Journal of Physics A: Mathematical and Theoretical

PAPER

The structure factor of primes

To cite this article: G Zhang *et al* 2018 *J. Phys. A: Math. Theor.* **51** 115001

IOP Publishing Journal of Physics A: Mathematical and Theoretical

J. Phys. A: Math. Theor. **51** (2018) 115001 (16pp) https://doi.org/10.1088/1751-8121/aaa52a

The structure factor of primes

G Zhang[1], F Martelli[1] and S Torquato[2]

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

[2] 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

E-mail: torquato@electron.princeton.edu

Received 4 October 2017, revised 19 December 2017
Accepted for publication 4 January 2018
Published 14 February 2018

CrossMark

Abstract

Although the prime numbers are deterministic, they can be viewed, by some measures, as pseudo-random numbers. In this article, we numerically study the pair statistics of the primes using statistical–mechanical methods, particularly the structure factor $S(k)$ in an interval $M \leqslant p \leqslant M+L$ with M large, and L/M smaller than unity. We show that the structure factor of the prime-number configurations in such intervals exhibits well-defined Bragg-like peaks along with a small 'diffuse' contribution. This indicates that primes are appreciably more correlated and ordered than previously thought. Our numerical results definitively suggest an explicit formula for the locations and heights of the peaks. This formula predicts infinitely many peaks in any non-zero interval, similar to the behavior of quasicrystals. However, primes differ from quasicrystals in that the ratio between the location of any two predicted peaks is rational. We also show numerically that the diffuse part decays slowly as M and L increases. This suggests that the diffuse part vanishes in an appropriate infinite-system-size limit.

Keywords: prime numbers, structure factor, hyperuniformity

(Some figures may appear in colour only in the online journal)

1. Introduction

The properties of prime numbers have been a source of fascination since ancient times. Euclid proved that there are infinitely many primes. Given the first n prime numbers $p_1, p_2, \cdots, p_n$, the subsequent prime can be found deterministically by sieving [1]. Nonetheless, there is no known deterministic formula that can quickly (polynomial in the number of digits in a prime) generate large numbers that are guaranteed to be prime. (So far, the largest known prime is

1751-8121/18/115001+16$33.00

$2^{77,232,917} - 1$, which is about 22 million digits long [2].) Let $\pi(x)$ denote the *prime counting function*, which gives the number of primes less than integer x. According to the prime number theorem [3], the prime counting function in the large-x asymptotic limit is given by

$$\pi(x) \sim \frac{x}{\ln(x)} \qquad (x \to \infty). \tag{1}$$

This means that for sufficiently large x, the probability that a randomly selected integer not greater than x is prime is very close to $1/\ln(x)$, which can be viewed as position-dependent number density $\rho(x)$ (number of primes up to x divided by the interval x). This implies that primes become sparser as x increases and hence can be regarded as a statistically inhomogeneous set of points that are located on a subset of odd integers.

While the prime numbers (except for 2) are a deterministic subset of odd integers, they can be viewed, by some measures, as pseudo-random numbers. Moreover, there are quick *stochastic* ways to find large primes [7–11], examples of which are based on variants of Fermat's little theorem [7–10]. To get a sense of how primes can be viewed as pseudo-random numbers, let us consider the *gap distribution function* $P(z)$, which gives the probability distribution of the gap size between two consecutive primes, z. Figure 1 compares the *gap probability distribution* $P(z)$ for primes to the *uncorrelated lattice gas* at the same number density. An 'uncorrelated lattice gas' refers to a lattice-gas system where each site has a certain probability of being occupied, independent of the occupation of other sites. For an uncorrelated lattice gas at number density $\rho = N/L$ with lattice spacing 2 in the infinite-system-size limit, the gap distribution is exactly given by

$$P(z) = f(1-f)^{z/2-1}, \tag{2}$$

where $f = 2\rho$ is the probability that a site is occupied. We see that by the gap distribution, primes cannot be clearly distinguished from an uncorrelated lattice gas. Indeed, probabilistic methods to treat primes have yielded fruitful insights about them [4–6]. For example, based on the assumption that primes behave like a Poisson process (uncorrelated lattice gas), Cramér (1920) conjectured that [4] for large x

$$g(x) \geqslant c\ln^2(x) \tag{3}$$

where $g(x)$ denotes the *largest prime gap* within an interval $[x,2x]$.

On the other hand, it is known that primes contain unusual patterns. Chebyshev observed in 1853 that primes congruent to 3 modulo 4 seem to predominate over those congruent to 1 [12]. Assuming a generalized Riemann hypothesis, Rubinstein and Sarnak [13] exactly characterized this phenomenon and more general related results. A computational study on the Goldbach conjecture demonstrates a connection based on a modulo 3 geometry between the set of even integers and the set of primes [14]. In 1934, Vinogradov proved that every sufficiently large odd integer is the sum of three primes [15]. This method has been extended to cover many other types of patterns [16–19]. Recently, it has been shown that there are infinitely many pairs of primes with some finite gap [20] and that primes ending in 1 are less likely to be followed by another prime ending in 1 [21]. Numerical evidence of regularities in the distribution of gaps between primes when these are divided into congruence families have also been reported [22–24], along with the observation of period-three oscillations in the distribution of increments of the distances between consecutive prime numbers [25].

The present paper is motivated by certain unusual properties of the Riemann zeta function $\zeta(s)$, which is a function of a complex variable s that is intimately related to primes. The zeta function has many different representations, one of which is the series formula

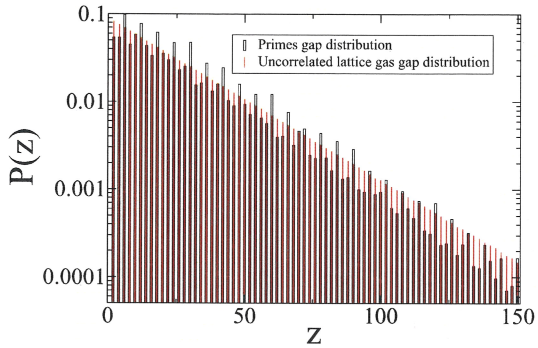

Figure 1. Comparison of the gap distribution for primes and the uncorrelated lattice gas with the same cardinality (occupation number) as the set of primes. The primes are taken to lie on an integer lattice with a spacing of 2, i.e. a subset of odd positive integers. We consider N primes in interval $[M,M+L]$ (M large and $M \gg L$). Here $N = 10^7$, $L = 244\,651\,478$ with $M = 42\,151\,671\,493$, the 1800 000 000th prime number.

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \tag{4}$$

which only converges for $\mathrm{Re}(s) > 1$. However, $\zeta(s)$ has a unique analytic continuation to the entire complex plane, excluding the simple pole at $s = 1$. According to the *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. The nontrivial zeros tend to get denser the higher they are on the critical line. When the spacings of the zeros are appropriately normalized so that they can be treated as a homogeneous point process at unity density, the resulting pair correlation function takes on the simple form $1 - \sin^2(\pi r)/(\pi r)^2$ [6]. The corresponding structure factor $S(k)$ (essentially the Fourier transform of $g_2(r)$) tends to zero linearly in the wavenumber k as k tends to zero but is unity for sufficiently large k. This implies that the normalized Riemann zeros possess a remarkable type of correlated disorder at large length scales known as hyperuniformity [26]. A hyperuniform many-particle system is one in which the structure factor approaches zero in the infinite-wavelength limit [27]. In such systems, density fluctuations are anomalously suppressed at very large length scales, a 'hidden' order that imposes strong global structural constraints. All structurally 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 [28–45]. The zeta function is directly related to primes via the following Euler product formula:

$$\zeta(s) = \left[\prod_{n=1}^{\infty}[1 - 1/p_n^s]\right]^{-1}, \tag{5}$$

which leads to a variety of *explicit* formulas that link primes on the one hand to the zeros of the zeta function on the other hand [46–48]. Thus, one can in principle deduce information about primes from information about zeros of the zeta function. Accordingly, one might expect primes to encode hyperuniform correlations that are seen in the Riemann zeros.

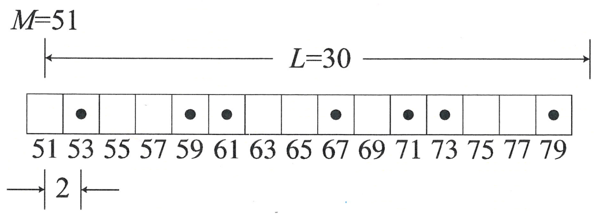

Figure 2. Schematic plot of a prime-number configuration with $M = 51$ and $L = 30$. Since we always use $M \geqslant 3$, any prime number in the interval $[M, M + L)$ is odd. Therefore, a prime-number configuration is a lattice gas with lattice spacing 2 in which the primes are the 'occupied' sites and the composites are 'unoccupied' sites.

In this article, we numerically study the pair statistics of configurations of primes, especially the structure factor $S(k)$ in an interval $M \leqslant p \leqslant M + L$ with M large, and L/M smaller than unity. As we will detail in section 2.3, this choice of intervals allow us to obtain prime configurations with virtually constant density from one end of the interval to the other. We show in section 3 that the structure factor exhibits well-defined Bragg-like peaks along with a small 'diffuse' contribution. This indicates that primes are appreciably more correlated than anyone has previously conceived. Our numerical results definitively suggest an explicit formula for the locations and heights of the peaks. This formula predicts infinitely many peaks in any non-zero interval. Although such a behavior is similar to that of quasicrystals [51–53], primes differ from the latter in that the ratio between the locations of any two predicted peaks is rational. We also show numerically that the diffuse part decays slowly as M or L increases. In a subsequent paper [55], we prove this to be true and investigate its consequences. This might indicate that the diffuse part vanishes in an appropriate infinite-system-size limit. We note that the structure factor has been used to study other number-theoretic systems [7, 26, 49, 50].

The rest of the paper is organized as follows: in section 2, we present relevant definitions and describe the simulation procedure. In section 3, we present results for the pair statistics of primes in certain intervals in both direct and reciprocal spaces. In section 4, we make concluding remarks.

2. Definitions and simulation procedure

We will study the pair correlation function as well as the structure factor of prime-number configurations. Similar to [54], we treat primes in some interval $[M, M + L]$ to be a special lattice-gas model: the primes are 'occupied' sites on a integer lattice of spacing 2 that contains all of the positive odd integers and the unoccupied sites are the odd composite integers. As detailed below, we consider the positive parameter L/M generally to be smaller than unity to ensure homogeneity. Within this interval, let N_s and N be the total number of sites and the total number of primes, respectively. In practice, we consider the half-open interval $[M, M + L)$ because it provides a simple relation between L and N_s, namely, $L = 2N_s$. For all cases considered in this paper, $M \geqslant 3$. Figure 2 illustrates an example of a prime configuration. The rest of this section details the mathematical tools that we use to treat such systems.

2.1. Discrete fourier transform

For a function $f(r)$ defined on an integer lattice with spacing a that is contained within a periodic box of length L, one may define its Fourier transform as follows:

$$\tilde{f}(k) = \sum_{r=0,a,2a,\cdots,L-a} f(r)\exp(\mathrm{i}kr), \tag{6}$$

where the parameter k is an integer multiple of $2\pi/L$. The inverse transform is given by

$$f(r) = \frac{1}{N_s}\sum_{j=0}^{N_s-1}\tilde{f}\left(\frac{2\pi j}{L}\right)\exp\left(-\mathrm{i}\frac{2\pi j}{L}r\right), \tag{7}$$

where $N_s = L/a$ is the number of sites.

2.2. Pair statistics and their basis properties

We define $\eta(r)$ as the indicator function such that $\eta(r) = 1$ if the site at r is occupied, and $\eta(r) = 0$ otherwise. Let $\tilde{\eta}(r)$ be its Fourier transform. We define occupation fraction to be $f =< \eta(r) >$, where $<>$ denotes an average over all r. We define the structure factor as

$$S(k) = |\tilde{\eta}(k)|^2/N - N\delta_{k,0}. \tag{8}$$

Define the pair correlation function $g_2(r)$ as

$$g_2(r) = \frac{1}{Nf}\sum_{n=0,a,2a,\cdots,L-a}\eta(n)\eta(n+r) - \frac{\delta_{r,0}}{f}. \tag{9}$$

By definition, $g_2(0) = 0$. For $r \neq 0$, $g_2(r)$ can be interpreted as the probability that the site at $p + r$ is occupied given that the site at p is occupied divided by f.

The structure factor and the pair correlation function are related as follows:

$$\begin{aligned}
N\delta_{k,0} + S(k) &= |\tilde{\eta}(k)|^2/N \\
&= \frac{1}{N}\sum_{m=0,a,2a,\cdots,L-a}\eta(m)\exp(\mathrm{i}km)\sum_{n=0,a,2a,\cdots,L-a}\eta(n)\exp(-\mathrm{i}kn) \\
&= \frac{1}{N}\sum_{m=0,a,2a,\cdots,L-a}\sum_{n=0,a,2a,\cdots,L-a}\eta(m)\eta(n)\exp[\mathrm{i}k(m-n)] \\
&= \frac{1}{N}\sum_{r=0,a,2a,\cdots,L-a}\sum_{n=0,a,2a,\cdots,L-a}\eta(n+r)\eta(n)\exp(\mathrm{i}kr) \\
&= \frac{1}{N}\sum_{n=0,a,2a,\cdots,L-a}\eta(n+0)\eta(n)\exp(\mathrm{i}k0) \\
&+ \frac{1}{N}\sum_{r=a,2a,\cdots,L-a}\sum_{n=0,a,2a,\cdots,L-a}\eta(n+r)\eta(n)\exp(\mathrm{i}kr) \\
&= \frac{N}{N} + \frac{1}{N}\sum_{r=a,2a,\cdots,L-a}Nfg_2(r)\exp(\mathrm{i}kr) \\
&= 1 + f\sum_{r=a,2a,\cdots,L-a}g_2(r)\exp(\mathrm{i}kr) \\
&= 1 + f\sum_{r=0,a,\cdots,L-a}g_2(r)\exp(\mathrm{i}kr).
\end{aligned} \tag{10}$$

This equation enables us to obtain a sum rule for both $g_2(r)$ and $S(k)$. For $g_2(r)$, plugging $k = 0$ into equation (10) yields

$$\frac{N-1}{f} = \sum_{r=a,2a,\cdots,L-a} g_2(r). \tag{11}$$

The sum rule for $S(k)$ is easily found by invoking the inverse Fourier transform equation:

$$g_2(r) = \frac{1}{N_s f} \sum_{j=0}^{N_s-1} \left[S\left(\frac{2\pi j}{L}\right) + N\delta_{k,0} - 1 \right] \exp\left(-\mathrm{i}\frac{2\pi j}{L} r\right). \tag{12}$$

At $r = 0$, this relation becomes

$$\begin{aligned} 0 &= \sum_{j=0}^{N_s-1} \left[S\left(\frac{2\pi j}{L}\right) + N\delta_{k,0} - 1 \right] \\ &= \sum_{j=1}^{N_s-1} S\left(\frac{2\pi j}{L}\right) + N - N_s, \end{aligned} \tag{13}$$

and hence the sum rule for the structure factor is given by

$$\sum_{j=1}^{N_s-1} S\left(\frac{2\pi j}{L}\right) = N_s - N. \tag{14}$$

For the primes, $L = 2N_s$, and hence the sum rule is specifically

$$\sum_{j=1}^{N_s-1} S\left(\frac{\pi j}{N_s}\right) = N_s - N. \tag{15}$$

The fact that all primes greater than 3 are odd integers lead to a few important properties of $S(k)$. First, $N\delta_{k,0} + S(k)$ is a periodic function of period π ,since

$$\begin{aligned} N\delta_{k+\pi,0} + S(k+\pi) &= \frac{|\sum_{j=1}^{N} \exp[-\mathrm{i}(k+\pi)(p_j - M)]|^2}{N} \\ &= \frac{|\sum_{j=1}^{N} \exp[-\mathrm{i}k(p_j - M)] \exp(-\mathrm{i}\pi(p_j - M))|^2}{N} \\ &= \frac{|-\exp(\mathrm{i}\pi M) \sum_{j=1}^{N} \exp[-\mathrm{i}k(p_j - M)]|^2}{N} \\ &= \frac{|\sum_{j=1}^{N} \exp[-\mathrm{i}k(p_j - M)]|^2}{N} \\ &= N\delta_{k,0} + S(k). \end{aligned} \tag{16}$$

Second, from equations (6) and (8), one can see that when $k = m\pi$, where m is any non-zero integer, $S(k) = N$ achieves the global maximum of this function. The function $S(k)$ therefore displays strong peaks at such k values. Third, from equations (6) and (8), one can see the function $S(k)$ has reflection symmetry $S(k) = S(-k)$. This reflection symmetry, combined with the periodicity (equation (16)), implies another reflection symmetry, $S(\pi/2 + k) = S(\pi/2 - k)$. With these properties in mind, we only need to study $S(k)$ in the range $0 < k \leqslant \pi$ in this paper.

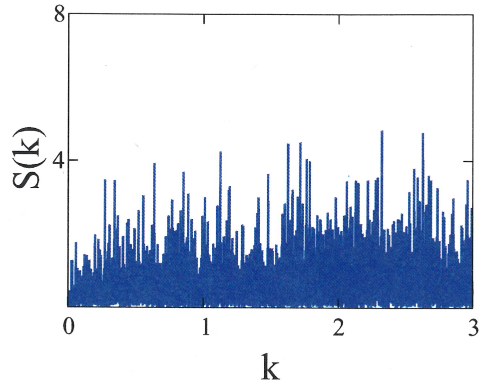

Figure 3. Structure factor $S(k)$ associated with $p/\ln(p)$ for all prime number p's in the interval $[3, 3 + 10^5)$.

2.3. Simulation procedure

In statistical mechanics, the study of $S(k)$ often focuses on statistically homogeneous systems. However, prime numbers are not homogeneous. Instead, in the vicinity of x, the density of prime numbers scales as $1/\ln(x)$. To overcome this difference, we focus on large M values and let L/M be a constant less than unity[3]. This implies that the 'local' density from the beginning to the end of the interval $[M, M + L)$ is virtually a constant. For example, we will study a system of $M = 10^{10}$ and $L = 10^7$. As x changes from M to $M + L$, $1/\ln(x)$ only changes from 0.043 429… to 0.043 427….

We minimize the problem of inhomogeneity by requiring sufficiently large M. Instead, one might think an even better solution to this problem is to rescale the configuration such that it is homogeneous. The natural scaling is to replace each prime number p with $p/\ln(p)$. However, it turns out that after performing such rescaling, the structure factor appears to be completely noisy with no obvious peaks with heights comparable to N; see figure 3[4]. This is to be contrasted with our findings reported in the rest of the paper, in which we choose to study primes in the interval $[M, M + L)$ with M large, and L/M smaller than unity.

We obtain a list of prime numbers from [56], and calculate $S(k)$ of prime numbers and uncorrelated lattice gases with the fast Fourier transform (FFT) algorithm using the kissFFT software [57]. This algorithm has the advantage of not only being 'fast'[5], but also being accurate, as the upper bound on the relative error scales as $\epsilon \log(L)$, where ϵ is the machine floating-point relative precision. We use double-precision numbers to further minimize ϵ.

More precisely, FFT allows one to efficiently calculate

$$X(q) = \sum_{n=0}^{T-1} x_n \exp(-2\pi \mathrm{i} qn/T) \tag{17}$$

[3] Strictly speaking, the requirement that $L/M < 1$ is not important. Densities from one end to the other would be nearly constant in the $M \to \infty$ limit even if $L/M = \beta > 1$. However, for computationally practical values of M, a much smaller L is desired to accurately ascertain constant density.

[4] Notice that in this interval, primes possess a sharp density gradient, and the $p \to p/\ln(p)$ rescaling is strongly non-linear. If we had chosen an interval with a negligible density gradient, then the $p \to p/\ln(p)$ rescaling would be almost linear, and $S(k)$ would essentially be a simple rescaling of the ones reported below (figure 5).

[5] The time complexity of calculating $S(k)$ for all k's using FFT algorithm scales as $L\log(L)$, while the time complexity of doing so using equation (10) scales as LN.

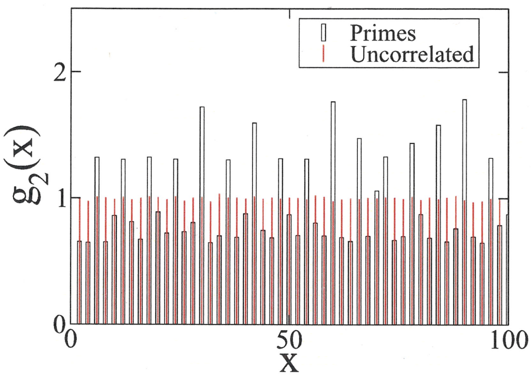

Figure 4. Pair correlation function $g_2(x)$, as defined in equation (9), for a prime number configuration of $M = 10^{10} + 1$, $L = 10^6$, and $N = 43\,427$, compared with $g_2(x)$ of a uncorrelated lattice gas configuration of the same L and N.

for arbitrary $x_0, x_1, \cdots, x_{T-1}$. Here, we simply let $T = L/2$, and let $x_j = 1$ if $M + 2j$ is a prime number and $x_j = 0$ if $M + 2j$ is composite. The structure factor is then calculated from:

$$S(2\pi q/L) = |X(q)|^2/N. \tag{18}$$

3. Results for the pair statistics of prime numbers

The pair correlation function as defined in equation (9), $g_2(x)$, of prime numbers is presented in figure 4 and compared with $g_2(x)$ of uncorrelated lattice gases. This quantity for uncorrelated lattice gas is simply

$$g_2(x) = \frac{N-1}{f(N_s-1)} \tag{19}$$

for any $x \neq 0$. This is because after one site is occupied, out of the remaining $N_s - 1$ sites, exactly $N - 1$ sites are occupied. We see that by this measure, the prime numbers appear to be distinctly different from uncorrelated lattice gases. We see that $g_2(x)$ for primes is higher than $g_2(x)$ of the uncorrelated lattice gas if and only if x is divisible by 3. However, we will see that the difference in pair statistics is much more obvious when we study $S(k)$ below.

We present and study numerically calculated structure factors of prime numbers for various M's and L's in this section. At first, let us examine $S(k)$ for $M = 10^6 + 1$ and $L = 5000$, which is presented in figure 5. At a larger scale, $S(k)$ appears to consist of many well-defined Bragg-like peaks of various heights, with the highest peak occurring at $k = \pi$ (left panel). As we zoom in, it becomes evident that besides those peaks, $S(k)$ also has a random, noisy contribution that is often below 1 (right panel). We will call the latter contribution the 'diffuse part' in the rest of the paper. Figure 5 also includes $S(k)$ for uncorrelated lattice gases, which consists of a diffuse part and a single peak at the trivial value of $k = \pi$. Away from the peak, $S(k)$ for uncorrelated lattice gases fluctuates around an average value of $\frac{N_s-N}{N_s-1}$ (this particular average value is required by the sum rule, equation (14)). A major conclusion is that the structure factor of primes is characterized by a substantial amount of order across length scales, relative to the uncorrelated lattice gas, as evidenced by the appearance of many Bragg-like peaks. At this stage, it seems that the existence of the diffuse part makes primes non-hyperuniform.

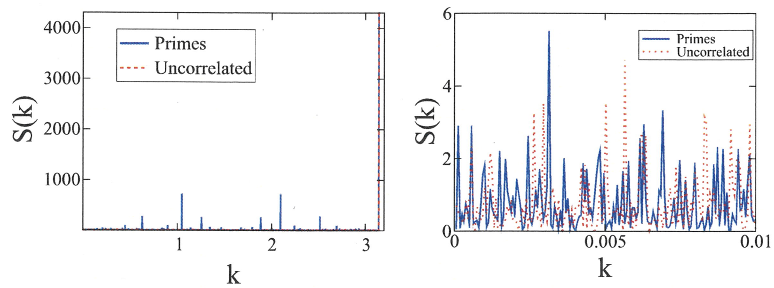

Figure 5. Left: $S(k)$ for prime numbers as a function of k (in units of the integer lattice spacing) for $M = 10^{10} + 1$ and $L = 10^5$ contains many well-defined Bragg-like peaks of various heights, creating a type of self-similarity. Right: a zoomed-in view revealing the existence of a small, noisy 'diffuse part' besides the peaks. We also plot $S(k)$ for uncorrelated lattice gases for comparison. As we have discussed in section 2, we only show $S(k)$ in the range $0 < k \leqslant \pi$, and therefore omit the peak at $k = 0$.

However, we will show in section 3.2 that the diffuse part decreases as L increases, and suggest that it vanishes in the infinite-system-size limit.

At this stage, the distinction between the peaks and the 'diffuse part' is somewhat unclear. Since $S(k)$ contains peaks of various heights, is it possible that the diffuse part is actually made of many smaller peaks? We can only answer this question after we study the peaks and the diffuse part more deeply later in this section.

3.1. Peaks

We move on to study the peaks. From figure 5, one sees that the highest peak is at $k = \pi$, which is trivially caused by the periodicity of the underlying lattice. The next highest two peaks are at $k = \pi/3$ and $k = 2\pi/3$[6]. Even lower peaks occur at $k = \pi/5$, $2\pi/5$, $3\pi/5$, and $4\pi/5$. Still lower peaks occur at $k = \pi/7$, $2\pi/7$, $3\pi/7$, $4\pi/7$, $k = 5\pi/7$, and $6\pi/7$. Examining $S(k)$ of a much larger system ($M = 10^{10} + 1$ and $L = 10^7$) revealed that there are even more peaks with locations that obey the formula $k = m\pi/n$, where m is any integer coprime with n and n is any square-free odd integer and hence has a distinct prime factorization, i.e. $n = \prod_{j=1}^{J} p_j$, where J is a positive integer, and $p_1, p_2, \cdots, p_J$ are non-repeating prime numbers larger than 2. If n is even or is not square-free, then we observe no peak at $k = m\pi/n$. We verified the existence of such peaks for n up to 300. As n increases beyond 300, however, the peaks become too weak to be distinguishable from the diffuse part.

Having an analytical formula of the peak locations, we move on to study the peak heights. As we have shown earlier, the height of the peak at $k = \pi$ is simply N. What can we say about the heights of the other peaks? In figure 6 we present computed peak heights at $k = \pi/3$ and $k = \pi/5$ for $M = 10^{10} + 1$ and various L's. We see that as L grows, the heights of the peaks at $k = \pi/3$ and $k = \pi/5$ also grow and remain roughly proportional to N. Looking at the inset, we see that both $S(\pi/3)$ and $S(\pi/5)$ oscillates periodically as L increases: $S(\pi/3)$ attains a maximum when L is divisible by 3, and $S(\pi/5)$ attains a maximum when L is divisible by 5. Examining the heights of other peaks, we find that the height of a peak at $k = m\pi/n$ is indeed highest when L is divisible by n.

[6] It should be noted that since k has to be integer multiples of $2\pi/L$, for $L = 10^5$, $k = \pi/3$ and $k = 2\pi/3$ cannot be chosen. The actually observed peaks occur at the closest allowed k points instead.

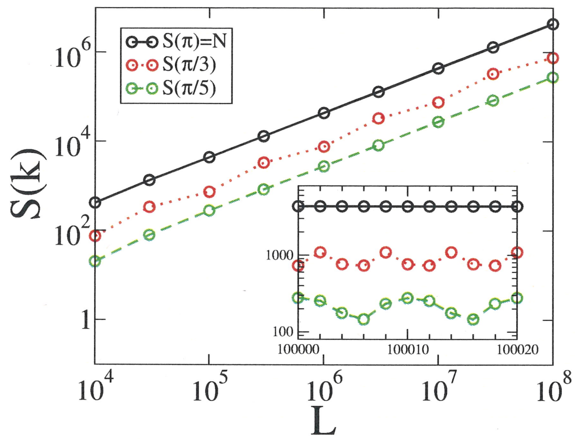

Figure 6. The structure factor $S(k)$ at $k = \pi$, $k = \pi/3$, and $k = \pi/5$, as a function of L at $M = 10^{10} + 1$. The inset presents more data for $10^5 \leqslant L \leqslant 10^5 + 20$.

Since the divisibility of L with n affects the peak heights and hence will introduce unintended errors if not chosen properly, we desire an L that is divisible by as many prime numbers as possible. We therefore chose $L = 2 \times 3 \times 5 \times 7 \times 11 \times 13 \times 17 \times 19 = 9699\,690$ and recomputed the heights of several peaks. The results are summarized in table 1. We find that when L is divisible by n, the height of the peak at $k = m\pi/n$ is very close to $N \prod_{j=1}^{J} (p_j - 1)^{-2}$, where p_j are the distinct prime factors of n.

Do these numerically generated peaks have finite or infinitesimal width? To answer this question, we present a close view of the peak at $k = \pi/3$ for three different L's in figure 7. It turns out that, if L is divisible by n, the peak at $k = m\pi/n$ has infinitesimal width, in the sense that $S(k)$ at one k value attains the local maximum and $S(k)$ at all adjacent k values are as low as the typical diffuse part. However, if L is not divisible by n, then the peak has a finite width, as $S(k)$ of all k values very close to $m\pi/n$ rises and become much higher than the typical diffuse part. In figure 7 one can also see that when peak widths are finite, a lower L results in a more broadly spread peak. Therefore, all of the peaks may have infinitesimal width in the infinite-L limit. However, in a finite-L simulation, choosing an L that is divisible by as many prime numbers as possible provides a better estimate of $S(k)$ in the infinite-L limit. As n increases, one finds an increasing number of lower peaks, resulting in a statistical self-similarity.

3.2. Diffuse part

The above analysis suggest that the structure factor of primes possess infinitely many Dirac-delta-function peaks of various heights in the infinite-system-size limit. Therefore, one might naturally ask, could the random, noisy 'diffuse part' be simply a superposition of many small peaks? The answer is no. In figure 8 we present $S(k)$ of two different L's in the range $0.2 \leqslant k \leqslant 0.24$. For $L = 5000$, $S(k)$ in this range appeared completely random and noisy, matching our definition of the diffuse part. To see if this diffuse part is actually a superposition of many small peaks, we compare it to the structure factor for $L = 500\,000$. The larger L allows more k points to be chosen, and therefore improves the k resolution, and reveals

Table 1. Peak heights at several different n and m's for $M = 2.5\times10^8 + 1$ and $L = 9699\,690$ and comparison with the predicted height from the analytical formula.

n	m	$S(m\pi/n)/N$	Postulated analytical formula
3	1	0.250 000 0003	$(3-1)^{-2} = 0.25$
5	1	0.062 682 935 36	$(5-1)^{-2} = 0.0625$
	2	0.062 318 335 26	
7	1	0.027 646 966 27	$(7-1)^{-2} = 0.027\,77\cdots$
	2	0.027 834 230 55	
	3	0.027 852 824 86	
$15 = 3\times5$	1	0.015 641 151 90	$[(3-1)(5-1)]^{-2} = 0.015\,625$
	2	0.015 832 663 09	
	4	0.015 518 143 12	
	7	0.015 509 640 66	
$105 = 3\times5\times7$	1	0.000 409 696 3803	$[(3-1)(5-1)(7-1)]^{-2} = 0.000\,434\,027\,77\cdots$
	2	0.000 441 802 5682	
	4	0.000 430 592 4622	
	8	0.000 387 997 4866	
	11	0.000 420 322 3484	
	13	0.000 441 110 7279	
	16	0.000 419 124 9498	
	17	0.000 389 371 6268	
	19	0.000 438 812 8207	
	22	0.000 419 303 6024	
	23	0.000 437 559 9695	
	26	0.000 420 361 3535	
	29	0.000 441 818 7004	
	31	0.000 445 765 0582	
	32	0.000 423 763 5619	
	34	0.000 450 097 9160	
	37	0.000 446 659 7486	
	38	0.000 466 330 4920	
	41	0.000 425 567 3779	
	43	0.000 484 593 3410	
	44	0.000 467 958 9962	
	46	0.000 457 298 5410	
	47	0.000 409 563 7658	
	52	0.000 452 196 2772	

some peaks in this k range. We see that although $S(k)$ of the smaller L appeared to be entirely random, the maximum at $k \approx 0.21$ corresponds to a strong peak of the larger system, and is therefore actually a peak. However, other maxima for the smaller system do not correspond to peaks for the larger system, and can only be explained by assuming the existence of a noisy contribution to $S(k)$, which we call the 'diffuse part'. To summarize, in figure 8 we show that for finite systems, there is clearly a noisy contribution to $S(k)$ other than the peak contribution, even though distinguishing these two components of $S(k)$ can be difficult without consulting the analytical formula for the peaks.

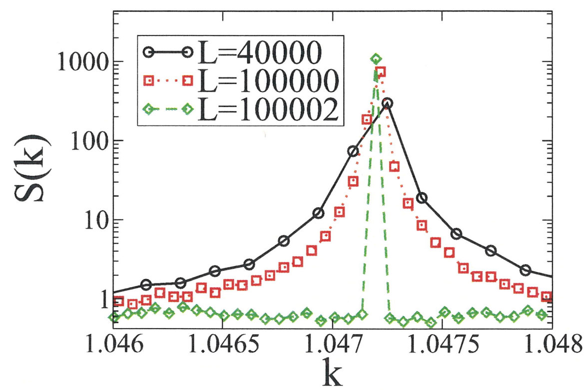

Figure 7. $S(k)$ near $k = \pi/3$ for three different L's. Each curve is averaged over 100 prime-number configurations, with the jth configuration consists of all prime numbers in the range $[10^{10} + (j-1)L + 1, 10^{10} + jL + 1)$.

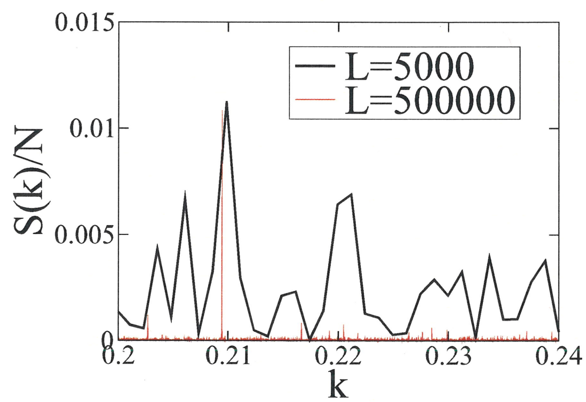

Figure 8. The structure factor, $S(k)$, normalized by N, for two different L's and $M = 10^6 + 1$.

The diffuse part contributes to not only k points where there are no peaks, but also to k points where there are peaks. In figure 9, we compare numerical peak heights, averaged over all allowed m for a particular n, with the analytic peak formula described in section 3.1. It turns out that the numerical average is always slightly higher than that predicted by this formula, and their difference is of the same order of magnitude as the diffuse part. Thus, $S(k)$ at predicted peak locations is actually the sum of the peak and diffuse contributions.

We can quantify the diffuse part using the median of $S(k)$ for all possible choices of k in the range $0 < k \leqslant \pi$. We present these medians in figure 10. As L increases, the median of $S(k)$ generally decreases. However, with our current data, it is unclear if the median of $S(k)$ approaches zero in the $L \to \infty$ limit.

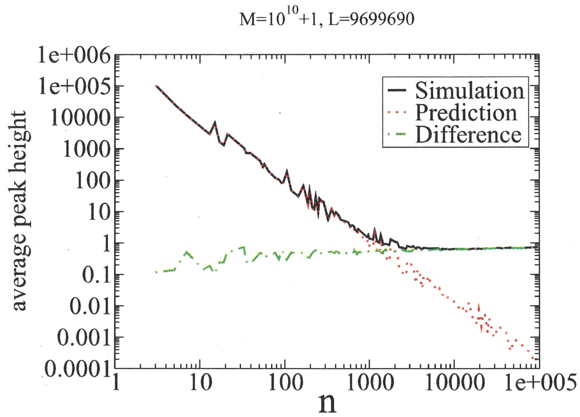

Figure 9. Average peak height of all peaks of a given n, the predicted peak heights, and their difference for all $n < 10^5$ that are odd, square-free, and divide L evenly. Here $M = 10^{10} + 1$ and $L = 9699\,690$. For each n, we find all m's that are coprime with n, and average the heights of peaks at $m\pi/n$. The average turns out to be always greater than the prediction, $N\prod_{j=1}^{J}(p_j - 1)^{-2}$. Their difference is between 0.1 and 1, which is of the same order of magnitude as the diffuse part.

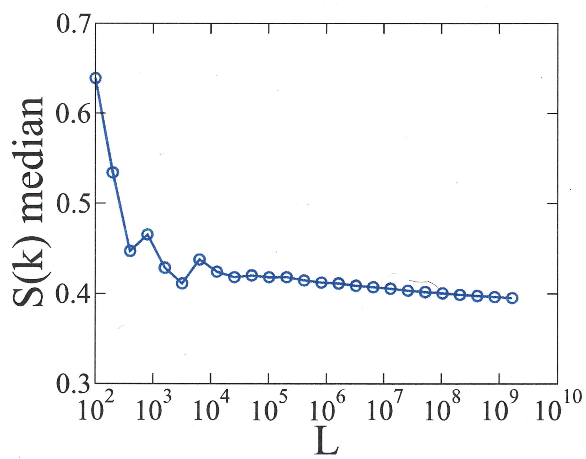

Figure 10. The median of the structure factor, $S(k)$, for all possible choices of k, as a function of L. Here M is chosen to be $10L$. The diffuse part of primes appears to be slowly decreasing as L increases. This is to be contrasted with the uncorrelated lattice gas with an appreciably larger predictable diffuse part in which there is no dependence on the system size.

4. Conclusions

In summary, we have numerically shown that the structure factor of prime numbers in the interval $M \leqslant p < M + L$ with large M and $L/M < 1$ exhibit well-defined Bragg-like peaks, alongside a very small diffuse part, which slowly decays as the system size increases.

In contrast, the peaks persist as the system size increases. Therefore, we have numerically shown in this paper that primes are characterized by a substantial amount of order, especially relative to an uncorrelated lattice gas that does not have any such peaks. We also show that peaks at $k = m\pi/n$ are sharper when L is divisible by n. Our numerical results definitively suggested an explicit formula for the peak locations and heights, which predicts dense Bragg peaks, as occurs in quasicrystals (e.g. Fibonacci chain [51] as well as other one-dimensional examples [53]), but unlike the latter, the peak locations occur at rational multiples of π. In an upcoming paper [55], we will show that the primes in the intervals studied here are similar to 'limit periodic' systems [50] but still distinct from them. Limit-periodic systems are unions of an infinite number of periodic systems with rational periods, and possess dense Bragg peaks with rational ratios between their locations. Nevertheless, primes are certainly not unions of periodic systems, since they possess a density gradient, and large primes are difficult to predict [2].

In [55], we will use the well-known Dirichlet's theorem on arithmetic progressions to show that primes in the intervals studied here are limit-periodic in a probabilistic sense. Based on this theorem, we will provide a number-theoretic explanation for the numerical observations here; specifically, the peak location and height formula, the decrease of peak widths, and how the diffuse part vanishes as L increases. Moreover, we will show there that if the interval of the primes is chosen to be appreciably smaller or larger than the one used here, the primes would not be characterized by a large degree of order, as measured by the τ order metric [40].

The numerical techniques that we employed here to investigate the structure factor of primes in finite intervals should be applicable to characterize other complex point configurations in which the shape and heights of peaks depend sensitively on the system size. These examples include quasicrystals [51] and limit-periodic systems [50]. For example, to study the shape and height of a peak located at $k = 2\pi/\alpha$, the best choice for the system size is a multiple of α. Moreover, the structure factor of any finite system with dominant peak structure will always numerically possess a 'noisy' diffuse contribution. A simple way to characterize the diffuse part is to calculate the median value of $S(k)$ for all allowed k values and then study its behavior as a function of the system size to see if it becomes negligible in the infinite-size limit.

Acknowledgments

We are very grateful to Matthew de Courcy-Ireland for helpful discussions. This work was supported in part by the National Science Foundation under Award No DMR-1714722.

References

[1] Horsley S 1772 Being an account of his method of finding all the prime numbers *Phil. Trans.* **62** 327–47

[2] Mersene Research, Inc. www.mersenne.org/report_exponent/?exp_lo=77232917

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

[4] Granville A 1995 Harald Cramér and the distribution of prime numbers *Scand. Actuar. J.* **1995** 12–28

[5] Gallagher P X 1976 On the distribution of primes in short intervals *Mathematika* **23** 4–9

[6] Montgomery H L 1973 The pair correlation of zeros of the zeta function *Proc. Symp. in Pure Mathematics* vol 24 pp 181–93

[7] Miller G 1976 Riemann's hypothesis and tests for primality *J. Comput. Syst. Sci.* **13** 300–17

[8] Rabin M 1980 Probabilistic algorithm for testing primality *J. Number Theory* **12** 128–38

[9] Pomerance C, Selfridge J L and Wagstaff S S 1980 The pseudoprimes to $25 \cdot 10^9$ *Math. Comput.* **35** 1003–26
[10] Baillie R and Wagstaff S S 1980 Lucas pseudoprimes *Math. Comput.* **35** 1391–417
[11] Atkin A O L and Morain F 1993 Elliptic curves and primality proving *Math. Comput.* **61** 29–68
[12] Granville A and Martin G 2006 Prime number races *Am. Math. Mon.* **113** 1
[13] Rubinstein M and Sarnak P 1994 Chebyshev's bias *Exp. Math.* **3** 173–97
[14] Martelli F 2013 Dealing with primes I.: on the Goldbach conjecture (arXiv:1309.5895 [math.NT])
[15] Vinogradov I M 1954 *The Method of Trigonometrical Sums in the Theory of Numbers* (New York: Interscience)
[16] Green B and Tao T 2008 The primes contain arbitrarily long arithmetic progressions *Ann. Math.* **167** 481–547
[17] Green B and Tao T 2010 Linear equations in primes *Ann. Math.* **171** 1753–850
[18] Green B and Tao T 2012 The Mobius function is strongly orthogonal to nilsequences *Ann. Math.* **175** 541–66
[19] Tao T and Ziegler T 2012 The inverse conjecture for the Gowers norm over finite fields in low characteristic *Ann. Comb.* **16** 121–88
[20] Zhang Y 2014 Bounded gaps between primes *Ann. Math.* **3** 1121–74
[21] Lemke Oliver R J and Soundararajan K 2016 Unexpected biases in the distribution of consecutive primes (arXiv:1603.03720 [math.NT])
[22] Maynard J 2015 Small gaps between primes *Ann. Math.* **181** 383–413
[23] Dahmen S R, Prado S D and Stuermer-Daitx T 2001 Similarity in the statistics of prime numbers *Physica* A **296** 523–8
[24] Wolflf M 1996 Unexpected regularities in the distribution of prime numbers *Proc. 8th Joint EPS—APS Int. Conf. Physics Computing*
[25] Kumar P, Ivanov P C and Stanley H E 2003 Information entropy and correlations in prime numbers (arXiv:cond-mat/0303110 [cond-mat.stat-mech])
[26] 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
[27] Torquato S and Stillinger F H 2003 Local density fluctuations, hyperuniformity, and order metrics *Phys. Rev.* E **68** 041113
[28] Donev A, Stillinger F H and Torquato S 2005 Unexpected density fluctuations in jammed disordered sphere packings *Phys. Rev. Lett.* **95** 090604
[29] 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
[30] Zachary C E and Torquato S 2009 Hyperuniformity in point patterns and two-phase random heterogeneous media *J. Stat. Mech.* P12015
[31] 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
[32] Kurita R and Weeks E R 2011 Incompressibility of polydisperse random-close-packed colloidal particles *Phys. Rev.* E **84** 030401
[33] Xie R, Long G G, Weigand S J, Moss S C, Carvalho T, Roorda S, Hejna M, Torquato S and Steinhardt P J 2013 Hyperuniformity in amorphous silicon based on the measurement of the infinite-wavelength limit of the structure factor *Proc. Natl Acad. Sci.* **110** 13250–4
[34] 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
[35] Lesanovsky I and Garrahan J P 2014 Out-of-equilibrium structures in strongly interacting Rydberg gases with dissipation *Phys. Rev.* A **90** 011603
[36] Muller N, Haberko J, Marichy C and Scheffold F 2014 Silicon hyperuniform disordered photonic materials with a pronounced gap in the shortwave infrared *Adv. Opt. Mater.* **2** 115–9
[37] Hexner D and Levine D 2015 Hyperuniformity of critical absorbing states *Phys. Rev. Lett.* **114** 110602
[38] 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
[39] De Rosa C, Auriemma F, Diletto C, Di Girolamo R, Malafronte A, Morvillo P, Zito G, Rusciano G, Pesce G and Sasso A 2015 Toward hyperuniform disordered plasmonic nanostructures for reproducible surface-enhanced raman spectroscopy *Phys. Chem. Chem. Phys.* **17** 8061–9

[40] Torquato S, Zhang G and Stillinger F H 2015 Ensemble theory for stealthy hyperuniform disordered ground states *Phys. Rev.* X **5** 021020
[41] Weijs J H, Jeanneret R, Dreyfus R and Bartolo D 2015 Emergent hyperuniformity in periodically driven emulsions *Phys. Rev. Lett.* **115** 108301
[42] Torquato S 2016 Hyperuniformity and its generalizations *Phys. Rev.* E **94** 022122
[43] Leseur O, Pierrat R and Carminati R 2016 High-density hyperuniform materials can be transparent *Optica* **3** 763
[44] Xu W-S, Douglas J F and Freed K F 2016 Influence of cohesive energy on the thermodynamic properties of a model glass-forming polymer melt *Macromolecules* **49** 8341–54
[45] Ma Z and Torquato S 2017 Random scalar fields and hyperuniformity *J. Appl. Phys.* **121** 244904
[46] Davenport H and Montgomery H L 1980 *Multiplicative Number Theory* vol 74 (New York: Springer)
[47] Tenenbaum G 1995 *Introduction to Analytic and Probabilistic Number Theory* vol 46 (Cambridge: Cambridge University Press)
[48] Iwaniec H and Kowalski E 2004 *Analytic Number Theory* vol 53 (Providence, RI: American Mathematical Society)
[49] Baake M, Moody R V and Pleasants P A B 2012 Diffraction from visible lattice points and *k*th power free integers *Discrete Math.* **221** 3
[50] Baake M and Grimm U 2011 Diffraction of limit periodic point sets *Phil. Mag.* **91** 2661–70
[51] Levine D and Steinhardt P J 1984 Quasicrystals: a new class of ordered structures *Phys. Rev. Lett.* **53** 2477
[52] Socolar J E S and Steinhardt P J 1986 Quasicrystals. II. Unit-cell configurations *Phys. Rev.* B **34** 617
[53] Oğuz E C, Socolar J E S, Steinhardt P J and Torquato S 2017 Hyperuniformity of quasicrystals *Phys. Rev.* B **95** 054119
[54] Vericat F 2013 A lattice gas of prime numbers and the Riemann hypothesis *Physica* A **392** 4516
[55] Torquato S, Zhang G and de Courcy-Ireland M Hidden multiscale order in the primes unpublished
[56] Primos 'First 2 billion prime numbers' http://www.primos.mat.br/2T_en.html
[57] Source Forge 'Kiss FFT' https://sourceforge.net/projects/kissfft/