Peter Watts and Peter Coles
School of Physics & Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom

Adrian Melott
Department of Physics and Astronomy, University of Kansas, Lawrence, KS 66045
(Received November 20, 2002)

The large-scale structure of the Universe is thought to evolve by a process of gravitational amplification from low-amplitude Gaussian noise generated in the early Universe. The later, non-linear stages of gravitation-induced clustering produce phase correlations with well-defined statistical properties. In particular, the distribution of phase differences $D$ between neighboring Fourier modes provides useful insights into the clustering phenomenon. Here we develop an approximate theory for the probability distribution $D$ and test it using a large battery of numerical simulations. We find a remarkable universal form for the distribution which is well described by theoretical arguments.

The large-scale structure of the Universe is a complex interconnecting pattern whose structural elements comprise filaments, sheets and clusters of galaxies surrounding large voids. According to standard theories this “cosmic web” develops by a process of gravitational instability from small initial fluctuations in the density of a largely homogeneous early Universe.

The physical description of an inhomogeneous Universe revolves around the dimensionless density contrast, $\delta(x)$, which is obtained from the spatially-varying matter density $\rho(x)$ via $\delta(x) = (\rho(x) - \rho_0)/\rho_0$ where $\rho_0$ is the global mean density. It is useful to expand the density contrast in Fourier series, in which $\delta$ is treated as a superposition of plane waves:

$$\delta(x) = \sum_k \delta(k) \exp(i k \cdot x). \quad (1)$$

The Fourier transform $\hat{\delta}(k)$ is complex and therefore possesses both amplitude $C(k) = |\hat{\delta}(k)|$ and phase $\Phi_k$ where

$$\hat{\delta}(k) = C(k) \exp(i \Phi_k) = A(k) + iB(k); \quad (2)$$

the real and imaginary parts are $A$ and $B$ so that $A(k) = C(k) \cos(\Phi_k)$ and $B(k) = C(k) \sin(\Phi_k)$. We also have $C^2(k) = A^2(k) + B^2(k)$ and $\tan \Phi_k = B(k)/A(k)$.

In theories of structure formation involving cosmic inflation [1], the initial fluctuations that seeded the structure formation process form a Gaussian random field [2] possessing the properties of statistical homogeneity and isotropy. In such fields the real and imaginary parts of $\hat{\delta}(k)$ are independent Gaussians so that the modulus $C(k)$ has a Rayleigh distribution and the phases $\Phi_k$ are uniformly random on the interval $[0, 2\pi]$. Obviously the distribution of $\delta$ in position space will also be Gaussian; indeed all finite-dimensional joint probabilities of $\delta$ in different locations are multivariate Gaussian in this case [2].

Even if the primordial density fluctuations were indeed Gaussian, the later stages of gravitational clustering must induce some form of non-linearity. One particular way of looking at this issue is to study the behavior of Fourier modes of the cosmological density field. If the hypothesis of primordial Gaussianity is correct then these modes began with random spatial phases. In the early stages of evolution, the plane-wave components of the density evolve independently like linear waves on the surface of deep water. As the structures grow in mass, they interact with other in non-linear ways, more like waves breaking in shallow water. These mode-mode interactions lead to the generation of coupled phases. While the Fourier phases of a Gaussian field contain no information (they are random), non-linearity generates non-random phases that contain much information about the spatial pattern of the fluctuations but which is ignored entirely in the usual clustering descriptors, such as the power spectrum (see below). There have been a number of attempts to gain quantitative insight into the behavior of phases in gravitational systems. Some studies [3] have concentrated on the evolution of phase shifts for individual modes using perturbation theory and numerical simulations. An alternative approach was adopted by Scherrer, Melott & Shandarin [4], who developed a practical method for measuring the phase coupling in random fields that could be applied to real data. More recent studies have established connections between phase evolution, clustering dynamics and morphology [5] and new methods have been developed for visualizing phase information [6]. Connections have also been demonstrated [7] between phase information and alternative measures of non-linear clustering such as the bispectrum [8].

One of the major barriers to the more widespread use of phase-based methods for probing cosmic nonlinearity is the lack of any well-established statistical framework for quantifying the information contained in the Fourier phases. Since phases are angular variables traditional statistical measures of location and dispersion are inappropriate. Moreover, probability distributions for circular variables have very different properties to those for
variables defined on the real line; see the monographs by Mardia [9] and Fisher [10] for more detailed discussion. The upshot of this is that it is not yet known whether there are useful standard distributions and limit theorems like those that make the Gaussian distribution so useful and so ubiquitous for statistical analysis. It is this deficiency we address in this paper.

Non-linear clustering is difficult to handle rigorously with analytic methods. Our approach is therefore to develop and approximate theory and test it using numerical experiments. The simulations we shall be comparing to theory are constructed within a finite cubic volume with periodic boundary conditions. The Fourier representation of the clustering pattern they reveal is therefore discrete. In particular phases are defined for wavevectors on a cubic lattice which, for simplicity, we take to have an integer spacing. In previous analyses of phase coupling a cubic lattice which, for simplicity, we take to have an

The upshot of this is that it is not yet known whether there are useful standard distributions and limit theorems like those that make the Gaussian distribution so useful and so ubiquitous for statistical analysis. It is this deficiency we address in this paper.

Non-linear clustering is difficult to handle rigorously with analytic methods. Our approach is therefore to develop and approximate theory and test it using numerical experiments. The simulations we shall be comparing to theory are constructed within a finite cubic volume with periodic boundary conditions. The Fourier representation of the clustering pattern they reveal is therefore discrete. In particular phases are defined for wavevectors on a cubic lattice which, for simplicity, we take to have an integer spacing. In previous analyses of phase coupling a cubic lattice which, for simplicity, we take to have an

...
FIG. 1. Histories $P(D)$ of phase differences $D$ are shown in grey along with the analytical model (8) for four different initial spectra as functions of epoch.

A better theory might be obtained by marginalizing over these variables, but the integrals involved are messy. On the other hand the distribution of amplitudes is relatively narrow, peaking around $c_1 = \sigma_1$ and $c_2 = \sigma_2$ and $\chi$ is presumably small for quasi-nonlinear stages. Under these circumstances we expect the distribution of phase differences formed over a large number of pairs to follow the form given above, i.e.

$$P(D) = \frac{1}{2\pi I_0(\kappa)} \exp\left[-\kappa \cos(D - \mu)\right].$$  \hspace{1cm} (8)

This distribution is well-known in the field of circular statistics where it is known as the von Mises distribution [9,10]. The mean $\mu$ is controlled by the positions of individual features of the distribution and will consequently vary from sample to sample. The parameter $\kappa$, related to $\chi$, describes the level of nonlinearity. When $\kappa \to 0$
is small the distribution is approximately uniform, while for small $\kappa$ it takes the form $P(D) \approx \frac{1}{\sqrt{2\pi}}[1 + \kappa \cos(D - \mu)]$ showing that initial departures from random phases manifest themselves as a sinusoidal perturbation of $P(D)$. In the limit $\kappa \to \infty$ the distribution tends towards a single spike at $\theta = \mu$; this corresponds to a single concentration in position space.

To test our hypothesis with fairly complete coverage of possible parameters, we used a large ensemble of gravitational clustering simulations [13]. These comprise sets of 128$^3$ particles and are not large by current standards, but the particular benefit they offer our analysis is a fairly complete coverage of parameter space. There are four realizations of each type of initial conditions, using different pseudorandom number generators to generate the initial phases. Initial power spectra were pure power laws, i.e. $P(k) = Ak^n$ with indices $n = -2, -1, 0, 1$. Data is taken every time the scale of clustering doubles, from $k_{\text{nl}} = 64k_f$ to $k_{\text{nl}} = 4k_f$, where $k_f$ is the fundamental mode of the box and $k_{\text{nl}}$ is defined by

$$\int_0^{k_{\text{nl}}} P(k)dk = 1. \quad (9)$$

The first stage may be significantly compromised by resolution effects; we have simulations with $k_{\text{nl}} = 2k_f$ but these almost certainly suffer from problems connected with the boundary conditions (which are, as usual, periodic).

Analysis of the properties of the phase differences was conducted on each realization of each spectrum and evolutionary stage, i.e. a total of 120 times altogether. We Fourier-transform each stage and extract phases for allolutionary phase, i.e. a total of 120 times altogether. We Fourier-transform each stage and extract phases for resolution effects; we have simulations with $k_{\text{nl}} = 2k_f$ but these almost certainly suffer from problems connected with the boundary conditions (which are, as usual, periodic).

Analysis of the properties of the phase differences was conducted on each realization of each spectrum and evolutionary phase, i.e. a total of 120 times altogether. We Fourier-transform each stage and extract phases for each wavevector $k$. From these we obtain differences $D_k$ in three orthogonal $k$-space directions from which we form histograms. The mean value $\mu$ of each distribution contains information about the specific spatial location of dominant features [5], which will differ from realization to realization and which also varies with direction. We therefore rotate individual distributions so that they have the same mean value and “stack” the resulting histograms. We also combine all three directional differences into an overall histogram $P(D)$ where $D = (D_1^2 + D_2^2 + D_3^2)^{1/2}/\sqrt{3}$. This approach, together with the large size of the ensemble, produces final histograms with relatively small error bars, as can be seen in Fig. 1. The results show extremely good agreement over the range of initial power-spectra and evolutionary stages, although it does break down at late stages for the case $n = -2$. This is not surprising, given the fact that such spectra have large amounts of power on large scales and phase correlations therefore develop extremely quickly. Even the earliest stage shown of this simulation shows a significantly non-uniform distribution of $D$. The development of phase correlations with evolutionary stage for each initial spectrum is represented by the increasing deviation from uniformity down each column starting, as expected, with sinusoidal departures.

The excellent match of the distribution (8) to the results of detailed numerical simulations may appear surprising given the very approximate nature of its derivation. But its validity is reinforced by the fact that it is the maximum entropy distribution on a circle for a fixed mean $\mu$ and fixed circular dispersion [9]. It should really therefore be regarded as the circular equivalent of a Gaussian distribution, which has maximal entropy for fixed mean and variance on the real line.

The universal behavior we have demonstrated will allow us in future to discriminate between gravitationally-induced mode-coupling and other forms, such as that induced by peculiar motions [14].

This work was supported by PPARC grant PPA/G/S/1999/00660; ALM gratefully acknowledges the support of NSF through grant AST-0070702.