CAMBRIDGE

Cambridge University Press
978-1-107-08969-3 - Inflation and String Theory
Daniel Baumann and Liam McAllister
Excerpt
More information

1

Inflation: theory and observations

A fundamental observational fact about our universe is that on large scales it is well-described by the spatially flat Friedmann–Robertson–Walker (FRW) metric

$$ds^2 = -dt^2 + a^2(t)\, d\boldsymbol{x}^2 . \tag{1.1}$$

In Section 1.1, we first explain why the homogeneity, isotropy, and flatness of the universe encoded in (1.1) are puzzling in the standard cosmology. We then show how an early phase of quasi-de Sitter evolution drives the primordial universe towards these conditions, even if it started in an inhomogeneous, anisotropic, and curved initial state. In Section 1.2, we argue that quantum fluctuations during inflation are the origin of all structure in the universe, and we derive the power spectra of scalar and tensor fluctuations. In Section 1.3, we describe the main cosmological observables, which are used, in Section 1.4, to obtain constraints on the inflationary parameters. We then review recent experimental results. Finally, in Section 1.5, we discuss future prospects for testing the physics of inflation with cosmological observations.

1.1 Horizon problem

1.1.1 The particle horizon

To discuss the causal structure of the FRW spacetime, we write the metric (1.1) in terms of *conformal time* τ:

$$ds^2 = a^2(\tau)\left[-d\tau^2 + d\boldsymbol{x}^2\right], \tag{1.2}$$

so that the maximal *comoving distance* $|\Delta\boldsymbol{x}|$ that a particle can travel between times τ_1 and $\tau_2 = \tau_1 + \Delta\tau$ is simply $|\Delta\boldsymbol{x}| = \Delta\tau$, for any $a(\tau)$. In the standard Big Bang cosmology, the expansion at early times is driven by the energy density of radiation, and by tracing the evolution backward one finds that $a \to 0$ at sufficiently early times, and the spacetime becomes singular at this point. We choose coordinates so that the initial singularity is at $t = 0$. At some time $t > 0$, the maximal comoving distance a particle can have traversed since the initial singularity (also known as the *particle horizon*) is given by

$$\Delta\tau = \int_0^t \frac{\mathrm{d}t'}{a(t')} = \int_{-\infty}^{\ln a(t)} \frac{\mathrm{d}\ln a}{aH}\,, \qquad \text{where} \quad H \equiv \frac{1}{a}\frac{da}{dt}\,. \tag{1.3}$$

During the standard Big Bang evolution, $\ddot{a} < 0$ and the comoving Hubble radius $(aH)^{-1} = (\dot{a})^{-1}$ grows with time. The integral in (1.3) is therefore dominated by the contributions from late times. This leads to the so-called *horizon problem*. The amount of conformal time that elapses between the singularity and the formation of the cosmic microwave background (an event known as recombination) is much smaller than the conformal time between recombination and today (see Fig. 1.1). Quantitatively, one finds that points in the CMB that are separated by more than one degree were never in causal contact, according to the standard cosmology: their past light cones do not overlap before the spacetime is terminated by the initial singularity. Yet their temperatures are observed to be the same, to one part in 10^4. Moreover, the observed temperature fluctuations are *correlated* on what seem to be superhorizon scales at recombination. Not only must we explain why the CMB is so uniform, we must also explain why its small fluctuations are correlated on apparently acausal scales.

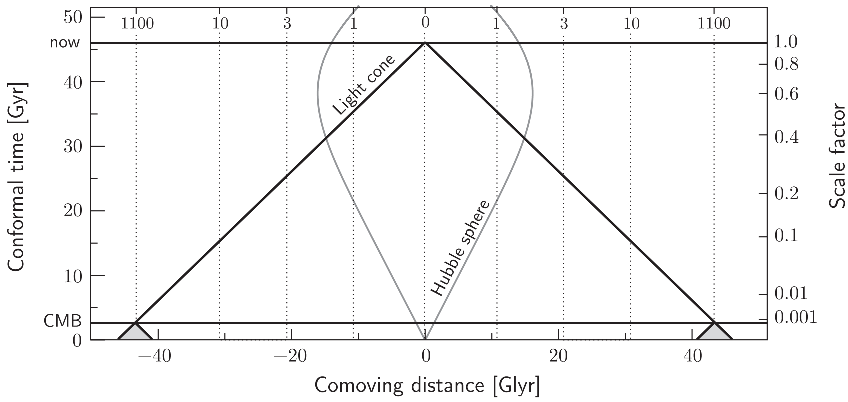

Fig. 1.1 Spacetime diagram illustrating the horizon problem in comoving coordinates (figure adapted from [52]). The dotted vertical lines correspond to the worldlines of comoving objects. We are the central worldline. The current redshifts of the comoving galaxies are labeled on each worldline. All events that we currently observe are on our past light cone. The intersection of our past light cone with the spacelike slice labeled CMB corresponds to two opposite points on the CMB surface of last-scattering. The past light cones of these points, shaded gray, do not overlap, so the points appear never to have been in causal contact.

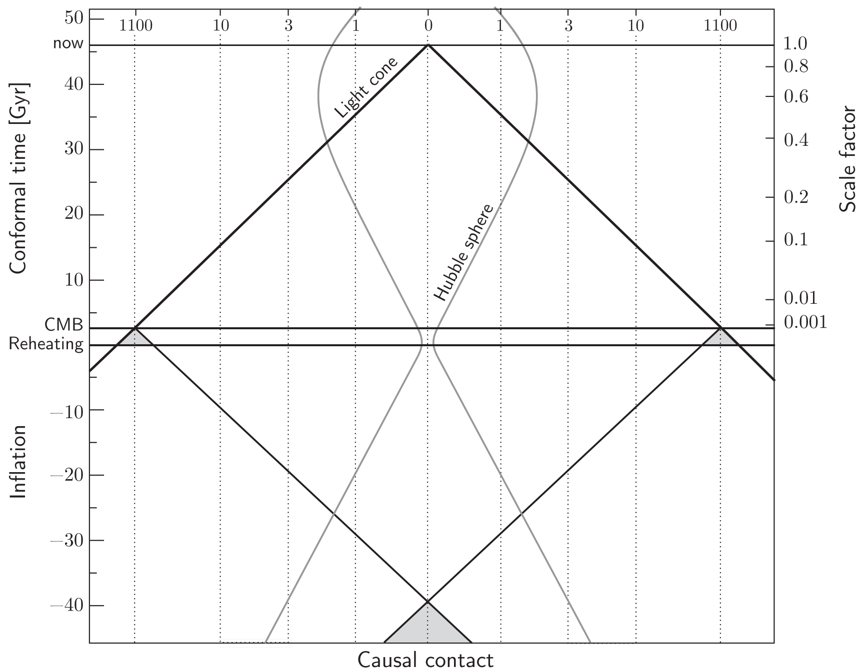

Fig. 1.2 Inflationary solution to the horizon problem. The comoving Hubble sphere shrinks during inflation and expands during the conventional Big Bang evolution (at least until dark energy takes over). Conformal time during inflation is negative. The spacelike singularity of the standard Big Bang is replaced by the reheating surface: rather than marking the beginning of time, $\tau = 0$ now corresponds to the transition from inflation to the standard Big Bang evolution. All points in the CMB have overlapping past light cones and therefore originated from a causally connected region of space.

1.1.2 Cosmic inflation

To address the horizon problem, we may postulate that the comoving Hubble radius was decreasing in the early universe, so that the integral in (1.3) is dominated by the contributions from early times. This introduces an additional span of conformal time between the singularity and recombination (see Fig. 1.2): in fact, conformal time now extends to negative values. If the period of decreasing comoving Hubble radius is sufficiently prolonged, all points in the CMB originate from a causally connected region of space. The observed correlations can therefore result from ordinary causal processes at early times.

In an expanding universe, a shrinking comoving Hubble sphere implies

$$\frac{d}{dt}(aH)^{-1} = -\frac{1}{a}\left[\frac{\dot{H}}{H^2}+1\right] < 0 \qquad \Rightarrow \qquad \varepsilon \equiv -\frac{\dot{H}}{H^2} < 1 \ . \tag{1.4}$$

We will take the slow evolution of the Hubble parameter, $\varepsilon < 1$, as our definition of inflation. This definition includes, but is not limited to, the dynamics of a slowly rolling scalar field (see Section 2.2.1). In the de Sitter limit, $\varepsilon \to 0$, the space grows exponentially,

$$a(t) \propto e^{Ht} , \tag{1.5}$$

with $H \approx const.$

Inflationary expansion requires a somewhat unconventional matter content. In a spatially flat FRW universe supported by a perfect fluid, the Einstein equations lead to the Friedmann equations,

$$3M_{\rm pl}^2 H^2 = \rho \, , \tag{1.6}$$

$$6M_{\rm pl}^2(\dot{H} + H^2) = -(\rho + 3P) \, , \tag{1.7}$$

where ρ and P are the energy density and pressure of the fluid. Combining (1.6) and (1.7), we find

$$2M_{\rm pl}^2 \dot{H} = -(\rho + P) \, , \tag{1.8}$$

and hence

$$\varepsilon = \frac{3}{2}\left(1 + \frac{P}{\rho}\right) \, . \tag{1.9}$$

Inflation therefore occurs when $P < -\frac{1}{3}\rho$, corresponding to a violation of the strong energy condition (SEC), which for a perfect fluid states that $\rho + P \geq 0$ and $\rho + 3P \geq 0$. One simple energy source that can drive inflation is a positive potential energy density of a scalar field with negligible kinetic energy, but we will encounter a range of alternative mechanisms.

1.2 Primordial perturbations

> With the new cosmology the universe must have been started off in some very simple way. What, then, becomes of the initial conditions required by dynamical theory? Plainly there cannot be any, or they must be trivial. We are left in a situation which would be untenable with the old mechanics. If the universe were simply the motion which follows from a given scheme of equations of motion with trivial initial conditions, it could not contain the complexity we observe. Quantum mechanics provides an escape from the difficulty. It enables us to ascribe the complexity to the quantum jumps, lying outside the scheme of equations of motion. The quantum jumps now form the uncalculable part of natural phenomena, to replace the initial conditions of the old mechanistic view.
>
> *P. A. M. Dirac* [53].

Inflation not only explains the homogeneity of the universe, but also provides a mechanism to create the primordial inhomogeneities required for structure formation [12–17]. This process happens automatically when we treat the inflationary de Sitter phase quantum mechanically. Here, we briefly sketch the quantum generation of primordial fluctuations. We also present the modern view of inflation as a symmetry-breaking phenomenon [50, 51]. For more details, see Appendices B and C.

1.2.1 Goldstone action

By definition, inflation is a transient phase of accelerated expansion, corresponding approximately, but not exactly, to a de Sitter solution. In order for inflation to end, the time-translation invariance present in an eternal de Sitter spacetime must be broken. The slow evolution of the Hubble parameter $H(t)$ serves as a clock that measures the progress of inflation, breaking time-translation invariance and defining a preferred time slicing of the spacetime. The isometries of de Sitter space, $SO(4,1)$, are spontaneously broken down to just spatial rotations and translations. It is often useful to think of the time slicing as being defined by the time-dependent expectation values $\psi_m(t)$ of one or more bosonic fields ψ_m (see Fig. 1.3).

As with spontaneously broken symmetries in flat-space quantum field theory (see e.g. [54]), the broken symmetry is *nonlinearly realized* by a Goldstone boson. Focusing on symmetry breaking and on the physics of the Goldstone boson allows a model-insensitive description of fluctuations during inflation [51]. In particular, we can defer consideration of the dynamics that created the background evolution $H(t)$, though ultimately we will return to explaining the background.

The Goldstone boson associated with the spontaneous breaking of time translation invariance is introduced as a spacetime-dependent transformation along the direction of the broken symmetry, i.e. as a spacetime-dependent shift of the time coordinate [50],

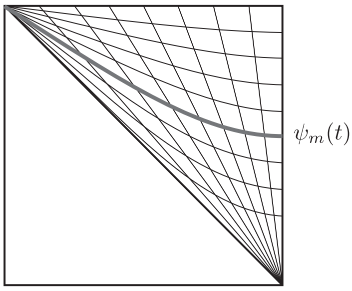

Fig. 1.3 Time-dependent background fields $\psi_m(t)$ introduce a preferred time slicing of de Sitter space.

$$U(t, \boldsymbol{x}) \equiv t + \pi(t, \boldsymbol{x}) \ . \tag{1.10}$$

The Goldstone mode π parameterizes *adiabatic fluctuations* of the fields ψ_m, i.e. perturbations corresponding to a common, local shift in time of the homogeneous fields,

$$\delta\psi_m(t, \boldsymbol{x}) \equiv \psi_m\big(t + \pi(t, \boldsymbol{x})\big) - \psi_m(t) \ . \tag{1.11}$$

The Einstein equations couple the Goldstone boson π to metric fluctuations $\delta g_{\mu\nu}$. A convenient gauge for describing these fluctuations is the *spatially flat gauge*, where the spatial part of the metric is unperturbed,

$$g_{ij} = a^2(t)\, \delta_{ij} \ . \tag{1.12}$$

The remaining metric fluctuations δg_{00} and δg_{0i} are related to π by the Einstein constraint equations. The dynamics of the coupled Goldstone-metric system can therefore be described by π alone.

A second description of the same physics is sometimes convenient, especially in the cosmological context. First, we note that, for purely adiabatic fluctuations, we can perform a time reparameterization that removes all matter fluctuations, $\delta\psi_m \mapsto 0$. This takes us to *comoving gauge*, where the field π has been "eaten" by the metric $g_{\mu\nu}$. The spatial part of the metric can now be written as

$$g_{ij} = a^2(t)\, e^{2\mathcal{R}(t,\boldsymbol{x})}\, \delta_{ij} \ , \tag{1.13}$$

where $\mathcal{R}$ is called the *comoving curvature perturbation*. The other components of the metric are related to $\mathcal{R}$ by the Einstein constraint equations (see Appendix C). The relationship between π (in spatially flat gauge) and $\mathcal{R}$ (in comoving gauge) is

$$\mathcal{R} = -H\pi + \cdots \ , \tag{1.14}$$

where the ellipsis denotes terms that are higher order in π. This links the comoving curvature perturbation $\mathcal{R}$ with the Goldstone boson π of spontaneous symmetry breaking during inflation [55, 56].

The Goldstone mode π exists in every model of inflation. In single-field inflation, π is the unique fluctuation mode [51], while in multi-field inflation, additional light fields can contribute to $\mathcal{R}$: see Appendix B. As we will see in Chapter 4, string theory strongly motivates considering scenarios in which multiple fields are light during inflation. However, from a purely bottom-up perspective, extra light fields during inflation are not required by present observations, and in this section we will focus on the minimal case of a single light field.

One can learn a great deal about the CMB perturbations by studying the Goldstone boson fluctuations alone. The physics of the Goldstone boson is described by the low-energy effective action for π, which can be obtained by writing down the most general Lorentz-invariant action for the field $U \equiv t + \pi$:

$$S = \int \mathrm{d}^4x \sqrt{-g}\, \mathcal{L}[U, (\partial_\mu U)^2, \Box U, \cdots] \,. \tag{1.15}$$

The action (1.15) is manifestly invariant under spatial diffeomorphisms, but because π transforms nonlinearly under time translations, one says that time translation symmetry is nonlinearly realized in (1.15). Expanding (1.15) in powers of π and derivatives gives the effective action for the Goldstone mode. We derive the Goldstone action in detail in Appendix B, via an alternative geometric approach [50, 51], and present only the main results here. At quadratic order in π, and to leading order in derivatives, one finds (cf. Eq. (B.77))

$$S_\pi^{(2)} = \int \mathrm{d}^4x \, \sqrt{-g}\, \frac{M_{\rm pl}^2|\dot{H}|}{c_s^2} \left[\dot{\pi}^2 - \frac{c_s^2}{a^2}(\partial_i\pi)^2 + 3\varepsilon H^2\pi^2\right] \,, \tag{1.16}$$

where $(\partial_i\pi)^2 \equiv \delta^{ij}\partial_i\pi\partial_j\pi$. Since Lorentz symmetry is broken by the time-dependence of the background, we have the possibility of a nontrivial *speed of sound* c_s; standard slow-roll inflation (see Section 2.2.1) is recovered for $c_s = 1$. The field π has a small mass term, which arises from the mixing between π and the metric fluctuations. Using (1.14), we can write (1.16) in terms of the curvature perturbation $\mathcal{R}$,

$$S_{\mathcal{R}}^{(2)} = \frac{1}{2}\int \mathrm{d}^4x \; a^3\, y^2(t) \left[\dot{\mathcal{R}}^2 - \frac{c_s^2}{a^2}(\partial_i\mathcal{R})^2\right] \,, \tag{1.17}$$

where

$$y^2 \equiv 2M_{\rm pl}^2\frac{\varepsilon}{c_s^2} \,. \tag{1.18}$$

The field $\mathcal{R}$ is therefore massless, implying – as we shall see – that it is conserved on superhorizon scales [55].

For simplicity, we will assume that ε and c_s are nearly constant, so that the overall normalization of the action can be absorbed into the definition of a new, canonically normalized, field

$$v \equiv y\,\mathcal{R} = \int \mathrm{d}^3k \, \left[v_k(t)\, a_{\boldsymbol{k}}\, e^{i\boldsymbol{k}\cdot\boldsymbol{x}} + c.c.\right] \,. \tag{1.19}$$

We have written v in terms of time-independent stochastic parameters $a_{\boldsymbol{k}}$ and time-dependent mode functions $v_k(t)$. The mode functions satisfy the *Mukhanov–Sasaki equation*,

$$\ddot{v}_k + 3H\,\dot{v}_k + \frac{c_s^2k^2}{a^2}v_k = 0 \,. \tag{1.20}$$

This is the equation of a simple harmonic oscillator with a friction term provided by the expanding background. The oscillation frequency depends on the physical momentum and is therefore time dependent:

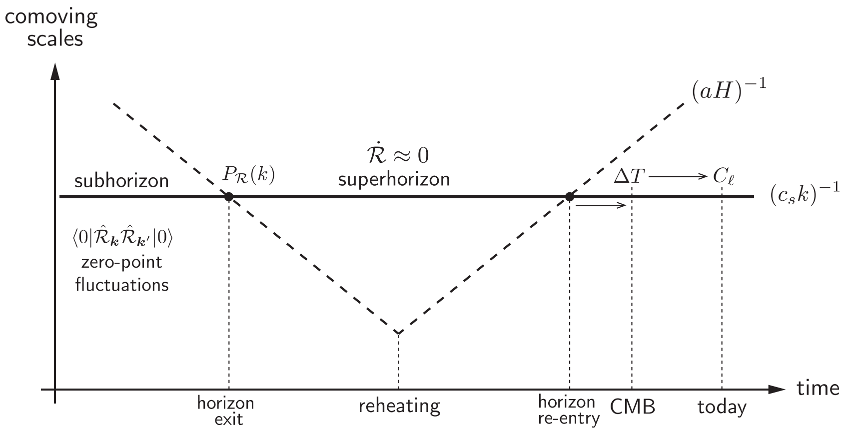

Fig. 1.4 The evolution of curvature perturbations during and after inflation: the comoving horizon $(aH)^{-1}$ shrinks during inflation and grows in the subsequent FRW evolution. This implies that comoving scales $(c_s k)^{-1}$ exit the horizon at early times and re-enter the horizon at late times. In physical coordinates, the Hubble radius H^{-1} is constant and the physical wavelength grows exponentially, $\lambda \propto a(t) \propto e^{Ht}$. For adiabatic fluctuations, the curvature perturbations $\mathcal{R}$ do not evolve outside of the horizon, so the power spectrum $P_{\mathcal{R}}(k)$ at horizon exit during inflation can be related directly to CMB observables at late times.

$$\omega_k(t) \equiv \frac{c_s k}{a(t)} . \tag{1.21}$$

At early times (small a), $\omega_k \gg H$ for all modes of interest. In this limit, the friction is irrelevant and the modes oscillate. However, the frequency of any given mode drops exponentially during inflation. At late times (large a), the dynamics is dominated by friction and the mode has a constant amplitude. We say that the mode "freezes" at *horizon crossing*, i.e. when $\omega_k(t_\star) = H$ or $c_s k = aH(t_\star)$. It is these constant superhorizon fluctuations that eventually become the density fluctuations that we observe in the CMB and in LSS (see Fig. 1.4).[1]

1.2.2 *Vacuum fluctuations*

The initial conditions for v (or $\mathcal{R}$) are computed by treating it as a quantum field in a classical inflationary background spacetime. This calculation has become textbook material [57, 58] and can also be found in many reviews (e.g. [27, 59]). We present the details in Appendix C. Here, we will restrict ourselves to a simplified, but intuitive, computation [60].

[1] Recall that we are assuming adiabatic initial conditions. The presence of entropy perturbations, as in multi-field models, can complicate the relation between the curvature perturbations at horizon crossing and the late-time observables – see Appendix C.

The Fourier modes of the classical field v are promoted to quantum operators

$$\hat{v}_{\boldsymbol{k}} = v_k(t)\hat{a}_{\boldsymbol{k}} + h.c. \tag{1.22}$$

At sufficiently early times, all modes of cosmological interest were deep inside the Hubble radius. In this limit, each mode behaves as an ordinary harmonic oscillator. The operators $\hat{a}_{\boldsymbol{k}}$ play the role of the annihilation operators of the quantum oscillators. The vacuum state is defined by $\hat{a}_{\boldsymbol{k}}|0\rangle = 0$. The oscillation amplitude will experience the same zero-point fluctuations as an oscillator in flat space, $\langle 0|\hat{v}_{\boldsymbol{k}}\hat{v}_{\boldsymbol{k}'}|0\rangle = (2\pi)^3|v_k|^2\delta(\boldsymbol{k}+\boldsymbol{k}')$, where

$$|v_k|^2 = \frac{1}{a^3}\frac{1}{2\omega_k} \ . \tag{1.23}$$

The factor of a^{-3} arises from the physical volume element in the Lagrangian (1.17) – note that the Fourier mode v_k was defined using the comoving coordinates rather than the physical coordinates. The second factor, $1/(2\omega_k)$, is the standard result for the variance of the amplitude of a harmonic oscillator in its ground state. (In inflation, this state is the *Bunch–Davies vacuum*.) As long as the physical wavelength of the mode is smaller than the Hubble radius, the ground state will evolve adiabatically. Equation (1.23) then continues to hold, and the precise time at which we define the initial condition is not important. Once a given mode gets stretched outside the Hubble radius, the adiabatic approximation breaks down and the fluctuation amplitude freezes at

$$|v_k|^2 = \frac{1}{2}\frac{1}{a_\star^3}\frac{1}{c_s k/a_\star} \ , \tag{1.24}$$

where $a_\star$ is the value of the scale factor at horizon crossing,

$$\frac{c_s k}{a_\star} = H \ . \tag{1.25}$$

Combining (1.25) and (1.24), we get

$$|v_k|^2 = \frac{1}{2}\frac{H^2}{(c_s k)^3} \ , \tag{1.26}$$

where from now on it is understood implicitly that the right-hand side is evaluated at horizon crossing.

1.2.3 Curvature perturbations

Using (1.19), we obtain the *power spectrum* of primordial curvature perturbations,

$$P_{\mathcal{R}}(k) \equiv |\mathcal{R}_k|^2 = \frac{1}{4}\frac{H^4}{M_{\rm pl}^2|\dot{H}|c_s}\frac{1}{k^3} \ . \tag{1.27}$$

The variance in real space is $\langle \mathcal{R}^2 \rangle = \int \mathrm{d}\ln k \; \Delta^2_{\mathcal{R}}(k)$, where we have defined the dimensionless power spectrum

$$\Delta^2_{\mathcal{R}}(k) \equiv \frac{k^3}{2\pi^2} P_{\mathcal{R}}(k) = \frac{1}{8\pi^2} \frac{H^4}{M_{\rm pl}^2 |\dot{H}| c_s} \; . \tag{1.28}$$

Since the right-hand side is supposed to be evaluated at horizon crossing, $c_s k = aH$, any time dependence of H and c_s translates into a scale dependence of the power spectrum. Scale-invariant fluctuations correspond to $\Delta^2_{\mathcal{R}}(k) = const.$, and deviations from scale invariance are quantified by the *spectral tilt*

$$n_s - 1 \equiv \frac{d \ln \Delta^2_{\mathcal{R}}}{d \ln k} = -2\varepsilon - \tilde{\eta} - \kappa \; , \tag{1.29}$$

where we have defined two additional expansion parameters,

$$\tilde{\eta} \equiv \frac{\dot{\varepsilon}}{H\varepsilon} \qquad \text{and} \qquad \kappa \equiv \frac{\dot{c}_s}{H c_s} \; . \tag{1.30}$$

Inflationary backgrounds typically satisfy $\{\varepsilon, |\tilde{\eta}|, |\kappa|\} \ll 1$ and hence predict $n_s \approx 1$. Inflation would not end if the slow-roll parameters vanished, so importantly, we also expect a finite deviation from perfect scale invariance, $n_s \neq 1$.

1.2.4 Gravitational waves

Arguably the cleanest prediction of inflation is a spectrum of primordial gravitational waves. These are tensor perturbations to the spatial metric,

$$g_{ij} = a^2(t)(\delta_{ij} + h_{ij}) \; , \tag{1.31}$$

where h_{ij} is transverse and traceless. Expanding the Einstein–Hilbert action leads to the quadratic action for the tensor fluctuations:

$$S_h^{(2)} = \frac{1}{2} \int \mathrm{d}^4 x \; a^3 \, y^2 \left[(\dot{h}_{ij})^2 - \frac{1}{a^2} (\partial_k h_{ij})^2 \right] , \tag{1.32}$$

where

$$y^2 \equiv \frac{1}{4} M_{\rm pl}^2 \; . \tag{1.33}$$

The structure of the action is identical to that of the scalar fluctuations, Eq. (1.17), except that tensors do not have a nontrivial sound speed and the relation to the canonically normalized field does not include ε, because at linear order tensors do not feel the symmetry breaking due to the background evolution. The quantization of tensor fluctuations is therefore the same as for the scalar fluctuations. In particular, Eq. (1.26) applies to each polarization mode of