Journal of Statistical Mechanics:
Theory and Experiment

PAPER: INTERDISCIPLINARY STATISTICAL MECHANICS

Uncovering multiscale order in the prime numbers via scattering

To cite this article: S Torquato *et al J. Stat. Mech.* (2018) 093401

Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

PAPER: Interdisciplinary statistical mechanics

Uncovering multiscale order in the prime numbers via scattering

S Torquato[1,4], G Zhang[2] and M de Courcy-Ireland[3]

[1] Department of Chemistry, Department of Physics, Princeton Institute for the Science and Technology of Materials, and Program in Applied and Computational Mathematics, Princeton University, Princeton NJ 08544, United States of America

[2] Department of Chemistry, Princeton University, Princeton NJ 08544, United States of America

[3] Department of Mathematics, Princeton University, Princeton NJ 08544, United States of America

E-mail: torquato@princeton.edu

Received 9 May 2018
Accepted for publication 24 July 2018
Published 5 September 2018

Online at stacks.iop.org/JSTAT/2018/093401
https://doi.org/10.1088/1742-5468/aad6be

J. Stat. Mech. (2018) 093401

Abstract. The prime numbers have been a source of fascination for millennia and continue to surprise us. Motivated by the hyperuniformity concept, which has attracted recent attention in physics and materials science, we show that the prime numbers in certain large intervals possess unanticipated order across length scales and represent the first example of a new class of many-particle systems with pure point diffraction patterns, which we call *effectively limit-periodic*. In particular, the primes in this regime are hyperuniform. This is shown analytically using the structure factor $S(k)$, proportional to the scattering intensity from a many-particle system. Remarkably, the structure factor of the primes is characterized by dense Bragg peaks, like a quasicrystal, but positioned at certain rational wavenumbers, like a limit-periodic point pattern. However, the primes show an erratic pattern of occupied and unoccupied sites, very different from the predictable patterns of standard limit-periodic systems. We also identify a transition between ordered and disordered prime regimes that depends on the intervals studied. Our analysis leads to an algorithm that enables one to predict primes with high accuracy. Effective limit-periodicity deserves future investigation in physics, independent of its link to the primes.

Keywords: random/ordered microstructures

[4] Author to whom any correspondence should be addressed.

1742-5468/18/093401+15$33.00

Contents

Introduction

X-ray and neutron scattering techniques provide powerful ways to probe the structure of matter [1]. The observed scattering intensity in Fourier (reciprocal) space is encoded in the structure factor $S(\mathbf{k})$, where $\mathbf{k}$ is the wave vector. A fundamental problem in condensed matter theory and statistical physics is the determination of the class of ordered many-body systems in d-dimensional Euclidean space $\mathbb{R}^d$ with pure point diffraction or scattering patterns, i.e. those described by a structure factor involving only a set of Dirac delta functions (Bragg peaks):

$$S(\mathbf{k}) = \sum_{j=1}^{\infty} a_j \delta(\mathbf{k} - \mathbf{k}_j), \tag{1}$$

where the a_j ($j = 1, 2, \ldots$) are positive weights. It is well-known that perfect crystals (periodic point patterns), which have scattering patterns that are quite different from disordered systems (e.g. gases and liquids) with continuous spectra, fall in this class; see figure 1 for illustrative examples. It came as a great surprise in the early 1980's that a family of noncrystalline (aperiodic) states of matter, called 'quasicrystals' [2], also have pure point diffraction, but with a twist, namely, the corresponding Bragg peaks densely fill reciprocal space exhibiting symmetries that would be prohibited for crystals. A less familiar phylum of point patterns obeying (1) are limit-periodic systems [3, 4]. These are deterministic point patterns that consist of a union of an infinite set of distinct periodic structures with different (rational) periods, and hence are also characterized by dense Bragg peaks. What differentiates limit-periodic systems from quasicrystals is that the ratio between any two peak locations is rational.

Where in this zoology of particle systems does one place the prime numbers? These are the numbers such as 2, 3, 5, 7, 11, $13, \ldots, 163, \ldots, 691, \ldots, 2^{77,232,917} - 1, \ldots$, having no proper factors, viewed as a one-dimensional (1D) point pattern. In the present paper, we show that the primes in judiciously chosen intervals have dense Bragg peaks, similar to a limit-periodic system, and hence satisfy (1). This is in astonishing contrast to the general understanding of primes as pseudo-random numbers [6]. The apparent

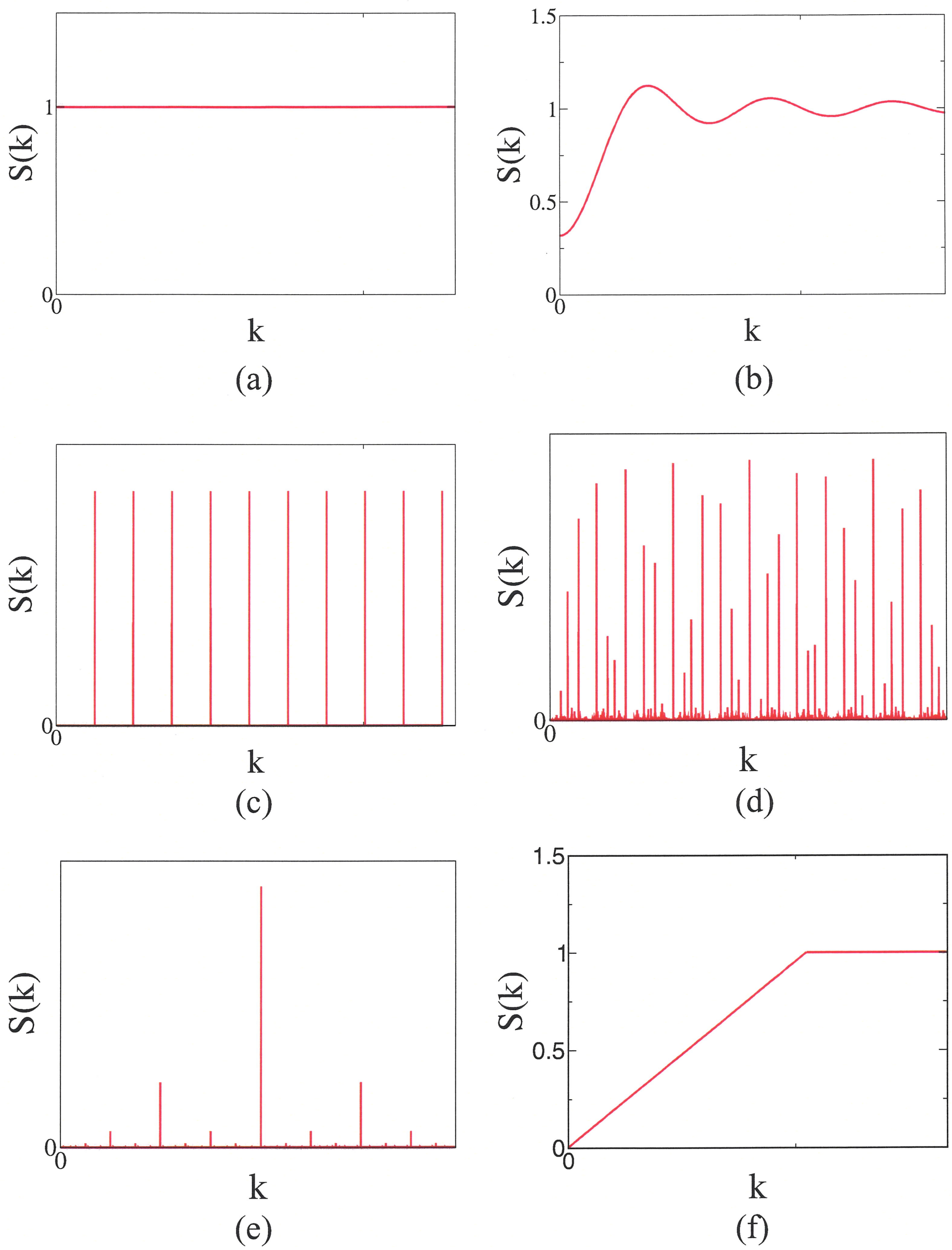

Figure 1. Illustrative examples of structure factors for ordered and disordered point patterns in 1D Euclidean space: (a) spatially uncorrelated (Poisson process or ideal gas); (b) Liquid; (c) Crystal (integer lattice); (d) Quasicrystal (Fibonacci chain); (e) Limit-periodic (period-doubling chain); (f) Nontrivial zeros of the Riemann zeta function. Whereas the cases (a) and (b) are examples of nonhyperuniform systems, the remaining cases represent hyperuniform examples.

difficulty of factoring large numbers into primes is basic to contemporary cryptography, and the lack of any obvious pattern is nicely summarized in a famous quotation attributed to Vaughan: ‘ *It is evident that the primes are randomly distributed but, unfortunately, we do not know what ‘random’ means.*’ [5]. In short intervals, say from a large number X to $X + \ln(X)$, Gallagher [6] proved that the primes have a pseudo-random spatial distribution with gaps following a Poisson process. In the present paper, we study longer intervals, such as X to $2X$, and find that the primes have multiscale order characterized by dense Bragg peaks. Thus they obey (1), but are distinguished from both quasicrystals and limit-periodic systems in ways that we detail below.

Much of the modern understanding of prime numbers is based on a fundamental insight of Riemann [7]. He introduced what we now call the Riemann zeta function $\zeta(s)$, for a complex variable s, and indicated an explicit formula relating the primes to its zeros. Our original motivation to study the scattering patterns of the primes is the fact that the nontrivial zeros[5] of the Riemann zeta function have an exotic hidden order on large length scales called *hyperuniformity*. A hyperuniform point configuration in $\mathbb{R}^d$ is one in which $S(k)$ tends to zero as the wavenumber k tends to zero or, equivalently, one in which the local number variance $\sigma^2(R)$ associated within a spherical window of radius R grows more slowly than R^{d} in the large-R limit [8]. All perfect crystals and quasicrystals are hyperuniform, but typical disordered many-particle systems, including gases, liquids, and glasses, are not. Disordered hyperuniform many-particle systems are exotic states of amorphous matter that have attracted considerable recent attention in physics and materials science because of their novel structural and physical properties [9–19]. According to the celebrated *Riemann hypothesis*, the nontrivial zeros of the zeta function lie along the *critical line* $s = 1/2 + it$ with $t \in \mathbb{R}$ in the complex plane and thus form a 1D point process. A resolution of this hypothesis is widely considered to be one of the most important open problems in pure mathematics [20]. Montgomery [21] advanced the conjecture that the pair correlation function $g_2(r)$ of the (normalized) zeros takes on the simple form $1 - \sin^2(\pi r)/(\pi r)^2$. Remarkably, this exactly matches the pair correlation function of the eigenvalues of certain random Hermitian matrices [22–24]. The corresponding structure factor $S(k)$ tends to zero linearly in k in the limit $k \to 0$, as shown in figure 1(f). This means that the Riemann zeros are disordered but hyperuniform [10]. In his famous essay entitled 'Birds and Frogs' [25], Dyson suggested an approach to the Riemann hypothesis where one would first classify all 1D quasicrystals and then show that one such quasicrystal corresponds to the non-trivial Riemann zeros.

An important aspect of the distribution of prime numbers is that larger ones become increasingly sparse. The drop-off is gradual enough that, in Gallagher's regime of short intervals or even for the longer intervals considered here, the density of prime numbers can be treated as constant [26]. According to the *prime number theorem* [27], the prime counting function $\pi(x)$, which gives the number of primes less than x, in the large-x asymptotic limit is given by

$$\pi(x) \sim \frac{x}{\ln(x)} \qquad (x \to \infty). \tag{2}$$

One can interpret this as indicating that the probability that a randomly selected integer less than a sufficiently large x is prime is inversely proportional to the number of digits of x. This implies a position-dependent number density $\rho(x) \sim 1/\ln(x)$. Thus the primes constitute a *statistically inhomogeneous* set of points in large intervals, becoming sparser as x increases. This means one must be careful in choosing the interval over

[5] The trivial zeros are −2, −4, −6,..., which have less bearing on prime numbers.

which the primes are sampled. This observation is crucial to the remarkable properties of the primes that we report here.

The hyperuniformity of the Riemann zeros led us to seek intervals in the primes in which they might be regarded as a hyperuniform point pattern. In a concurrent numerical study [28], we examined the structure factor $S(k)$ for primes in an interval $[M, M+L]$ with M large (say, 10^{10}) and L/M a small positive number. These simulations strongly suggest that the structure factor in such finite intervals exhibits many well-defined Bragg-like peaks dramatically overwhelming a small 'diffuse' contribution, indicating that the primes are more ordered than previously known.

Motivated by this numerical study, here we apply the tools of statistical physics to understand the nature of the primes as a point process by quantifying the structure factor, pair correlation function, local number variance, and the τ order metric in various intervals. Our main results are obtained for the interval $M \leqslant p \leqslant M+L$ with M very large and the ratio L/M held constant. This enables us to treat the primes as a homogeneous point pattern. We also consider appreciably larger and smaller intervals for purposes of comparison. We prove that the primes are characterized by unanticipated multiscale order; see [26] for details. Specifically, an analytical formula that we derive for their limiting structure factor $S(k)$ has dense Bragg peaks, as in the case of quasicrystals [2]. Unlike quasicrystals, however, the prime peaks occur at certain rational multiples of π, which is similar to limit-periodic systems [3]. But the primes show an erratic pattern of occupied and unoccupied sites, very different from the predictable and orderly patterns of standard limit-periodic systems. Hence, the primes are the first example of a point pattern that is *effectively* limit-periodic.

Our analysis is rooted in the *circle method* of Hardy–Littlewood [29], in particular their conjecture on prime k-tuples, but we emphasize the perspective of statistical physics and the new consequences that arise in the limit of infinite system size. Our analytical formula (19) expresses the pair correlation function g_2, including the density of twin primes, as an infinite sum, whereas the celebrated Hardy–Littlewood representation was originally presented as a product over primes. Using a scalar order metric τ numerically calculated from $S(k)$, we identify a transition between the order exhibited when L is comparable to M and the uncorrelated behavior when L is only logarithmic in M. Our formulation also yields an algorithm that enables one to predict (reconstruct) primes with high accuracy.

Background and definitions

For a statistically homogeneous point process in d-dimensional Euclidean space $\mathbb{R}^d$ at number density ρ, let $g_2(\mathbf{r})$ denote the standard pair correlation function, which is normalized such that it decays to unity for large $|\mathbf{r}|$ in the absence of long-ranger order. The corresponding structure factor $S(\mathbf{k})$ is defined as follows:

$$S(\mathbf{k}) = 1 + \rho\tilde{h}(\mathbf{k}), \tag{3}$$

where

$$\tilde{h}(\mathbf{k}) = \int_{\mathbb{R}^d} h(\mathbf{r}) \exp\left[-\mathrm{i}(\mathbf{k}\cdot\mathbf{r})\right] \mathrm{d}\mathbf{r} \tag{4}$$

is the Fourier transform of the total correlation function $h(\mathbf{r}) \equiv g_2(\mathbf{r}) - 1$ so that

$$h(\mathbf{r}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \tilde{h}(\mathbf{k}) \exp\left[\mathrm{i}(\mathbf{k} \cdot \mathbf{r})\right] \mathrm{d}\mathbf{k}. \tag{5}$$

While $S(\mathbf{k})$ is a nontrivial function for spatially correlated point processes, it is exactly equal to unity for all $\mathbf{k}$ for a homogeneous Poisson point process.

As noted earlier, we consider the primes in the interval $[M, M + L]$ as a 1D lattice-gas in which the primes occupy a subset of the odd integers. Henceforth, we will focus on 1D systems. A particular lattice-gas configuration under periodic boundary conditions is characterized by the *local* density:

$$\eta(\mathbf{r}) = \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{x}_i), \tag{6}$$

where N is the number of primes in the interval. The Fourier transform of the local density, called the *complex collective density variable* $\tilde{\eta}(\mathbf{k})$, is given by

$$\tilde{\eta}(\mathbf{k}) = \sum_{j=1}^{N} \exp(-\mathrm{i}\mathbf{k} \cdot \mathbf{r}_j). \tag{7}$$

The corresponding structure factor is given by

$$S(\mathbf{k}) \equiv \frac{|\tilde{\eta}(\mathbf{k})|^2}{N} (\mathbf{k} \neq 0) \tag{8}$$

where the wavenumber $k = |\mathbf{k}|$ ranges from zero to $2\pi/a$, extended periodically; and a is the lattice spacing. Ultimately, we pass to the limit $N \to \infty$ and assume ergodicity. A statistically homogeneous point process is *ergodic* if any single realization of the ensemble is representative of the ensemble in the infinite-system-size limit.

A useful scalar quantity that is capable of capturing the degree of translational order of a point process in d-dimensional Euclidean spaces across length scales is the τ order metric [15]:

$$\tau \propto \int_{\mathbb{R}^d} \left[S(\mathbf{k}) - 1\right]^2 \mathrm{d}\mathbf{k}, \tag{9}$$

where $S(\mathbf{k})$ is the ensemble-averaged structure factor in the thermodynamic limit. In this paper, we use the discrete-setting counterpart of this order metric for a 1D lattice gas in a fundamental cell of length L under periodic boundary conditions [33]:

$$\tau = \frac{1}{N_s} \sum_{j=1}^{Ns-1} \left(S\left(\frac{j\pi}{N_s}\right) - (1 - f) \right)^2, \tag{10}$$

where N_s is the number of lattice sites within the fundamental cell and f is the occupation fraction.

We define an uncorrelated (Poisson) lattice gas as the discrete counterpart of an ideal gas in $\mathbb{R}$: each site has a probability f of being occupied, independent of other sites. For such systems, the ensemble-averaged $S(\mathbf{k})$ is a constant $1 - f$, so that $\tau = 0$ in the infinite-system-size limit. Thus, we see from (10) that a deviation of τ from zero measures translational order with respect to the fully uncorrelated case.

The local number variance $\sigma^2(R)$ associated within an interval (window) of length $2R$ for a 1D homogeneous point process in $\mathbb{R}$ [8] depends on an integral involving the structure factor $S(k)$:

$$\sigma^2(R) = \frac{2\rho R}{\pi} \int_0^\infty S(k)\tilde{\alpha}_2(k;R)\mathrm{d}k, \tag{11}$$

where $\tilde{\alpha}_2(k;R) = 2\sin^2(kR)/(kR)$. Integrating by parts leads to an alternative representation of the number variance [31]:

$$\sigma^2(R) = -\frac{\rho R}{\pi} \int_0^\infty Z(k)\frac{\partial\tilde{\alpha}_2(k;R)}{\partial k}\mathrm{d}k, \tag{12}$$

where $Z(K)$ is defined by (16). The quantity $Z(k)$ has advantages over $S(k)$ in the characterization of quasicrystals and other point processes with dense Bragg peaks [31]. This is the formula that we employ to determine the local variance of the number of primes. The local number variance for uncorrelated lattice gases grows linearly with R. A hyperuniform point process in $\mathbb{R}$ has a local number variance $\sigma^2(R)$ that grows more slowly than R in the large-R limit [8].

Results

We consider the primes in the interval $[M, M+L]$ to be a special 'lattice-gas' model: the primes and odd composite integers are 'occupied' and 'unoccupied' sites, respectively, on an integer lattice of spacing 2 that contains all of the positive odd integers. We study the pair statistics between primes in such intervals. If L is much larger than M, the density $1/\ln(n)$ drops off appreciably as n ranges from M to $M+L$, and then the system becomes inhomogeneous (nonuniform), which is diametrically the opposite of hyperuniform. On the other hand, if the interval is small such that $L \sim \ln(M)$, one enters Gallagher's regime in which the primes are Poisson distributed.

One of our main analytical results is formula (13) for the structure factor of the primes, valid in the regime $M \to \infty$ with L/M converging to a fixed positive value denoted β. Here $\ln(M+L) = \ln(M) + \ln(1+L/M)$ is asymptotic to $\ln(M)$. Our $M \to \infty$ limiting form for $S(k)$, given by equation (14), is valid for any positive value of $\beta > 0$. These results lead to several significant consequences, which we describe below.

We study various prime intervals, but we show that when $L \sim \beta M$, the major contribution to the structure factor $S(k)$ is a set of dense Bragg peaks that are located at certain rational values of k/π. Let N be the number of primes in the interval from M to $M+L$. For finite but large N, the structure factor at special rational values of k/π is given by Bragg peaks with heights given by [26]

$$S(\pi m/n) \sim \frac{N}{\phi(2n)^2}\mu(2n)^2. \tag{13}$$

Here $|m| \leqslant n$ and n are co-prime integers (share no common divisors, except 1), $\phi(n)$ is Euler's totient function [30], which counts the positive integers up to a given integer n that are co-prime to n, and $\mu(n)$ is the Möbius function [30] so that $\mu^2(2n)$ is one

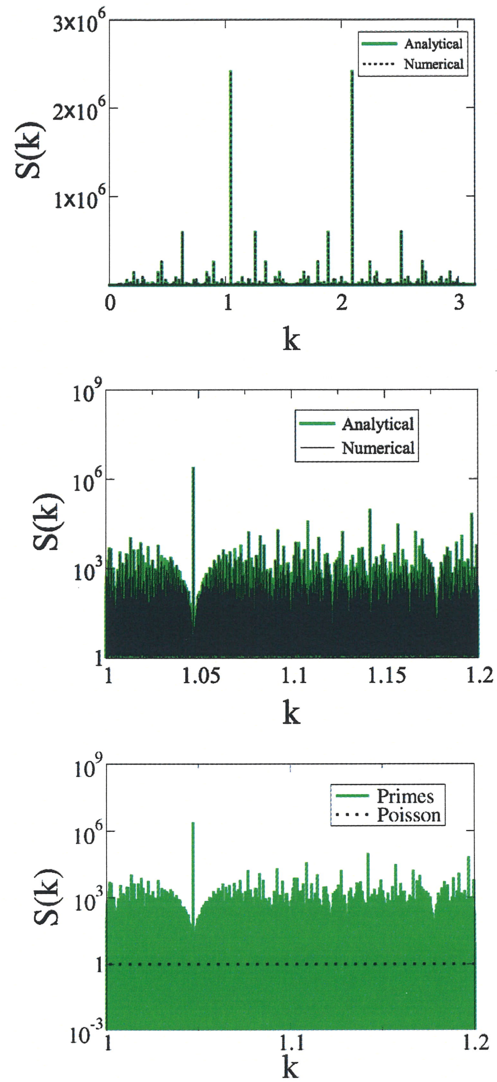

Figure 2. Top: $S(k)$ for the primes as a function of k (in units of the integer lattice spacing), as predicted from formula (13) for $M = 10^{10} + 1$, $L = 2.23 \times 10^8$ and $n_{\max} = 100\ln(M)$ is in excellent agreement with the corresponding numerically computed structure factor obtained in [28]. Note the many Bragg peaks of various heights with a self-similar pattern. Middle: same as the top panel, but at a smaller horizontal scale and log vertical scale to reveal the dense peak structure. Bottom: the prediction from (13) but with a different vertical scale than in the middle panel to again reveal the dense peak structure and its stark contrast with the uncorrelated (Poisson) lattice gas.

whenever $2n$ is square-free and zero otherwise. Notice that as n grows, the size of the peak shrinks because of the denominator $\phi(n)^2$. Figure 2 depicts the structure factor of the primes obtained from (13) at different horizontal scales, where $M = 10^{10} + 1$, $L = 2.23 \times 10^8$ and n is truncated at $n_{\max} = 100\ln(M)$. This is in excellent agreement with the corresponding numerically computed structure factor of the actual primes configuration in this interval (top and middle panels of figure 2). The heights of some peaks in the numerical result are slightly lower than their analytical predictions.

J. Stat. Mech. (2018) 093401

This is caused by the fact that numerical results are limited to a discrete choice of k ($k = 2\pi z/L$, where $z \in \mathbb{Z}$), and hence involve slight mismatches with the appropriate exact peak locations ($k = \pi m/n$). The numerically determined $S(k)$, which is calculated at slightly off-peak locations, can have peak heights that are slightly lower than the analytically predicted ones in the limit that $L \to \infty$. The structure factor contains many well-defined Bragg-like peaks of various intensities characterized by a type of self-similarity. This self-similarity becomes exact in the limit $M \to \infty$, due to the fact that $\phi(n_1 n_2) = \phi(n_1)\phi(n_2)$ for relatively prime n_1 and n_2, so that rescaling preserves the relative heights of the peaks given by equation (13). The bottom part of the figure uses a different vertical scale to compare $S(k)$ to that of an uncorrelated lattice gas. This definitively demonstrates that the primes in this interval exhibit order across multiple length scales, making them substantially more ordered than the uncorrelated lattice gas found by Gallagher for shorter intervals of primes.

J. Stat. Mech. (2018) 093401

'Effective' limit-periodicity

The proof of (13), assuming the first Hardy–Littlewood conjecture [29], is given in [26]. It is based on grouping the terms in the sum defining $S(\pi m/n)$ according to their remainder after division by $2n$. A fundamental heuristic about prime numbers is that each of the allowed remainders occurs roughly equally often, so that the primes are evenly distributed in arithmetic progressions. This breaks down if the modulus n is too large relative to the primes under consideration. Increasingly precise versions of this statement have been established rigorously, starting from Dirichlet's theorem that there are infinitely many primes for each remainder to a fixed modulus. A strong interpretation conjectured by Elliott–Halberstam plays a role in some of the work of the Polymath project pushing progress on gaps between primes to the limit. The key calculation underlying our proof is that the roughly even distribution across all possible remainders causes constructive or destructive interference in the sum $S(\pi m/n)$, depending on the fraction m/n, and leads to (13). Referring to (13), note that if n is even or if n has a repeated factor, then $\mu(2n) = 0$ so $S(\pi m/n)$ vanishes up to the accuracy in comparing the number of primes in different progressions. If n is odd and square-free, then the structure factor has a peak of size $N/\phi(n)^2$ (since, n being odd, $\phi(2n) = \phi(n)$). This explains the peaks observed numerically [28] at, for example, $S(\pi/3)$. A value on the order of N should indeed be viewed as a peak since $S(k)$ is a sum of length N and ignoring all cancellation shows that $|S(k)| \leqslant N$. Thus the largest values of $S(k)$ are the peaks when k/π is a rational number with odd, square-free denominator. Taken together, these locations correspond to effective periodicities, as illustrated in figure 3.

In the limit of infinite system size, the peaks will become Dirac delta functions at rational numbers with odd, square-free denominators, and the discrete formula (13) (scaled by $2\pi\rho$) tends to

$$\lim_{M\to\infty} \frac{S(k)}{2\pi\rho} = \sum_{n\in\mathbb{Z}^+}^{\flat} \sum_{m\in\mathbb{Z}}^{\times} \frac{1}{\phi(n)^2} \delta\left(k - \frac{m\pi}{n}\right), \tag{14}$$

where the symbol $\flat$ is meant to indicate that the sum over n only involves odd, square-free positive values of n (excluding 1 to eliminate forward scattering) and the symbol $\times$ indicates that m and n have no common factor. The double sum on the right hand

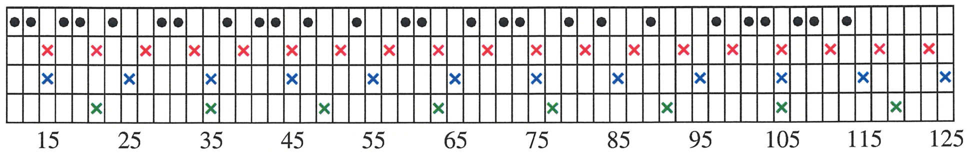

Figure 3. Illustration of the superposition of effective multiple periodicities in the primes. We take the primes to be 'occupied' sites (black dots) on an integer lattice of spacing 2 that contains all of the positive odd integers. The crosses indicate sites that cannot be occupied because of a certain periodicity $2n$ (n sites on the odd integer lattice), where n is a square-free odd number. For example, the peak at $\pi/3$ with $n=3$ and $m=1$ corresponds to remainders when dividing by 6. A prime must leave a remainder of 1 or 5 or else it would be divisible by 2 or 3. The lattice in the figure has a spacing of 2, so these allowed sites appear with a period of 3 instead of 6. The forbidden sites appear as red crosses. The other two sites may or may not be prime, and if one averages over many periods in a sufficiently large interval, each of them will have an equal occupation probability (due to Dirichlet's theorem on arithmetic progressions). The overall effect of this equidistribution of occupied sites is an effective periodicity of 6. Similarly, the primes show an effective periodicity of 10 (blue crosses), 14 (green crosses), and even larger periods (not shown in the figure). The superposition of all of the effective periodicities leads to a pattern of dense Bragg peaks located at $m\pi/n$ (where m are the positive integers and n is odd and square-free), reminiscent of a limit-periodic system even though each local period is subject to erratic disruptions. These peaks are illustrated in figure 2. However, if the interval is too small or too large, then the effective periodicity would not be seen. It is a distinctive feature of primes in intervals from M to $M+L$ with M and L large numbers of comparable magnitude.

side of (14) weakly converges in the sense that it is well-defined when integrated against test functions, including its action under Fourier transformation. Equation (14) implies that the 'diffuse' part observed numerically in [28] vanishes in the infinite-system-size limit. Hence, the primes become effectively limit-periodic, despite the variable pattern of occupied sites, as proved in [26] and illustrated in figure 3. Such multiscale order in the primes appears to be a new discovery.

This is to be distinguished from a standard limit-periodic system, which is a deterministic point process characterized by dense Bragg peaks at rational multiples of π. A prototypical example is the *period-doubling chain* [3]. In this model, there are sites of two types, a and b, forming a point pattern on the integer lattice defined by the following iterative substitution rule, initialized with a single site a: $a\to ab$ and $b\to aa$ [3]. The locations of the b's are given by a superposition of an infinite number of arithmetic progressions of the form $2+4j$, $8+16j$, $32+64j$, …, with a factor of 4 from one to the next. Thus, the infinite-size limit is a union of periodic systems in which $S(k)$ consists of dense Bragg peaks at certain rational values k/π. The structure factor associated with the a's (assuming unit lattice spacing) is given by

$$S(k)=\frac{4\pi}{3}\left[\sum_{m=1}^{\infty}\delta(k-2\pi m)+\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}2^{-2n}\delta\left(k-\frac{(2m-1)\pi}{2^{n-1}}\right)\right]. \tag{15}$$

Figure 1(e) shows the structure factor for the period-doubling chain.

J. Stat. Mech. (2018) 093401

Hyperuniformity

We now show that the effective limit-periodic form (14) of $S(k)$ implies that the primes are hyperuniform. The structure factor $S(k)$ is not a continuous function because there are dense Bragg peaks arbitrarily close to 0, so we do not have $S(k) \to 0$ as $k \to 0$ in the usual sense. We follow the practice of [31] in such instances and pass to a cumulative version of the structure factor, $Z(K)$, defined by

$$Z(K) = 2\int_0^K S(k)\mathrm{d}k, \tag{16}$$

which is the *cumulative* intensity function within a 'sphere' of radius K of the origin in reciprocal space. If $Z(K)$ tends to 0 as a power $K^{\alpha+1}$, any positive power $\alpha > 0$ yields hyperuniformity and distinguishes the primes from a Poisson distribution of points with the same density. Using relations (14) and (16), we find

$$\lim_{M\to\infty} \frac{Z(K)}{2\pi\rho} = 2\sum_{n}^{\flat} \sum_{m\pi/n<K}^{\times} \frac{1}{\phi(n)^2}. \tag{17}$$

Using (17) together with some results from analytic number theory [26], we can show that $Z(K) \sim K^2$ as $K \to 0$.

Recall that a 1D hyperuniform point process is one in which $\sigma^2(R)$ grows more slowly than R in the large-R limit [8]. Using formula (12) that relates $\sigma^2(R)$ to $Z(K)$, we find relation (17) implies the primes have a number variance $\sigma^2(R)$ that scales logarithmically with R in the large-R limit, which makes them hyperuniform of class II [32]. This is precisely the same growth rate exhibited by the Riemann zeros [10], but as we will see, the latter are appreciably less ordered than the former.

Transition between order and disorder

Here we compute the order metric τ via the discrete formula (10). For general lattice gases characterized by Bragg peaks, τ/ρ^2 grows linearly with L for sufficiently large L:

$$\tau/\rho^2 \sim cL, \tag{18}$$

where c is dependent on the system. Based on this order metric, the primes are substantially more ordered than the uncorrelated lattice gas and appreciably less ordered than an integer lattice, but similar in order to the period-doubling chain [3]. For example, consider the integer lattice with spacing of occupied sites such that $f = 0.1$, chosen to match the density of our system of primes. The lattice has $c = 18$, much larger than the value $c = 0.1674$ for primes, which are closer to the period-doubling chain ($c = 0.1429$). (However, strictly speaking, comparing the magnitudes of the values of τ between the primes and the period-doubling chain and drawing conclusions about their relative degrees of order is not appropriate due to their very different occupation fractions.) In all of these ordered examples, τ grows with the system size L. This indicates multiscale order in the primes, absent from a case such as the Riemann zeros in which, assuming Montgomery's pair correlation conjecture, τ converges to 2/3 and in particular does not grow with L [26].

J. Stat. Mech. (2018) 093401

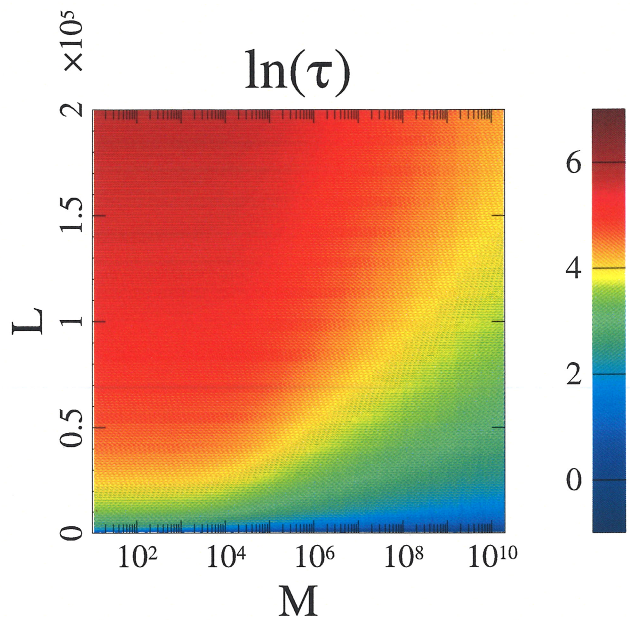

Figure 4. Natural logarithm of the order metric τ of prime numbers for $10 < M < 2\times10^{10}$ and $8 < L < 2\times10^{5}$. The 'warmer' colors in the upper left corner indicate systems with a larger value of τ, and hence stronger order. The 'cooler' colors in the lower right indicate relatively disordered systems comparable to an uncorrelated lattice gas. The interfaces separating these layers have a parabolic shape, which is explained by (18).

Now we study the τ order metric for prime-number configurations with different M and L. Recall that a deviation of τ from zero measures translational order with respect to the fully uncorrelated case. As derived from (18) and illustrated by figure 4, the constant-τ level curves have the form $L \sim \ln^2(M)$, i.e. level curves appear as quadratic curves in the log plot. For an uncorrelated lattice gas for any L and ρ, τ is very small. Thus, $L \sim \ln^2(M)$ is the boundary between regions where primes can be considered to be uncorrelated versus correlated. For the uncorrelated regime in which Gallagher's results apply, $L \sim \ln(M)$, and τ diminishes as M increases. As L increases, prime-number configurations move from the uncorrelated regime ($\tau \sim 1$, $L \leqslant \ln^2(M)$) to the limit-periodic regime we studied in this paper ($\tau \sim L$, $L \propto M$), and then to the inhomogeneous and nonhyperuniform regime, where the density gradient is no longer negligible (e.g. if $L \sim M^2$).

Recovery of Hardy–Littlewood conjecture

We can get the pair correlation function $g_2(r)$ of the primes via the limiting form of the structure factor (14) by performing the inverse Fourier transform of $S(k) - 1 \equiv \rho\tilde{h}(k)$ using (5), where $\tilde{h}(k)$ is the Fourier transform of $h(r) \equiv g_2(r) - 1$ for $r \neq 0$:

$$g_2(r) = 1 + \sum_{n\in\mathbb{Z}^+}{}^{\flat} \frac{1}{\phi^2(n)} \sum_{1\leqslant m\leqslant n-1}{}^{\times} \exp(rm\pi\mathrm{i}/n). \tag{19}$$

The expression given in equation (19) for distinct values of $r = 2, 4, 6, \ldots$ is a different representation of the constants in the famous Hardy–Littlewood k-tuple conjecture

J. Stat. Mech. (2018) 093401

Table 1. Hardy–Littlewood constants for $k = 2$ calculated with our formula (19) using $n_{\max} = 10\,000$ for $r = 2, 4, 6, 8$ and 10 compared to their accurate values [29]. The case $r = 2$ corresponds to the 'twin' primes. Notice that $g_2(2) = g_2(4) = g_2(8)$, which has a simple explanation in terms of the Hardy–Littlewood product over primes. This product includes a special contribution from primes dividing r, and the only such prime is 2 when r is a power of 2.

r	Prediction of (19)	Reference [29]
2	0.660 161 536	0.660 161 816
4	0.660 161 536	0.660 161 816
6	1.320 323 071	1.320 323 632
8	0.660 161 536	0.660 161 816
10	0.880 215 710	0.880 215 754

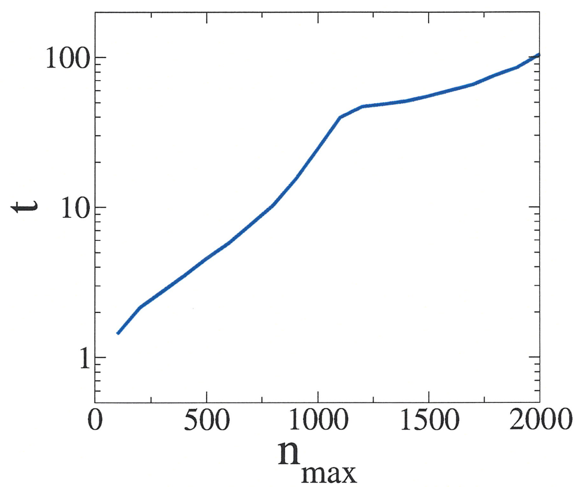

Figure 5. Prime prediction accuracy parameter t versus $n_{\max}$ for $M = 10^6$.

(theorem X, p 61 in [29]) for the special case $k = 2$. The Hardy–Littlewood conjecture is beyond doubt in the mathematical community, and yet far from a rigorous proof. Spectacular progress towards it was made recently by Maynard, Zhang, Tao, and the massive online collaboration Polymath 8 [34–36]. In particular, Maynard's theorem 1.2 from [36] shows that there must be infinitely many prime shifts for a positive proportion of k-tuples.

Summation of (19) for $r = 2, 4, 6, 8$ and 10 with a cut-off $n < 10\,000$ yields predictions that are in agreement with established values, as shown in table 1. Indeed, we can prove [26] that equation (19) is equivalent to Hardy and Littlewood's original expression [29]. This adds to the validity of our effective limit-periodic form of the structure factor of the primes, which has heretofore not been identified.

Reconstruction of the prime numbers

Note that we not only have an analytical formula for $S(k)$, but also for the complex density variable $\tilde{\eta}(k)$, defined by (7), which includes phase information. This analytical expression for $\tilde{\eta}(k)$ of the primes enables us to reconstruct, in principle, a prime-number configuration within an arbitrary interval $[M, M + L]$ by obtaining the inverse Fourier transform of $\tilde{\eta}(k)$. In practice, one is computationally limited by the fact that one can only include a finite number of peaks for which $n < n_{\max}$; see [26] for details. This leads to an algorithm to reconstruct primes in a dyadic interval with high accuracy provided that $n_{\max}$ is sufficiently large and M is not too large. A measure of the accuracy of the reconstruction algorithm is given by the ratio $t = N_c/N_i$, where N_c and N_i are the number of correctly predicted primes and incorrectly predicted primes (composite odd numbers), respectively. Figure 5 shows that the reconstruction procedure for $M = 10^6$ becomes highly accurate as the number of Bragg peaks incorporated, as measured by $n_{\max}$, increases.

Discussion

In summary, by focusing on the scattering characteristics of the primes in certain sufficiently large intervals, we have discovered that prime configurations are hyperuniform of class II and characterized by an unexpected order across length scales. In particular, they provide the first example of an effectively limit-periodic point process, a hallmark of which are dense Bragg peaks in the structure factor. The discovery of this hidden multiscale order in the primes is in contradistinction to their traditional treatment as pseudo-random numbers.

Effective limit-periodic systems represent a new class of many-particle systems with pure point diffraction patterns that deserve future investigation in physics, apart from their connection to the primes. For example, the formulation of other theoretical structural models of effectively limit-periodic point processes in one and higher dimensions and the study of their physical properties are exciting areas for further exploration.

Acknowledgments

We are grateful to Peter Sarnak for valuable discussions. This work was supported in part by the National Science Foundation under Award No. DMR-1714722. de Courcy-Ireland was supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] Chaikin P M and Lubensky T C 1995 *Principles of Condensed Matter Physics* (New York: Cambridge University Press)
[2] Levine D and Steinhardt P J 1984 Quasicrystals: a new class of ordered structures *Phys. Rev. Lett.* **53** 2477
[3] Baake M and Grimm U 2011 Diffraction of limit periodic point sets *Phil. Mag.* **91** 2661–70
[4] Socolar J E S and Taylor J M 2011 An aperiodic hexagonal tile *J. Comb. Theory* **118** 2207–31
[5] Granville A 1995 Harald Cramér and the distribution of prime numbers *Scand. Actuarial J.* **1995** 12–28
[6] Gallagher P X 1976 On the distribution of primes in short intervals *Mathematika* **23** 4–9
[7] Riemann B 1859 Ueber die anzahl der primzahlen unter einer gegebenen grosse *Ges. Math. Werke Wissenschaftlicher Nachlaß* **2** 145–55
[8] Torquato S and Stillinger F H 2003 Local density fluctuations, hyperuniform systems, and order metrics *Phys. Rev.* E **68** 041113
[9] Donev A, Stillinger F H and Torquato S 2005 Unexpected density fluctuations in disordered jammed hard-sphere packings *Phys. Rev. Lett.* **95** 090604

[10] Torquato S, Scardicchio A and Zachary C E 2008 Point processes in arbitrary dimension from fermionic gases, random matrix theory, and number theory *J. Stat. Mech.* P11019

[11] Batten R D, Stillinger F H and Torquato S 2008 Classical disordered ground states: super-ideal gases, and stealth and equi-luminous materials *J. Appl. Phys.* **104** 033504

[12] Florescu M, Torquato S and Steinhardt P J 2009 Designer disordered materials with large complete photonic band gaps *Proc. Natl Acad. Sci.* **106** 20658–63

[13] Zachary C E, Jiao Y and Torquato S 2011 Hyperuniform long-range correlations are a signature of disordered jammed hard-particle packings *Phys. Rev. Lett.* **106** 178001

[14] Jiao Y, Lau T, Hatzikirou H, Meyer-Hermann M, Corbo J C and Torquato S 2014 Avian photoreceptor patterns represent a disordered hyperuniform solution to a multiscale packing problem *Phys. Rev.* E **89** 022721

[15] Torquato S, Zhang G and Stillinger F H 2015 Ensemble theory for stealthy hyperuniform disordered ground states *Phys. Rev.* X **5** 021020

[16] Jack R L, Thompson I R and Sollich P 2015 Hyperuniformity and phase separation in biased ensembles of trajectories for diffusive systems *Phys. Rev. Lett.* **114** 060601

[17] Ghosh S and Lebowitz J L 2016 Fluctuations, large deviations and rigidity in hyperuniform systems: a brief survey (arXiv:1608.07496)

[18] Hexner D, Chaikin P M and Levine D 2017 Enhanced hyperuniformity from random reorganization *Proc. Natl Acad. Sci.* **114** 4294–99

[19] Ricouvier J, Pierrat R, Carminati R, Tabeling P and Yazhgur P 2017 Optimizing hyperuniformity in self-assembled bidisperse emulsions *Phys. Rev. Lett.* **119** 208001

[20] Bombieri E 2000 *The Riemann Hypothesis-Official Problem Description* (Providence, RI: American Mathematical Society and Clay Mathematics Institute)

[21] Montgomery H L 1973 The pair correlation of zeros of the zeta function *Am. Math. Soc.* **24** 181–93

[22] Dyson F J 1962 Statistical theory of the energy levels of complex systems. I *J. Math. Phys.* **3** 140–56

[23] Mehta M L 1991 *Random Matrices* (New York: Academic)

[24] Rudnick Z and Sarnak P 1996 Zeros of principal L-functions and random matrix theory *Duke Math. J.* **81** 269–322

[25] Dyson F J 2009 Birds and frogs *Not. Am. Math. Soc.* **56** 212–23

[26] Torquato S, Zhang G and de Courcy-Ireland M 2018 Hidden multiscale order in the primes (arXiv:1804.06279)

[27] Hadamard J 1896 Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques *Bull. Soc. Math. France* **24** 199–220

[28] Zhang G, Martelli F and Torquato S 2018 Structure factor of the primes *J. Phys. A: Math. Theor.* **51** 115001

[29] Hardy G H and Littlewood J E 1923 Some problems of partitio numerorum; III: on the expression of a number as a sum of primes *Acta Math.* **44** 1–70

[30] Tenenbaum G 1995 *Introduction to Analytic and Probabilistic Number Theory* vol 46 (Cambridge: Cambridge University Press)

[31] Oğuz E C, Socolar J E S, Steinhardt P J and Torquato S 2017 Hyperuniformity of quasicrystals *Phys. Rev.* B **95** 054119

[32] Torquato S 2018 Hyperuniform states of matter *Phys. Rep.* **745** 1–95

[33] DiStasio R A, Zhang G, Stillinger F H and Torquato S 2018 Rational design of stealthy hyperuniform patterns with tunable order *Phys. Rev.* E **97** 023311

[34] Zhang Y 2014 Bounded gaps between primes *Ann. Math.* **3** 1121–74

[35] Polymath D H J 2014 Variants of the Selberg sieve, and bounded intervals containing many primes *Res. Math. Sci.* **1** 12

[36] Maynard J 2015 Small gaps between primes *Ann. Math.* **181** 383–413

J. Stat. Mech. (2018) 093401