The Detectability of Departures from the Inflationary Consistency Equation

Yong-Seon Son and Lloyd Knox

Department of Physics, One Shields Avenue, University of California, Davis, California 95616
(Dated: March 20, 2022)

We study the detectability, given CMB polarization maps, of departures from the inflationary consistency equation, \( r \equiv T/S \sim -5n_T \), where \( T \) and \( S \) are the tensor and scalar contributions to the quadrupole variance, respectively. The consistency equation holds if inflation is driven by a slowly-rolling scalar field. Departures can be caused by: 1) higher-order terms in the expansion in slow-roll parameters, 2) quantum loop corrections or 3) multiple fields. Higher-order corrections in the first two slow-roll parameters are undetectably small. Loop corrections are detectable if they are nearly maximal and \( r \gtrsim 0.1 \). Large departures (\(|\Delta n_T| \gtrsim 0.1\)) can be seen if \( r \gtrsim 0.001 \). High angular resolution can be important for detecting non-zero \( r + 5n_T \), even when not important for detecting non-zero \( r \).

PACS numbers: draft

I. INTRODUCTION

Inflation is the most promising paradigm for explaining our flat, old and structure-filled Universe. Recent observations of the CMB temperature power spectrum confirm that the scalar perturbation spectrum is nearly scale-invariant, as predicted. Further, the detection of the temperature/polarization E-mode anti-correlation on degree scales is a strong indication of correlations on length scales larger than the classical horizon.

In addition to the scalar spectrum of perturbations studied so far, inflation also produces tensor perturbations. Several exciting possibilities could come from study of the amplitude and shape of this tensor perturbation spectrum: 1) single-field slow-roll inflation can be verified through confirmation of the consistency equation; 2) the presence of loop corrections can be inferred (from small, but detectable, departures from the consistency equation) and used to constrain the more fundamental physics underlying the effective field theory description of the inflaton; or 3) single-field slow-roll inflation can be ruled out if the departures from the consistency equation are larger than can come from loop corrections. Here we quantify how well the tensor spectrum can be measured (for varying sensitivity and angular resolution of CMB observations) and discuss these various possibilities.

That there must be a consistency equation can be seen from a degrees-of-freedom counting argument. For a single slowly-rolling scalar field, to leading order in the expansion in slow roll parameters, there are only three important parameters: the Hubble parameter and its first two derivatives with respect to the scalar field \( \phi \). Since these three parameters control four observables (the amplitude and power-law indexes of the tensor and scalar perturbation spectra) these four observables cannot be independent and indeed are related by \( r \sim -5n_T \).

The scalar field driving inflation is most likely the scalar field of an effective field theory (EFT), which may receive large corrections near some high energy scale, \( M \). The degrees of freedom at this higher scale can generate quantum loop corrections to the effective Lagrangian which will lead to departures from the consistency equation.

We show here that these departures may be large enough to be detectable, via CMB polarization observations, if \( r \gtrsim 0.1 \). Thus there may be an observational window on the more fundamental physics underlying the inflaton. For this window to be open, \( M^2 \) must not be much larger than \( H^2 \). Holographic considerations may place an upper bound on \( M \).

We further show that the quantum loop corrections envisioned in cannot lead to large departures from the consistency equation. Thus large departures cannot be confused with these loop corrections, but would clearly signal the failure of a single-field description.

We discuss the implications of our results for observation strategies. Although the high resolution required to reduce the gravitational lensing contamination of the tensor signal is not necessary for measurement of the amplitude of the tensor spectrum when \( r > 0.1 \), it can make a significant difference for measurement of the shape. However, for \( r \sim 0.01 \) high resolution has very little benefit since other noise sources dominate.

We concentrate solely on CMB observations because these are likely the only observations that can be used to detect the influence of tensor perturbations from inflation. Although direct detection by space-based interferometers has been discussed, such a mission is at least several decades away and it is likely that foreground signals (from merging massive black hole binaries) will dominate the primordial signal. If direct detection were possible, in combination with CMB observations it would be enormously valuable for measuring the shape of the tensor spectrum since the length scales probed differ by 10 orders of magnitude.

In section I we introduce the consistency equation to...
leading order and its next order corrections. In section II we discuss the loop corrections. In section III we discuss CMB observations and how the tensor spectrum can be recovered from them. In section IV we present our detectability limits for departures from the consistency equation and the presence of loop corrections.

II. THE CONSISTENCY EQUATION

We consider a single scalar field slow-roll inflation model \cite{12}. For a review of the spectrum of scalar and tensor perturbations produced by a slow-roll scalar field, see \cite{13}. The equation of motion for the single scalar field in an expanding universe is

\[
\ddot{\phi} + 3H\dot{\phi} + V' = 0,
\]

where a dot denotes the derivative in terms of a physical time and a prime denotes the derivative in terms of \(\phi\). The Friedmann equation with the scalar field is

\[
H^2 = \frac{8\pi}{m_{pl}^2} \left( \frac{1}{2} \dot{\phi}^2 + V(\phi) \right).
\]

The slow-roll parameters, \(\epsilon_0\) and \(\eta_0\), which we use here are defined with the Hubble parameter \(H\) rather than potential \(V\) \cite{14}, in order to clarify the relation with the loop corrections. They are given by

\[
\epsilon_0 = \frac{m_{pl}^2}{4\pi} \left( \frac{H'}{H} \right)^2,
\]

\[
\eta_0 = \frac{m_{pl}^2}{4\pi} \frac{H''}{H}.
\]

The square roots of the resulting scalar and tensor power spectra are, to next order in the slow-roll parameters \(\epsilon_0\) and \(\eta_0\) \cite{15,16},

\[
A_S^0(k) = \frac{2}{5} \left[ 1 - (2C + 1)\epsilon_0 + 2\eta_0 \right] \frac{2}{m_{pl}|H|} \left| H \right|_{k=aH},
\]

\[
A_T^0(k) = \frac{1}{10} \left[ 1 - (C + 1)\epsilon_0 \right] \frac{4}{\sqrt{\pi}} \frac{H}{m_{pl}} \left| \frac{H}{k} \right|_{k=aH},
\]

where \(C \simeq -0.73\). The ratio between \((A_S^0(k))^2\) and \((A_T^0(k))^2\) is simply \(\epsilon_0\) to zeroth order. The full description to the next order is \cite{13}

\[
\epsilon_0 \simeq \frac{(A_T^0)^2}{(A_S^0)^2} \left[ 1 - 2C(\epsilon_0 - \eta_0) \right].
\]

This is the first step toward getting the consistency equation in a simple algebraic form. The ratio of both amplitudes turns into a slow-roll parameter \(\epsilon_0\).

The tensor power spectrum is determined by an inflationary energy scale represented by \(H\). Thus its spectral index only includes the first order derivative to leading order. The spectral index \(n_T^0\) for the tensor power spectrum in the next order is \cite{17}

\[
\epsilon_0 \simeq -2\epsilon_0 \left[ 1 + (3 + 2C)\epsilon_0 - 2(1 + C)\eta_0 \right].
\]

Two degrees of freedom of the primordial perturbations, \(H\) and \(\epsilon_0\), describe the three observables, \(A_S^0\), \(A_T^0\) and \(n_T\). There is a single equation which relates these observables. It can be written as

\[
n_T^0 + 2\frac{(A_T^0)^2}{(A_S^0)^2} \simeq 0,
\]

to zeroth order. If we consider the possible departure in the next order, the consistency equation is \cite{13}

\[
n_T^0 + 2\frac{A_T^0}{A_S^0} - 2\frac{A_T^4}{A_S^0^2} + 2\frac{A_T^2}{A_S^0} \left( 1 - n_S^0 \right) \simeq 0.
\]

The next order corrections can be estimated from the observables, such as \((A_T^0/A_S^0)^2\) and \(n_S\). Such a departure is not unknown quantity like the loop corrections which we present in the next section.

We expect higher-order derivatives to be negligible. If they are not then we could tell from observing the scalar spectrum and modify our analysis accordingly.

III. THE LOOP CORRECTIONS

The large corrections to EFT appearing near the higher energy scale \(M\) may leave an imprint on the CMB. They appear in the effective Lagrangian due to loop corrections proportional to the even powers of \(H/M\) where \(H < M\) \cite{18}. The detectability of such ‘short distance physics’ will be possible only if \(M\) is far less than \(M_{Pl}\). Some theories allow such a low scale of \(M\).

The variance of density perturbations \(\delta \rho/\rho\) is proportional to the mean-square spectrum of fluctuations of the scalar field as

\[
\left( \frac{\delta \rho}{\rho} \right)^2 \propto \left( \frac{H}{\phi} \right)^2 \left( \frac{\delta \phi}{\phi} \right)^2.
\]

The quantum fluctuation \((\delta \phi)^2\) is given by the equal time two point function \(\langle \phi(\rho)\phi(-\rho) \rangle\) when the mode crosses out the horizon. The physical momentum \(p\) is equal to \(H\) at the horizon crossing. Thus we have

\[
(\delta \phi)^2 \sim \langle \phi(\rho)\phi(-\rho) \rangle \big|_{p=H},
\]

where the two point function is defined as

\[
\langle \phi(\rho)\phi(-\rho) \rangle = \frac{1}{2}(p^2 + H^2),
\]

in de Sitter space.

The loop corrections to to \(\langle \phi(\rho)\phi(-\rho) \rangle\) are \cite{19}.

\[
\langle \phi(\rho)\phi(-\rho) \rangle \big|_{p=H} = H^2 \left( 1 + \frac{H^2}{M^2} + \cdots \right).
\]
where the curvature is kept close to $H^2$ by fine-tuning.

The appearance of the loop corrections in the density perturbations leads to a shift in the observable quantities from the zeroth order, as

$$
A_S = A_S^0 \left(1 + \chi_S \frac{H^2}{M^2}\right)
$$

$$
A_T = A_T^0 \left(1 + \chi_T \frac{H^2}{M^2}\right)
$$

$$
n_T = n_T^0 \left(1 + 2\chi_T \frac{H^2}{M^2}\right),
$$

where we separate $\chi$ into $\chi_S$ for the scalar and $\chi_T$ for the tensor. The shift in $A_S$ leads to an alteration of the consistency equation such that

$$
n_T + 2 \left(\frac{A_T}{A_S}\right)^2 \approx -2\epsilon_0 \chi_S \frac{H^2}{M^2}.
$$

Thus an observed violation of the consistency equation is possibly a signature of short distance degrees of freedom affecting the inflaton $\phi$.

There are many different ways in which the EFT could receive large corrections near some mass scale $M$. It could happen simply from a Yukawa interaction with a fermion of mass $M$. Or, if we have the high dimensional theory with proper compactification, then $M$ is the reduced Planck scale. In $M$ theory, the fundamental scale possibly approaches the scale of $H$ and can give us the detectable loop corrections, $\chi(H^2/M^2) > 0.1$. Holographic considerations also suggest large corrections to EFT at short distances.

With $|\chi| < 1$, we have the maximum departure from the loop corrections as

$$
\left| n_T + 2 \left(\frac{A_T}{A_S}\right)^2 \right| < 2\epsilon_0,
$$

since $H^2/M^2$ is always less than unity. At $|\chi(H^2/M^2)| \sim 1$, the consistency equation is maximally broken by the loop corrections. As we discuss below, this maximal departure is larger than any other contribution in the single scalar field slow-roll inflationary model.

**IV. FORECASTING DETECTABILITY LIMITS**

The cosmic microwave background (CMB) anisotropies are polarized at the last-scattering surface due to the quadrupolar temperature fluctuations seeded by the almost scale-invariant density fluctuations at horizon crossing. The different types of polarization are expected from the different sources of the quadrupolar temperature fluctuations: scalar, vector and tensor quadrupole anisotropy. The observable polarization patterns are separated into the gradient E-mode polarization having even parity and the curly B mode polarization having odd parity.

![FIG. 1: The tensor B mode power spectrum for $r = 0.1$ and $\tau = 0.17$ (solid) and the scalar B mode power spectrum (long dashes). The noise power spectrum for an experiment with $\Delta_P = 3\sqrt{2}\mu K \cdot \text{arcmin}$ and $\theta_b = 3'$ is shown with the dotted line. The short-dashed curve is the residual scalar B mode power spectrum after cleaning of the lensing contaminant by such an experiment.](image)

The different sources of the quadrupolar temperature fluctuations contribute the polarization patterns in their own way. The scalar source leads to E-mode primarily, but the weak lensing effect due to the mass distribution along lines of sight between the observer and the last scattering leads to the transformed secondary B mode of which amplitude is naturally smaller than the overall amplitude of the primary polarization. Vector perturbations have no growing modes in linear perturbation theory and thus are not expected to be significant in inflationary models for either E or B modes. The tensor source expected from the inflationary model has both E and B patterns in roughly equal magnitude. Thus the B mode has the highest ratio of tensor-to-scalar fluctuation power and is what we consider.

We show the scalar power spectrum $C_{BB}^s$ and the tensor power spectrum $C_{rr}^s$ in Fig. 1. With $r = 0.1$, $C_{BB}^s$ is greater than $C_{BB}^s$ at $l < 100$. The bump in $C_{BB}^s$ at $l < 20$ appears due to reionization. The amplitude is proportional to the square of the optical depth, $\tau$. Here we have set $\tau = 0.17$. The detectability of the tensor B mode is enhanced by the reionization bump.

The primary temperature and polarization maps are distorted by the gradient of the projected lensing potential $\phi$. We can estimate $\phi$ from the 4-point function of the temperature and polarization fields. With $\phi$ estimated from the lensed maps, we can clean the lensed B
mode. The cleaned and uncleaned $C_{lB}^S$ are shown in Fig. 1. The residual lensing-induced B mode after cleaning is up to 10 times smaller than the uncleaned lensing-induced B mode. The minimum detectable limit of $r$ which we can achieve from the mass reconstruction is close to $2 \times 10^{-5}$ which is 10 times better than with no cleaning.\cite{23,24}.

We use the the following cosmological parameters : $\Omega_m = 0.34$, $\Omega_b = 0.05$, $\Omega_V = 0.66$, $h = 0.66$, $\tau_{\text{reion}} = 0.17$ and $\sigma_8 = 0.86$, and let $n_T$ and $r$ be free parameters. We use the high sensitive future CMB experiments with noise levels, $\Delta_P/\sqrt{2} = \Delta_T$ in the range $\Delta_T = 1 \mu K \cdot \text{arcmin}$ to $15 \mu K \cdot \text{arcmin}$, angular resolution in the range $\theta_h = 1.0'$ to $30.0'$ and full sky coverage $(f_{s,sky} = 1)$. Here $\Delta_P = \omega^{-1/2}$ and $\omega$ is the weight per solid angle for the $Q$ and $U$ linear polarization Stokes parameters.

The variance of B mode CMB power spectrum, $\Delta C_l^B$, is given by

$$\Delta C_l^B = \sqrt{\frac{2}{(2l+1)f_{s,sky}} \left( C_l^{B,T} + C_l^{B,S} + N_l \right)},$$

where $C_l^{B,T}$ is the tensorial B mode and $C_l^{B,S}$ is the scalar B mode from lensing either before or after cleaning as in $\cite{24}$. $N_l$ is the experimental noise which is given by

$$N_l = \left( \frac{\pi}{180 \times 60 T_{\text{CMB}}} \right)^2 e^{2\theta_h^2/(8 \ln 2)}.$$

Based on the Gaussianity of perturbations, we use the Fisher matrix analysis to estimate the detectability

$$F_{pp'} = \sum_l \frac{\partial C_l^B}{\partial p} \left( \Delta C_l^B \right)^{-2} \frac{\partial C_l^B}{\partial p'},$$

where $p$ and $p'$ enumerate the cosmological parameters we consider here. The diagonal elements of the $F_{pp'}^{-1/2}$ are the 1-$\sigma$ errors for the parameters $\cite{27}$.

The slow-roll parameter $\epsilon$ can be replaced by the measurable quantity $r$, the tensor to scalar ratio $T/S$, by using the fitting formula $\cite{28,29}$,

$$\epsilon = \frac{1}{14} \frac{1.04 - 0.82 \Omega_V + 2 \Omega_V^2}{1.0 - 0.03 \Omega_V - 0.1 \Omega_V^2} r.$$

The above fitting formula includes the non-negligible contribution to the quadrupole anisotropy from the late-time ISW effect.

\section{RESULTS AND DISCUSSION}

Before turning to our detectability limits, we survey the $r$ vs. $n_T + 2 (A_T)^2 / (A_S)^2$ plane. We first rewrite the consistency equation Eq. $\cite{19}$ with Eq. $\cite{19}$

$$n_T + \frac{1}{4.8} r = - \frac{1}{4.8} r r_l,$$  

FIG. 2: The solid lines bound the possible departures from the consistency equation for models with a single scalar field. The shaded area is excluded by observations of the temperature power spectrum $\cite{34}$.

where $r_l$ is the loop corrections, $r_l = \chi_S H^2 / M^2$. The solid line in Fig. 2 is given by setting $r_l = 1$, its maximum value. These are largest possible departure to the consistency equation. If the observed value of $n_T + r/4.8$ is outside the two solid lines in Fig. 2, the single scalar field slow-roll inflation model can be excluded. If the observed value of $n_T + r/4.8$ is inside the two solid lines and yet different from zero, we may have a chance to study physics at distances shorter than $1/M$. The current upper bound on $r$ from the temperature power spectrum is 0.71. $\cite{31}$.

Our detectability limits are shown in Fig. 3. We fixed all parameters except $r$ and $n_T$ and calculated the Fisher matrix with $r$ varying from $10^{-5}$ to 0.71. We show the error on the combination $n_T + r/4.8$ since we are interested in testing the consistency equation, but $\sigma(n_T) \simeq \sigma(n_T + r/4.8)$ since $\sigma(n_T) \gg \sigma(r)/4.8$. As expected, $\sigma(n_T)$ decreases as $r$ increases and raises the signal level.

From Fig. 3 we see that at $r < 0.1$, any possible theoretical correction to the consistency equation for slow-roll single field models is beneath the detectability limit. Therefore, observing non-zero $n_T + r/4.8$ and $r < 0.1$ will lead us to exclude the single slow-roll scalar field inflation scenario. At $r > 0.1$, the loop correction terms begin to show up above the detectability limit. In this case, the broken consistency equation does not necessarily mean the failure of the single slow-roll scalar field inflation. The solid straight line indicates the maximal loop cor-
The cosmic variance limit is the solid line in both panels. In the upper panel, we fix $\Delta_P = 3\sqrt{2}\mu K \cdot \text{arcmin}$ and vary the angular resolution: $\theta_b = 1'$ (long dash), $\theta_b = 3'$ (dash) and $\theta_b = 5'$ (dots). In the lower panel, we fix the angular resolution at $\theta_b = 3'$ and vary the weight per solid angle: $\Delta_P = 2(\Delta_n^2/A_n^2) = \sigma_n^2$. We get the solid line by cleaning the scalar B mode power spectrum. The shaded area is excluded by observations of the temperature power spectrum [30].

In the upper panel of Fig. 3 we show the variation of $\Delta_P$ with fixed angular resolution. As $\Delta_P$ increases, the noise power becomes larger than the cleaned, and then even the uncleaned, scalar B mode power. This increased noise reduces the maximum observable tensor $l$ and adds significant noise all across the tensor B mode spectrum rise from $l \approx 20$ to $l = 100$. Thus there is a strong sensitivity to increases in the noise above our fiducial value.

The cosmological parameter with the most impact on the tensor spectrum is $\tau$. Although $\tau$ can make a big difference for the detectability limit of $\Delta_P$, it has little impact on the detectability limits for $n_T + r/4.8$. The $r$ limit is improved by increased $\tau$ since the ‘reionization bump’ at $l < 20$ has $C_l \propto \tau^2$. In contrast, for acceptable values of $\tau$, and $r \gtrsim 0.01$, the error in $n_T + r/4.8$ is dominated by uncertainties in the tensor power spectrum at $l > 20$. Here the only effect of reionization is a suppression of power by $\exp(-2\tau)$.

Our analysis has ignored polarized emission from galactic and extragalactic sources. Multi–frequency observations can be used to clean out these signals based on their distinct spectral shapes. However, residual contamination is unavoidable and will also limit the ability of observations to study the consistency equation. As we learn more about polarized foreground emission, these will likely have a big impact on observing strategies and forecasted $n_T + r/4.8$ detectability limits below some value of $r$. Our forecasts should therefore be viewed as lower limits.

We now turn our attention to the signature of the short distance physics; i.e., how well can we detect a non-zero $\phi_H^2$? Fig. 4 shows the narrow window for the detectability of $r_1$ which can be seen at $r \gtrsim 0.1$. As Kaloper et. al. pointed out, we cannot see the new physics at $r_1 < 0.1$, but some M theories with proper compactification give $r_1 > 0.1$ [7].

The window for probing the new physics is very narrow at $r > 0.1$ and $r_1 > 0.1$. But it is a clean window from any next order correction in the slow roll parameters, since the next order corrections in Eq. [5] is beneath the detectability limit at this window. We conclude it is possible to probe physics at distances near $1/M$ as long as $r > 0.1$ and $M$ is very close to $H$. The dotted line in Fig. 4 is the forecasted constraint from the uncleaned B mode with the same experiment ($\Delta_P = 3\sqrt{2}\mu K \cdot \text{arcmin} & \theta_b = 3'$). We get the solid line by cleaning the scalar B mode with the estimated lensing potential. If the reduced Planck mass is truly such a small amount, i.e. $r_1 > 0.1$, then even a small amount of improvement in $\sigma(r_1)$ will be valuable. As we see in Fig. 4, the lensing potential reconstruction improves the detectability of $r_1$ by about 50%. If we can detect the tensor power spectrum at $r > 0.1$, the uncleaned B mode can probe the tensor power spectrum well. But we will want the cleaned B mode to probe to...
even shorter distances.

Wands et al. [32] have studied the consistency equation for two and more scalar fields. For two scalar fields there is a generalized consistency equation giving the tensor perturbation to scalar curvature perturbation ratio proportional to \( n_T \) times an isocurvature correlation angle. For more than two fields this becomes an inequality, providing an upper bound on this tensor-to-scalar ratio. Thus observation of departure from this generalized consistency equation (by more than allowed by loop corrections) would rule out two–field models and an observed violation of the inequality would rule out all slow-roll models of inflation.

VI. CONCLUSIONS

The tensor perturbation spectrum is a more direct probe of inflation than the scalar perturbation spectrum and may provide us with highly valuable information about the physics of inflation. We have quantified how well idealized versions of future experiments can probe the tensor perturbation spectrum and, in particular, test the inflationary consistency equation. Detectable departures may come from additional fields or short–distance corrections to the inflaton’s effective field theory. We have shown that for \( r > 0.01 \), even though high resolution does not improve the detectability of \( r \), high resolution does improve the detectability of departures from the consistency equation.

We thank Nemanja Kaloper for his feedback on a draft version of this manuscript and Andreas Albrecht, Eugene Lim and Sayan Basu for useful conversations. This work was supported in part by NASA NAG5-11098.
B. P. Flannery, *Numerical Recipes* (Cambridge University Press, Cambridge, UK —c1992, ADDRESS, 1992).

[28] L. Knox, Phys. Rev. D 52, 4307 (1995).
[29] M. S. Turner and M. White, Phys. Rev. D 53, 6822 (1996).
[30] D. N. Spergel et al., (2003), astro-ph/0302209.
[31] C. L. Bennett et al., ArXiv Astrophysics e-prints 2207 (2003).
[32] D. Wands, N. Bartolo, S. Matarrese, and A. Riotto, ArXiv Astrophysics e-prints 5253 (2002).

[33] N. Turok, Phys. Rev. D 54, 3686 (1996).
[34] U. Seljak and M. Zaldarriaga, Astrophys. J. 469, 437 (1996).
[35] In particular this anti-correlation rules out the ‘mimic’ models [33] designed to create inflation-looking temperature power spectra without such long-range correlations.
[36] The shifts in $n_T$ and $A_T$ cancel each other.
[37] We use CMBfast [34] for all our angular power spectrum calculations.