Contents

Part I
Inflation

1 Length scales and horizons

1.1 Particle and event horizons

In general, horizon is a surface that you cannot see beyond. Particle horizon is a sphere drawn around an observer, such that the light from sources on that sphere, if emitted at $t = 0$ will just reach the observer now, at t_0. Since the null geodesic tells us that $dt = a(t)dr/\sqrt{1 - r^2/R^2}$, the coordinate radius of that sphere is

$$d_p(t) = \int_0^t \frac{dt}{a(t)} = \int_0^r \frac{dr}{(1 - r^2/R^2)}. \tag{1}$$

The scale factor in the denominator of the middle portion of eq. 1 says that during a fixed dt light covers larger coordinate distances in the past, when a was small, and smaller coordinate distances at later times. (The amount of proper distance covered during a fixed dt is the same at all times, as it should be.) The proper radius corresponding to eq. 1 is $D_p = a(t)\,d_p$, and is called the (proper) particle horizon.

Consider a case where the equation of state of the fluid filling the Universe is determined by w: $p = w\rho$, and w's value is not restricted,

$$d_p(t) = \int_0^t \frac{dt}{a(t)} = \int_0^a \frac{da}{\dot{a}a} \propto \int_0^a \frac{da}{[a^2(1 - \Omega_w) + \Omega_w a^{(1-3w)}]^{1/2}} \tag{2}$$

Depending on the value of w, and the presence/absence of curvature, one of the two terms in the dominator will dominate at early times (small a). If $w > -1/3$ the second of the two terms (the one that scales as $a^{(1-3w)}$) dominates and the integral $d_p(t)$ converges, hence the distance to the particle horizon is finite, which means the horizon exists, and our view of distant things is limited. If $w < -1/3$ the first of the two terms (curvature) dominates, so the integral does not converge, instead, it diverges on the small a side, i.e. there is no particle horizon and we, as observers, could see all of the Universe at some time in the past, that means everything was in causal contact at very early times.

If $a(t) \sim t^{2/[3(1+w)]}$, as we had for an arbitrary component of the Universe with $w > -1/3$, the particle horizon is finite and equal to $D_p = [3(1 + w)/(1 + 3w)]\ t$. For matter and radiation dominated cases, this translates into $D_p = 3t$ and $D_p = 2t$, respectively. So even during the early epochs in Universe's history, the expansion in a standard, non-inflationary cosmology (i.e. matter or radiation dominated) ensures that you can see only a limited portion of the Universe around you. This region, D_p, grows with time, and in an non-inflationary Universe it will grow *faster* than the rate at which two observers are carried away from each other by the Hubble expansion. Because of that, all new stuff (particle, galaxies,...) that comes into your view as time goes by has never been seen by you before. In other words, you are seeing stuff that you have never had a chance to establish causal contact with.

Event horizon asks the question complimentary to that answered by the particle horizon: How far into the future can light emitted by me reach? The calculation is the same as in eq. 1, except that we are now interested in the future, so the limits of integration are from the present time, $t = t_0$, to $t = \infty$. Event horizon can be calculated for various cosmological models. Event horizon is also important in black hole physics ("will my flash light signal ever reach the outside world, so I can get help?"), and so is the more well known of the two types of horizons.

1.2 Hubble scale/horizon/length

Universe is always expanding and hence its physical characteristics are constantly changing. However, on short time-scales these changes are small, even negligible. What is the largest time-scale (and hence length-scale) on which we can claim the Universe does not change appreciably? A convenient order-of-magnitude definition is provided by setting the fractional change in scale factor to 1, $da = a$. How long does that take? Remember the definition of the Hubble parameter: $H = \frac{\dot{a}}{a} = (da/a)dt^{-1}$, so the time during which the Universe changes appreciably is just H^{-1}, the Hubble time, the characteristic expansion time scale. The size of the "observable Universe" is then roughly c/H.

At any given epoch H^{-1} defines the physical size of the region that could have been in 'relatively good' causal contact with itself. We can send a signal out to somebody at a distance H^{-1} and receive a reply before the Universe's size changed too much. So, one can have microphysics go on inside such a region, like gas could come to a common temperature, pressure waves can propagate, heat conduction can carry energy across the whole Hubble length, etc.

The Hubble radius and particle horizon are not the same: If two objects are within each other's particle horizons they will forever 'retain the memory' of that. It is not so for the Hubble horizon: objects can go in and out of Hubble radius, depending on how fast the Universe is expanding, i.e. the value of H.

As we saw above, if the scale factor is a power law in time, as it is for most of the evolution of the Universe, then the particle horizon size is a multiple of H^{-1}, i.e. same as the Hubble radius, to within a numerical factor of order 1. So for most of the evolution of the Universe, i.e. during the standard Big Bang evolution, the two concepts are more or less interchangeable. This breaks down dramatically when the scale factor grows not as a power law of time, but, say, exponentially, $a \sim e^{Ht}$, which is the case during inflation.

2 Problems with the Standard Big Bang Evolution, and how to solve them

2.1 Problems

Overall, standard cosmology must be on the right track: its successes include nucleosynthesis microwave background, large scale structure, agreement of various estimates of age of the Universe, agreement between H_0 derived by very different observational methods, etc. As successful of the standard model of cosmology is, it has a few problems.

Flatness: According to observations, the current density in all the components, matter, radiation, cosmological constant is $\Omega_{tot} = \sum_w \Omega_w$ is of order 1 today. (This excludes the curvature term, which is $\Omega_K = 1 - \sum_w \Omega_w \approx 0$ today.) As we will see below Ω_{tot} rapidly evolves away from 1 as the Universe expands, so there is no reason why it should be close to 1 today. Given that our Universe is very old (the ratio of scale factors now vs., say, at nucleosynthesis, is enormous), it is surprising that the density is so close to $\Omega_{tot} = 1$.

Horizon/Smoothness: The size of the particle horizon, in comoving coordinates, is about 150 Mpc at the time of recombination and subtends an angle of $\sim 2^\circ$ on the sky. So any two points on the observed CMB today separated by more than a couple of degrees were not in causal contact at the time of recombination, and there is no reason why they should look the same. Contrast that with the observed smoothness of the CMB across the whole sky: the fractional temperature fluctuations are within 10^{-5} of average, and the spectrum of the fluctuations is the same across the sky—not only do all the $\sim 2^\circ$ patches agree on what temperature to have, they also agree on how, in a statistical sense, and by how much to deviate from the average temperature. The same type of reasoning applies to a number of other observations, for example we could have used the size of the horizon at the time of nucleosynthesis, as apposed to CMB. When nucleosynthesis took place the comoving size of the Universe was about the size of a small present day galaxy. And yet the abundances of light elements—the products of the nucleosynthesis epoch—are the same everywhere we look, in our galaxy and well beyond. This amazing agreement between different parts of our currently observable Universe require that these different parts establish causal contact in the distant past, when they would have agreed on the initial conditions. Such agreement is not possible within the standard big bang (SBB) cosmology, because of the existence of the particle horizon in cosmologies dominated by matter or radiation (or any $w > -1/3$) all the way to the earliest epochs of the Universe. So SBB dynamical evolution cannot possibly hold throughout the past epochs, and non-standard evolution, one that ensures a very very large particle horizon early on.

Magnetic Monopoles: A generic prediction of the GUT phase transition is the production of monopoles. One expects to have of order of one monopole per Hubble horizon at the time of GUT transition (because Hubble horizon is the largest causally connected region). Being non-relativistic, monopole number decreases as a^{-3} and therefore they come to dominate over the radiation density, which drops faster in the early Universe, as a^{-4}. As a consequence, today's Universe should be filled with monopoles, with the space density not too different from that of baryons. (There might have been other phase transitions, associated with supersymmetry, etc. Those transitions could have also produced relics of their own.)

Structure in the Universe: The present day galaxies, clusters of galaxies, etc. grew from initially small density perturbations, through gravity. So, some physical mechanism must have produced these early-epoch 'seed' fluctuations. But that's not all. The present-day observed characteristics of the spatial clustering of galaxies, etc. define a *specific spectrum* and *amplitude* of density perturbations that must have populated the early Universe. In other words, to get today's distribution of galaxies it is not enough to have just any type of distribution of early density fluctuations. A specific power spectrum and amplitude of density fluctuations is needed as a 'seed' field of future structure. What physical mechanism could have given rise to these fluctuations? Topological defects could have, in principle, seeded the structure, but the amplitude of the resulting fluctuations would have been very different from what is observed. We will see that inflation produces just the right power spectrum.

One could argue that not all of the above problems are of the same degree of severity. In particular, there are problems and there are fine-tunings. Horizon problem is a real problem: one absolutely needs to find a way to bring the different regions of today's Universe into causal contact in the distant past, and one absolutely needs to get rid of the monopoles, if GUT-like theories are correct. The flatness aspect of the Universe is a fine-tuning problem: there is no physical law that prevents Ω_{tot} from being very very close to unity. In particular, one cannot absolutely rule out that Ω_{tot} just happened to be exactly 1 in the distant past, in which case it would have remained exactly 1 to the present day. (But why would it be *exactly* 1...? Wish we had a sample of universes to look at...)

2.2 Solutions

The Friedman equation can be rewritten as

$$1 - \sum \Omega_t = \frac{c^2}{(aHR)^2} = \frac{c^2}{(R\dot{a})^2}, \tag{3}$$

where Ω_t is the contribution of the usual w components, matter, radiation and cosmological constant, but evaluated not at present, but at some other cosmic time t. The only 'component' on the right hand side is the curvature. $Ha = \dot{a}$ and is always decreasing in standard BB evolution because of the retarding influence of matter and radiation on the expansion. (Λ does not become important until much later, and is irrelevant at early times.) For example, in the matter dominated case, $H^2 \sim a^{-3}$, and so $H^2a^2 \sim a^{-1}$. If radiation dominates, $H^2 \sim a^{-4}$, and so $H^2a^2 \sim a^{-2}$. The Universe has doubled in size a very large number of times since the early epochs, and so the evolution of curvature term even as slow as a^{-1} would drive the curvature contribution (i.e. the RHS) very far from 0 during the lifetime of the Universe. Since we observe a very small curvature today, $(1 - \sum_w \Omega_w) < 10^{-5}$, the curvature term must have started out being truly tiny. It seems the curvature could have picked any initial value it wanted, so why did it settle on a value so close to 0? This is the flatness problem.

Here's another way of thinking about it. The RHS of eq. 3 contains a ratio of two relevant length scales: the Hubble scale, H^{-1} and the curvature scale, $a|R^2|^{-1/2}$, the scale below which curvature effects are small. (R^2 is constant for any given Universe model.) Equation 3 says that the ratio of Hubble length to curvature scale length is always increasing during SBB, so the curvature of the Universe should become more prominent with time, whereas we seem to detect no curvature at this late stage of the evolution of the Universe.

Also note that during SBB evolution, the scale factor grows slower with time ($a \sim t^{2/[3(1+w)]}$) than H^{-1} ($H^{-1} \sim t$) which means that the size of the "observable" Universe, as measured in comoving coordinates, is increasing: If you are an observer, then as time goes by, you see more and more galaxies come into your view, galaxies that were never visible to you before. (The implicit assumption here is that galaxies are fixed in comoving coordinates.) This is a restatement of the horizon problem.

If we want to correct the flatness problem, the general nature of solution presents itself: make the evolution of H^2a^2 with time (or a) such that $1 - \sum_w \Omega_w$ is driven towards zero; i.e. make $Ha = \dot{a}$ get smaller in the past, or equivalently, make $\ddot{a} > 0$. In other words, make the Universe do just the opposite of what it does during the SBB evolution. In such a situation the Hubble length will grow slower than the curvature scale and the scale factor, and as an observer you will count less and less galaxies around you as time goes by; during this stage of evolution galaxies will be disappearing beyond your (future) event horizon.

To solve the flatness problem we need to have an accelerating Universe, which is obtained if $w < -1/3$ (see the Friedmann eqn.) so, among others, cosmological constant-type components ($w = -1$) will work for this period of the evolution.

Recall our earlier discussion of particle horizon. We noticed that if a universe is dominated by a component with $w < -1/3$ then there is no particle horizon in such a universe, so that all the material was in causal contact at some point in the very distant past. That is exactly what we need to solve the smoothness problem: if all of the CMB $\sim 2°$ patches were in causal contact with one another long before the production of the CMB then the same physical constants/laws, as well as the amplitude and spectrum of fluctuations could have been established throughout that whole region. *Thus $w < -1/3$ solves flatness as well as smoothness problems.*

3 The inflationary epoch

3.1 The big picture

Let us assemble a general picture of what the early history of the Universe should be like. Very very early on, before the inflationary period commences there were small causally connected patches in the Universe. Their size was $1/H_i$, and i stands for 'just before inflation'. Causal connection allows the whole patch to agreed upon the common values for the physical constants, like G, α, etc, and get thermalized. Then there was a period of accelerated expansion that was driven by a component with $w < -1/3$. This component blows a small patch (we can only talk with any certainty about *our* patch) to a size much much larger than the Hubble length at the time. During this period we see comoving observers move out of our Hubble scale. The density of monopoles is diluted. Flatness and homogeneity within this patch are ensured. This inflationary period cannot last forever and must end before baryogenesis, nucleosynthesis, and CMB production. After inflation ends, the Hubble length resumes its rapid (compared to the scale factor) growth, and so more and more comoving volume comes into our Hubble horizon.

To summarize, to solve flatness, smoothness and monopole problems we have proposed a period of inflation (just a general idea so far, no specifics): One needs to have a $w < -1/3$ component. After the inflation, having done its job, this component must decay away. Note: the present-day state of accelerated expansion may or not have anything to do with the field that gave rise to inflation. Even though inflation and current accelerated expansion probably arise from different physical causes, the dynamical evolution characterizing the two epochs is very similar.

3.2 The main classes of inflation theories

Old Inflation:
This is Guth's (1981) original proposal. The early Universe is stuck in a state of false vacuum, whose energy density is much greater than that of the true vacuum. But the Universe cannot classically evolve to a state of true vacuum because there is a potential barrier that's in the way. In other words, the Universe must undergo a first order phase transition for which latent heat is required. But this energy is not available, and so the Universe, or more correctly parts of it must quantum tunnel to the new, lower energy state, the true vacuum. Quantum tunneling is required to circumvent the barrier. Because this is a quantum mechanical process it cannot take place simultaneously over macroscopic regions of space. Small portions the Universe make the transition independently, creating "bubbles" of the new, lower energy state. Unfortunately, these bubbles can never get together in large enough numbers to make a large enough universe, because the bubbles are being continuously driven apart by exponential expansion. Thus, this inflation never comes to an end; old inflation suffers from a "graceful exit problem".

New Inflation:
This version was suggested soon after Guth's original paper came out. It proposes that the shape of the inflaton potential does not have any barriers in it (in other words, the transition is second order, and no latent heat is required.) However, new inflation suffers from a severe fine-tuning problem: the shape of the $V(\phi)$ potential must be rather flat, or else inflation will not last long enough. The exact mechanism that could give us this flat potential is yet to be spelled out.

Chaotic Inflation:
In this version of inflation various regions of the early Universe had their own shapes of potential,

and because of that some regions did not inflate at all, some inflated a little, etc. Given a large diversity of possibilities, one of the regions would inflate for a long time, long enough to explain our region of the Universe. Other regions, for example, started out with a net positive spatial curvature and a short inflationary period, or no inflation at all; those regions would have collapsed back within Planck time. This version of inflation has no latent heat, no topological defects, and no fine-tuning problems.

There are many many more versions of inflation, which have postulated various shapes for the inflaton potential, based on a variety of premises.

3.3 The mechanism of inflation

Suppose there existed a field ϕ; call it "inflaton field". Let it be a scalar field, associated with spinless bosons. The field could have been a product of some early phase transition associated with some spontaneous symmetry breaking—or not. As of now there is no agreed-upon and accepted link between any specific unified GUT-type theory and the inflaton. There may not be any. Scalar fields can be thought of as a coherent superposition of ϕ-particles with zero spin angular momentum. The Universe is filled with this field at early epochs; each point of space has some value ϕ, and hence the corresponding $V(\phi)$; ϕ can be a function of spatial position, but even if spatial gradients did exist they would have been made small by the expansion, so we will ignore them altogether. Thus we assume that a single value of ϕ characterizes the Universe at any given epoch. The energy density of the field is the sum of its kinetic and potential energy densities. Let the corresponding potential be $V(\phi)$. The density and pressure are given by

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi) \quad \text{and} \quad p_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi) \tag{4}$$

When we plug these into the Friedman and continuity equations we get (dots represent derivatives with respect to time and primes are derivatives with respect to ϕ):

$$H^2 \propto V(\phi) + \frac{1}{2}\dot{\phi}^2 \tag{5}$$

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0 \tag{6}$$

respectively. In the first equation we have ignored the curvature term for simplicity; we anticipated the outcome of one of the main features of inflation: that the curvature within the Hubble horizon will become negligible soon after the start of inflation. The second one of the pair is the equation of motion. It is just the same as an equation of a ball rolling down a hill, with friction. In that analogy ϕ is the position coordinate of the ball on the hill, and $V(\phi)$ is the "gravitational potential". As usual, the friction term, $3H\dot{\phi}$, is proportional to the velocity of the ball, $\dot{\phi}$. "Friction" here is provided by the expansion of the Universe.

We want the inflationary period to last for a long time, since we need lots of increase in the scale factor. Therefore ϕ has to change slowly with time, and so $\ddot{\phi}$ should be negligible compared to $\dot{\phi}$ during this period. In the hill/ball analogy that corresponds to the ball hardly moving, but moving nonetheless, down towards the bottom of the hill, with small positive velocity, $\dot{\phi} > 0$, but with hardly any acceleration, $\ddot{\phi}$.

Using eq. 4, the definition of the equation of state parameter, $w = p/\rho$, and the fact that $w < -1/3$ for inflation, we get $\dot{\phi}^2 < V$. In fact let us assume that $\dot{\phi}^2 \ll V$, in which case $w \approx -1$. This is known as the slow-roll approximation:

$$H^2 \sim V \quad \text{and} \quad 3H\dot{\phi} \approx -V' \tag{7}$$

Slow roll conditions can also be expresses as follows,

$$\epsilon = \frac{M_{pl}^2}{16\pi}\Big(\frac{V'}{V}\Big)^2 << 1, \qquad \eta = \frac{M_{pl}^2}{8\pi}\Big(\frac{V''}{V}\Big)^2 << 1, \tag{8}$$

where we have used 'natural' units, $M_{pl} = 1/\sqrt{G}$, and M_{pl} is the Planck mass. the two parameters, ϵ and η quantify the slope and curvature of the shape of the $V(\phi)$ potential. The slow roll conditions say that for inflation to last long and produce nearly exponential growth of the scale factor, the shape of the potential, as quantified by its slope, V'/V, and curvature, V''/V must be small.

Since the potential is nearly flat, i.e. its value stays roughly constant during the period of inflation, H is also nearly constant, so $\dot{a} \sim a$, and

$$a(t) = e^{Ht}; \quad H = \text{const.} \tag{9}$$

Dynamically, this is the same as the cosmological constant-dominated Friedman equation, i.e. the de Sitter solution. So during the epoch of slow roll inflation we have exponential growth of the scale factor, accelerated Hubble expansion, and a nearly constant Hubble parameter. Notice, however, that even though the dynamics are similar, vacuum energy density that supposedly drives today's accelerated expansion could not have caused inflation. Today's expansion is characterized by $a \sim e^{H_\Lambda t}$, while inflation had $a \sim e^{H_I t}$, but the corresponding Hubble parameters, H_Λ and H_I are many orders of magnitude different, with $H_\Lambda \ll H_I$. As a consequence, inflation achieved many e-foldings increase in the scale factor in less than 10^{-35} seconds, while the effects of today's exponential expansion can be discerned only if we observe SNIa's at $z \approx 1$, i.e. objects whose light took billions of years to get to us. Another way to compare inflation and today's cosmological constant driven expansion is to compare the energy densities contained in corresponding field. At the time of inflation the energy density in what we call "cosmological constant" was minute compared to the energy density in the inflaton field. Finally, inflation had to come to an end, whereas vacuum energy density (if that is in fact the correct identity of the cosmological constant) can never go away, and the current accelerated expansion will continue forever.

3.4 How much to inflate? Number of e-foldings

Potential must be relatively flat, V' small, so that inflation lasts for long enough time to (many enough number of e-foldings) to make sure that all of the presently observable Universe was in causal contact at some point in the past so that all the $\sim 2^\circ$ CMB patches must agree on details, etc.

We require that inflation establish causality for all the observable Universe. The spatial scale equal to the horizon at the beginning of the inflation was the first scale to leave the horizon as the inflation commenced. Because the scales reenter the horizon in reverse order, this scale is the one reentering the horizon right now (last one, as far as we are concerned, but there could be many more of course). So we require that the comoving size of the largest scale to participate in inflation must be much larger than the horizon today, i.e. $r_i \gg r_0$, with the 'much much greater' requirement guaranteeing that we do not live at special time and there are many many more inflation-processed scales out there that have not yet come into our horizon. Approximately, the proper size of the Hubble radius at any time is $ar \sim 1/H$, so the following must hold:

$$H_i a_i \ll H_0 a_0. \tag{10}$$

From Friedman equation we know that

$$Ha \sim a^{-(1+3w)/2} \tag{11}$$

Rewriting eq. 10,

$$\frac{H_i a_i}{H_f a_f} \ll \frac{H_0 a_0}{H_f a_f} = \frac{H_0 a_0}{H_{eq} a_{eq}} \frac{H_{eq} a_{eq}}{H_f a_f}. \tag{12}$$

Substituting from eq. 11

$$\left(\frac{a_f}{a_i}\right)^{-(1+3w)} \gg \left(\frac{a_0}{a_{eq}}\right)\left(\frac{a_{eq}}{a_f}\right)^2 \approx z_{eq}\left(\frac{T_{Planck}}{T_{eq}}\right)^2 \left(\frac{T_f}{T_{Planck}}\right)^2, \tag{13}$$

where between the first and the middle part of the equation we have plugged in the corresponding w factors for the two epochs. $T_{eq} \sim 10^5$K, and $T_{Planck} \sim 10^{32}$K; $T_f/T_{planck} \sim 10^{-5} - 1$, and $z_{eq} \sim 25000$, so $a_f/a_i \sim 10^{26}$, corresponding to about 60 e-foldings.

Large number of e-foldings is also necessary to make sure that the Universe is spatially flat. The above criterion turns out to be sufficient for that purpose as well.

Flatness of the inflaton potential $V(\phi)$ is also required to achieve appropriate shape of mass density fluctuations, as discussed towards the end of Section 4.6. It turns out that this is actually a more stringent condition compared to the number of e-foldings; in other words if the amplitude of fluctuations is correct then the sufficient number of e-foldings is automatically achieved.

3.5 Just after the inflation: reheating/preheating

During inflation all conventional forms of matter and radiation, if they existed before inflation, are rapidly redshifted (or, diluted) away, as a^{-3} or a^{-4}, whereas the energy density in the inflaton field stays the same if $w = -1$. Even in the "worse case scenario", where $w = -1/3$, the bare minimum needed to cure SBB's problems, the energy density is diluted as a^{-2}. So at the end of inflation the potential energy in the inflaton field is by far the most dominant energy component.

When ϕ reaches the value where $V(\phi)$ attains a minimum, inflation comes to an end. At this point the field ϕ rolls to be bottom of the V hill, and then oscillates about the minimum, just as you would expect a ball to do. The energy in the inflaton field is converted into radiation. Just as with many aspects of inflation, the exact mechanism is yet to be settled:
Reheating: As the field oscillates a friction term, $\Gamma\dot{\phi}$, transfers the energy of the inflaton field into bosons, like photons, and other types of bosons. The friction term, which can be made to operate only during this phase in the evolution is small and the transfer of energy is slow and inefficient. In effect, this mechanism assumes that individual inflatons decay independently.
Preheating: The oscillations transfer energy through a parametric resonance, and so the transfer is fast and efficient. Here, the inflaton bosons act coherently, leading to oscillations of the inflaton field. The energy of the inflaton is rapidly transferred to other boson fields.

In either case, the Universe 'heat ups', i.e. undergoes an injection of entropy, and becomes dominated by relativistic matter. The reheated temperature must be high enough, so that the entropy, "heat" produced at the end of inflation is sufficient for baryogenesis to take place. But the final temperature must not be as high at the GUT epoch temperature, otherwise the magnetic monopoles will be produced again.

From now on standard evolution takes over, and the scale factor a increases as t raised to some positive power.

4 The spectrum of density perturbations

So far in the course we have been assuming that the Universe is completely smooth, with matter and radiation distributed perfectly uniformly. From now on we will often be dealing with the inhomogeneities, or fluctuations in the distribution of matter. The most immediate need to consider these arises from our discussion of inflation. Inflation naturally produces the spectrum of density fluctuations that we need to explain present day observations of galaxy clustering and the like.

4.1 A brief introduction to power spectra of density fluctuations

A reasonably continuous density fluctuation field can be decomposed into Fourier modes, i.e. sines and cosines, specified by their wave number, k, or wavelength $\lambda = 2\pi/k$. Then the density excess/deficit at point $\vec{r}$ in space is given by,

$$\frac{\delta\rho}{\rho}(\vec{r}) = \frac{\rho(\vec{r}) - \rho_{\rm ave}}{\rho_{\rm ave}} \sim \sum \delta_{\vec{k}} \exp(-i\vec{k}\cdot\vec{r}) \sim \int_{-\infty}^{\infty} \delta_{\vec{k}} \exp(-i\vec{k}\cdot\vec{r})\, d^3k, \tag{14}$$

where all length scales are comoving. The real space and the Fourier space are both 3D and extend to infinity in all 3 dimensions, and $\vec{k}$ and $\vec{x}$ are real. The exponential factor $e^{-i\vec{k}\cdot\vec{x}}$ is complex, and so is $\delta_{\vec{k}}$, but since density and $\delta(\vec{x})$ have to be real there is a condition imposed on $\delta_{\vec{k}}$, the so-called reality condition. For the density field to be real, the imaginary part of eq. 14 must always be zero. In the following we treat negative and positive portions of $\vec{k}$ space separately, so $\vec{k}$ is non-negative:

$$Im(\delta_{\vec{k}})\ \cos(\vec{k}\cdot\vec{x}) - Re(\delta_{\vec{k}})\ \sin(\vec{k}\cdot\vec{x}) + Im(\delta_{-\vec{k}})\ \cos(-\vec{k}\cdot\vec{x}) - Re(\delta_{-\vec{k}})\ \sin(-\vec{k}\cdot\vec{x}) = 0, \tag{15}$$

or,

$$\left.\begin{array}{r} Im(\delta_{\vec{k}}) = -Im(\delta_{-\vec{k}}) \\ Re(\delta_{\vec{k}}) = +Re(\delta_{-\vec{k}}) \end{array}\right\} \quad \Rightarrow \quad \delta^{\dagger}_{\vec{k}} = \delta_{-\vec{k}},$$

where $\dagger$ denotes complex conjugate. This is the reality condition.

The cosmological principle implies that space is statistically the same in all directions, and hence the same is true of $\vec{k}$ space. Therefore we can simplify the situation considerably by assuming that δ's are functions of k, and not $\vec{k}$. The power spectrum is then a record of the squares of the amplitudes for each k, $P(k) = |\delta_k|^2$.

A quantity of fundamental observational importance (because it has a clear physical meaning) is the rms fluctuations in density filtered (or windowed) on a given comoving scale λ,

$$\Big\langle\Big|\frac{\delta\rho}{\rho}\Big|^2\Big\rangle_\lambda \sim \Big\langle\int_0^\infty |\frac{\delta\rho}{\rho}|^2\, |W_x|^2 x^2 dx\Big\rangle_{x-{\rm space}} \sim \int_0^\infty |\delta_k|^2 |W_k|^2 k^2 dk. \tag{16}$$

$W_\lambda(r)$ is the spatial filter, and defines the size and shape of the regions we survey to get $\Big\langle\Big|\frac{\delta\rho}{\rho}\Big|^2\Big\rangle_\lambda$. $W_k(k)$ is the Fourier transform of the spatial filter. Two commonly used shapes for $W_\lambda(r)$ are top-hat (the 'ice-cream scoop'), and Gaussian, which weights regions at larger r less than the central regions. The gentler shape of the Gaussian filter is preferable because its Fourier transform does not have the high frequency (large k) modes. The sharp edges of the top-hat filter imply that in k space large k modes contribute significantly, making the convergence of the integral in k-space (right most part of eq. 16) difficult. The quantity $\Big\langle\Big|\frac{\delta\rho}{\rho}\Big|^2\Big\rangle$ is the rms dispersion in the average density excess in spheres of radius λ and having some shape $W_\lambda(r)$. The fractional density excess is equal to the fractional excess in mass, therefore

$$\Big\langle\Big|\frac{\delta\rho}{\rho}\Big|^2\Big\rangle = \Big\langle\Big|\frac{\delta M}{M}\Big|^2\Big\rangle \equiv \sigma_M^2 \tag{17}$$

Eq. 16 can be integrated; it is most easily done for the Gaussian window. The result is,

$$\Big\langle\Big|\frac{\delta\rho}{\rho}\Big|^2\Big\rangle_\lambda \simeq \frac{1}{2}\Big[\frac{(n+1)}{2}\Big]!\ k_G^3|\delta_{k_G}|^2, \tag{18}$$

and k_G is width of the Gaussian window in k-space; $k_G x_G = 2\pi$. For other filter choices the integral will still be proportional to $k^3|\delta_k|^2$ but the constant upfront will be different.

Let us assume that $P(k)$ is a power law, i.e. $P(k) = |\delta_k|^2 = Ak^n$. Power laws are common in nature, and for good reason: many phenomena are scale-free for many orders of magnitude. It turns out that it is an appropriate choice here as well, as observations show. The constant A, the normalization of the power spectrum can be obtained from observations. For example, one can measure σ_8, which is σ_M (see eq. 17) obtained using a top-hat filter of radius $r = 8h^{-1}$Mpc.

Using eq. 18 and the definition of the power spectrum, $\left\langle\left|\frac{\delta\rho}{\rho}\right|^2\right\rangle_\lambda \sim k^{n+3}$, where again $\lambda \sim 2\pi/k$. A comoving scale λ encloses a certain average mass, $M \sim \lambda^3$

$$\left\langle\left|\frac{\delta\rho}{\rho}\right|^2\right\rangle_\lambda^{1/2} = \sigma_M \sim M^{-n/6-1/2}. \tag{19}$$

Let us sum up the main results and the notation used so far:

$$P(k) \sim k^n; \qquad \sigma_M \sim M^{-\alpha}; \quad \alpha = n/6 + 1/2. \tag{20}$$

The last equation is a relation between two power law exponents: n is the power law index of the power spectrum, and α is the power law index specifying how rms fluctuations in mass (or density) scale with the mass scale. Here, M is used in lieu of a comoving scale.

When the amplitude of fluctuations is small, as is the case in the early Universe, the density fluctuation field can be decomposed into its constituent Fourier modes (sines and cosines) and each mode can be considered independently of all other modes. This is what we will do in most of what follows. As we will see, the growth rate of a mode is independent of its wavenumber, k, therefore, the growth rate of any and all density fluctuations, which consist of many k-modes, depends on global things, like average density, scale factor of the Universe at the time and the like.

4.2 The growth rate of density fluctuations on super-horizon scales

When a given comoving scale is outside the Hubble horizon (i.e. larger than the horizon) its evolution is not influenced by microphysical processes, like fluid pressure, so the only thing to consider is gravity, i.e. GR. We can use the Friedman equation which describes the global evolution of the Universe, because it is derived from Einstein Field equation, and so is fully relativistic. The question we ask is: if a given super-horizon scale happens to be slightly overdense or underdense compared to average density, $\bar{\rho}$, will the amplitude of that fractional deviation from the average density grow with time, diminish with time, or stay the same?

For simplicity, assume that the background geometry is flat, which is not an unreasonable assumption for our Universe. A piece of the Universe can be described by its value of ρ, which will be higher/lower than average for overdense/underdense regions. The Friedman equations for the background (i.e. average) and the overdense pieces are,

$$H^2 = \frac{8\pi G\bar{\rho}}{3} \quad (K=0) \quad \text{and} \quad H^2 = \frac{8\pi G\rho}{3} - \frac{K}{a^2} \quad (K \neq 0, \text{ but small}). \tag{21}$$

Comparing these two patches when their expansion rates are the same we get,

$$\delta\rho/\rho = \frac{\rho - \bar{\rho}}{\bar{\rho}} = \frac{K/a^2}{8\pi G\bar{\rho}/3} \sim (a^2\bar{\rho})^{-1}, \tag{22}$$

where the last part tells us how each region's $\delta\rho/\rho$ will grow with time. Since the growth rate depends only on the average density and the scale factor, the growth rate will be the same for all regions, and the growth rate for the rms of the fluctuations will be same as well. We identify two relevant cases,

• Radiation dominated: $\sigma_M \sim a^2$, since $\bar{\rho} \sim a^{-4}$
• Matter dominated: $\sigma_M \sim a$, since $\bar{\rho} \sim a^{-3}$.

4.3 When to specify the spectra?

In Section 4.1 we derived the rms amplitude of mass density fluctuations as a function of scale (k, or λ, or r), and as a function of average mass enclosed by that scale. What we have not said is at what cosmic time these measurements/definitions are being done. In fact, there are two physically relevant ways of picking the time. **One**, is to specify the amplitude of all spectral modes at one given cosmic epoch (usually chosen to be early on, say, right after inflation ends), and **Two**, to specify the amplitude of a given Fourier k-mode at a time when that mode just crosses inside the Hubble horizon, $1/H$, and so becomes subject to pressure forces, conductivity, etc.[1] The latter definition (**Two**) implies that we are specifying the amplitudes of various modes at different cosmic epochs. This turns out to be useful, as we'll see shortly.

Suppose we know what the spectrum looks like at some early (initial) time, and it is characterized by indeces n_i and α_i. Consider one k-mode. What amplitude will that mode have as it crosses inside the Hubble horizon? The rms value of the fluctuations on that scale, σ_M, and the average mass enclosed by that scale, M_H, grow continuously, but at two different rates. Furthermore, these rates depend on whether the Universe is matter or radiation dominated. So in all we have to specify four different growth rates. The two growth rates for σ_M were calculated in the last section.

The growth rate of average mass enclosed inside the Hubble horizon is $M_H \sim \; r_H^3 \, \bar{\rho}$, where only the non-relativistic matter contribution to $\bar{\rho}$ is considered; r_H scales as $1/H \sim t$, so:

• Radiation dominated: $M_H \sim t^3 \, a^{-3} \sim a^3$ (recall : $a \sim t^{1/2}$)
• Matter dominated: $M_H \sim t^3 \, a^{-3} \sim a^{3/2}$ (recall : $a \sim t^{2/3}$)

First, consider a mode that crosses inside the horizon before matter-radiation equality. That means we have to consider only one growth rate for fluctuations and one growth rate for the enclosed mass. These are:

$$\sigma_M \sim a^2 \sim t \sim (1+z)^{-2}; \quad M_H \sim a^3 \qquad [z > z_{eq}] \tag{23}$$

Combining these (eliminating a) and remembering the initial spectrum shape we get,

$$\sigma_M \sim M_H^{-\alpha_i} a^2 \sim M_H^{-\alpha_i + 2/3} = M_H^{-\alpha_H}. \tag{24}$$

Now the slightly more involved case of modes that cross inside the horizon *after* matter-radiation equality. There modes have longer comoving wavelengths that we ones we just dealt with. These grow according to eq. 23 before z_{eq} and and according to

$$\sigma_M \sim t^{2/3} \sim a; \quad M_H \sim a^{3/2} \qquad [z_{eq} > z] \tag{25}$$

after equality. So for these modes

$$\sigma_M \sim M^{-\alpha_i} a_{eq}^2 a \sim M_H^{-\alpha_i + 2/3} = M_H^{-\alpha_H}, \tag{26}$$

and we've dropped the constant factor of a_{eq}^2, since we are only interested in relative scalings. So regardless of when the mode crosses inside the horizon its amplitude at that time will be given by

[1] Recall that during the standard evolution the scale factor grows slower than the Hubble scale, so with time, Hubble scale grows bigger and bigger in terms of comoving length, and so larger and larger comoving scales λ become horizon-sized, and then sub-horizon-sized.

$\sigma_M \sim M_H^{-\alpha_i+2/3} = M_H^{-\alpha_H}$. In all there are four relevant spectral indices:

• n_i and n_H specify single-epoch and horizon-crossing $P(k)$'s dependence on k;
• α_i and α_H specify single-epoch and horizon-crossing σ_M's dependence on M_H.
The relation connecting n's is: $n_H = n_i - 4$, and the relation connecting α's is: $\alpha_i - 2/3 = \alpha_H$.

4.4 Harrison-Zeldovich spectrum

Now, notice that if $n_i = 1$, then $\alpha_H = 0$, and the rms amplitude of mass fluctuation, σ_M are *constant on all scales.* This spectrum was first proposed independently by Harrison and Zel'dovich in 1970 and 1972, because of its special physical significance of *scale-invariance.* The initial, single epoch perturbation spectrum looks like,

$$\sigma_M \sim M^{-2/3}, \quad \text{and} \quad P(k) \sim k. \tag{27}$$

A nice feature of HZ is that because $\alpha_H = 0$, no matter how large or how small the mass scale of the mode, the fluctuations are guaranteed to be finite, in other words, the spectrum does not 'blow up' at either end, and so does not have to be artificially truncated at either long or short wavelength side. Two types of physical mechanisms could give rise to such a spectrum: inflation (see Section 4.6), or topological defects, like cosmic strings. The latter are ruled out as the source of cosmic structure on observational grounds (for example, CMB).

4.5 P(k) from observations

Can we say anything about the form of the power spectrum from observations? Can we constrain n? Consider three mass scales and the corresponding fluctuations in mass (M_H, σ_M) when those scales crossed the horizon, i.e. when the corresponding size becomes smaller than H^{-1}.

1. Present day horizon: The present day size of the Universe, $\sim H_0^{-1} \sim 3000h^{-1}$ Mpc is just now coming into causal contact with itself. What are the rms fluctuations on this scale? If fluctuations were large then the Cosmic Microwave Background photons coming to us from different directions in the sky would have suffered large temperature changes, i.e. large $\delta T/T$, which is not observed: CMB is nearly a perfect blackbody whichever direction in the sky you look at. Fluctuations are tiny, $\sim 10^{-4...-5}$, and set an upper limit on the rms fluctuations on the present day horizon scale. The mass scale here, i.e. the mass contained within the observable Universe today is about $M \sim 10^{22} M_\odot$.

2. Comoving 10 Mpc scales: This is roughly the comoving size scale of horizon at the epoch of matter-radiation equality. Since present fluctuations on this scale are of order 1, linear growth tells us[2] the amplitude must have been few $\times 10^{-5}$ at z_{eq}. Mass contained within this scale is about $10^{15} M_\odot$.

3. Primordial black holes: Black holes smaller than about 10^{15} grams, or $10^{-18} M_\odot$ would have evaporated by now, but not larger ones. Suppose primordial $P(k)$ was such that fluctuations were close to 1 at this low mass end. Then as soon as these fluctuations cross inside the horizon gravity will quickly take over and make BH's before pressure has enough time to dissipate perturbations. We do not see numerous BH in that mass range, so fluctuations must have been such that $\sigma_M < 1$ as that mass scale crossed inside horizon.

[2] We will derive this result—growth of fluctuations on sub-horizon scales during the matter dominated epoch—a little later in the course.

Points *1*, *2*, *3* give us a direct observational way to constrain the slope of the fluctuation spectrum. They span about 40 orders of magnitude in mass (!), and yet the rms amplitude of fluctuations changes by at most 4-5 orders of magnitude. This implies that $\sigma_M \sim M_H^{\pm 0.1}$, or, approximately, the fluctuations are scale-invariant over a huge range of scales. In other words, $\sigma_M \sim M^{-\alpha_H}$, with $\alpha_H \sim 0$.

4.6 P(k) from inflation

As we saw earlier, during inflation the scale factor grows exponentially, $a \sim e^{Ht}$, while the Hubble parameter stays about the same (one of the slow roll conditions); the latter defines the largest possible physical scale, H^{-1} that a light signal (or any other information carrying signal) can traverse in a time it takes the Universe to change considerably (double in size).

Since H is huge in the early Universe, Hubble scale is tiny, and quantum fluctuations are important on that scale. H^{-1} is the largest scale that can be affected by causal physics, like quantum fluctuations. Now, consider a certain comoving k-mode, or k-scale, at a time when its proper size is smaller than the Hubble scale. The fluctuations arise and disappear, new fluctuations appear, etc. This goes on until that comoving scale increases in proper size to the size of the Hubble radius. This is the last chance for quantum fluctuations to make an imprint on that comoving k-scale. Once the proper size of the k-mode becomes super-horizon causal quantum fluctuations cannot affect it. So the comoving scale whose proper size is super-horizon will retain the quantum fluctuation amplitude that was 'stamped' on it as it was leaving the Hubble horizon. Once that comoving scale is outside the horizon, and out of causal contact with different parts of itself, its amplitude cannot be changed by any microphysical process, so it is said to be frozen in. Furthermore, as that comoving scale continues to get inflated it reaches macroscopic size and the fluctuations that arose as quantum fluctuations become 'classical'. These later provide the seeds for large scale structure in the Universe, like galaxies, clusters and superclusters. Remember that even though we considered a single region there are in fact many such regions in the Universe, and each gets its own different fluctuation amplitude on each comoving scale. Note that because of the nature of quantum fluctuations the phases of the perturbation modes are uncorrelated; this is a feature which appears to be well borne out by the observations of the large scale structure today. The large scale density fluctuations in today's Universe are well described by a Gaussian random field, completely consistent with their inflationary origin.

From quantum mechanics we know that the energy fluctuation amplitude is inversely proportional to the duration of the fluctuation, $\Delta E\ \Delta t \geq \hbar$, by the uncertainty principle. Because $\Delta t \sim H^{-1}$ as a given scale crosses outside the Hubble horizon, the energy imparted to it is $\Delta E \sim H$. When the fluctuation becomes classical, ΔE translates into $\Delta\Phi_{grav}$, creating a density (or curvature) dimple in space-time. So each scale gets the same amplitude of $\Delta\Phi_{grav}$. In the following we assumed that average density $\bar{\rho}$ stays constant during inflation, as is expected for a component with $w \approx -1$. So,

$$\Delta\Phi_{grav}(r) \approx \frac{G\delta M}{r} \sim r^2\delta\rho/\rho \sim r^2\sigma_M \sim M^{2/3}\sigma_M = \text{const}, \tag{28}$$

since this has to hold for all scales, we see that $\sigma_M \sim M^{-2/3}$. Consulting eq. 27 we conclude that quantum fluctuations during inflation lead to scale-invariant Harrison-Zeldovich perturbation spectrum, with $\alpha_i = 2/3$, $\alpha_H = 0$, and $n_i = 1$.

Harrison-Zel'dovich spectrum of fluctuations arises naturally from inflation. Does that mean that inflation took place? Not necessarily; there could be other physical mechanisms that would produce scale invariant spectrum. (Note that Harrison and Zel'dovich papers came out in 1970-72, whereas inflation first appeared on the scene in 1981.)

Hubble parameter H does not stay exactly constant during inflation: it decreases a little because $V(\phi)$ decreases (one of the slow-roll conditions gives us a relation between H and $V(\phi)$). This means that the later comoving scales to participate in inflation will get smaller ΔE and $\Delta\Phi_{grav}$ imprinted on them. These scales are the last ones to leave the horizon during inflation and are the first ones to come back into our horizon during the post-inflationary standard evolution. Therefore these are the 'small scales', or large k-modes, as viewed by us now. Because of the gradual decrease in H during inflation we would expect these smaller scales to have less power. In effect, $P(k)$ emerging from inflation should have a small reddish tilt, i.e. n_i should be slightly smaller than 1.

5 Discussion

Inflation is still a working model. It needs a proper context within the rest of physics, as well as unique and testable predictions that can be compared to observations. Two of the most natural prediction of inflation are actually confirmed by observations: the Harrison-Zel'dovich spectrum of perturbations, and the Gaussian random field nature of the density perturbation field. However, these could also be the outcome of some other theory, not inflation. Furthermore, even though these features come out naturally from the most straightforward versions of inflation, one can come up with many "modified" versions of inflation that would not produce these two features.

Another prediction of inflation is the production of a gravitational wave background. In principle, these can be detected using gravity wave detectors, like LIGO, or indirectly, from the Cosmic Microwave Background. Neither have been detected yet.

What took place before inflation? It doesn't really matter, because once inflation starts its exponential increase in scale factor sweeps away any prior populations of particles, relics, etc. that might have existed. It also irons out completely whatever pre-inflation space-time wrinkles our patch of the Universe might have had.

Inflation was first proposed as a means of getting rid of magnetic monopoles, an unavoidable by-product of GUT. Not only does inflation get rid of monopoles, it also nicely explains at least three other important puzzles in cosmology: horizon and flatness problems and the spectral index of the matter density fluctuations. Even though it has its own problems (fine-tuning) inflation has a hall-mark of a good theory: it explains more observational problems than it was originally designed to explain. One must also keep in mind that inflation has successfully stood the test of time: in the nearly 30 years since it first appeared, no other theory has been able to replace it, or even stand alongside inflation as a serious contender.

If inflation is to prove to be the correct theory of very early Universe, it must be shown to be a natural consequence of GUT, supersymmetry, or the like.