Gravitation and Cosmology (GC8)
- Compact Object Formation and Evolution -

Michela Mapelli

2020-2021

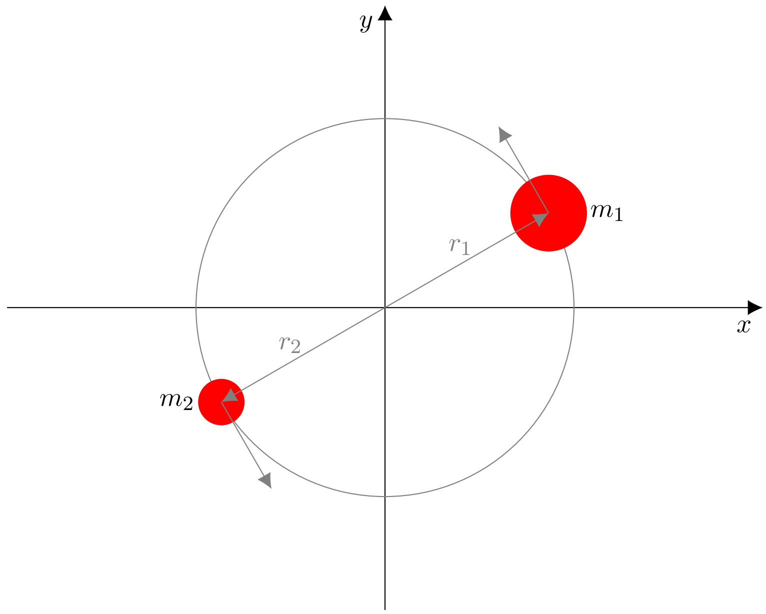
integration notes to presentations

written by Ulyana Dupletsa

Contents

I. Gravitational Wave Astrophysics

i. GWs emitted by a binary system in circular orbit



Here we are considering a circular orbit, although in general it is an ellipse. So we can derive the semi-major axis a as:

$$(r_1 + r_2) = 2a \tag{1}$$

In this reference system also the center of mass coincides with the origin, so that the following relations hold:

$$m_1 r_1 = m_2 r_2 \implies \begin{cases} r_1 = \frac{m_2}{M} 2a \\ r_2 = \frac{m_1}{M} 2a \end{cases} \tag{2}$$

where $M = m_1 + m_2$ is the total mass of the binary.
The orbital frequency can be found using Kepler's third law:

$$\omega^2 = \frac{GM}{(2a)^3} \tag{3}$$

In order to compute the quadrupolar mass moment

$$I^{ij} = \int d\vec{x}^3 \rho(t, \vec{x}) x^i x^j \tag{4}$$

of our system, we need to describe its density:

$$\rho(t, \vec{x}) = \sum_{i=1}^{2} m_i \delta(x - x_i)\delta(y - y_i)\delta(z) \tag{5}$$

since the masses are in the xy plane.
So we need to describe the coordinates of our masses:

$$\begin{cases} x_1 = r_1 \cos(\omega t) \\ y_1 = r_1 \sin(\omega t) \\ x_2 = r_2 \cos(\omega t) \\ y_2 = r_2 \sin(\omega t) \end{cases} \implies \begin{cases} x_1 = 2\frac{m_2}{M} a \cos(\omega t) \\ y_1 = 2\frac{m_2}{M} a \sin(\omega t) \\ x_2 = 2\frac{m_1}{M} a \cos(\omega t) \\ y_2 = 2\frac{m_1}{M} a \sin(\omega t) \end{cases} \tag{6}$$

Gravitational wave astrophysics is the branch of astrophysics that deals with the characterization of gravitational wave sources. Since LIGO and Virgo just started observing mergers, this branch of astrophysics is quite young and fast evolving. It is mostly about binary compact objects: BBHs and BNSs and NSBHs.
The main open questions regard the formation channels of the binary compact objects observed by LIGO/Virgo.

II. GW theory

Gravitational waves can be derived as a solution to Einstein field equations, either numerically (fully non linear way) or analytically after linearizing them. This last case, under some assumptions leads to a wave equation:

$$\Box \bar{h}_{\alpha\beta} = \frac{16\pi G}{c^4} T_{\alpha\beta} \tag{7}$$

A rather general solution of this equation for slowly moving sources can be expressed as a function of time and space:

$$\bar{h}^{ij}(t,\vec{x}) \sim \frac{2}{r}\frac{G}{c^4}\frac{d^2}{dt^2} I^{ij}\left(t - \frac{r}{c}\right) \tag{8}$$

r is the distance between the source and the observer and I^{ij} is the second mass moment:

$$I^{ij} = \int d^3x \rho(t,\vec{x}) x^i x^j \tag{9}$$

If there is no second time variation of the quadrupole moment then there are no GWs.

We can make the solution explicit for the case of a binary system. Suppose we have a binary star in the xy plane with orbital angular momentum along z, and $2a$ the orbital separation with a the semi-major axis. During the motion the binary describes the angle ωt. We can assume that both members have mass M and also assume zero eccentricity.
Under these assumptions and in the reduced particle system, we can write the second mass moment in a simple way:

$$\begin{aligned} I^{xx} &= 2Ma^2\cos^2(\omega t) = Ma^2[1+\cos(2\omega t)] \\ I^{xy} &= 2Ma^2\sin(\omega t)\cos(\omega t) = Ma^2\sin(2\omega t) \\ I^{yy} &= 2Ma^2\sin^2(\omega t) = Ma^2[1-\cos(2\omega t)] \end{aligned} \tag{10}$$

By taking the second time derivative we obtain:

$$\bar{h}^{ij} \sim -\frac{2}{r}\frac{G}{c^4}(2\omega)^2 Ma^2 \begin{pmatrix} \cos(2\omega(t-r/c)) & \sin(2\omega(t-r/c)) & 0 \\ \sin(2\omega(t-r/c)) & -\cos(2\omega(t-r/c)) & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{11}$$

Using Kepler's third law $\omega^2 = \frac{G2M}{a^3}$, we can write:

$$\bar{h}^{ij} \sim -\frac{16G^2}{c^4}\frac{M^2}{ra} \begin{pmatrix} \cos(2\omega(t-r/c)) & \sin(2\omega(t-r/c)) & 0 \\ \sin(2\omega(t-r/c)) & -\cos(2\omega(t-r/c)) & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{12}$$

This expression encodes everything we will need. We see that the angular frequency of GWs is twice the orbital frequency:

$$\omega_{GW} = 2\omega_{orb} \tag{13}$$

We can also see that GWs have two polarizations:

$$\begin{aligned} h_+ &\sim \frac{16G^2}{c^4}\frac{M^2}{ra}\cos(2\omega(t-r/c)) \\ h_\times &\sim \frac{16G^2}{c^4}\frac{M^2}{ra}\sin(2\omega(t-r/c)) \end{aligned} \tag{14}$$

We can also derive the amplitude of GWs:

$$h = \frac{1}{2}\left(h_+^2 + h_\times^2\right)^{1/2} \sim \frac{8G^2}{c^4}\frac{M^2}{ra} \sim 10^{-21}\left(\frac{M}{M_\odot}\right)^2\left(\frac{1\text{kpc}}{r}\right)\left(\frac{1R_\odot}{a}\right) \tag{15}$$

We immediately see that:

- The bigger the amplitude the easier the detection
- The farther the binary, the smaller the amplitude
- The larger the masses, the larger the amplitude
- The smaller the semi-major axis, the larger the amplitude

This implies some important considerations about the evolution of the binary. The binding energy of a binary system is $E_{orb} = -\frac{Gm_1m_2}{2a}$. The emission of GWs comes from the orbital energy of the binary itself. So the system loses binary energy, therefore it shrinks. a becomes smaller and smaller the more efficiently the binary emits GWs. This also means that the frequency of GWs becomes higher and higher.
A very useful back of the envelope calculation, is the derivation of a timescale for the coalescence by GW emission, which is the timescale over which a binary system will merge just by emission of GWs, no other processes.

$$P_{GW} = \frac{32}{5}\frac{G^4}{c^5}\frac{1}{a^5}m_1^2m_2^2(m_1+m_2) = \frac{dE_{orb}}{dt} = \frac{Gm_1m_2}{2a^2}\frac{da}{dt} \tag{16}$$

which implies that:

$$\frac{da}{dt} = \frac{64}{5}\frac{G^3}{c^5}\frac{1}{a^3}m_1m_2(m_1+m_2) \tag{17}$$

We are assuming that the masses are not changing yet, we are far from the merger.
From this we can find a timescale for GW emission (replacing general eccentricity):

$$t_{GW} = \frac{5}{256}\frac{c^5}{G^3}\frac{a^4(1-e^2)^{7/2}}{m_1m_2(m_1+m_2)} \tag{18}$$

If we put some numbers in the expression of this timescale, we see that for most of the systems this is a really long timescale. If we calculate it for two neutron stars: $m_1 = m_2 = M_\odot$, and $a = 1$A.U. and $e = 0$, we obtain:

$$t_{GW} \sim 2\times10^{17}\text{yr} \gg t_{Hubble} \sim 13\times10^9\text{yr} \tag{19}$$

These equations are not true for all the stages of the life of a binary. They are correct during the so called **inspiral**. A binary system evolving through GW emission is characterized by 3 stages: inspiral (a stage where the system is still a binary, which can be considered pretty Newtonian), merger and ringdown (the resultant compact object is re-adjusting to its final equilibrium). The three stages correspond to very different modeling: the inspiral is almost Newtonian (post-Newtonian techniques), the merger is described by Numerical Relativity and the ringdown can be described with some pertubative approach.

We can consider the inspiral to be well approximated by Newtonian physics till the last stable circular orbit (LSCO):

$$r_{LSCO} = 3r_{SW} = 3\frac{2G(m_1+m_2)}{c^2} \tag{20}$$

and we can also estimate the frequency at the LSCO as twice the orbital frequency at LSCO:

$$\omega_{GW,LSCO} = 2\sqrt{\frac{G(m_1+m_2)}{r_{LSCO}^3}} \sim 460\text{Hz}\frac{60M_\odot}{m_1+m_2} \tag{21}$$

We will focus on compact binaries which emit GWs from 10^{-4}Hz up to hundreds and thousands of Hz. Above 10Hz we have the current ground-based GW detectors.

III. Gravitational facts

LIGO, Virgo and KAGRA have been observing the sky for GWs. We know 50 GW candidates from the observations so far.
What can we actually observe? What physical properties of compact binaries can we get from the observation of GWs? There are 15 observable (plus eccentricity, which is not included in the LIGO/Virgo analysis). The most important are the 2 masses, 6 spin components, redshift of the merger, the 2 positions on the sky (RA, DEC).

iii.1 Masses and spins

BHs are very simple in principle, uniquely defined by their mass and spin. LIGO and Virgo do not observe m_1 and m_2 straightforwardly, but they observe combinations of them:

- **Chirp mass**: easily derived because the evolution of frequency scale with a power of this chirp mass:
$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1+m_2)^{1/5}} \tag{22}$$
- **Total mass** that sets the frequency during inspiral. We have seen that the frequency at the LSCO depends on M.
- **Mass ratio**:
$$q = \frac{m_2}{m_1} \tag{23}$$
with $m_2 < m_1$. It affects the phase of GWs

In the mass graph we have blue points which are pulsars for which we know the mass, because they are in binary systems, or with other neutrons stars or just stars.
The green dots are BHs for which we know the mass from EM spectrum. They are BHs in X-ray binaries for which we have a dynamical measurement of mass. This means that we do not only know from X rays that there is a BH there, but we have also identified the companion, form optical or infrared radiation. We can estimate the properties of the luminous companion, like its mass. We have measured from spectra the line of sight velocity of the binary. So we know the velocity, we know the mass of the companion and we can reconstruct the mass of the dark object, using Keplerian equations.
Finally the orange dots (47) are the BBHs from GWs. While almost all of the green dots are below $20M_\odot$, a large fraction of the orange dots are above $20M_\odot$. We need to understand what astrophysical processes can produce BHs of these sizes, assuming that they come from massive stars.

The spin is defined by the vector $\vec{S}$, which is the spin angular momentum normalized to the quantity below:
$$\vec{S} = \frac{\vec{J}c}{Gm^2} \tag{24}$$
For a Schwarzschild BH $S = 0$, and it is nearly 1 for a maximally rotating BH.
LIGO and Virgo usually give us information about the spins via two combinations:

1. **Effective spin**: it is the mass weighted sum of the components of the spins of the two objects along the orbital angular momentum of the binary:
$$\chi_{eff} = \frac{(m_1\vec{S}_1 + m_2\vec{S}_2)}{m_1+m_2} \cdot \vec{L} \tag{25}$$
According to this definition it is a scalar quantity, with $-1 < \chi_{eff} < 1$. $\chi_{eff} = +1$ means the two objects are maximally spinning and are aligned with the orbital angular momentum. $\chi_{eff} = -1$ means that both objects are maximally spinning but anti-aligned with the orbital angular momentum. $\chi_{eff} = 0$ either means zero spins or spins that are in the plane of the binary system.
It affects mostly the phase of GWs.

2. **Effective precession spin**: it's the combination of other components of the spins. It considers the components in the plane of the binary:
$$\chi_p = \frac{1}{B_1 m_1^2} \max\left(B_1 S_{1,\perp}, B_2 S_{2,\perp}\right) \tag{26}$$
It is a good tracer of one of the two spins of the binary, generally the spin of the most massive or the one that has the largest component in the plane.

Almost all of the events are consistent with $\chi_{eff} \sim 0$ within 90% of confidence interval a part from two. We can interpret it in two different ways: either all spins are small and we get $\chi_{eff} \sim 0$, or it is an indicator of an excess of spins lying in the plane or a combination of small spins lying in the plane.
In the χ_p plot we have almost no constraint, from the first two runs. Some events of O3-A report some good measurements. The white part is the prior, the colored one is the posterior. If they are the same this means that we have no constraints. We see a tendency for large χ_p for GW190521.

LIGO and Virgo have an alternative approach to spins. The violin plots measure and constrain the single event. What if we analyze the all population? If we adopt priors for the properties for the binary compact objects that are informed by an entire population of BBHs or BNSs? This is a population study. We can see what we can get for spins for a population perspective: we can get stronger constraints.
The orange line is the mean distribution of the posterior: we have some indications of χ_p clustering around $0.1 - 0.2$ with a long tail, for larger χ_p.
The same exercise for χ_{eff} produces a prior distribution which seems to be squeezed to lower values, but still a little bit off with respect to zero, pointing towards larger than zero χ_{eff} for the bulk of the population.
We can then obtain even some constraints on χ, which is the actual spin dimensionless spin magnitude (S previously) of the primary component of the binary. The posterior distribution obtained for the whole population shows a mean (black line) which shows the best χ values around $0 - 0.4$ and a long tail at larger χ. After O3-A we started inferring a distributions for the spin distribution under some assumptions for the populations.

iv. Merger rates

A crucial information that can be inferred from LIGO and Virgo is the rate of BNSs and BBHs and NSBHs.

II. Massive Star Evolution

i. Eddington luminosity

For any object there is a maximum luminosity beyond which radiation pressure will overcome gravity, and material outside the object will be forced away from it rather than falling inwards.
We want to find the luminosity at which the gravitational force inwards balances with the radiation force outwards. The gravitational force is given by:

$$F_{\text{grav}} = \frac{GMm}{R^2} \tag{27}$$

where M is the mass of the radiating object and m is the mass of the radiated object at distance R. The radiation pressure at distance R is given by:

$$P_{\text{rad}} = \frac{L}{c}\frac{1}{4\pi R^2} \tag{28}$$

To calculate the radiation force on the object we need to know its opacity κ. Radiation pressure is force per unit area, **opacity is the cross-sectional area per unit mass for radiation scattering**:

$$F_{\text{rad}} = P_{\text{rad}}\kappa m \tag{29}$$

Balancing the two forces gives the Eddington luminosity:

$$\frac{GMm}{R^2} = \frac{L\kappa m}{4\pi c R^2} \implies L_{\text{edd}} = \frac{4\pi cGM}{\kappa} \tag{30}$$

The Eddington luminosity depends only on the mass of the radiating object and we have assumed spherical symmetry. When the accreting material is mostly ionized hydrogen, then the opacity is given by Thompson scattering. The cross-section comes mostly exclusively from radiation pressure on electrons, but the mass lies mostly exclusively in protons. There are electrostatic forces between electrons and protons: if we exert a radiation pressure that is mostly felt by electrons, they will drag the protons along with them. Under this approximation:

$$\kappa = \frac{\sigma_T}{m_p} \tag{31}$$

with $\sigma_T = 6.65 \times 10^{-25}\text{cm}^2$ and $m_p = 1.67 \times 10^{-24}\text{g}$.
The Eddington luminosity is important to describe stellar winds.

ii. Effective temperature

The emission of a star is always described by a black body. The Stefan-Boltzmann law relates the rate of radiation per unit area to the effective temperature:

$$j = \sigma T_{\text{eff}}^4 \tag{32}$$

On the other hand, by definition:

$$j = \frac{L}{4\pi R^2} \tag{33}$$

Therefore, we can directly relate the luminosity of a star to its effective temperature:

$$L = 4\pi R^2 \sigma T_{rmeff}^4 \tag{34}$$

This is what is represented in the **HR diagram**. For the Sun:

$$L_\odot = 3.83 \times 10^{26}\text{W} \tag{35}$$

Massive stellar evolution is a total mess!

III. Stellar evolution

Before ZAMS a star is not a star, it is a protostar. It's just a clump of accreting gas which is contracting trying to reach the central temperature to start burning hydrogen. Therefore, the ZAMS mass of a star is the initial mass of the star.
In the Hertzsprung gap the luminosity is almost constant and the temperature is changing fast. It's almost out of equilibrium: the star is not burning hydrogen in the core anymore and it has not started yet to burn hydrogen in a shell.
When the star starts to burn hydrogen in a shell, it passes to the red giant branch. It is a branch because we observe a steep branch in the luminosity-temperature diagram.
At some point, the star, which now has a helium core, can start burning helium in its core and because a core He burning star in the HR diagram goes to the blue, that is why it is called blue loop. Although it is better the physical definition of helium burning star.
When the star is relatively low mass ($\sim 1 - 8M_\odot$), after He burning it undergoes the asymptotic giant branch, which is a scaled-up luminosity version of the red giant branch. An asymptotic giant branch (AGB) star has a carbon-oxygen core which is not yet burning. It burns He in a shell and H in a more external shell.

We consider now massive stars (more than $20M_\odot$ for ZAMS mass).
The HD limit is a diagonal line which becomes horizontal at low temperature and it is an empirically derived region in the HR diagram. It is derived from observations, NOT theory. It is based on the fact that there are very few stars in the Milky Way and nearby galaxies that lie in that region (zone of avoidance). The simplest way to interpret the HD limit is that this region is where the radiation pressure cannot be balanced by gravity. It's the region where the star would emit above its Eddington limit. This region though is not completely empty: the $50M_\odot$ enters it as well as the $250M_\odot$. What are these stars when they lie in this region? We call them the **luminous blue variable stars** because they are super luminous and they are reasonably blue (they have a temperature between $10000 - 50000$K, so they still have some H envelope). When these LBV stars evolve, they go almost immediately to the region of high effective temperatures. They almost immediately transform into a Wolf-Rayet star. The physical process that pushes them to this region is the stellar wind.
The WR stars have an extremely high effective temperature because they are naked He cores, no H around.

We have seen that effective temperature connects the radius and the luminosity. We observe that at fixed luminosity a high effective temperature means a smaller radius.
Depending on their radii a star binary may evolve in very different ways according to the fact that the stars touch or not. For a red giant star is very likely to get in contact with another star in a binary system, than for a WR star. Less massive stars will die in the red giant phase, whereas more massive ones will go to the left towards the WR phase, to the high temperature.

IV. Stellar winds

Stellar winds are very effective in more massive stars and they drive the transformation of an O-type star in a WR star.
The mass of a star changes a lot during its life.

Some photons reach the photosphere of the star and there they can interact with the ions and the interaction can produce a transfer of linear momentum from photons to ions. Basically these interactions are mostly due to resonant metal lines and they are mainly connected to iron lines. So the loss of these ions will depend on metallicity.
At the stellar surface there can be some local violation of two fundamental equations of stellar equilibrium: mass conservation and the law of hydrostatic equilibrium.

dm and dr depend on time, the equation of mass conservation becomes an equation of mass loss:

$$\dot{m} = 4\pi\rho r^2 v \tag{36}$$

In the same way we can rewrite the hydrostatic equilibrium equation in a way that it is no longer in equilibrium: it becomes an equation of motion:

$$\frac{dv(r)}{dt} = -\frac{1}{\rho}\frac{dP}{dr} - \frac{Gm(r)}{r^2} \tag{37}$$

What kind of pressure acts in a star? Inside a star we have two sources of pressure: the pressure of gas and the radiation pressure. At the stellar surface the gas pressure is basically negligible because we do not have gas on top. So we will have only the radiation pressure in the equation, given by the photons which are coming from the inside of the star.

$$\frac{1}{2}\frac{dv^2}{dr} = -\frac{1}{\rho}\frac{dP_{\rm rad}}{dr} - \frac{Gm(r)}{r^2} \tag{38}$$

If the LSH term is positive, it means that we have a net movement from the stellar surface to the outside of the star. If we do not neglect the gas pressure at the surface of the star, the line driven wind equation is:

$$\frac{1}{2}\left(1 - \frac{v_s^2}{v^2}\right)\frac{dv^2}{dr} = -g_{\rm eff} \tag{39}$$

What are the physical terms that give a sufficiently large $g_{\rm rad}$ to have a positive outflow? For massive hot stars (progenitors of BHs), we have two terms: g_{ES} and g_{LINES}. These terms can counteract and win gravity only because at the stellar surface we have an anisotropy between the radiation field which is mostly directed outward, it is not balanced.
Electron scattering is Thomson scattering: it is dominant when the luminosity of the star is close to its Eddington limit. This term does not depend on the metallicity (even if the luminosity can slightly depend on it). Therefore, even a metal poor star will have stellar winds powered just by Thomson scattering.

On the contrary the other line term is powered by resonant metal lines. In this case, the ions at the surface of the star acquire momentum from the photons because they absorb photons through the resonant lines. In this case we do have a dependence on metallicity: we have either one line that dominates the mass loss (just one metal species) or we have the sum of multiple lines. We have that g_{LINES} scales with the effective number of metal lines. This in turn scales with metallicity with a given index $1 - \alpha$.
These two terms can be placed in the wind equation and we can get the wind velocity if there is any.
About the metallicity dependence of g_{LINES} there is an important thing to say. The early models assume that each photon can interact with a single ion only a single time and then the photon is either totally absorbed or it flies away. Under these assumptions, the most important lines are the C IV, Si IV, N IV. The predictions for mass loss $\dot{m}$ (we have the velocity and then we get $\dot{m}$, i.e. how much mass is lost by the star at a given time as an effect of stellar winds):

$$\dot{m} \propto Z^{0.5} \tag{40}$$

The higher the metallicity the higher the mass loss rate.
The new models (Vink et al.) change the approach to stellar winds. They use Monte Carlo approach that allow for multiple scatterings for each photon. In this analysis the Fe lines are quite always the most important lines. In this case, the prediction for the mass loss becomes a little bit different:

$$\dot{m} \propto Z^{0.85} \tag{41}$$

This is a much steeper trend with metallicity. Vink et al. also give useful fitting formulae to estimate the mass loss rate in different cases. Vink et al. worked for O-type stars. What happens for stars that are out of the main sequence, like WR and LBV? The bottom line is that also for post main sequence massive hot stars the rate is

$\dot{m} \propto Z^{0.85}$. In this case the models are really approximate: they rely on the idea that mass loss is a smooth function of time. In reality, the astronomical images show that winds from massive post main sequence stars are really clumpy, they look more like ejections of matter rather than smooth winds. There is ground for further developments of these theories.

How can we take into account the effect of electron scattering? There is an important correction to do to this scaling when the star is close to be radiation pressure dominated. Then the impact of electron scattering on stellar winds becomes more and more important (models by Graefener and Hamann). They take into account both g_{LINES} and g_{ES}. Here even at low metallicity we have a higher mass loss because of the contribution of electron scattering.

The simplest model of mass loss by stellar winds that takes into account both the contribution of resonant metal lines and the contribution of electron scattering, is a model that scales with metallicity but not in a constant way. It depends on Γ, which is the ratio between the actual stellar luminosity and its Eddington limit. The closer the star to its Eddington limit, the higher the mass loss even if the metallicity is low. This is the model that is commonly used in stellar evolution models to describe the evolution of stars that will bring to the formation of black holes.

What about stars with a low effective temperature? In this case the situation is even more complicated.

A solar metallicity massive star will lose more than 2/3 of its ZAMS mass as a result of stellar winds. The metal poor stars preserve most of their mass to the very end. This is one of the main ingredients to understand the mass of the compact object.

Now we can interpret better the track of a massive star (from $20M_{\odot}$ to $350M_{\odot}$). At high metallicity the stars have an extreme evolution from right to left in the HR diagram. We have a lot of WR stars: only a naked He core, no H left! They go to the blue part.
In contrast, at low Z, almost none of these stars goes to the left at the end of their lives. They die inside the HD limit, as LBV or red giant stars.

At solar Z, the winds are really effective and remove the outermost envelopes of the star and so in the end the star is peeled off and dies as a WR. At lower Z, but not so low, the star dies when it is still in the red, when it is very large and it was able to retain most of its envelope. Because of stellar winds, a low Z star, will end its life with a very different mass wrt a high Z star.

III. Formation of Compact Objects from Single Star Evolution

The scope is to understand the plots which show the expected mass of the compact object as a function of the initial mass of the progenitor star for stars of different metallicities. Metal poor stars will leave more massive black holes, with a caveat that is connected to pair instability.

Our objective is to understand how can we produce black holes from few to tens of solar masses ($60 - 80M_{\odot}$). When the paper about the first GW event was published, there were also papers trying to explain how can we produce black holes with masses as the ones observed. There were two models trying to explain it:

- The final mass as a function of metallicity for a model with strong and one with a weak wind.
- The expected mass of the compact object vs the mass of the progenitor star for different Z. The largest component mass matches only with the lines with relatively low sub solar metallicity.

Why this kind of models can explain the formation of BHs with $20 - 30M_{\odot}$? And why the stellar metallicity needs to be so important?
The two critical physical ingredients of current BH and NS formation models are the mass evolution of the progenitor star during its life (basically stellar winds) and the final fate of progenitor star which means essentially the SN explosion or direct collapse.

i. Core-collapse SNe

When the star is sufficiently massive to form an iron core at the end of its life (generally it means that the star has an initial total mass larger than $8M_{\odot}$), the iron group atoms have a binding energy that is maximum, we cannot release any more energy by fusion of these elements. The stellar core starts collapsing because the pressure drops. The collapse goes on until the electron degeneracy pressure starts to become important. When we compress the core of the star enough the electrons becomes degenerate and we have the degeneracy pressure trying to stop the collapse. We have a critical mass, the Chandrasekhar mass: if the mass is above it, then the electron degeneracy pressure is not enough, because the pressure exerted by the outer layer is so strong that electron and proton capture removes the electrons and the electron pressure decreases. The core keeps collapsing till nuclear densities are reached. At this density the neutron degeneracy pressure settles in. The inner part of the core already collapsed to nuclear densities forms a proto-neutron star. Here the surface is supported by neutron degeneracy pressure.
We still have a part of the core which is pre-collapsed, which has not yet collapsed to the center and we have already a PNS formed at its center. When the rest of the core hits the surface of the PNS we have a bounce shock.

Up to the bounce shock the processes that happen are known form the '60s. From now on things are uncertain.
We have the bounce shock that starts expanding. We have like $30 - 40M_{\odot}$ that are still collapsing and the shock that is propagating. Is the shock able to reverse the collapse of the outer layers? If it is not we will have just a collapse without a SN. If the shock is enough to reverse the collapse then we have a core-collapse SN.

The energy of the bounce shock is mostly propagating as kinetic and thermal energy of neutrinos. The density is so high that even neutrinos are trapped and they can exchange energy. Then we have the outer part, where the collapse is still ongoing.
If just consider the bounce shock there will be a radius inside the star where the energy of the bounce shock and the energy of the infalling material (ram pressure) are almost equal. At this place the shock cannot propagate any more: shock stalling region. The initial bounce shock stalls here. The situation is complicated because the density of the star decreases from internal to external part. So at this shock stalling region we reach a density where neutrinos do not interact with the rest of the matter of the star and they simply leak out without exchanging energy anymore: the energy of the shock is lost without contributing any more to reverse the collapse of the outer regions.

Does it mean all SN are going to fail? It is not possible that all the stars just collapse because we do observe core-collapse SN. There must be something that revives the initial bounce shock and allows the shock to win over the imploding outer layers and produce the explosion.
There are some mechanisms that allow for the revival of the shock:

- Rotational power form the star
- Description based on the magnetic field

The most popular oldest model is the **convective engine**: if the neutrinos do not propagate just by diffusion but at the surface of the proto-neutron star we produce convective bubbles. If the energy is efficiently transferred not by diffusion but by convective bubbles, then the convective bubbles can be so efficient to drive the energy above the stalling region and to allow the shock to revive out of the stalling region and to reverse the collapse. If convection is effective, it can push the shock out of the shock stalling region and lead to explosion.

Let's assume the convective engines are effective (we DO observe SN), do we expect that this shock revival mechanism is efficient for all possible stars, or there will be stars able to collapse anyway?
Let's try with a simple energetic argument to see if there is a good reason why some stars should collapse anyway without SN explosion. Our supernova will have a maximum energy, E_{SN}. For the SN to succeed this maximum energy, that can be released by the SN, should be larger than the binding energy of the outer layer of the star. Naively we can write the binding energy as a function of PNS (the initial core) plus the mass of the envelope (initially infalling) divided by the radius of the envelope of the star. Making some estimates, we find that the mass of the envelope should about $50M_{\odot}$. If the star has a larger mass, than the energy would not be enough to produce an explosion. The star collapses directly to a BH!
If the star explodes as a SN or if the star collapses directly has a huge difference for the final mass of the compact object. In fact, if we have an explosion of the outer layer, we expect that what will produce the final compact object will have the mass of the PNS plus possibly some additional mass from some falling back material. We expect that if the core-collapse SN occurs, the mass of the final object will be small, either a NS or a BH.
If the SN fails, the final compact object will have a mass comparable to the mass of the star at the onset of the collapse. We expect to produce more massive BHs, possibly above $30M_{\odot}$.

II. How to study a SN explosion?

How can we go from the back of the envelope calculation to the complete physics? The only way to know if there will be an explosion or not is to have a full calculation that gives us the binding energy at each layer of the star at the onset of core-collapse and during the core-collapse. We do this with hydrodynamical simulations of core-collapse SN. We have three families of simulations: 1d, 2d and 3d. The 3d are super expensive computationally. We cannot run hundreds of models which we need to understand if a star is going to explode. Moreover, we have to simplify the physical ingredients. We really need 3d models!
All 3d models show the development and importance of asymmetries, which make it easier and easier for the SN to explode. The bottom line is that we have simulations, they contain approximations, the heaviest of which is that most of the literature is 1d model.
We need to find some simplifications based on 1d models. Here we are talking about models that try to find a relationship that connects properties like the mass of the star at the onset of collapse and the mass of the BH or NS after the collapse. There are three families:

- Mass of the C-O core. The massive star at the end of its life is like an onion with several shells. The most internal one is the iron core, which drives the collapse and will produce the PNS. The next core is the C-O one, the last important core developed by the massive star. We can infer form 1d models a correlation between the C-O core mass and the final mass of the BH. Fryer et al noted that if the mass of the C-O core is larger than $8-12M_{\odot}$ then the SN will fail. If the mass is below the SN succeeds and we will have a smaller compact object.

- Compactness. The mass of the C-O core is not enough: the correlation is not so strong, there are several uncertainties, it is not possible to describe everything with just the C-O core. To do better, in 2011, O'Connor and Ott proposed the compactness, which is defined as one reference mass over the radius R that encloses this mass inside the star at the onset of collapse. If we take as a reference mass $2.5M_{\odot}$, we look at the stellar interior and find the radius which encloses this mass at the onset of collapse and calculate the compactness ξ. How do they employ this definition?
  They calculate C at $2.5M_{\odot}$ and they see if there is a correlation between the internal compactness and the explodability. They consider different properties: they show the time for the formation of the BH as a function of compactness for different models (different EOS, different formalisms for the transfer of neutrinos). We are looking at the time for the formation of the BH, because if this time is infinite it means that the BH does not form, it means that we will have just a NS. If this time is very short after the onset of core-collapse it means that we will form a BH very efficiently. They note a correlation between compactness and the formation time: in particular, this time goes to infinity when the compactness drops below 0.25. Here is the criterion: the SN fails if the compactness is larger than 0.2, the SN succeeds and I have a small compact object if the compactness is below this value.
  This is the compactness model. This model has a problem: the compactness should be measured at the onset of collapse and only if we know all the structure. Very few stellar evolution models reach this point of the evolution. What can we do?
  Here comes the reply of C-O model. Is there any correlation between the compactness and C-O core mass, which is so nice because all the stellar evolution models know the C-O core mass? Some plots say yes, others no. Let's analyze one by Limongi: he shows the compactness at the onset of collapse as a function of the C-O core mass at the onset of collapse. Different points refer to different stellar models, whereas the colors refer to different stellar rotational velocities. There is reasonable correlation. Therefore, the C-O core mass criterium is not bad because it correlates with compactness.
  Another school say that we do not have any strong confirmation that compacness is enough. Here comes the third model of the 2 criteria.
- 2-parameter criteria. There are two fundamental quantities that determine the success or the failure of the SN:
  - $\dot{M}$, which is the rate of mass infall of the outer layers: it describes the gravity of the outer layers
  - Luminosity of neutrinos: the higher the luminosity of neutrinos, the higher the energy radiated in neutrinos, the higher the energy available to the SN and the higher the chance that the SN succeeds.

  We need two criteria, so Ertl et al. look at the simulation and find these 2 criteria: M_4, which is the enclosed mass, and μ_4 which is the mass gradient, both calculated at dimensionless entropy per nucleon equal to 4. Why this value? They checked all the quantities they had in the simulations and they found that these are good criteria. They found that the mass gradient seems a very good proxy for $\dot{M}$, while the product $M_4\mu_4$ seems to scale with the luminosity of neutrinos. They give a fit between models that explode and the ones that do not based on the usage of M_4 and μ_4.
  What is the net result of this complex model? Let's see the plot that shows the ZAMS mass of a star (at solar Z) and M_4. Each bar represents the value of M_4 for a single model. If the ZAMS mass is large than it is more likely that the SN fails. However, we have also successful SN at high ZAMS mass, and failed ones at low ZAMS mass. The probability that a SN succeedsis NOT a monotonic function of a star's mass, it is neither a monotonic function of the C-O core mass of the star. We can have models that collapse even at low masses, as well as models that explode even at high masses.
  This is why these models are referred to in literature as models that predict islands of direct collapse or SN explosion. The basic idea is that the final fate of a massive star is rather a patchwork. On the other hand, we see that for a massive star it is more common to have a direct collapse, so there is a dependence on the mass of the star.

Are these the only open questions about the final explosion of a core-collapse SN? NO, there is also one open question about the fallback: we can expect that even if there is a core-collapse SN explosion, not all the outer layers of the

star get a velocity that is larger than the escape velocity from the PNS surface. Some of the ejecta will have a lower than the escape velocity and so at some point they will fall back on to the PNS. If this fallback is efficient, it can produce remnants which are neither low mass NS nor very massive BHs, but something in between, like BHs with $3-7M_{\odot}$.
How efficient is this fallback? It depends on the explosion energy, which is ill constrained, on the angular momentum transfer during the SN, which is ill constrained, on the properties of the progenitor.

Another uncertainty is the rapidity of the explosion: how much time it takes for the mechanism that revives the SN to relaunch the actual shock? If it is rapid (less than 200ms), we have a larger amount of energy available. If it is more than 200ms than we have an energy that is a factor of 10 lower. It determines a huge difference. It determines what is called the rapid and the delayed models of explosion.

iii. Pair instability SNe

There other kinds of SN, not just core-collapse. These are the pair-instability SN. If a star is very massive, it is able to produce a He core with mass larger than $64M_{\odot}$ at the end of C burning. In this case, if we look at the plot of the central temperature as a function of the central density, if a star can build a He core that big, the central T goes above 10^7K.
So we have an efficient production of γ rays in the core. They are not very energetic ($1-10$ MeV or close), but able to produce electron-positron pairs when scattering atomic nuclei. These are essentially at rest when produced: we lose γ ray photons and the pairs we produce do not carry enough energy and pressure as the γ ray photons did (they are almost at rest).
The missing photon pressure from the γ ray photons produces an unbalance between gravity and pressure, which leads to a collapse. Here the star is not the one that has developed an iron core, it is the star at the end of C burning, it just has a C-O core inside the He core.
A collapse of the core at this stage leads to the increase of the T (if we compress and the density increases the T increases as well). The T becomes sufficiently high to ignite all the remaining atomic species. We have a lot of C, that burns to O, we have O, we have Si. All of them start igniting earlier than they should and this release of energy can lead to an explosion. Since this explosion happens when the PNS has not formed, it leaves no compact object. If a star evolves in a way that it reaches the instability region at this stage we expect a SN without a compact object. In this case and differently from core-collapse SN, the physics is under control. It is complete from early papers.

iv. Pulsational pair instability SNe

These are the younger sisters of pair instability SNe. What happens if the star is massive but not that massive? If a star develops, at the end of C burning, a He core that is between $30-60M_{\odot}$ then we still have some production of γ ray photons, but not as effective as for more massive stars. We will still have an instability due to the missing pressure. It triggers a collapse, that increases the density, the temperature (in this case we have the switching on of O burning, Ne burning and Si burning), but in this case the input of energy from these additional nuclear reactions rates enhances the mass loss from the star.
The new reactions lead to a re-expansion of the core, which leads to an enhanced mass loss, which happens like a pulsation. The core pulses up, the entire star pulses up. This is an oscillation. At the end of one oscillation, the mass of the outer layer is lost and the star finds either a new equilibrium or evolves for a little bit longer until pair instability triggers a contraction of the core which switches on again O, Ne and Si burning, which means again expansion of the core, which leads again to an oscillation with mass loss. So we can have a set or one single oscillation, at the end of which the star has lost some mass which allows it to find a new equilibrium to a smaller mass. The star basically loses mass till it finds a new equilibrium. Since there is this enhancement of the mass loss, the final mass of the star will be lower.
And this pulsational pair instability, which is NOT a SN, just an instability! After this instability, the star will evolve a little bit more and we will end up with a core-collapse SN or a direct collapse.

v. Sum up

Can we say something about the formation of compact objects? Let's start with a rule of thumb: let's say our massive star has low metallicity, less than $0.5Z_{\odot}$, then stellar winds are quenched, $\dot{m}$ is lower, the star at the onset of SN will have a larger mass, then it is more likely that it will collapse to a black hole directly. We can allow for some uncertainties due to models of islands of explodability or other criteria for core-collapse SN. Nevertheless, If the star is massive, it is more likely that it will have a higher compactness, a higher C-O core mass.

Let's analyze the plots by Heger et al.. They show the most likely final fate of a solar Z star and of a zero Z star. Let's look at the simplified cartoons derived form these plots, that encode just the essential information.

- On the left plot we have the final mass as a function of the initial mass, with solar Z. The black line is an ideal line when the end mass and the ZAMS mass are the same. The blue line is the actual pre-SN mass of the star. The red line is the final mass of the compact object from the models by Heger et collaborators. At solar Z, blue line is definitely lower than the black line and the difference is due to stellar winds that is always super efficient at solar Z. The red line is also much lower than the blue one. The final mass of the compact object at solar Z will be always small because the mass loss by stellar winds removes a lot of mass and then the core-collapse SN always succeeds.
- On the right hand side the plot is the same but for zero Z. This is the metal free stars, the pop III stars. Since Z is zero, we do not have stellar winds form metal lines, we only have Tomson scattering, the g_{ES} component. So the blue line almost overlaps with the black one. Stellar winds are really inefficient. The red line presents two different regimes: below $40M_{\odot}$ final mass is sufficiently small so that we expect a core-collapse SN to take place. For masses above $40M_{\odot}$ the star is so massive at the onset of collapse that the SN fails anyway and the star collapses directly.

So at solar Z we produce really small compact objects, whereas at zero Z we can produce BH with masses similar to those detected by LIGO/Virgo, above $30M_{\odot}$.
But what happens to the rest of the universe? Solar Z and zero are really the extremes. The metal free universe is extremely simple to describe (just Tomson scattering). Solar Z is not that easy, but we have a lot of observations. We can a have a lot of measurements of mass loss and calibrate the stellar evolution models and the core-collapse SN models to give a good description.
What happens at Z between solar and zero is a totally different story, because it's more difficult to have observations and do the models. Let's see a couple of hints about the models that describe the intermediate Z cases.
The first plot shows the mass of the pre-SN star as a function of the initial mass for different Z: there's clear trend where lower Z stars end their lives with a higher pre-SN mass. How does this translate into the final compact object mass? We can have either a core-collapse SN explosion or a direct collapse. Here the prescription to have a core-collapse SN is the simplest one: we know the C-O core mass of the star, we know the total mass of the star at the end of stellar life and we infer the mass of the compact object.
If we assume the stellar wind models as we have seen and if we assume that we can model the outcome of the core-collapse SN as a function of the mass of the C-O core then we expect the following trend: the lower Z stars will lead to higher maximum BH mass than metal rich one. All the details have a dependence that is more or less strong on the stellar winds, for which we have good models.

vi. Lower mass gap

We should be quite worried about the dependence on core-collapse SN models. Let's analyze this plot where different such models are employed. Different models have a huge impact on the low mass end of the stellar masses. Here we do not use the islands of explodability which would make it more complex.
If we zoom on the low mass region we see some important features. If the progenitors are between $9M_{\odot}$ and $24M_{\odot}$ the different core-collapse SN models change a lot the region of the plot where we go from NS to the light BH. The rapid model has a basically vertical line: we have nothing with $3-4M_{\odot}$ and then we produce BHs. In the delayed

model we have a very mild slope going between the NS regime to the BHs. The fallback is very imporatnt and produces BHs with $3 - 4M_\odot$ black holes. We expect a continuum between the masses of NS and the masses of light BH.
We can see this even better in the following plot, where we simulate a population of stars with an initial mass distribution that resembles the observed one (Kroupa, Sal Peter). The delayed model (green) predicts the existence of a lot of compact objects between 2 and $6M_\odot$. On the other hand, the rapid model (yellow) does not predict any compact object in the same mass interval. It's the so called lower mass gap in the BH mass spectrum.

Some models predict a gap between 2 and $5M_\odot$ i the BH mass spectrum. There are not many objects in there as observed by LIGO/Virgo. Furthermore, they do not count because they are the result of the merger. In 190814, we observe an object of about $2.6M_\odot$ which was determined with a very high accuracy because this was an event with a very high SNR. The origin of this object is very puzzling and puts a really big question mark about the existence of this lower mass gap.
Is the existence of a lower mass gap a real thing? It is more observational than theoretical. From the point of view of core-collapse SN, they are not predictive so we cannot tell whether the mas gap exists or not, since it depends on which model we are using. From the point of view of observations, it seems that this region is rather not populated with compact objects with respect to the other regions.

VII. Upper mass gap

So far we have put in our models only stellar evolution (with winds) and the core-collapse SN. We have mentioned pair instability and what it causes to a star.

Let's analyze the usual plot of the expected mass of the compact object as a function of the ZAMS mass for different Z, but we will extend the x axes to super huge stars, up $350M_\odot$, because to understand the impact of pair instability it is better to include them. Do they exist somewhere in the universe? We do not know, the most massive stars that we can observe, limited to the local universe, have a mass of about $150 - 200M_\odot$. We know a couple of these stars in the Magellanic cloud.
This plot without pair instability and pulsational pair instability, it just includes stellar evolution and core-collapse SN and we see the usual trend: the lower the Z the higher the final BH mass. For low Z we build BH between $3M_\odot$ and $280M_\odot$ in a continuous way. If we switch on pair instability we have a very strong difference.
In the region of ZAMS mass between $110M_\odot$ and $230M_\odot$, the low Z cases, the mass of the remnant is zero: we have a full pair instability SN that destroys the star completely. Even for a few slightly higher Z we have a region with zero mass compact objects, but the region becomes narrower and narrower till $Z = 0.01$ and higher. The reason is that the mass of the He and C-O core depends on Z. For low Z we can reach high masses, with high central T and density, while for large Z the core remains smaller. This is mostly due to the fact that stellar winds remove mass, and prevent the core from growing significantly.
The second difference depending on Z is when we look at a star of $\sim 60M_\odot$ that leaves a BH of about $55M_\odot$, and then we have a drop, which was not present in the model without pair instability. This dip is where pulsational pair instability becomes important. The star loses a lot of mass because of pulsational pair instability, but it finds a new equilibrium so it does not die as pair instability SN, but when it collapses to become a BH it becomes a smaller BH because there was this mass loss induced by the pair instability. So in this region we expect a local maximum for the BH mass, but it is model dependent ($\sim 55M_\odot$ for this model).

Then we have another region: if somewhere in the universe exist super massive stars that are sufficiently metal poor (above $200M_\odot$) with $Z < 5 \times 10^{-4}$ we can have the formation of massive BH via direct collapse? Isn't pair instability efficient? It is efficient, but it does not drive any SN. When the He core mass is above $130M_\odot$ then pair instability occurs, because of course such huge stars can produce electron-positron pairs in their core and they can undergo pair instability, but the gravity force of the outer layer of the star is so strong that even the switch on of the Ne and Si

burning in the stellar core during pair instability cannot drive a SN explosion. The pair instability in this case just leads to the total collapse of the star. We expect to form intermediate mass BH via pair instability. These are BH with masses higher than $100M_\odot$ and below the mass of SMBH. It is a gap and not a threshold, because if we have massive enough stars, they are expected to form BH in the intermediate mass regime. The gap is approximately between $60M_\odot$ and $120M_\odot$.
How confident are we about these boundaries? It's a current hot topic. We should allow for $20-25M_\odot$ uncertainty in the lower edge of the mass gap, and even for larger mass uncertainties in the upper bound of the mass gap. This uncertainties are not originated much from uncertainties in the pair instability formalism itself, but it the result of uncertainties in the massive star evolution. The boundaries of the mass gap can change if we assume that the stars are fast rotators and we don't know, if we treat in a different way the convection inside massive stars and there are several different models for convection in massive stars, if we change a little bit some nuclear reaction rates within the experimental uncertainties.

Overlapping the mass gap with the LIGO/Virgo mass cartoon, we see that a conservative assumption of the mass gap is that it is between $65M_\odot$ and $150M_\odot$. Is it confirmed? Not so, but all the BH are just the merger products, not the result of the collapse of a single star, with possibly one exception, GW190521. It can be though that it is the result of a previous merger but if so how could it find another companion to merge with. This require the hierarchical merger scenario.

viii. The role of spin

The most uncertain thing is not the mass, but the spin and especially its magnitude. The spin of a compact object if it comes from a star should be related to the spin of the core of the star at the end of stellar evolution. Otherwise we would need a total dissipation of angular momentum. If we have a successful SN explosion we have a lot of transfer of angular momentum and we do expect a part which can be large of the spin that can be lost.
In fact, when we observe the spin of a relatively young pulsar we observe a relatively low spin. In the case of SN we have a lot of dissipation of the final rotation of the core and relatively low spinning NS, less than 0.1 in terms of the dimensionless spin parameters.

If the SN is not totally successful, we may go through a stage of accretion before producing the BH. We may have a jet or a disk around the proto compact object. So we expect some transfer of angular momentum, if the BH does not come from a total direct collapse.
If the BH comes from a direct collapse we expect the angular momentum to be preserved. Let's take the final angular momentum of the core of a massive star in most of the stellar evolution models that are available, assume that everything directly collapses to a BH, then we would have only fast spinning BH. But what we observe with LIGO/Virgo is relatively slow spinning.

One of the most interesting ways out is to consider the impact of stellar magnetic field, which is neglected in most stellar evolution models, because the physics is already complicated. There are some processes that happen if we include the magnetic field and that favor the transfer of angular momentum from the stellar core to the outer part of the star. One of the them is the so called Taylor-Spruit dynamo, where if the core of a star is really fast rotating it powers its own magnetic field like in a dynamo, the field then couples the core with the envelope of the star, and this allows to transfer angular momentum form the fast rotating core to the envelope. Once the angular momentum is in the envelope it is extremely easy to dissipate it, because part of the envelope is lost by stellar winds and the angular momentum with it.
Fuller and Ma apllied the Taylor and Spruit mechanism to their stellar evolution model. The plot shows the spin parameter (between 0 and 1) of the He core of the star at the onset of collapse, and they overlapped the data points with errors. The important point is that the model predicts that the progenitor cannot spin fast if the Taylor-Spruit dynamo removes the spin. There are only a couple of points with a high spin, but those belong to stars that were

evolved in binary systems, where we have mechanisms to spin up stars. This is a model to explain why even massive black holes that are born from direct collapse have a small spin.
Does this mechanism work for all stars? We do have BH that are nearly maximally spinning! The Taylor-Spruit dynamo cannot work for all stars. The only way out is binary evolution.

What about the directions of the spin in the sky? The spin direction is a little bit more clear, because we believe that it is very difficult to change the orientation of the spin of the progenitor star when it becomes a BH. The only process that can significantly change the direction of the spin form the progenitor to the BH or NS is the SN, which can give a kick to the BH or NS. This cannot change the direction of the spin of the compact object but it re-adjusts the direction of the orbital plane of the binary. The relative inclination between the spin of the compact object and the angular momentum changes.

IV. Binary Evolution Processes

So far we have seen how black holes and neutron stars form from massive stars. LIGO and Virgo do not observe single BH or NS, but binaries with an initial distance sufficiently short to make them merge. For a binary to shrink and merge via GW only in less than a Hubble time the initial separation between the two objects should be really really small. For two relatively massive BH we need an initial separation of less than $100R_{\odot}$, which is much less than the distance between Earth and the Sun.

i. Isolated binary evolution

Let's take two massive stars and assume they are members of a binary system since their formation from the gas cloud, they evolve, they become two BH that merge via GW emission. We know from observations that almost all of the massive stars form binaries.
There is one problem: evolving two stars as two single stars or as members of a loose binary system (more than $10000R_{\odot}$) is very different from evolving two stars in a tight binary with an initial orbital separation of less than $10000R_{\odot}$, because if the binary is tight enough the two stars are affected by binary evolution processes that can completely change the final stages of the evolution of the binary. We form no BH at all, or just one.
We will discuss the most important theories for the formation of two binary BH. We will just focus on mass transfer processes. This kind of evolution can be described with the so called binary population synthesis code.

Before getting started, let's refresh some definitions from Newtonian binary system: reduced mass, angular frequency, energy, angular momentum.

ii. Mass transfer

We have three different channels of mass transfer:

- Via wind
- Via Roche lobe overflow
- Via common envelope

iii. Wind mass transfer

The idea is simple: if at least one of two stars loses mass by stellar winds, it loses mass in all the possible directions. The secondary component of the binary can catch some of this mass and accrete it. How much? Not so much actually. The wind is lost isotropically in all directions and the companion cannot be isotropically in all directions simultaneously. It will only accrete what it intersects during its orbit, eventually with some gravitational focus if it is massive.
If the mass lost by the donor star is $\dot{M}_{1W}$ the mass acquired by the companion will be only a tiny fraction of the mass lost by the primary. Namely it will scale with the fourth power of the orbital velocity over the wind velocity. If the orbital velocity is higher, the accretion is higher because the secondary has more chance to sweep all the matter lost by stellar wind before it goes away. If the wind escapes really fast, on the other hand, there is not enough time for the companion to accrete it. The size of this wind velocity is order of thousands km per s. Whereas the orbital velocity is order of hundreds km per s. Therefore, the speed of the wind wins. Accretion by stellar wind is usually very inefficient.
The fact that it is inefficient does not mean that it is not important. Actually, a good fraction of X ray binaries (the donor mass is much larger than that of the accretor) we observe in the sky are wind accreting systems, they are wind powered systems. It is not efficient in the sense that the accreted mass $\dot{M}_{2A}$ is much less than the lost mass, but it is important from the observational point of view.

IV. Roche lobe overflow

Let's start visualizing the assumptions of our system: binary star with no eccentricity. The two stars are point masses and anyway spherical (reasonably smaller than the orbit). Let's consider a stationary test particle in a frame that co-rotates with the binary. This test particle will feel an acceleration that we can define as an acceleration given by an effective potential ϕ_R (where R stands for Roche):

$$\phi_R = \frac{Gm_1}{|\vec{r} - \vec{d_1}|} + \frac{Gm_2}{|\vec{r} - \vec{d_2}|} + \frac{1}{2}|\vec{\omega} \times \vec{r}|^2 \tag{42}$$

where the first terms is the gravitational pull exerted by star 1, the second the gravitational pull form star 2 and the last one is the centrifugal term, since we are in co-rotating system. $\vec{d_1}$ and $\vec{d_2}$ are the positions of the two stars in our reference frame, $\vec{r}$ is the position of the test particle in the same reference frame.
Because of the definition of the center of mass of a system, we have some useful equivalences:

$$\frac{\vec{d_1}}{m_2} = -\frac{\vec{d_2}}{m_1} = \frac{\vec{d}}{m_1 + m_2} \tag{43}$$

with $\vec{d} \equiv \vec{d_1} - \vec{d_2}$, the relative distance between the two stars.
If we replace these in the potential equation, we find that ϕ_R does not depend individually on $\vec{d_1}$, $\vec{d_2}$ or m_1 and m_2, but it depends only on $\vec{d}$ the relative distance between the two stars and q the mass ratio, defined as the ratio between the donor an d the accretor.

Let's now take this potential and draw equi-potential surfaces around the two stars. At some point we have an equi-potential surface that is 8-shaped. Each drop of this shape is one of the two Roche lobes. The only single point that connects the two lobes is the L_1 inner Lagrangian point, a point of unstable equilibrium. It can be shown that it sufficient that the mass ratio is larger than zero for the 8-shaped surface to exist.
Making the usual approximations, the lobes are not perfectly spherical but we can approximate them to two perfect circles. It is possible to calculate an analytic version of the size the Roche lobes (accurate to $\sim 1\%$). For example, the size of the Roche lobe r_1 depends only on the semi-major axis a and on the mass ratio q:

$$\frac{r_1}{a} = \frac{0.49q^{2/3}}{0.6q^{2/3} + \ln 1 + q^{1/3}} \tag{44}$$

To calculate r_2 we just revert the q.
From the physical point of view, because of the definition of equipotential surface, if the radius of a star reaches the surface of the Roche lobe, which we say in technical words, the star fills its Roche lobe, then matter can flow without the need of additional energy from one lobe to the other and it can be accreted on the surface of the other star or compact object. The star that fills its Roche lobe is called the donor and the one that receives matter is called accretor.
In this case, we also have a significant geometry for the accretion and this is also the reason why it is way more efficient than stellar winds. Since we are in a rotating system there will be some Coriolis force that we have neglected so far and this means that matter that reaches the accretor will have a non zero angular momentum and it will produce a disk-like structure around the accretor. The processes that happen in this disk because of viscosity for example increase the efficiency of accretion onto the companion.
Note that the accretor is not necessarily a compact object. In case it is then we have an X ray binary. The accretor can be a non degenerate star as well.

V. Stability of the Roche lobe

We have three fundamental timescales in stellar evolution:

- Dynamical timescale, it's the shortest one, it is the time the star starts to collaps under its own gravity if we assume that all the pressure is removed instantaneously, which is unphysical in most cases. For solar values it is of order of ~ 3000s, amazingly short
- Kelvin-Helmholzt, a little bit longer. It's the time for a star to contract by radiating all of its gravitational potential energy assuming that it is the gas that provides the radiation, not the nuclear processes. It is the ratio between the gravitational potential energy and rate of energy loss, which is the luminosity. For the Sun this is about ~ 30Myr. It depends on the star mass squared.
- Nuclear timescale: it's the timescale for nuclear burning to change the stellar properties. It's way longer, and it also depends on the fuel that is burning. The H is the one with the longest timescale. If calculate it for the Sun and for H we get $\sim 10^{10}$yr. Note that f is the fraction of mass which is locked inside the core, X is the fraction of mass made of H.

These timescales are important because the Roche overflow mass transfer can become unstable either on the dynamical timescale, or on the Kelvin-Helmholtz timescale or on the nuclear timescale. If it becomes unstable on the nuclear timescale it basically means that it is stable, because it means that the star itself has to change its properties for the Roche lobe to stop or to begin. In the other two cases, it becomes unstable but two different kinds of instability.

Let's see how we can evaluate whether the Roche lobe is stable or unstable and on which timescale. The best treatment that has been proposed over the years is to look at the ζ coefficients. ζ_L is the change of the size of the Roche lobe as a response of mass loss via Roche lobe mass transfer. Here we are talking about the donor:

$$\zeta_L = \frac{d \ln R_L}{d \ln M} \tag{45}$$

so R_L and M are the radius and the mass of the donor. The Roche lobe changes when the donor loses mass. It can shrink or increase. ζ_L will be larger when the Roche lobe decreases.

ζ_{ad} is the change of the radius of the star to re-adjust to a new hydrostatic equilibrium as an effect of mass loss. The star is trying to find again its basic equilibrium, hydrostatic equilibrium. If it does not re-adjust fast enough to a new equilibrium it goes unstable dynamically.

$$\zeta_{ad} = \left(\frac{d \ln R}{d \ln M} \right)_{ad} \tag{46}$$

ζ_{th} is the parameter that says what is the change of the radius needed for the star to re-adjust to a new thermal equilibrium, to a new Kelvin-Helmholtz equilibrium because of the mass loss

$$\zeta_{th} = \left(\frac{d \ln R}{d \ln M} \right)_{th} \tag{47}$$

If we can evaluate these three ζ, we have:

- If $\zeta_L > \zeta_{ad}$ the Roche lobe shrinks faster than how the star shrinks in response to the mass loss, the star will always be filling the Roche lobe. In this case, the star will always be overfilling the Roche lobe, and the mass transfer will become unstable over a dynamical timescale, because this is associated with hydrostatic equilibrium and so dynamical timescale.
- If $\zeta_{ad} > \zeta_L > \zeta_{th}$ the Roche lobe will shrink faster than the time the star needs to re-adjust to a new Kelvin-Helmholtz equilibrium. It means that the Roche lobe will be thermally unstable and this meas that it will be quite large. During the time the star is thermally unstable it will shred more mass through the Roche lobe.
- If $\min(\zeta ad, \zeta_{th}) > \zeta_L$ we have to wait for a nuclear evolution timescale for the star to start feeling or to stop feeling the Roche lobe and so we can say that the Roche lobe mass transfer is stable.

What we are more interested in is the dynamically unstable case.

v.1 Modeling a Roche lobe overflow

We assume the simplest possible mass transfer, a conservative one. All the mass that is lost from the donor, star 1, will be accreted by star 2. Also the total angular momentum of the system will be conserved. If we have $L=$const nad $m_1+m_2=$const, no mass losses, then we have another constant obtained rearranging the two conserved quantities: $(m_1m_2)^2a=$const. This means that the size of the semi-major axis will have a minimum for $m_1=m_2$. If at the beginning of the Roche flow m_1, the donor, is larger than m_2, as an effect of conservative Roche lobe the orbital separation decreases.
If initially $m_1<m_2$ the orbital separation increases. The orbital separation has in turn an obvious impact on the Roche lobe, because it scales as the semi-major axis a: $R_L\propto a$.

How to model a very simplified Roche lobe overflow: assume conservative mass transfer, assume that star 1 fills the Roche lobe all the time, and that we have a mass transfer of amount $dm=1M_\odot$ until m_1 goes from $100M_\odot$ to $15M_\odot$ then we assume that magically the mass transfer stops. How do a, m_1 and m_2 evolve?
We see that m_1 is the donor and it decreases and m_2 increases taking exactly the mass lost by m_1 step by step. If we look at the semi-major axis, it first decreases, then it reaches a minimum when $m_1=m_2$ and then it starts increasing again when the accretor becomes the more massive one.
At the beginning, in most mass transfers we observe in Nature, we have the mass of the donor larger than the mass of the accretor, then a initially shrinks. The Roche lobe of the donor also shrinks, because the orbital separation is connected to the Roche lobe. Star 1 has to shrink otherwise it overfills its Roche lobe. If it cannot make it then the Roche lobe becomes dynamically unstable and we will see that this means a common envelope phase. If the star shrinks fast enough to avoid a dynamically unstable situation but cannot avoid to be out of thermal equilibrium, the donor for a while is not in thermal equilibrium and it has to re-establish it at some point. It triggers a mass rate $\dot{m}_1$ which scales with the mass of the donor over the KH timescale, which is a pretty large amount of mass loss. So if the donor loses its thermal equilibrium we do not have a dramatic instability, but the mass loss is quite large because it is proportional to how the star is out of its KH equilibrium.
When the two masses flip, if the system is still in Roche lobe, the orbital separation starts to increase and so it is easier for the donor to re-establish its thermal equilibrium because the Roche lobe increases faster than the radius of the star. Mass transfer can still continue but on a nuclear timescale.

As a general consideration we know that the thermal timescale is much sorter than the nuclear one. It means that the mass transfer will be larger when $m_1>m_2$ but the time of this thermally unstable mass transfer will be also shorter than the other one, because the thermal timescale is shorter than the nuclear one. So the thermal mass transfer is more efficient but shorter.
It is much easier to observe a system with a thermally stable mass transfer than one with a thermally unstable mass, because the first one lasts longer. Nevertheless, most of the mass transfer episodes we observe (X ray binaries for example) are with $m_1<m_2$: it is the one that happens on a longer timescale.

VI. Common envelope

Common envelope is what happens in most of the cases when we have a dynamically unstable Roche lobe flow. We have two massive stars initially under-filling their Roche lobes, then one of two fills its Roche lobe, the mass transfer becomes unstable over a dynamical timescale, which means that the envelope of the Roche lobe filling star becomes very very large and engulfs also the other member of the binary. We have a system where the two cores orbit around each other inside a very extended halo which does not co-rotate anymore with the internal structure.
The key point is that initially the two stars orbit each other in the interstellar medium and the ISM has a super low density, less than 1 particle per cm^3. What we expect here is that the two of them orbit without any friction. In the case the two cores are in the common envelope they orbit about each other in a much higher density (10 orders of magnitude higher). There will be some friction that will make the two cores spiraling in, because it will extract kinetic energy from them and it will make the binary shrink. The removed energy will not disappear, but it will be

converted in another form of energy. One of the possible outcomes is that it goes into the internal energy of the envelope. As a consequence, the envelope inflates and becomes less and less bound to the rest of the configuration. So at this point we have 2 possibilities:

- If the energy we transfer to the envelope is big enough, the envelope will be completely unbound and ejected at some point and we are left with a new binary which is a naked binary, just the two central cores and no envelope anymore.
- If there is not enough energy to unbound the envelope, nothing will prevent the two cores from spiraling in and at some point they merge and we have a single object left.

Why is this important for the specific case of binary compact objects? Here we do an example with a BH, but it works the same for a NS. It is important because if the energy is enough to eject the envelope then we are left with a new binary system. If it is not we are left with a single object, either a BH or NS, which unfortunately is not a source of GW. If the two bodies survive, we have a new binary composed of the compact object and the core of the companion which is naked He star, similar to a WR star. The orbital separation initially is two orders of magnitude larger than after the envelope ejection. This means that since the two bodies are closer they are a brighter X ray binary, plus if also the companion collapses and forms a BH or NS the binary will a short binary separation ($1 - 100R\odot$).
This is essential for the binary to coalesce in less than a Hubble time. In fact, is it possible for a massive binary star with an initial separation less than $100R_\odot$ and then form in the end a binary BH with an orbital separation less than $100R_\odot$? We said at the beginning that in order for a system to coalesce via GW the initial orbital separation should be less than $100R_\odot$. One could answer that we could just start with a stellar system with an orbital separation less than $100R_\odot$, but it does not work like this, because the maximum radius of a star during its evolution is of order of $100 - 1000R_\odot$ or even more. If we started with such a small separation the two stars would merge before forming two BH or NS. We need the common envelope episode to shrink the binary.

There's one exception which is the chemically homogeneous scenario which can avoid the common envelope.

The common envelope is the least understood binary evolution process. We can divide it in 4 different stages and we only understand two of them:

1. Loss of corotation: when the companion plunges into the envelope of the donor. This is not yet consistently modeled.

2. Fast spiral in: in the beginning we have a lot of envelope and the two cores spiral in basically on a dynamical timescale (hundred days or so). This has been simulated in 3d, we understand something. But when the system reaches a configuration when inside the orbit of the two cores there is the same mass of the envelope as the total mass of the cores then it is almost impossible to shrink them further. We do not have enough mass budget. What happens is that the common envelope slows down. The evolution goes from dynamical to KH timescale (hundred thousand yr)

3. Slow spiral in: poorly understood, what removes the envelope?

4. The merger: we understand this one

What has been done so far is the following: α formalism. It is based on one simple idea: we have only 2 kinds of energies and the transfer is only from one kind to the other. There is no change in angular momentum nor other sources of energy.
One is the orbital kinetic energy of the cores:

$$E_{orb} = -\frac{1}{2}\frac{GM_{c,1}M_{c,2}}{a} \tag{48}$$

We assume that a fraction of the orbital energy goes into unbinding the envelope:

$$E_{bind,i} = -\frac{G}{\lambda}\left(\frac{M_1 M_{env,1}}{r_1} + \frac{M_2 M_{env,2}}{r_2}\right) \tag{49}$$

where λ is a parameter that describes the profile of the envelope. Therefore we have that the envelope is ejected if the energy of the envelope is smaller or at maximum equal to a fraction α of the variation of the orbital energy:

$$E_{bind,i} = \Delta E_{orb} = \alpha \left(E_{orb,f} - E_{orb,i}\right) \tag{50}$$

α is the dimensionless parameter that tells how much of the variation of the orbital energy efficiently goes into unbinding the envelope. so it should be between zero and 1. The problem is that we can try to calibrate the value of α with observations and in some cases we get a value that is larger than one. How is it possible? It means we create energy!
It is possible indeed, because this, the orbital energy, is not the only source of energy that plays here. There are other things: if the common envelope engulfs the compact object it will accrete part of matter and also generate outflows which will participate into unbinding the envelope.
Another example is the recombination: when the envelope becomes super extended its outer part reaches T well below few thousands K, some molecules and atoms start recombining which generate energy that contributes to unbound the envelope even more.

We are pretty aware that the α formalism does not capture the physics of the common envelope, which includes many other sources of energy. What can we do then? There are 2 ways:

- Insist on the modeling of the 4 stages adding more and more physics into the hydrodynamical models, but this is challenging computationally
- Complementary approach, consider α a parameter free to be larger than 1, as if it represents all the energy available to the system. This approach is fine if we want to stude a statistically large sample of these binary systems.

Our decision of the $\alpha\lambda$ parameter will change the spectrum of BBH masses and the rate of BBH and BNS mergers. Do we have some observations? We do observe some post common envelope systems: the Cat's nebula for example. But all of these systems are relatively low mass binaries.

VII. Chemically homogeneous model

What happens if we try to avoid common envelope? Is there an alternative way to shrink the binary without the common envelope? This leads to the development of the chemically homogeneous evolution model. The basic idea is that if a star remains chemically homogeneous, that means that it does not develop a gradient with C at the center, He a little bit outside and H in the envelope, but remains all mixed together then the radius of the star will be smaller. Maybe sufficiently small that two stars can orbit each other at a distance of about $100R_\odot$, without merging too soon.
In order for a star to remain all chemically mixed the star has to rotate really fast. Fast rotation induce mixing of the species inside the star. If the star rotates fast it can be chemically homogeneous and the radius can be much smaller. Two stars of this kind can start very close to each other and form BH that are very close to each other.

The problem is that not all stars can be fast rotators, because if Z is too high stellar winds are important and so no fast rotation, because winds remove angular momentum. Even if the star starts fast rotating at some point it is not fast rotating any more.
The rotator mechanism works only for metal poor stars. It also has other features that make it quite peculiar. Most of these predict equal mass binary BH, and does not predict BH with masses below $25M_\odot$, so we need another way to form smaller BH. It also prefers aligned spins and maximally rotating spins, which again is observed only in very

few systems.
This does not mean that the chemically homogeneous model is not good, maybe it accounts only for a fraction of systems that we observe with LIGO/Virgo.

VIII. SN kicks

SN explosion when it happens it has not only the effect of removing a lot of mass and leaving a small compact object, it has also the effect of giving a natal kick to the compact object. Why should it give a kick?
If the SN has some asymmetry (which is very likely) there will be asymmetry in the final outcome and a kick. If for example, most of the ejecta go in one direction, then for conservation of momentum the compact object will get a kick in the opposite direction.
Even if the ejecta are perfectly symmetric, but not so the neutrino loss, we will get a kick
Even if the SN is totally symmetric but happens in a binary system it will change the configuration of the binary system. If the star that undergoes SN loses more that 50% of its mass this will change instantaneously the total mass of the binary and so the semi-major axis and the eccentricity and this is what is called the Blaauw kick. This last one is less important. Still all of these kicks mean a possible unbinding of the binary.

From the observational point of view we can measure kicks for neutron stars, namely for pulsars. If we observe the proper motion of young pulsars, which is mainly affected by SN, then we can reconstruct the size of the kick. For the moment we have a sample of ~ 200 pulsars, but only about 50 of them are young and can be used to derive the kick.
We obtain a histogram and see that the distribution of the kicks expressed as pdf perfectly fits a Maxwellian distribution with a 1d velocity of about 265km/s. This is what we know for NS. What about BH? Drama! It's super difficult to observe the natal kicks of BH. The only thing we have are the proper motion of BH in X ray binaries in the Milky Way, but these could be affected by many other things. What people do is one of three options:

1. Assume the energy available for the kick of a NS and a BH is the same and we need to conserve linear momentum and so the BH will get a lower kick than the NS just because it is more massive
2. Assume that BH form without SN get no kick, while BH formed with some fallback get a lower kick modulated by the amount of fallback
3. Assume there is linear momentum conservation and that the kick depends only on the asymmetry of the mass of the ejecta. In this case the size of the kick will depend on the mass of the ejecta itself. This model is universal, valid for both BH and NS.

IX. GW decay

We use the formalism by Peters 1964. We have a trend for the semi-major axis evolution and also for eccentricity evolution.

X. Sum up

V. MOBSE - Massive Object Binary Stellar Evolution -

There are three different approaches to study this kind of binaries:

1. Very low statistics high accuracy hydro-dynamical simulation of a specific process. We are pushing here all the possible physics. For just one binary we need like ten million CPU hours.

2. Intermediate approach: binary evolution with an hydrostatic stellar evolution code + binary evolution processes. A million binaries for the million hours of CPU time. Yet we expect one BBH merger every $10^4 - 10^7$ binary stars. One needs to be very lucky.

3. Very high statistics and more simplifications, this is MOBSE. It's a semi-analytic evolution. We can simulate millions of binary in less than one day with a single CPU. Here we have many simplifications, but a very large statistics.

The ideal situation is the one in which we use the three possible approaches to calibrate each other.
We use the MESA code to calibrate the efficiency of mass transfer.
The challenge is to translate a lot of physics into the simplest possible model.

In practice, what are the main simplifications?

- Single stellar evolution: we cannot integrate the equations of stellar structure, the ones that give the properties of star per every shell. We use the polynomial fitting formulas (Harley et al.). But ones they have been done for a given set, we need to redo everything with the updated sets.

We need to have a binary population synthesis code that interpolates these quantities, evolution of $M(t)$, $L(t)$, $R(t)$, from tables of stellar evolution obtained from other codes.
How to re-map the properties of the star at the end of its life and the SN properties to the ones of the compact object. We have to take prescriptions from people who run hydro-dynamical simulations.

For the Roche-Lobe we use the effective potential, definition of stability based on ζ parameters, α formalism for the common envelope.
Analytic model means that everything is described by analytic formulas with no free parameters to calibrate.
Semi-analytic requires some parameters to be calibrated on data or other simulations, like the treatment of the Roche-Lobe.

VI. Dynamics of Stars and Black Holes in Dense Stellar Systems

We will see how to assemble compact binaries in dense stellar environments. We mean especially BBHs, because dynamics is way more important for BHs than for smaller compact objects like NSs.
Why we should care about dynamics? One of the most convincing reasons is that massive stars which are the progenitors of BHs and NSs form preferentially in cluster environments, in star clusters, which are dynamically active places. In the solar neighbor the density of stars is less 1 per pc^3, so it is extremely unlikely that two stars get sufficiently close to perturb their orbit unless they are in a binary system. In a star cluster the central density is higher that a thousand stars per pc^3, so it is likely that stars have close flybys with each other.

- When we talk about star clusters, we usually think of **globular clusters** as opposite to stellar fields, like the solar neighbor, the low density regions of our Galaxy. Globular clusters are very massive ($> 10^4 M_\odot$), they are long lived (12Gyr) and they are one the densest places in the Universe, so that here encounters are important. On the other hand, globular clusters are also rare systems: they represent less than 1% of the baryonic mass in the Universe.
- There are also other flavors of star clusters, important from the point of view of dynamics like **young star clusters** and **open clusters**. They are more short lived ($<$ 1Gyr), they are less massive (down to $100 M_\odot$, i.e. a few stars), and they are also the most common place of massive stars formation (about 80% of star formation happens in these young star clusters).
  The importance of young star clusters and open star cluster lies in the fact that when they are young (few hundreds of Myr up to 1Gyr) they are dynamically active environments and when they die (1Gyr), because they are tidally disrupted in the field of their host galaxy, they provide the stars and compact objects that were inside the star cluster to the field. Therefore, a portion of the population of a field was born and evolved in young star clusters and open clusters.
- The last flavor of star clusters we have in the Universe are **nuclear star clusters**. They are more rare but very peculiar. They lie in the very center of Galaxies, not in all of them (we have one in the Milky Way). They are very massive (as massive as the most massive globular clusters or even more $> 10^6 M_\odot$). Sometimes they coexist with the central SMBH, which means a lot of dynamics.

In star clusters stars can have close encounters. We will tackle collisional dynamics of stars. We will focus on binary stars. A binary star, in fact, is more active in the field of close encounters, because the geometric cross section is larger. For a single star the geometric cross-section is πR^2, whereas for a binary it is πa^2, where a is the semi-major axis. Moreover, a binary has another feature which is its internal energy:

$$E_{\mathrm{int}} = \frac{1}{2}\mu v^2 - \frac{G m_1 m_2}{r} \tag{51}$$

where v is the modulus of the relative velocity between the two bodies and r their relative distance. The internal energy is something the binary can exchange with compact object when it interacts with them via close encounters. Basically the internal energy is a sort of energy reservoir that can be used to change the relative kinetic energy between binary system and single stars in dense environments as star clusters. For a binary:

$$E_{\mathrm{int}} = -\frac{G m_1 m_2}{2a} = -E_b \tag{52}$$

i. 3-body encounters zoology

If we have a close encounter between a binary and a single star we have that some part of the kinetic energy of the binary can be exchanged with the kinetic energy of the single star. If the star acquires kinetic energy from the binary, the star acquires recoil velocity from the encounter. On the other hand, the binary has to provide this kinetic energy to the single star and the only way the binary can do this is that a of the binary changes (as long as the two masses do not change). This means that $a_f < a_i$ and this means that the final binding energy of the binary is greater than the initial one. If the star acquires kinetic energy from the binary, the binary shrinks. Also the center of mass of the binary acquire the same recoil velocity in the opposite direction.

There is also another way that allows the single to acquire kinetic energy with respect to the binary: by changing masses instead of the semi-major axis. This means we can have a **dynamical exchange** in which the single star kicks off one of two components of the binary and becomes a member of the binary itself. If $m_3 > m_2$ we increase the binding energy of the binary. This kind of exchange more likely happens (Hills and Fullerton) as a function of the ratio between the mass of the intruder and the primary component of the binary, when the intruder mass is as big as the primary component mass. This means essentially that exchanges tend to produce more and more massive binary stars, because they are more energetically stable, they have a higher binding energy. Physical systems naturally evolve to more energetically stable configurations.
Vice versa we can have the case where the single star loses some kinetic energy and transfers it to the binary system, which will expand.
It is also possible that the star removes so much energy from the binary that it basically **ionizes** the binary. We can calculate the critical velocity the intruder needs to have to unbind the binary completely.
In a resonant flyby the intruder completes some orbits around the binary and its orbit is much more perturbed than in a prompt flyby. We can also have a fast exchange or a resonant exchange.

II. Heggie's law

A binary can shrink or become wider or even break up. Statistically, thanks to the Heggie's law, we can know which scenario is more likely. It states that hard binaries tend to become harder and soft binaries tend to become softer as a result of averaged 3-body encounters.
We define hard and soft binaries according to their binding energy, which is compared to the average kinetic energy of a star in the star cluster.

- If the binary binding energy is larger than the average kinetic energy of a star in the cluster, then the binary is **hard**:
$$\frac{Gm_1m_2}{2a} > \frac{1}{2}\langle m\rangle\sigma^2 \tag{53}$$
where $\langle m\rangle$ is the average mass of a star in the cluster and σ is the velocity dispersion (in a cluster we do not have a bulk motion, we do not have rotation or a small amount of it, so the average kinetic energy is just given by the random motion of velocity dispersion).
- The binary is **soft** if:
$$\frac{Gm_1m_2}{2a} < \frac{1}{2}\langle m\rangle\sigma^2 \tag{54}$$
The fate of soft binaries is to break up.

III. Rate of 3-body encounters

We can derive the evolution of the semi-major axis due to these 3-body encounters.
We start by defining the **impact parameter** which is the component of the distance between the intruder and the center of mass of the target perpendicular to the intruder velocity.
We can now define the simplest cross-section for 3-body encounters as

$$\Sigma = \pi b_{\max}^2 \tag{55}$$

where $b_{\max}$ can be defined as the maximum impact parameter to have a non zero energy exchange between the star and the binary. How can we decide what is $b_{\max}$? A reasonable method is to use energy and angular momentum conservation and assume that the maximum pericenter distance is $p_{\max} = a$ where a is the semi-major axis, then the value of $b_{\max}$ can be written as:

$$\Sigma = 2\pi G \frac{m_T a}{v_\infty^2} \tag{56}$$

where m_T is the total mass of the binary, a its semi-major axis and v_∞ is the velocity of the intruder at infinity.
We write energy conservation at infinity (before the encounter) and at the distance of maximum approach (when the star and the binary are the closest). We balance them and we have one equation to try to derive the value of $b_{\max}$. We derive the second equation using angular momentum conservation. The two equations allow to express $b_{\max}$. A reasonable average approximation is that $v_\infty = \sigma$. We obtain:

$$b_{\max}^2 = \frac{2Gm_T}{\sigma^2} a \tag{57}$$

We see that this makes sense because if the binary mass, m_T, is larger it will gravitationally attract more bodies and so the cross section is larger. Also if the σ is smaller there will be a higher chance to interact (we are always assuming stars as point objects).
Starting with this Σ we can derive the interaction rate: number of interactions the binary will undergo per unit time. Under general physical consideration, it will be given by:

$$R = \frac{dN}{dt} = n\Sigma v_\infty \tag{58}$$

The denser the environment, the larger the cross section the more likely the encounters. If we substitute the expression for Σ we find that the rate of encounters is:

$$R = 2\pi G n \frac{m_T a}{v_\infty} \tag{59}$$

So the binary star in a star cluster will interact more if the density in the neighbor of the binary, if the binary is massive, if the semi-major axis of the binary is large and if the velocity between the bodies is small.

Can we also calculate on average how much energy is extracted during each encounter? Yes and no. We need some numerical calibration. Up to R we have used only analytical expressions, now we need some help from simulations. It's not easy to find a relation between energy exchange and encounters. From empirical encounters we find out that:

$$\frac{\Delta E_b}{E_b} \propto \frac{m_3}{m_1 + m_2} \tag{60}$$

The more massive is the binary the more difficult is to perturb it. This scaling can be expressed with a dimensionless constant usually called ξ, which can be calibrated via numerical scatterings:

$$\xi = \frac{m_1 + m_2}{m_3} \frac{\langle \Delta E_b \rangle}{E_b} \tag{61}$$

This constant should be between 0.1 and 1. Now we can express the rate of change of binding energy per encounter as the average variation of energy during a single encounter times dN/dt which is the rate we estimated.
We can explicitly write down the average variation using the numerically calibrated ξ:

$$\frac{dE_b}{dt} = \xi \frac{\langle m \rangle}{m_1 + m_2} E_b \frac{2\pi G(m_1 + m_2) n a}{\sigma} \tag{62}$$

Simplifying a little bit, $\langle m \rangle n = \rho$:

$$\frac{dE_b}{dt} = \pi \xi G^2 \frac{\rho}{\sigma} m_1 m_2 \tag{63}$$

So the rate of energy exchange depends on two properties of the cluster, mass density and the velocity dispersion, and on the product of the masses of the binary. If we assume that the properties of the cluster do not change with time and also that the masses of the binary stay the same (hard binary), then all of this tells us that **a hard binary hardens at a constant rate**.

Derivation of $b_{\max}$

We find $b_{\max}$ imposing energy and angular momentum conservation:

- **Energy conservation**:

$$\frac{1}{2}\frac{m_3(m_1+m_2)}{M_{tot}}v_f^2 - \frac{Gm_3(m_1+m_2)}{p} = \frac{1}{2}\frac{m_3(m_1+m_2)}{M_{tot}}\sigma^2 - \underbrace{\frac{Gm_3(m_1+m_2)}{D_{in}}}_{0} \tag{64}$$

where p is the point of closest approach. From this we find:

$$\frac{1}{2}\frac{\sigma^2}{M_{tot}} = \frac{1}{2}\frac{v_f^2}{M_{tot}} - \frac{G}{p} \tag{65}$$

- **Angular momentum conservation**: the angular momentum parallel to the impact parameter is conserved:

$$\Delta\vec{J} = \Delta\vec{J}_\perp = (pv_f - b\sigma)\frac{m_3(m_1+m_2)}{M_{tot}} \tag{66}$$

which implies:

$$b\sigma = pv_f \tag{67}$$

where we have used:

$$\begin{aligned} \vec{J}_i &= \frac{m_1m_2}{m_1+m_2}\sqrt{G(m_1+m_2)a}\hat{b}_\parallel + b\sigma\left(\frac{m_3(m_1+m_2)}{M_{tot}}\right)\hat{b}_\perp \\ \vec{J}_f &= \frac{m_1m_2}{m_1+m_2}\sqrt{G(m_1+m_2)a}\hat{b}_\parallel + pv_f\left(\frac{m_3(m_1+m_2)}{M_{tot}}\right)\hat{b}_\perp \end{aligned} \tag{68}$$

Combining the two equations we obtain an expression for the pericenter:

$$v_f = \frac{b\sigma}{p} \implies \frac{1}{2}\frac{\sigma^2}{M_{tot}} = \frac{1}{2}\frac{b^2\sigma^2}{p^2M_{tot}} - \frac{G}{p} \tag{69}$$

So we obtain a second order equation in p:

$$p^2\sigma^2 + 2GM_{tot}p - b^2\sigma^2 = 0 \tag{70}$$

which solutions are given by:

$$p = \frac{-G \pm G\sqrt{1+\frac{b^2\sigma^4}{G^2M_{tot}^2}}}{\frac{\sigma^2}{M_{tot}}} \sim \frac{b^2\sigma^2}{2GM_{tot}} \tag{71}$$

Therefore, we finally get an expression for the value of $b_{\max}$:

$$b_{\max}^2 = \frac{2GM_{tot}}{\sigma^2}p_{\max} = \frac{2FGM_{tot}}{\sigma^2}a \tag{72}$$

IV. Hardening timescale

If we express the derivative with respect to time of $1/a$ deriving the expression of the binary binding energy:

$$\frac{d}{dt}\left(\frac{1}{a}\right) = \frac{2}{Gm_1m_2}\frac{dE_b}{dt} = 2\pi G\xi\frac{\rho}{\sigma} \tag{73}$$

This is usually called hardening rate because it tells us how fast a binary shrinks because of close encounters. It does not even depend on the properties of the binary!
If we express this as the variation of the semi-major axis itself:

$$\frac{da}{dt} = -2\pi G\xi\frac{\rho}{\sigma}a^2 \tag{74}$$

We have a minus sign because the semi-major axis shrinks as the result of the encounters. If the binary is large it will shrink very efficiently, but when it gets small it becomes difficult to keep shrinking because the cross-section of the binary becomes smaller.
We can extract a timescale from this formula, the hardening timescale:

$$t_h = \left|\frac{a}{\dot{a}}\right| = \frac{1}{2\pi G\xi}\frac{\sigma}{\rho}\frac{1}{a} \tag{75}$$

V. Sum up

Binary stars interact in star clusters with single intruders and binary interactions mean exchange of energy. The exchanges can even change the component of the binaries leading to more and more massive binaries. Hard binaries, which are already quite energetic in the star cluster when they form, tend to become more hard as the semi-major shrinks.

VI. Effects of dynamics on BBHs

We have seen that hard binaries harden as the result of 3-body encounters. So if we have a BBH whose semi-major axis is too large to merge via GW emission but the BBH is in the core of a dense star cluster repeated encounters with other stars will shrink the orbit till the binary will become sufficiently tight to merge via GW.
We can make a back-of-the-envelope calculation using the hardening and the GW timescale to evaluate the maximum value for the semi-major axis for GW to dominate ($t_h = t_{GW}$):

$$a_{GW} = \left[\frac{256}{5}\frac{G^2m_1m_2(m_1+m_2)\sigma}{2\pi\xi(1-e^2)^{7/2}c^5\rho}\right]^{1/5} \tag{76}$$

This value a signals the transition from the regime where 3-body encounters are more effective to the regime where GW emission is more effective in shrinking the binary. For $a < a_{GW}$ the binary stops being dominated by newtonian 3-body encounters and starts being dominated by GW emission. This depends both on the properties of the binary (mass and eccentricity) and on the properties of the cluster (density and velocity dispersion).
We can also use directly the expression for the evolution of the semi-major axis:

$$\frac{da}{dt} = -2\pi\xi\frac{G\rho}{\sigma}a^2 - \frac{64}{5}\frac{G^3m_1m_2(m_1+m_2)}{c^3(1-e^2)^{7/2}}\frac{1}{a^3} \tag{77}$$

Because of the very different dependence on a we will distinguish between the two regimes: the 3-body encounters regime is the same for all binaries because it only depends on the properties of the cluster, then it differs for GW emission depending on mass. The harder the binary the better this approximation works. A binary of two SMBHs is very hard and this approximation is pretty good. For a stellar BBH this approximation is less good, mainly because we can have exchanges.

With this approximation we can also estimate the number of encounters the binary has to undergo to enter the regime where GW are important. We just use the interaction rate.
The first process that does affect the evolution of a BBH in a star cluster is the hardening. The second are dynamical exchanges: suppose we have a binary system composed of a BH and a low mass companion, like a solar star. If this binary is in a field, it is unlikely that it will become a BBH, because the star is small, but if it is in a dense star cluster then it will have a chance to interact with other stars and BHs and a BH may kick out the star from the system and form a BBH. We expect to form more BBHs than to destroy them by exchange because BHs are in general more massive than other stars in a cluster. So BHs are more efficient in acquiring a companion via dynamical exchanges because they are more massive than other celestial bodies in a star cluster and exchanges tend to favor the formation of more and more massive binaries.
Exchanges are very efficient in forming BBHs: more than 90% of BBHs form via dynamical exchange in young star clusters. Exchanges also leave a strong imprint on the properties of BBHs. BBHs that form via exchange present the following features:

- They tend to be more massive
- At least at the beginning, they have a larger eccentricity, then GWs tend to circularize the binary
- We expect these dynamical encounters to completely randomize the spins. From binary evolution we have mentioned that if a BBH forms from binary evolution the spins tend to be aligned with the orbital angular momentum of the binary (only the SN can misalign them a little bit). But dynamical encounters tend to make the system to lose memory of the previous spin. So we expect isotropically oriented spins. The most efficient way to produce binary system with isotropically oriented spins (large χ_p) is via dynamics.

A nice feature that distinguish field and star cluster binaries is the masses. The chirp mass distribution is such that the low mass systems are not present in the star cluster. This is always due to exchanges which produce more massive BBHs.
The same feature is present also in young star clusters, where we compare isolated binary and dynamically formed BBHs.
The isolated systems because of the effect of mass transfer tend to lie close to the region of nearly equal mass. Dynamics can produce systems with small mass ratio, as GW190412.
While isolated BBHs form up to masses $40 - 45M_\odot$, dynamically formed ones reach larger masses. Where do the very massive BBHs come from? There is a combination of 2 factors:

- From single stellar evolution we can form BHs even up to $50 - 60M_\odot$.
- The more massive BBHs can form from single star evolution or from loose binary evolution. Otherwise if the binary is tight there will be mass transfer, the system will undergo a common envelope phase. All these processes will remove matter. In particular, the common envelope removes the entire envelope of the binary. Even if the binary survives the mass of the final BHs will be essentiallly given by the mass of the stellar cores, as the envelopes are lost. With dynamics, even we have a BH that forms with a $50M_\odot$, in a star cluster it can acquire a companion via exchange.

We have even systems with masses $80 - 90M_\odot$. There is a boost to the mass of a BH in a star cluster due to merger of stars. If we merge massive stars in very peculiar stages of their lives they can produce BHs even more massive than $60M_\odot$.

VII. Kozai-Lidov resonance

What about eccentricity? The eccentricity is high after an exchange. At 10Hz, when the binary enters the LIGO/Virgo band, this eccentricity is almost zero.
Nevertheless, there is an important process to take into account, which is GW capture, the possibility that two BHs

capture each other in a system because of GW emission. If include the GW capture process, then there are a few system with eccentricity close to 1 even in the LIGO/Virgo band.

A very peculiar process, called Kozai-Lidov, that takes place in both star clusters and fields, requires to have three bodies in a **hierarchical triple**, i.e. a stable triple where we have an inner binary orbited about by a third body. If the orbital plane of the inner binary is a little bit offset with respect to the orbital plane of the outer binary (a non zero angle), then we have the formation of Kozai-Lidov resonance, which leads the eccentricity of the inner binary and the inclination between the two orbital planes to oscillate in a surprisingly regular way.
The semi-major axis does change: the Kozai-Lidov effect does not change the energy of the system. This is a purely newtonian approach. Eccentricity oscillations triggers timescale changes in the GW timescale. So we can conclude that this is important for GW.

From observations, 25% of massive stars are in triples: so they cannot be neglected.

If we include some relativistic corrections (2.5 post-newtonian term), the semi-major axis at some point drops because the GW timescale also drops. Kozai-Lidov effect is very important because it predicts BBH mergers with extreme eccentricity even in the LIGO/Virgo band.

VIII. Formation of intermediate mass BHs

Dynamics is one of the main ways, if not the only one, that we have to form intermediate mass BHs (masses between $100 - 10^5 M_\odot$). This definition is not physically motivated, but just historical: these are BHs with a mass intermediate between stellar mass BHs and SMBHs.
They are very important because they may explain the formation of SMBHs, via merging of many intermediate mass BHs with each other. They are extremely elusive. There is at least one strong candidate: the merger product of GW190521. It's quite low mass, but in the regime of intermediate mass BHs. This also means that GWs are good probes of intermediate mass BHs.

There are candidates from electro-magnetic emission: the strongest one is HLX-1 (HyperLuminousX-ray source number 1). It is a source with a luminosity in excess of 10^{40}ergs, which cannot be associated to a SMBH. Nevertheless, it has a pick luminosity which is much higher, 10^{42}ergs. If we measure the light curve from X-rays, it has modulations. It cannot be a bunch of fainter X-ray sources that we dot distinguish, but a modulated single source. It has also a radio jet, which allows to constrain the mass of the BH, between $9000 - 90000 M_\odot$, way more massive than GW190521, but both are in the regime of intermediate mass BHs.

Another strong but less direct is a BH in a globular cluster in the Andromeda galaxy. We infer its presence by the mass to light ratio, not a robust evidence.

How can they form?

- **Runaway collision of stars**: requires dynamics
- **Repeated mergers of BHs**: requires dynamics
- **Remnants of very massive ($> 260 M_\odot$) extremely metal-poor stars**: we have already seen it with pair instability, it doesn't require dynamics. Still in order to form these monster stars is to merge massive stars, which can happen only in star clusters, and this means dynamics.
- **Low-mass end of SMBHs**: then the question becomes how SMBHs form, which may require dynamics.

viii.1 Runaway collisions

In a star cluster we have massive stars and low mass stars and we have a dynamical process, which is dynamical friction. Massive bodies which are evolving in an environment of light bodies tend to decrease their velocity because of a drag force exerted by lighter bodies, which is called **dynamical friction**. Because of dynamical friction the massive bodies sink to the center of the star cluster where they produce an excess of massive stars in the center of the cluster.
All these massive stars packed together in a fraction of a parsec means that the probability that they undergo close encounter is very high. These close encounters can very easily trigger collisions: so we have runaway collisions, a very efficient series of collisions, that produces a supermassive star. If this star is lucky enough it will collapse to an intermediate mass BH directly.

The main problem is that during the merger of two stars we lose mass and we need more hydrodynamical simulations to quantify how much mass, it can be a substantial fraction of the total mass. Or after the merger, the supermassive star will be radiation pressure dominated so stellar winds will be important. We expect a lot of mass loss, which may prevent the formation of an intermediate mass BH, and just leave a smaller BH.

viii.2 Repeated mergers

It is the only one for which we have observational confirmation, which is GW190521. The idea is that if we have two BHs that merge in the field, then they leave a single BH which will not find another companion. But if we merge them in a dense star cluster, and the merger remnant remains inside the star cluster, it will have a chance to find another companion to pair up with again dynamically.
If we can repeat this a sufficiently large number of times we can grow an intermediate mass BH. The big issue is retaining the BH, because after every dynamical encounter we have Newtonian recoil velocity which could be larger than the escape velocity from the cluster. Even worse, when the two BHs merge we will have a relativistic kick, i.e. asymmetries in the emission of GWs, which produce a kick that can give to the remnant a natal kick of hundreds up to thousand km/s, which is definitely larger than the escape velocity from most of the star clusters (only the nuclear star cluster have a large escape velocity and can retain a very large fraction of BHs).
We can calculate the properties for a BBH for which it avoids ejection by Newtonian kick. On the other side, from NR we can see the distribution of the relativistic kick (we go up to 3500km/s). In the most pessimistic case we have extreme kicks.

One additional problem, besides the kicks, Newtonian and relativistic, is the problem of efficiency: we can hardly build BHs this way. We need roughly 300Myr to build up an intermediate mass BH this way.
From simulations we have that both runaway collisions and repeated mergers take place in a globular cluster, but the runaway happen only at early times (we still need alive stars) and they are more efficient with respect to repeated mergers, which form later on and form less massive BHs.