Implications of the cosmic microwave background power asymmetry for the early universe

Article  (Accepted Version)


This version is available from Sussex Research Online: http://sro.sussex.ac.uk/id/eprint/62039/

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher’s version. Please see the URL above for details on accessing the published version.

Copyright and reuse:
Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.
Implications of the CMB power asymmetry for the early universe  
Christian T. Byrnes, Donough Regan, David Seery, and Ewan R. M. Tarrant
Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH, UK

Observations of the microwave background fluctuations suggest a scale-dependent amplitude asymmetry of roughly 2.5σ significance. Inflationary explanations for this ‘anomaly’ require non-Gaussian fluctuations which couple observable modes to those on much larger scales. In this Letter we describe an analysis of such scenarios which significantly extends previous treatments. We identify the non-Gaussian ‘response function’ which characterizes the asymmetry, and show that it is non-trivial to construct a model which yields a sufficient amplitude: many independent fine tunings are required, often making such models appear less likely than the anomaly they seek to explain. We present an explicit model satisfying observational constraints and determine for the first time how large its bispectrum would appear to a Planck-like experiment. Although this model is merely illustrative, we expect it is a good proxy for the bispectrum in a sizeable class of models which generate a scale-dependent response using a large η parameter.

INTRODUCTION

The statistical properties of the cosmic microwave background (CMB) show remarkable consistency with the paradigm of early universe inflation. But a number of troubling anomalies persist, including the large cold spot, the quadrupole–octupole alignment, and a hemispherical amplitude asymmetry. If these anomalies are primordial it is not yet clear whether they can be compatible with the simplest inflationary models which typically predict statistical independence of each multipole (see Ref. [1] and references therein). In this Letter we report results for a special set of inflationary scenarios which can accommodate the hemispherical asymmetry.

Working with the Planck 2013 temperature data, Aiola et al. demonstrated that the asymmetry could be approximately fit by a position-dependent power-spectrum at the last-scattering surface of the form [2]

$$\mathcal{P}_{\text{obs}}(k) \approx \frac{k^2 P(k)}{2 \pi^2} \left( 1 + 2 A(k) \hat{p} \cdot \hat{n} + \cdots \right),$$ (1)

where \( \hat{p} \) represents the direction of maximal asymmetry, \( \hat{n} \) is the line-of-sight from Earth, and \( A(k) \) is an amplitude which Aiola et al. found to scale roughly like \( k^{-0.5} \). Averaged over \( \ell \approx 2 \sim 64 \) it is of order 0.07. In this paper our primary objective is to explain how an inflationary model can produce an asymmetry which replicates this scale dependence.

The effect is seen in multiple frequency channels and in the older WMAP data, which makes it less likely to be attributable to an instrumental effect or foreground. Future improvements in observation are likely to be driven by polarization data, which provide an independent probe of the largest-scale modes [3].

Inflationary explanations.—Erickcek, Carroll and Kamionkowski proposed that (1) could be produced during an inflationary epoch if the two-point function at wavenumber \( k \) is modulated by perturbations of much larger wavelength [4]. This entails the presence of a bispectrum with nonempty squeezed limit, and if the amplitude is sufficiently large it would be the first evidence for multiple active light fields in the inflationary era.

This is an exciting possibility but there is significant concern that a bispectrum of this type may already be ruled out by observation. Current experiments do not measure the bispectrum on individual configurations, but rather weighted averages over related groups of configurations—and at present are most sensitive to modestly squeezed examples. Averaged over these configurations, Planck observations require the non-Gaussian component to have amplitude \( |f_{\text{NL}}|/10^5 \lesssim 0.01\% \) [5] compared to the leading Gaussian part. Meanwhile, ignoring all scale dependence, Refs. [6–9] showed that an inflationary origin would require

$$\frac{|a_{20}|}{6.9 \times 10^{-6}} \frac{|f_{\text{NL}}|}{10} \approx 6 \left( \frac{A}{0.07} \right)^2 \beta \tag{2}$$

where \( a_{20} \) is the quadrupole of the CMB temperature anisotropy, measured to be approximately \( |a_{20}| \approx 6.9 \times 10^{-6} \) [10], and \( \beta \) is a model-dependent number which would typically be rather larger than unity. Therefore Eq. (2) suggests that an inflationary scenario may require \( |f_{\text{NL}}| \gtrsim 60 \), in contradiction to measurement. If so, we would have to abandon the possibility of an inflationary origin, at least if produced by the Erickcek–Carroll–Kamionkowski mechanism. To evade this Eq. (2) could be weakened by tuning our position on the long-wavelength background to reduce \( \beta \), but clearly we should not allow ourselves to entertain fine-tunings which are less likely than the anomaly they seek to explain.

Averaged constraints.—The requirement that \( A(k) \) varies with scale gives an alternative way out which has yet to be studied in detail. It could happen that the bispectrum amplitude is large on long wavelengths but runs to small values at shorter wavelengths in such a way that the wavelength-averaged values measured by CMB experiments remain acceptable. Eq. (2) might then apply for
a small number of wavenumber configurations but would have no simple relation to observable quantities.

In this Letter we provide, for the first time, an analysis of the CMB temperature bispectrum generated by a scale- and shape-dependent primordial bispectrum which is compatible with the modulation $A(k)$. We do this by constructing an explicit model which can be contrived to match all current observations, and also serves as a useful example showing the complications which are encountered. Despite its contrivance, we expect the bispectrum produced by this model to be a good proxy for the bispectrum generated in a much larger class of successful scenarios producing scale-dependence through a large, negative $\eta$-parameter. If the model can be embedded within a viable early universe scenario, we show that it can explain the asymmetry without introducing tension with the $f_{\text{NL}}$ or low-$\ell$ amplitude constraints (the ‘Grischkum-Zel’dovich’ effect).

In this Letter we focus on the simplest possibility that the non-Gaussian fluctuations of a single field generate the asymmetry, although we allow a second field to generate the Gaussian part of the curvature perturbation. Generalizations and further details are presented in a longer companion paper [11].

**GENERATING THE ASYMMETRY**

We denote the field with scale-dependent fluctuations by $\sigma$, and take it to substantially dominate the bispectrum for the observable curvature perturbation $\zeta$. The $\zeta$ two-point function $\langle \zeta(k_1)\zeta(k_2) \rangle = (2\pi)^3\delta(k_1 + k_2)P(k)$ can depend on $\sigma$, or alternatively on any combination of $\sigma$ and other Gaussian fields. The question to be resolved is how $P(k)$ responds to a long-wavelength background of $\sigma$ modes which we write $\delta \phi^\sigma$.

**Response function.**—In Ref. [11] we show that this response can be computed using the operator product expansion (OPE), and expressed in terms of the ensemble-averaged two- and three-point functions of the inflationary model. We focus on models in which the primary effect is due to the amplitude of the long-wavelength background rather than its gradients. Since the perturbation is small it is possible to write

$$P(k, x) = P(k)\left(1 + \delta \phi^\sigma(x)\rho_\sigma(k)+ \cdots\right). \quad (3)$$

We call $\rho_\sigma(k)$ the ‘response function’. It can be regarded as the derivative $d \ln P(k)/d\sigma$. The OPE gives [11]

$$\rho_\sigma(k) \simeq \frac{1}{P(k)}[\Sigma^{-1}(k_L)]_{\lambda\lambda}B^\lambda(k, k, k_L) \quad \text{if } k \gg k_L, \quad (4)$$

where a sum over $\lambda$ is implied, and $\Sigma^{\alpha\beta}$ and $B^\alpha$ are spectral functions for certain mixed two- and three-point correlators of $\zeta$ with the light fields of the inflationary model (and their momenta), which we collectively denote $\delta \phi^\sigma$.

\begin{align}
\langle \delta \phi^\sigma(k_1)\delta \phi^\sigma(k_2) \rangle &= (2\pi)^3\delta(k_1 + k_2)\Sigma^{\alpha\beta} \quad (5a) \\
\langle \delta \phi^\sigma(k_1)\zeta(k_2)\zeta(k_3) \rangle &= (2\pi)^3\delta(k_1 + k_2 + k_3)B^\alpha. \quad (5b)
\end{align}

Eq. (4) is one of our principal new results. It enables us to extend the analysis of inflationary models beyond those already considered in the literature to cases with non-trivial, scale-dependent correlation functions. The conditions under which it applies are discussed in more detail in Ref. [11].

In the special case of a slow-roll model in which a single field generates all perturbations, it can be shown that the right-hand side of (4) is related to the reduced bispectrum,

$$\rho_\sigma(k) = \frac{12}{5} f_{\text{NL}}(k, k, k_L), \quad k \gg k_L, \quad (6)$$

where $f_{\text{NL}}(k_1, k_2, k_3)$ is defined by

$$\frac{6}{5} f_{\text{NL}}(k_1, k_2, k_3) = \frac{B(k_1, k_2, k_3)}{P(k_1)P(k_2) + 2 \text{ cyclic perms}}. \quad (7)$$

Notice that, for a generic $B(k_1, k_2, k_3)$, the reduced bispectrum defined this way has no simple relation to any of the amplitudes $f_{\text{NL}}^{\text{local}}, f_{\text{NL}}^{\text{quint}}, \text{etc.}$, measured by experiment. Eq. (7) reproduces earlier results given in the literature [6, 7, 9, 12] but does not apply for the more realistic models considered in this Letter.

**Long wavelength background.**—To model the long-wavelength background we take

$$\delta \sigma(x) \approx E P_\sigma^{1/2}(k_\lambda)\cos(k_\lambda \cdot x + \vartheta) \quad (8)$$

where $E$ labels the ‘exceptionality’ of the amplitude, with $E = 1$ being typical and $E \gg 1$ being substantially larger than typical. We take the wavenumber $k_\lambda$ to be fixed. The phase $\vartheta$ will vary between realizations, and the Earth is located at $x = 0$.

The last-scattering surface is at comoving radius $x_{ls} \approx 14,000 h \text{ Mpc}^{-1}$. Evaluating (3) and (8) on this surface at physical location $x = x_{ls} \hat{n}$, and assuming $\alpha \equiv x_{ls} k_\lambda / 2\pi \ll 1$ so that the wavelength associated with $k_\lambda$ is somewhat larger than $x_{ls}$, we obtain

$$P(k, x) = P(k)\left(1 - C(k) + 2 A(k)\frac{x \cdot \hat{k}_\lambda}{x_{ls}} + \cdots\right). \quad (9)$$

The quantities $A(k)$ and $C(k)$ are determined in terms of the response $\rho_\sigma$ and long-wavelength background by

\begin{align}
A(k) &= \pi \alpha E P_\sigma^{1/2}(k_\lambda)\rho_\sigma(k)\sin \vartheta \\
C(k) &= -\frac{A(k)}{\pi \alpha \sin \vartheta}. \quad (10a) \quad (10b)
\end{align}

\[1\] In the restricted setup we are describing, where only $\sigma$ has a non-negligible bispectrum, the sum over $\lambda$ in Eq. (4) would include the field $\sigma$ and its momentum.
Both \( A \) and \( C \) share the same scale-dependence, so it is possible that \( C(k) \) could be used to explain the lack of power on large scales \([7, 12]\). If so, the model could simultaneously explain two anomalies—although this would entail a stringent constraint on \( \alpha \) in order that \( C(k) \) does not depress the power spectrum too strongly at small \( \ell \). The relative amplitude of \( A(k) \) and \( C(k) \) depends on the unknown phase \( \vartheta \) and our assumption of the form (8), but the observation that they scale the same way with \( k \) constitutes a new and firm prediction for all models which explain the power asymmetry by modulation from a single super-horizon mode.

**BUILDING A SUCCESSFUL MODEL**

**Single-source scenarios.**—In the case where one field dominates the two- and three-point functions of \( \zeta \), the bispectrum is equal in squeezed and equilateral configurations \([13, 14]\). Therefore

\[
\rho_\sigma = \frac{12}{5} f_{\text{NL}}(k, k, k_L) = \frac{12}{5} f_{\text{NL}}(k, k, k),
\]

and the asymmetry scales in the same way as the equilateral configuration \( f_{\text{NL}}(k, k, k) \). The asymmetry can be computed using \([15]\)

\[
\frac{d \ln |f_{\text{NL}}|}{d \ln k} = \frac{5}{6f_{\text{NL}}} \sqrt{\frac{M^4 V'''}{8 \, 3H^2}},
\]

where \( r \lesssim 0.1 \) is the tensor-to-scalar ratio. To achieve strong scaling we require \( M^4 V'''/(3H^2) \gg 1 \). But within a few e-folds this will typically generate an unacceptably large second slow-roll parameter \( \eta_\sigma \), defined by

\[
\eta_\sigma = \frac{M^2 V'''}{3H^2}.
\]

Therefore it will spoil the observed near-scale-invariance of the power spectrum.

As a specific example, a self-interacting curvaton model was studied in Ref. \([16]\). This gave rise to many difficulties, including logarithmic running of \( f_{\text{NL}}(k, k, k) \) with \( k \)—which is not an acceptable fit to the scale dependence of \( A(k) \) \([2]\). Even worse, because (12) is large only when \( f_{\text{NL}} \) is suppressed below its natural value, both the trispectrum amplitude \( g_{\text{NL}} \) and the quadrupolar modulation of the power spectrum were unacceptable. In view of these difficulties we will not pursue single-source models further.

**Multiple-source scenarios.**—In multiple-source scenarios there is more flexibility. If different fields contribute to the power spectrum and bispectrum it need not happen that a large \( \eta_\sigma \) necessarily spoils scale-invariance. In these scenarios \( \rho_\sigma \) no longer scales like the reduced bispectrum, but rather its square-root \( f_{\text{NL}}(k, k, k)^{1/2} \). Therefore

\[
\frac{d \ln A}{d \ln k} = \frac{1}{2} \frac{d \ln |f_{\text{NL}}(k, k, k)|}{d \ln k} \approx \frac{d \ln (P/P)}{d \ln k} \approx 2\eta_\sigma - (n_s - 1),
\]

where \( P \) is the dimensionless power spectrum, \( n_s - 1 \approx -0.03 \) is the observed scalar spectral index and \( \eta_\sigma \) was defined in Eq. (13). If we can achieve a constant \( \eta_\sigma \approx -0.25 \) while observable scales are leaving the horizon then it is possible to produce an acceptable power-law for \( A(k) \).

For further details of these scaling estimates for \( A(k) \) see Kenton et al. \([17]\) or Ref. \([11]\).

A simple potential with large constant \( \eta_\sigma \) is

\[
W(\phi, \sigma) = V(\phi) \left( 1 - \frac{1}{2} \frac{m^2 \sigma^2}{M^2} \right).
\]

The inflaton \( \phi \) is taken to dominate the energy density and therefore drives the inflationary phase. Initially \( \sigma \) lies near the hilltop at \( \sigma = 0 \), so its kinetic energy is subdominant and \( \epsilon \approx \frac{M^2 V^3}{2V^2} \). (Here \( \epsilon = -\dot{H}/H^2 \) is the conventional slow-roll parameter.) As inflation proceeds \( \sigma \) will roll down the hill like \( \sigma(N) = \sigma_0 e^{-\eta_\sigma N} \), where \( \star \) denotes evaluation at the initial time and \( N \) measures the number of subsequent e-folds.

To keep the \( \sigma \) energy density subdominant we must prevent it rolling to large field values, which implies that \( \sigma_0 \) must be chosen to be very close to the hilltop. But the initial condition must also lie outside the diffusion-dominated regime, meaning the classical rolling should be substantially larger than quantum fluctuations in \( \sigma \). This requires \( |d\sigma/dN| \gg H_s/2\pi \). In combination with the requirement that \( \sigma \) remain subdominant in the observed power spectrum, we find that \( \sigma_0 \) should be chosen so that \( |\sigma_0| \geq \sqrt{\epsilon_s \mathcal{P}_s}/|\eta_\sigma| \). For typical values of \( \epsilon = 10^{-2} \) and \( \eta_\sigma = -0.25 \) this requires \(|\sigma(60)| \gtrsim 100\mathcal{P}_s \) which is much too large. The problem can be ameliorated by reducing \( \epsilon \), but then \( \sigma \) contributes significantly to \( \epsilon \) during the inflationary period. This reduces the bispectrum amplitude to a tiny value, or causes \( \sigma \) to contaminate the power spectrum and spoil its scale invariance \([11]\).

**A working model.**—To avoid these problems, consider a potential in which the effective mass of the \( \sigma \) field makes a rapid transition. An example is

\[
W = W_0 \left( 1 + \frac{1}{2} \eta_\sigma \frac{\phi^2}{M^2} \right) \left( 1 + \frac{1}{2} \eta_\sigma(N) \frac{\sigma^2}{M^2} \right),
\]

where \( \eta_\sigma(N) \) is chosen to be \(-0.25 \) while observable scales exit the horizon, later running rapidly to settle near \(-0.08 \). (For a concrete realization see Ref. \([11]\).) We take the transition to occur roughly 16 e-folds after the largest observable scales exited the horizon. The field \( \phi \) will dominate the Gaussian part of \( \zeta \) and its mass should be chosen to match the observed spectral index.
Although simple and illustrative, this model is not trivial to embed in a fully realistic early universe scenario. The required initial value of $\sigma$ is only a little outside the quantum diffusion regime which may lead to unwanted observable consequences. Also, not all isocurvature modes decay by the end of the inflationary epoch so (16) should be completed by a specification of the reheating model, and it is possible this could change the prediction for the $n$-point functions. But, assuming these problems are not insurmountable, we can accurately compute the bispectrum generated by (16). Our predictions then apply to any successful realization of this scenario.

**Estimator for $f_{\text{NL}}^{\text{local}}$**—The most urgent question is whether the bispectrum amplitude is compatible with present constraints for $f_{\text{NL}}^{\text{local}}$, $f_{\text{NL}}^{\text{equiv}}$, etc., which as explained above are weighted averages over the bispectrum amplitude on groups of related configurations. At present the strongest constraints apply to $f_{\text{NL}}^{\text{local}}$ which averages over modestly squeezed configurations.

To determine the response of these estimators we construct a Fisher estimate. We numerically compute $\sim 5 \times 10^6$ bispectrum configurations for (16) covering the range from $\ell \sim 1$ to $\ell \sim 7000$ and use these to predict the observed angular temperature bispectrum.

For a choice of parameter values which generate the correct amplitude and scaling of $A(k)$, we find that a Planck-like experiment would measure order-unity values,

$$f_{\text{NL}}^{\text{local}} = 0.25, \quad f_{\text{NL}}^{\text{equiv}} = 0.6, \quad f_{\text{NL}}^{\text{ortho}} = -1.0. \quad (17)$$

These estimates are our second principal result. They are one to two orders of magnitude smaller than previous estimates based on Eq. (2), and are easily compatible with present-day constraints. The difference comes from the strong running of the bispectrum amplitude required for compatibility with $A(k)$, and also the growing number of bispectrum configurations available at large $\ell$. This means that the signal-to-noise tends to be dominated by the largest-$\ell$ configurations where the amplitude is small, depressing the final weighted average; in fact, we find that the reduced bispectrum amplitude near the Planck pivot scale $\ell \sim 700$ is a fair predictor for the averages (17).

We find that it is possible to simultaneously satisfy observational constraints on the amplitude of low-$\ell$ multipoles of the power spectrum [11]. For example, choosing $\alpha = 0.01$ (which makes the wavelength of the modulating mode roughly 100 times the distance to the last-scattering surface) requires an exceptionality $E \approx 300$ to match the measured amplitude of $A(k)$. A value for $E$ in this range would likely require further new physics, but it could perhaps be reduced to a value of order 10 by increasing the bispectrum amplitude. For these parameter choices the low-$\ell$ suppression $C(k)$ may be larger than the approximate bound $C(k) \lesssim 0.14$ suggested by Contaldi et al. [18]. (There is some uncertainty regarding the precise numerical bound, because the result of Contaldi et al. assumed the BICEP measurement of $r$ which is now known to have been confused by dust.) If necessary this would apparently have to be mitigated by tuning our position on the long-wavelength mode.

Finally, we note that although the precise bispectrum used in our analysis applies to the specific model (16), any model which generates scale dependence through a large $\eta_0$ is expected to produce a similar shape. Therefore, despite the contrivances of our example, we expect our conclusions to be robust and apply much more generally.

**CONCLUSIONS**

The CMB power asymmetry is a puzzling feature which may impact on our understanding of the very early universe. The most popular inflation-based explanations deploy the Erickcek–Carroll–Kamionkowski mechanism, in which a single super-horizon mode of exceptional amplitude modulates the small-scale power spectrum (see Ref. [19] for a generalization to include all superhorizon modes). But until now, comparisons of the scenario with observation have not accounted for the scale-dependence of the asymmetry—or the bispectrum which is responsible for it. This is a necessary feature of the model. Previous analyses based on Eq. (2) have suggested the required bispectrum amplitude may be incompatible with observation, but it has not been clear how the inclusion of scale-dependence would modify this conclusion.

In this Letter we have presented a direct determination of the response function which couples the asymmetry to the ensemble-averaged bispectrum and the super-horizon mode. We have presented an illustrative example which satisfies all current observational constraints, and which can be used to obtain precise predictions for the primordial bispectrum. Using this to predict the angular bispectrum of the CMB temperature anisotropy we have confirmed that the bispectrum amplitude is well within the bounds set by current Planck data.

Although this bispectrum strictly applies for the step model (16) we believe it to be a good proxy for any inflationary explanation of the asymmetry which uses a large $\eta$ parameter to generate the scale dependence. Our results show that such scenarios involve much less tension with observation than would be expected on the basis of (2). Nevertheless, this does not mean that an inflationary explanation is automatically attractive. To build a successful model we have been forced to make a number of arbitrary choices, including the initial and final values of the $\sigma$ mass, and the location and rapidity of the transition. It is also unclear whether this inflationary model can be embedded within a viable early universe scenario, which should include at least initial conditions for the inflationary era and a description of how reheating connects it to a subsequent radiation epoch. In our present state
of knowledge it seems challenging to construct a scenario including all these features, and capable of explaining the hemispherical asymmetry, which does not involve choices at least as unlikely as the asymmetry itself.

Acknowledgements

DS acknowledges support from the Science and Technology Facilities Council [grant number ST/L000652/1]. CTB is a Royal Society University Research Fellow. The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP/2007–2013) / ERC Grant Agreement No. [308082]. This work was supported in part by National Science Foundation Grant No. PHY-1066293.

Data availability statement.—Please contact the authors to obtain the bispectrum for the step model (16), which was used to estimate the responses (17).

* C.Byrnes@sussex.ac.uk
† D.Regan@sussex.ac.uk
‡ D.Seery@sussex.ac.uk