Lecture notes on inflation and primordial black holes

Christian T. Byrnes[a1]

[1]*Department of Physics and Astronomy,*

Pevensey II Building,

University of Sussex, BN1 9RH, UK

Lecture notes for the GGI school about Inflation and Primordial Black Holes, held virtually during March 2021.

The first 2 lectures are about inflation and covered by these notes.

The third lecture will be held using slides.

The final 2 lectures about primordial black holes are also covered by these notes.

Please let me know about any typos you spot by email to c.byrnes@sussex.ac.uk. Many thanks to Andreas Mantziris who sent me a very helpful list of corrections.

CONTENTS

[a] E-mail: c.byrnes@sussex.ac.uk

I. INTRODUCTION AND THE AIMS OF THESE LECTURE NOTES

In this short lecture course, I will aim to address two topics about two of the most fundamental questions in cosmology, what is the universe made of, and what were the initial conditions? We will study the leading theory of the early universe called inflation, and primordial black holes, which are a dark matter candidate.

We will first briefly study the historical motivations for inflation, before understanding how it can generate the initial density and temperature perturbations. Remarkably, inflation does this by invoking quantum mechanics - the theory of the smallest scales - and general relativity, to explain how the largest scale objects in the universe today are a consequence of physics on the smallest scales during the epoch of inflation. I will sketch the calculation of how the initial perturbations were generated and how to relate them to the perturbations observed at much later times via the cosmic microwave background and large-scale structure.

There are many pieces of evidence that the majority of gravitationally clustered matter in the universe is non-baryonic (e.g. galaxy rotation curves, structure formation and the bullet cluster), the so-called dark matter. Primordial black holes (PBHs) are the unique dark matter candidate which does not invoke a new particle to explain this, and instead invokes non-standard initial conditions to form black holes within a second of the big bang and the end of inflation. Such non-standard initial conditions are perhaps most likely generated by inflation, which is then required to generate a much larger amplitude of perturbations on the small scales relevant for PBH formation compared to the perturbations observed on larger scales where we have precision observations. A study of PBHs therefore naturally relates to theories of inflation and provides a window onto the early universe on smaller scales than can be probed by any other means. Even if PBHs are only a negligible fraction of the dark matter, the detection of even a single PBH would have major implications on our understanding of the early universe, and current constraints on their existence are used to constrain the physics of inflation.

The suggestion of PBHs predates inflation, but interest in them has never been higher following the detection by LIGO of merging black holes through their gravitation wave signature. Could some of those black holes be primordial?[1]

Either of these subjects could easily be the subject of a much longer lecture course. I will therefore give only a summary of many of the details and derivations, with the goal being to provide a practical and modern knowledge of both subjects, with references being given to help you fill in the gaps. Because these are lecture notes, we are not aiming to provide a comprehensive list of references and will instead focus on review articles and books when possible. The emphasis will be on providing intuition and understanding rather than technical details, while the workshops led by Philippa Cole will be used to fill in some of the technical details and help build an understanding of

[1] `https://www.quantamagazine.org/black-holes-from-the-big-bang-could-be-the-dark-matter-20200923/`

some of the calculational techniques.

In these lectures I will use natural units where $c = \hbar = 1$ unless otherwise stated, and use the reduced Planck mass which is related to Newton's gravitational constant by $M_{\rm Pl}^2 = 1/(8\pi G)$.

II. A CRASH COURSE IN BACKGROUND COSMOLOGY

Although I will focus only on inflation and primordial black holes in these lectures, some knowledge of cosmology at the level of the homogeneous and isotropic background is essential before we can begin. If you have not studied much cosmology I highly recommend the very readable text book "Introduction to Cosmology" by Barbara Ryden which does not require any prior knowledge of general relativity.[2] Isotropic means that something looks the same in every direction, and this is true for the observed universe, as is best evidenced from observations of the cosmic microwave background. Homogeneous means the same everywhere, and 3D galaxy surveys confirm that statistically, and on very large scales (scales larger than a cluster of galaxies) the universe is homogeneous.

Although the universe today is very inhomogeneous on galactic scales, with large structures having grown due to the power of gravitational attraction applied over billions of years, the early universe was much closer to be homogeneous on all scales which are large enough to be observationally probed today. Once again, the best probe comes from observations of the temperature differences observable on the CMB, which is the last scattering surface of cosmic photons which finally became free to travel long distances through the universe about 400,000 years after the Big Bang shortly after the background temperature/energy had reduced sufficiently to allow electrons to bind to atomic nuclei in the epoch known as recombination,[3] due to the large number of high energy photons which could strip the electrons away from the (mainly hydrogen and helium) nuclei. The density/temperature perturbations observed in the CMB have a characteristic amplitude of a few parts per hundred thousand, showing

[2] https://www.cambridge.org/core/books/introduction-to-cosmology/7E9E7C9C717570F1FFB3BA70F864A8FA

[3] but which should really be called 'combination' since electrons and atomic nuclei had never been bound before this time

that the universe really was very close to being "smooth" at those times, and because gravitational attraction acts to make the density perturbation amplitude grow with time, it is essentially sure that the earlier universe (at least on the scales which can be observed today) was even closer to being perfectly smooth.

Three key equations describe the evolution of a homogeneous and isotropic universe, modelled in terms of the growth of the cosmic scale factor $a(t)$ which relates physical and comoving scales via

$$r(t) = a(t)x$$

where the comoving distance x between two comoving observers (those which are comoving with the expanding universe) is constant, while the physical distance $r(t)$ grows proportionally with the growth of the universe. The three equations are the **Friedmann equation**

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{1}{3M_{\rm Pl}^2}\rho - \frac{\kappa}{R_0^2 a^2}, \tag{II.1}$$

the **fluid equation**

$$\dot{\rho} + 3H\left(\rho + P\right) = 0, \tag{II.2}$$

and the **acceleration equation**

$$\frac{\ddot{a}}{a} = -\frac{1}{6M_{\rm Pl}^2}\left(\rho + 3P\right). \tag{II.3}$$

The energy density is $\rho(t)$, the pressure $P(t)$, the curvature κ and radius of curvature measured today is R_0. As usual a dot denotes a derivative with respect to time.

These are three of the most important equations in cosmology, so they deserve some explanation. I will not attempt to derive them, and a rigorous derivation requires a reasonable background in GR. Perhaps the first thing to stress is that only two of the three equations are independent, so it is possible (and a good exercise) to combine the first two equations to derive the third one.

Because only two of the three equations written above to describe the evolution of an FLRW universe are independent, we need a third equation to solve for the three unknowns, $a(t)$ (or equivalently $H(t)$), $\rho(t)$ and $P(t)$, even if we assume we know the value of the curvature. We need an extra ingredient, which is the equation of state of the fluid(s) in the universe, i.e. a relationship of the form

$$P = \omega\rho, \tag{II.4}$$

or, in practise, multiple such relationships if the universe has multiple components contributing to the energy density. Note that this linear relationship is obviously not the most general form of $P(\epsilon)$ possible, especially since we will assume ω is a constant for each component. However, in practise, this simplifying assumption is perfectly adequate for describing our real universe during most of its evolution, and it allows us to make substantial analytical progress in solving the Friedmann equation and hence determining how the different components of the universe evolved with time, and the history of the Hubble parameter, etc.

We do not have time to derive the different properties of the different types of energy components of the universe, but to briefly summarise, the important components after inflation has ended are matter (which is pressureless, $\omega_{\rm mat} = 0$), radiation ($\omega_{\rm rad} = 1/3$), curvature ($\omega_\kappa = -1/3$) and the cosmological constant ($\omega_\Lambda = -1$).

By solving the fluid equation, component by component, you can (quite straightforwardly) derive

$$\epsilon_{\rm rad} \propto a^{-4},\ \epsilon_{\rm mat} \propto a^{-3},\ \ \epsilon_\kappa \propto a^{-2},\ \ \epsilon_\Lambda \propto a^0 = \text{constant}.$$

Hence in the future Λ will dominate because it cannot be diluted, while in the past radiation must have dominated. In between, we will see that there was a long period when matter dominated. It is not believed that the curvature was ever important, except possibly before an epoch of inflation during the extremely early universe. But that is a different story which we will come back to near the end of this module. For now, it is important to note these results and that treating the universe as if it only had one energy component is a useful exercise which provides a good approximation to the expansion of the universe during significant time periods. We will go on to study such cases, before considering the more complete and complex case of a multifluid universe.

III. INFLATION

Cosmologists can probe the epoch of the cosmic microwave background (CMB) from about 400,000 years after the Big Bang relatively directly by measurements of the photons which last scattered when the CMB formed. Less directly we can keep extrapolating back in time until about 1 second after the Big Bang, to the epoch of Big Bang nucleosynthesis (BBN), and once again there is observational evidence that the predictions made for the abundances of the light elements are correct, giving us confidence that we understand the history of the universe back to a time when the energy scale was above an MeV and when neutrinos started free streaming.

Going even further back in time is speculative, but there is good (albeit inconclusive) evidence that much less than a second after the Big Bang there was a period when the universe underwent accelerated expansion, popularly known as inflation. The three classic (and original) pieces of evidence for a period of inflation were that it could answer the following three riddles:

1. Why does the universe obey the cosmological principle, i.e. why is it homogeneous and isotropic? Although this minimal assumption has turned out to be true, to the good fortune of cosmologists around the world, it is not easy to explain why the CMB temperature on opposite sides of the sky, which should never have been in causal contact, have nearly the same temperature.

2. Why is the universe so close to spatially flat? Since the effective energy density of curvature dilutes like a^{-2} with the expansion, which is slower than both matter and radiation, it seems surprising that it didn't come to dominate the evolution of the universe before dark energy, e.g. Λ, became important.

3. Why don't we see any magnetic monopoles or other massive relics from the high energy early universe? The basic issue is that if we keep extrapolating the laws of physics back to ever higher energy scales, then it appears "likely" that stable particles with a large mass would have formed in huge quantities. But such particles have never been seen, so how did they evade detection?

In most textbooks, the above three problems are presented one at a time, followed by an explanation of what inflation is, and then a new list of how inflation solves the three classic problems. In appendix A we instead discuss each 'problem' in order, providing both more details of the challenge and how inflation offers a solution. The reader familiar with the original motivations from inflation may skip this part.

More importantly, we will discuss how inflation is believed to have also generated the primordial perturbations, which are observed as temperature perturbations on the CMB sky. This is a remarkable story which potentially explains the origin of all cosmological structures - including galaxy clusters which are the biggest objects in the universe - as being due to quantum mechanical perturbations present during the brief period of inflation. There could not be a grander story of mighty oaks growing from tiny acorns than this one!

A. What is inflation?

The simplest answer is that inflation refers to an early period when the expansion of the universe accelerated, meaning that $\ddot{a} > 0$. Due to the cosmological constant Λ (or more generally, dark energy), the universe also appears to be inflating now, but at a vastly lower energy scale (recall how much the temperature and the energy of the universe have decreased since BBN, and inflation must have occurred before BBN since otherwise the abundance of the primordial elements such as hydrogen and helium would have been diluted to nearly zero.

From the acceleration equation, (II.3), we can see that $\ddot{a} > 0$ means that the total equation-of-state parameter must satisfy

$$\omega < -\frac{1}{3}.$$

However, it turns out to be extremely difficult to get an extended duration of inflation[4] unless the equation-of-state parameter satisfies $\omega \simeq -1$, so we will use this as a practical definition of inflation. The good news is that this means the Friedmann equations are

[4] I do not mean that it lasts a long time, but rather that the universe grows by many orders of magnitude. We will soon learn a good way to quantify the amount of expansion during inflation.

quite straightforward to solve for such a scenario, because inflation means the universe approximately behaves as if it was dominated by a cosmological constant. However, inflation cannot be caused by a cosmological constant because the energy scale must be greater than the energy scale of the universe at the time of BBN, which is a vastly larger energy scale than the scale of dark energy/Λ today. During inflation the scale factor grows at an exponential rate,

$$a(t) \propto e^{H_{\rm inf} t}. \tag{III.1}$$

When discussing how much inflation is needed to fix the horizon, flatness and monopole problems, the question is not how long (in nano seconds, or any other units) inflation lasted, but rather by how much the scale factor grew during inflation, when $\omega \simeq -1$. A common and convenient measure of this growth is the efolding number, where 1-efolding means that the universe has grown by a factor of $e \simeq 2.72$, 2-efoldings mean by a factor of e^2, and in general N-efoldings mean by a factor of e^N. The total number of efoldings of inflation is defined to be

$$N_{\rm inf} \equiv \ln\left(\frac{a_{\rm inf,end}}{a_{\rm inf,initial}}\right) = H_{\rm inf}(t_{\rm inf,end} - t_{\rm inf,initial}), \tag{III.2}$$

where the equality follows from (A.11). By defining the duration of inflation in terms of the growth of the scale factor, we remove the degeneracy between the (unknown) energy scale of inflation (which is related to $H_{\rm inf}$) and how long (in time) inflation took place.

1. Length scales and the comoving Hubble scale

We will soon go on to calculating properties of the perturbations generated during inflation. An important concept for this calculation will be the horizon scale, which is related to the Hubble scale and Hubble time. The full units for the Hubble constant measured today are the weird sounding km/s/Mpc, so it has units of inverse time and the Hubble time $1/H$ gives an estimate for the age of the universe. The Hubble distance, c/H is the distance that light can travel in one Hubble time, making it a good estimate for the 'scale of causal interactions' as a function of time. Recalling that physical scales are related to a comoving scale by a factor of a and that we are using units with $c = 1$, we can deduce that the comoving Hubble scale is given by

$$\frac{1}{aH}. \tag{III.3}$$

During inflation $H \simeq$ constant and hence $1/(aH) \propto 1/a$. During radiation domination $a \propto t^{1/2}$, $H \propto 1/t$ and therefore $1/(aH) \propto t^{1/2} \propto a$ while during matter domination $a \propto t^{2/3}$, $H \propto 1/t$ and therefore $1/(aH) \propto t^{1/3} \propto a^{1/2}$. The evolution of the comoving Hubble scale as a function of efolding number $N = \ln(a)$ is shown in Fig. 1. Notice that even though the universe is today dominated by the cosmological constant, this has only become dominant in very 'recent' times when time is measured on a log scale.

Because it is normal to measure the statistical properties of the primordial density perturbations in terms of the power spectrum in Fourier space, it is normal for cosmologists to talk about 'length' scales in terms of the comoving wavenumber $k \sim 1/(\text{comoving length})$ which is often measured in units of inverse megaparsecs (Mpc^{-1}). Notice that large values of k corresponds to small scales, and vice versa. An important concept for calculations of inflationary perturbations is whether the length scale is smaller or larger than the comoving Hubble scale, with perturbations on length scales which are larger than the comoving Hubble scale being called super-Hubble or super horizon. This corresponds to $k < aH$ while a sub-Hubble mode satisfies $k > aH$. The mode crosses the Hubble scale when $k = aH$. As shown by Fig. 1, modes start off inside the Hubble scale, then exit during inflation and later re-enter the Hubble scale after inflation ends, with larger scale modes (corresponding to smaller values of k) exiting earlier and re-entering later than small scale modes. For reasonable choices of the inflationary energy scale, the number of efoldings from when our observable horizon exited during inflation until the end of inflation is between 50 and 60 [1].

B. What could cause inflation?

We will start by considering a single-scalar field ϕ with a potential energy given by $V(\phi)$. Because the background universe was close to being homogeneous and isotropic, we will make the assumption that the background inflaton field only depends on time

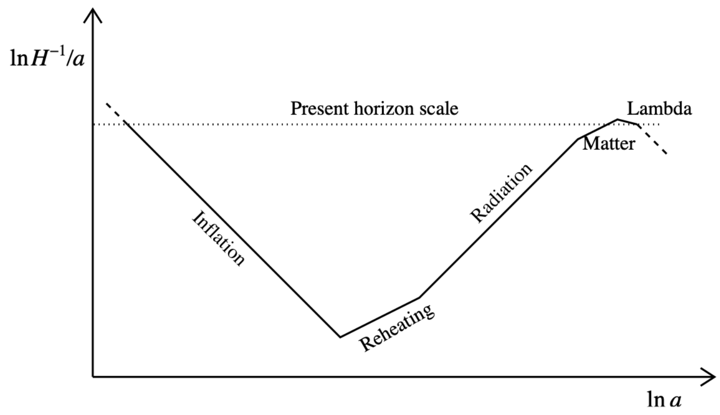

FIG. 1. The evolution of the comoving Hubble horizon as a function of $N = \ln(a)$, showing how modes (with a constant comoving length) become super-Hubble ($k < aH$) during inflation and then re-enter the horizon after inflation has ended when $1/(aH)$ grows. The possibility of an early epoch with zero pressure during the time when the inflaton field decays into radiation is also shown and labelled as 'reheating'. This plot is taken from [1].

but is spatially invariant, hence $\phi = \phi(t)$. The energy density and pressure of this field[5] are

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi), \tag{III.4}$$

$$P_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi), \tag{III.5}$$

so we can see that the background equation-of-state parameter will be close to that of a cosmological constant, $w \simeq -1$, whenever the potential energy dominates over the kinetic energy of the field, i.e. if

$$V(\phi) \gg \frac{1}{2}\dot{\phi}^2. \tag{III.6}$$

Having an equation of state close to -1 means that the background density is only reducing very slowly, with the 'extreme' case of $\omega = -1$ corresponding to a constant background energy. A good way to parametrise how quickly the background energy is

[5] This comes from the energy-momentum tensor of general relativity

diluting is via the slow-roll parameter

$$\epsilon_H \equiv -\frac{\dot{H}}{H^2}, \tag{III.7}$$

and with a bit of algebraic manipulation using the (flat) Friedmann and acceleration equations we can write this in terms of the equation-of-state parameter as

$$\epsilon_H = -\frac{1}{H^2}\frac{\ddot{a}}{a} + 1 = \frac{3M_{\rm Pl}^2}{\rho}\frac{1}{6M_{\rm Pl}^2}(\rho + 3P) + 1 = \frac{1}{2}(1+3\omega) + 1 = \frac{3}{2}(1+\omega). \tag{III.8}$$

From the acceleration equation we can see that $\ddot{a} > 0$ if $\omega < -1/3$, and therefore acceleration (which is the definition of inflation) corresponds to $\epsilon_H < 1$. The commonly considered case of $\omega \simeq -1$ corresponds to $\epsilon_H \ll 1$.

C. Slow-roll inflation

We will now consider in more detail how to describe the (background) motion of a scalar field which causes inflation. The background equation of motion for such a field in an expanding background is given by[6]

$$\ddot{\phi} + 3H\dot{\phi} + V' = 0, \tag{III.9}$$

where the overdots refer to time derivatives and the prime denotes a derivative of the potential with respect to the inflaton field, ϕ. The middle term represents the 'Hubble friction' caused by the expansion of the universe, which would be zero if the background was static.

Taking the time derivative of the Friedmann equation for a scalar field in a flat background

$$H^2 = \frac{1}{3M_{\rm Pl}^2}\left(V + \frac{1}{2}\dot{\phi}^2\right), \tag{III.10}$$

$$\Rightarrow\; 3M_{\rm Pl}^2 \times 2H\dot{H} = V'\dot{\phi} + \dot{\phi}\ddot{\phi} = \dot{\phi}\left(V' - (3H\dot{\phi} + V')\right) = -3H\dot{\phi}^2 \tag{III.11}$$

where we used the equation of motion (III.9) and hence we can rewrite the slow-roll parameter (without making any approximation) as

$$\epsilon_H \equiv -\frac{\dot{H}}{H^2} = \frac{\dot{\phi}^2}{2H^2M_{\rm Pl}^2}. \tag{III.12}$$

[6] A derivation of this result goes beyond the scope of this course, but can be derived by applying the Euler-Lagrange equation to the action of the scalar field.

It should not be difficult to convince yourself that this slow-roll parameter is therefore measuring the ratio of the kinetic and potential energies (up to order unity numerical factors) and hence inflation will occur when the kinetic energy is subdominant, as we had previously shown in another way.

There are a huge number of ways of parametrising and classifying the slow-roll conditions. At lowest order in slow roll, two commonly used parameters are

$$\epsilon_V = \frac{M_{\rm Pl}^2}{2}\left(\frac{V'}{V}\right)^2 , \qquad \text{(III.13)}$$

$$\eta_V = M_{\rm Pl}^2 \frac{V''}{V}, \qquad \text{(III.14)}$$

where the primes imply differentiation with respect to the inflaton field ϕ. The slow-roll conditions imply that

$$\epsilon_V \simeq \epsilon_H \ll 1\,, \qquad |\eta_V| \ll 1, \qquad \text{(III.15)}$$

and one can show that these approximations are equivelant to dropping the following terms from the equations governing the evolution of the inflaton field

$$\cancel{\ddot{\phi}} + 3H\dot{\phi} + V' = 0 \qquad \text{(III.16)}$$

$$H^2 = \frac{1}{3M_{\rm Pl}^2}\left(V + \cancel{\frac{1}{2}\dot{\phi}^2}\right) \qquad \text{(III.17)}$$

which results in the simplified slow-roll equations (which can be solved analytically for some simple choices of $V(\phi)$)

$$3H\dot{\phi} \simeq -V' \qquad 3H^2 M_{\rm Pl}^2 \simeq V. \qquad \text{(III.18)}$$

D. Perturbation generation

The reason for the continued interest in inflation 4 decades after it was first suggested has little to do with its ability to solve the classic horizon and flatness problems and is almost all to do with its ability to also generate the primordial perturbations. The story of how this works is quite complex and wonderful, with the basic idea being that the quantum mechanical uncertainty principle means that the inflaton field - and hence the universe - cannot become completely smoothed out by inflation, with quantum mechanical perturbations (which are normally only relevant on really tiny scales) becoming also important on large, classical, scales due to the quasi-exponential expansion of the universe which rapidly makes small scales grow into large scales. We won't attempt to derive the quantum mechanical perturbations in these lectures, but simply state the results and give some explanations. For a detailed treatment, see e.g. 'The primordial density perturbation' textbook by Lyth and Liddle.

The relevant energy scale of inflation is given by the Hubble parameter[7] and it is not so surprising that the typical amplitude of the scalar field perturbations ($\delta\phi$) as well as the metric (tensor) perturbations (h) are both linearly proportional to H, but a derivation of this important result goes beyond the scope of these lectures.

More precisely, we may write the power spectrum amplitude of the scalar (inflaton) and tensor (T) perturbations at horizon crossing during inflation as

$$\mathcal{P}_{\phi,*} = \left(\frac{H_*}{2\pi}\right)^2, \qquad \mathcal{P}_{T,*} = \frac{8}{M_{\rm Pl}^2}\left(\frac{H_*}{2\pi}\right)^2. \tag{III.19}$$

How do the $\delta\phi$ power spectrum relate to the observed temperature perturbations, $\Delta T/T$? It turns out in quite an indirect manner. The quantity which is closely related to observations is rather the dimensionless curvature perturbation[8]

$$\mathcal{R} = \frac{H}{\dot{\phi}}\delta\phi = \frac{1}{M_{\rm Pl}\sqrt{2\epsilon_H}}\delta\phi \simeq \frac{V'}{V}\delta\phi \simeq \frac{\delta V}{V} \simeq \frac{\delta\rho}{\rho}$$

and hence the power spectrum of $\mathcal{R}$ measured at horizon crossing is given by

$$\mathcal{P}_{\mathcal{R},*} = \frac{1}{2M_{\rm Pl}^2\epsilon_{H,*}}\left(\frac{H_*}{2\pi}\right)^2, \tag{III.20}$$

where H and ϵ_H should be evaluated at horizon crossing when $k = aH$, but this time (or equivelantly scale) dependence is often not shown explicitly.

[7] You may point out that the dimensions of the Hubble parameter are actually time^{-1} which is true, but in natural units time (and length) scales are both the inverse of mass and energy scales.

[8] I am skipping lots of details here, including the issue of gauges. The perturbed quantity $\delta\phi$ is a gauge dependent quantity but it turns out that $\mathcal{R}$ is gauge independent, provided that it is defined carefully. There also exist numerous definitions and sign conventions for the curvature perturbation, which is often also denoted by ζ, called 'zeta'.

E. Observational tests of inflation

The CMB temperature perturbations have been measured to high accuracy over three decades in length scales, with the best constraints on most scales coming from the Planck satellite. Although the spectrum looks complicated, it is a remarkable fact that the statistical properties of the CMB map, which consists of about 10 million pixels, can be parametrised in terms of a primordial power spectrum which has only two free parameters,

$$\mathcal{P}_{\mathcal{R}} = A_s \left(\frac{k}{k_{\text{pivot}}} \right)^{n_s - 1} , \tag{III.21}$$

where the pivot scale is chosen to be in the 'middle' of the range of data which can be constrained[9], $A_s \simeq 2 \times 10^{-9}$ is the amplitude of the primordial power spectrum and $n_s - 1$ is the spectral index, with $n_s = 1$ corresponding to a scale-invariant spectrum because this would imply that $\mathcal{P}_{\mathcal{R}}$ would be independent of k. Observations favour $n_s \simeq 1$ but exact scale-invariance has been ruled out with high significance.

Given the definitions of H, the slow-roll parameters, and using $k = aH$, you can check (and this is a good exercise to do so - beware it is easy to get the second relation wrong by an overall minus sign) that

$$\frac{d \ln H^2}{d \ln k} = -2\epsilon_H \tag{III.22}$$

$$\frac{d \ln \epsilon_V}{d \ln k} = 2\epsilon_V \left(2\epsilon_V - \eta_V\right) \tag{III.23}$$

and hence to leading order in slow roll (which means we can approximate $\epsilon_H = \epsilon_V$) we find the following, very useful result for the spectral index

$$n_s - 1 \equiv \frac{d \ln \mathcal{P}_{\mathcal{R}}}{d \ln k} = -6\epsilon_V + 2\eta_V , \tag{III.24}$$

where the slow-roll parameters should be evaluated around the time when the pivot scale $k_{\text{pivot}} = aH$ during inflation, which corresponds to 50-60 efoldings before the end of inflation, as discussed in section III A 1. Although the slow-roll parameters vary slowly-during slow-roll inflation and hence are normally essentially constant while the

[9] note that this is not a free parameter of the model.

decades of observable length scales cross the (comoving) Hubble scale, they do normally vary significantly during 50 efolds of inflation.

This is an incredibly useful formula. It means that for any single-field slow-roll model of inflation, you can determine the scale dependence of the primordial power spectrum in terms of the derivatives of the potential with respect to the scalar field and nothing more than that. This normally makes the calculation quite straightforward. One does however need to know for which value of the field to determine the derivative.

We can find the relation by starting with the formula $N \propto Ht$ during inflation, and we will use t_* to denote the time when k equals the comoving Hubble scale, $k = aH$, while t_e denotes the end of inflation. We therefore have

$$N = \int_{t_*}^{t_e} H dt = \int_{\phi_*}^{\phi_e} \frac{H}{\dot{\phi}} d\phi \simeq \frac{1}{M_{\rm Pl}^2} \int_{\phi_e}^{\phi_*} \frac{V}{V'} d\phi, \tag{III.25}$$

where one needs to use the slow-roll equations of motion in order to get to the final result. It is often a good approximation to use $\phi_* \gg \phi_e$ for values of $N \gg 1$.

The amplitude of the primordial tensor perturbations is often quoted as a ratio compared to the amplitude of the scalar perturbations, with the tensor-to-scalar ratio defined as

$$r \equiv \frac{\mathcal{P}_T}{\mathcal{P}_\mathcal{R}} = 16\epsilon_H. \tag{III.26}$$

Although this quantity also depends on scale (and should again be evaluated at horizon crossing) because ϵ_H varies slowly and r is constrained to be small but has not been detected, the mild scale dependence of r is normally not important.

The Planck constraints on inflation have determined that

$$n_s - 1 = 0.965 \pm 0.004, \qquad r \lesssim 0.1$$

while the addition of ground based data from Bicep-Keck strengthens the constraint on the tensor-to-scalar ratio to be $r < 0.044$ at the 95% confidence level [2]. This implies that $\epsilon_H \simeq \epsilon_V < r/16 \lesssim 0.003$ and therefore we can see from the formula for the spectral index, (III.24), that it deviates too far away from being scale-invariant for the deviation to be due to the ϵ slow-roll parameter, and hence the data requires

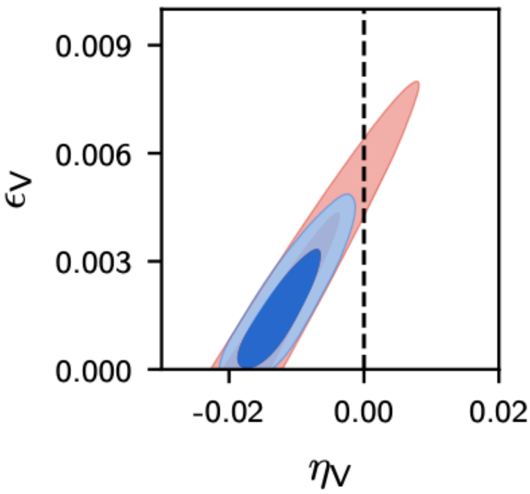

FIG. 2. The constraints on the two slow-roll parameters showing the 1 and 2-σ confidence levels. The red contours are based on the Planck data alone, while the tighter blue contours also include Bicep-Keck data. Notice that the addition of the Bicep-Keck data, which tightens on the constraint on r by about a factor of 2 is required in order to conclude that $\eta_V < 0$. This figure is taken from [2].

$\eta_V < 0$, which implies that $V'' < 0$. The observational constraints on ϵ_V and η_V are shown in Fig. 2. This demonstrates that even with such a limited number of non-zero inflationary parameters we can deduce something very non-trivial about the inflaton potential. However, all of this analysis has been made assuming the simplest case of single-field slow-roll inflation.

Lecture 2

IV. ULTRA SLOW-ROLL INFLATION

Recall the equation of motion for the inflaton field

$$\ddot{\phi} + 3H\dot{\phi} + V' = 0, \tag{IV.1}$$

where the prime in this equation refers to a derivative with respect to ϕ and the dot refers to a derivative with respect to (cosmic) time. The normal SR approximation is to drop the $\ddot{\phi}$ term and hence turn the second order equation of motion into a (much simpler) first-order equation, giving the result $3H\dot{\phi} \simeq -V'$. This approximation is normally valid provided that $\epsilon_V \ll 1 \Rightarrow V' \ll M_{\rm Pl}V$, but not always.

What happens if $V' = 0$? Then the SR approximation would say that $\dot{\phi} = 0$ implying that the field is not rolling at all. This does not imply that inflation has ended, rather the opposite. If the inflaton field is not at the bottom of the potential then the field will not move again and hence the universe is dominated by the potential energy of the inflaton field, which at least classically is the same as a cosmological constant and hence this gives rise to de Sitter expansion which is eternal in the future (called eternal inflation). We clearly do not live in such a universe.

However, if $V' = 0$ then it would be wrong to neglect $\ddot{\phi}$ in comparison to V', and we should instead study the following equation of motion:

$$\ddot{\phi} + 3H\dot{\phi} = 0. \tag{IV.2}$$

This is often described as the equation of motion for ultra-slow-roll (USR) inflation. Why USR? Because the field velocity decreases extremely quickly during USR inflation. We can solve (IV.2) by using the substitution $v = \dot{\phi}$ and recalling $H = \dot{a}/a$ to find the solution

$$\phi \propto a^{-3} \propto e^{-3N} \tag{IV.3}$$

which shows that the kinetic energy (KE) decreases like

$$\text{KE} \propto \dot{\phi}^2 \propto a^{-6}.$$

Hence the potential energy very quickly dominates over the KE, but $\dot{\phi}$ remains important both in order to get the inflaton past the flat part of the potential in order to avoid eternal inflation and also in order to calculate the inflaton perturbations.

Note that the first slow-roll parameter

$$\epsilon_H = \frac{\dot{\phi}^2}{2M_{\rm Pl}^2 H^2} \propto a^{-6} \ll 1 \tag{IV.4}$$

is small and rapidly decreasing during inflation, which implies that H is very close to constant. In contrast, $\epsilon_V = 0$ precisely if $V' = 0$, so whilst both versions of the epsilon slow-roll parameter are small during USR inflation, they do not have the same order of magnitude or time dependence. Hence, unlike in SR inflation, one cannot use them interchangeably.

We now introduce a "second" slow-roll parameter which measures how quickly ϵ_H varies with time,

$$\eta \equiv \frac{\dot{\epsilon_H}}{H\epsilon} \simeq \frac{\frac{d\epsilon_H}{dN}}{\epsilon_H} \tag{IV.5}$$

where we have used $N \simeq Ht$ to get to the final equality, which is valid during inflation. We can hence see the important results $\eta \simeq 0$ during SR inflation while $\eta \simeq -6$ during USR inflation.

In Fig. 3 we show an example of a potential which has an inflection point. USR inflation will occur while the inflaton field traverses the inflection point and for a brief period either side of the inflection point.

A. The amplitude of the perturbation

The standard formula for the amplitude of the perturbations generated during single-field slow-roll inflation is

$$\mathcal{P}_{\mathcal{R}} = \frac{H^2}{8\pi^2 M_{\rm Pl}^2 \epsilon}. \tag{IV.6}$$

It turns out that this remains partially true even when USR inflation occurs (only partially because it is not valid on all scales), but subject to some very important caveats. As we will see later, during USR inflation the perturbations do not freeze out

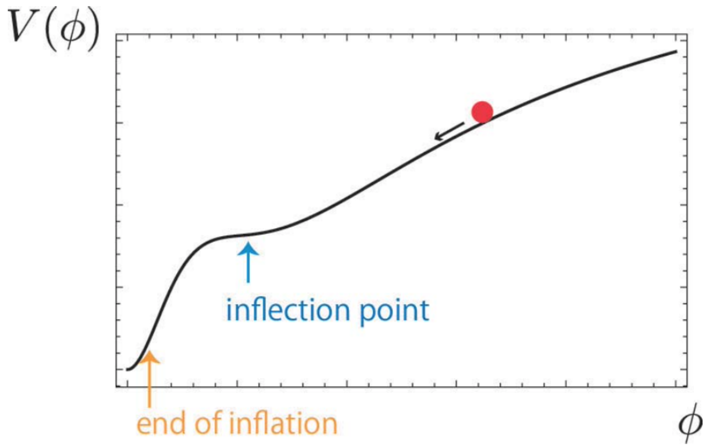

FIG. 3. An example of an inflaton potential that includes an inflection point. This plot is taken from a review article about PBHs [3].

at horizon exit, and hence the formula cannot be evaluated at horizon crossing but rather only after USR inflation ends. Another key difference is that while SR inflation implies that $\epsilon_V \simeq \epsilon_H$ and hence you can choose to use either definition of ϵ, during USR the two become extremely different and indeed $\epsilon_V = 0$ if $V' = 0$, but this does not imply that $\mathcal{P} \to \infty$. In the formula for the power spectrum, (IV.6), must be evaluated using $\epsilon = \epsilon_H$.

Recalling that $\epsilon \propto a^{-6}$ during USR inflation we can see that $\mathcal{P}_\mathcal{R} \propto e^{6N}$ during USR inflation since H is very close to constant, i.e. the perturbations (at least on some scales) grow very rapidly during USR inflation. This should make it clear that USR inflation cannot last very 'long', i.e. for a large number of efolds of inflation, because if it did then $\mathcal{P}_\mathcal{R} \to 1$ would be reached at which point perturbation theory breaks down since the perturbations are no longer small.

More seriously, $\mathcal{P}_\mathcal{R} \to 1$ also corresponds to the onset of eternal inflation, because this is when

$$\sqrt{\mathcal{P}_\mathcal{R}} \sim \frac{H}{\sqrt{\epsilon}} \sim \frac{H}{\frac{\Delta\phi}{\Delta N}} \sim 1$$

which means that the quantum fluctuations of the field – which have amplitude $\delta\phi \sim H$ – have the same magnitude as the distance which the inflaton field will (at the background level) roll down the potential during 1 efold of inflation, $\Delta\phi \sim H\Delta N$. At this point the inflaton field is about as likely to quantum mechanically "jump" up the potential (meaning that it moves backwards towards the starting point) as it is to continue moving towards the minimum of the potential where inflation can end. Hence inflation will never end (at least not in all locations) and this corresponds to the regime of eternal inflation.

Key messages:

1. You have to be very careful calculating the perturbations during USR inflation

2. An inflection point can lead to eternal inflation: It is then important to check is that the inflaton field has enough kinetic energy to get past the flat part of the potential and avoid eternal inflation

3. If $V' \neq 0$ exactly then USR inflation normally won't last long, because $\dot{\phi}$ is decreasing so quickly and hence slow roll with $3H\dot{\phi} \simeq V'$ will typically be quickly reached.

The (quasi) scale invariance of the perturbations generated during inflation is often intuitively explained as being a consequence of the (quasi) time translation invariance of the quasi de Sitter space expansion caused by inflation. Whilst these things do correspond to $H \simeq$ constant and hence both $\delta\phi$ and the tensor perturbations are nearly scale invariant, the scale invariance of the curvature perturbation is not guaranteed. During SR inflation the near constancy of ϵ does obviously imply from (IV.6) that the curvature perturbations are also nearly scale invariant, but when ϵ_H varies then scale invariance is normally lost. However, and quite remarkably, USR inflation can give rise to a scale-invariant spectrum of perturbations as well.

Warning: Do not confuse the time dependence with the scale dependence of the primordial curvature perturbations during USR inflation. We can only observe (at best) the scale dependence of the primordial power spectrum, evaluated after inflation has ended. For purely SR inflation it is enough (and much simpler) to just evaluate the primordial power spectrum at the time when modes exit the horizon during inflation, because the perturbations thereafter have a constant amplitude. During USR inflation this is not true, for reasons which will be explained in the next subsection.

B. USR perturbations on large scales

We will now study more quantitatively the evolution of the curvature perturbation, by starting with (but not deriving) its equation of motion. This is most conveniently written and solved in terms of conformal time τ,[10] which is related to cosmic time by the scale factor as

$$a d\tau = dt.$$

The name "conformal time" comes from the fact that the scale factor $a(t)$ becomes a conformal (overall) factor of the Friedmann-Le Maitre-Robertson-Walker metric when written in terms of this time coordinate. This is true for any global geometry (flat, closed or open) but specialising to the flat case for simplicity we can see this explicitly from

$$ds^2 = -c^2 dt^2 + a^2(t) \sum_{i=1}^{3} dx_i^2 \quad \text{(IV.7)}$$

$$= a^2(\tau) \left(-c^2 d\tau^2 + \sum_{i=1}^{3} dx_i^2 \right). \quad \text{(IV.8)}$$

In terms of conformal time, the equation of motion of the curvature perturbation, $\mathcal{R}(k) \equiv \mathcal{R}_k$, is

$$\frac{\partial^2 \mathcal{R}_k}{\partial \tau^2} + 2 \frac{\frac{\partial z}{\partial \tau}}{z} \frac{\partial \mathcal{R}_k}{\partial \tau} + k^2 \mathcal{R}_k = 0, \quad \text{(IV.9)}$$

where

$$z^2 = 2a^2 M_{\rm Pl}^2 \epsilon_H. \quad \text{(IV.10)}$$

In the $k \to 0$ limit one can find a partially analytic solution by first substituting $v_k = \partial \mathcal{R}_k / \partial \tau$, solving for v_k and then integrating, with the general solution (which is

[10] Conformal time is just as often called η in the literature, but I am already using η for one of the 'SR' parameters.

written in terms of two k dependent constants that depend on the initial conditions) to be

$$\mathcal{R}_{k\to 0} = C_k + D_k \int^{\tau} \frac{d\tau'}{a^2 \epsilon_H} \tag{IV.11}$$

$$= C_k + D_k \int^{t} \frac{dt'}{a^3 \epsilon_H}. \tag{IV.12}$$

The constant mode C_k corresponds to the usual constant mode, while D_k corresponds to the mode usually called the decaying mode. During SR inflation $\epsilon \simeq$ constant and hence the decaying mode does decay like a^{-3}, showing that in this case $\mathcal{R}_k$ does indeed freeze out at around the time of horizon crossing, after which $k \ll aH$.[11] For constant values of η it is straightforward to show that $\epsilon \propto a^{\eta}$ and therefore the "decaying" mode will in fact grow for $\eta \leq -3$. In particular, during USR inflation we have $\eta = -6$ and therefore

$$\mathcal{R}_k \propto D_k \int^t dt' a^3 \propto \int^t dt' e^{3Ht'} \propto e^{3Ht} \propto a^3, \tag{IV.13}$$

where we have used the fact that $a \propto e^{Ht}$ during inflation. Hence we see that the power spectrum during USR inflation grows like

$$\mathcal{P}_{\mathcal{R}}^{\rm USR} \sim \mathcal{R}_k^2 \propto a^6$$

as we had previously argued it must based on (IV.6) and the behaviour of $\epsilon_H \propto a^{-6}$ during USR inflation.

End of lecture 2

[11] Note that while this statement is true, this has not been shown rigorously - like many steps in these lecture notes - because we have only solved (IV.9) for $k = 0$ rather than finding the general solution and then taking the $k \to 0$ limit.

V. PRIMORDIAL BLACK HOLES

Lecture 3 will be held using slides and a pdf copy of those slides will be made available

Note that Laura Covi covered the observational evidence for the existence of a substantial fraction of the energy budget of the universe today consisting of cold dark matter in her lectures at the same school,[12] so I will not repeat that here. I will advertise though that the special nature of primordial black holes (PBHs) as a dark matter candidate is that they uniquely do not require the existence of any new particle. They are also (by far) the most massive dark matter candidate, being the only candidate I am aware of whose mass is larger than the Planck mass. We will see later that the main window for PBHs to make up all of the dark matter is if they have roughly asteroid mass, but it might be possible that PBH Planck mass relics are the dark matter. Furthermore, PBHs with almost any mass could make up a small fraction of the total dark matter, and there is no observational evidence for or against a mixed dark matter model with multiple components of dark matter.

It is important to realise that the evidence for dark matter does not just come from observations of the 'late' universe, such as the observations of galaxy clusters, galactic rotation curves and the bullet cluster, but also from the growth of perturbations between the time when the CMB formed and the time of the first galaxies as well as consistent measurements of the baryon-to-photon ration from the time of BBN about 1 minute after the Big Bang and the CMB formation about 400,000 years later. Hence, dark matter must have already existed during the early universe before BBN took place and this proves that 'stellar' black holes which form from the collapse of stars are not a viable DM candidate. However, PBHs form very early (we later determine the relation between their mass and formation time) and hence PBHs in certain mass ranges could be the DM.

[12] https://agenda.infn.it/event/24368/page/5548-scientific-program

Lecture 4

A. PBH formation

Recall the discussion in section III A 1 about the comoving Hubble scale $1/(aH)$. Modes with a constant comoving length scale $\sim 1/k$ are initially smaller than the comoving Hubble scale but this changes when $k = aH$ and they later have a longer wavelength than the comoving Hubble scale, meaning that $k < aH$ until after inflation ends and the comoving Hubble scales starts to grow. While $k < aH$ the mode is described as being 'super-horizon' or 'super-Hubble', and after inflation ends the modes will at some point 're-enter' the horizon when $k = aH$ again. The significance of horizon entry is that this time corresponds to when the corresponding horizon scale comes into causal contact, i.e. the time when information travelling at the speed of light can travel across the comoving scale $1/k = 1/(aH)$ in one Hubble time, $1/H$.

PBH formation is a causal process. Gravity (which travels at the speed of light) needs to communicate the existence of an overdensity in order for gravitational collapse to begin. Therefore, a PBH of scale $1/k$ cannot form while $k < aH$. The horizon scale is a key concept in PBH formation. PBHs form with a mass comparable to the horizon mass M_H, which means that there is an approximate 1-2-1 relation between the PBH mass $M_{\rm PBH} \sim M_H$ and k and time. We know that before matter domination began about 50,000 years after the Big Bang the universe was radiation dominated, and that it was radiation dominated at the time of BBN when the primordial elements (mainly hydrogen and helium) were formed about a minute after the Big Bang. At even earlier times we cannot be sure but in the standard model of cosmology the universe was radiation dominated at even much earlier times from shortly after the time when inflation ended and the inflaton field reheated the universe by decaying into radiation.[13] The horizon mass at the time of radiation-matter equality was an enormous $M_H = M_{\rm eq} \sim 10^{16} M_\odot$ and we will therefore focus on PBH formation during radiation domination during these lectures, when the horizon mass was smaller.

[13] For a review article about the equation of state of the early universe see [4].

The background pressure is very large during radiation domination, $P = \omega\rho = \rho/3$, meaning that only large amplitude perturbations will have a strong enough gravitational attraction to overcome the pressure forces and collapse into a black hole. The typical amplitude of the density perturbations on CMB scales is

$$\delta = \frac{\delta\rho}{\rho} \sim \mathcal{R} \sim \sqrt{\mathcal{P}_{\mathcal{R}}} \sim \sqrt{A_s} \sim 5 \times 10^{-5},$$

which is far too small to lead to PBH formation. To form a PBH we instead need $\delta\rho/\rho \sim 1$ at the time of horizon entry and hence PBH formation requires special initial conditions. If the spectral index satisfies $n_s - 1 \simeq 0$ on all scales then the power spectrum will also be small on all scales and hence zero PBHs will form.[14]

The original estimate for the collapse threshold for PBH formation was made by Bernard Carr in 1975 (while he was Hawking's PhD student) using the Jean's length and time and using Newtonian gravity, who found that an overdensity would collapse if

$$\delta \equiv \frac{\delta\rho}{\rho}|_{k=aH} > \delta_c = c_s^2, \tag{V.1}$$

where c_s is the sound speed of perturbations, which is an important quantity because this determines how quickly a pressure wave caused by the overdensity can travel from the centre to the edge of the perturbation. During radiation domination, $c_s = 1/\sqrt{3}$ so $\delta_c = c_s^2 = \omega = 1/3$. Both one-dimensional, and even recent three-dimensional GR simulations have shown that

$$\delta_c \simeq 0.45, \tag{V.2}$$

which is quite close to Carr's original estimate, and they have also shown that the collapse threshold has only a mild dependence on the initial density profile. However, (V.1) does not remain accurate in the limit of a matter dominated universe with $c_s \to 0$ because his estimate assumed the initial overdensity was spherically symmetric. Whilst this is a good approximation for rare peaks in the density field [6], this is not valid for small values of c_s.

[14] Note that there are other mechanisms which could form PBHs, for example the collisions of cosmic strings. However, all of these alternatives require new (beyond the standard model) physics to form, and we won't consider them further. For a brief review see [5].

We would now like to estimate the amplitude of $\mathcal{P}_{\mathcal{R}}$ which can generate an 'interesting' number of PBHs. As is normal for models of DM production, we only need a tiny fraction of the total energy density to be in the form of PBHs at the time of PBH formation in order to get a significant fraction of the DM to consist of PBHs, where this fraction is normally parametrised by

$$f_{\rm PBH} \equiv \frac{\rho_{\rm PBH}}{\rho_{\rm DM}}, \tag{V.3}$$

where $f_{\rm PBH} = 1$ means that all of the DM is made out of PBHs and therefore DM would not be a new particle.

During radiation domination $\rho_{\rm tot} = \rho_{\rm rad} \propto a^{-4}$ while after formation $\rho_{\rm PBH} \propto a^{-3}$ and therefore the PBH fraction $\rho_{\rm PBH}/\rho_{\rm tot}$ grows proportional to the scale factor a from the time of formation until matter-radiation equality. Denoting the fraction of the universe's energy density in PBHs at the time of formation as β, we therefore have

$$f_{\rm PBH} = \frac{\rho_{\rm PBH}}{\rho_{\rm DM}}|_0 \simeq \frac{\rho_{\rm PBH}}{\rho_{\rm tot}}|_{\rm eq} \simeq \frac{a_{\rm eq}}{a_{\rm form}}\beta. \tag{V.4}$$

PBH formation is 'fast' because the initial over density is so huge, unlike galaxy formation which takes billions of years because they form from a tiny initial overdensity with $\delta_{\rm initial} \sim 10^{-4}$. PBH formation takes about ten Hubble times, meaning that the collapse occurs at a time $10/H_{k=aH}$ where $1/H_{k=aH}$ is the Hubble time when the mode re-entered the horizon. Notice that this corresponds to the universe growing by only about e-folding during radiation, because $a \propto t^{1/2}$ during this time and therefore $\Delta N_{\rm PBH\ formation} \sim \ln(10)/2 \sim 1$. We can therefore approximate the time of PBH formation as being equal to the time when the mode of the overdensity which will form the PBH re-enters the horizon, i.e. the time when $k = aH$.

As stated before, the PBH mass is comparable to the horizon mass at the time of formation, so

$$M_{\rm PBH} \sim M_H = \rho V = \frac{4}{3}\pi\rho\left(\frac{1}{H}\right)^3 \propto \rho^{-1/2} \propto a^2 \propto t \tag{V.5}$$

where it is important to realise we used the physical (not comoving) Hubble scale $1/H$ as an estimate of the radius of the volume V and we also used $H^2 \propto \rho \propto a^{-4}$ to derive some of the relations in the above equation. Note that $M_{\rm PBH} \propto a^2$ rather than a^3

because the density is decreasing while the horizon volume increases. Inserting the numerical factors one can find

$$M_{\rm PBH} = \left(\frac{a_{\rm form}}{a_{\rm eq}}\right)^2 M_{\rm eq} \simeq \left(\frac{a_{\rm form}}{a_{\rm eq}}\right)^2 10^{16} M_\odot, \tag{V.6}$$

$$M_{\rm PBH} \sim 10^{15}{\rm g}\,\frac{t}{10^{-23}{\rm s}}. \tag{V.7}$$

The reason why the mass to formation time is often written in the form seen above[15] is because PBHs with an initial mass of 10^{15}g will be evaporating today due to Hawking evaporation of the black hole, while those which form with a smaller initial mass will have already evaporated completed and we can neglect the impact of Hawking radiation for BHs those which form with a significantly larger initial mass.

In terms of the comoving wavenumber k measured at horizon entry after inflation,

$$k = aH \propto t^{1/2} \propto a^{-1} \quad \Rightarrow \quad M_H \propto k^{-2} \tag{V.8}$$

and inserting numerical factors leads to

$$M_{\rm PBH} \simeq M_H \sim 10^{13} M_\odot\, k^{-2}\,{\rm Mpc}^2. \tag{V.9}$$

For the case of a solar mass PBH, $M_{\rm PBH} = M_\odot = 2 \times 10^{33}$g, we can make order of magnitude estimates that they form when

$$k \sim 10^7{\rm Mpc}^{-1}, \;\; {\rm t} \sim 10^{-6}{\rm s}, \;\; {\rm a_{form}} \sim 10^{-8}{\rm a_{eq}}. \tag{V.10}$$

The corresponding energy at this time is about 200 MeV which corresponds to the time of the QCD transition when quarks bind into hadrons. We will see the significance of this coincidence later. Because $a_{\rm form} \sim 10^{-8} a_{\rm eq}$ only 10 parts per billion of the universe needs to be in the form of PBHs at the formation time (i.e. $\beta \sim 10^{-8}$) in order for all of the DM to be made out of solar mass PBHs.

B. Estimating the collapse fraction β

The easiest method is to use the Press-Schechter formalism, in which the collapse fraction of the universe into PBHs at the time of formation (or horizon entry of the

[15] For example, see the very first equation in the PBH review article by Green and Kavanagh [7].

relevant mode) is estimated by calculating the fraction of the universe with $\delta > \delta_c$,

$$\beta(M_{\rm PBH}) = \frac{\rho(M_{\rm PBH})}{\rho_{\rm tot}}|_{\rm formation} = \int_{\delta_c}^{\infty} P(\delta) d\delta, \tag{V.11}$$

where P is the pdf, not to be confused with the power spectrum.

For simplicity in these lectures, we will assume $\mathcal{R} = \delta$ and hence they have the same power spectrum, although we caution that this is not a very accurate approximation. Technically one should also use a window function to smooth the density contrast δ on the scale of PBH formation, $R \sim 1/k$, whose variance is related to the primordial power spectrum of the density contrast by

$$\sigma^2(R) = \int_0^{\infty} \widetilde{W}^2(kR)\mathcal{P}_\delta(k) d\ln k, \tag{V.12}$$

where $\widetilde{W}$ is the Fourier transform of a real space window function and $\mathcal{P}_\delta(k)$ is the dimensionless power spectrum of the dimensionless density perturbation.

For simplicity we will neglect these complications and instead use

$$\sigma = \mathcal{P}_{\mathcal{R}}$$

as a rough estimate. Then if we assume that the perturbations are Gaussian distributed with variance σ^2, and using the fact that $\beta \ll 1$, we can make an asymptotic expansion in the limit $\delta_c/\sigma \gg 1$ to show that

$$\beta \simeq \frac{1}{2}{\rm erfc}\left(\frac{\delta_{\rm c}}{\sqrt{2}\sigma}\right) \simeq \frac{\sigma}{\sqrt{2\pi}\delta_{\rm c}} {\rm e}^{-\delta_{\rm c}^2/(2\sigma^2)}. \tag{V.13}$$

As a **very** rough estimate, we can invert this to estimate

$$\mathcal{P}_{\mathcal{R}} \sim \sigma^2 \sim \frac{\delta_c^2}{\ln(1/\beta)} \sim \frac{0.2}{\ln(1/\beta)}. \tag{V.14}$$

Notice that $\mathcal{P}_{\mathcal{R}}$ is only logarithmically sensitive to β, whilst instead β is exponentially sensitive to the power spectrum amplitude. Hence, relatively small changes to δ_c, or the relation between $\mathcal{R}$ and δ, or changes to the choice of the window function all lead to exponentially large changes in β. Normally such changes are unimportant, despite appearing to change the answer by orders of magnitude.

Consider again the case of solar mass PBHs and assuming that $f_{\rm PBH} = 1$, it is straightforward to estimate that the required power spectrum amplitude is

$$\mathcal{P}_{\mathcal{R}} \sim \frac{0.2}{\ln(10^8)} \simeq 0.011 \sim 10^{-2}. \qquad \text{(V.15)}$$

For much smaller PBH masses the relevant values of β also become much smaller (due to the longer period of expansion between PBH formation and radiation-matter equality). The tightest observational constraint on β is $\beta \lesssim 10^{-28}$ for $M_{\rm PBH} \sim 10^{15}$g [8], i.e. the PBH initial mass which corresponds to them decaying today and hence having a large observational signature through Hawking radiation. How much impact does this twenty order-of-magnitudes tighter constraint have on the consequent constraint on the power spectrum amplitude? The answer is not much, only by a factor of $28/8 = 3.5$. Hence the range of power spectrum amplitudes which are of interest for PBH formation is quite limited, lying in the range $\mathcal{P}_{\mathcal{R}} \sim 10^{-3} - 10^{-2}$, independently of the PBH mass and for any potentially observable value of $f_{\rm PBH}$. However, primordial non-Gaussianity or an early matter dominated epoch can strongly change the constraints on the power spectrum amplitude, but a study of these topics goes beyond the scope of these lectures.

How sensitive are these results to changes in the collapse threshold δ_c? Phrased in terms of β the answer appears to be huge, for example if $\sigma^2 = 3 \times 10^{-3}$ (the value relevant for the formation of PBHs which are decaying today) then

$$\frac{\beta(\delta_c = 1/3)}{\beta(\delta_c = 0.45)} \simeq 10^{-6}.$$

However, the change in the value of $\mathcal{P}_{\mathcal{R}}$ required to get a particular value of β, is only

$$\frac{\sigma^2(\delta_c = 1/3)}{\sigma^2(\delta_c = 0.45)} \simeq \frac{0.33^2}{0.45^2} \simeq 0.5.$$

Authors of PBH papers like to phrase the difference in terms of β in order to make the importance of their results exponentially large, but in reality, this normally overstates the importance of the change.

One exception is during the QCD transition when the equation-of-state parameter ω drops from 1/3 to 0.25, about a 25% reduction, at the time when the horizon mass is one solar mass [9]. This reduction in the background pressure makes PBH formation 'easier', and the collapse threshold drops from $\delta_c = 045$ to $\delta_c \simeq 0.4$, a reduction by about 10% [10]. This relatively small reduction leads to a 2–3 orders-of-magnitude enhancement in the production of solar mass PBHs compared to the number of PBHs on similar mass scales where $\omega = 1/3$ and $\delta_c = 0.45$, provided that the amplitude of the primordial power spectrum is sufficiently large and constant over the relevant range of scales. This mass range is of special interest both because they are below the Chandrasekhar mass (about 1.4 solar masses) which is the smallest mass with which a compact object (a neutron star or BH) can form in the late universe through standard astrophysical processes and also because this mass range can be probed by ground based gravitational wave detectors such as the current LIGO and Virgo instruments. The observation of a sub solar-mass compact object would be a smoking gun signature for a PBH and the detection of even just one such object would have huge implications for our understanding of dark matter and the physics of the very early universe.

Lecture 5

VI. OBSERVATIONAL SEARCHES FOR PBHS

When studying PBHs as a DM candidate the constraints are normally best phrased in terms of $f_{\rm PBH}$ but for PBHs which are currently evaporating, or which have already evaporated it is normal to use constraints in terms of the initial collapse fraction β. Recall that neglecting accretion and evaporation the two are related by

$$f_{\rm PBH} \simeq \frac{a_{\rm eq}}{a_{\rm form}}\beta \tag{VI.1}$$

and this is sometimes used to define $f_{\rm PBH}$ even in the case that the PBHs have evaporated, even though this means that $\rho_{\rm PBH} = 0$ today. Roughly speaking, the observational constraints can be divided into two categories:

1. Gravitational constraints which more directly relate to $f_{\rm PBH}$ constraints.

2. Hawking evaporation constraints which are best phrased in terms of β.

1. Microlensing constraints

If a PBH (or any other sufficiently compact object) passes close to the line of sight to a more distant luminous object, then the light from the source will be gravitationally focused and enhanced. Hence, PBHs can make stars look brighter while they are close to the line of sight. The strength of the lensing magnification is larger for larger mass objects and also largest when the lensing object passes closest to the line of sight to the luminous source. Therefore, the total duration of the luminosity enhancement depends on the mass of the compact object as well as its velocity transverse to the line of sight, with the timescale of the magnification signal varying from a few hours for a $10^{-6}M_\odot$ compact object to a timescale of months for a $10M_\odot$ mass object.

Historically searches for compact objects focused on repeating observations of the same patch of the sky every night and repeating the process for days or years, so these surveys such as OGLE and EROS were most sensitive to compact objects in the mass range $10^{-6} - 1M_{\odot}$ and they have found a few lensing events, but not more than would (probably) be expected from compact objects created by standard astrophysical processes such as freely floating planets, and the overall constraint in the mass range they could probe was $f_{\rm PBH} \lesssim 0.1$.

Recently the Hyper-Suprime Cam (HSC) on the Subaru telescope made a very detailed search over just one night for low mass lenses, and this greatly improved the constraints to the lower mass end, being sensitive enough to provide the constraint of $f_{\rm PBH} \lesssim 10^{-2}$ for $M_{\rm PBH} \sim 10^{-9} M_{\odot}$ and ruling out $f_{\rm PBH} = 1$ for the mass range $10^{-12} M_{\odot} \lesssim M_{\rm PBH} \lesssim 10^{-6} M_{\odot}$. Originally the HSC collaboration claimed to have constrained even smaller mass compact objects, but this has now been accepted not to be correct because of the finite source effect, which basically means that once the apparent size in the sky of the lensing object becomes comparable to the apparent size of the object being lensed then gravitational lensing is only effective on part of the surface of the lensed star and not the entire surface, which means that the relative enhancement in the luminosity becomes too small to be detectable [11].

2. Other gravitational constraints

For larger masses (more than about a solar mass) there are two other constraints which have been considered for a long time, accretion and DM discreteness effects, plus a much newer gravitational wave constraint.

The accretion of gas onto black holes emits high energy radiation which can be detected, for example there was recently a lot of publicity about the event horizon telescope "image" of a supermassive black hole.[16] Accretion involves highly non-linear physics and is therefore hard to model robustly, but the constraints from accretion (both at $z = 0$ or during recombination when the CMB formed) appear to rule out $f_{\rm PBH} = 1$ for $M_{\rm PBH} \gtrsim$ few $\times M_{\odot}$ and the constraints become much tighter for larger masses, e.g. $f_{\rm PBH} \lesssim 10^{-4}$ for $M_{\rm PBH} \gtrsim 10^{2} M_{\odot}$. In general, accretion is not expected to be significant for PBHs with an initial mass much below ten solar masses, and hence

[16] For a review of astrophysical black holes see [12].

light PBHs are expected to have an essentially constant mass (unless they are so small that Hawking evaporation is important) and hence also a constant spin [13].

Discreteness effects (sometimes called dynamical constraints) are caused by very massive PBHs not looking like a 'smooth' density field on small scales. For example, dwarf galaxies with masses $10^7 - 10^9 M_\odot$ could be modified or even destroyed if DM was made out of PBHs with very large masses. In practise these constraints are not as tight as the accretion constraints and hence are not so widely discussed.

The gravitational wave bound is based on the LIGO and Virgo observations of merging compact objects. These detectors are most sensitive in the $10^2 - 10^3 M_\odot$ mass range and hence the tightest constraints are also found in this mass range, with the constraint being $f_{\rm PBH} \lesssim 10^{-3}$ which is the tightest constraint in this mass range, but it is also quite model dependent since it is not straightforward to estimate the current merger rate of a large population of PBHs. The standard calculation assumes that PBHs form binary pairs in the very early universe, shortly after formation and long before matter-radiation equality, and that many of these binary pairs remain relatively undisturbed until today. Estimating the disruption rate of these 'primordial' binaries pairs is a numerically challenging task but there appears to be a consensus that although disruption is a very important effect when $f_{\rm PBH} \simeq 1$ and hence there are many PBHs which can disrupt each other, that for $f_{\rm PBH} \lesssim 10^{-2}$ that disruption becomes relatively rare [14]. Requiring that all of the observed LIGO Virgo merger events were due to primordial black holes requires $f_{\rm PBH} \simeq 3 \times 10^{-3}$ (assuming a lognormal mass distribution with a central mass around the $20 M_\odot$ but this scenario is strongly disfavoured compared to the alternative (standard) scenario that all of the mergers are due to astrophysical black holes [15]. A mixed scenario with astrophysical and a subdominant population of primordial black hole mergers works [16, 17]. In practise astrophysical models of black hole formation and merger are sufficiently uncertain that it would be very hard to prove any specific merger of two compact objects was caused by primordial black holes unless at least one of the component masses is clearly below the Chandrasekhar mass, which is the lowest mass compact object that can form through standard astrophysical processes.

To the lower mass range than the microlensing lower bound there is a relatively wide mass window where $f_{\rm PBH} = 1$ is possible, which is from $10^{17} - 10^{22}$ g or equivelantly $10^{-16} - 10^{-12} M_\odot$. This window, in which all of the DM could consist of PBHs, is sometimes referred to as the 'asteroid mass window'.

Although the constraints discussed here are based on an unrealistic monochromatic (single) mass function, in practice the constraints do not change by more than a factor of order unity when considering more realistic and broader mass functions [18]. Hence there is a reasonable (but not complete) agreement that even a broad mass function would not allow for all the DM being made out of PBHs with masses far above the 'asteroid' mass range.

3. Evaporation constraints and PBH relics

Black holes are not perfectly black when taking quantum mechanical effects into account, and in fact they radiate energy away with a temperature set by the Hawking radiation from a black hole, which satisfies

$$T \propto M_{\rm PBH}^{-1}.$$

Recalling that the radius of a black hole is proportional to its mass, meaning its surface area is proportional to $M_{\rm PBH}^2$, and given that the energy radiated follows a blackbody distribution with total energy proportional to T^4 per unit area, the total energy radiated away is proportional to the area times T^4, which is proportional to $M_{\rm PBH}^{-2}$. Using $E = mc^2$ we can deduce that the rate of mass loss from the black hole satisfies

$$\frac{dM_{\rm PBH}}{dt} \propto \frac{1}{M_{\rm PBH}^2}$$

and this can be integrated to find that the lifetime of a BH satisfies

$$t_{\rm evap} \propto M_{\rm PBH}^3.$$

Inserting numerical factors one can check that the evaporation timescale equals the age of the universe for a PBH with initial mass $M_{\rm PBH} \simeq 10^{15}$ g, and that the Hawking evaporation is non-negligible today if the initial mass satisfies $M_{\rm PBH} \lesssim 10^{17}$ g. PBHs which decay today or during the time of recombination when the CMB formed are very tightly constrained by observations, and there are also constraints on the allowed decay of PBHs which would decay during BBN. However, the decay of PBHs before BBN begins (corresponding to those with initial mass $M_{\rm PBH} \lesssim 10^{10}$ g) is extremely hard to constrain using any observations.

When the PBH comes close to completely evaporating the energy being evaporated comes close to the Planck energy at the time the mass becomes comparable to the Planck mass. At this point the semi-classical physics used to derive the Hawking temperature might not remain valid and it remains an open question whether black holes evaporate completely or whether a relic remains. If a relic remains it would presumably have approximately a Planckian mass. If Planck mass relics can form then they could be a DM candidates, potentially a 'nightmare' DM scenario in which the DM is too hard to detect by any known technology. However, there has recently been a theoretical argument made that relics would gain a large peculiar velocity during the decay process and this could rule them out as a cold dark matter candidate [19].

Figure 4 shows a summary of the constraints on $f_{\rm PBH}$ over a wide range of masses. For more details of the observational constraints see [7] and references therein, or for a comprehensive list of constraints phrased in terms of β see [20].

A. Constraints on the primordial power spectrum

In this section we will briefly summarise the primary alternative methods to constraining the power spectrum amplitude. On scales larger than about a Mpc, $k \lesssim {\rm Mpc}^{-1}$, observations of the CMB and LSS have provided an accurate measurement of the amplitude. Over a huge range of smaller scales, PBHs provide a weaker constraint as discussed in section V B.

1. *Spectral distortions*

Because the early universe before recombination was in thermal equilibrium, the CMB photons follow a black body distribution. Confirmation of this fact by the

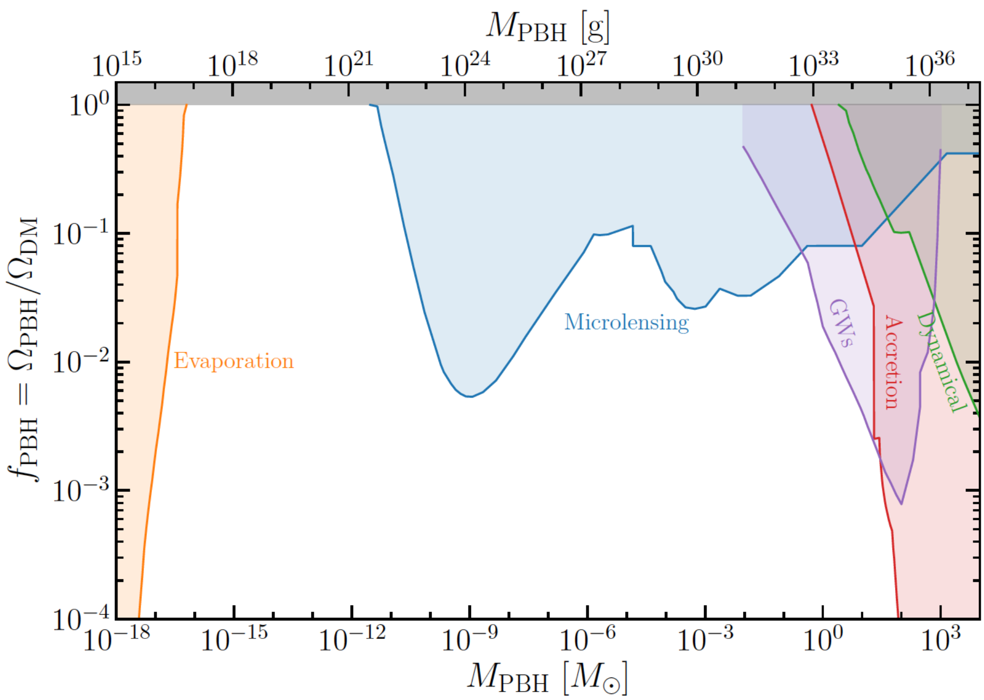

FIG. 4. Some of the key constraints on $f_{\rm PBH}$ showing which types of constraints are most important over a wide range of mass scales. Taken from the review article [7].

COBE FIRAS instrument was a very important piece of evidence in favour of the Big Bang theory. However, the damping of large amplitude density perturbations during certain redshifts/temperatures would act as an energy injection into the baryon-photon plasma, which could lead to a deviation from thermal equilibrium and hence a deviation from a blackbody spectrum [21].

The relevant constraint from the non-detection of a cosmic-μ distortion based on the COBE FIRAS results are roughly $\mathcal{P}_\mathcal{R} \lesssim 10^{-4}$ in the range $k \sim 1 - 10^4$ Mpc^{-1}, where the smallest constrained scale corresponds to a PBH mass of about $10^4 M_\odot$. Hence, PBHs cannot have a larger mass than this if they were formed by the collapse of large amplitude perturbations shortly after horizon entry, unless the perturbations were strongly non-Gaussian or if there was an early matter dominated epoch taking place while the relevant scales were entering the horizon.

2. Gravitational waves

As you may have learnt in an introductory cosmology course, at linear order the scalar, vector and tensor perturbations all decouple. On large scales, the perturbations observed via the CMB are so small that we know linear perturbation theory is an excellent approximation. However, on smaller scales where the perturbation amplitude becomes much larger, there could be a significant non-linear coupling between different types of perturbations. Of particular interest are the second-order tensor perturbations which are generated by the square of linear scalar perturbations. The full equations showing the coupled evolution of the non-linear perturbations are very complicated, but we can write their order of magnitude schematically as $h^{(2)} \sim \mathcal{R}^2$, where $h^{(2)}$ denote the intrinsically second-order tensor (gravitational wave) perturbations. These induce a power spectrum amplitude of the tensor perturbations given by

$$\mathcal{P}_T \sim \left(h^{(2)}\right)^2 \sim \mathcal{R}^4 \mathcal{P}_{\mathcal{R}}^2. \tag{VI.2}$$

The corresponding frequency of the waves is given in terms of the scale at horizon entry, $k = aH$, by

$$f \sim ck.$$

These second-order tensor perturbations could appear as a stochastic background of gravitational waves. By a nice coincidence the scale corresponding to a horizon mass of 1 solar mass (and the QCD transition) has a frequency in the range which pulsar timing arrays (PTA) can constrain, and the current PTA constraints on the amplitude of the primordial scalar power spectrum are at almost exactly the same amplitude as is required to generate PBHs with this mass. Therefore, if LIGO and Virgo have or do detect any PBHs then we should expect to see a corresponding signal of stochastic gravitational waves. The coincidence of scales and constraints is shown in the upper plot of figure 5. We note that the NANOGrav collaboration have recently reported a detection of excess noise in pulsar timing residuals, but they don't have a strong enough signal to determine whether this data is caused by gravitational waves [22].

It is also of interest to note that if the DM does consist of asteroid mass PBHs then the LISA space based gravitational wave detector, due to launch in the mid 2030's, is expected to detect an associated signature of stochastic gravitational waves at high significance [23]. The lower plot of figure 5 shows this forecasted sensitivity of the LISA instrument, as well as the Einstein Telescope (ET) and the Square Kilometre Array (SKA) constraints on pulsar timing observations. The forecasted constraints on the μ-distortions are based on assuming a PIXIE like survey [24].

B. Possible signatures

Every observational constraint is also a possible signature. However, whilst some signatures could be relatively clearly identified as being due to a PBH, others would be much harder to interpret. For example, a detection of microlensing events could be due to freely floating planets or other astrophysical objects, and it is hard to make a robust estimate of how many such objects could form out of baryonic objects. Likewise, a detection of gamma rays, which could be due to evaporating PBHs, could also be caused by high energy astrophysics or either decaying or annihilating dark matter particles. Nonetheless, there are some tentative hints of PBH signatures in existing data, see [26] for an overview.

The most promising direct detection signature which is accessible with current instruments would be the discovery for a sub-solar mass compact object [27].[17] If such low mass PBHs are not detected, an alternative direct gravitational wave probe - in the far future with a detector such as the Cosmic Explorer - would be the detection of a very high redshift merger ($z \gtrsim 40$), which would be a signature from such early times that stellar objects would not yet have had time to collapse into compact objects [28].

A less direct, but still promising, gravitational wave probe is via the stochastic background of gravitational waves generated at high redshift. The most promising current probe are the PTA searches which form a synergy with the characteristic mass of the LIGO-Virgo detections and the QCD transition, as shown in the upper plot of figure 5. This opens the possibility that an analysis of the frequency dependence of the PTA gravitational wave background (if detected and not astrophysical in origin)

[17] For a popular science article about this, see `https://www.quantamagazine.org/black-holes-from-the-big-bang-could-be-the-dark-matter-20200923/?fbclid=IwAR2GgvelVyYAEkvZTfitIVjAEgwnvTwpgp6TNHfNrtS1SHin6hX6Gy7L5BY`.

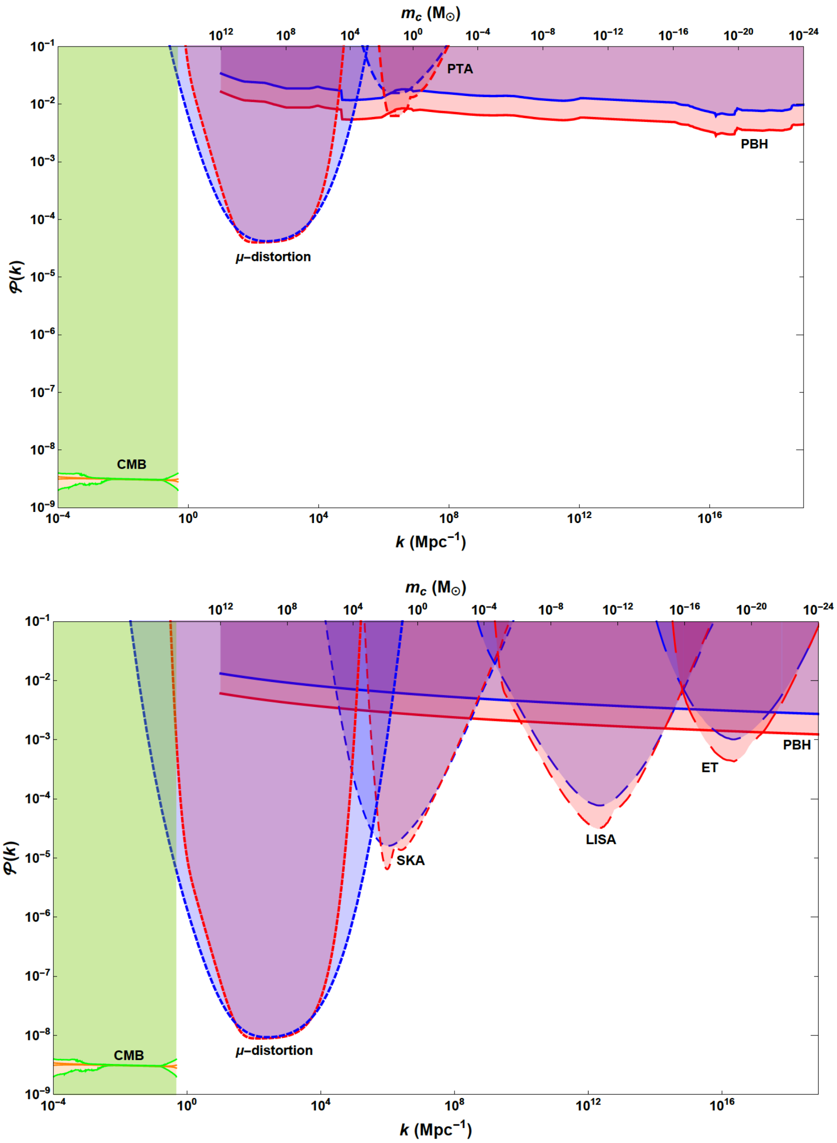

FIG. 5. Constraints on the amplitude of the primordial power spectrum. On both figures the left hand lines show the amplitude detected by current CMB observations. All other shaded regions show upper bounds on the amplitude, focusing on current constraints in the upper plot and future constraints in the lower plot. For both plots the blue lines (which are weaker for PBH constraints and broader for all other constraints) are based on a narrowly peaked power spectrum, while the red lines show the equivelant constraints for a power spectrum with a broader peak. The PBH constraint lines in the lower plot are the best theoretically possible constraints, based on zero PBHs forming inside the observable universe. The top horizontal axis, m_c, shows the central PBH mass corresponding to each value of k shown in the lower axis. For more details see [25].

could be tested against the corresponding mass range of LIGO and other ground based gravitational wave detectors.

In the longer term (on a time scale of a few decades) the space-based LISA gravitational wave detector should see a clear signal of a stochastic gravitational wave background if the DM is made out of asteroid mass PBHS. Recalling the logarithmic sensitivity of the scalar power spectrum amplitude to $f_{\rm PBH}$ (and hence also the associated tensor power spectrum) it becomes clear that the non-detection of a primordial stochastic background by LISA would not only rule out $f_{\rm PBH} = 1$ but even the formation of any PBHs at all in the same mass range.[18] Given that the lower plot of figure 5 shows the upper bound on the power spectrum amplitude below which zero PBHs would form in today's observable universe, we can see that in the future there will be almost no remaining mass windows for non-evaporated (or relic) PBHs to have formed, assuming that none of the experiments shown detect the relevant signature used to constrain the power spectrum amplitude. Hence, future searches for PBHs have a bright future, and we can realistically hope to determine whether or not DM is made out of a new particle or primordial black holes.

In summary – whether or not PBHs exist – the search for them has led to new understandings about the nature of the contents and initial conditions of the universe.

---

[1] A. R. Liddle and S. M. Leach, *How long before the end of inflation were observable perturbations produced?*, *Phys. Rev. D* **68** (2003) 103503 [astro-ph/0305263].

[2] Planck collaboration, *Planck 2018 results. X. Constraints on inflation*, *Astron. Astrophys.* **641** (2020) A10 [1807.06211].

[3] M. Sasaki, T. Suyama, T. Tanaka and S. Yokoyama, *Primordial black holes—perspectives in gravitational wave astronomy*, *Class. Quant. Grav.* **35** (2018) 063001 [1801.05235].

[4] R. Allahverdi et al., *The First Three Seconds: a Review of Possible Expansion Histories of the Early Universe*, 2006.16182.

[5] A. M. Green, *Primordial Black Holes: sirens of the early Universe*, *Fundam. Theor. Phys.* **178** (2015) 129 [1403.1198].

[18] Assuming, as is done in these lectures, that the primordial perturbations are close to Gaussian distributed and that PBHs form from the direct collapse of large amplitude density perturbations shortly after horizon entry.

[6] J. M. Bardeen, J. R. Bond, N. Kaiser and A. S. Szalay, *The Statistics of Peaks of Gaussian Random Fields*, *Astrophys. J.* **304** (1986) 15.

[7] A. M. Green and B. J. Kavanagh, *Primordial Black Holes as a dark matter candidate*, *J. Phys. G* **48** (2021) 043001 [2007.10722].

[8] B. J. Carr, K. Kohri, Y. Sendouda and J. Yokoyama, *New cosmological constraints on primordial black holes*, *Phys. Rev. D* **81** (2010) 104019 [0912.5297].

[9] S. Borsanyi et al., *Calculation of the axion mass based on high-temperature lattice quantum chromodynamics*, *Nature* **539** (2016) 69 [1606.07494].

[10] C. T. Byrnes, M. Hindmarsh, S. Young and M. R. S. Hawkins, *Primordial black holes with an accurate QCD equation of state*, *JCAP* **08** (2018) 041 [1801.06138].

[11] H. Niikura et al., *Microlensing constraints on primordial black holes with Subaru/HSC Andromeda observations*, *Nature Astron.* **3** (2019) 524 [1701.02151].

[12] C. Bambi, *Astrophysical Black Holes: A Review*, *PoS* **MULTIF2019** (2020) 028 [1906.03871].

[13] V. De Luca, G. Franciolini, P. Pani and A. Riotto, *Constraints on Primordial Black Holes: the Importance of Accretion*, *Phys. Rev. D* **102** (2020) 043505 [2003.12589].

[14] M. Raidal, C. Spethmann, V. Vaskonen and H. Veermäe, *Formation and Evolution of Primordial Black Hole Binaries in the Early Universe*, *JCAP* **02** (2019) 018 [1812.01930].

[15] A. Hall, A. D. Gow and C. T. Byrnes, *Bayesian analysis of LIGO-Virgo mergers: Primordial vs. astrophysical black hole populations*, *Phys. Rev. D* **102** (2020) 123524 [2008.13704].

[16] G. Hütsi, M. Raidal, V. Vaskonen and H. Veermäe, *Two populations of LIGO-Virgo black holes*, *JCAP* **03** (2021) 068 [2012.02786].

[17] V. De Luca, G. Franciolini, P. Pani and A. Riotto, *Bayesian Evidence for Both Astrophysical and Primordial Black Holes: Mapping the GWTC-2 Catalog to Third-Generation Detectors*, 2102.03809.

[18] N. Bellomo, J. L. Bernal, A. Raccanelli and L. Verde, *Primordial Black Holes as Dark Matter: Converting Constraints from Monochromatic to Extended Mass Distributions*, *JCAP* **01** (2018) 004 [1709.07467].

[19] S. Kováčik, *Hawking-Radiation Recoil of Microscopic Black Holes*, 2102.06517.

[20] B. Carr, K. Kohri, Y. Sendouda and J. Yokoyama, *Constraints on Primordial Black Holes*, 2002.12778.

[21] J. Chluba et al., *New Horizons in Cosmology with Spectral Distortions of the Cosmic Microwave Background*, 1909.01593.

[22] NANOGrav collaboration, *The NANOGrav 12.5 yr Data Set: Search for an Isotropic Stochastic Gravitational-wave Background*, *Astrophys. J. Lett.* **905** (2020) L34 [2009.04496].

[23] R. Saito and J. Yokoyama, *Gravitational-Wave Constraints on the Abundance of Primordial Black Holes*, *Prog. Theor. Phys.* **123** (2010) 867 [0912.5317].

[24] A. Kogut et al., *The Primordial Inflation Explorer (PIXIE): A Nulling Polarimeter for Cosmic Microwave Background Observations*, *JCAP* **07** (2011) 025 [1105.2044].

[25] A. D. Gow, C. T. Byrnes, P. S. Cole and S. Young, *The power spectrum on small scales: Robust constraints and comparing PBH methodologies*, *JCAP* **02** (2021) 002 [2008.03289].

[26] S. Clesse and J. García-Bellido, *Seven Hints for Primordial Black Hole Dark Matter*, *Phys. Dark Univ.* **22** (2018) 137 [1711.10458].

[27] R. Magee, A.-S. Deutsch, P. McClincy, C. Hanna, C. Horst, D. Meacher et al., *Methods for the detection of gravitational waves from subsolar mass ultracompact binaries*, *Phys. Rev. D* **98** (2018) 103024 [1808.04772].

[28] S. M. Koushiappas and A. Loeb, *Maximum redshift of gravitational wave merger events*, *Phys. Rev. Lett.* **119** (2017) 221104 [1708.07380].

Appendix A: The classic motivations for inflation

I here explain the three original motivations for inflations in more detail than in the main part of the lectures. This appendix is an optional addition to lecture 1.

1. The flatness problem and how inflation can solve it

The flatness problem is easier to state and understand than the horizon problem. In a nutshell, the problem is "Why is the universe so close to flat?" and this is a problem because the universe is expected to become less and less close to flat during radiation and matter domination. This is because the effective equation of state of curvature is $\omega = -1/3$ which corresponds to a dilution rate of a^{-2} as the universe expands, as can be seen directly from the behaviour of the curvature term in the Friedmann equation (II.1). Therefore, during radiation domination when $\epsilon \simeq \epsilon_r \propto a^{-4}$ and during matter domination when $\epsilon \simeq \epsilon_m \propto a^{-3}$ we should expect the curvature to become relatively more important as the universe grows, and the fact that the universe is still not dominated by the curvature term may point to a fine tuning of the initial conditions of the universe. We will estimate how much of a fine tuning this represents below.

The critical density defines the density of a spatially flat universe, i.e.

$$\rho_c \equiv 3M_{\rm Pl}^2 H^2. \tag{A.1}$$

It is useful to use this to rewrite each energy component as its contribution towards the critical energy density of the universe, i.e.

$$\Omega_i \equiv \frac{\rho_i}{\rho_c} \tag{A.2}$$

and the total density density is therefore

$$\Omega = \Omega_m + \Omega_r + \Omega_\Lambda. \tag{A.3}$$

We can then rewrite the Friedmann equation in yet another form, as

$$\frac{\kappa}{R_0^2} = a^2 H^2 \left(\Omega - 1\right) \tag{A.4}$$

where $\kappa = 0$ corresponds to a (spatially) flat universe, and we therefore see that a flat universe has $\Omega = 1$ at all times.

Recalling from equation (A.4) that the total density parameter (which determines the flatness, or otherwise of the universe) scales as

$$|\Omega - 1| \propto \frac{1}{a^2 H^2} = \begin{cases} a = t^{2/3} & \text{during matter domination} \\ a^2 = t & \text{during radiation domination} \end{cases} \tag{A.5}$$

we can see that the curvature term at the time of radiation-matter equality should have been smaller by a factor of

$$\frac{|\Omega(t_{\rm rm}) - 1|}{|\Omega_0 - 1|} \propto \frac{a_{\rm rm}}{a_0} \simeq \frac{1}{3500}. \tag{A.6}$$

Although I have neglected the impact of Λ on this estimate, recall that the scale factor at matter-Λ equality is close to unity and hence this correction is not very important.

Since the Planck satellite measurements of the position of the first peak in the CMB perturbation spectrum constrain the universe today to be within a percent of being flat, meaning that

$$|\Omega_0 - 1| \lesssim 10^{-2}$$

we can see that the universe needed to satisfy the condition that

$$|\Omega(t_{\rm rm}) - 1| \lesssim 10^{-6},$$

which looks quite finely tuned. Going back further in time, for example to the time of 1 second after the Big Bang when BBN was just beginning, the fine tuning was even worse, by a factor of $(a_{\rm rm}/a(t=1{\rm s}))^2 = 50{,}000$ years / 1 second $\simeq 1.6 \times 10^{12}$, so at this time the flatness must have satisfied

$$|\Omega(t = 1\,{\rm second}) - 1| \lesssim 10^{-18},$$

and it keeps getting more and more fine tuned as we go further back in time. Until, of course, the epoch of inflation.

During inflation the energy density of the universe decreases at a negligible rate, so instead the curvature will become less important at a rate of

$$a^{-2} = e^{-2N}$$

during inflation. Hence, provided that inflation lasted long enough to drive the curvature of the universe down to a tiny value, then all of the subsequent growth during radiation and matter domination will still not be enough to make the universe deviate significantly from being flat today.

How much inflation we need (i.e. what is the minimum required value of $N_{\rm inf}$) depends both on the energy scale of inflation, because that determines the energy scale/time when the universe becomes radiation dominated again, and also on the initial curvature of the universe before inflation begun. Given that we don't have a good theory for what preceded inflation, perhaps the most natural assumption is that just before inflation begun the curvature made an order unity contribution to the Friedmann equation, meaning that

$$|\Omega(t_i) - 1| \sim 1, \tag{A.7}$$

where t_i corresponds to the time when inflation begun. Given that the current constraint on the curvature is two orders of magnitude tighter than this, inflation must have proceeded for long enough that $|\Omega - 1|$ is a factor of hundred smaller today than it was at the time that inflation begun.

If we (unrealistically) make the energy scale of inflation to be as low as possible, the MeV scale of BBN, which corresponds to inflation ending about 1 second after the Big Bang, then we require

$$|\Omega(t_i) - 1|e^{-2N_{\rm inf}} \sim e^{-2N_{\rm inf}} \sim |\Omega(t = 1\,\text{second}) - 1| \lesssim 10^{-18} \tag{A.8}$$

which implies

$$N_{\rm inf} \gtrsim \frac{1}{2} \times 18\ln(10) \simeq 21. \tag{A.9}$$

If - as expected - radiation domination begun much earlier than the time of BBN and hence lasted for longer, then the universe expands by a larger factor during radiation domination and hence there is more time for the curvature to become important. This would require a larger number of efoldings of inflation in order that the universe remains sufficiently close to flat today.

To be concrete, let us consider the opposite extreme where inflation takes place at an energy scale of 10^{15} GeV, which is close to the grand unified theory (GUT) energy scale. This corresponds to a temperature 18 orders of magnitude larger than the temperature when BBN begun and hence the scale factor is 18 orders of magnitude smaller, meaning that the universe must be closer to flat by 36 orders of magnitude[19] compared the case when inflation ended just before BBN begins. In this case we require

$$N_{\rm inf} \gtrsim 21 + \frac{1}{2} \times 36\ln(10) \simeq 62, \tag{A.10}$$

and since inflation at a high energy scale is generally considered more natural than inflation at a low energy scale (based on theoretical arguments rather than observations), values of $N_{\rm inf} \gtrsim 50 - 60$ are normally considered the "benchmark" lower bounds.

---

[19] Recall that the curvature becomes relatively more important than radiation at a rate of a^2 during radiation domination.

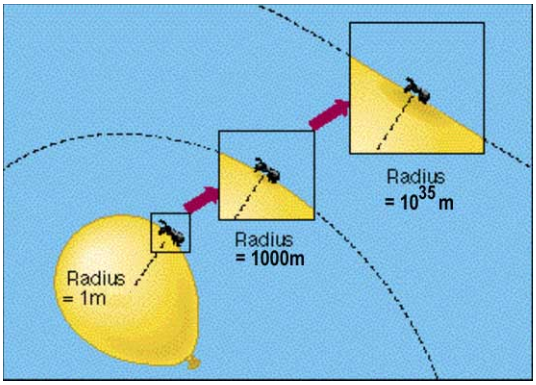

FIG. 6. A pictorial way to understand the impact of inflation. The radius of curvature is massively increased in a short time, leading the ants to be unable to see that they live on a non-flat surface. The figure is from `http://ircamera.as.arizona.edu/NatSci102/NatSci102/lectures/eraplanck.htm`

An intuitive way to understand the impact of inflation is to realise that the exponential growth of the scale factor – by a large factor of $e^{N_{\rm inf}}$ – means that the radius of curvature grows by this same huge factor, whilst the Hubble distance c/H barely increases because H is nearly constant during inflation. Therefore, thinking again of spatial curvature in two dimensions, in terms of ants walking on a curved surface and subject to a maximum local speed, it becomes clear that they will be unable to notice the impact of the global spatial curvature once the radius of curvature becomes much larger than the distance they can walk in period of time shorter than a Hubble time, and hence their universe will look flat to them. This is depicted in figure 6.

2. The horizon problem and how inflation can solve it

Although I have argued at length that the early universe was in thermal equilibrium, and hence different particle species shared a common temperature, it should be clear that reaching thermal equilibrium is a dynamical process which does not happen instantaneously. We do not need to to calculate exactly how long it takes a given region to reach thermal equilibrium, but we can at least find an absolute minimum by requiring that the process to reach thermal equilibrium is a causal process, meaning that the process cannot proceed more quickly than an upper bound set by the speed of light, which is the (local) cosmic speed limit. When calculating the maximum region that could have reached thermal equilibrium at any given time, we need to take the expansion rate of the universe into account, and the answer is given by the physical horizon distance.

The most interesting epoch to consider when thinking about the horizon problem is the last scattering surface of the CMB. This is the oldest light we can observe, and it is known to be extremely uniform in all directions, with a temperature that varies by about 1 part in 100,000 and which follows the blackbody distribution, meaning that the CMB must have formed when it was close to thermal equilibrium, locally and globally across the observable universe. Looking back to the week 9 lecture question, the physical horizon scale at last scattering was 0.25 Mpc with a corresponding angular scale on the sky of $1.1°$,[20] meaning that there is no explanation for how the temperature on two parts of the sky separated by much more than 1 degree could have been at the same temperature at the time of last scattering. Using the fact that the total area of a sphere is about 40,000 square degrees, we can see that causal processes in a universe which was radiation and matter dominated before last scattering can only explain how patches of the sky with a scale of about 1 degree reached thermal equilibrium, and hence it appears to be a truly enormous coincidence that about 40,000 causally disconnected patches of the sky all had the same temperature at the time of last scattering. Recall that although we observe the temperature of the CMB photons as measured today, which have redshifted and hence decreased by a factor of about 1,100, the last scattering surface formed at the same temperature everywhere and hence the incredible similarity of the relative CMB temperatures seen today must be reflected in the same relative (tiny) fluctuations of the temperature at the time of last scattering. This is the horizon problem.

In order to give a more concrete answer, let's consider a case where the universe was radiation dominated immediately before and after inflation[21], and that inflation

[20] Recall that we did not consider inflation when estimating these values.

[21] An instant transition is of course unrealistic, but provided that the scale factor grew by a much larger factor during inflation than during the transition to or from inflation, this approximation is unlikely to be important.

started at time t_i and ended at time t_e, such that the scale factor before radiation-matter equality satisfied

$$a(t) = \begin{cases} a_i(t/t_i)^{1/2} & \text{if } t < t_i \\ a_i e^{H_{\rm inf}(t-t_i)} & \text{if } t_i < t < t_e \\ a_i e^{H_{\rm inf}(t_e-t_i)}(t/t_e)^{1/2} = a_i e^{N_{\rm inf}}(t/t_e)^{1/2} & \text{if } t > t_e \end{cases} \tag{A.11}$$

and we used (III.2) in the final line of the equation above. The most striking fact about this solution is that for any given value of H (and hence for a given energy/temperature as well as a given Hubble time and Hubble distance) after inflation has ended, the scale factor is larger by a factor of e^N than it would be if inflation had never taken place.

Recall that the horizon distance at any general time (hence including the factor of $a(t)$ in the formula below) is given by

$$d_{\rm hor}(t) = a(t)c\int_0^t \frac{{\rm d}t'}{a(t')} \tag{A.12}$$

and hence the horizon scale at the start of inflation is

$$d_{\rm hor}(t_i) = 2ct_i = \frac{c}{H_i}$$

which is the Hubble distance at that time. By the end of inflation, the horizon distance has grown to become

$$d_{\rm hor}(t_e) = a_i e^{N_{\rm inf}} c\left(\int_0^{t_i} \frac{{\rm d}t'}{a_i(t'/t_i)^{1/2}} + \int_{t_i}^{t_e} \frac{{\rm d}t'}{a_i e^{H_{\rm inf}(t'-t_i)}}\right) \tag{A.13}$$

and in the case of interest, with $N_{\rm inf} \gg 1$, and assuming $H_{\rm inf} \sim t_i^{-1}$ we can approximate this as

$$d_{\rm hor}(t_e) \simeq e^{N_{\rm inf}} 3ct_i, \tag{A.14}$$

meaning the horizon scale has grown by more than a factor of $e^{N_{\rm inf}}$ in a time when the energy density did not significantly decrease. This is only possible when $\omega \simeq -1$.

The horizon distance after the end of inflation remains much larger than it would have been if inflation had not taken place. The horizon scale becomes

$$d_{\rm hor}(t > t_e) = a_i e^{N_{\rm inf}} c\left(\int_0^{t_i} \frac{{\rm d}t'}{a_i(t'/t_i)^{1/2}} + \int_{t_i}^{t_e} \frac{{\rm d}t'}{a_i e^{H_{\rm inf}(t'-t_i)}} + \int_{t_e}^{t} \frac{{\rm d}t'}{a_i e^{H_{\rm inf}(t_e-t_i)}(t'/t_e)^{1/2}}\right) \tag{A.15}$$

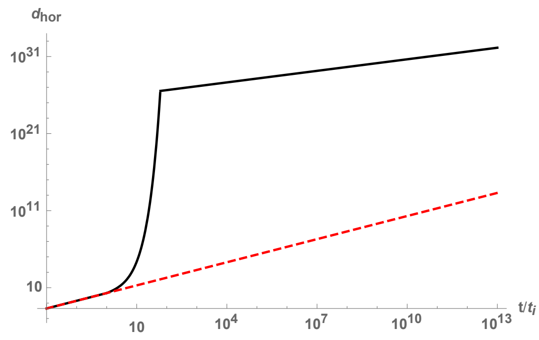

FIG. 7. A log-log plot of the horizon distance (in arbitrary units) in a radiation dominated universe with or without inflation. To make the plot I chose $N_{\rm inf} = 60$ and $H_{\rm inf} = 1/t_i$. The red-dashed line shows the result if no inflation had taken place, while the solid-black line shows the result including inflation. Notice that the two lines to not grow at the same rate after inflation has ended, the red-dashed line scales like t while the solid-black line scales like $t^{1/2}$.

and although the full expression is long, you can check that the middle integral (which is time independent) dominates over the two other integrals (assuming $N_{\rm inf}$ is sufficiently large) and hence to a good approximation the late time horizon distance (but still during radiation domination) becomes

$$d_{\rm hor}(t > t_e) \simeq \left(\frac{t}{t_e}\right)^{1/2} e^{N_{\rm inf}}\, 3\, c t_i, \tag{A.16}$$

so it scales like $t^{1/2}$ (which is slower than the scaling with t for a universe which has always been dominated by radiation) but it is boosted by about a factor of $e^{N_{\rm inf}} \ggg 1$ compared to the value it would have had if inflation had not taken place. Figure 7 shows a plot of how the horizon distance is changed by an early period of inflation, which is based on using the full expression given by equation (A.15).

Just as was the case with the flatness problem, how much inflation is needed depends on when inflation ended, and hence how long the subsequent radiation dominated epoch lasted. The rough idea is that the horizon scale at the time of last scattering (which you should recall occurs reasonably shortly after radiation-matter equality) should be big enough to explain how all the different Hubble patches at that time managed to reach the same temperature through thermalisation, which roughly means the angular scale of regions which were in causal contact at the time of last scattering should reach across the whole sky and hence have a size of 360° instead of about 1°, meaning it should be about a factor of 360 times larger.[22] A detailed calculation of exactly how much inflation is required to solve the horizon problem, as a function of the energy scale of inflation goes beyond the scope of this course, but it turns out the required amount of inflation is not very different from the amount of inflation required to solve the flatness problem, so the 'typical' value is considered to be $N_{\rm inf} \gtrsim 50$.

3. The monopole problem and how inflation can solve it

In brief, at very high energies (perhaps at the grand unified theory (GUT) scale where the electroweak and strong forces are postulated to be unified - but don't worry about this 'GUT' scale if it is not familiar) it is believed that magnetic monopoles could and would have formed. Because these particles would be heavy, they would quickly become non-relativistic and hence dilute like matter (a^{-3}) which is more slowly than the background radiation would dilute, and hence they would become increasingly important with time. Yet no magnetic monopole has ever been observed.

The way inflation solves this problem is very simple, it would dilute their number density by a factor of $e^{-3N_{\rm inf}}$ which could (for large enough values of $N_{\rm inf}$) mean that not a single magnetic monopole remains within our observable horizon today.

Of course, in order to be a successful explanation, inflation must take place at a lower energy scale than the scale associated with the production of monopoles. You might be concerned at this point that if monopoles formed before inflation then the universe would have been matter dominated instead of radiation dominated before inflation begun, in contradiction with the assumed behaviour of the scale factor in equation (A.11). That is true, but the qualitative picture would not change and even the quantitative picture would only change by (at most) a factor of order unity, so the uncertainty of what preceded inflation is not an important unknown.

[22] You might object that if the entire observable CMB sky had been in causal contact before last scattering then the position of the first peak of the CMB perturbations should also be on this much larger scale. That's a reasonable concern, but this is incorrect because the oscillations in the baryon-photon plasma can only begin around the time of BBN when the atomic nuclei have formed, and hence the horizon scale of last scattering which we calculated for that purpose (which neglected the possibility of inflation), still gave the correct estimate for that purpose.

4. A comparison of the problems inflation may have solved

In passing I want to mention that textbooks often present the horizon, flatness and monopole problems on an equal footing, suggesting they are all equally important. I would argue that is not justified, because two of the problems can be explained away relatively "easily" without invoking a period of inflation.

First of all, the monopole problem is only a problem if monopoles could really have formed, and there are only indirect theoretical arguments to suggest they should (in principle) exist. If those arguments are wrong, and they are based on extrapolating known physics up to much higher energy scales than can be experimentally tested, then the monopole problem is simply a problem of us not knowing the correct theory of very high energy physics.

Secondly, the flatness problem is only a problem if the universe is not exactly flat. But clearly it might be exactly flat, in which case $\Omega = 1$ at all times and there never was and never will be a flatness problem. I am not aware of any strong theoretical arguments to suggest the universe should or should not have global spatial curvature, and hence it seems more reasonable to suggest the universe happens to be exactly flat rather than the initial condition for the universe was so incredibly close to flat that the deviation from flatness at the time BBN begun was less than one part in 10^{16}.

However, the horizon problem is different. There is no reasonable mechanism known to explain why the universe had the same temperature everywhere at very early times, unless there was time for the early universe to reach thermal equilibrium on a large enough scale to explain why the last scattering surface has such a uniform temperature. Hence, I would single out the horizon problem as the strongest piece of evidence for inflation when only considering the homogeneous background universe.

If you demand that there were sufficient efoldings of inflation that the horizon problem is solved, it turns out that this amount of expansion also means that the flatness and monopole problems are naturally solved, assuming that they were problems which needed to be solved.

With all of these problems, notice that we have calculated how much inflation was required. The total duration of inflation may of course be longer, potentially vastly longer, but there is no observational way known to determine a reliable upper bound on the duration of inflation.