3.5 Horizons in the Universe

Due to the speed of light being finite, there are regions of the Universe that are and will be unaccessible to us, even if we wait an infinite amount of time. These regions are defined by *horizons*. In cosmology, there are three of these boundaries, which define the causal structure of our Universe.
Before we continue, let's recall the FRW metric:

$$\begin{aligned} ds^2 &= -c^2dt^2 + a^2(t)\left(\frac{dr^2}{1-kr^2} + r^2d\Omega^2\right) \\ &= ds^2 = a^2(\eta)(-c^2d\eta^2 + (d\chi^2 + S_k^2(\chi)d\Omega^2)) \end{aligned} \tag{3.41}$$

with the conformal time defined by:

$$d\eta = \frac{dt}{a(t)} \tag{3.42}$$

and the spatial geometry:

$$S_k(\chi) \equiv \begin{cases} \sinh\chi & k=-1 \\ \chi & k=0 \\ \sin\chi & k=+1 \end{cases} \tag{3.43}$$

The first of these horizons is what is called the *particle horizon* or the *cosmological horizon*. It is the maximum distance light could have traveled from the "beginning" of the Universe, i.e. the observable Universe. To be precise, we could ask the following question: given a comoving observer at coordinates $(\chi_0, \theta_0, \phi_0)$, for what values of (χ, θ, ϕ) would a light signal emitted at time $t = 0$ (the "beginning" of the Universe) reach the observer at time t? Given the FRW metric, we can answer this question rather easily. First of all, let's place the receiving observer at a convenient point in space (given our freedom of choice granted to us by homogeneity), setting $\chi_0 = 0$ (so the observer is us, Earth). Then, a light signal satisfies the geodesics equation with $ds^2 = 0$ and $d\phi = d\theta = 0$ (because the light moves on great circles). If this light signal reaches us at time t, then the comoving distance from the observer (who emitted the photon at the "beginning" of the Universe) is given by:

$$d_{h,comoving}(\eta) = \chi = \eta = c\int_0^t \frac{dt'}{a(t')} \tag{3.44}$$

In other words, the value of conformal time now η_0 is the observable Universe today. We can also define the *proper particle horizon*:

$$d_h(\eta) = a(\eta)c\int_0^t \frac{dt}{a(t)} \tag{3.45}$$

Notice that we have implicitly assumed $t = 0$ to be the beginning of the Universe. This notion, however, is not really defined, since we don't actually know what happened really far in the past above 100 GeV, as remarked two lectures ago. To put it another way, the equation (3.44) should really be written as:

$$\eta = c\int_0^{t_{100\text{ Gev}}} \frac{dt'}{a(t')} + c\int_{t_{100\text{ Gev}}}^t \frac{dt'}{a(t')} \tag{3.46}$$

Because the Universe was really small and really hot, we have no reason to trust classical general relativity as such high energies, so that the definition of "observable Universe" may make not much sense at all. In fact, the name "observable Universe" itself is usually meant to be the distance light could have traveled from some time $t_{100\text{ Gev}}$, where extrapolating the classical theory is still sensible.

Another type of horizon is the *event horizon*, which differs from the particle horizon, in that it is the maximum distance light, emitted at given time t, could travel to the infinite future. Again, the FRW metric allows us to easily calculate this distance, proceeding in the following way. Given a comoving observer at coordinates $(0, \theta_0, \phi_0)$ (again, homogeneity permits to set $\chi_0 = 0$) that emits a light signal at time t, the distance it travels up to $t = \infty$ is:

$$d_{e,comoving}(\eta) = \int_{\eta}^{\eta_f} d\eta = \eta_f - \eta = c\int_t^{\infty} \frac{dt'}{a(t')} \tag{3.47}$$

with its *proper* counterpart being:

$$d_e(\eta) = a(\eta)c\int_t^{\infty} \frac{dt'}{a(t')} \tag{3.48}$$

After a moment's thought, one notices that this definition really is just the "reverse" of the one for the particle horizon, in the sense that if we exchanged the beginning with the end of the Universe, the two concepts would switch. This is shown more clearly in the following figure, which depicts the two horizons:

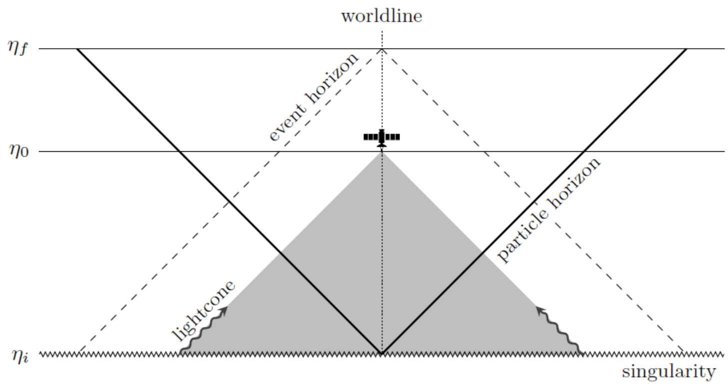

Figure 3: The particle and the event horizon, in conformal time.

We have, however, just glossed over some subtleties. For the moment, in fact, we have defined $\eta_i = \eta(t = 0) = 0$ and $\eta_f = \eta(t = \infty) \neq \infty$ (infinite time does not mean infinite conformal time). These two identities will be derived later on, where the dynamics of the Universe are discussed in detail. Moreover, notice the specularity of the definitions for $d_e(\eta)$ and $d_h(\eta)$: just as the particle horizon is sensitive to the initial conditions (the integral (3.45) goes down to $t = 0$), the event horizon is sensitive to the "final conditions" (the integral (3.47) goes until $t = \infty$). Practically, this is because we don't know what the fate of the Universe will be, this time not because of general relativity itself, but because the entity that controls its future is dark energy, whose nature is still really unresolved.

The final concept we linger on is the *Hubble horizon*. This is a confusing concept, for starters because it's not actually an horizon, so let's define it from the basics. As we have seen, Hubble's law dictates that the velocity of galaxies on the Hubble flow is proportional to their proper distance:

$$v_{\rm phys} = H(t) r_{\rm phys} \tag{3.49}$$

Replacing $v_{\rm phys}$ with the speed of light c, we find a proper distance d_{HR} (not comoving) above which galaxies (or objects in general) recede with a speed that's greater than that of light:

$$d_{HR} = \frac{c}{H(t)} \tag{3.50}$$

The *comoving Hubble horizon* is instead:

$$d_{HR} = \frac{c}{a(t)H(t)} \tag{3.51}$$

This is not in contrast with special relativity, as this velocity is not measured in any inertial frame. No observer is overtaking a light beam and locally, in the galaxy's position, observers measure the speed of light as c. Taking d_{HR}^3, we get what is called the *Hubble volume* or the *Hubble sphere*.
On the one hand, the Hubble sphere is not a measure of causality, because if two objects are separated by a distance greater than a Hubble length, it would be still possible for them to communicate. In fact, this is the case if the Universe is expanding and $\ddot{a} < 0$, i.e it is also *decelerating*, meaning that the Hubble sphere is actually increasing. After a while, the Hubble sphere will catch up with the light ray, which in turn will enter a zone where space is expanding slower than the speed of light, making it possible to freely flow to destination. If instead the Universe was expanding in an accelerating fashion, the Hubble sphere would decrease, and two objects separated by more than a Hubble length would forever be out of causal contact.
On the other hand, this discussion implies that the Hubble horizon has got at least something to do with causality, despite not really being an horizon. We can see how it is related to the particle horizon by rearranging the integral (3.44) starting at some time t_i:

$$d_h(\eta) = \int_{t_i}^{t} \frac{dt}{a} = \int_{a_i}^{a} \frac{da}{a\dot{a}} = \int_{\log a_i}^{\log a} (aH)^{-1} d\log a \tag{3.52}$$

Basically, the particle horizon is the logarithmic integral of the comoving Hubble horizon.

3.6 Distances in cosmology

The measure of distances in cosmology is obviously fundamental for our understanding of the Universe's properties. There are, however, a lot of subtleties that one needs to consider when trying to formulate rigorous definitions.

3.6.1 Luminosity distance

First, let's review Hubble's argument briefly, as it is really the crux of the whole topic. By looking at relatively close galaxies, Hubble found them to be slightly redshifted. At the time, the Universe was thought to be static, so he interpreted the redshift as being caused by the peculiar motion of galaxies. Estimating the velocity of galaxies as cz (relativistic redshift) and measuring their distances, he found that the two were proportional, i.e. $v = cz = H_0 d$.

We have seen that, by interpreting the galaxies' movement as caused by an expansion of spacetime, the FRW metric actually predicts Hubble's law, as we did in (3.6). There is a problem here though. Hubble couldn't possibly have measured the proper distance, as it is *not* an observable, because it is the distance between two objects at a *fixed time* (in fact, we derived Hubble's law from the FRW metric by setting $dt = 0$), i.e. the distance someone would measure by taking a snapshot of the Universe at a certain time. On the other hand, we have also rightly seen how Hubble's law is just an approximation for very close objects. So what was Hubble actually measuring? Clearly, there is something deeper going on here.

A useful way of measuring distances in cosmology uses the so-called *standard candles*. These are objects of known *intrinsic luminosity*, so that, by measuring their *observed luminosity*, we can infer their distance, which is called the *luminosity distance*.
Hubble used *Cepheids* to discover his law. These are stars whose brightnesses vary periodically, with periods being a known function of intrinsic luminosity of the stars themselves. Therefore, by measuring their periods, Hubble could infer their distance rather easily.
Let's formalize this argument. Let's say we have identified a certain source with a known intrinsic luminosity L. The observed flux F (energy per unit time per unit area) can then be used to deduce the luminosity distance. Consider a source at redshift z, with comoving distance being:

$$\chi(z) = c\int_{t_1}^{t_0} \frac{dt}{a(t)} = c\int_0^z \frac{dz}{H(z)} \tag{3.53}$$

We assume that the source emits light isotropically, so that in a Euclidian static space, the energy would be spread on the surface of a sphere of radius $4\pi\chi^2$ around the source. Then, the relation between the absolute luminosity and the observed flux is:

$$F = \frac{L}{4\pi\chi^2} \quad \text{static Euclidian space} \tag{3.54}$$

In a more complex expanding spacetime, this formula is modified in three fundamental ways:

- First, the radius of the sphere is not χ, but rather $S_k(\chi)$, as we have written in the FRW metric. The sphere surface is then $4\pi S_k(\chi)^2$. Furthermore, when light reaches Earth at time t_0, the sphere has an actual area stretched by $a^2(t_0)$, because of expansion. In total, the denominator of (3.54) becomes:
$$4\pi a^2(t_0) S_k(\chi)^2 \tag{3.55}$$
- The photons reach Earth with a decreased rate, again because of expansion. This effect reduced the observed flux F in (3.54) by a factor of $a(t_1)/a(t_0) = 1/(1+z)$.
- The observed photons lose an energy proportional to $1/(1+z)$ during their trip to Earth, reducing the observed flux once again.

For these reasons, the correct formula relating observed flux and the intrinsic luminosity of a source at redshift z is (remember that $a(t_0) = 1$):

$$F = \frac{L}{4\pi S_k(\chi)^2(1+z)^2} \equiv \frac{L}{4\pi d_L^2(z)} \tag{3.56}$$

where we have defined the luminosity distance as:

$$d_L(z) = (1+z)S_k(\chi(z)) \tag{3.57}$$

Now, we can actually find a perturbative expression for $d_L(z)$, by finding an expansion for $\chi(z)$. We first expand $a(t)$ for $z \ll 1$:

$$a(t) = 1 + H_0(t-t_0) - \frac{1}{2}q_0 H_0^2 (t-t_0)^2 + ... \tag{3.58}$$

where we have defined the *deceleration parameter* at present time q_0:

$$q_0 \equiv -\frac{\ddot{a}}{aH^2}\Big|_{t_0} \tag{3.59}$$

Given the definition of cosmological redshift, we have that:

$$z = \frac{1}{a(t_1)} - 1 = H_0(t_0 - t_1) + \frac{1}{2}(2+q_0)H_0^2(t_0-t_1)^2 + ... \tag{3.60}$$

By inverting it:

$$H_0(t_0 - t_1) = z - \frac{1}{2}(2+q_0)z^2 + ... \tag{3.61}$$

We can then rewrite the comoving distance making use of (3.58) and (3.61):

$$\begin{aligned} \chi(z) = c\int_{t_1}^{t_0} \frac{dt}{a(t)} &= c\int_{t_1}^{t_0} dt(1 - H_0(t-t_0) + ...) \\ &= c(t_0 - t_1) + \frac{1}{2}\frac{H_0}{c}c^2(t_1 - t_0)^2 + ... \\ &= \frac{c}{H_0}\left(z - \frac{1}{2}(1+q_0)z^2 + ...\right) \end{aligned} \tag{3.62}$$

Knowing the geometry of the Universe, by computing $S_k(\chi(z))$ with (3.62), we can estimate the luminosity distance of a source at redshift z. For example, if the Universe is spatially flat $k=0$, then $S_k(\chi) = \chi$, and:

$$d_L(z) = \frac{c}{H_0}\left(z + \frac{1}{2}(1-q_0)z^2 + ...\right) \tag{3.63}$$

We see that the first order approximation is $d_L(z) \simeq cz/H_0$, which is exactly Hubble's law. Now, however, we can understand it better. When doing the expansion in the last lecture, we found $d = cz/H_0$, so we can now identify that d with d_L. Not only that, but our expression (3.63) is only true for Euclidian space. We avoided this fact in the previous lecture by assuming that $d/c = t_0 - t_e$, which is exactly the relation for Euclidian space. In general, however, the Hubble's law will be true, at first order, for all geometries, since the curvature of space will not be a factor for nearby sources. Finally, remember how we found a *non perturbative* Hubble law in (3.6). It happens that, for nearby objects, light takes a relatively short time to travel, so that the luminosity distance can roughly be taken as an instantaneous measure, making it very similar to the proper distance. In other words, Hubble's law is only $d_L(z) \simeq cz/H_0$ (for nearby sources); the relationship $\dot{d}_{proper} = H(t)d_{proper}$ is, on the other hand, a perfect relationship. However, for close sources, light takes a short amount of time to travel, making the two concepts interchangeable.

Summarizing, given an observed flux of photons from a certain source on our telescope, if we happen to precisely know its intrinsic luminosity, we could deduce its luminosity distance using (3.56). The important question here clearly is: how would we know the intrinsic luminosity of a source? As we have said above, in general cosmologists use standard candles, objects for which empirical relations between their properties and their intrinsic luminosities are known. Hubble used Cepheids, but they only work for relatively close sources. Instead, for more distant ones, we need brighter standard candles. Remarkably, type Ia supernovae are exactly the right candidates. These objects are white dwarfs (degenerate stars that are made of oxygen or carbon, for which the electron degeneracy pressure is keeping the star from collapsing on itself) in a binary system. It turns out that white dwarfs have a critical mass, the *Chandraseckhar mass*, of about 1.4 $M_\odot$, above which the star will collapse and explode, creating a supernova event. This critical mass could be reached, for instance, if the white dwarf accretes matter from the companion in the binary. The crucial point here is that the explosion will happen at the Chandraseckhar mass independently of where the binary system is located in the Universe. Therefore, because the same physical process underlies the explosions, the intrinsic luminosities of type Ia supernovae are well known.

3.6.2 Angular diameter distance

An alternative way to measure the distance of a given source is by measuring its angular amplitude. The angular size is linked with the distance because, for instance in Euclidian static space, farther objects appear smaller.
Let's follow the same steps as before. Given a source at comoving distance $\chi(z)$ which has emitted a photon at time t_1, and with a transverse size of D, its angular size is, in a static Euclidian space:

$$\delta\theta = \frac{D}{\chi} \quad \text{static Euclidian space} \tag{3.64}$$

where we assumed $\delta\theta \ll 1$, which is true for very distant objects, like those we deal with in cosmology. If light travels in a radial way, then from the FRW metric it follows that the angular diameter of the source is:

$$\delta\theta = \frac{D}{a(t_1)S_k(\chi)} \equiv \frac{D}{d_A(z)} \tag{3.65}$$

where we have defined the *angular diameter distance*:

$$d_A(z) = \frac{S_k(\chi)}{1+z} \tag{3.66}$$

Notice the difference with the definition of the luminosity distance (3.57). Whereas for d_L we were interested in the moment of observation $a(t_0)$, we are now taking the time of emission $a(t_1)$. Furthermore, we can work out a perturbative expression for $d_A(z)$, just by noticing its relationship with $d_L(z)$:

$$d_A = \frac{d_L}{(1+z)^2} \tag{3.67}$$

Therefore:

$$d_A(z) = \frac{c}{H_0}\left(z - \frac{1}{2}(1+q_0)z^2 + ...\right) \tag{3.68}$$

Notice how, for really close sources, the first order term gives $d_L = d_A = cz/H_0$, i.e. Hubble's law. This, as remarked above, is true because spacetime is locally flat, and in Euclidian static space we know that there is only one distance, as our daily-life experience tells us. The Euclidian static limit also reveals something deep. A rightful question is in fact the following: given all these distances (comoving, proper, angular diameter, luminosity), which is the "correct" one? Clearly, this question only makes sense because we are used to the Euclidian static limit, where all distances coincide. However, if space is curved and expanding, different definitions yield different results for very distant objects.

We have completed the overview of the properties of the FRW metric. Noticing that most of these actually depend on the function $a(t)$, the *dynamics* of the Universe, we now turn to the Einstein equations.