\partial_x v\left(v - \frac{c_s^2}{v}\right) &= \frac{c_s^2}{A}\partial_x A
\end{aligned}$$

This can be cast in a slightly more suggestive form

$$\boxed{\partial_x v = \frac{c_s^2}{Av}\frac{1}{1-\frac{c_s^2}{v^2}}\partial_x A} \tag{101}$$

Consider a nozzle with a position-decreasing section ($\partial_x A < 0$). Two scenarios may arise

7 Sometimes I may be using EoS as a more compact notation.

- (i) the fluid is *subsonic* ($v < c_s$). This implies that the fluid accelerates $\partial_x v > 0$
- (ii) the fluid is *supersonic* ($v > c_s$). This implies that the fluid decelerates $\partial_x v < 0$

If the section gradient is flipped, the implications are reversed. In principle you could then imagine a nozzle with decreasing section up to a certain x_* where the fluid becomes supersonic, and with increasing section after that point.

In such a configuration, the fluid is always accelerating.

5.4 SCHWARTZSCHILD STABILITY CONDITION

Suppose we have a stratified plane parallel atmosphere approximately in hydrostatic equilibrium. Consider now a blob of fluid which has been displaced upwards as shown in the picture below.

Initially, the blob has the same density and pressure as the surrounding fluid, respectively ρ_0, P. After the displacement, the external fluid density and pressure are ρ', P' which are, in principle, different from density and pressure inside the bubble, ρ^*, P^*.

However, we'll work under the assumption that the bubble is always in pressure equilibrium with the surrounding fluid and does not exchange energy with it[8].

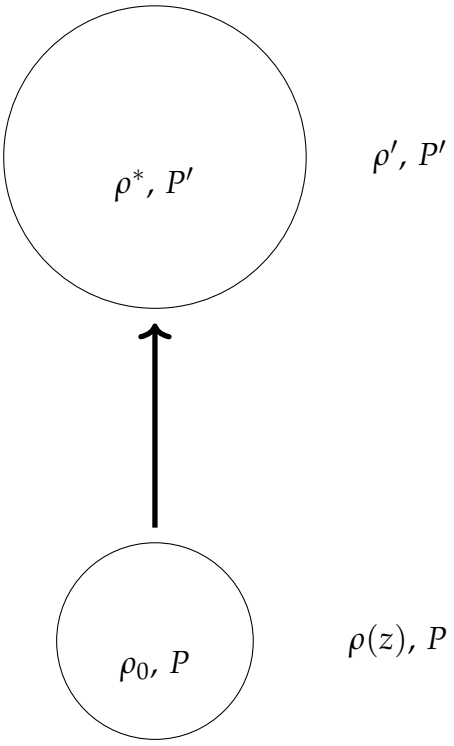

Figure 22: Vertical displacement of a blob of gas in a stratified atmosphere.

8 Typically, pressure imbalances are quickly removed by *acoustic waves*, while heat exchange takes more time. In short, we'll assume $P^* = P'$.

Let's consider the infinitesimal displacement ε, so that the equation of motion is

$$\rho\ddot{\varepsilon} = -g(\rho' - \rho)$$

We shall now take two independent Taylor expansions: One for the density of the blob, one for the density of the surrounding fluid.

$$\rho^* = \rho_0 + (\nabla_{ad}\,\rho)\varepsilon$$
$$\rho' = \rho_0 + (\nabla_{amb}\,\rho)\varepsilon$$

Plugging this two equations into the equation of motion yields a familiar expression

$$\rho\ddot{\varepsilon} = -g(\nabla_{ad}\,\rho - \nabla_{amb}\,\rho)\varepsilon \tag{102}$$

which is that of a harmonic oscillator. We identify the quantity

$$\omega_{BV}^2 = g(\nabla_{ad}\,\rho - \nabla_{amb}\,\rho)$$

as the *Brunt-Väisälä frequency*, which regulates the buyouancy of the blob. In particular we observe that if $\omega_{BV}^2 > 0$, the bubble is buyouantly stable. Conversely, if $\omega_{BV}^2 < 0$ the atmosphere is unstable and *convective instability* will arise.

Up to a factor of g (that would cancel out anyway), studying the sign of ω_{BV}^2 for $\rho \propto T^{-1}$ does retrieve the famous *Schwartzchild stability condition*.

How good is the "bubble picture" we've just used? Not much actually, as we'll see in the next chapter, but it still allows to get a grasp of how convection may work.

One thing we have neglected (but is worth considering when dealing with stars) is that the bubble is likely to be made of hot plasma. That implies that the bubble is able to radiate and affect different layers of the star through means that are not only its physical displacement.

5.4.1 Convection in presence of rotation

We've seen that density gradients in a stratified atmosphere are expected to generate convection. What happens if we add rotation to the picture?

To make things a little simpler, we're going to assume that the star rotates with angular velocity $\vec{\Omega}$ **as a rigid body**. This is a rather poor assumption, that may be true only for the core region of stars; in general, *differential rotation* applies, so $\Omega = \Omega(r, t)$.

As we've already seen, in presence of rotation, the potential of a self-gravitating object must be modified to include the contribution of centrifugal acceleration

$$\phi \to \phi + \phi_C$$

so that when we write down Euler's equation we find

$$\frac{\mathrm{D}\vec{v}}{\mathrm{D}t} + 2\vec{\Omega} \wedge \vec{v} = -\frac{\nabla p}{\rho} + \vec{F} \tag{103}$$

Here $2\vec{\Omega} \wedge \vec{v}$ represents Coriolis' forces[9]. At this point it's convenient to define a new quantity, the *vorticity* as $\vec{\omega} = \nabla \wedge \vec{v}$.

9 Coriolis' forces arise only in presence of latitudinal motion and imply the presence of shear.

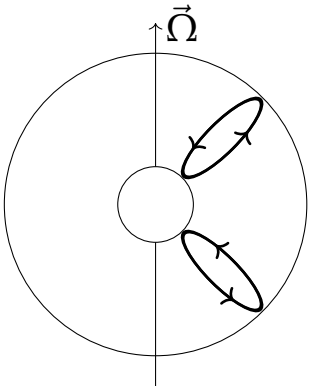

Figure 23: Convective motion in a rotating stellar atmosphere.

In a sense, vorticity represents the amount of angular momentum per unit mass transported by the fluid. If we decompose the stress tensor in a symmetric and antisymmetric part

$$\sigma_{ij} = \sigma_{ij}^S + \sigma_{ij}^A = \frac{1}{2}(\partial_j v_i + \partial_i v_j) + \frac{1}{2}(\partial_j v_i - \partial_i v_j)$$

we recognize that the vorticity generates the antisymmetric part.

We can take the curl of Euler's equation, finding a differential equation for the transport of vorticity

$$\frac{\mathrm{D}\vec{\omega}}{\mathrm{D}t} + \nabla \wedge (2\vec{\Omega} \wedge \vec{v}) = -\frac{1}{\rho^2}\nabla p \wedge \nabla \rho + \nabla \wedge \vec{F} \tag{104}$$

Hence, vorticity is always non-zero in presence of rotation. In hydrostatic equilibrium vorticity is always stable and conserved.

6 TURBULENCE

"When I meet God, I am going to ask him two questions: why relativity? And why turbulence? I really believe he will have an answer for the first."

Werner Heisenberg

Turbulence is quite an interesting feature arising in fluids. Turbulent flow is fluid motion characterized by chaotic changes in pressure and flow velocity, both spatially and temporally.

Turbulence's ability to transport and mix fluid is unparalleled, compared to stardard laminar flows. But it comes with a price: It's bloody difficult to deal with it.

There's not a unique definition of when a fluid becomes turbulent rather than staying in a chill (and much less chaotic) laminar flow.

Usually, we say that a flow is turbulent when the Reynolds' number (96) is much larger than 1. Essentially, what happens is that viscosity is no longer able to damp and constrain the fluid motion, which ends up going nuts. It should be clear, however, that for that to happen, you need energy (ε) to be injected in the fluid.

Before going through the essential features of turbulence, we shall briefly consider again our naive bubble description, only to find out that what happens is actually much more complex.

6.1 INSTABILITIES

Consider the bubble system depicted in Fig.22. As the bubble rises, it will have a non-zero vertical velocity. Let's see what this implies.

Consider a stratified, ideal, incompressible fluid with no vorticity [1] moving with constant velocity along the x axis as shown in Fig.24.

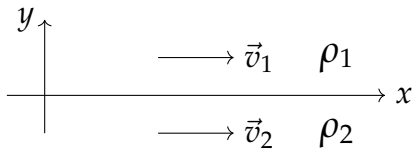

Figure 24: Stratified fluids may give raise to instabilities.

Given these assumptions, the velocity fields may be written in terms of a scalar potential

$$\vec{v}_i = -\nabla \phi_i$$

1 Actually, vorticity is divergent on the boundary.

Plugging this definition in Euler's equation yields (for example for fluid 1)

$$-\nabla(\partial_t \phi_1) + \nabla\left(\frac{v^2}{2}\right) = -\frac{\nabla p}{\rho}$$

Let's assume that the scalar potentials are composed of a constant term (temporally speaking) and of a small space and time dependent fluctuation

$$\phi_i = -v_i x + \delta\phi_i$$

The condition for the fluid to be incompressible then becomes

$$\nabla^2 \delta\phi_i = 0$$

Say that the boundary between the two fluids can be parametrized by $y = \xi(x,t)$, then in the Lagrangian frame we have that

$$\frac{D\xi}{Dt} = \partial_y \delta\phi_1 = -\partial_y \delta\phi_2$$

If we assume the pressure to be continuous at the interface and $\delta\phi_i$ to be plane waves of the form

$$\delta\phi_i = B_i \exp(i(kx - \omega t) + k_y)$$

we're able to find a dispersion relation for the system

$$\frac{\omega}{k} = \frac{\rho_1 v_1 + \rho_2 v_2}{\rho_1 + \rho_2} \pm \left[-\frac{\rho_1 \rho_2 (v_1 - v_2)^2}{(\rho_1 + \rho_2)}\right]^{1/2} \tag{105}$$

Unless $v_1 = v_2$, the argument of the square root is always negative, so the oscillation frequency has a purely imaginary component. When you plug that into the plane wave solution, you get an exponentially increasing amplitude for the wave, meaning that the system is unstable.

This is called *Kelvin-Helmoltz instability*. The KH instability has characteristic vortices arising, which normally are damped at some point by viscosity, stopping the endless formation of vortices.

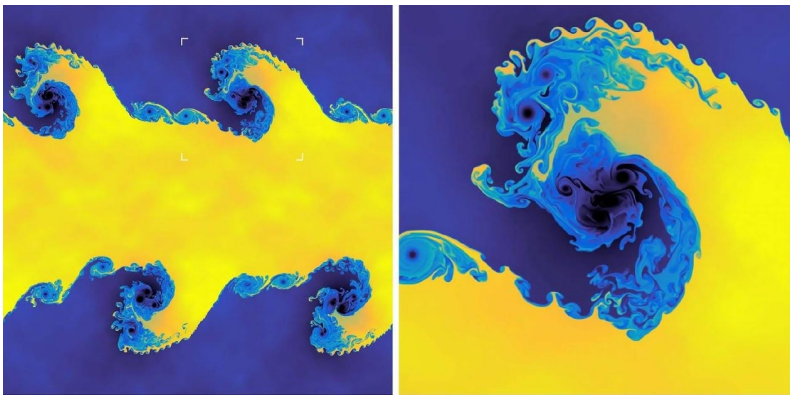

Figure 25: Vortices formation due to KH instability. Note that the process is self-similar. Credits: this video.

If we introduce a gravitational field in the vertical direction, we introduce a stabilizing factor in eq.105

$$\frac{\omega}{k} = \frac{\rho_1 v_1 + \rho_2 v_2}{\rho_1 + \rho_2} \pm \left[\frac{g}{k}\frac{\rho_2 - \rho_1}{\rho_1 + \rho_2} - \frac{\rho_1 \rho_2 (v_1 - v_2)^2}{(\rho_1 + \rho_2)}\right]^{1/2} \tag{106}$$

Please note that even if $v_1 = v_2$, the system may still be unstable. In fact, in such a scenario, the system is stable only if $\rho_1 < \rho_2$ (referring to Fig.24), which actually makes sense: The less dense fluid must float over the denser fluid.

Conversely, the fluid is unstable if $\rho_1 > \rho_2$ (*Rayleigh-Taylor instability*). The RT instability presents characteristic "fingers" (or jets) extending from the less dense layer to the denser one.

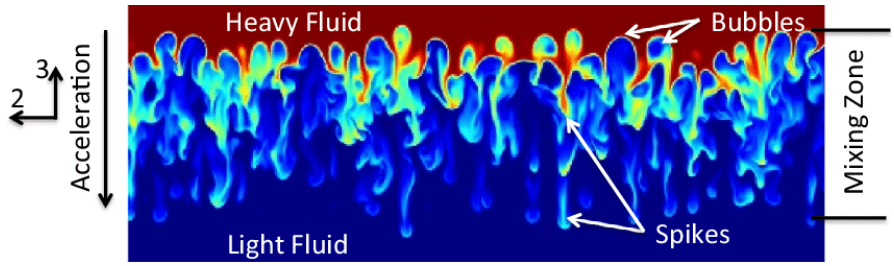

Figure 26: Jet formation due to RT instability. Credits: [21].

6.2 PROPERTIES OF TURBULENCE

To try giving a description of turbulence, we'll often perform a Reynolds' decomposition

$$\vec{v} = \langle \vec{v} \rangle + \delta\vec{v}$$

This allows us to write down an "averaged" version of the Navier-Stokes equation (95), in which a new term is coming out

$$\partial_t \langle v_i \rangle + (\langle v_j \rangle \cdot \partial_j)\langle v_i \rangle = -\frac{\langle \partial_i p \rangle}{\rho} - \partial_i \phi + \nu \partial_j^2 \langle v_i \rangle + \nu \langle \delta v_i \delta v_j \rangle$$

The last term $\langle \delta v_i \delta v_j \rangle$ representing correlations between velocity fluctuations is notably absent in laminar flows. It adds a third source of stress to the stress tensor. If we assume the fluid to be incompressible, then eq.94 becomes

$$\sigma_{ij} = -\langle p \rangle \delta_{ij} + \mu \left(\left[\partial_j \langle v_i \rangle + \partial_i \langle v_j \rangle \right] - \rho \langle \delta v_i \delta v_j \rangle \right)$$

The term containing the pressure is the *isotropic stress*, the term containing the gradients of v is the *viscous stress* while the third new term is the *Reynolds stress*, caused by fluctuations.

Making use of Wiener-Chinčin's theorem, we can link velocity fluctuations to fluctuations in kinetic energy. Usually the Reynolds' stress term is unpacked in a symmetric and an antisymmetric part defining the *turbulent kinetic energy tensor*

$$\kappa(\vec{x}, t) = \frac{1}{2} \langle \delta v_i \, \delta v_i \rangle$$

representing the mean kinetic energy (per unit mass) in the fluctuating velocity field. The *deviatoric anisotropic part* is thus

$$a_{ij} = \langle \delta v_i \delta v_j \rangle - \frac{2}{3} \kappa \delta_{ij}$$

which is the only component effectively transporting momentum. For a more detailed discussion about turbulence, you can check out the Fluid Dynamics course or give a look to [18].

6.2.1 Self-Similarity

Suppose to have a quantity dependent on two independent variables $Q(x, y)$. Characteristic scales $Q_0(x)$ and $\delta(x)$ are defined for Q, y respectively. Then the scaled variables are just

$$\xi = \frac{y}{\delta(x)} \quad \tilde{Q}(\xi, x) = \frac{Q(x, \xi)}{Q_0(x)}$$

If the scaled dependent variable is independent of x, i.e.

$$\exists \hat{Q}(\xi) : \tilde{Q}(\xi, x) = \hat{Q}(\xi)$$

then $Q(x, y)$ is self-similar.

Self-similar fluids have really interesting properties that more often than not are able to greatly simplify what we're working with. Since understanding what a self-similar solution actually is is often more complicated than using it, we'll give two examples.

The first we've already seen: It's eq.75. If you recall, we've defined *scales* for both the density and the radial distance and proceeded to tranform our messy second order ODE in a still messy but at least adimensional second order ODE.

Doing this has the perk that once you've found a solution for $\phi(\eta)$, you already have a solution for whatever density and/or radial distance scale may come to your mind.

Another example, you can find in the Schwartzschild metric for non-rotating, spherically symmetric black holes. General relativity and Birkhoff's theorem grants us that

$$\mathrm{d}s^2 = -\left(1 - \frac{R_s}{r}\right)\mathrm{d}t^2 + \left(1 - \frac{R_s}{r}\right)^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2_{(2)} \tag{107}$$

is the only possible spherically symmetric solution to the Einstein's equation in vacuum [2] (well, assuming there actually *is* something somewhere in spacetime inducing this metric).

From (107) is perhaps even simpler understanding what a self-simlar solution is: Just plug $y = R_s/r$ and you'll find that the metric is completely specified up to a factor of R_s, which contains all the details of the source of the curvature.

6.2.2 Kolmogorov's scales

In a revolutionary article of 1941, the Russian physicist A. V. Kolmogorov developed one of the most incredible theories about the structure of turbulence in incompressible fluids [15].

2 Please note that we're using Carroll's [5] convention for the signature of the metric, which, incidentally, isn't the same used in the GR course. Moreover, I've set $c = 1$.

We shall start assuming we're in presence of a fully turbulent flow with $Re = UL/\nu \gg 1$ and Richardson's theory about the "energy cascade" to hold: That means turbulence can be considered to be composed of eddies of different sizes which can be described through the tern of parameters l, $u(l)$, $\tau(l) = l/u(l)$.

The eddies of the largest scales are characterized by lengthscale l_0 comparable to L (the dimension of the system housing the turbulent liquid) and have characteristic velocity $u(l_0) \approx (2\kappa/3)^{1/2} \approx U$.

Larger eddies are unstable and, like boybands in the prime of their careers, they end up breaking apart, transferring energy to "smaller" eddies. This *energy cascade* continues until the Reynolds' number $Re(l)$ is sufficiently small that the eddy motion is stable.

The energy dissipation rate ε is determined by the transfer of energy from the largest eddies and is **independent of the kinematic viscosity** ν.

Here comes to play Kolmogorov's key hypothesis. At sufficiently high Re, the small-scale turbulent motions are claimed to be *isotropic*. Let's call l_I the lengthscale demarcation between anisotropic large and isotropic small eddies.

Isotropy in this case means that all directional information is lost, as well as all information about the geometry of larger eddies. In a sense then, the statistics of small-scale motions are *universal*.

1st hypothesis of self-similarity

In every turbulent flow with $Re \gg 1$, the statistics of the small-scale range have a universal form determined by ε, ν alone. We can define *Kolmogorov's scale* so that $Re = 1$

$$
\begin{aligned}
l_\eta &= \left(\frac{\nu^3}{\varepsilon}\right)^{1/4} \\
u_\eta &= (\varepsilon\nu)^{1/4} \\
\tau_\eta &= \left(\frac{\nu}{\varepsilon}\right)^{1/2}
\end{aligned}
$$

characterizing the smallest dissipative eddies.

The implications are remarkable. Consider a point $\vec{x}_0$ at a time t_0 for a fully-developed turbulent flow. In terms of Kolmogorov's scales, we can define adimensional coordinates and velocities

$$
\begin{aligned}
\vec{y} &= \frac{\vec{x} - \vec{x}_0}{\eta} \\
\omega(\vec{y}) &= \frac{\vec{U}(\vec{x}, t_0) - \vec{U}(\vec{x}_0, t_0)}{u_\eta}
\end{aligned}
$$

Since it is not possible to form a non-dimensional quantity with ε, ν, the universal form of the statistics of the non-dimensional field $\omega(\vec{y})$ **cannot depend on** ε, ν. That means that on not-too-large scales, $\omega(\vec{y})$ is statistically isotropic and identical at all points.

That implies that all fully-developed turbulent flows have velocity fields statistically similar.

2nd hypothesis of self-similarity

In every turbulent flow with $Re \gg 1$, the statistics of motion of scale l in the range $l_0 \gg l \gg \eta$ have a universal form determined by ε alone, independent of ν, which means that the effect of viscosity is relevant only when we're

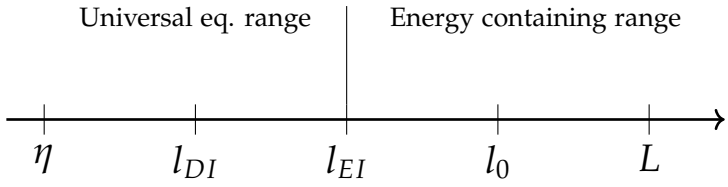

Figure 27: The different lengthscale's interplay in Kolmogorov's theory.

dealing with lengthscales smaller than η. It seems like a honest assumption: A Reynolds' number smaller than 1 (96) requires viscous forces to be greater than the inertial driving. By construction, Kolmogorov's scales identify the conditions for which $Re = 1$, so everything is consistent.

We're not going to show explicitly, but from the two hypothesis of self-similarity we can conclude something really important. In the equilibrium range ($l < l_{EI}$, see Fig.27) the spectrum is a universal function of ε, ν. But then, the second hypothesis of self-similarity tells us that in the inertial subrange statistics cannot depend on ν.

If we were to calculate explicitly velocity correlations to plug them into Wiener-Chinčin's theorem, we'd get an expression for how energy is distributed into modes κ (the wavenumber)

$$E(\kappa) = C\varepsilon^{2/3}\kappa^{-5/3} \tag{108}$$

Here C is a universal constant. What's most fascinating about (108) is that not only it has been tested to be actually true for many types of different fluids, but it is also valid if we relax the condition of incompressibility[3].

The typical velocity fluctuation in a fluid can be evaluated from the root mean square

$$\delta\bar{v} = \sqrt{\langle \delta v^2 \rangle} > 0$$

introducing a non-thermal broadening in the spectrum lines (see "Line Broadening" in Chapter 2). In principle, we have only access to velocity-fluctuations on the line of sight (since that's what we can directly measure in observations), hence the difficulty of recognizing the occuring of turbulence in most objects of astrophysical interest. And that is most frustrating since, for example, the ISM[4] has a Reynolds' number of roughly 10^8.

To actually resolve the spatial structure of turbulence we require a resolution of astrophysical proportions (pun intended), at least proportional to Re^{-1}.

As a concluding note, in the presence of magnetic fields, eq.108 is modified $E_M \propto \kappa^{-2}$. In principle, both dependencies can be present at once, allowing us to differentiate between neutral and magnetized phases.

3 The κ dependence is unchanged, but a dependence on the *Mach number* $\mathcal{M}$ is introduced.
4 i.e. *InterStellar Medium*.

6.3 SOUND WAVES AND SHOCKS

At this point, it should come to no surprise that we're coming back once more to our initial, easy rising-bubble model to further complicate it. We've already seen that as a consequence of its vertical motion, instabilities arise (106), transforming our cool little bubble in a swirly mess.

Another fairly important consequence of the bubble's motion is the arising of pressure imbalances between the material itself and the wake[5] coming after the bubble.

For this section (and many more in the Gravitation Part) I'll be following [10]. Relevant paragraphs for this section are §2.4 and §3.8.

Let's write Euler's equation and the continuity equation (93) in a conservative form

$$\partial_t(\rho\vec{v}) + (\vec{v}\cdot\nabla)(\rho\vec{v}) = -\nabla p$$
$$\partial_t\rho + \nabla(\rho\vec{v}) = 0$$

and look for perturbations around the equilibrium solution.

$$\vec{v} = \vec{v}_0 + \vec{v}_1$$
$$p = p_0 + p_1$$
$$\rho = \rho_0 + \rho_1$$

Since Euler's equation (and by extension, Navier-Stokes' equation) is invariant under Galileian transformations, we can set $\vec{v}_0 = 0$. We shall also assume a one dimensional motion and a barotropic equation of state, allowing us to linearize our equations.

Quantities with the "o" subscript are assumed to be uniform and stationary. Keeping only first order quantities, we find

$$\partial_t\rho_1 + \rho_0\partial_x v_1 = 0$$
$$\rho_0\partial_t v_1 + \left(\frac{\partial p}{\partial\rho}\right)_0 \frac{\partial\rho_1}{\partial x} = 0$$

Cross derivating the two equations and plugging one equation into the other, we end up with

$$\partial_t^2\rho_1 - \left(\frac{\partial p}{\partial\rho}\right)_0 \partial_x^2\rho_1 = 0 \tag{109}$$

which is a hyperbolic equation describing propagating density waves!

We can identify

$$c_s^2 = \left(\frac{\partial p}{\partial\rho}\right)_0$$

as the *speed of sound* in the fluid. Sometimes c_s is also said to be the compressibility of the fluid.

Since c_s is the speed at which pressure (or density) perturbations travel through the gas, it limits the rapidity with which the fluid can react to such changes. For example, if the pressure in one part of a region of the fluid of characteristic size L is suddenly changed, the other parts of the region

5 *N.d.T.*: in Italian, it should sound something like "scia".

cannot respond to this change until a time of order L/c_s, the sound crossing time, has elapsed.

In a sense, c_s is also the velocity at which causality travels in the fluid. Thus, if we were to consider a *supersonic flow*, then the fluid wouldn't be able to respond on the flow time $L/|\vec{v}|$, so pressure gradients have little effect on the flow. At the other extreme, for subsonic flows the fluid can adjust in less than the flow time, so to a first approximation the fluid behaves as if in hydrostatic equilibrium.

The density dependence of the sound speed implies that regions of higher than average density have higher than average sound speeds, a fact which gives rise to the possibility of *shock waves*.

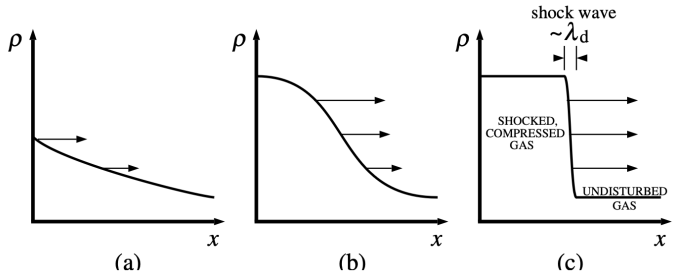

Figure 28: The formation of a shock wave. Credits: Frank, King and Raine.

In a shock, relevant fluid quantities change on lengthscales of the order of the mean free path and this is represented as a discontinuity in the fluid.

6.3.1 Rankine-Hugoniot junction conditions

Since discontinuities are introduced into the picture, we can no longer use differential quantities to describe conservations. The solution to this problem is easily found looking into *integral quantities*.

To do so, it is convenient to choose a reference frame in which the shock is at rest. Since the shock thickness is small, we can regard it as locally plane; moreover, since the fluid flows through the shock so quickly, changes in the fluid conditions cannot affect the details of the transition across the shock, so we can regard the flow into and out of the shock as steady.

All we have to do is apply the various conservation laws in their integral form

$$\rho_1 v_1 = \rho_2 v_2 \tag{110}$$

$$p_1 + \rho_1 v_1^2 = p_2 + \rho_2 v_2^2 \tag{111}$$

$$\frac{1}{2} v_1^2 + \varepsilon_1 = \frac{1}{2} v_2^2 + \varepsilon_2 \tag{112}$$

which are respectively mass flux, momentum flux and energy flux conservation. Especially in the latter, we've assumed radiative losses, thermal conduction, etc., across the shock front to be negligible.

This is the adiabatic assumption, and the resulting junction conditions are known as the adiabatic shock conditions or *Rankine-Hugoniot junction conditions*.

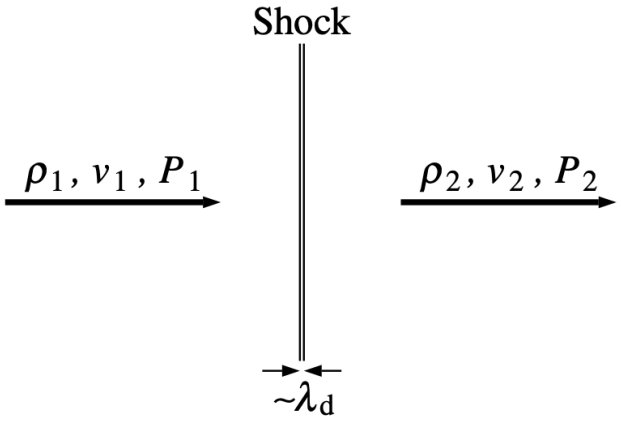

Figure 29: Calculation of the Rankine-Hugoniot junction conditions. Credits: Frank, King and Raine.

Keeping Fig.29 in mind, we can proceed to find a unique relation connecting the ratio of the downstream density (or velocity) ρ_2 and upstream density to the pressure

$$\frac{\rho_2}{\rho_1} = \frac{v_1}{v_2} = \frac{(\gamma+1)p_1 + (\gamma-1)p_2}{(\gamma-1)p_1 + (\gamma+1)p_2}$$

Note that it is not assured for the adiabatic coefficient γ to come out unscathed from the shock, even more so if it is energetic enough to ionize the fluid.

Usually, RH junction conditions are expressed in terms of the *Mach number* $\mathcal{M}_i = v_i/c_{s,i}^{\mathrm{ad}} = (3\rho v^2/5p)^{1/2}$. For a monoatomic gas $\gamma = 5/3$, the RH junction conditions yield

$$1 + \frac{v_2}{v_1} = \frac{5}{4}\left[\frac{3}{5\mathcal{M}_1^2} + 1\right]$$

In the limit of a strong shock $\mathcal{M}_1 \gg 1$[6], we find out that

$$\frac{v_2}{v_1} = \frac{\rho_1}{\rho_2} = \frac{1}{4}$$

i.e. the velocity drops to one-quarter of its upstream value across the shock while the gas is compressed by a factor 4 in a strong shock.

For more general gases, assuming adiabatic compression, the limit compression rate is

$$\frac{v_2}{v_1} = \frac{\rho_1}{\rho_2} = \frac{\gamma-1}{\gamma+1}$$

If the shock is *isothermal* rather than adiabatic, there's no limit to compressibility.

Note that without proper injection of energy, shocks are eventually stalled by viscosity, then degenerating into normal soundwaves. As we have anticipated in the last section, Kolmogorov's spectrum has the same κ dependency,

6 Here $\mathcal{M}$ is the upstream Mach number.

but a proportionality to $\mathcal{M}^8$ is introduced, meaning that shocks dissipate energy very efficiently[7].

Most of the times, assuming shocks to be adiabatic is questionable *at best* since there's a strong dependence on the cooling timescales through radiative losses.

6.3.2 Sedov-Taylor blastwave solution

There's a (arguably) cool and educational application of shockwaves, which incidentally made the fortune of a rather famous film producer.

Explosions.

An explosion is essentially a localized injection of a given quantity of energy that induces an adiabatic expansion of the gas (or the fluid). To find a proper description of explosions we're one theorem short, which we'll briefly discuss in the following paragraphs.

Buckingham-Pi theorem

Assume to have a set of n variables so that $q_1 - f(q_2, ...q_n) = 0$. That means we can define a function g that satisfies

$$g(q_1, ...q_n) = 0$$

Given a relation like the one above, the independent variables of the system can be reduced to $n - m$ with m typically of the order of the independent dimensions (called Π).

The theorem then claims that we may pass from function $g(q) = 0$ to $G(\Pi) = 0$ following this procedure

- List all parameters describing the system;
- Choose the primary dimensions (e.g.: M, L, T and so on);
- Perform dimension analysis of the parameters;
- Choose parameters that contain all the primary dimensions;
- Set dimensional equations and find the explicit relations for the Π parameters.

On top of helping us create a model for explosions, this allows us to cast Euler's equation into a self-similar ODE form which can be solved much more easily (numerically, of course). Defining $\eta \equiv r^\lambda t^{-\mu}$, λ and μ will be set by the particular constrains we have, but in general we can express

$$\begin{aligned} v &= rt^{-1}U(\eta) \\ \rho &= r^{-3}D(\eta) \\ p &= r^{-1}t^{-2}\Pi(\eta) \\ c_S &= rt^{-1}C(\eta) \end{aligned}$$

7 I tried looking for an article or something talking about this, but found none. Or, at least, I found none that convinced me. If you were to find something, please contact me so I can reference it.

so that relevant derivatives take the following form

$$\frac{\partial}{\partial t} = -\mu\eta t^{-1}\frac{\mathrm{d}}{\mathrm{d}\eta}$$
$$\frac{\partial}{\partial r} = \lambda\eta r^{-1}\frac{\mathrm{d}}{\mathrm{d}\eta}$$

With this set of equations we can cast Euler's, continuity and energy conservation equations in a self-similar form as anticipated

$$(\lambda U-\mu)\eta\frac{\mathrm{d}D}{\mathrm{d}\eta} + (n-2)DU + \lambda D\eta\frac{\mathrm{d}U}{\mathrm{d}\eta} = 0$$
$$(\lambda U-\mu)\eta\frac{\mathrm{d}D}{\mathrm{d}\eta} = D^{-1}\left(\lambda\eta\frac{\mathrm{d}\Pi}{\mathrm{d}\eta} - \Pi\right) - U(U-1)$$
$$(\lambda U-\mu)\eta\frac{\mathrm{d}(\Pi D^{-\gamma})}{\mathrm{d}\eta} + ((3\gamma-1)U-1)\Pi D^{-\alpha} = 0$$

ODEs have the nice convenience of being generally easier to solve than PDEs.

As promised, here comes at long last the description of blastwaves.

The relevant parameters are just three if we assume spherical symmetry

$$[\rho] = \mathrm{M}\,\mathrm{L}^{-3} \quad [t] = \mathrm{T} \quad [E_0] = \mathrm{M}\,\mathrm{L}^2\,\mathrm{T}^{-2}$$

which means our η parameter is just $\eta = r^5 t^{-2}$, since $[E_0]/[\rho] = \mathrm{L}^5\,T^{-2}$. Hence, simply inverting the expression yields

$$R \sim \left(\frac{E_0}{\rho_0}\right)^{1/5} t^{2/5} \qquad v = \dot{R} \sim \frac{2}{5}Rt^{-1} \tag{113}$$

incidentally, the proportionality constant is of order unity. Note that the velocity is decreasing, since the acceleration $\ddot{R} < 0$. This can be roughly explained recalling the conservation of momentum: As the wavefront expands it "gains" more mass [8] hence the need to slow down to conserve momentum.

So far we've been assuming the process to be adiabatic, but is it actually a sensible guess? It certainly depends on timescales: If the cooling timescale is much larger than the dynamical timescale, then we're good.

On top of that, especially in the first instants of explosions, the medium can be considered *optically thick* thanks to the density compression induced by shockwaves (for this, you might want to recall that $\mathrm{d}\tau_\nu = \alpha_\nu\,\mathrm{d}s$ and the absorption coefficient can be expressed in terms of the density and the opacity of the system as $\alpha_\nu = \rho\kappa_\nu$, hence the conclusion).

Since no radiation is coming out of there, we can assume that no energy is lost to radiation and is conserved (see Fig.30).

We can follow a similar line of reasoning and find out that the pressure of the blastwave scales as [9]

$$p \sim E_0 R^{-3}$$

That means that there is a radius for which every pressure is reached. That holds until the shockwave's pressure reaches ambient pressure [10] and equilibrium is established. When that happens, the shock stalls and propagates with constant velocity, usually some multiple of the sound speed.

8 The proper expression would probably be "entrains more material".

9 You just have to recall that pressures are dimensionally equal to energy densities.

10 Which is fated to happen sometime. As the blastwave expands, pressure continues to drop.

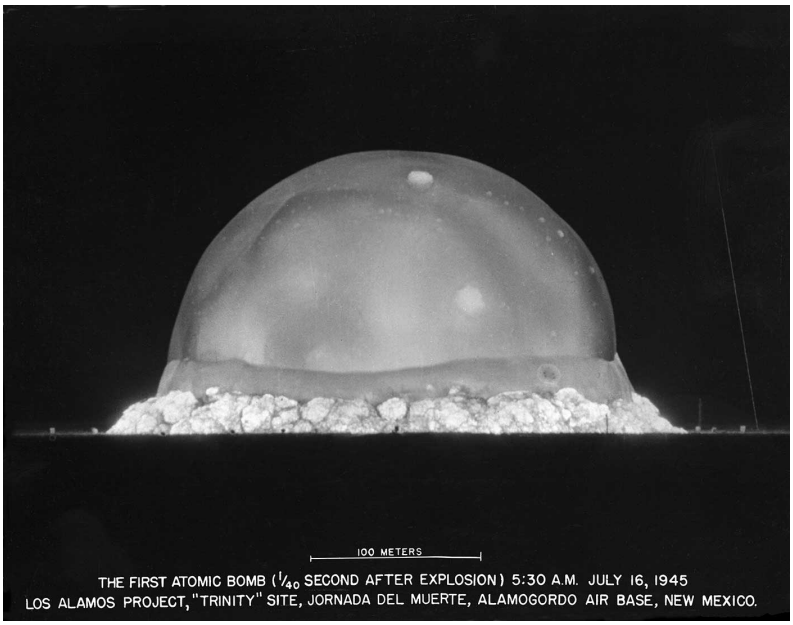

Figure 30: A real photograph portraying the first instants of the Trinity atomic test conducted in 1945 in New Mexico, Alamogordo.
Note how, after $1/40$ of a second, the blastwave still retains (partial) spherical symmetry and the medium is still optically thick.

As the shock expands, it cools down, hence E is not constant anymore. This causes the shock to slow down and *entrain* material in its course.

Eventually, the expansion will have to become subsonic due the arising of viscous forces, which we had completely ignored about in our model. But that was not without a logic. Viscous forces enter at second order in velocity, so in a first order model they can be initially neglected.

Since E is not conserved anymore, the evolution is governed by conservation of momentum, and the expansion law becomes

$$R \sim \left(\frac{MV}{\rho}\right)^{1/4} t^{1/4} \tag{114}$$

this is usually called the *Snowplow phase*.

The expression we've found to describe the Sedov-Taylor blastwave (113) also allows us to predict quite accurately the evolution of the expansion of Supernovae explosions. For this highly energetic phenomena, we can consider an initial injection of energy of order $E_0 \sim 10^{50}\,\mathrm{erg}$ and a typical density of $\rho_0 \sim 10^{-24}\,\mathrm{g/cm^3}$. This means that the typical radius of the explosion is given by

$$R(t) \sim 10^{-0.7} \left(\frac{E_{50}}{\rho_{24}}\right)^{1/5} t_{yr}^{2/5} \ [\mathrm{pc}]$$

where the subscripts indicate that quantities are to be expressed in terms of the typical order of magnitude for SN events. So for example, assuming a radius of 1 parsec after a 100 years long expansion, you'd get a typical velocity for the wavefront of order $v \sim 10^3\,\mathrm{km\,s^{-1}}$.

What have we learned from this section?

If you (really) have to test a new bomb (and you're supposed to keep it a secret), don't go around showing images of the explosion with convenient space and time references.

Part III

GRAVITATION

7 ACCRETION PHYSICS

Everything we've seen in the last few chapters will come into play here.

In this chapter we're considering the two main possibilities for accelerated plasma in presence of a gravitational field: It's either they receive a kick and then fall back on the surface (*accretion*) or just make it out safely from the gravitational field (*stellar winds*).

In the next sections, we'll try to give a punctual and precise description so to not underplay the important physical processes lying underneath.

7.1 STELLAR WINDS

Starting from last century it had become clearer and clearer that an outgoing, corpuscolar flow from the Sun was needed to explain certain phenomena, like cometary tail's disruptions and Earth's geomagnetic storms.

In order to start approaching this subject, let us approximate the Sun (or a star, in general) as an autogravitating sphere of mass M with spherical symmetry, ignoring whatever drift may be caused by magnetic fields or rotation. We're also going to neglect the self-gravitating potential of the stellar corona hovering around the central star.

Following the original article by Parker [17] we may first discuss whether the stellar corona can be in hydrostatic equilibrium at all distances. The easy answer is that it (probably) cannot. If we were to do the full computation,

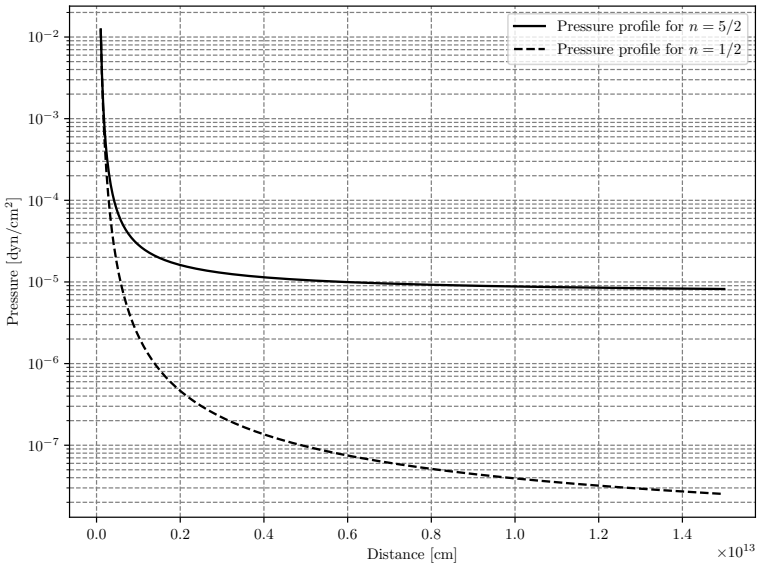

Figure 31: Pressure profile under the assumption of hydrostatic equilibrium.

we'd find a pressure profile like the one in Fig.31, which presents asymptotic values for both cases considered (neutral and ionized hydrogen). Such asymptotic pressure is much larger than any other source of pressure we may expect finding out there. For example, the pressure exerted by the ISM, the only sensible and suitable candidate, is about five orders of magnitudes less than what would be needed to mantain hydrostatic equilibrium at all distances.

Stellar coronas must then be expanding. Whether inwards or outwards that remains to be seen.

Since it is not possible to find a stable solution for the coronal gas to be in hydrostatic equilibrium at large distances from the central star, we now turn our eye towards a *dynamical model* able to describe the process.

The starting point are as always Euler's equation and the continuity equation, both expressed under the assumption of spherical symmetry and *stationarity*, so that we can put all time derivatives to zero. Eq.95 may be then written as

$$\rho v \frac{\mathrm{d}v}{\mathrm{d}r} = -\frac{\mathrm{d}p}{\mathrm{d}r} - \frac{G\rho M}{r^2} \tag{115}$$

The continuity equation in spherical symmetry yields

$$\frac{d}{dr}(r^2 \rho v) = 0 \tag{116}$$

which may be solved exactly

$$\dot{M} = 4\pi r^2 \rho v(r) = \text{const.} \tag{117}$$

Assuming a barotropic equation of state (100)[1]:

$$v^2 \partial_r \log(v) = -c_s^2 \partial_r \log(\rho) - \frac{GM}{r^2}$$

which can be promptly restated as

$$(v^2 - c_s^2)\partial_r \log(v) = \frac{2c_s^2}{r}\left(1 - \frac{GM}{2c_s^2 r}\right)$$

In this form it's easy to see that if $v > c_s$ the fluid must be accelerating[2].

It may be worth pointing out that the latter equation may be cast in a self-similar fashion. This is easier seen if we assume the plasma to behave like a perfect gas: Under the assumption of constant temperature in the region we're observing, the equation of state is barotropic

$$p = nkT = \frac{kT}{\mu m_H}\rho$$

where μ is the mean molecular weight

$$\mu = \left(2X + \frac{3}{4}Y + \frac{1}{2}Z\right)^{-1}$$

1 This is not a bad assumption: The energy transported by the wind is much greater than that transported by heat conduction mechanisms [25]. We can thus assume temperature to be negligible beyond a given distance b and constant up to that point.

2 It dawned on me that stellar winds are much similar to a spherically symmetric de Lavalle's nozzle (Fig.21) with increasing section (the spherical shells). I'm pretty sure this was mentioned during class, though.

Expressed in this fashion, the sound speed c_s is simply equal to

$$c_s = \left(\frac{\mathrm{d}p}{\mathrm{d}\rho}\right)^{1/2} = \left(\frac{kT}{\mu m_H}\right)^{1/2} \approx \left(\frac{kT_0}{\mu m_H}\right)^{1/2} \tag{118}$$

at least to leading order. Plugging all this in the original expression for Euler's equation reads something like

$$\partial_\xi \psi \left(1 - \frac{\tau}{\psi}\right) = -2\xi^2 \partial_\xi \left(\frac{\tau}{\xi^2}\right) - \frac{2\lambda}{\xi^2} \tag{119}$$

This last equation was obtained setting $\xi = r/a$, $\tau = T(r)/T_0$, $\lambda = GMm_H/2kT_0a$, $\psi = m_H v^2/2kT_0$, where we also assumed the corona to be made of (ionized) Hydrogen only $\mu \approx 2$[3].

To integrate (119) we shall assume that temperature is uniform and equal to T_0 in the range $r \in (a, b)$, with b a distance beyond which heating mechanism are negligible. The interesting physics is in the region $r < b$.

Here temperature is constant, so $\tau = 1$ and the ODE is immediately solved

$$\psi - \ln(\psi) = \psi_0 - \ln(\psi_0) + 4\ln(\xi) - 2\lambda\left(1 - \frac{1}{\xi}\right) \tag{120}$$

where the integration constant has been chosen so that $\psi(\xi = 1) = \psi_0$.

Discarding all non-physical solutions (negative magnitudes and complex solutions) requires $\lambda > 2$. Once you've got rid of all the unwanted nasty solutions, the final expression is something like

$$\psi - \ln\psi = -3 - 4\ln\frac{\lambda}{2} + 4\ln\xi + \frac{2\lambda}{\xi} \tag{121}$$

which is plotted for different values of temperature T_0 in Fig.32.

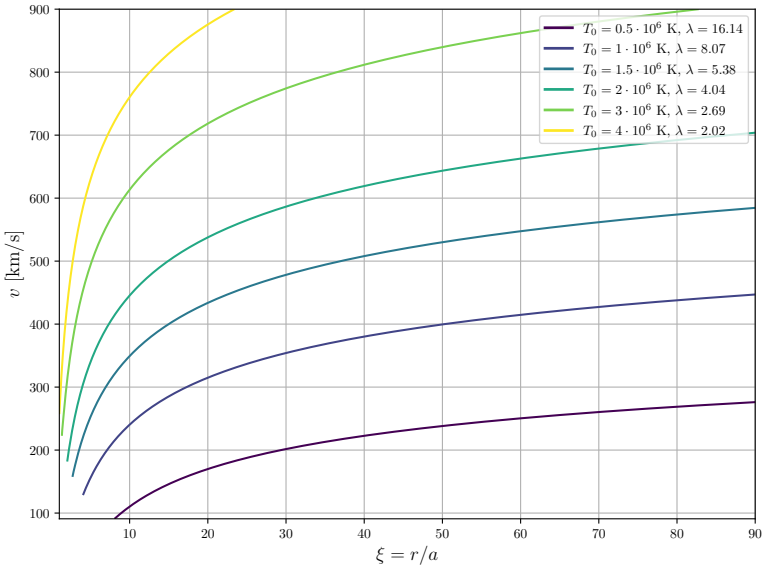

Figure 32: Velocity profile for eq.121 for different values of T_0.

3 Please be warned that I tried to adapt the notation I've used for my Batchelor thesis to the one of these notes. Please forgive me if there's some subscript not *subscripting* properly.

Closed this self-complacent excursus, let's turn back to the expression shown in class

$$(v^2 - c_s^2)\partial_r \log(v) = \frac{2c_s^2}{r}\left(1 - \frac{GM}{2c_s^2 r}\right)$$

If $v > 0$ we have at first bounded motion, pretty much like an oscillation of some sort, but then, as the fluid accelerates, it gets closer and closer to the escape velocity [4] until the fluid finally escapes.

But what if the the fluid is *ingoing* rather than outgoing? We won't have winds, proper, but rather *spherical accretion*, which is discussed in the next section.

7.2 SPHERICAL ACCRETION

We consider a star of mass M accreting spherically symmetrically from a large gas cloud. This would be a reasonable approximation to the real situation of an isolated star accreting from the interstellar medium, provided that the angular momentum, magnetic field strength and bulk motion of the interstellar gas with respect to the star could be neglected.

We'll follow the description given by Bondi, Hoyle and Lyttleton [13], [4] as presented by [10].

Recall now the expression we've found for stellar winds

$$(v^2 - c_s^2)\partial_r \log(v) = \frac{2c_s^2}{r}\left(1 - \frac{GM}{2c_s^2 r}\right)$$

We'll now cast it in a slightly different form that makes more evident the relationships between the various elements interplaying

$$\frac{1}{2}\left(1 - \frac{c_s^2}{v^2}\right)\frac{\mathrm{d}(v^2)}{\mathrm{d}r} = -\frac{GM}{r^2}\left[1 - \frac{2c_s^2 r}{GM}\right] \tag{122}$$

As we have mentioned at the end of last section, accretion can be considered as a stellar wind process with negative velocity. This means that we can write down an expression for the (constant) accretion rate

$$\dot{M} = 4\pi r^2 \rho(-v) = \text{const.}$$

First, we note that at large distances from the star the factor $\left[1 - \frac{2c_s^2 r}{GM}\right]$ on the right hand side must be negative, since c_s^2 approaches some finite asymptotic value $c_s^2(\infty)$ related to the gas temperature far from the star, while r can increase, in principle, without limit. This means that for large r the right hand side of (122) is positive. On the left hand side, the factor $\frac{\mathrm{d}(v^2)}{\mathrm{d}r}$ must be negative, since we want the gas far from the star to be at rest, accelerating as it approaches the star with r decreasing.

Hence, for large values of r, the fluid must be *subsonic*. This is, of course, a very reasonable result, as the gas will have a non-zero temperature and hence a non-zero sound speed far from the star (118).

4 Note that in this regard, thermal fluctuations of the velocity often give the plasma the needed bump to escape the gravitational pull.

As made obvious from the equations, as the gas approaches the star, r is decreasing and the expression in square brackets of (122) must tend to increase, eventually reaching zero.

Such condition is verified when

$$r_s = \frac{GM}{2c_s^2} \simeq 7.5 \cdot 10^{13} \left(\frac{T(r_*)}{10^4\,\mathrm{K}} \right)^{-1} \left(\frac{M}{M_\odot} \right) \mathrm{cm} \tag{123}$$

which is curiously resemblant of the Schwartzschild radius. Note that the order of magnitude of r_s is about 100 times that of the solar corona, which is much larger than the radius R_* of any compact object. Technically, a very high temperature would be needed in order to make $r_s < R_*$. This condition can be achieved, for example, if we consider a standing shock wave close to the stellar surface.

If we perform a similar analysis of the signs of eq.122, we find out that near the star the gas must be *supersonic*.

The existence of a point described by (123) is of great importance in characterizing the accretion flow. The direct mathematical consequences are two

$$v^2 = c_s^2 \quad r = r_s \tag{124}$$

$$\frac{\mathrm{d}(v^2)}{\mathrm{d}r} = 0 \quad r = r_s \tag{125}$$

so that all solutions of (122) can be classified by their behavior at r_s, given by two expressions above, together with their behaviour at large r. This behaviour can be summarized in a (mildly confusing) plot as shown in Fig.33.

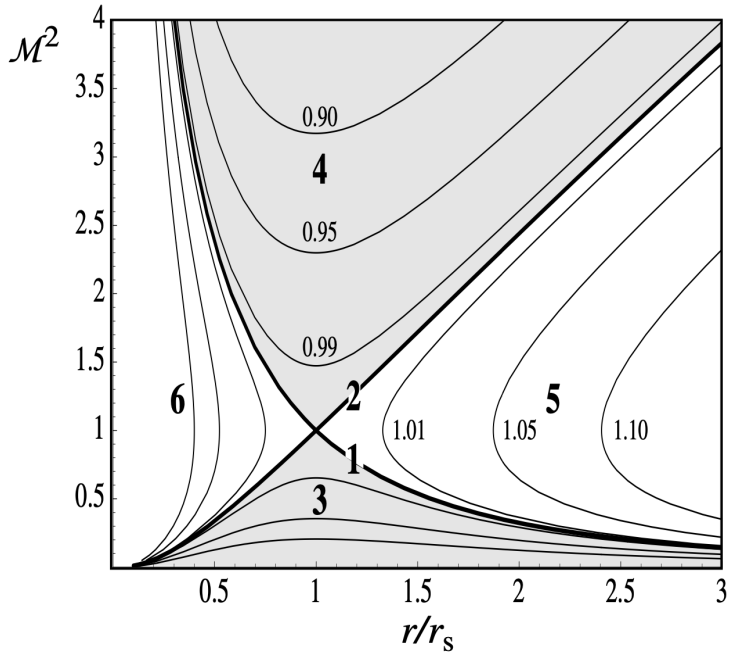

Figure 33: Mach number squared $\mathcal{M}^2 = v^2(r)/c_s^2(r)$ as a function of radius r/r_s for spherically symmetrical adiabatic gas flows in the gravitational field of a star. Credits: Frank, King and Raine [10].

An exact description yields six families of solutions, but such precision is beyond the intentions of these notes. The interested reader can refer to [10] for a more detailed discussion.

Excluding "unwanted" solutions leaves us with solutions of the form

$$v^2(r_s) = c_s^2(r_s) \quad v^2 \to 0 \text{ as } r \to \infty \quad (v^2 < c_s^2,\ r > r_s;\ v^2 > c_s^2,\ r < r_s)$$

which allows us to integrate directly Euler's equation and evaluate some nice quantities which, although fairly interesting, we'll skip right away. I'll just point out that such solution allows us in principle to infer that accretion from the interstellar medium is unlikely to be an observable phenomenon.

It should be pretty obvious that, as the gas infalls into the star, it builds up energy. This process is eventually brought to a stop if the star has a *hard boundary* at, say, $r = \bar{r}$. The sheer variation in gravitational potential is (assuming gas starting from infinity) $\Delta\phi = GM/r$.

The energy dissipated, either by radiation or by the gas heating up, if the gas bounces on the surface of the star is $\dot{e} = GM/r$; if the gas doesn't stop, the Virial theorem yields $\Delta\phi = GM/2r$.

Let us assume for the moment adiabatic accretion. The internal energy of the gas could then be expressed as

$$e = \frac{p}{(\gamma - 1)\rho} \quad p = \frac{\rho RT}{\mu m_H}$$

After complete energy dissipation, we can use the Virial temperature to write

$$T = \frac{1}{2}(\gamma - 1)T_{vir}$$

The *Virial temperature* is defined as the temperature that accreted material would reach if its gravitational potential energy were turned entirely into thermal energy

$$T_{vir} = \frac{GM\mu}{R\bar{r}} \tag{126}$$

Since the sound speed of a perfect gas can be written as $c_s = (\gamma RT/\mu)^{1/2}$, this implies that at a temperature $T = T_{vir}$, the sound speed is roughly equal to the escape velocity from the star. In this scenario, accretion is not possible, since density fluctuations can escape the gravitational pull. Similarly, if the gas is too hot, accretion stops[5]. In a sense, accretion is a *self-regulating process*.

What if we consider now other mechanisms to dissipate energy?

First and foremost, it's safe to assume that in most scenarios the temperature will always be such that $T \ll T_{vir}$. Still, accretion can't go on for an indefinite amount of time. Eventually it will have to stop.

If we increase the accreting velocity

$$\dot{M} = -4\pi\rho r^2 v \quad (v < 0)$$

the variation of optical depth in respect to the rate of accretion will be positive $\mathrm{d}\tau/\mathrm{d}\dot{M} > 0$, thus there must be a critical accretion rate ($\dot{M}_c$) beyond

5 When the temperature raises to or over the virial temperature the sound's speed becomes roughly equal to the escape velocity and so the gas cannot be pulled inward.

which photons can't escape, making the process of accretion always adiabatic.

Consider a medium with opacity κ_ν and flux F_ν. The force exerted by photons will approximately be proportional to $\kappa_\nu F_\nu / c$, and will be the main contribution opposing to accretion. From here we can define an "equilibrium" flux for which gravity is balanced by photon pressure alone

$$F_{\nu,eq} = \frac{c}{\kappa_\nu} \frac{GM}{r^2}$$

Hence, under the assumption of spherical symmetry, we can define *Eddington's critical luminosity* as

$$L_E = 4\pi r^2 F_{\nu,eq} = \frac{4\pi c G M}{\kappa_\nu} \tag{127}$$

Eddington's luminosity plays a crucial role in accretion processes, for it defines a limiting luminosity beyond which accretion can't take place.

If we assume electrons to play the major contribution to the luminosity, the opacity is entirely dominated by Thomson scattering, and Eddington's luminosity is roughly equal to $L_E \simeq 4 \cdot 10^4 M / M_\odot L_\odot$.

Please note, however, that **accretion is seldom spherically symmetric**. In these cases, is much more convenient defining a Eddington's accretion rate rather than a luminosity. Since the luminosity of an accreting flow can be written as

$$L = \frac{GM\dot{M}}{R}$$

Eddington's accretion rate will then be

$$\dot{M}_E = \frac{4\pi R c}{\kappa} \tag{128}$$

that plays as the boundary between radiation dominated and advection dominated accretion flows (ADAF).

In the next chapter, we'll take a closer look as to what happens when we include non-zero angular momentum into the picture.

8 ACCRETION IN BINARY SYSTEMS

8.1 INTRODUCTION

Binary systems of stars (or black holes, as we'll see in the next chapter) are one of the most common configurations, were we take a good look at the sky (that is, assuming we're able to optically resolve the binary system).

On top of that, differently from their single counterparts, binaries reveal more about themselves, notably their masses and dimensions, than do other astronomical objects; this is particularly true in the case of *eclipsing binaries*.

It may come as no surprise hearing that angular momentum is going to play a major role in the study of accretion in binary systems. For example, in many cases, the transferred material cannot land on the accreting star until it has rid itself of most of its angular momentum, leading to the formation of *accretion disks*, which we'll be the focus of quite a few sections among the following.

8.2 INTERACTING BINARIES

As for last chapter, I'll be referring to [10] for this part as well.

There are two main reasons many binaries transfer matter at some stage of their evolutionary lifetimes:

- (i) one of the stars in a binary system may increase in radius (for example because it's entering the *giant phase* of its evolution), or the binary separation shrink (this will be treated more carefully in the last chapter of this notes), to the point where the gravitational pull of the companion can remove the outer layers of its envelope;
- (ii) one of the stars may eject much of its mass in the form of a stellar wind (again, perhaps because its a particularly massive giant star); some of this material will be captured gravitationally by the companion.

We'll focus on the (i) scenario.

8.3 ROCHE LOBES

We shall now try to understand what is it of a small test particle living in the combined gravitational potential of the two stars that make the binary, orbiting each other under the influence of their mutual gravitational attractions. The two bodies of the binary are assumed to be so massive (in respect to the test particle) that the test particle has no influence over their orbits.

Assuming this to hold, the system is reduced to a Physics 1 problem of two bodies interacting with a radial potential. So we already know the solution: The two stars execute Kepler orbits about each other in a plane.

As a further simplification, we'll assume these orbits to be circular. This is usually a good approximation for binary systems, since tidal effects tend to circularize originally eccentric orbits on timescales short compared to the time over which mass transfer occurs.

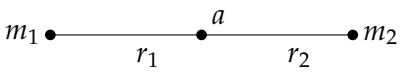

Figure 34: Crude depiction of two objects of mass m_1, m_2 orbiting around their common center of mass.

Consider a binary system as crudely depicted in Fig.34. For the purposes of this section, we're going to assume that the two stars of the binary can be assumed to be point-like sources of potential with separation a.

That being said, the binary period[1] is inferred from Kepler's third law

$$\Omega^2 = \frac{G(m_1 + m_2)}{a^2} \tag{129}$$

A more aesthetically pleasing version of Fig.34 is given below.

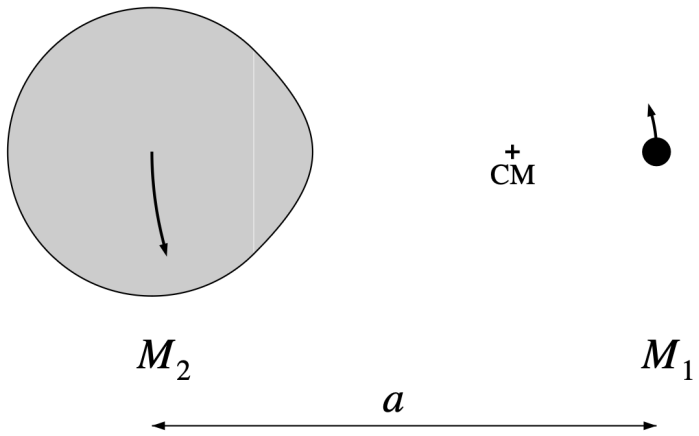

Figure 35: A binary system of two stars orbiting around their common center of mass. Credits: Frank, King and Raine.

In the corotating frame, the potential generated from the system is

$$\phi(\vec{r}) = -\frac{Gm_1}{|\vec{r} - \vec{r}_1|} - \frac{Gm_2}{|\vec{r} - \vec{r}_2|} - \frac{1}{2}(\vec{\Omega} \wedge \vec{r})^2$$

where $\vec{r_1}$, $\vec{r_2}$ are the position vectors of the centres of the two stars.

It's clear that any gas flow between the two stars we should be able to describe, in principle, through Euler's equation, which will be of the form

$$\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla) \cdot \vec{v} = -\frac{\nabla p}{\rho} - \nabla\phi - 2\vec{\Omega} \wedge \vec{v}$$

1 Here we're using the angular frequency rather than the period. It's the same up to some power of 2π.

Considerable insight about accretion problems can be gained if we consider the equipotential surfaces of the the gravitational potential, and in particular their section on the orbital plane.

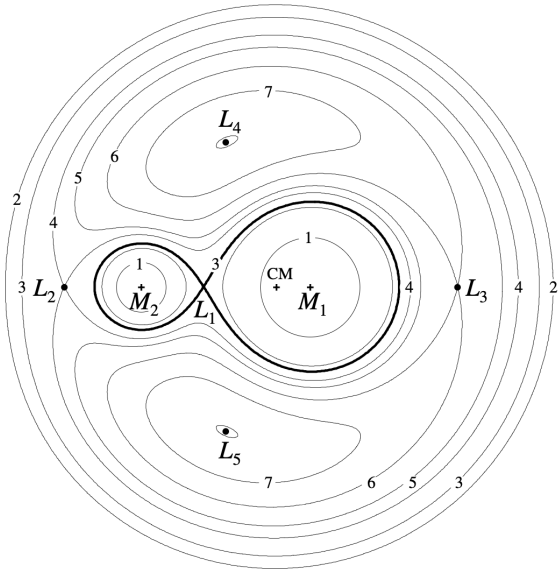

Figure 36: Sections in the orbital plane of the Roche potential ϕ. The five Lagrange point where the potential is zero are marked in the picture with a L_i. The saddle point L_1 is said "inner Lagrange point" and forms a "pass" from the two Roche lobes (3). Note that the points L_1, L_2, L_3 are always points of unstable equilibrium, while L_4 and L_5 might be stable provided that $M_1/M_2 > 24.96$. Credits: Frank, King and Raine.

What if one of the two stars' external layer bloates up and steps out of its Roche lobe? Since the lobes are essentially demarcation lines between the two (single star) potentials, once a Roche lobe is overflown, material will fall on the other object, captured by the gravitational potential of the other star.

This type of mass transfer is called *Roche lobe overflow*. Since mass is removed from the star quite readily when the lobe is filled, the star cannot never grow significantly larger than its Roche lobe[2]. The star losing mass is typically named *donor*, while the one gaining mass is called *accretor*.

Generally, there's a non-zero pressure gradient near the L_1 Lagrange point, with material often flowing at supersonic speeds[3]. If we introduce angular momentum back into the picture, Coriolis' forces will arise, deviating the trajectory of said gas until it settles into a ring.

To a good approximation we can take the stream trajectory as the orbit of a test particle released from rest at L_1, and thus with a given angular momentum, falling in the gravitational field of the accreter alone. This would give an elliptical orbit lying in the binary plane: The presence of the donor causes this to precess slowly. A continuous stream trying to follow this orbit will therefore intersect itself, resulting in dissipation of energy via shocks.

2 This only applies to binary systems whose Roche lobes are smaller than the radius of the object after its expansion.

3 For those who may be interested in the subject, I do suggest reading into something of I. S. Shklovskii, for example [23] or search for something here.

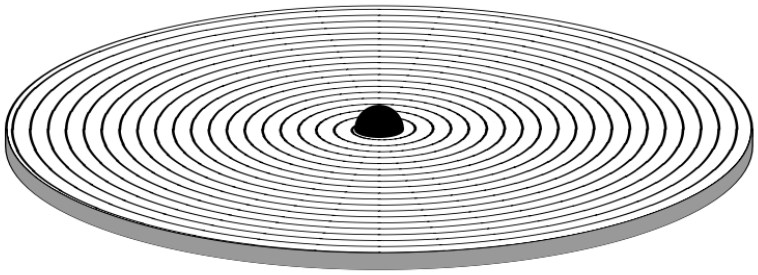

Figure 37: Perspective view of the spiraling gas. Credits: Frank, King and Raine.

On the other hand, the gas has little opportunity to rid itself of the angular momentum it had upon leaving L_1, so it will tend to the orbit of lowest energy for a given angular momentum, a circular orbit. We thus expect the gas to initially orbit the accretor in the binary plane at a radius R_c such that the Kepler orbit at R_c has the same specific angular momentum as the transferring gas had on passing through L_1.

If we consider Euler's equation for the vorticity in presence of viscosity

$$\frac{\mathrm{D}\vec{\omega}}{\mathrm{D}t} = \nu\nabla^2\vec{\omega}$$

we diskover that a sensible solution may be of the form

$$\omega \propto \exp\left(-\frac{r^2}{2\nu\tau}\right)$$

so that material diffuses all over until eventually forms a *disk*, which will be the focus of the next section.

AGNs and BEBs (Big Energetic Bros)

Jokes aside, as we'll have the pleasure to diskuss in the next few sections, accretion isn't limited to "small" objects with hard boundaries[4] like stars.

As far as our understanding goes, the very things at the center of galaxies are constantly accreting the nearby matter, depleting in the process gravitational energy in (often) spectacular forms which we're able to detect!

On top of that, we're aware of the existence of binary systems made of a "small" black hole and something else; this is the case, for example, of the Sco-X1 binary system.

AGNs (*Active Galactic Nuclei*) are particularly relevant in the picture of accretion since they've been observed to be able to accrete both spherically (no angular momentum) and non-spherically (non-zero angular momentum), with the formation of a disk. Our common understanding of AGNs relies on the (fairly tested) belief that they're powered by supermassive black holes with masses lying in the range of $10^5 - 10^8\, M_\odot$.

To give the world the illusion that we're really understanding what's going on in there, AGNs have been given a nice classification based on the type of galaxy they're found in and based on their spectral features (when needed):

4 The notion of "hard" for something made of hot plasma may be somewhat deceptive, but we'll stick with the notation.

- Blazers → Elliptical galaxies
- Type I Seyfert → (Late type) Spiral Galaxies with narrow absorption lines
- Type II Seyfert → (Late type) Spiral Galaxies with broad absorption lines[5]
- Radio Loud Quasars → (Early type) Elliptical galaxies
- Radio Quiet Quasars → (Early type) Elliptical galaxies

It may be worth pointing out that this classification is a hand-wavingly way to present the same object from different points of view. In fact, all the entroes of the previous list represent essentially the same object, with the relative tilt from the line of sight the only thing that is really changing.

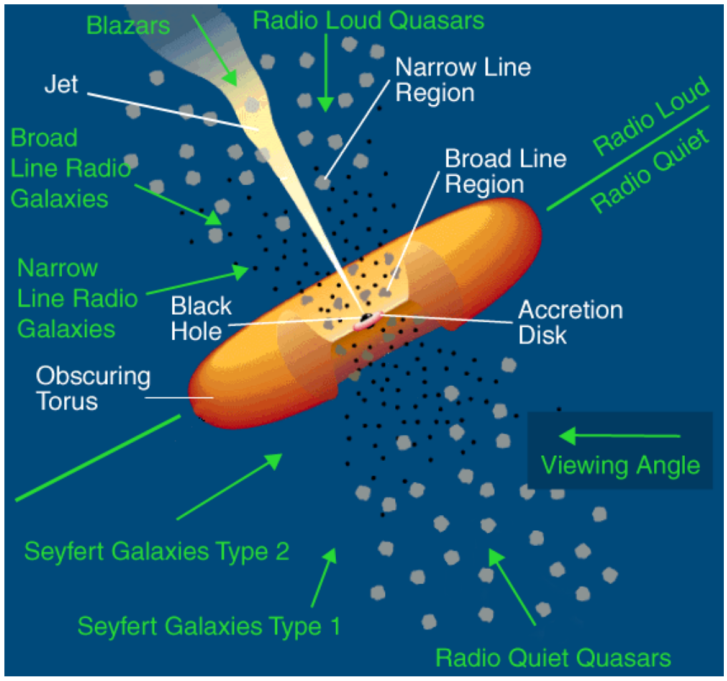

Figure 38: Explicative pictorial representation of the unfied model of AGNs. Credits: [24]

But this should not surprise you: Relativistic beaming is heavily frequency dependent and direction dependent, so the direction you're looking from will impact significantly on what you're going to see, leading to really different results. As you can easily cross-check in your favorite book of electrodynamics (mine, for example, is [14]), the angular distribution of radiation emitted by an accelerated charge can be calculated starting from the relativistic-adjusted Poynting's vector.

5 The broadness or narrowness of the lines is related to their velocity in respect to us. Narrow absorption lines mean that the AGN is far from us, while broad absorption lines are typical of closer AGNs.

Doing this, you'll find out that when the velocity and the acceleration of the emitting particle are parallel, the emitted power per solid angle is given by

$$\frac{\mathrm{d}P}{\mathrm{d}\Omega} = \frac{e^2\dot{v}^2}{4\pi c^3}\frac{\sin^2\theta}{(1-\beta\cos\theta)^5}$$

where θ is the angle of observation measured from the direction of $\vec{\beta} = \vec{v}/c$. From this expression it's clear the heavy direction-dependence; to find the frequency spectrum you need to put in some more work.

Although not entirely nor strictly related to this, I think you can get an idea of how perspective affects what we'll be able to see by watching this short clip by Kip Thorne on the film "Interstellar".

Another interesting fact is that most of these systems present powerful jets which may remain coherent for *kiloparsecs*.

How? Now, that's a good question. The most accredited theory has been developed by R. D. Blandford and R. L. Znajek in 1977 [3]; we won't go much deeper than this here[6], but they have essentially speculated the formation of magnetic funnels around spinning black holes.

When you start considering spinning black holes, a whole other handful of problems arises. For once, you can no longer use the nice and simple Schwarzschild metric to describe the BH, but you have to use Kerr's. Even then, it has been shown that in presence of Kerr's black holes, if the accretion disk is (somehow) tilted in respect to the rotation axis of the BH, an effect called *Bardeen precession* [1] arises, leading to a considerable part of the radiation emitted in the central part of the disk to be reabsorbed, inducing observable changes in the X-ray spectrum.

And then you have curious fellas like *obscure black holes*, which are essentially geometrically thick (we'll see more about this in the next section) tori through which only infrared radiation may pass.

8.4 DISK FORMATION

If we ignore viscosity in Euler's equation, we may write

$$\partial_t\vec{v} + (\vec{v}\cdot\nabla)\cdot\vec{v} = -\frac{\nabla p}{\rho} - \frac{GM}{r^2}\hat{r}$$

Assuming ideal gas behavior

$$\frac{1}{\rho}\nabla p = \frac{RT}{\mu}\nabla\left(\log(p)\right)$$

For convenience's sake, we're going to switch to adimensional coordinates

$$\tilde{r} = r/r_0 \quad \tilde{v} = v/\Omega_0 r_0 \quad \tilde{t} = \Omega_0 t \implies \tilde{\nabla} = r_0\nabla$$

this way Euler's equation can be restated in terms of adimensional quantities only. Introducing once again the Virial temperature (126)

$$\frac{\partial\tilde{v}}{\partial\tilde{t}} + (\tilde{v}\cdot\tilde{\nabla})\cdot\tilde{v} = -\frac{T}{T_{vir}}\tilde{\nabla}\log(p) - \frac{1}{\tilde{r}^2}\hat{r}$$

6 To quote a book I've studied from: "The author of these notes does not claim to have any particular insight on the subject."

From this last equation we're learning something important already: If $T \ll T_{vir}$ there's no feedback from the gas, and so the disk's geometry will be determined by gravity alone.

Assuming that there's little pressure pushing on the "vertical" direction, we may deduce that the vertical size will be "geometrically thin", and its dynamic will be dominated by gravity alone. In fact, at first order, the vertical component of the acceleration can be assumed to be

$$g_z = -\Omega_0^2 z$$

which gives raise to a non-negligible velocity dispersion along the vertical direction. In short: Particles start oscillating.

Although we'd be contradicting ourselves, we're going to assume hydrostatic equilibrium to hold along the vertical direction. This allows us to find an analytic expression for the density profile along the vertical axis

$$\rho(z) = \rho_0 \exp\left(-\frac{z^2}{2H^2}\right)$$

with $H = c_s/\Omega_0$ the disk's scale height and $c_s = (TR/\mu)^{1/2}$ the sound speed. If we define the *aspect ratio* δ of the disk as

$$\delta = \frac{H}{r} = \frac{c_s}{\Omega_0 r} = \left(\frac{T}{T_{vir}}\right)^{1/2} = \frac{1}{\mathcal{M}} < 1 \tag{130}$$

we conclude that the gas in the disk must be supersonic[7]. Thus, the disk is pretty much a continuum of Keplerian orbits along which the gas has different velocities.

Now, if we switch to the comoving frame and recall Fig.24, we may notice that we are in the exact conditions for shear instability to arise. But since angular momentum is *fixed* for any given orbit, each displacement caused by shear instabilities will be simply ignored, for angular momentum conservation will eventually bring back the displaced element to its original position.

In fact, as long as

$$\frac{\partial J}{\partial r} < 0$$

the disk is stable against small element-displacement.

If that's the case, how can accretion be possible? Viscosity.

Although we have no clues (sorta[8]) of how the viscosity gets high enough to allow the rings of gas to migrate towards the center and lose angular momentum, it's clear that if that is the case, then accretion may indeed be possible.

8.5 THE SHAKURA–SUNYAEV MODEL

As we have mentioned *en passant* at the end of the last section, a differentially rotating medium (like our Keplerian-continuum) must develope tangential

7 For $\Omega_0 r$ can be considered the tangential velocity to the (circular) orbit.

8 As for many other things in the Universe, it is probably turbulence's doing. We'll take a look at it in the next section.

stresses between adjacent layers, which are believed to be connected to the presence of magnetic fields, turbulence and molecular and/or radiative viscosity acting as the primary mechanisms of transport of angular momentum.

In the conditions that are of interest to us, nor molecular nor radiative viscosity can lead to accretion. So that leaves us with only magnetic fields and turbulence. As noticed by Shakura and Sunyaev [22], the magnetic field flowing into the disk may have a regular structure at the distances comparable to the outer radius of the disk. But since $\nabla \cdot \vec{H} = 0$, the radial component of the magnetic field must have alternating sign, hence differential rotation leads to division of the large magnetic loops into smaller ones.

Within the disk is then most probable for the field to be chaotic and on the small scales. The energy density in the magnetic field cannot exceed the thermal energy of matter

$$\frac{|\vec{H}|^2}{8\pi} < \rho \frac{c_s^2}{2}$$

Although we have no complete theory of turbulence, we have observational checks of its existence in accretion disks, so we can assume turbulence to be indeed present.

In fact, as computed by [10] (§4.7), a safe estimate of the Reyndolds' number evaluates roughly at $\mathrm{Re} > 10^{14}$. Hence, molecular viscosity is far too weak to bring about the viscous dissipation and angular momentum transport we require. We may then conjecture that the gas flow in accretion disks is also turbulent, although there is, as yet, no proof that this is so.

What we're going to do is invoking Boussinesq approximation and treat turbulence as a source of viscosity[9].

Although we know pretty much nothing, we can still make some sensible claims on the quantities limiting the onset of turbulence: First, the typical size of the largest turbulent eddies cannot exceed the disk thickness H; second, it is unlikely that the turnover velocity (that is, the velocity beyond which the fluid can be considered turbulent) v_{turb} is supersonic, for, in this case, the turbulent motions would probably be thermalized by shocks. Thus, we can write

$$\nu = \alpha c_s H \tag{131}$$

and can probably expect with a certain degree of confidence that α will turn out to be less than 1. A way to see this is by considering two adjacent elements of the disk that are going to generate eddies of, say, size $l \approx H$. The tangential stress tensor $\sigma_{r\phi}$ is going to be roughly equal to

$$\sigma_{r\phi} \sim \nu_t \rho r \partial_r \Omega \sim -\frac{3}{2} \nu_t \rho \Omega \implies \sigma_{r\phi} \sim -\rho c_s^2 \frac{v_t}{c_s}$$

where we've both used the Boussinesq approximation $\nu_t \approx v_t H$ (v_t is the characteristic velocity of the turbulent flow) and the definition of the scale height $H = c_s/\Omega$.

Since the velocity of the turbulent flow can't get as high as being causality-breaking, $v_{max} \sim c_s$. Hence the viscous stresses

$$-\sigma_{r\phi} \sim \rho c_s^2 \frac{v_t}{c_s} = \alpha \rho c_s^2$$

9 For more details about this you may check the Fluid Dynamics class or [18].

can be characterized by only one parameter α, which in the case of a turbulent mechanism must be $\alpha < 1$ always (since $v_t/c_s \lesssim 1$).

This is the celebrated *Shakura-Sunyaev α-prescription*. Note how this is but a mere parametrization: We have simply dumped everything we don't know about turbulence in the α parameter and then adjusted the remaining quantities on the basis of dimensional analysis only.

8.5.1 The structure of the disk

Let's assume the accretion disk to be geometrically thin (along the vertical axis) with cylindrical symmetry. All calculations will be carried out with Newtonian mechanics. It is necessary to take into account GR corrections only in the region $r < 3R_s$, where no stable circular orbits are possible[10].

We can define the surface density of the disk $\Sigma(r)$ as

$$\Sigma(r) = \int_{-\infty}^{+\infty} \rho(z, r)\, \mathrm{d}z \approx 2H\rho_0 \tag{132}$$

with ρ_0 the density on the plane of the disk[11].

We shall now consider the conservation equations

$$\partial_t(r\Sigma) + \partial_r(r\Sigma v_r) = 0 \tag{133}$$

$$v_\phi^2 = \frac{GM}{r} \tag{134}$$

while the equation of conservation of angular momentum will be

$$\partial_t(\rho v_\phi) + \partial_r(\rho v_r v_\phi) = F_\phi \tag{135}$$

with F_ϕ the viscous "forces". Integrating over the vertical direction yields

$$\partial_t(r\Sigma\Omega r^2) + \partial_r(r\Sigma v_r \Omega r^2) = \partial_r(S r^3 \partial_r \Omega) \tag{136}$$

with

$$S = \int_{-\infty}^{+\infty} \rho\nu\, \mathrm{d}z \approx \Sigma\nu$$

if we assume ν weakly dependent on z (*isothermal disk approximation*). We can use the conservation of mass equation to simplify the expression above

$$r\Sigma v_r \partial_r(r^2\Omega) = \partial_r(\Sigma\nu r^3 \partial_r \Omega)$$

where we've assumed $\partial_t\Omega = 0$, condition that will hold for orbits in a fixed gravitational potential. We can use once more the continuity equation of mass to eliminate v_r

$$r\partial_t\Sigma = -\partial_r\left[\frac{1}{\partial_r(r^2\Omega)}\partial_r(\Sigma\nu r^3\partial_r\Omega)\right]$$

Now we can use the assumption of Keplerian orbits $\Omega \propto r^{-3/2}$ to obtain the *thin disk equation*

$$r\partial_t\Sigma = 3\partial_r\left(r^{1/2}\partial_r\left(\nu\Sigma r^{1/2}\right)\right) \tag{137}$$

10 You can see [5] §5, for more details.

11 We have plugged in the Gaussian $\rho(z)$ behavior we have hand-wavingly pulled out of the cylinder in the last section.

This is the basic equation governing the time evolution of surface density in a Keplerian disk. In general, it is a nonlinear diffusion equation for Σ, because ν may be a function of local variables in the disk.

Given a solution of this equation, we can backtrack and find a solution for v_r, which will simply be

$$v_r = -\frac{3}{\Sigma r^{1/2}} \partial_r \left(\nu \Sigma r^{1/2} \right)$$

Clearly, to make further progress we need some prescription for ν.

Let's see what happens if we allow the disk to settle into a steady-state structure and set all $\partial_t = 0$ in the conservation equations. From here it's clear that the mass transfer rate $\dot{M}$ is given by

$$\dot{M} = -2\pi \Sigma |v_r| r \quad (v_r < 0)$$

which is also equal to

$$\dot{M} = 6\pi r^{1/2} \partial_r \left(\nu \Sigma r^{1/2} \right)$$

by means of substituting the backtracked value of v_r we've written down a few lines above. From the conservation of angular momentum we get

$$-\nu \Sigma r^3 \partial_r \Omega = -|v_r| \Sigma \Omega r^3 + C$$

with C an arbitrary constant of integration, related to the rate at which angular momentum flows into the compact star, or equivalently, the couple exerted by the star on the inner edge of the disk.

If the system is stationary, all the properties of the disk are determined by the (constant) accretion rate alone

$$\nu \Sigma = \frac{1}{3\pi} \dot{M} \left(1 - \beta \left(\frac{R_*}{r} \right)^{1/2} \right) \tag{138}$$

where β is an integration constant and R_* is the inner radius of the disk. From here we deduce that the flux of angular momentum going in the accretor is

$$F_J = -\dot{M} \beta \Omega_{Kepl}(R_*) R_*^2$$

For example, let us suppose that the disk extends all the way down to the surface $r = R_*$ of the central star[12]. Hence, there must exist a radius $\tilde{R} > R_*$ where $\partial_r \Omega = 0$. In a realistic situation, the star must rotate more slowly than break-up speed at its equator

$$\Omega(R_*) < \Omega_{Kepl}(R_*)$$

this implies $\beta = 1$. This means there will be an inward flux of angular momentum; accretion is therefore going to speed up the rotation of the central object. This is in particular the case of objects commonly known as "recycled pulsars".

12 As we already mentioned, this can't be the case if the central object is a BH. The minimum radius permitted in that case is that of the ISCO, the Innermost Stable Circular Orbit predicted by Schwarzschild.

8.5.2 Stochastic background of GW

Recycled pulsars are thought to be related to low-mass X-ray binary systems. It is speculated that X-rays emissions in these systems are generated by the accretion disk of a neutron star. The transfer of angular momentum from this accretion event can increase the rotation rate of the pulsar to hundreds of times per second, as is observed in millisecond pulsars.

This category of object is particularly relevant because of their extreme regularity: They have very short periods (typically of the order of the millisecond) with very low rate of change in time[13]. As such, it has become customary to use such pulsars to "build" a PTA (*Pulsar Timing Array*[14]) in order to detect correlations of the pulses of the pulsars due to the transit of a gravitational wave.

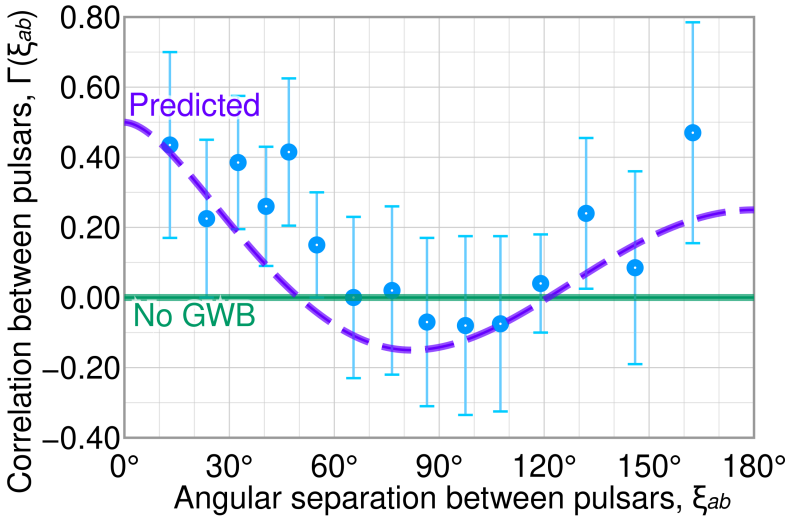

Figure 39: Plot of the correlations observed by NANOGrav (2023) vs angular separation between pulsars, compared with a theoretical model (dashed purple) and if there were no gravitational wave background (solid green) Credits: Wikipedia.

This is particularly true when the wavelength of the GW is comparable with the relative distance between the pulsars.

This way, it has been possible to gather evidence of a stochastic background of GW, as depicted in Fig.39 and predicted by Hellings and Downs in 1983 [12].

13 The interested reader can refer, for example, to [9].

14 Basically, a set of galactic pulsars that is monitored and analyzed to search for correlated signatures in the pulse arrival times on Earth.

8.6 SPECTRUM AND TEMPERATURE OF THE DISK

Consider again eq.138 for $\Omega_* < \Omega_{Kepl}$, so that $\beta = 1$. Under the assumption of LTE, we'd like to see if we're able to determinate the surface temeprature of the disk. You should recall the 1st principle of Thermodynamics

$$\rho T \frac{\mathrm{d}S}{\mathrm{d}t} = -\nabla \cdot \vec{F}_H + Q_\nu$$

On the left hand side we have the variation of entropy, while on the right hand side we have the heat flux and viscous dissipation[15]. If we consider almost stationary processes, the dissipation must happen over a timescale roughly equal to the orbital period $t_D \sim \Omega^{-1}$. Hence

$$\rho T \frac{\mathrm{d}S}{\mathrm{d}t} \approx 0 \text{ for } t > t_D$$

Note that this condition is automatically satisfied if LTE holds, since it assures energetic equilibrium. This way we can integrate over the vertical direction the 1st principle above and find

$$\int_{-\infty}^{+\infty} (-\nabla \cdot \vec{F})\, \mathrm{d}z + \int_{-\infty}^{+\infty} Q_\nu \, \mathrm{d}z \approx 0$$

Because of the thin disk approximation, the disk medium is essentially "plane-parallel" at each radius, so that the temperature gradient is effectively in the z-direction; thus, the left hand side is simply equal to 2 times $F(z)$ (up to an additive constant).

Assuming that the disk is optically thick in the z-direction holds as a consequence that each element of the disk face radiates roughly as a blackbody with a temperature $T(r)$ given by equating the dissipation rate $D(r)$ per unit face area to the blackbody flux (4)

$$\sigma T_s^4 = \int_{-\infty}^{+\infty} Q_\nu \, \mathrm{d}z = D(r)$$

What this equation is saying is just that all energy dissipated by viscosity is radiated away. At this point we may write down the 8 equations for the disk structure, but, in the interest of time, I won't. The interested reader may refer, once again, to [10], §5.4, equation(s) 5.41.

Let's write down the dissipation rate

$$D(r) = \frac{G\partial_r \Omega}{4\pi r} = \frac{1}{2}\nu\Sigma(r\partial_r\Omega)^2$$

and evaluate it where the angular frequency has the Keplerian form

$$D(r) = \frac{9}{8}\nu\Sigma\frac{GM}{r^3} = \frac{9}{8}\nu\Sigma\Omega^2$$

At this point we can equate the blackbody flux with the viscous dissipation and, plugging in the result from eq.138, we finally obtain

$$\sigma T(r)^4 = \frac{3}{8\pi}\dot{M}\frac{GM}{r^3}\left(1 - \left(\frac{R_*}{r}\right)^{1/2}\right) \tag{139}$$

15 As in stars, the vertical energy transport mechanism may be either radiative or convective, depending on whether or not the temperature gradient required for radiative transport is smaller or greater than the gradient given by the adiabatic assumption. See Chapter 5, §4.4.

which is notably independent from viscosity.

Far from the inner radius of the disk $r \gg R_*$, the surface temperature of the disk can be expressed as

$$T(r) = T_* \left(\frac{r}{R_*}\right)^{-3/4} \qquad T_* = \left(\frac{3GM\dot{M}}{8\pi\sigma R_*^3}\right)^{1/4}$$

with T_* curiously of the same order of the blackbody temperature defined as follows

$$T_B = \left(\frac{L_{acc}}{4\pi\sigma R_*^2}\right)^{1/4}$$

that is the temperature the accreting source would have if it were to radiate the given power as a blackbody spectrum.

Recalling the plane parallel approximation for radiative diffusion (33) with no flux coming from the outside

$$-\frac{\mathrm{d}}{\mathrm{d}r}(\sigma T^4) = \frac{3}{4} F \rho \kappa_R \implies F = \sigma T^4 (\tau = 2/3)^{16}$$

hence if we assume the flux to be roughly constant along the vertical direction

$$F = \sigma T_s^4 = \sigma T(r)^4 = \frac{3}{8\pi}\dot{M}\frac{GM}{r^3}\left(1 - \left(\frac{R_*}{r}\right)^{1/2}\right)$$

If we assume the opacity to be constant along the vertical direction

$$\tau = \frac{\kappa\Sigma}{2} \gg 1 \implies T^4 = \frac{27}{64}\sigma^{-1}\Omega^2\nu\Sigma^2\kappa$$

Despite this being a really bad take, it still gives us an idea of how the central temperature of the disk may depend on the viscosity.

8.6.1 The structure of steady α-disks

The disk structure we're now investigating was, as you may have guessed from the title, proposed by Shakura and Sunyaev. All we need is a prescription for the viscosity and one for the opacity.

In order to do this, we're going to assume ideal gas behavior with negligible radiation pressure, with viscosity given by the α-prescription

$$\nu = \alpha c_s H$$

and assume that ρ and T_c (a.k.a. the central temperature) are such that the Rosseland mean opacity is well approximated by Kramers' law[17]

$$\kappa_R = 5 \cdot 10^{24} \rho T_c^{-7/2}\,\mathrm{cm}^2\,\mathrm{g}^{-1}$$

With these choices for ν and κ_R, the system of equations for the disk is algebraic and can be solved straightforwardly.

The Shakura-Sunyaev disk solution is expressed in the system of equations below.

16 This one holds true because for a plane parallel grey atmosphere with the Eddington approximation in LTE you can show that $T^4 = 3/4\, T_{eff}^4(\tau + 2/3)$.

17 We haven't actually seen this in class, so take it at face value.

$$
\left.
\begin{aligned}
\Sigma &= 5.2\alpha^{-4/5}\dot{M}_{16}^{7/10}m_1^{1/4}R_{10}^{-3/4}f^{14/5}\,\mathrm{g\ cm^{-2}},\\
H &= 1.7\times10^{8}\alpha^{-1/10}\dot{M}_{16}^{3/20}m_1^{-3/8}R_{10}^{9/8}f^{3/5}\,\mathrm{cm},\\
\rho &= 3.1\times10^{-8}\alpha^{-7/10}\dot{M}_{16}^{11/20}m_1^{5/8}R_{10}^{-15/8}f^{11/5}\,\mathrm{g\ cm^{-3}},\\
T_\mathrm{c} &= 1.4\times10^{4}\alpha^{-1/5}\dot{M}_{16}^{3/10}m_1^{1/4}R_{10}^{-3/4}f^{6/5}\,\mathrm{K},\\
\tau &= 190\alpha^{-4/5}\dot{M}_{16}^{1/5}f^{4/5},\\
\nu &= 1.8\times10^{14}\alpha^{4/5}\dot{M}_{16}^{3/10}m_1^{-1/4}R_{10}^{3/4}f^{6/5}\,\mathrm{cm^2\,s^{-1}},\\
v_R &= 2.7\times10^{4}\alpha^{4/5}\dot{M}_{16}^{3/10}m_1^{-1/4}R_{10}^{-1/4}f^{-14/5}\,\mathrm{cm\ s^{-1}},\\
\text{with } f &= \left[1-\left(\frac{R_*}{R}\right)^{1/2}\right]^{1/4}.
\end{aligned}
\right\}
$$

Figure 40: Disk equations for the Shakura-Sunyaev model. You'll pardon me if I didn't write this down. Credits: Frank, King and Raine.

A number of interesting remarks can be made just by observing these equations. First of all, in none of the equations above the α-parameter (that we know pretty much nothing about) enters with a high power. This means that all these quantities are not particularly sensitive to the actual value of α. Note, however, that this also means that we cannot expect to diskover the typical size of α by direct comparison of steady-state disk theory with observations.

If we compute the aspect ratio δ

$$\delta = \frac{H}{r} = \left(\frac{R}{\mu}\right)^{2/5}\left(\frac{3}{64\pi^2\sigma}\right)^{1/10}\left(\frac{\kappa}{\alpha}\right)^{1/10}GM^{-7/10}r^{1/20}(f\dot{M})^{1/5} \qquad (140)$$

so the disk is indeed thin.

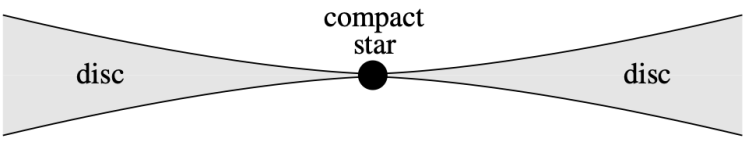

Figure 41: Section (not scaled) of the disk. The weak r dependence is greatly accentuated.

In agreement with this, the radial drift velocity v_r is highly subsonic while the Kepler velocity v_ϕ is very supersonic. The disk is certainly optically thick for any reasonable accretion rate; the central and surface temperatures are in the ratio $\sim \tau^{1/4} \sim 3.7$, so the disk is roughly uniform in the vertical direction and can extend out to quite large radii, of the order of the Roche lobe of the accreting star.

Earlier we've computed the disspation due to viscous forces

$$W_\nu = \frac{1}{2}D(r) = \frac{9}{4}\nu\Sigma\Omega^2$$

and we can compare it to the gravitational potential W_G

$$W_G = \frac{1}{2\pi r}\frac{GM\dot{M}}{2r^2}$$

which is defined so that

$$\int_{R_*}^{\infty} 2\pi r W_G = \frac{GM\dot{M}}{2R_*}$$

Their ratio evaluates to

$$\frac{W_\nu}{W_G} = 3\left(1 - \left(\frac{R_*}{r}\right)^{1/2}\right) \tag{141}$$

from which is clear that at large distances (respect to the inner radius of the disk) $W_\nu \approx 3W_G$.

At a fixed distance we should then see a monochromatic intensity which is a convolution of all the blackbody radiation from that point on

$$I_\nu = \int B_\nu(T_s(r))\,\mathrm{d}r$$

A first correction would be to include radiation pressure $p = aT^4/3$ if it is too big to be neglected. The aspect ratio would then change

$$\delta \approx \frac{3}{8\pi}\frac{\kappa}{cR} f\dot{M} = \frac{3}{2} f \frac{\dot{M}}{\dot{M}_E} \quad \dot{M}_E = \frac{4\pi R c}{\kappa}$$

If this is the case, there's a serious "risk" of incurring into ADAF.

9 BLACK HOLES BINARIES AND COALESCENCE

9.1 INTRODUCTION

This last part of the notes will focus on (Astrophysical) Black Holes, going through the formation of BH binaries and all the processes that make possible the formation of such systems.

I'll closely follow my own notes, as well as those made available by Professor M. Crisostomi [6]. However, I won't fail to suggest books and/or article for those who may be interested.

It is common to identify astrophysical black holes according to their mass

- *Stellar-mass Black Holes* (SBHs); they have masses in the range $M_\odot \lesssim M \lesssim 10^2 M_\odot$. SBHs are what remains of stars with initial mass of the order $M \gtrsim 20 M_\odot$.
- *Intermediate-mass Black Holes* (IMBHs); they have masses in the range of $10^2 M_\odot \lesssim M \lesssim 10^6 M_\odot$. Since we have only scarce observative evidence about these guys, they're the ones we know the least about. We know that they can form at high redshift either from PopIII star remnants[1] or via direct collapse from a marginally stable protogalactic disc or quasistar, or at lower redshifts via dynamical processes in massive star clusters.
- *(Super) Massive Black Holes* (MBHs or SMBHs); they have masses in the range of $10^6 M_\odot \lesssim M \lesssim 10^{10} M_\odot$. They are observed at the center of almost all galaxies, AGNs and Quasars.

Most of the objects of astrophysical interest (stars, neutron stars, BHS etc.) are typically observed in binary systems. For example, half of the total amount of stars and, more interesting, 70% of the stars with mass greater than ten solar masses, that are relevant to the formation of compact objects, are in binaries. On top of that, as shown in any GR course, two compact objects bound in a binary system have a non-zero quadrupole moment and are thus able to emit GWs.

After a short digression about redshift, more for the sake of completeness than anything, we'll start turning our attention to the physical features of binary systems. In particular, we'll consider the Keplerian motion of the binary, where two compact objects can be regarded as point masses and, as long as the orbital velocities are small compared to the speed of light.

1 "Population III stars are a hypothetical population of extremely massive, luminous and hot stars with virtually no "metals", except possibly for intermixing ejecta from other nearby, early population III supernovae. Such stars are likely to have existed in the very early universe (i.e., at high redshift) and may have started the production of chemical elements heavier than hydrogen, which are needed for the later formation of planets and life as we know it." Definition from Wikipedia.

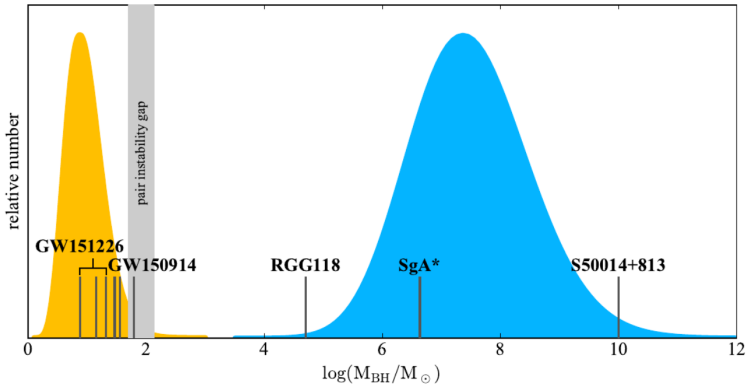

Figure 42: Pictorial representation of the mass ranges of known black holes. Credits: [8].

In short, unless pointed out otherwise, our discussion will be mainly classical.

A short detour on redshift

In our description of the Universe we'd like to observationally determine a number of quantities to try to infer which of the models proposed by the FRW metric[2]

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a(t)^2\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\Omega^2_{(2)}\right) \tag{142}$$

corresponds to our Universe. First and foremost, we could perform a slight manipulation of the FRW metric (142) and define a conformal time

$$a(\eta)\mathrm{d}\eta = \mathrm{d}t$$

so that the metric is, guess what, conformal. The scale factor a is therefore factorized and negligible if we consider null-like trajectories $\mathrm{d}s^2 = 0$. Note that in

$$\mathrm{d}s^2 = a(\eta)^2\left(-\mathrm{d}\eta^2 + \frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\Omega^2_{(2)}\right) \tag{143}$$

the quantity within parentheses is independent of the conformal time. We could try with a hand-waving approach and obtain (out of thin air) the expression for the cosmological redshift, but we'll take a somewhat more rigorous approach.

Consider the four-velocity of comoving observers $U^\mu = (1,0,0,0)$, then the metric admits a Killing tensor of the form

$$K_{\mu\nu} = a^2(g_{\mu\nu} + U_\mu U_\nu)$$

that you can check that it satisfies $D_{(\sigma}K_{\mu\nu)} = 0$ where D_σ is the covariant derivative. By definition of Killing vector (the generalization to a K. tensor is straightforward), the quantity

$$K^2 = K_{\mu\nu}V^\mu V^\nu = a^2\left[V_\mu V^\mu + (U_\mu V^\mu)^2\right] \tag{144}$$

2 As per any GR subject in this notes, I'll follow [5]. You can find this discussion at §8.5.

is conserved along the geodesics. Here $V^\mu = \mathrm{d}x^\mu/\mathrm{d}\lambda$. For massive particles, since $V_\mu V^\mu = -1$ (note the convention on the signature of the metric), we also have

$$(V^0)^2 = 1 + |\vec{V}|^2 \quad |\vec{V}|^2 = g_{ij}V^iV^j$$

Recalling the formula for the four-velocity of the observer, we have $U_\mu V^\mu = -V^0$, so (144) implies

$$|\vec{V}| = \frac{K}{a}$$

Notice how this means that the particle "slows down" with respect to the comoving coordinates as the Universe expands. If we instead consider null geodesics (that is what we're actually interested in)

$$V_\mu V^\mu = 0 \implies U_\mu V^\mu = \frac{K}{a}$$

but for null-like particles, like photons, we can write the four-velocity as $V^\mu = (\omega, \vec{k})$. So the frequency of the photon as measured by a comoving observer is $\omega = -U_\mu V^\mu$ (remember that $c = 1$). The frequency emitted will still be given by the same functional expression but calculated in the local inertial frame of the emitter, so $\omega_{em} = -K/a_{em}$. We then find out that

$$\frac{\omega_{obs}}{\omega_{em}} = \frac{a_{em}}{a_{obs}}$$

To make things easier we then define a quantity called *redshift* defined as

$$z_{em} = \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}}$$

so that, if the observation takes place today ($a_{obs} = a_0 = 1$), we can just write

$$a_{em} = \frac{1}{1 + z_{em}}$$

and

$$\boxed{\frac{\nu_{obs}}{\nu_{em}} = \frac{1}{1 + z_{em}}} \tag{145}$$

So the redshift tells us two things: The scale factor of the universe when the photon was emitted and, in a sense, the distance of the emitter when the photon was emitted. In fact, using Hubble's law (assuming we've somehow computed the Hubble constant at its present value) we can easily calculate the *instantaneous physical distance*

$$v = cz = H_0 d_P$$

Was all of this really necessary for what we'll be going through in the next sections? No, not really. But I thought it might have been cool having a clue on where this curious quantity called redshift comes from and, more importantly, knowing what it actually means and implies. That being said, we can hop on back on tracks.

9.1.1 Keplerian Motion

Let us consider two point-like objects of mass M_1, M_2 moving along elliptical orbits around their center of mass as shown in Fig.43.

As customary, we can define the total mass M_{tot} and the reduced mass μ of the system as well as the relative coordinate $\vec{r} = \vec{r}_1 - \vec{r}_2$, with $\vec{r}_i$ the distance of the mass M_i from the center of mass.

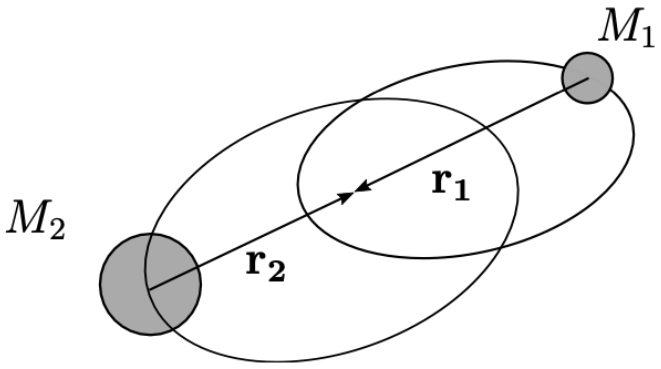

Figure 43: Binary system religiously moving as predicted by Newtonian dynamics. Credits: [6].

$$\vec{r}_1 = \frac{M_2}{M_{tot}}\vec{r} \quad \vec{r}_2 = -\frac{M_1}{M_{tot}}\vec{r}$$

The velocities relative to the center of mass can be obtained straightforwardly by deriving the expression above in respect to time.

The total energy of the system is

$$E = \frac{M_1 v_1^2}{2} + \frac{M_2 v_2^2}{2} - \frac{GM_1M_2}{r} = \frac{\mu v^2}{2} - \frac{GM\mu}{r} = -\frac{GM\mu}{2a}$$

In the last step, we have used the conservation of energy along the orbit and wrote the energy in terms of the semi-major axis a. Hence, the binary system is equivalent to considering a single body with mass μ moving in an effective external gravitational potential. The orbital frequency is given by Kepler's third law

$$\omega = \frac{2\pi}{T} = \sqrt{\frac{GM}{a^3}}$$

which is, incidentally, the same result you obtain considering the (proper time) orbital period of a test-body in the gravitational potential induced by a Schwartzschild metric[3].

It is useful to calculate the orbital angular momentum $J_{orb} = \vec{r} \wedge \mu\vec{v}$, which has magnitude

$$J_{orb} = \mu\sqrt{GMa(1-e^2)}$$

where e is the eccentricity of the orbit. Thus, for a circular orbit ($e = 0$) the magnitude of the angular momentum is just

$$J_{orb,C} = \mu\sqrt{GMr_C}$$

3 It's not difficult to show. You have to make use of the various conserved quantities the Killing vectors kindly gift you and then calculate $d\phi/d\tau$ with τ the proper time of the particle.

hence the tangential velocity is

$$v = \sqrt{\frac{GM}{r_C}}$$

9.1.2 GW radiation from a binary system

As you should probably know if you've attended a GR class, a system of sufficiently[4] compact objects bound in a binary system can start emitting GWs. A full description of how GWs are produced in the weak field limit from the linearized Einstein's equation is beyond the scope of these notes. The interested reader can refer to [5], §7.

For Newtonian sources localized in a compact region of space, the gravitational power radiated is governed by the *quadrupole momenta* of the binary and given by

$$\frac{\mathrm{d}E_{rad}}{\mathrm{d}t} = \frac{G}{5}\langle \dddot{I}_{ij}\dddot{I}_{ji}\rangle \tag{146}$$

where the brackets stand for average over the solid angle and the tensor I_{ij} is the mass quadrupole moment given by the following integral ($c = 1$)

$$I_{ij}(t) = \int_{\text{source}} \mathrm{d}^3 r \left(r_i r_j - \frac{1}{3}\delta_{ij} r^2\right) T_{00}$$

For binary systems, it is easy to prove that the averaged power emitted can be expressed as

$$\frac{\mathrm{d}E_{rad}}{\mathrm{d}t} = \frac{32}{5}\frac{G^4}{c^5}\frac{\mu^2 M^3}{a^5}F(e)$$

where the factor

$$F(e) = (1-e^2)^{-7/2}\left(1 + \frac{73}{24}e^2 + \frac{37}{96}e^4\right)$$

depends only on the eccentricity and shows that highly eccentric binaries are much more efficient at radiating away energy in form of GWs.

We can also compute the averaged angular momentum flux

$$\frac{\mathrm{d}J_{rad}}{\mathrm{d}t} = \frac{32}{5}\frac{G^{7/2}}{c^5}\frac{M_1^2 M_2^2 (M_1+M_2)^{1/2}}{a^{7/2}}(1-e^2)^{-2}\left(1 + \frac{7}{8}e^2\right)$$

So GWs extract both energy and angular momentum out of the binary. As a consequence, the two masses are drawn closer, spiraling around each other until they eventually merge.

From the previous equations is possible to show that

$$\frac{\mathrm{d}a}{\mathrm{d}t} = -\frac{64}{5}\frac{G^3}{c^5}\frac{M_1 M_2 (M_1+M_2)}{a^3}F(e) \tag{147}$$

$$\frac{\mathrm{d}e}{\mathrm{d}t} = -\frac{304}{15}\frac{G^3}{c^5}\frac{M_1 M_2 (M_1+M_2)}{a^4}e(a-e^2)^{-5/2}\left(1 + \frac{121}{304}e^2\right) \tag{148}$$

4 Please note that there's not a threshold, proper, on the mass or density of the objects to start emitting GW, but rather their intensity would be so faint that we couldn't possibly dream about observing them.

from which we see that GWs drive the binary towards circularization along the inspiral.

We can integrate in time the expression for the separation and, neglecting that the eccentricity varies in time, find an estimate for the *coalescence time* of the binary

$$t_{\rm coal} = \frac{5}{256}\frac{c^5}{G^3}\frac{a_0^4}{M_1 M_2 (M_1+M_2)}\frac{1}{F(e)} \tag{149}$$

We can then rewrite the initial separation a_0 as

$$a_0 = 1.6 R_\odot \left(\frac{M_1}{M_\odot}\right)^{3/4} \left[q(1+q)F(e)\left(\frac{t_{\rm coal}}{1\,{\rm Gyr}}\right)\right]^{1/4}$$

where $q = M_2/M_1$. From the expression above it's immediate to see that the initial separation required for two objects bound in a binary to merge within an Hubble time[5] is $a_0 \approx 10^{11}$ cm ≈ 0.01 AU, which is quite the narrow range, actually.

This is also quite problematic if you think about it. For once, during their evolution, stars may undergo a giant phase characterized by $R_G \approx 10^{14}$ cm, so stars with initial separation $a_0 < R_G$ would simply engulf each other during their giant phase and merge in a single star *before* forming our nice binary of compact objects.

The scope of the next sections is trying to find a way to have our compact objects come sufficiently close so that they can get bound in a binary system and then, eventually, merge.

9.2 COMMON EVOLUTION CHANNEL (SBHS)

As you may have guessed from the title of this section, we're now turning our attention to stellar-mass BHs; in the following, we'll try to elaborate and write down a sensible mechanism to bring the two objects close enough to form our cute binary. Two formation channels consistent with the first GWs observations of SBH binaries mergers have been proposed: the *common evolution of field binaries* and the *dynamical capture in dense environments*. Let's focus on the first one.

The common evolution is the astrophysical scenario in which the two stars, eventually producing the SBH binary, form as a stellar binary system and evolve together through the different phases of stellar evolution. To achieve this, we have to look into four key ingredients, which will be the subject of the four next subsections.

9.2.1 Gravitational potential

As we've already done somewhere above, let us consider two point-like objects of mass M_1 and M_2 separated by a and moving, for simplicity, in circu-

5 The Hubble time, if you recall, is $t_H = H_0^{-1} \approx 14.4\,{\rm Gyr}$ and is a decent approximation of the age of the Universe.

lar orbits about their common center of mass. In the corotating frame, the energy potential of the system is just

$$\begin{aligned} U &= -\frac{Gm(M_1+M_2)}{r_{CM}} - \frac{1}{2}m\omega^2 r_{CM}^2 \\ &= -Gm\left(\frac{M_1}{s_1}+\frac{M_2}{s_2}\right) - \frac{1}{2}m\omega^2 r_{CM}^2 \end{aligned}$$

where s_i is the distance of a test-mass m from mass M_i and r_{CM} is the distance of said test-mass from the center of mass.

As we've already seen in the last chapter, the equilibrium points are the Lagrange points of the potential, which satisfy

$$\vec{F} = -m\nabla\phi = 0$$

There exists a critical equipotential surface, forming a two-lobed figure-of-eight, with one of the two object at the centre of each lobe and intersecting itself at the L_1 Lagrangian point, known as Roche lobes. An approximate formula for the radius of the Roche lobe around M_1 was derived by Eggleton

$$\frac{R_1}{a} = \frac{0.49q^{2/3}}{0.6q^{2/3}+\ln(1+q^{1/3})}$$

up to a 1% accuracy.

9.2.2 Mass transfer

When one of the two objects fills its Roche lobe, matter may overflow the lobe and infall onto the other object, without any need of energy exchange.

Suppose that M_2 loses material at a rate $\dot{M}_2 < 0$ and let $\beta \in [0,1]$ be the fraction of the ejected matter leaving the system, so that

$$\dot{M}_1 = -(1-\beta)\dot{M}_2 \geq 0$$

Clearly, if $\beta = 0$, then all the mass lost by M_2 is captured by M_1 and the mass transfer is fully conservative. Keeping it as a free parameter, we are instead considering the more general case in which a fraction of the mass can be lost and escape the system, as may be the case, for example, in the presence of stellar winds.

The angular momentum was $J = \mu\sqrt{GMa}$; if we now differentiate it in respect to time and use $\dot{M}_1 + \dot{M}_2 = \dot{M} = \beta\dot{M}_2$, we obtain

$$\frac{\dot{a}}{a} = -2\left(1-\frac{M_2}{M_1}\right)\frac{\dot{M}_2}{\dot{M}_1} \tag{150}$$

In short, if $M_1 > M_2$, the orbit expands ($\dot{a} > 0$), otherwise it shrinks. From Kepler's third law we also know that

$$\frac{\dot{\omega}}{\omega} = -\frac{3}{2}\frac{\dot{a}}{a}$$

and thus the angular frequency unsurprisingly increases as the orbit shrinks.

9.2.3 Supernova kicks

At the end of all the subsequent stages of nuclear burning[6], there's a chance that the more massive stars can undergo a supernova (SN) explosion[7], after which much of the stellar material is expelled and leaving behind only a shell of its former self, typically in the form of a neutron star or a stellar-mass black hole. The mass loss is practically instantaneous as the typical timescale for the explosion is much shorter than the orbital period.

In general, the collapse is not perfectly symmetric nor isotropic. As a result, the SN imprints a kick to the object characterised by a recoil velocity $v_{\rm kick}$. Since mass ejection decreases the total mass of the binary, also the gravitational potential changes and, if enough mass is ejected, the SN explosion can unbind the binary (which would suck for our purposes' sake).

Moreover, $v_{\rm kick}$ is generally (much) greater than the orbital velocity, so the kick may as well destroy most of the binaries.

To describe the effect of a SN explosion, we start by considering the setup we've used before, the notation is identical.

Prior to the SN-explosion, the relative velocity is

$$v_i = \sqrt{\frac{G(M_1 + M_2)}{a_i}}$$

After the explosion of, say, the giant star M_1, what remains of that star has mass $M_c < M_1$, and $\Delta M = M_1 - M_c$ is ejected. We can assume that, after an instantaneous explosion, the position of the once-M_1 has not changed, but the reduced mass has

$$\mu_f = \frac{M_c M_2}{M_c + M_2}$$

and the final (relative) velocity is $\vec{v}_f = \vec{v}_i + \vec{v}_{\rm kick}$. The final energy of the system will then just be

$$E_f = \frac{1}{2}\mu_f v_f^2 - \frac{G M_c M_2}{a_i}$$

The system will remain bound only if the final velocity is smaller than the escape velocity

$$v_f \leq v_e = \sqrt{\frac{2G(M_c + M_2)}{a_i}}$$

If the SN-explosion is perfectly spherically symmetric, isotropy leads to having no kick at all, and the condition for "boundness" implies

$$\Delta M \leq \frac{M_1 + M_2}{2}$$

Even in absence of kicks, a binary can be disrupted due to mass loss only, if the SN explosion ejects more than half of the initial mass of the binary system.

6 Note that nothing is *actually* burning in the common sense of "chemical burning". It's just an astrophysical slang to refer to nuclear fusion.

7 As shown in Chapter 6, §6.3.2, SN explosions are described fairly well with the Sedov-Taylor blastwave model.

The case of asymmetric SNs is more complicated and usually characterized by kicks of order $10^2\,\mathrm{km/s}$. More precisely, the natal kick distribution is typically modelled by a Maxwellian probability distribution with velocity dispersion $\sigma_v = 190\,\mathrm{km/s}$, even though, a bimodal distribution may be somewhat more appropriate, especially when describing neutron stars' kicks.

For SBHs the situation is less clear and the problem of binary disruption due to SN kicks might be less severe. In this case, there is no hard surface to bounce onto and, as a consequence, the kicks might be smaller, typically of order of $50\,\mathrm{km/s}$, which is, however, of the order of the the escape velocity for most globular clusters in the Milky Way.

9.3 COMMON ENVELOPE

Giant stars are composed of a core and an envelope, and to a good approximation we may think of decomposing the mass of the star as $M_{\mathrm{gs}} = M_{\mathrm{core}} + M_{\mathrm{env}}$. In the core, Hydrogen has been fully converted into Helium and the nuclear reactions have (momentarily) stopped, causing a core contraction under the action of gravity. The envelope, on the other hand, is mainly made of Hydrogen. It is then safe to assume, and treat, the two objects as well-separated objects.

When the giant star overfills its Roche lobe, mass transfer is allowed and, depending on the mass of the two stars, the orbit may shrink causing even more material to overflow the Roche lobe. This eventually leads to the runaway process of dynamically unstable mass transfer.

It is therefore possible that the mass transfer rate from the donor is so high that the SBH cannot accommodate all the accreting matter. In this situation, the envelope continues to expand, eventually engulfing the companion SBH and leading to the formation of a *common envelope*. The common envelope can extract energy from the orbit of the binary system, formed by the SBH and the core of the massive star, via dynamical friction, eventually unbinding itself from the system.

We can calculate the initial binding energy of the stellar envelope

$$E_{\mathrm{env},i} = -\frac{GM_{\mathrm{gs}}M_{\mathrm{env}}}{\lambda R_L}$$

where R_L is the typical Roche lobe's radius, while λ is the concentration parameter that depends on the density profile of the envelope. Usually, the density of the envelope may be well-described as a power law $\rho(r) \propto r^{-\gamma}$, $\gamma > 0$. If we assume that, at the end of the common envelope stage, the envelope unbinds, formally reaching infinity with zero velocity, we can write

$$\Delta E_{\mathrm{env}} = -E_{\mathrm{env},i} = \frac{GM_{\mathrm{gs}}M_{\mathrm{env}}}{\lambda R_L}$$

In short, the more concentrated the envelope is the more binding energy is possible to extract from it. As a consequence, the separation between the SBH and the core of the giant star decreases. We can compute the variation of the binary orbital energy

$$\Delta E_{\mathrm{orb}} = \alpha_{ce}\left[-\frac{GM_{\mathrm{core}}M_{\mathrm{BH}}}{2a_f} - \left(-\frac{G(M_{\mathrm{core}}+M_{\mathrm{env}})M_{\mathrm{BH}}}{2a_i}\right)\right] \tag{151}$$

where α_{ce} is the common envelope parameter describing the efficiency of the expenditure of orbital energy upon expulsion of the envelope.

Since conservation of energy implies

$$\Delta E_{\text{orb}} + \Delta E_{\text{env}} = 0$$

if we use the core-envelope decomposition, we can calculate the ratio between the initial and final semi-major axes of the orbit

$$\frac{a_f}{a_i} = \frac{M_{\text{core}}}{M_{\text{gs}}} \left(1 + \frac{2}{\lambda \alpha_{ce}} \frac{a_i}{R_L} \frac{M_{\text{env}}}{M_{\text{BH}}}\right)^{-1} = \frac{M_{\text{core}}}{M_{\text{gs}}} \left(\frac{M_{\text{BH}}}{M_{\text{BH}} + \frac{2M_{\text{env}}}{\lambda \alpha_{ce}} \frac{a_i}{R_L}}\right) \tag{152}$$

Despite the large uncertainties, it is possible to estimate the product $\lambda \alpha_{ce}$ by modelling specific systems or well-defined samples of objects corrected for observational selection effects. In general, the envelope is quite concentrated and typical values are of the order $\lambda \alpha_{ce} \ll 1$ and $M_{\text{core}}/M_{\text{gs}} \approx 0.2 - 0.3$. Thus, the order of magnitude approximation of (152) is

$$\frac{a_f}{a_i} \sim 10^{-3} - 10^{-2}$$

This process takes about $10^2 - 10^3$ years and upon completion the system has shrunk to $a_f \approx R_\odot$, close enough to merge in a Hubble time due to GW emission.

What we've described so far works particularly well to predict what happens when all other gravitational sources are absent.

Let's assume that the average stellar mass density in the Universe is of order $\rho_* \approx 3 \cdot 10^8 \, M_\odot/\text{Mpc}^3$; from Salpeter's initial mass function (IMF) we know that $n \sim M^{-2.35}$. This implies that roughly 0.3% of the total stars in the Universe are over $30 M_\odot$, thus leaving a SBH as relic at the end of their life. Since 70% of the massive stars are observed in binaries, and we need two stars to form a binary, we can give an estimate and claim that there are about $3 \cdot 10^5$ massive binary stars per cubic megaparsec.

We can make the (extreme) assumption that those binaries are produced in a continuous, steady-state star formation process over about 10 Gyr, thus resulting in a formation rate of $3 \cdot 10^{-5} \, \text{yr}^{-1} \, \text{Mpc}^{-3}$; assuming that all those massive binaries give rise to SBHBs that merge in a short timescale, then we get a SBHB merger rate of[8] $3 \cdot 10^4 \, \text{yr}^{-1} \, \text{Gpc}^{-3}$.

9.4 BH BINARIES IN GLOBULAR CLUSTERS

Another channel to form SBH binaries is via dynamical processes; this is closely related to the fact that most stars are observed to form in clusters and associations. However, the vast majority of the stars we observe today in the MW[9] are field stars and therefore do not belong to stellar associations[10].

8 Note that this doesn't take into account the binary systems that get disrupted by SN kicks, so the number of mergers will be rather lower. On top of that, we also assumed a constant star formation rate across cosmic time, although it's known that it peaks at about $z \approx 1.5$.

9 Short for Milky Way.

10 A "stellar association" is a group of more than 10^3 stars, like globular clusters, young massive star clusters and open clusters.

In order to understand the dynamical formation scenario, it is therefore important to start from the physics of star cluster formation and evolution.

Let's start with the basics.

9.4.1 Jeans' mass

Consider a cloud of gas. The virial theorem states that in order to achieve and maintain hydrostatic equilibrium the following relation must hold

$$-2\langle K\rangle = \langle U\rangle \tag{153}$$

where brackets now denote time-averaged quantities. From here it's clear that the condition for gravitational collapse is simply

$$2\langle K\rangle < \langle |U|\rangle$$

Under the assumption of spherical symmetry, we can write the infinitesimal gravitational potential energy over a thin shell of mass $\mathrm{d}m$

$$\mathrm{d}U = -\frac{GM(r)\mathrm{d}m}{r} = -GM(r)4\pi\rho r\mathrm{d}r$$

If density is constant, then $M(r) = 4\pi\rho r^3/3$ and the mass of a globular molecular cloud of given radius R is $M = 4\pi\rho R^3/3$. Integrating the potential energy yields

$$U = -\frac{3}{5}\frac{GM^2}{R}$$

The total internal energy of the cloud is $K = 3NkT/2$. We can define the mean molecular weight as

$$\mu = \frac{\langle m\rangle}{m_H}$$

so that $N = M/\mu m_H$. The condition for gravitational collapse can therefore be rewritten as

$$\frac{3MkT}{\mu m_H} < \frac{3}{5}\frac{GM^2}{R}$$

Solving for M and plugging back the expression for the radius of a spherical and constant distribution of matter, we obtain

$$M > M_J = \left[\frac{375}{4}\left(\frac{k}{Gm_H}\right)^3\frac{T^3}{\mu^3\rho}\right]^{1/2} \tag{154}$$

$$\approx 3\cdot 10^3 M_\odot\left[\frac{1}{\mu^4}\left(\frac{T}{100\,\mathrm{K}}\right)^3\left(\frac{10^3\,\mathrm{cm}^{-3}}{n}\right)\right]^{1/2} \tag{155}$$

where we've used $\rho = n\langle m\rangle = n\mu m_H$. If the mass of the cloud exceeds the Jeans' mass, the cloud will be unstable against gravitational collapse.

9.4.2 Free-fall timescale

Let's consider a spherical cloud with constant density. In presence of gravity alone, we can write

$$(\dot{r})\ddot{r} = -\frac{GM}{r^2}(\dot{r})$$

which can be immediately integrated to yield

$$\begin{aligned}\frac{1}{2}\dot{r}^2 &= \frac{GM}{r} + \text{const.} \\ &= \frac{GM}{r} - \frac{GM}{R}\end{aligned}$$

where we have set the initial conditions for $t = 0$: $r = R$ and $\dot{r} = 0$.

The free-fall timescale can be thus obtained by committing a mathematical atrocity, but I (and, for extension, you too) don't care[11]

$$\begin{aligned}\tau_{ff} &= \int_0^{t_{ff}} \mathrm{d}t = \int_R^0 \frac{1}{\dot{r}}\,\mathrm{d}r \\ &= \int_0^R \frac{\mathrm{d}r}{\sqrt{2GM\left(\frac{1}{r} - \frac{1}{R}\right)}} = \frac{\pi}{2\sqrt{2}}\sqrt{\frac{R^3}{GM}}\end{aligned}$$

thus, if we use $R^3/GM = 3/4\pi G\rho$

$$\tau_{ff} = \sqrt{\frac{3\pi}{32G\rho}} \approx 1.63\,\mathrm{Myr}\cdot\left(\frac{10^3\,\mathrm{cm}^{-3}}{\mu n}\right)^{1/2} \tag{156}$$

It shouldn't surprise you that under this approximation, all points get to the center at the same time, and density *at all points* increases at the same rate. This is clearly unphysical, but serves well as a tentative solution to understand what's going on in our system.

Similarly, we can calculate the explosion time, or sound time, as

$$\tau_{\text{sound}} = \frac{R}{c_s} \approx 0.5\,\mathrm{Myr}\left(\frac{R}{0.1\,\mathrm{pc}}\right)\left(\frac{0.2\,\mathrm{km\,s}^{-1}}{c_s}\right) \tag{157}$$

If $\tau_{ff} > \tau_{\text{sound}}$ the system returns to a stable equilibrium as the pressure forces can overcome gravity. Conversely, for $\tau_{ff} < \tau_{\text{sound}}$ gravitational collapse takes place.

We can define the Jeans' length as the distance over which (in spherical symmetry) is containted a Jeans' mass

$$R = \lambda_J = c_s\sqrt{\frac{3\pi}{32G\rho}}$$

More precisely, if we were to involve relativistic corrections, the Jeans' length is

$$\lambda_J = c_s\sqrt{\frac{\pi}{G\rho}}$$

11 It is actually less atrocious than I'm making it sound like.

9.4.3 Cloud fragmentation

Jeans' instability plays a very important role in star formation, because it is responsible for the fragmentation of the molecular cloud. We've seen that the Jeans' mass depends on the temperature and the density

$$M_J \propto \left(\frac{T^3}{\rho}\right)^{1/2}$$

Suppose that the cloud is big and massive enough that potential energy overcomes internal energy and the cloud starts to collapse. During the collapse, R decreases and T increases. The process can be either *isothermal* or *adiabatic*, depending on whether the gas can efficiently cool or not, which is ultimately related to the metallicity content of the gas.

Whether a cloud can cool efficiently or not depends on its metallicity. This is because at typical temperatures ($T \approx 100\,\mathrm{K}$), collisions in the gas cannot excite the atomic levels of Hydrogen and Helium ($T > 10^3\,\mathrm{K}$ is needed). Heavier elements, however, have much smaller energy gaps between atomic states. So, also at low temperatures, collisions in the gas excite atomic/molecular levels that return to the ground state by emitting photons that can effectively escape, thus cooling the cloud. Such cooling is possible if the metallicity $Z > Z_{\mathrm{crit}}$ with $10^{-4} \lesssim Z_{\mathrm{crit}}/Z_\odot \lesssim 10^{-3}$.

Adiabatic collapse

If the metallicity is below the critical threshold, the cloud cannot cool efficiently and the collapse is approximately adiabatic. So

$$pV^\gamma = \text{const.} \quad \text{or} \quad p = K\rho^\gamma \tag{158}$$

with γ the adiabatic index of the gas and K a constant. Using the ideal gas EoS we find a relation between temperature and density

$$T \propto \rho^{\gamma-1}$$

so the Jeans' mass is

$$M_J \propto \left(\frac{T^3}{\rho}\right)^{1/2} = \rho^{3\gamma/2-2}$$

For atomic Hydrogen $\gamma = 5/3$ and $M_J \propto \rho^{1/2}$, so that the Jeans' mass increases as the density increases. As a consequence, the cloud cannot fragment and the collapse is approximately *monolithic*[12]; in other words, during an adiabatic collapse, the cloud contracts without fragmenting, thus potentially forming a single massive protostar of the order of $10^3 M_\odot$. The reason for this upper limit comes from the fact that, when the cloud collapses adiabatically, the temperature eventually increases up to $T \approx 10^2\,\mathrm{K}$ ($M_J \approx 10^3 M_\odot$); at that point, molecular hydrogen transitions start to cool the cloud, and the Jeans' mass does not increase anymore.

12 The Jeans mass sets a lower bound to the mass of an object to be stable against gravitational collapse. Hence, if M_J increases as the cloud collapses (for ρ is always increasing during a collapse), only objects with mass $M > M_J$ may be subject to gravitational collapse. In short, only the cloud as a whole can collapse.

Adiabatic collapse is held responsible as one of the main formation channels of IMBHs. Remarkably, early on in structure formation, at redshift $15 < z < 20$, the gas metallicity was very low, so the first protostellar clouds likely collapsed with little fragmentation, leaving behind a first generation of stars (PopIII stars) that were very massive and could potentially evolve into IMBHs characterized by mass of the order of $10^2 - 10^3 M_\odot$.

Isothermal collapse

Conversely to what we've seen in the last section, for Z over the critical threshold the collapse is essentially isothermal: Temperature remains approximately constant and the mass limit for instability decreases when the density of the cloud increases as $M_J \propto \rho^{-1/2}$. Any initial density perturbations will then cause individual regions within the cloud to cross the instability threshold independently and collapse locally, forming a large number of smaller objects: Isothermal collapse naturally leads to *fragmentation*.

It is obvious, however, that in order for protostars to *actually* form, the processes cannot remain isothermal indefinitely, for temperatures of order $T \sim 10^7$ K are needed to turn on nuclear reactions.

This happens when the clump density becomes high enough that gas becomes opaque to infrared photons. At this point, the cloud cannot cool efficiently, the isothermal approximation breaks down and further evolution is approximately adiabatic. As a consequence, temperature increases and so does M_J, thus leading to a minimum fragment size into which the cloud can break up. This is known as the *opacity limit*.

Opacity limit

The energy released during the collapse of the protostellar cloud obeys the virial theorem (153), so

$$\langle E \rangle = \langle U \rangle + \langle K \rangle = \frac{\langle U \rangle}{2}$$

Only half the change in gravitational potential energy can be radiated away. Since (assuming spherical symmetry)

$$U = -\frac{3}{5}\frac{GM^2}{R}$$

the energy released is $\Delta E_g = 3GM^2/10R$, from which we can calculate the emitted (average) luminosity over the free-fall time

$$L_{ff} = \frac{\Delta E_g}{\tau_{ff}} = \frac{3}{10}\frac{GM^2}{R}\left(\frac{3\pi}{32G\rho}\right)^{-1/2} = \frac{3\sqrt{2}}{5\pi}G^{3/2}\left(\frac{M}{R}\right)^{5/2} \tag{159}$$

This is the power that, in the opacity limit, is absorbed by the gas. For the sake of simplicity, we're going to assume that the gas emits radiation as a grey-body, thus the cloud can radiate away energy at a rate given by

$$L_{\text{rad}} = 4\pi R^2 e_f \sigma T^4$$

where we have introduced the efficiency factor e_f ($0 < e_f < 1$), because the collapsing cloud is not in thermodynamic equilibrium. The gas will start to

heat up (becoming adiabatic) as soon as the cooling becomes slower than the heating rate

$$L_{rad} \lesssim L_{ff}$$

The isothermal approximation then breaks down where the two luminosities are equal

$$M^5 \lesssim M^5_{\rm crit} = \frac{200}{9}\frac{\pi^4\sigma^2}{G^3}e_f^2 R^9 T^8 \tag{160}$$

This gives us a further constraint on the mass of the cluster: $M_J < M < M_{\rm crit}$. The opacity limit is reached when $M_J = M_{\rm crit}$. For typical values of $\mu = 1$, $e_f = 0.1$, $T = 10^3\,\mathrm{K}$ we have a mass for the single fragment of order $M_{\rm frag} \approx 0.2 M_\odot$. Hence, fragmentation ceases when individual fragments are approximately solar-mass objects.

9.5 EVOLUTION OF STAR CLUSTERS

Consider an idealized cluster of size R consisting of N (for simplicity) identical stars with mass m uniformely distributed.

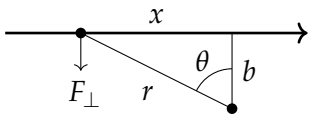

Figure 44: A field star approaches another star at speed v and impact parameter b.

9.5.1 Two-body relaxation timescale

Given this setup, depicted in Fig.44, we're now going to show that stars in clusters can reach equilibrium through mutual interactions in a process named *two-body relaxation*, which is essentially analogous to thermalization. Therefore, a very important timescale in collisional dynamics in clusters is the time for a star to completely lose memory of its initial velocity by means of gravitational encounters.

The gravitational force is

$$\vec{F} = -\frac{Gm^2}{r^3}\vec{r} = \vec{F}_\parallel + \vec{F}_\perp$$

where the orthogonal component is

$$F_\perp = -\frac{Gm^2}{r^2}\frac{b}{r} = -G\frac{m^2}{b^2}\left[1 + \left(\frac{vt}{b}\right)^2\right]^{-3/2} = -m\dot{v}_\perp$$

since $r^2 = b^2 + (vt)^2$. The change in velocity integrated over one entire encounter is

$$\delta v_\perp = \int_{-\infty}^{\infty} \dot{v}_\perp \,\mathrm{d}t = \frac{Gm}{b^2}\int_{-\infty}^{\infty}\left[1 + \left(\frac{vt}{b}\right)^2\right]^{-3/2}\mathrm{d}t$$

Note that for symmetry reasons it must be $\delta v_{\parallel} = 0$. The integral above is trivial and evaluates to[13]

$$\delta v_{\perp} = \frac{2Gm}{bv}$$

Now taking into account all stars in the system, the surface density of stars in an idelized cluster is $N/\pi R^2$, and the number of interactions per unit element is

$$\delta n = \frac{N}{\pi R^2} \mathrm{d}(\pi b^2) = \frac{2bN}{R^2} \mathrm{d}n$$

Defining $\delta v_{\mathrm{tot}}^2 = \int \delta v_{\perp}^2 \, \delta n$, where the integral is performed over all possible impact parameters, we get

$$\delta v_{\mathrm{tot}}^2 = 8N \left(\frac{Gm}{Rv} \right)^2 \log \left(\frac{b_{\mathrm{max}}}{b_{\mathrm{min}}} \right)$$

The integration limit b_{max} is of the order of the size of the system, while b_{min} corresponds to the smallest b to avoid stellar collisons, which can be shown to be equal to

$$b_{\mathrm{min}} = \frac{2Gm}{v^2}$$

thus leading to

$$\delta v_{\mathrm{tot}}^2 = 8N \left(\frac{Gm}{Rv} \right)^2 \log \left(\frac{Rv^2}{2Gm} \right)$$

The typical speed of a star in a virialized system[14] is given by

$$Nmv^2 = \frac{G(Nm)^2}{R}$$

Replacing for v in the equation for δv_{tot} we get

$$\delta v_{\mathrm{tot}}^2 = v^2 \frac{8}{N} \log \left(\frac{N}{2} \right)$$

The number of crossings for which $\delta v_{\mathrm{tot}}^2 / v^2 \approx 1$ (that is the number of crossings after which the star has changed its initial velocity completely and lost memory of its initial conditions) is given by

$$n_{\mathrm{cross}} \approx \frac{N}{8} \frac{1}{\log(N/2)}$$

The time needed to cross the system is

$$\tau_{\mathrm{cross}} = \frac{R}{v} = \sqrt{\frac{R^3}{GNm}} \propto \frac{1}{\sqrt{G\rho}} \propto \tau_{ff}$$

while the time necessary for stars in a system to completely lose the memory of their initial velocity, called the relaxation time, is

$$\tau_{rlx} = n_{\mathrm{cross}} \tau_{\mathrm{cross}} = \frac{N}{8} \frac{1}{\log(N/2)} \frac{R}{v} \approx 10^7 - 10^{10} \, \mathrm{yr} \tag{161}$$

Collisional systems have $\tau_{rlx} \ll t_{\mathrm{Hubble}}$ and are therefore "relaxed" by two body interactions to a state that does not retain memory of their initial conditions.

13 Before you slander me: Set $vt/b = \sinh(y)$, then unpack the hyperbolic cosine that pops out and set $z = e^{2y}$. The integral is indeed trivial and evaluates to $2b/v$.

14 We're referring to stars that obey the virial theorem (153) and the equation that follows is merely a restating of the more general theorem.

9.5.2 Infant mortality

Most stars form in clusters, but nowadays we're observing only a minority of stars in clusters and associations. This happens because most clusters actually dissolve as they go through certain processes, as, for example, infant mortality.

Clusters are essentially bound systems of stars and gas, as so, they're not immune to aging[15]. As the more massive stars go through their life cycle, it often happens that stellar winds or, more dramatically, SN explosions disrupt the cluster and blow away its internal gas, which won't ever be turned into new stars. This leads to a decreasing of the total gravitational potential holding the stars together, and so stars can escape in an irreversible runaway process that can completely dissolve the cluster.

For a cluster with mass $M = M_g + M_*$, if the mass of the gas getting lost is $M_g > M_*$, then the cluster dissolves. From the virial theorem, the velocity dispersion of the stars before gas removal is estimated as

$$\sigma_0^2 = \frac{G(M_g + M_*)}{R_0}$$

where R_0 is the initial size of the cluster. Assuming instantaneous gas removal

$$E_f = \frac{1}{2} M_* \sigma_0^2 - \frac{G M_*^2}{R_0} = -\frac{G M_*^2}{2 R_f}$$

This implies that

$$-\frac{M_*}{R_f} = \frac{1}{2}\frac{M_g + M_*}{R_0} - \frac{M_*}{R_0}$$

Finally, solving for R_f yields

$$\frac{R_f}{R_i} = \left[1 - \frac{1}{2}\left(1 + \frac{M_g}{M_*}\right)\right]^{-1} \qquad R_f > 0 \iff M_g < M_* \tag{162}$$

Therefore, the cluster survives only if the expelled gaseous mass is smaller than the total mass in stars. This result is a strong lower limit to the maximum mass of gas which can be expelled without destroying the star cluster, because we assume instantaneous expulsion of the gas component. With more accurate calculations, assuming an actual timescale for the gas expulsion, we'd get $M_g \lesssim 4 M_*$.

9.5.3 Evaporation

Evaporation of a cluster is yet another process able to disrupt a cluster. Describing the star cluster as a self-gravitating system, we can define an escape velocity for any given distance from its center

$$v_e^2(r) = -2\phi(r)$$

The mean square escape velocity is then obtained by averaging over the (spherically symmetric) mass density distribution $\rho(r)$

$$\langle v_e^2 \rangle = \frac{\int \rho(r) v_e^2(r)\, \mathrm{d}r}{\int \rho(r)\, \mathrm{d}r} = -\frac{2}{M}\int \rho(r)\phi(r)\, \mathrm{d}r = -\frac{4E_g}{M}$$

15 So to say.

where E_g is the potential self-energy and M is the total mass. Using the virial theorem once more leads to

$$\langle v_e^2 \rangle = 4\langle v^2 \rangle$$

Therefore, stars with velocities exceeding twice the root-mean-square (RMS) velocity of the distribution are unbound. For a typical Maxwellian velocity distribution, this amounts to a fraction $\epsilon = 7.4 \cdot 10^{-3}$ of all stars.

Roughly, evaporation removes $\mathrm{d}N = -\epsilon N$ stars on a relaxation timescale, so the system gets slightly hotter and contracts. The velocity distribution adjusts to another Maxwellian distribution, so that in every relaxation time ϵN stars are removed by evaporation

$$\frac{\mathrm{d}N}{\mathrm{d}t} = -\frac{\epsilon N}{t_{rlx}} := -\frac{N}{t_{\mathrm{evap}}} \tag{163}$$

Note that the evaporation timescale is much greater than the relaxation timescale.

9.5.4 Core Collapse of the Cluster

As we're going to see, evaporation is usually enough to lead to *core collapse*. In the last section we've shown that stars with velocity $v \geq v_e = 2\sqrt{\langle v^2 \rangle}$ are unbound, adding up to the 0.74% of the total population, if the velocity distribution were exactly Maxwellian, that is. Although evaporation is generally not a steady-state process, we can search for a self-similar solution that can describe evaporation with sufficient accuracy. In the self-similar regime, we expect a constant rate of mass loss

$$\dot{M} = -\xi_e \frac{M(t)}{\tau_{rlx}(t)}$$

Neglecting changes in the logarithm in (161), the relaxation time is proportional to

$$\tau_{rlx} \propto \sqrt{M} R^{3/2}$$

so that

$$\tau_{rlx} = \tau_{rlx}(0) \left(\frac{R(t)}{R(0)} \right)^{3/2} \left(\frac{M(t)}{M(0)} \right)^{1/2}$$

Using everything we've found so far, we get to

$$\dot{M} = -\xi_e \frac{M(0)}{\tau_{rlx}(0)} \left(\frac{R(t)}{R(0)} \right)^{-3/2} \left(\frac{M(t)}{M(0)} \right)^{1/2}$$

Each star escaping from the cluster carries away a certain kinetic energy per unit mass

$$\frac{\mathrm{d}E_{tot}}{\mathrm{d}M} = \zeta E_m = \zeta \frac{E_{tot}}{M}$$

As a consequence

$$\frac{\mathrm{d}E_{tot}}{\mathrm{d}t} = \frac{\mathrm{d}E_{tot}}{\mathrm{d}M} \dot{M} = \zeta \frac{E_{tot}}{M} \dot{M}$$

But recall that $E_{tot} \propto -M^2/R$, so we have

$$\zeta \frac{E_{tot}}{M}\dot{M} \propto -\zeta \frac{M}{R}\dot{M}$$

On the other hand, we can also implicitly differentiate the relation for the total energy with respect to time

$$E_{tot} \propto -\frac{M^2}{R} \implies \frac{\mathrm{d}E_{tot}}{\mathrm{d}t} \propto -\frac{2M}{R}\dot{M} + \frac{M^2}{R^2}\frac{\mathrm{d}R}{\mathrm{d}t}$$

In conclusion

$$(2-\zeta)\frac{\mathrm{d}M}{M} = \frac{\mathrm{d}R}{R}$$

This we can integrate right away

$$\frac{R(t)}{R(0)} = \left(\frac{M(t)}{M(0)}\right)^{2-\zeta} \implies \frac{\rho(t)}{\rho(0)} \propto \left(\frac{M(t)}{M(0)}\right)^{3\zeta-5}$$

For realistic clusters $\zeta < 1$; this makes sense since that most of the stars are ejected just barely over v_e, so that their velocity at infinity is generally $v_\infty < \sigma$. Since the typical energy per unit mass of particles is of order σ^2, we have $v_\infty^2 < \sigma^2$ and thus $\zeta < 1$.

Note that $\rho(t) \to \infty$ as $M(t) \to 0$. We can now solve the equation for the mass flux $\dot{M}$, which bears as a result

$$M(t) = M(0)\left[1 - \frac{\xi_e(7-3\xi)}{2}\frac{t}{\tau_{rlx}(0)}\right]^{2/(7-3\zeta)} := M(0)\left[1 - \frac{t}{\tau_0}\right]^{2/(7-3\zeta)}$$

where we've defined the collapse time τ_0 that satisfies

$$M(\tau_0) = R(\tau_0) = 0$$

For a cluster composed of equal-mass stars, the collapse time is $\tau_0 \geq 10\tau_{rlx}$.

Post Core-Collapse

What happens after the core has collapsed due to the evaporation of the cluster?

We've seen that mass loss due to evaporation leads to collapse, a runaway process called gravothermal instability. If the system contracts, it becomes denser and the two-body encounter rate increases along with the evaporation rate. As a consequence, the core of the cluster loses energy (the kinetic energy of the evaporated stars) to the halo, and $\mathrm{d}E_{\mathrm{core}} < 0$. Since for any bound, finite system in which the dominant force is gravity we have

$$C := \frac{\mathrm{d}E}{\mathrm{d}T} < 0$$

the temperature of the core must increase. Therefore, stars exchange more energy and become dynamically hotter, and faster stars tend to evaporate at a higher rate. This runaway process leads to the unphysical situation of star clusters with infinite core density.

To avoid this catastrophic scenario, we consider the possibility that the core collapse is reversed by an external source injecting kinetic energy into

the core ($dE_{\text{core}} > 0$), thus cooling it ($dT_{\text{core}} < 0$) until the temperature gradient declines to zero, thus halting the collapse. There are multiple ways this can happen: Mass loss by stellar winds and/or SNs; formation of binaries; three-body encounters between single stars and binaries extracting kinetic energy from the internal energy of the binary system.

Consider this last scenario with the three stars initially far away from each other. The initial energy will then be just the sum of the respective kinetic energies $E = K_1 + K_2 + K_3$.

Once the binary is formed, the total energy will become $E = K_{\text{bin}} + E_{\text{bin}} + K_3'$. Since $E_{\text{bin}} < 0$, conservation of energy implies

$$K_{\text{bin}} + K_3' > K_1 + K_2 + K_3$$

from which we see that the kinetic energy after the interaction is larger than the initial kinetic energy of the three stars. Therefore, the formation of binaries can pump kinetic energy into single stars crossing the core of the cluster (where most of the binaries form). Those stars then share the acquired extra kinetic energy with other stars through two-body relaxation, heating up the cluster.

9.5.5 Mass Segregation and Spitzer's instability

All we've gone through up until now works decently for any stellar population, given they have the same mass. In reality, not surprisingly, not all the stars in a cluster will have the same mass, and will rather sit somewhere into a spectrum ranging from $\sim 0.5 M_\odot$ to $\sim 150 M_\odot$.

First of all, as we'll see in the next sections, massive-above-than-average stars are expected to go through a process called *dynamical friction*. This means that a massive star walking through a sea of lighter stars feels a drag force, which decelerates its motion. The timescale of dynamical friction for a star of mass M is approximately

$$\tau_{\text{segr}} = \frac{\langle m \rangle}{M_{tot}} \tau_{rlx}$$

where $\langle m \rangle$ is the average star mass and τ_{rlx} the two-body relaxation timescale. The effect of dynamical friction is that the most massive stars in a star cluster lose kinetic energy in favor of the light stars and segregate toward the centre of the star cluster. This generates the phenomenon called *mass segregation*: The radial distribution of massive stars tends to be more centrally concentrated than the average stellar distribution in dense star clusters. Notice how such a process may be able to bring two stars close enough to form a binary system.

At equilibrium, energy is shared equally by all masses

$$\frac{1}{2} m_i \langle v_i^2 \rangle = \frac{1}{2} m_j \langle v_j^2 \rangle \tag{164}$$

If the velocities of all stars are initially drawn from the same distribution, massive stars are thus expected to transfer kinetic energy to lighter stars and slow down, till they reach equipartition. But can equilbrium always be reached?

Consider the case of a cluster composed by N_1 stars of mass m_1 and N_2 of mass m_2 so that $m_2 \gg m_1$ and $N_2 \ll N_1$, in a way that, if we define $M_i = N_i m_i$, then $M_2 \ll M_1$. Let ρ_i be the local density of stars of mass m_i. From the virial theorem

$$\langle v_i^2 \rangle = \alpha \frac{GM_i}{r_i} + \frac{G}{M_i} \int_0^\infty \rho_i \frac{M_j(r)}{r} 4\pi r^2 \, \mathrm{d}r$$

where the first term describes the self-gravity of the population (with α a parameter that describes the density distribution throughout the cluster) and the second describes the gravitational energy that comes through the interaction of the two popolations; r_i is the half-mass radius of population.

As a consequence of segregation, the more massive stars become centrally concentrated compared to the distribution of low-mass stars. We can thus assume that the density of lighter stars is constant $\rho_1(r) \approx \rho_1(0)$ throughout the region occupied by the heavier stars, so that

$$M_1(r) = \frac{4\pi}{3} r^3 \rho_{cl,1}$$

where $\rho_{cl,1}$ is the central density of stars of mass m_1. Under these assumptions, the equations for the RMS-velocities are simplified[16]

$$\begin{aligned} \langle v_1^2 \rangle &= \frac{\alpha G M_1}{r_1} \\ \langle v_2^2 \rangle &= \frac{\alpha G M_2}{r_2} + \frac{4\pi G}{3} \rho_{cl,1} R_{sp,2}^2 \end{aligned}$$

where we've defined Spitzer's radius as

$$R_{sp,2}^2 = \frac{1}{M_2} \int_0^\infty r^2 \rho_2 (4\pi r^2) \, \mathrm{d}r \tag{165}$$

Defining the mean density of stars of each type within their half-mass radius

$$\rho_{m,i} = \frac{3}{4\pi r_i^3} M_i$$

we can invert it and express the half-mass radius in terms of the mean density of the i-th population, so that substituting into the equipartition condition (164) yields

$$\chi = \frac{M_2}{M_1} \left(\frac{m_2}{m_1} \right)^{\frac{3}{2}} = \left(\frac{\rho_{m,1}}{\rho_{m,2}} \right)^{\frac{1}{2}} \left[1 + \beta \left(\frac{\rho_{m,1}}{\rho_{m,2}} \right) \right]^{-\frac{3}{2}} \qquad \beta = \frac{\rho_{c,1}}{\rho_{c,2}} \frac{1}{2\alpha} \left(\frac{R_{sp,1}}{R_{sp,2}} \right)^2 \tag{166}$$

The expression above has the maximum value

$$\frac{\rho_1}{\rho_2} = (2\beta)^{-1} \implies \chi_{\max} = \sqrt{\frac{4}{27\beta}}$$

A realistic value for β is ≈ 5.8, so that $\chi_{\max} \approx 0.16$. So, Spitzer's condition (166) predicts equilibrium against mass segregation only if

$$\frac{M_2}{M_1} \left(\frac{m_2}{m_1} \right)^{3/2} < 0.16 \tag{167}$$

16 In the equation for $\langle v_2^2 \rangle$ we substitute the relation for M_1 we've written above.

For a realistic cluster, $m_2 \approx 10m_1$ and $M_2 \approx 0.03M_1$. This implies $\chi \approx 1 > \chi_{\max}$, so there's no reaching equilibrium and massive stars are dragged towards the center as a result of mass segregation. This system will thus be prone to the formation of tight SBH binaries via capture and other processes.

9.6 STELLAR AND BH BINARIES HARDENING

Recall that the internal energy of a binary is

$$E_{\text{bin}} = -\frac{GM_1M_2}{2a} = -E_b$$

Consider now a 3-body interaction between the binary and a third object m_3 (a "intruder"). If the original binary is preserved in the encounter, there are two possibilities

- the single body extracts internal energy from the binary, so that the final kinetic energy of center of mass of the intruder and of the binary is higher than the initial one;
- the single body loses a fraction of its kinetic energy, which is converted into internal energy of the binary.

In the first case, the object and the binary acquire recoil velocity and the binding energy increases

$$K_i - E_{b,i} = K_f - E_{b,f} \implies E_{b,f} - E_{b,i} = K_f - K_i$$

for $K_f > K_i$[17]

$$E_{b,f} = \frac{GM_1M_2}{2a_f} > \frac{GM_1M_2}{2a_i} = E_{b,i} \implies a_f < a_i$$

The result is flipped in the other case.

Another possibility for the binary to increase the binding energy during a 3-body interaction is the exchange, for example if the intruder replaces one of the members of the binary. This usually happens when $M_2 < m_3 < M_1$, in which case, after the exchange the binary is formed by M_1 and m_3

$$E_{b,f} = \frac{GM_1m_3}{2a} > \frac{GM_1M_2}{2a} = E_{b,i}$$

The final binary can also becomes less bound and can even be "ionized" if its velocity at infinity exceeds the critical velocity v_c. In fact

$$E_f = \frac{1}{2}\frac{m_3(M_1+M_2)}{(M_1+M_2+m_3)}v^2 - \frac{GM_1M_2}{2a}$$

and the system is unbound if $E_f = 0$, so if

$$v_c = \sqrt{\frac{GM_1M_2(M_1+M_2+m_3)}{am_3(M_1+M_2)}} \tag{168}$$

We define *hard binaries* those with a binding energy greater than $\langle m\rangle\sigma^2/2$, where σ is the average velocity of the stars and $\langle m\rangle$ the average mass. Conversely, *soft binaries* satisfies the inverse relation.

Heggie's law states that hard (soft) binaries only tend to get harder (softer).

17 Note that this are the total kinetic energies of the 3-body system.

Cross Section for 3-body Encounters

To define the cross section for 3-body encounters, let us consider the maximum impact parameter $b_{\max}$ for a non-zero energy exchange between the single object m_3 and the binary. To estimate the impact parameter, we need to consider gravitational focusing, that is the fact that the trajectory of the intruder is significantly deflected by the presence of the binary, thus approaching it with an effective pericentre p much smaller than the formal impact parameter b at infinity.

Conservation of energy implies

$$\Delta E = 0 = \frac{m_3(M_1+M_2)}{M_1+M_2+m_3}(v_f^2 - v_i^2) + Gm_3(M_1+M_2)\left(\frac{1}{D} - \frac{1}{p}\right)$$

where D is the initial distance between the single object and the binary, and for the initial velocity of the single object we consider $v_i = \sigma$. Assuming $D \ll p$ (the periastron), the equation simplifies to

$$\frac{1}{2}\frac{\sigma^2}{M_1+M_2+m_3} = \frac{1}{2}\frac{v_f^2}{M_1+M_2+m_3} - \frac{G}{p}$$

On the other hand, we also have to consider angular momentum conservation

$$\Delta J = 0 = (pv_f - b\sigma)\frac{m_3(M_1+M_2)}{(M_1+M_2+m_3)}$$

so that $pv_f = b\sigma$. Combining the two conservation equations and solving for p

$$p = \frac{G(M_1+M_2+m_3)}{\sigma^2}\left[\sqrt{1+\frac{b^2\sigma^4}{G^2(M_1+M_2+M_3)^2}} - 1\right]$$

which can be Taylor expanded for

$$\frac{b^2\sigma^4}{G^2(M_1+M_2+M_3)^2} \ll 1$$

to finally yield

$$p \simeq \frac{b^2\sigma^2}{2G(M_1+M_2+M_3)} \tag{169}$$

The 3-body cross section is then defined just as

$$\Sigma = \pi b_{\max}^2 \simeq \pi\left[\frac{2G(M_1+M_2+m_3)}{\sigma^2}\right]p_{max} \simeq \frac{2\pi G(M_1+M_2+m_3)a}{\sigma^2}$$

where we have approximated $p_{\max} \simeq a$, which holds only for very energetic 3-body encounters.

9.6.1 3-body Hardening

Since we now have an expression for the 3-body cross section, we can estimate the interaction rate

$$\frac{\mathrm{d}N}{\mathrm{d}t} = n\Sigma\sigma = \frac{2\pi G(M_1+M_2+m_3)na}{\sigma}$$

We now make a series of simplifying assumptions that characterize those binaries that will eventually become GW sources. Importantly, those assumptions are relevant for SBHs and SMBHs alike, thus providing a useful description to the dynamics of SBH binaries and MBH binaries. We assume that

1. the binary is hard;
2. the effective pericentre satisfies $p \lesssim 2a$;
3. the mass of the intruder is small in respect to the binary $m_3 \ll M_1, M_2$.

This way, the average binding energy variation per encounter reads something like

$$\langle \Delta E_b \rangle = \xi \frac{m_3}{M_1 + M_2} E_b = \xi \frac{m_3}{M_1 + M_2} \frac{G M_1 M_2}{2a}$$

where $\xi \approx 0.2 - 1$ is a parameter that can be extracted from 3-body scattering experiments. The rate of binding energy exchange for a hard binary is

$$\frac{\mathrm{d}E_b}{\mathrm{d}t} = \langle \Delta E_b \rangle \frac{\mathrm{d}N}{\mathrm{d}t} = 2\pi\xi \frac{M_1 M_2 m_3 (M_1 + M_2 + m_3)}{M_1 + M_2} \frac{G^2 n}{\sigma}$$

Supposing a single mass population of intruders characterized by $m_3 = \langle m \rangle$, we can write the rate of binding energy exchange in terms of the local mass density $\rho = n\langle m \rangle$. Exploiting the mass "hierarchy" condition

$$\frac{\mathrm{d}E_b}{\mathrm{d}t} = \frac{2\pi\xi G^2 M_1 M_2 \rho}{\sigma}$$

Therefore, hard binaries harden at a constant rate!

Expressing a in terms of E_b the hardening rate is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{a}\right) = \frac{2}{G M_1 M_2} \frac{\mathrm{d}E_b}{\mathrm{d}t} = 4\pi G \xi \frac{\rho}{\sigma} = \frac{G H \rho}{\sigma^2}$$

which can be written as

$$\frac{\mathrm{d}a}{\mathrm{d}t} = -\frac{G H \rho}{\sigma} a^2 \tag{170}$$

where we've defined $H \approx 15 - 20$ as a dimensionless hardening rate[18].

9.6.2 Hardening and GWs

How is all this stuff connected to GWs?

From (170) we can see that hardening in a given stellar background proceeds at a constant rate determined by the properties of the background. The evolution of semi-major axis can be thought as composed of two contributions

$$\frac{\mathrm{d}a}{\mathrm{d}t} = \left(\frac{\mathrm{d}a}{\mathrm{d}t}\right)_{3b} + \left(\frac{\mathrm{d}a}{\mathrm{d}t}\right)_{GW} = -Aa^2 - Ba^{-3}$$

where we have expliticted the dependence on a of both terms and defined

$$A = \frac{G H \rho}{\sigma} \quad B = \frac{64}{5} \frac{G^3}{c^5} M_1 M_2 (M_1 + M_2) F(e)$$

18 The lower bound is for circular orbits, while 20 is for very eccentric orbits.

Since stellar hardening is $\propto a^2$ and the GW hardening is $\propto a^{-3}$, binaries spend most of their time at the transition separation obtained by imposing $(\mathrm{d}a/\mathrm{d}t)_{3b} = (\mathrm{d}a/\mathrm{d}t)_{GW}$

$$\bar{a} = \left[\frac{64G^2\sigma M_1 M_2(M_1+M_2)F(e)}{5c^5 H\rho}\right]^{1/5}$$

and their lifetime can be written as

$$\tau(\bar{a}) = \frac{\sigma}{GH\rho\bar{a}} \approx 3\,\mathrm{Gyr}\left(\frac{\sigma}{10\,\mathrm{km\,s^{-1}}}\right)^{1/5}\left(\frac{\rho}{10^5 M_\odot\,\mathrm{pc^{-3}}}\frac{\bar{a}}{0.15\,\mathrm{AU}}\right)^{-1}$$

Given this expression, is it possible to enter the GW hardening regime in less than a Hubble time? Yes!

For typical[19] values, $M_1 + M_2 \approx 60M_\odot$, $\sigma = 10\,\mathrm{km\,s^{-1}}$, $\rho = 10^5 M_\odot\,\mathrm{pc^{-3}}$, you can get a hardening time $\tau(\bar{a}) \approx 3\,\mathrm{Gyr} < t_{\mathrm{Hubble}}$.

Moreover, the mechanism efficiency increases with the BH binary mass[20]. If IMBHs can indeed form in star clusters, stellar hardening provides an efficient mechanism to merge them with SBHs or with a companion IMBH.

9.7 SUPERMASSIVE BLACK HOLES

We now turn to discuss some relevant astrophysical aspects of (S)MBHs. Those objects has been observed at the centre of massive galaxies, and inhabit virtually all nuclei of galaxies with $M_* > 10^{11}M_\odot$, whereas their ubiquity in lighter galaxies is much debated. Our discussion will be mainly classical (Newtonian), although some results of GR will be implemented when necessary.

9.7.1 Bondi accretion

In order to accrete, a MBH needs to capture gas from its surroundings at a sufficient rate. We've already discussed about this in Chapter 7, but a brief refresher won't do much harm.

The model assumes an object (a black hole in our case) of mass M surrounded by an infinite cloud of gas, accreting with stationary and spherically symmetric motion. The model neglects any self-gravity effects of the cloud, magnetic fields, angular momentum and viscosity due to the accretion mechanism.

The gas is assumed to be perfect and polytropic

$$p = p_\infty\left(\frac{\rho}{\rho_\infty}\right)^\gamma$$

with γ the polytropic index. Because of stationarity and spherical symmetry, conservation of mass $D_\mu(\rho u^\mu) = 0$ implies a constant accretion rate

$$\dot{M} = 4\pi r^2\rho v = \text{const.}$$

19 Typical as long as massive SBH binaries are concerned.

20 Recall that $\bar{a} \propto M_1M_2(M_1+M_2)$, while $\tau(\bar{a}) \propto \bar{a}^{-1}$.

where r is the radial coordinate and v in the inward velocity of the gas. We can also define the sound speed of the gas $c_s^2 = \gamma p_\infty / \rho_\infty$ at infinity and a characteristic lengthscale of the problem, called *Bondi radius*

$$r_B = \frac{GM}{c_s^2}$$

I hope all these things are actually ringing some bells, no matter how out of tune they may be. We can the rescale the variables

$$\begin{aligned} r &= x r_B \\ v &= y c_s \\ \rho &= z \rho_\infty \end{aligned}$$

All the dynamic of the system can be thus summarized in just two equations

$$\begin{aligned} x^2 y z &= \lambda \\ \frac{y^2}{2} + \frac{z^{\gamma-1} - 1}{\gamma - 1} - \frac{1}{x} &= 0 \end{aligned}$$

where the dimensionless accretion rate parameter λ is just

$$\lambda = \frac{\dot{M}_B}{4\pi r_B^2 c_s \rho_\infty} = \frac{\dot{M} c_s^3}{4\pi G^2 M^2 \rho_\infty}$$

For $\lambda = 1$, we get Bondi's accretion rate

$$\dot{M}_B \approx 2.5 \cdot 10^2 \left(\frac{c_s}{100\,\mathrm{km\,s^{-1}}} \right)^{-3} \left(\frac{M}{10^8 M_\odot} \right)^2 M_\odot\,\mathrm{yr}^{-1}$$

Therefore $\dot{M}_B \propto c_s^{-3} M^2$. The mass dependence, in particular, implies that the mass of the compact object diverges to infinity in a finite amount of time for an infinite "fuel" supply.

It should not surprise you that this is largely unphysical, for once we're assuming *perfect* spherical symmetry and neglecting all feedbacks due to the radiation emitted by the accretion flow.

9.7.2 The Eddington Limit

Once again, here's something we've already got acquainted with back in Chapter 7.

The radiation reaction onto the accretion flow is the physical rationale behind the Eddington accretion limit, and sets the maximum luminosity L that an AGN (or, in fact, any astrophysical object) can emit when the radiation force acting outward equals the gravitational force acting inward. Beyond this limit, the radiation force overwhelms the gravitational force and the accretion process is considerably softened or halted.

Let's assume a spherically symmetric cloud of fully ionized Hydrogen around our BH, so that the main channel of interaction for photons is the Thomson scattering.

The flux of energy through a spherical surface of radius r is

$$\Phi = \frac{L}{4\pi r^2}$$

and the momentum flux (pressure) is

$$p_{\text{rad}} = \frac{\Phi}{c} = \frac{L/c}{4\pi r^2}$$

Therefore, the force exerted by the radiation on a single electron is given by

$$F_{\text{rad}} = p_{\text{rad}}\sigma_{Th} = \frac{L}{4\pi r^2 c}\frac{8\pi}{3}\left(\frac{e^2}{m_e c^2}\right)^2$$

The dependence of the cross section on the particle mass justifies our assumption to neglect scattering with heavier particles, namely protons.

We can write the gravitational force too

$$F_g = \frac{GM(m_e + m_p)}{r^2} \approx \frac{GMm_p}{r^2}$$

Equating the two forces and solving for L gives us our longed for Eddington's luminosity

$$L_{\text{Edd}} = \frac{3GMm_p c}{2}\left(\frac{m_e c^2}{e^2}\right)^2 \approx 1.26 \cdot 10^{28}\left(\frac{M}{M_\odot}\right)\,\text{erg}\,\text{s}^{-1} \tag{171}$$

Now, assume that the accretion process occurs at a given rate $\dot{M}$, and that only a fraction[21] ϵ of the rest mass energy of the accreted matter is radiated away. Then, the luminosity can be expressed as

$$L = \epsilon \dot{M} c^2$$

thus setting a limit on the accretion rate

$$\dot{M}_{\text{Edd}} = \frac{4\pi GMm_p}{c\sigma_{th}}\frac{M}{\epsilon} \approx 2.2 \cdot 10^{-8}\left(\frac{\epsilon}{0.1}\right)\left(\frac{M}{M_\odot}\right) M_\odot\,\text{yr}$$

For a gas element in a wide initial orbit, we can safely assume $E_i = 0$, while upon reaching the inner rim of the accretion disk in an (approximately) circular orbit, its energy is given by $E_{\text{rim}} = -GMm/2r_{\text{rim}}$. We can then evaluate the luminosity of the disk as

$$L = -\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{G\dot{m}M}{2r_{\text{rim}}}$$

where $\dot{m} = -\mathrm{d}m/\mathrm{d}t$ it the mass accretion rate. At this point we can identify r_{rim} with the ISCO (Innermost Stable Circular Orbit) whose expression is a standard GR result[22]

$$r_{\text{rim}} = r_{\text{ISCO}} = \beta\frac{2GM}{c^2} = \beta R_s$$

so that

$$L = \frac{1}{4\beta}\dot{M}c^2 = \epsilon\dot{M}c^2 \implies \epsilon = 4\beta$$

The value of β depends on the particular specimen of BH: For Schwartzschild's BHs $\beta = 3$, while for Kerr's extremal (maximally spinning) BHs $\beta = 1$.

21 The fraction ϵ is the radiative efficiency of the accretion process and can be estimated from the energy loss of the accreted material.

22 At least for Schwartzschild BHs you can try calculate r_{ISCO} as an exercise. The reader too lazy to do it him/herself can see [5], §5 and §6.

9.7.3 MBH Growth

Our most recent understanding of MBHs involves the existence of BH *seeds* at high redshift that had become supermassive as they accreted gas, stars and possibly even merged with other compact objects.

The need to start from seeds with $M < 10^6 M_\odot$ is dictated by the fact that no physical mechanism able to monolithically form a billion stellar-mass compact object is known at present day.

From Eddington's luminosity (171) we can define Eddington's timescale

$$\tau_{\rm Edd} = \frac{Mc^2}{L_{\rm Edd}} = \frac{\sigma_{th} c}{4\pi G m_p} \approx 0.45\,{\rm Gyr} \tag{172}$$

We can then write down the evolution equation for the mass of the MBH as

$$\dot{M} = (1-\epsilon)\dot{M}_{acc} = \frac{1-\epsilon}{\epsilon}\frac{f_{\rm Edd}}{\tau_{\rm Edd}} M$$

where $f_{\rm Edd} = L/L_{\rm Edd}$ is the fraction of the Eddington luminosity being radiated. The last equation can be integrated right away

$$M(t) = M_0 \exp\left(\frac{1-\epsilon}{\epsilon}\frac{f_{\rm Edd}}{\tau_{\rm Edd}}(t-t_0)\right)$$

where M_0 is the initial mass of the MBH at time t_0. In the limit $f_{\rm Edd} \to 1$,

$$M(t) = M_0 \exp\left(\frac{1-\epsilon}{\epsilon}\frac{t-t_0}{\tau_{\rm Edd}}\right)$$

we can now see what is the fastest rate at which a MBH can grow. This evidently depends on the efficiency parameter ϵ, which, we've seen, is closely related to the spin parameter of the BH β

- $\epsilon = 1/12$ (S) $\to M \propto M_0 \exp(\frac{t}{3\cdot 10^7\,{\rm yr}})$
- $\epsilon = 1/4$ (K) $\to M \propto M_0 \exp(\frac{t}{3\cdot 10^8\,{\rm yr}})$

Therefore, in both scenarios, continuous Eddington-limited accretion allows 50-to-500 e-folds in mass, depending on the BH's spin: This is more than enough to grow a $10^9 M_\odot$ BH starting from any reasonable M_0 in less than a Hubble time.

Seeding Mechanisms

We've shown how, given the existence of a seed, we're able to accrete a BH up to a MBH, but we've never discussed about what actually are these seeds we're talking about.

In a very Hegelian way[23], seeds mainly comes in three "flavors":

- *PopIII Remnant*: This scenario relies on the fact that the first generation of essentially metal free stars is expected to have a very top-heavy mass function. As we've seen, the absence of metal disfavors fragmentation and massive stars can form. The natural relics of such massive stars are BHs of several hundred solar masses. The viability of popIII remnant as seeds of the most massive MBHs has been recently questioned;

23 Or PokéMon-ian kind of way, if you're more inclined towards a modern time comparison.

- *Runaway PopIII Mergers*: If clusters of massive stars are a common occurrence at high redshifts, runaway mergers might still result in the formation of a metal free star with mass high enough to leave behind a $10^3 M_\odot$ BH remnant;

- *Direct Collapse*: The key idea is that in the most massive protogalactic halos at $z \approx 15$, gas accreted from the cosmic web can be supplied to the very centre at a rate of $\approx 1 M_\odot\,\mathrm{yr}^{-1}$. Such extreme conditions can prompt the formation of a seed BH with mass in the range $10^4 - 10^5 M_\odot$ either via collective infall from a marginally stable massive disk or via the formation of a quasi star or via direct collapse.

One challenge to this simple picture is that billion-solar-masses MBHs are not observed only in the local Universe (i.e., low redshift), but also at high redshift, which issues serious questions since they would have had really short times to accrete to such huge masses (assuming the model we've described is actually correct).

This is a problem, because prolonged accretion inevitably results in highly spinning MBHs after about an e-fold in mass growth. It seems therefore unlikely that the highest redshift quasars grew by Eddington limited, prolonged accretion, independently on the seeding model.

Now, this problem is somewhat mitigated with an argument the interested reader can find in [6] (pag. 59).

9.7.4 Massive BH Binaries

We start this new subsection with a question: Can Supermassive BHs really merge?

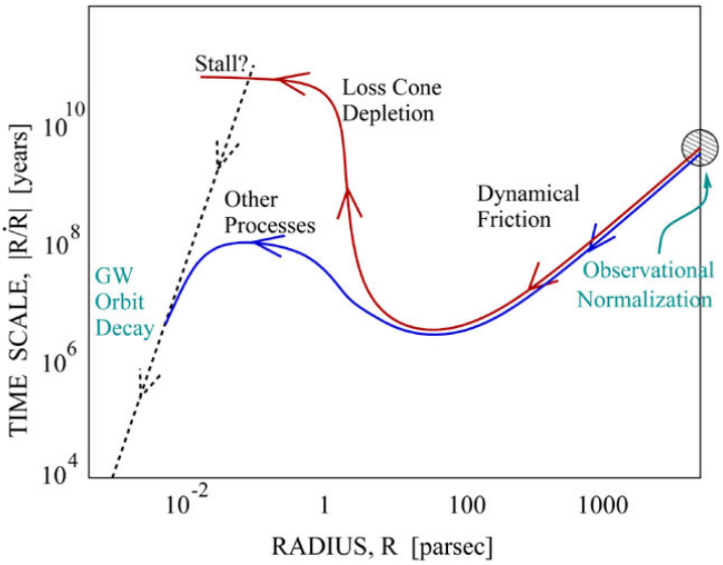

Figure 45: Different phases of the evolution of a MBH binary after a galaxy merger. Credits: [6]

Turns out, they probably can. Despite growing most of their mass through accretion along the cosmic history, in the hierarchical structure formation

framework MBHs will also acquire some mass because of mergers with other MBHs. Galaxies are in fact observed to merge quite regularly, with massive galaxies experiencing at least a major merger. If each of the progenitor galaxies host a MBH, the outcome of a galaxy merger will be the formation of a MBHBs that eventually coalesces due to GW emission.

The evolution of the system is pictorially represented in Fig.45.

We already had a look at some of the processes that can reduce the distance under the pc separation and at GW emission, that lead to swift coalescence at sub-parsec sepratation. What we're missing is a process able to bring two BHs from kpc to pc separation. An effective process to do so is *dynamical friction*, and we're going to take a look at it in the next (and last) subsection.

9.7.5 Dynamical Friction

Consider a lonely massive object of mass M and velocity V wandering in a sea of particles of mass m and velocity v[24]. The idea is having these small particles interacting with the massive intruder through a collective force that goes under the name of *dynamical friction*.

A first detailed calculation was carried out by Chandrasekhar

$$\frac{\mathrm{d}\vec{V}}{\mathrm{d}t} = -16\pi^2 \ln(\Lambda) G^2 m(M+m) \frac{\vec{V}}{V^3} \int_0^V f(v) v^2 \,\mathrm{d}v \tag{173}$$

where $f(v)$ is the velocity distribution function of the light particles and $\ln \Lambda$ is the Coulomb logarithm $\Lambda = b_{\max}/b_{\min}$.

In the limit $V \to 0$, we can approximate $f(v)$ roughly as constant, and the expression in (173) evaluates to

$$\frac{\mathrm{d}\vec{V}}{\mathrm{d}t} = -\frac{16\pi^2}{3} \ln(\Lambda) G^2 m(M+m) f_0 \vec{V}$$

which has the form typical of a viscous friction $\mathrm{d}V/\mathrm{d}t \propto -V$.

Conversely, in the limit $V \gg \bar{v}$, with $\bar{v}$ the typical velocity of the distribution, the integral is performed over the whole distribution, thus returning the number density $n/4\pi$. Thus

$$\frac{\mathrm{d}\vec{V}}{\mathrm{d}t} = -4\pi \ln(\Lambda) G^2 M \rho \frac{\vec{V}}{V^3}$$

where we've made use of $\rho = nm$ and $M \gg m$. At high velocities we see that $\mathrm{d}V/\mathrm{d}t \propto -V^{-2}$ and DF quickly becomes ineffective.

Note that, due to the proportionality with M, the force $F \propto M^2$; this is usually interpreted as the pull of a wake formed by DF behind the massive object.

Given these two limits, it's evident that DF is most efficient when $V \approx \bar{v}$.

Let us now consider the ideal situation of an isothermal sphere of particles. If we assume the massive object in circular orbit, we can calculate

24 The velocity of the massive object is to be considered with respect to the Center of Mass of the sea of particles.

the frictional force by explicitly solving Chandrasekhar's equation (173)[25], finding

$$F \approx -0.428 \ln(\Lambda) \frac{GM^2}{r^2}$$

The angular momentum of the circular object is simply $J = Vr$, so that

$$\frac{\mathrm{d}J}{\mathrm{d}t} = \frac{\mathrm{d}V}{\mathrm{d}t} r + V \frac{\mathrm{d}r}{\mathrm{d}t}$$

Now, we're going to perform what may look like a magic trick. Since, instantaneously, DF does not act on r and only changes the velocity of the massive object, we can write $\mathrm{d}J/\mathrm{d}t = (\mathrm{d}V/\mathrm{d}t)r = (F/M)r$. On the other hand, the massive object is in circular orbit in an isothermal sphere. Under the approximation that the orbit remains circular, V cannot change so that the actual result of the interaction would be to move the object onto a tighter orbit, thus shrinking r. In practice, DF does not change the kinetic energy of M, but it eventually extracts its potential energy, so that we can write

$$\frac{\mathrm{d}J}{\mathrm{d}t} = \frac{F}{M} r = V \frac{\mathrm{d}r}{\mathrm{d}t}$$

We can plug in the expression for the force we've found earlier

$$V \frac{\mathrm{d}r}{\mathrm{d}t} = -0.428 \ln(\Lambda) \frac{GM^2}{r^2}$$

that we can integrate by parts to get

$$\tau_f = \frac{1.17}{\ln \Lambda} \frac{r_i^2 v_c}{GM} = \frac{19}{\ln \Lambda} \,\mathrm{Gyr} \left(\frac{r_i}{5\,\mathrm{kpc}} \right)^2 \frac{\sigma}{200\,\mathrm{km\,s^{-1}}} \frac{10^8 M_\odot}{M} \tag{174}$$

where r_i is an integration constant we can identify with some kind of initial distance. For typical $\ln \Lambda \approx 10 - 15$, MBHs can inspiral in the centre of the stellar remnant from a 10 kpc initial distance in less than a Gyr (which is less than a Hubble time!).

This is the time it takes to bring a single MBH to the centre of an isothermal distribution of stars. It can be applied to the two MBHs inspiralling in the aftermath of a galaxy merger so long as they evolve independently of each other as individual objects interacting with the stellar distribution. This is no longer true when the two MBHs start to "see each other."

For an isothermal sphere this amounts to a mutual separation of

$$a \approx 30\,\mathrm{pc} \left(\frac{M}{10^8 M_\odot} \right)^{1/2}$$

Once the two BHs overlap with the cloud's center of mass, DF cannot bring them any closer. To do so, we need Stellar Hardening and all the other nice processes we've discussed earlier.

The End?

25 We're assuming a Maxwellian distribution $f(v)$.

ABOUT ME

Hi there, I'm Giacomo, sometimes known as Cael by a very small niche of people, the author of the notes you just finished reading.

I'm a Batchelor-graduate student at the University of Pisa now in the process of undertaking a major (Master degree) in Astronomy and Astrophysics at the same university. In my (meager) free time I'm an avid reader, gamer, anime-enthusiast as well as an independent author in the very process of publishing his first novel.

For those of you that may be interested I'll leave both the link to the novel's page as well to my instagram profile in the QRcodes right below.

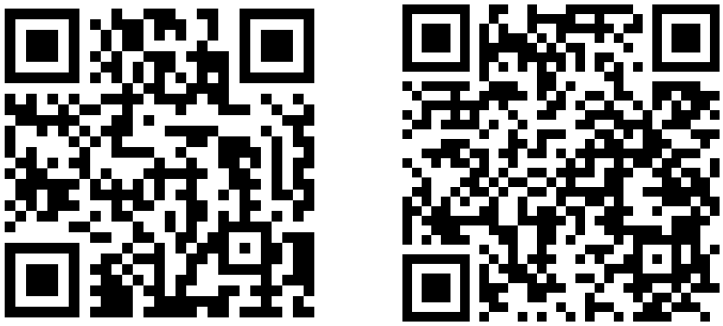

Figure 46: On the left is the link to my instagram profile, on the right the one to my novel's page (Python really has a library for anything, huh?).

Come say hi if you fancy action-fantasy stories with magic, swords, some comedy to garnish and a lot of other stuff :)

At last but not the least, I want to thank all my collegues and fellow students that helped (and hopefully will help me again) during the drafting of this text, pointing out errors and occasions for further expansion.

If you'd like your names to be included and take your fair share of credits, feel free to contact me and I'll add your name right away in the few lines that are left of this page.

I hope my understading of the complex interplayings going on in our Universe has managed to facilitate your own understading.

If not, well... oopsie.

See you, space cowboys...

Giacomo

BIBLIOGRAPHY

[1] James M. Bardeen and Jacobus A. Petterson. The Lense-Thirring Effect and Accretion Disks around Kerr Black Holes. *Astrophysical Journal Letters*, 195:L65, January 1975.

[2] G. K. Batchelor. *An Introduction to Fluid Dynamics*. Cambridge Univeristy Press, 1967.

[3] R. D. Blandford and R. L. Znajek. Electromagnetic extraction of energy from Kerr black holes. *Monthly Notices of the Royal Astronomical Society*, 179:433–456, May 1977.

[4] H. Bondi and F. Hoyle. On the mechanism of accretion by stars. *Monthly Notices of the Royal Astronomical Society*, 104:273, January 1944.

[5] Sean M. Carroll. *Spacetime and Geometry: An Introduction to General Relativity*. Cambridge University Press, 2019.

[6] Marco Celoria, Roberto Oliveri, Alberto Sesana, and Michela Mapelli. Lecture notes on black hole binary astrophysics, 2018.

[7] A. R. Choudhuri. *Astrophysics for Physicists*. Astrophysics-Textbooks. Cambridge University Press, 2010.

[8] Monica Colpi and Alberto Sesana. *Gravitational Wave Sources in the Era of Multi-Band Gravitational Wave Astronomy*, pages 43–140. WORLD SCIENTIFIC, February 2017.

[9] Thibault Damour and J. H. Taylor. On the Orbital Period Change of the Binary Pulsar PSR 1913+16. *Astrophysical Journal*, 366:501, January 1991.

[10] Juhan Frank, Andrew King, and Derek Raine. *Accretion Power in Astrophysics*. Cambridge University Press, 3 edition, 2002.

[11] D. L. Goodstein. *States of Matter*. Dover Books on Physics. Dover Publications, 2014.

[12] R. W. Hellings and G. S. Downs. Upper limits on the isotropic gravitational radiation background from pulsar timing analysis. *Astrophysical Journal Letters*, 265:L39–L42, February 1983.

[13] F. Hoyle and R. A. Lyttleton. The evolution of the stars. *Proceedings of the Cambridge Philosophical Society*, 35(4):592, January 1939.

[14] Jackson, John David. *Classical electrodynamics*. Wiley, 3 edition, 2021.

[15] A. N. Kolmogorov. The Local Structure of Turbulence in Incompressible Viscous Fluid for Very Large Reynolds Numbers. *Proceedings of the Royal Society of London Series A*, 434(1890):9–13, July 1991.

[16] Donald E. Osterbrock and Gary J. Ferland. *Astrophysics of gaseous nebulae and active galactic nuclei*. 2006.

[17] E. N. Parker. Dynamics of the Interplanetary Gas and Magnetic Fields. *Astrophysical Journal*, 128:664, November 1958.

[18] Stephen B. Pope. *Turbulent Flows*. Cambridge University Press, 2000.

[19] Giulia Ricciardi. *Notes of Astrophysical Processes*. 2021-2022.

[20] George B. Rybicki and Alan P. Lightman. *Radiative Processes in Astrophysics*. 1986.

[21] J. Schwarzkopf, Daniel Livescu, Jon Baltzer, R. Gore, and J. Ristorcelli. A two-length scale turbulence model for single-phase multi-fluid mixing. *Flow, Turbulence and Combustion*, 96, 09 2015.

[22] N. I. Shakura and R. A. Sunyaev. Black holes in binary systems. Observational appearance. *Astronomy and Astrophysics*, 24:337–355, January 1973.

[23] I. S. Shklovskii. Nature of X-Ray Sources of the Scorpius X-1 Type. *Soviet Astronomy*, 15:886, June 1972.

[24] C. Megan Urry and Paolo Padovani. Unified schemes for radio-loud active galactic nuclei. *Publications of the Astronomical Society of the Pacific*, 107:803, September 1995.

[25] H. C. van de Hulst. The Chromosphere and the Corona. In Gerard Peter Kuiper, editor, *The Sun*, page 207. 1953.