Inflation Physics from the Cosmic Microwave Background and Large Scale Structure

Topical Conveners: J.E. Carlstrom, A.T. Lee


Executive Summary

Precision cosmological measurements push the boundaries of our understanding of the fundamental physics that governs our universe. In the coming years, cosmologists will be in a position to make major breakthroughs in our understanding of the physics of the very early universe and be able to probe particle physics and gravity at the highest energy scales yet accessed. A major leap forward in the sensitivity of cosmological experiments is within our technological reach, leveraging past and current experience to tackle some of the most interesting fundamental physics questions.

Cosmic inflation, the theory that the universe underwent a violent, exponential expansion during the first moments of time, is the leading theoretical paradigm for the earliest history of the universe and for the origin of the structure in the universe. Current measurements of the cosmic microwave background (CMB) and observations of the large scale distributions of dark matter and galaxies in the universe are in stunning agreement with the concept of inflation. The next generations of experiments in observational cosmology are poised to decide central questions about the mechanism behind inflation. In this short document, we highlight the importance of experimentally determining the nature of inflation in the early universe and the unique opportunity these experiments provide to explore the physics of space, time, and matter at the highest energies possible: those found at the birth of the universe.
Although the landscape of possible models for inflation is potentially large—and sensitive to quantum gravity corrections to the low energy quantum field theory—the phenomenology is sufficiently well understood to make concrete distinctions between fundamentally different classes of models that we can test observationally.

One key and generic prediction is the existence of a background of gravitational waves from inflation that produces a distinct signature in the polarization of the CMB, referred to as “B-mode” polarization. The amplitude of primordial gravitational waves, or tensor modes, which can be detected or constrained by observations of the $B$-mode polarization in the CMB, is fundamentally interesting for several basic reasons. It is proportional to the energy scale of inflation and tied to the range of the inflaton field. In particular, observations promise to reach the level of sensitivity that will enable them to determine whether the field range is larger than the Planck scale in the simplest versions of inflation [1]. This provides a striking ultraviolet-sensitive probe of quantum field theory and quantum gravity, and an observational test of string theoretic large-field inflation. Additionally, in one theoretically developed (though currently speculative) alternative to inflation, the ekpyrotic scenario, the authors of [2, 3] find no mechanism for generation of the tensor perturbations; hence, if these calculations are correct, detection of $B$-modes would present a convincing refutation of that model. Last but not least, a detection of tensor modes would constitute a measurement—for the first time—of the quantum mechanical fluctuations of the gravitational field.

This motivates a next-generation CMB experiment with the sensitivity and systematics control to detect such a polarized signal at $\geq 5\sigma$ significance, thus ensuring either a detection of inflationary gravitational waves or the ability to rule out large classes of inflationary models. A program to meet these goals by developing a Stage IV CMB experiment, CMB-S4, with $O(500,000)$ detectors by 2020 is described in the companion cosmic frontier planning document *CF5 Neutrino Physics from Cosmic Microwave Background and Large Scale Structure*. Such an experiment would also contribute to inflationary science by strongly constraining the spectrum of primordial density fluctuations, allowing us to distinguish different families of inflationary models.

Possibilities for self-interactions of the inflaton and for additional fields are tested by different limits of the correlation functions of the perturbations. Despite important recent progress, we require substantial improvements before observational constraints on these quantities limit the interactions to be small corrections to slow-roll, or to detect non-Gaussianity if it is present. A concerted theoretical effort combined with observations of large scale structure promises to fill this gap. A detection of primordial non-Gaussianity of the so-called local shape would effectively rule out all models of inflation that involve a single scalar field [4, 5, 6]. The CMB bound on local-model non-Gaussianity is now limited by having only one sky to observe; further improvements will come from measurements of the large scale structure of the universe. The next generation of large scale structure measurements will produce non-Gaussianity constraints that are an important cross-check of the CMB bound and will pave the way for more stringent bounds from future large scale structure measurements.

1 Inflation Science: theoretical motivations

Cosmic inflation, the idea that the universe underwent a period of exponential expansion in the first $10^{-34}$ seconds of its existence, was proposed in the early 1980s to explain the apparent smoothness and flatness of the universe and the absence of relics such as magnetic monopoles [7]. It was soon realized that quantum fluctuations generated during inflation would eventually evolve into the distributions of dark matter and galaxies we observe today. Inflation drives the spatial curvature to nearly zero, and introduces density perturbations that are adiabatic with a nearly scale invariant spectrum that depends on the details of the inflationary potential.
The frontier of inflation research currently lies in measurement of the polarization of the CMB and in searching for non-Gaussianity in the distribution of dark matter and galaxies in the late universe. The CMB offers a unique window between the late-time universe dominated by dark matter and dark energy, and the early universe when the energy density was dominated by the potential that drove cosmic inflation. The amplitude of tensor $B$-mode polarization in the CMB is proportional to the energy of inflation and tied to the range of the inflaton field. The rich phenomenology of non-Gaussianity in the distribution of dark matter and galaxies in the late universe offers opportunities to directly study the dynamics of inflation.

In the context of inflationary paradigm, we can be precise about the significance and interpretation of these measurements. Here we briefly summarize some highlights.

The predictions of most inflationary models can be characterized in terms of the statistical properties of perturbations to the metric away from a homogeneous background solution $a(t) \approx e^{Ht}$:

$$ds^2 = -dt^2 + h_{ij}dx^idx^j, \quad h_{ij} = a(t)^2 [e^{2\zeta} \delta_{ij} + \gamma_{ij}].$$

(1)

Here $\zeta$ contains the scalar perturbation (in a gauge where the inflaton perturbation has been gauged away via time reparameterization), and $\gamma$ the tensor perturbation. The CMB and LSS are sufficiently linear in the regimes of interest that these primordial metric perturbations can be inferred directly from observations. The greater challenge is to make inferences about the physics of inflation from knowledge of $\zeta$ and $\gamma_{ij}$.

1.1 Tensor Modes

Determining the tensor to scalar ratio

$$r = \frac{\langle \gamma \gamma \rangle}{\langle \zeta \zeta \rangle}$$

via a measurement of the primordial $B$-mode polarization [8, 9, 10, 11] is important for three simple reasons.\(^1\)

(1) A detection would constitute a measurement—for the first time—of the quantum mechanical fluctuations of the metric: in the absence of classical inhomogeneities ($\langle \gamma_{ij} \rangle = 0$) inflation generates a nonzero variance

$$\langle \gamma_s \gamma_{s'} \rangle' = 2 \frac{H^2}{M_P^2} \delta_{ss'},$$

(3)

where $s, s'$ label graviton polarizations and $\langle \ldots \rangle'$ denotes dropping the momentum conserving delta function.

(2) The formula (3) also exhibits a direct connection between the tensor signal and the scale of inflation ($H$ or equivalently $V \sim 3M_P^2 H^2$), with the observable level fortuitously corresponding to GUT scale inflation. However, since $V^{1/4} \sim (r/0.01)^{1/4} \times 10^{16}$ GeV, improvements in $r$ of the order we consider in this document do not translate into a large improvement in the constraint on the scale $V^{1/4}$. However, the improvements do correspond to reaching a very significant threshold in the field range of the inflaton as described next.

(3) Either a detection or a constraint at the level of $r \sim 0.01 - 0.001$ would determine whether the inflaton $\phi$ rolled a super-Planckian or sub-Planckian distance in field space. In single field slow-roll inflation, we have

\(^1\)Exceptions to these going beyond single field slow-roll inflation, for example postulating rapid variation of the slow-roll parameter $H/H^2$ [12, 13], or including gravitational waves from sources produced during inflation [14], bring in their own motivations on par with that derived from the relation (4) in the simplest cases. Although these more exotic possibilities are interesting, with space constraints we will not lay out the various caveats.
the following simple relation [1] relating the number of e-foldings $N_e$ to the tensor to scalar ratio $r$

$$N_e = \int \frac{da}{a} = \int H dt = \int \frac{H M_p}{\dot{\phi}} \frac{d\phi}{M_p} = \sqrt{8} r^{-1/2} \Delta \phi / M_p \tag{4}$$

where we used the slow-roll result $\langle \zeta \zeta \rangle' = \frac{H^2}{2 \rho}$, (3), and (2). A detection of $r$ corresponding to $\Delta \phi \geq M_P$ would imply sensitivity to an infinite sequence of Planck-suppressed operators in the effective action for the inflaton. At a field range of $M_P$, a minimal reheating temperature roughly corresponds to $N_e \sim 30$ and $r \sim 0.01$, and a maximal reheating temperature corresponds to $N_e \sim 60$ and $r \sim 0.002$. The former is the reach of near-term experiments, and the latter the goal of the Stage IV experiment, CMB-S4, as discussed in the experimental sections of this report.

An approximate shift symmetry—if present in the ultraviolet completion of gravity—may protect against such terms. Such a symmetry over a large field range yields radiative stability from the low energy quantum field theory point of view. Two now-classic examples are $m^2 \phi^2 \text{ inflation}$ [15] and $\text{Natural Inflation}$ [16] (traditional quantum field theory axions, requiring a super-Planckian period), as well as a generalization known as $\text{N-flation}$ [17] which employs multiple fields each with a small range. A recent theoretical development is the recognition of a very broad class of different models, including the Starobinsky model [18], $\text{Higgs inflation}$ [19, 20], and a broad class of models based on spontaneously broken conformal and superconformal symmetry [21], which give identical, model-independent predictions of $n_s$ and $r$, with the caveat that these models still await ultraviolet completion. In particular, these models predict tensor perturbations with $r \sim 0.004$, within reach of the experimental program described in this report.

It is important to understand if a symmetry over a super-Planckian range operates in a consistent ultraviolet completion of gravity. In string theory, a leading candidate for quantum gravity, a way in which large-field inflation protected by an underlying shift symmetry arises naturally is via monodromy [22, 23, 24] for scalar fields such as axions: in the presence of generic fluxes and branes, potential energy builds up along axion directions. This yields potentials which flatten out relative to $m^2 \phi^2$ at large field range, a simple result of massive degrees of freedom adjusting in an energetically favorable way [25]. This mechanism predicts observable $B$ modes and a tilt distinct from $m^2 \phi^2$ inflation and $\text{Natural Inflation}$. Other string theoretic mechanisms can produce small field inflation, a perfectly viable possibility as well.

Another interesting direction is to use tensor perturbations to constrain additional sectors of light fields. For various scenarios involving additional axions, this has been extensively explored in the literature; see for example [26][27] and references therein.

To summarize, the CMB $B$-mode measurements will, according to present forecasts, reach an important threshold, the Planck range in field. A null result on $r$ would be qualitatively important, showing that small field inflation is the operative case. A detection would provide an important lever to quantum gravitational physics.

On the experimental side, the Stage-IV CMB experiment CMB-S4 described below will significantly improve the sensitivity over current efforts, and provide a conclusive experimental discrimination of large-field and small-field inflation. The proposed goal for CMB-S4 on the precision of the tensor to scalar ratio $r$ is $\sigma(r) = 0.001$ (statistical), with a similar level of systematic uncertainty budgeted. This goal is found to be achievable with a wide range of design parameters for CMB-S4 using only the degree-scale feature ($\ell \sim 100$) known to be accessible from the ground. With this level of uncertainty, CMB-S4 will unambiguously detect ($> 5 \sigma$, as universally accepted in the HEP community) tensor modes from any large-field inflation model with $r \gtrsim 0.01$. If $r$ is near the $2\sigma$ limit set by data from the $\text{Planck}$ satellite, as in the case of $m^2 \phi^2$ inflation, current funded experiments will be able to detect $B$-mode polarization with high significance. In that scenario, the CMB-S4 surveys can be reconfigured to measure tensor amplitude with high precision. Further
characterization of tensor properties, such as scale-invariance and Gaussianity, would also be possible with CMB-S4’s superb sensitivity. These tests, even with relatively modest precision, would constitute striking qualitative verifications of inflation.

1.2 Scalar Tilt, Running, and Beyond

During inflation $|\dot{H}|, |\dddot{H}| > 0$ and therefore the background evolves with time. Perturbations produced at different times (and hence with different $k$) see different Hubble scales and will freeze with slightly different amplitudes. As a result, most inflation models predict a power law dependence of the scalar power spectrum

$$k^3 \langle \zeta_k \zeta_{-k} \rangle \propto k^{n_s - 1}$$

with a value of $n_s$ slightly different from the value $n_s = 1$ in the scale-invariant case. Current measurements from the Planck satellite give [28]

$$n_s = 0.960 \pm 0.0073$$

excluding $n_s = 1$ by more than $5\sigma$. This is concrete evidence that cosmic structure was generated by a physical quantum field theory rather than by any putative mechanism imposing precise scale-invariance.

In the slow-roll approximation, the exponent $n_s$ is given by

$$n_s - 1 = -6\epsilon + 2\eta$$

where

$$\epsilon = \frac{M_p^2}{2} \left( \frac{V'}{V} \right)^2, \quad \eta = M_p^2 V'' / V,$$

with $' \equiv \frac{\partial}{\partial \phi}$. Thus, the precision measurement of $n_s$ gives specific information that can be used as a constraint on a model of the inflaton potential. This is especially relevant when combined with Planck measurements of $r$. From figure 1, we can see that various classes of large-field models make distinct predictions within the $n_s$-$r$ plane. Of particular interest is the fate of the minimal $m^2 \phi^2$ model, which is on the edge of the

![Figure 1. Current CMB constraints on the combined $n_s$-$r$ parameter space [28]. Constraints from WMAP satellite are essentially the same, since this large scale temperature measurements are cosmic variance limited.](image-url)
currently allowed region. Therefore, we gain critical information through \( n_s \) and its running, \( \frac{\partial n_s}{\partial \log k} \), that is complimentary to the effort to measure \( r \).

Finally, inflation predicts small, but non-zero curvature. In a general inflationary scenario, one expects spatial curvature on the horizon patches to receive contributions from large scale modes and hence the curvature parameter \( \Omega_k \) is naturally of the order \( 10^{-4} \). Measurements of the value of \( \Omega_k \) that would differ from naive inflationary expectation would be very informative about the process responsible for inflation.

In particular, if \( |\Omega_k| \) is found to be considerably larger than this value, it will tell us that inflation was not proceeding in a slow-roll when scales just larger than our observable horizon exited their inflationary horizons. More concretely, observations of negative and large \( \Omega_k \) (positive curvature) would falsify eternal inflation, while observation of positive and large \( \Omega_k \) (negative curvature) would be consistent with false vacuum eternal inflation [29, 30].

Our current constraints on this parameter from the CMB alone are \( \Omega_k = 0.042^{+0.027}_{-0.018} \) and improve significantly to \( \Omega_k = (0.5 \pm 3.3) \times 10^{-3} \) upon the addition of baryon acoustic oscillation (BAO) data [31]. While these numbers are already interesting, future constraints approach an even more interesting regime. For example, an optimistic forecast for the error on \( \Omega_k \) stemming from combination of current Planck priors, Euclid and Dark Energy Spectroscopic Instrument (DESI) data (assuming \( \Lambda \)CDM, but varying neutrino masses) is \( 5 \times 10^{-4} \), improving current constraints by an order of magnitude. If curvature were to be detected at this level, it would have profound implications for the inflationary paradigm.

Further potential observables include features in the potential or oscillations in the power spectrum (see e.g., [32, 33, 34, 35, 36]) which may be correlated with measurable \( B \) modes [34]. The latter analysis is subtle because of theoretical backreaction effects, offering an interesting challenge at the interface of theory and data analysis.

Finally, anomalies reported by the Planck collaboration, which would also require additional parameters beyond \( \Lambda \)CDM, deserve further work. These may indicate additional structure in the power spectrum and non-Gaussianity, perhaps related to pre-inflationary relics or reheating dynamics in the presence of additional fields. However, their statistical significance is currently low. The wavenumber of the modes which contribute the most to the multipoles of the anomalies is not particularly small. Next generation LSS surveys, such as the Large Synoptic Survey Telescope (LSST) and DESI, will be able to observe some of these modes, although it is still unclear if they will be able to control systematics effects at such large distances. Since for a three-dimensional survey error bars from cosmic variance are inversely proportional to the wavenumber of the mode to the 3/2 power, there is, in principle, enough information to make the Planck anomalies significant by many standard deviations. Clearly, this would be a very important discovery that requires further study.

1.3 Non-Gaussianity

More direct evidence of the nature of the inflaton as a physical interacting quantum field would come from the observation of nonlinear inflaton interactions. These would show up as correlations in the CMB fluctuations beyond the Gaussian approximation (see [37] for a recent review). The presence of non-Gaussian correlations in the initial condition for structure formation should also be visible in direct observations of cosmic structure, measured by large scale surveys such as LSST and DESI. They would appear as non-Gaussian correlations of galaxies (or dark matter) on very large scales.
Any observation of a non-zero correlation function beyond the power spectra (i.e., non-Gaussianity) would include a wealth of new information about the primordial universe. Many studies have focused on a non-zero three point function of primordial perturbations, which can be expressed in terms of the bispectrum $B(k_1, k_2, k_3)$,

$$
\langle \zeta_{k_1} \zeta_{k_2} \zeta_{k_3} \rangle = B(k_1, k_2, k_3) (2\pi)^3 \delta^{(3)}(k_1 + k_2 + k_3) .
$$

(9)

The four-point function has been explored as well, though to a lesser extent. After considering translation and rotation invariance (as assumed in (9)), the bispectrum is a function of three variables. Many models produce scale invariant correlation functions, which further reduces it to a function of two variables that can be described in terms of the shape of the triangle in momentum space formed from $k_1, k_2, k_3$ \cite{38}. Large violations of scale invariance can arise in the bispectrum (e.g., \cite{39, 40}) and may \cite{41} (or may not \cite{42}) have correlated signatures in the power spectrum.

These higher correlation functions contain full functions worth of information about the inflationary epoch. This information would be invaluable to our understanding of the mechanism of inflation and the origin of the primordial fluctuations. Some important implications include

(1) **Probing interactions:** Inflation requires a slowly varying Hubble expansion, for example a slow variation in the potential energy of a field. This can arise roughly speaking in two ways—either via a flat potential, or on a steep potential with sufficient interactions or dissipation to slow the field evolution. The latter class of mechanisms tends to produce significant non-Gaussianity, peaked on equilateral triangles $k_1 \sim k_2 \sim k_3$ (with other shapes possible). UV complete mechanisms exhibiting all these possibilities have been studied \cite{43, 44}, helping stimulate the much more systematic understanding of the observables \cite{45} that have been classified using the effective field theory (EFT) of the inflationary perturbations \cite{46, 47}.

In the EFT, the inflationary perturbations can be realized as Goldstone modes ($\pi$) of time translation symmetry, whose dynamics is described by a universal Lagrangian, schematically

$$
S_\pi = \int d^4 x \sqrt{-g} \left[ M^2_\pi H \left( \pi^2 - \frac{(\partial \pi)^2}{a^2} \right) + M^4_\pi \left( \frac{\pi^2 + \pi^3 - \pi \left( \frac{\partial \pi}{a^2} \right)^2}{a^2} \right) + M^4_\pi \pi^3 + \ldots \right],
$$

(10)

where $\pi$ is the Goldstone boson associated to the non-linearly realized time-translations. As typical of Goldstone bosons, this Lagrangian is characterized by higher-dimension interactions that violate perturbative unitarity at a high energy scale $\Lambda_U$ \cite{47, 48}. Similarly to what happens in particle physics with the precision electroweak tests, limits on non-Gaussianities can be mapped, in a model-independent way, into limits of the parameters of this Lagrangian, as recently done by the WMAP and *Planck* collaborations. The amplitude on non-Gaussianities take the form $(H/\Lambda_U)^\alpha$ where $\alpha \sim \mathcal{O}(1)$. In the case of single field inflation, existing constraints from *Planck* on the equilateral and the “orthogonal” shapes require $\Lambda_U \gtrsim \mathcal{O}(10)H$. It would be extremely interesting to raise the bounds on $\Lambda_U$ in order to constrain it to be larger than the scale associated to the speed of inflaton in standard slow-roll models, given by $\dot{\phi} \sim \mathcal{O}(100)H$ (or equivalently $f^{\text{equilateral}}_{NL} \sim 1$). In this case, we would have shown that the EFT can be extended to such a high energy to be best described by standard slow-roll inflation.

(2) **“Discovering” extra fields:** It can be shown that single-field inflation produces no signal in the “squeezed limit” of the bispectrum where one momentum is much smaller than the other two, $k_1 \ll k_2 \sim k_3$ \cite{4, 5}. The constraints on this limit are often characterized in terms of the local shape, $f^{\text{local}}_{NL}$. The detection of a non-zero $f^{\text{local}}_{NL}$, or any non-trivial squeezed shape, would require additional fields, beyond the inflaton, and thus would rule out single-field inflation. Furthermore, precise measurements in the limit can uncover non-trivial information regarding these extra fields. For example, additional sectors containing weakly interacting massive fields \cite{49} (as in supersymmetric theories \cite{50}), as well as additional sectors of strongly coupled CFTs \cite{51}, both produce bispectra of the form $B_{\text{squeezed}}(k_1, k_2, k_3)(k_1/k_2)^\Delta$ (where the constant $\Delta$
is determined by the mass, via $\Delta = \frac{3}{2} - \sqrt{\frac{9}{4} - \frac{m^2}{H^2}}$, or is the dimension of a conformal primary operator). The measurement of $\Delta$ would thus tell us about the spectrum of masses / operators in such a sector.

Given the Planck constraint on $f_{NL}$, it is possible to turn things around and derive precision constraints on hidden sectors coupled to the inflaton perturbations via higher dimension operators suppressed by a mass scale $M_*$; this scale must be several orders of magnitude larger than the Hubble scale during inflation [51, 52]. For large-field inflation that provides a lever to the Planck scale.

Finally, there are several additional forms of non-Gaussianities that are not characterized by a peculiar squeezed limit, that are largely unconstrained, and that are possible only in multifield models. They have emerged in the context of the effective field theory approach to multi field inflation [53], and they can be mapped into signatures of the symmetries that explain the lightness of additional fields during inflation.

(3) “Discovering” multiple sources: Relations between the bispectrum and the trispectrum (4-point function) can further indicate that multiple fields produce the observed curvature perturbations (i.e., there are multiple sources of curvature). Specifically, if the amplitude of the four point function, $\tau_{NL}$, is found to satisfy $\tau_{NL} \gg f_{NL}^2$, then one must have multiple sources of the curvature perturbation [54, 55, 56, 57]. Furthermore, the consistency of many models requires that the precise shape of the bispectrum and trispectrum are not independent [47]. Precise measurements of both would then strongly constrain the origin of the curvature perturbations.

From all this it is clear that an improved determination of the constraints on non-Gaussianity, or a detection, would help distinguish very different mechanisms for inflation and possibilities for additional sectors in the early universe. Moreover, it is worth stressing that other forms of non-Gaussianity, not yet searched for, may lie in the data.

To date, the best constraints on non-Gaussianity have come from observations of the CMB. However, current and planned large scale structure surveys like the Baryon Oscillation Spectroscopic Survey (BOSS) [58], its successor eBOSS, DESI, Dark Energy Survey (DES) and LSST have the potential to dramatically improve our sensitivity to non-Gaussian effects beyond the limits set by Planck. Realizing the promise of these surveys for inflation requires a community effort dedicated to both data analysis for non-Gaussianity and improving our theoretical understanding of the mildly non-linear regime of large scale structure (including dark-matter clustering, tracer biasing and redshift-space distortions). These efforts will be relatively inexpensive, and their value could be enormous. Since the amount of information grows like the cube of the smallest scale measured, a factor of two improvement in our understanding of the mildly non-linear regime is equivalent to improving almost tenfold the size of a survey. A recently developed approach, based on applying effective field theory techniques to LSS [59], seem to show that this is indeed possible. Such efforts would pay similar dividends to the study of dark energy and are therefore complimentary to the existing effort.

2 Constraining Inflation Physics With Cosmological Probes

As detailed in Section 1, the theory of cosmic inflation is the most promising model for the dynamics of the universe at very early times and high energies. Inflation supplies a natural explanation for the smoothness and geometrical flatness of our observable universe, and the predicted consequences of an inflationary beginning to our universe, including a nearly Gaussian distribution of primarily adiabatic density fluctuations with a nearly scale-invariant spectrum, have been spectacularly confirmed in recent years. The strong acoustic features in the power spectrum of fluctuations in the CMB temperature and polarization fields and the correlations between the temperature and polarization fluctuations demonstrate that the primordial density
fluctuations are almost entirely adiabatic [60]. The power-law slope of the primordial fluctuations has recently been shown to be very close to the scale-invariant value but with a significant (5σ) deviation in the direction predicted by inflation [31]. The measured level of non-Gaussianity in the CMB has remained consistent with zero, even as experimental sensitivity has dramatically increased [61, 62].

Now that the general picture of an early inflationary epoch is well-established, we can begin to ask questions about the specifics of inflation physics, such as: What was the energy scale of inflation? Was inflation caused by a single scalar field, or were there multiple fields involved? What was the shape of the inflationary potential? The next generation of cosmological probes, particularly measurements of the polarization of the CMB and large-area galaxy surveys, are poised to deliver the first answers to these fundamental questions.

2.1 Probing High Energy Scale Physics with the Polarized Cosmic Microwave Background

The study of the CMB is a unique and powerful tool for learning how our universe works at the most fundamental level. Ever since the epoch of recombination, when the CMB was released, most of the photons in the universe have been freely streaming. We can therefore use the CMB to directly test precise predictions of cosmological models and directly probe the basic physics governing our universe up to the epoch of recombination, including models of cosmic inflation in the very early universe.

One of the inescapable predictions of inflation is that the period of violent expansion produced a spectrum of relic gravitational waves (i.e., tensor-mode perturbations). The ratio of the gravitational-wave (tensor) to density-fluctuation (scalar) power in the CMB, \( r \), is related to the energy scale of inflation and the range of the inflaton field as discussed above in the theory section. Even though these tensor modes are produced at all wavelengths, the signal is far too small to be seen directly with any conceivable gravitational-wave interferometer. However, these gravitational waves leave a potentially observable imprint in the CMB at large angular scales. If the energy scale of inflation were high enough, the primordial gravitational-wave background would have been seen by the WMAP or Planck satellites via the extra fluctuations that would have been left behind in the CMB temperature power spectrum. Planck has placed a 95% upper limit on the tensor-to-scalar ratio of \( r < 0.11 \) [31]. Due to cosmic variance (the fact that we only have one universe to observe), we will never be able to improve this limit on \( r \) significantly through CMB temperature measurements alone.

This leaves us with only one avenue to probe the energy scale of inflation: precision measurements of the polarization of the CMB. Primordial gravitational waves generated as a result of inflation would have imprinted a faint but unique signal in the polarization of the CMB, the amplitude of which directly scales with the energy scale of inflation. The polarization of the CMB can be decomposed into a curl-free component and a divergence-free component, called \( E \) modes and \( B \) modes in analogy to electromagnetism. The polarization imprinted on the CMB at the last scattering surface by a scalar-generated quadrupole can only be an \( E \)-mode signal and has been well-measured [63, 64, 65, 66]. In contrast, the only source of \( B \)-mode signal in the primary CMB is gravitational waves from inflation. If inflation happened on the GUT scale (\( \sim 10^{16} \) GeV), the signal will be easily detectable, and with a future generation of CMB experiments we will be able to directly map the quantum fluctuations generated by inflation.

A secondary signal from gravitational lensing of the CMB by structure along the line of sight can produce \( B \) modes by distorting modes that were originally pure \( E \).

This lensing \( B \)-mode signal is predominantly at smaller angular scales than the inflationary signal and will have to be well-measured and cleaned (or “delensed”) from CMB maps in order to continue pushing
Figure 2. Expected signal levels for the $E$-mode (red, solid), inflationary gravitational-wave $B$-mode (blue, solid), and lensing $B$-mode (green, solid) signals. The gravitational-wave $B$-mode signals are shown for tensor-to-scalar ratios of $r = 0.001$ (the Stage-IV goal) and $r = 0.01$ (the boundary between small-field and large-field inflation models). The lensing $B$-mode signal is shown as a band encompassing the predicted signal for values of the sum of neutrino masses $0 \leq \sum m_\nu \leq 0.1$ eV. Defensing by a factor of 4 in amplitude is shown schematically by the green arrow, with the residual signal at $\ell \leq 200$ (where the delensing is critical to the constraint on $r$) shown by the green, long-dashed line. The black, short-dashed line shows the level of current 95% upper limits on $B$ modes from WMAP at the largest scales, the BICEP experiment at degree scales, and the QUIET and QUaD experiments at smaller scales. The brown, long-dashed lines show the expected polarized foreground contamination at 95 GHz for the cleanest 1% and 25% of the sky.

Figure 2 shows the expected signal levels for the scalar $E$-mode, inflationary gravitational-wave $B$-mode, and lensing $B$-mode signals (with and without delensing by a factor of four in amplitude or 16 in power) for tensor-to-scalar ratios of $r = 0.001$ and $r = 0.01$. Also shown are current limits on the $B$-mode amplitude and expected foreground contamination at an observing frequency of 95 GHz for the cleanest 1% and 25% of the sky. As Figure 2 shows, the inflationary $B$-mode signal is accessible with significant improvements in instrument sensitivity and careful control of foreground and lensing contamination.
Sections 2.1.1 through 2.1.3 provide an overview of the current generation of CMB polarization experiments and the inflationary science reach of a potential future experiment. We have adopted a nomenclature to categorize these current and future efforts: we categorize generations, or stages, of experiments roughly by the number of detectors that are on the sky, which is a good proxy for sensitivity. Stage-I experiments have finished observing, and have roughly $\sim 100$ detectors. Stage-II experiments are currently observing with $\sim 1000$ detectors, and Stage-III experiments are currently under development with $\sim 10000$ detectors. A Stage-IV experiment would have yet another order of magnitude more detectors and would harness the resources and experience of the CMB community to produce a cohesive suite of experiments targeting some of the most interesting and fundamental questions in the study of the nature of our universe.

Figure 3 shows the results of forecasts for a few representative configurations of a Stage-IV CMB experiment (for details, see Section 2.1.3). The Stage-IV goal of $\sigma(r) = 10^{-3}$ – which would result in an unambiguous $>5\sigma$ detection for large-field inflation – is achieved or exceeded for a variety of design parameters, fractions of sky covered ($f_{\text{sky}}$), and foreground assumptions, with a broad minimum around $f_{\text{sky}} = 1\%$.

**Figure 3.** Expected statistical uncertainty on tensor-to-scalar ratio for a few Stage-IV CMB configurations, plotted as a function of sky coverage. Forecast for two beam sizes (1' and 8' FWHM) and two overall sensitivity goals (3.5 and 1.1 $\mu$K/$\sqrt{s}$) are displayed. Foregrounds are assumed to be cleaned to 10% or 5% residual level (in map). We find the stated goal of $\sigma(r) < 10^{-3}$ is achieved or exceeded.

### 2.1.1 Current Experimental Efforts and Upgrades: Stages II and III

The current generation CMB polarization experiments (Stage II) consists of a suite of complementary experiments that employ a variety of experimental approaches to take steps toward searching for an inflationary $B$-mode signal. There are a number of experiments that observe the sky on degree angular scales, and are
specifically designed to directly measure inflationary $B$ modes. If the level of tensor modes is near the upper limit of what is still allowed by recent Planck results [31], these Stage II experiments have the sensitivity required to measure a $B$-mode signal.

The BICEP/Keck Array [67] program at the South Pole is a set of compact, refractive polarimeters currently taking data on the sky at degree angular scales and has already accumulated high raw sensitivity at 150 GHz. The BICEP program will expand its frequency coverage to include 100 GHz and 220 GHz, adding a Stage-III experiment in 2015. The balloon borne EBEX [68] experiment has already collected data over 6000 square degrees with three frequency bands between 150 and 410 GHz in a flight launched in December 2012. SPIDER [69] and PIPER, also balloon-borne, will have first light at the end of 2013 and 2014, respectively, and will observe large areas of the sky at frequencies between 100 and 600 GHz. Although the integration time of balloon-borne experiments is much shorter than ground-based experiments, they benefit from dramatically reduced atmospheric contamination and detector loading. The ABS [70] experiment observes in Chile at degree angular scales and CLASS will begin observations in 2014. All of these experiments employ different and complementary approaches to mitigation of systematic uncertainties.

Because of their large beams and limited frequency coverage, these degree-angular-scale Stage-II and Stage-III programs are expected to be limited ultimately by foregrounds and the gravitational lensing signal. Pushing harder on inflationary science requires input from experiments with higher angular resolution and broader frequency coverage to disentangle a primordial $B$-mode signal from a foreground or gravitational lensing signal. Joint analysis with higher angular resolution data from Stage-II and Stage-III experiments such as POLARBEAR [71], SPTpol [72], and ACTpol [73] will become critical in the near future when lensing becomes the limiting factor in the search for the inflationary signature. Partnerships have already formed between complementary experiments to begin joint analysis of data sets for delensing. Of these experiments that have sensitivity on arcminute angular scales, POLARBEAR and SPTpol have been on-sky for over a year—SPTpol recently reported the first detection of lensing $B$ modes [74]—and ACTpol is currently being deployed. Stage-III upgrades for all three of these experiments are planned in the next few years.

2.1.2 The Stage-IV Experimental Goal on the Tensor-to-Scalar Ratio: $\sigma(r) = 0.001$.

The Stage-IV CMB experiment aims to significantly improve sensitivity over current efforts and to provide a conclusive experimental discrimination of large-field and small-field inflation. The proposed sensitivity goal for a Stage-IV CMB experiment is to measure the tensor-to-scalar ratio to $\sigma(r) = 0.001$ (stat.), with a similar level of systematic uncertainty budgeted. With this level of uncertainty, a Stage-IV CMB experiment will unambiguously detect (CL > 5$\sigma$) tensor modes from any large-field inflation model with $r > 0.01$. Conversely, a null result at the level of $r < 0.002$ would rule out large-field inflation. A Stage-IV experiment will determine if inflation happened on the GUT scale or not. This goal is achievable with a wide range of Stage-IV design parameters targeted at measuring the degree-scale feature in the $B$-mode spectrum ($\ell \sim 100$) known to be accessible from the ground.

If $r$ is near the 2$\sigma$ Planck limit, as in the case of $m^2\phi^2$ inflation, current funded Stage-III experiments will detect $B$-mode polarization with high significance. In that scenario, a Stage-IV survey can be designed to measure the tensor amplitude with high precision and directly map quantum fluctuations produced during inflation. Further characterization of tensor properties, such as scale-invariance and level of non-Gaussianity, will be possible with Stage IV’s superb sensitivity. These tests, even with relatively modest precision, will constitute striking qualitative verifications of inflation.
2.1.3 Survey Considerations for a Stage-IV experiment

To achieve the design goals of the Stage-IV experiment outlined in Section 2.1.2 will require a major advancement in raw sensitivity and tight control of instrumental systematics. Measuring the tensor-to-scalar ratio \( r \) below a level of \( \sigma(r) = 0.01 \) requires mapping at least hundreds of square degrees of sky to noise levels well below 10 \( \mu \)K in a 1-arcmin pixel (or 10 \( \mu \)K-arcmin), which is the lowest noise level achieved in any current CMB observation of a patch of sky larger than a few square degrees. Effectively cleaning the lensing \( B \) modes from the inflationary signal also places stringent requirements on sensitivity. Significant delensing is only possible with a \(<10'\) beam and \(<5 \mu\)K-arcmin noise in the \( B \)-mode map \[75\]. This noise level and resolution across a significant area of sky (at least 1000 deg\(^2\)) would reduce the lensing limit on the measurement of \( r \) to below \( 10^{-3} \).

A Stage-IV polarized CMB experiment must be designed with a few key features in mind. It must have multiple observing bands to separate polarized foreground signals. It must have the sensitivity required on degree angular scales to observe the primordial signature and on arcminute angular scales for effective delensing. It must maintain a clean instrumental and optical design to reduce spurious polarized signals. A balance must be found between the size of the observed sky region and the depth to which that region is mapped.

The size of the observed patch is determined by the angular scale of the primordial \( B \)-mode feature accessible to ground based instruments, the so-called “recombination bump” at multipole \( \ell \sim 100 \), or angular scale \( \theta \sim 2° \). According to theoretical predictions, the inflationary tensor signal at very large angular scales, the “reionization bump” at \( \ell < 10 \), exceeds lensing even for low values of \( r (r < 10^{-3}) \). Predicted galactic dust foreground at these large angular scales is expected to be a factor of 30 (in temperature) larger than the signal for \( r = 0.001 \) (at 90 GHz, for 5% dust polarization and for 75% of the sky \[76\]). Information from \textit{Planck} will provide more information about polarized foregrounds. In this document we are describing a CMB-S4 experiment that will extract the primordial \( B \)-mode signal from a deep survey over only a few percent of relatively clean sky.

Lensing-induced \( B \) modes have the same frequency dependence as the CMB and cannot be distinguished by multi-frequency observations. The lensing signal peaks at \( \ell \sim 1000 \) and tends to increase the optimal survey width, because covering more sky allows lensing confusion to be debiased (i.e., subtracted in power-spectrum space) in the same way instrumental noise is removed in temperature power spectrum measurements. This is possible because the lensing amplitude is determined to within 3% by constraints on cosmological parameters from \textit{Planck}’s temperature and \( E \)-mode measurements, and the uncertainty is expected to decrease dramatically, using the arcminute-scale \( B \)-mode survey enabled by a Stage-IV experiment. Alternatively, the lensing deflection field can be reconstructed from arcminute-scale \( B \)-mode measurements, and the expected lensing contamination to degree-scale \( B \) modes predicted and subtracted from the observed \( B \)-mode map. This procedure requires arcminute-resolution \( B \)-mode surveys (which in turn require large (>1m) primary apertures), but it complements debiasing because very different assumptions are made in the lensing removal process.

For simplicity we consider a survey with uniform coverage at a given angular resolution. The degree-scale and arcminute-scale measurements are provided by the same Stage-IV experiment: an array of platforms each fed by > 10k background-limited detectors. We use the \textit{Planck} Sky Model (PSM) to predict the level of foreground contamination at various frequencies. For each \( f_{\text{sky}} \) (fraction of sky covered) considered, the cleanest patch is identified in terms of foreground level at 95 GHz, which is near the minimum in foreground contamination vs. observing frequency \[76\]. In lieu of map-based component separation, we assume a level of residual foregrounds corresponding to 10% or 5% of the PSM. It is expected that this level of foreground removal will be achievable with good frequency coverage. We follow the prescription in \[77\] to predict
residual lensing contamination after ideal maximum-likelihood delensing [75]. We conservatively remove the first Fourier bin in each $f_{\text{sky}}$ considered to simulate loss of information caused by map boundaries.

Results of these forecasts for a few representative Stage-IV configurations are shown in Figure 3. The goal of $\sigma(r) = 10^{-3}$ is achieved or exceeded for a variety of configurations. The constraint relies more significantly on delensing for smaller sky fractions and on debiasing for larger sky fractions. Both methods are equally valid, and as a result, the formal $\sigma(r)$ evaluated at the optimal $f_{\text{sky}}$ is relatively insensitive to the resolution of the experiment: for an experiment with $1.1\mu K/\sqrt{s}$ sensitivity, $\sigma(r)$ improves modestly from $2.9 \times 10^{-4}$ to $1.8 \times 10^{-4}$ when the beam size decreases from 8′ to 1′ FWHM. However, it is important to point out that only delensing enables high S/N mapping of the inflationary $B$ modes if $r < 0.02$, a level that below which debiasing can only provide statistical measurements.

### 2.2 Measuring the Spectrum of Primordial Density Fluctuations

While the tensor-to-scalar ratio $r$ is in many ways the holy grail of inflationary CMB science, it is not the only constraint on inflation models provided by the CMB. The dynamics of inflation also leave an imprint on the power spectrum of scalar density fluctuations. The simplest inflation models predict a nearly scale-invariant power-law spectrum, and the most basic observational information about the primordial power spectrum is the power-law slope.

Inflation is a nearly time-translation invariant state; however this invariance must be broken for inflation to eventually come to an end. In the inflationary paradigm, the wavelength of perturbations depends solely on the time that they were produced, thus a time-translation invariant universe would produce scale-invariant power spectrum of scalar density fluctuations. The simplest inflation models predict a nearly scale-invariant power-law spectrum, and the most basic observational information about the primordial power spectrum is the power-law slope.

The value of $dn_s/d\ln k$ is often referred to as the “running” of the scalar spectral index. The running parameter is predicted to be undetectable by most inflationary theories, and a detection of non-zero running could provide information about the inflationary potential or point to models other than inflation. The Stage-IV experiment described in the previous section should be able to significantly improve the current constraints on running, particularly through an exquisite measurement of the $E$-mode damping tail. Similar measurements can be made using clustering of galaxies in combination with CMB data. Current measurements from Planck [28] indicate $dn_s/d\ln k = -0.0134 \pm 0.009$. For Planck in combination with BOSS and eBOSS, we forecast error on running to 0.006 improving to 0.004 for combination with DESI. Forecasts for how well the Lyman-$\alpha$ forest measurements will do are highly uncertain and depend on the modeling.

\[ n_s(k) = n_s(k_0) + d n_s/d \ln k \ln \left( \frac{k}{k_0} \right). \]  

\[ (11) \]

\[ 2 \text{Scale invariance here means that the contribution to the rms density fluctuation from a logarithmic interval in } k, \text{ at the time when } k = aH, \text{ is independent of } k. \text{ Here } a(t) \text{ is the scale factor and } H \equiv \dot{a}/a \text{ is the Hubble parameter.} \]
advances but we are predicting error of 0.002 for DESI Lyman-\(\alpha\) forest measurements. These errors are larger than naturally predicted by inflation, so new physics will have to be at work if we detect running at these levels. Experiments sensitive to running will also have the power to search for features in the power spectrum beyond the first few terms of a Taylor expansion. Such features are predicted in certain inflation models with sharp features in the inflationary potential or interactions during inflation.

2.3 Beyond the Power Spectrum: Constraining Inflation through Higher-order Correlations

In the inflationary paradigm, microscopic quantum fluctuations are the seeds of all the structure we see in the CMB and traced by collapsed objects like galaxies and clusters of galaxies. Due to the nature of these quantum fluctuations (specifically their random phase), we expect the resulting distribution of density perturbations in the post-inflation universe to be very nearly Gaussian (as long as they are still in the linear regime). In particular, the standard class of single-field, slow-roll inflation models predict a level of non-Gaussianity of \(f_{\text{equi}}^{\text{NL}} < 1\) [4, 78] and \(f_{\text{local}}^{\text{NL}} = 0\) [79], where \(f_{\text{local}}^{\text{NL}}\) and \(f_{\text{equi}}^{\text{NL}}\) are parameters encoding the level of correlation between sets of three Fourier modes in particular triangle configurations.

However, many models of inflation do predict a detectable level of non-Gaussianity. This means that a detection of primordial non-Gaussianity would rule out single-field slow-roll models, and that increasingly tight upper limits will rule out many other models.

The current best limits on primordial non-Gaussianity are obtained using data from the Planck satellite [62]: \(f_{\text{local}}^{\text{NL}} = 2.7 \pm 5.8\), \(f_{\text{equi}}^{\text{NL}} = -42 \pm 75\) and \(f_{\text{orthogonal}}^{\text{NL}} = -25 \pm 39\). At the angular scales that contribute most of the weight to the \(f_{\text{NL}}\) constraints, Planck has measured the CMB temperature fluctuations as well as they can be measured (i.e., the constraints on \(f_{\text{NL}}\) is now limited by cosmic variance, not noise). Adding CMB polarization information will improve this constraint, but at most by \(\sqrt{3}\).

2.4 Non-Gaussianity from large scale Structure Measurements

Non-Gaussian features in the primordial fluctuations modulate the subsequent evolution of structure in the universe and give rise to effects that are present in the statistics of the density field at all redshifts. The signature of non-Gaussianity is therefore expected to appear in the abundance of massive clusters and in the large scale distribution of galaxies and other tracers of dark-matter fluctuations. Large spectroscopic programs such as BOSS measure the LSS field in three dimensions over a very large comoving volume, and thus offer a considerably larger number of modes than the CMB. The suite of large-area, spectroscopic BAO surveys expected to occur between 2009 and 2022 (BOSS, eBOSS and DESI) will therefore provide valuable constraints that will complement the current Planck bispectrum limits and future CMB limits on the tensor to scalar ratio.

As opposed to the CMB, where non-Gaussianity is derived primarily from considering higher-order correlators, in wide-field optical spectroscopic surveys, non-Gaussianity can be constrained through the two-point function. The easiest way to understand this is to consider the tracer formation process as a non-linear transformation that brings higher-order moments into the two-point function. In particular, it has been shown that the typical tracers of cosmic structure produce a quadratic divergence in power at large scales that is proportional to \(f_{\text{local}}^{\text{NL}}\) [80]. This has been used several times to place independent constraints on non-Gaussianity (see e.g., [81, 82, 83, 84, 85] and references therein). However, even the most competitive
limits from this technique ($-37 < f_{\text{NL}}^\text{local} < 25$ in [85]) are dwarfed by the \textit{Planck} results. Predictions for BOSS, eBOSS and DESI indicate 68\% confidence limits $\Delta f_{\text{NL}}^\text{local} = 24$, 12 and 5, respectively. While there is some scope for improvement with better techniques that reduce the impact of sample variance ([86]), the non-Gaussianity limits from galaxy power-spectra remain useful mostly as a completely independent cross-check of CMB limits.

Non-linear evolution of the dark-matter fluctuations and the details of how sources trace the underlying field give rise to additional, non-primordial, sources of non-Gaussianity in the evolved galaxy field. This evolution must obey certain physical constraints (for example, gravitational evolution and redshift space distortions can only affect certain triangles of the 3-point correlation function). Due to these constraints on non-linear density growth, information from primordial non-Gaussianity is preserved even at smaller scales. Higher-order statistics such as the 3-point and 4-point correlation functions should provide superior constraints to those derived from the power spectrum. Constraining these correlation functions is especially important for equilateral and orthogonal type non-Gaussianity, which cannot be tested through the power spectrum [87]. While we do not have a reliable code to project non-Gaussianity constraints in this way, it is believed that these should provide constraints at least as competitive as the best projected CMB constraints, provided that systematics in target selection and uniformity of spectroscopic observations over large areas can be brought under control. We note however, that no constraint on non-Gaussianity from higher order statistics has yet been demonstrated with current data.

3 Conclusions

The next generation of CMB and LSS experiments are poised to dramatically increase our understanding of fundamental physics and the early universe by probing the inflationary epoch. In particular, constraints on the amplitude of tensor modes will provide unique insight into the physics of inflation only available through CMB observations. A Stage-IV CMB experiment such as CMB-S4 that surveys > 1\% of the sky to a depth of $\sim 1 \mu\text{K-arcmin}$ will deliver a constraint on the tensor-to-scalar ratio $r$ that will result in either a 5\sigma measurement of the energy scale of inflation or will rule out all large-field inflation models, even in the presence of foregrounds and the gravitational lensing $B$-mode signal. Such an experiment is technically feasible on the timescale envisioned in this document. The next generation of LSS measurements will complement the CMB effort by providing improved constraints on the cosmological parameters associated with inflationary models. Including all spectroscopic surveys through DESI, LSS measurements are expected to improve current constraints on running of the spectral index by up to a factor of four, improve constraints on curvature by a factor of ten, and provide non-Gaussianity constraints that are competitive with the current CMB bounds.
References


Community Planning Study: Snowmass 2013


