Statistical properties of the set of sites visited by the two-dimensional random walk

1997 J. Phys. A: Math. Gen. 30 507

(http://iopscience.iop.org/0305-4470/30/2/017)

J. Phys. A: Math. Gen. **30** (1997) 507–531. Printed in the UK PII: S0305-4470(97)77004-5

Statistical properties of the set of sites visited by the two-dimensional random walk

F van Wijland, S Caser and H J Hilhorst

Laboratoire de Physique Théorique et Hautes Energies†, Bâtiment 211, Université de Paris-Sud, 91405 Orsay cedex, France

Received 1 August 1996

Abstract. We study the support (i.e. *the set of visited sites*) of a t-step random walk on a two-dimensional square lattice in the large t limit. A broad class of global properties, $M(t)$, of the support is considered, including for example the number, $S(t)$, of its sites; the length of its boundary; the number of islands of unvisited sites that it encloses; the number of such islands of given shape, size, and orientation; and the number of occurrences in space of specific local patterns of visited and unvisited sites. On a finite lattice we determine the scaling functions that describe the averages, $\overline{M}(t)$, on appropriate lattice size-dependent time scales. On an infinite lattice we first observe that the $\overline{M}(t)$ all increase with t as $\sim t/\log^k t$, where k is an M-dependent positive integer. We then consider the class of random processes constituted by the *fluctuations around average* $\Delta M(t)$. We show that, to leading order as t gets large, these fluctuations are all proportional to a *single universal random process*, $\eta(t)$, normalized to $\overline{\eta^2}(t) = 1$. For $t \to \infty$ the probability law of $\eta(t)$ tends to that of Varadhan's *renormalized local time of self-intersections*. An implication is that in the long time limit all $\Delta M(t)$ are proportional to $\Delta S(t)$.

1. Introduction and summary

The reason for the multiple connections between the random walk and many questions of current and of permanent interest in science is evidently mathematical. The random walk models the action of the Laplace operator, which is common to all those problems. Indeed, important early results on the random walk are due to mathematicians. Among them, the famous Polya theorem [1] asserts that in spatial dimensions $d \leqslant 2$ a random walk is certain to return to its initial position, whereas in $d > 2$ it will escape to infinity. The random walk in $d = 2$ is, therefore, at its critical dimension for a return to the origin. Many features associated with this fact make the random walk in two dimensions particularly interesting. For example, the probability distribution of the time interval, τ_0, between two successive visits of the walk to its initial position decays for large intervals as $\sim 1/\tau_0 \log^2 \tau_0$, so that very long excursions occur away from the point of departure. Excellent recent monographs on random-walk theory have been written by Weiss [2] and by Hughes [3].

In this work we consider the *simple random walk* on a two-dimensional square lattice. The walk starts at time $t = 0$ at the origin $\boldsymbol{x} = \mathbf{0}$ and steps at $t = 1, 2, 3, \ldots$ with equal probability to one of the four neighbouring lattice sites. Our investigation focuses on the statistical properties of the *support* of the walk at time t, i.e. of the set of sites that have been visited during the first t steps.

† Laboratoire associé au Centre National de la Recherche Scientifique–URA D0063.

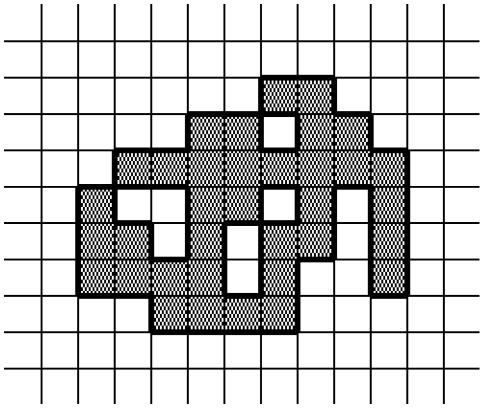


Figure 1. A random walker makes t random steps between the centres of neighbouring squares. The squares visited are coloured black and constitute the *support* of the walk. In the example of this figure the support encloses four islands of unvisited sites. The boundary of the support is represented by a heavy line.

At any given time t the support can be visualized as a set of black sites (the visited ones) in a lattice of otherwise white (unvisited) sites, as shown in figure 1. The set of unvisited sites is divided into components that are disjoint (i.e. not connected via any nearest neighbour link) and that we call *islands*. In the course of time existing islands will be reduced in size and single-site islands will eventually be destroyed; the number of islands increases each time that a step of the walk cuts an existing island, or the outer region surrounding the support, into disjoint components.

Questions about the statistical properties of the islands are natural and our interest in them was raised by a simulation study by Coutinho *et al* [4]. These authors investigated the evolution of the number of islands, $I(t)$, and several of their properties on *finite* lattices of up to 1200^2 sites. In a short report [5] we showed that an analytic calculation is possible, both on finite and infinite lattices, for *some* (not all) of the quantities considered by Coutinho *et al* [4], and that good agreement between theory and computer simulation is obtained. In this work we present a full account of most of the results announced in [5] and consider a great many related questions. We are, in particular, led to consider the infinite lattice again.

It appears that the number of islands, $I(t)$, is but one member of a much wider class of observables, generically to be denoted as $M(t)$, with closely related properties. This class includes the total number, $S(t)$, of sites in the support as well as the total length, $E(t)$, of the boundary of the support (on an infinite lattice the boundary length is the sum of the external perimeter and the perimeters of the islands enclosed).

The main variable that characterizes the support is its total number of sites, $S(t)$, sometimes called its *range*, which was first studied by Dvoretzky and Erdös [6]. In the limit of large times, t, it has the average value [6–9]

$$\overline{S}(t) \simeq \frac{\pi t}{\log 8t} \tag{1.1}$$

and the root mean square deviation [10, 11]

$$\overline{\Delta S^2(t)}^{1/2} \simeq \mathcal{A}\frac{\pi t}{\log^2 8t} \tag{1.2}$$

where $\mathcal{A} = 1.303\ldots$ and where we write $\Delta S \equiv S - \overline{S}$. The typical support is known to be far from spherical, and its principal moments of inertia and asphericity have been studied

[12, 13]. Expressions also exist [14] for its *span*, that is, the smallest rectangular box that it fits in.

The question of calculating the average $\overline{S}(t)$ becomes, in general dimension and in an appropriately taken continuum limit, the celebrated *Wiener sausage* problem: what is the volume swept out in time t by a d-dimensional Brownian sphere of finite radius? The Wiener sausage appears naturally in certain applications of the random walk such as, for example, the study by Kac and Luttinger [15] of the Bose–Einstein condensation in the presence of impurities. Historically this continuum problem precedes its lattice counterpart. It was considered as early as 1933 by Leontovitsh and Kolmogorov [16] and is still today an active subject of investigation in probability theory ([17]; see [18] for a recent overwiew) and in mathematical physics [19].

The islands in the support of a lattice random walk have their counterpart in the connected components into which a two-dimensional Brownian motion path divides the plane. These have been studied by Mountford [20] and Le Gall [17], and, most recently, by Werner [21], who determines, among other things, how many there are larger than a given size, ε, in the limit of $\varepsilon \to 0$. The outer boundary of the Brownian motion path has been considered very recently by Lawler [22].

We characterize as follows the class of observables, $M(t)$, to be studied below. For each lattice site, $\boldsymbol{x}$, we introduce an *occupation number*, $m(\boldsymbol{x}, t)$, equal to 0 if site $\boldsymbol{x}$ is visited by the walker before or at time t, and equal to 1 if it is not. Now let A_1 and A_2 be disjoint finite sets of lattice vectors. Then the product

$$\prod_{\boldsymbol{a}_1 \in A_1} m(\boldsymbol{x} + \boldsymbol{a}_1, t) \prod_{\boldsymbol{a}_2 \in A_2} [1 - m(\boldsymbol{x} + \boldsymbol{a}_2, t)] \tag{1.3}$$

codes for a specific spatial *pattern* $\alpha \equiv (A_1, A_2)$ of white (unvisited) and black (visited) sites. When $\boldsymbol{x}$ runs through the lattice, the product (1.3) equals 1 when the pattern, α, is encountered and 0 otherwise; hence this product summed on all $\boldsymbol{x}$ represents the total number $N_\alpha(t)$ of occurrences in space of the pattern α. The observables that this work deals with are these *pattern numbers*, $N_\alpha(t)$, and their linear combinations, $M(t)$. It is easy to see that suitably chosen $M(t)$ may represent, e.g. the total number, $S(t)$, of sites in the support; the total boundary length, $E(t)$, of the support; the total number, $I(t)$, of islands of unvisited sites enclosed by the support; or the total number, $I_\beta(t)$, of islands of a given type β (where *type* stands for shape, size, and orientation). Our main results concerning the observables, $M(t)$, are of two kinds.

(i) *Average behaviour on a finite lattice.* On a finite lattice of N sites the averages, $\overline{M}(t)$, approach their limiting values on the time scale $t \sim N \log^2 N$, whereas their main variation occurs on the earlier time scale $t \sim N \log N$. We calculate the scaling functions that describe the time dependence on both time scales. Several of the observables, $M(t)$, of interest just mentioned are treated as examples. A comparison is made, where possible, with the simulations by Coutinho *et al* [4]

(ii) *Average behaviour and fluctuations on the infinite lattice.* On an infinite lattice the fluctuating properties of the support manifest a universality that can be described as follows. Let M and M' be two linear combinations of pattern numbers. Then, with the same notation as before for deviations from average, we show by explicit calculation that asymptotically for $t \to \infty$

$$\overline{M}(t) \simeq m_k \frac{\pi^{k+1} t}{\log^{k+1} 8t} \qquad \overline{M'}(t) \simeq m'_{k'} \frac{\pi^{k'+1} t}{\log^{k'+1} 8t} \tag{1.4}$$

$$\overline{\Delta M(t) \Delta M'(t)} \simeq \mathcal{A}^2 (k+1)(k'+1) m_k m'_{k'} \frac{\pi^{k+k'+2} t^2}{\log^{k+k'+4} 8t}. \tag{1.5}$$

Here k and k' are non-negative integers that depend on M and M', respectively, and are called their *order*; m_k and $m'_{k'}$ are known proportionality constants; and the same number, $\mathcal{A}$, appears that was first encountered in the study of the variance of $S(t)$. The above equations imply that the normalized deviations from average

$$\eta_M(t) = \frac{\log 8t}{(k+1)\mathcal{A}} \frac{\Delta M(t)}{\overline{M}(t)} \tag{1.6}$$

have a correlation matrix, $\overline{\eta_M(t)\eta_{M'}(t)}$, whose elements all equal to unity. This can be true only if all η_M are equal to a *single random variable* to be called $\eta(t)$. As a consequence the random variables $\Delta M(t)$ are, to leading order as $t \to \infty$, all proportional to $\eta(t)$. Explicitly,

$$\Delta M(t) \simeq (k+1)m_k \frac{\pi^{k+1}t}{\log^{k+2} 8t} \mathcal{A}\eta(t) \tag{1.7}$$

where $\eta(t)$ is *universal* (independent of M). Hence what seemed to be a large number of independent fluctuating degrees of freedom of the support is hereby reduced, to leading order as $t \to \infty$, to only a single degree of freedom.

We now briefly summarize the contents of the successive sections. Since all quantities of interest are, in the end, expressed in terms of the random walk Green's function, we collect in section 2 the basic knowledge required about this function. In section 3 we express the total boundary length, $E(t)$, and the total number of islands, $I(t)$, in terms of the occupation numbers and discuss the wider class of observables, $M(t)$, of which they are examples. In sections 4.1 and 4.2 we calculate the time-dependent averages $\overline{M}(t)$ of such observables. In sections 4.3–4.6 the general expression for the result, equation (4.16), is analysed for large times and explicitly worked out for several examples, both on the infinite and the finite lattice. In sections 5.1 and 5.2 we show, for the infinite lattice, how to calculate the correlation between two observables $M(t)$ and $M'(t)$. In section 5.3 we arrive at the final results concerning the random variable $\eta(t)$. In the discussion in section 6 we present various comments, compare where possible our lattice results to their continuum analogues, and speculate about further connections between several quantities.

2. The random walk Green's function

2.1. Definition

A random walker starts at time $t = 0$ at the origin $\boldsymbol{x} = \mathbf{0}$ of a square lattice and steps at each instant of time $t = 1, 2, 3, \ldots$ with probability $\frac{1}{4}$ to one of its four nearest-neighbour sites. The Green's function $G(\boldsymbol{x}, t)$ denotes the probability that at time t the walker is at lattice site $\boldsymbol{x}$, and

$$\hat{G}(\boldsymbol{x}, z) = \sum_{t=0}^{\infty} z^t G(\boldsymbol{x}, t) \tag{2.1}$$

its generating function. In this section we collect, in concise form, those properties of the generating function that will be needed later. An elementary calculation gives, for a finite periodic lattice of $L \times L = N$ sites,

$$\hat{G}(\boldsymbol{x}, z) = \frac{1}{N} \sum_{\boldsymbol{q}} \frac{\mathrm{e}^{-\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}}}{1 - \frac{1}{2}z(\cos q_1 + \cos q_2)} \tag{2.2}$$

where $\boldsymbol{q} = (q_1, q_2) = 2\pi(\kappa_1, \kappa_2)/L$ with the κ_i running through the values $0, 1, 2, \ldots, L-1$. In the limit of an infinite lattice expression (2.2) becomes

$$\hat{G}(\boldsymbol{x}, z) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \mathrm{d}q_1 \int_{-\pi}^{\pi} \mathrm{d}q_2 \, \frac{\mathrm{e}^{-\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x}}}{1 - \frac{1}{2}z(\cos q_1 + \cos q_2)}. \tag{2.3}$$

At each point in this work it will be clear whether we are discussing the finite or the infinite lattice; in some cases we shall denote corresponding quantities in the two geometries by the same symbol, as for example in equations (2.2) and (2.3), and not explicitly indicate their N dependence on a finite lattice.

2.2. *Expansion near $z = 1$*

The long-time behaviour of the physical quantities of interest is determined by the behaviour of $\hat{G}(\boldsymbol{x}, z)$ in the complex plane near $z = 1$. Expressions (2.2) and (2.3) both have the property that for $z \to 1$ the function $\hat{G}(\boldsymbol{x}, z)$ diverges. In order to study this divergence it is convenient to write

$$\hat{G}(\boldsymbol{x}, z) = \hat{G}(\mathbf{0}, z) - g(\boldsymbol{x}, z) \tag{2.4}$$

where on the right-hand side the term $g(\boldsymbol{x}, z)$ contains all the $\boldsymbol{x}$ dependence and remains finite for $z \to 1$. We shall discuss $\hat{G}(\mathbf{0}, z)$ and $g(\boldsymbol{x}, z)$ separately.

2.2.1. The function $\hat{G}(\mathbf{0}, z)$ *Finite lattice.* Expression (2.2) has a simple pole as a function of z whenever one of the denominators inside the sum on $\boldsymbol{q}$ vanishes. This leads to a sequence of poles on the real axis for $z \geqslant 1$, of which the first one is located exactly at $z = 1$. The interval between two successive poles is of $\mathcal{O}(N^{-1})$ and contains a zero. Upon setting $\boldsymbol{x} = \mathbf{0}$ in equation (2.2) and expanding each term for small $1 - z$ one gets [23]

$$\hat{G}(\mathbf{0}, z) = \frac{1}{N(1-z)} + a(N) - a_1(N)(1-z) + \mathcal{O}((1-z)^2) \tag{2.5}$$

in which the coefficients are functions of N that in the limit $N \to \infty$ behave as [23]

$$\begin{aligned} a(N) &= \frac{1}{\pi} \log cN + \mathcal{O}(N^{-1}) \\ a_1(N) &= c_1 N + \mathcal{O}(\log N) \end{aligned} \tag{2.6}$$

with $c = 1.8456\ldots$ and $c_1 = 0.061\,87\ldots$

The expansion in equation (2.5) represents well the behaviour of $\hat{G}(\mathbf{0}, z)$ near the pole at $z = 1$, but is certainly not valid on approach of the next pole. It can be used, however, to determine the location of the zero $z = z_0$ between the first two poles. Upon solving equation (2.5) in successive orders for z_0 one finds

$$z_0 = 1 + \frac{1}{Na(N)} \left[1 - \frac{c_1}{a^2(N)} + \cdots \right]. \tag{2.7}$$

From equations (2.7) and (2.6) we conclude that when $N \to \infty$ this zero is separated from the pole at $z = 1$ by a distance only of $\mathcal{O}(1/N \log N)$. All other zeros of $\hat{G}$ are separated from $z = 1$ by a distance of at least $\mathcal{O}(N^{-1})$. Due to its exceptional proximity to $z = 1$ the zero z_0 plays a special role in the long-time behaviour of the random walk on finite lattices. This fact was first noted by Weiss *et al* [24] and has also been exploited [25] in the study of the covering time of a finite lattice by a random walk.

Infinite lattice. When $N \to \infty$ the poles of $\hat{G}$ become dense to a branch cut and the expansion near $z = 1$ is [26]

$$\hat{G}(\mathbf{0}, z) = \frac{1}{\pi} \log \frac{8}{1 - z} + \mathcal{O}((1 - z) \log(1 - z)). \tag{2.8}$$

Later on in this work, for brevity, we shall also denote $\hat{G}(\mathbf{0}, z)$ as $G_0(z)$.

2.2.2. The function $g(\boldsymbol{x}, z)$ Expanding $g(\boldsymbol{x}, z)$ for finite N around $z = 1$ gives

$$g(\boldsymbol{x}, z) = g_N(\boldsymbol{x}) + g'_N(\boldsymbol{x})(1 - z) + \mathcal{O}((1 - z)^2) \tag{2.9}$$

where we now explicitly indicate that the expansion coefficients are N dependent. We shall set

$$g(\boldsymbol{x}) = \lim_{N \to \infty} g_N(\boldsymbol{x}). \tag{2.10}$$

Spitzer [27] shows how to calculate the $g(\boldsymbol{x})$ for $\boldsymbol{x}$ close to the origin. Letting $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ denote the unit vectors we have the following values:

$$\begin{aligned} &g(\mathbf{0}) = 0 \qquad g(2\boldsymbol{e}_1) = 4 - 8/\pi \\ &g(\boldsymbol{e}_1) = 1 \qquad g(2\boldsymbol{e}_1 + \boldsymbol{e}_2) = 8/\pi - 1 \\ &g(\boldsymbol{e}_1 + \boldsymbol{e}_2) = 4/\pi. \end{aligned} \tag{2.11}$$

When combining preceding results we see that for finite N the quantity $\hat{G}(\boldsymbol{x}, z)$ has the expansion

$$\hat{G}(\boldsymbol{x}, z) \simeq \frac{1}{N(1 - z)} + [a(N) - g_N(\boldsymbol{x})] + \mathcal{O}(1 - z) \tag{2.12}$$

whose second term behaves for large N as

$$a(N) - g_N(\boldsymbol{x}) \simeq \frac{1}{\pi} \log cN - g(\boldsymbol{x}) + \cdots \tag{2.13}$$

with the dots representing terms that vanish as $N \to \infty$.

2.3. Scaling limit

In our study of the fluctuations in section 5 we shall also need $\hat{G}(\boldsymbol{x}, z)$ in the scaling limit $z \to 1$, $x \to \infty$ with $x^2(1 - z)$ fixed. The behaviour in this limit is [9]

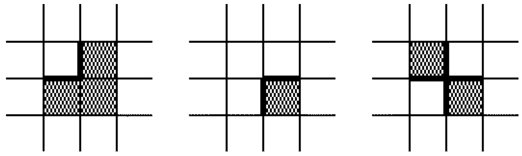
$$\hat{G}(\boldsymbol{x}, z) \simeq \frac{2}{\pi} K_0(2x(1 - z)^{1/2}) \tag{2.14}$$

where K_0 is the modified Bessel function of order zero. As is well known, for $z \to 1$ the dominant contribution to the sum (2.1) comes from values of t that are of $\mathcal{O}((1 - z)^{-1})$, and therefore the scaling limit corresponds to focusing on distances x of $\mathcal{O}(\sqrt{t})$.

3. Islands and other observables

3.1. Islands

If starting at an arbitrary element, one follows the boundary of the support such that the white sites are on the left and the black ones on the right, then one will return to the point of departure after having turned either through an angle 2π (if that part of the boundary encloses an island in the support) or through an angle -2π (if it is the outer boundary of



(*a*) (*b*) (*c*)

Figure 2. Let the boundary of the support be oriented such that as one proceeds along it the white sites are on the left and the black ones on the right. Then in (*a*) the boundary turns through an angle $+\frac{1}{2}\pi$ and in (*b*) through an angle $-\frac{1}{2}\pi$. In (*c*) the black squares are (by convention) considered to separate the white squares from one another and the part of the boundary shown makes two turns through $+\frac{1}{2}\pi$.

the support). The number of islands is therefore obtained from the number of turns in the boundary, by adding those of figure 2(*a*) with weight $+\frac{1}{4}$ and those of figure 2(*b*) with weight $-\frac{1}{4}$. The diagram of figure 2(*c*) corresponds to two turns of $\pi/2$ and therefore counts with weight factor $\frac{1}{2}$. This procedure counts the outer boundary with weight -1 and therefore 2 has to be added to obtain the final result. It is now easy to construct the expression for the total number of islands in terms of the occupation numbers $m(\boldsymbol{x}, t)$. We have to shift a 2×2 window across the lattice, check all 2×2 local site configurations and add all those that are of the types of figure 2 with the proper weights. Let $\boldsymbol{x}$ denote the lower left-hand site in the 2×2 window. Then, for example, the expression

$$m(\boldsymbol{x}, t)[1 - m(\boldsymbol{x} + \boldsymbol{e}_1, t)]\,[1 - m(\boldsymbol{x} + \boldsymbol{e}_2, t)]\,m(\boldsymbol{x} + \boldsymbol{e}_1 + \boldsymbol{e}_2, t) \tag{3.1}$$

(which is of type (1.3)) equals 1 or 0 according to whether the local 2×2 configuration is or is not equal to the diagram of figure 2(*c*). Writing down analogueous expressions for all other diagrams, summing these, and subsequently summing them on $\boldsymbol{x}$ yields $I(t)$ as a linear combination of pattern numbers. After rearranging terms one finds

$$\begin{aligned} I(t) = 2 + \sum_{\boldsymbol{x}} [&m(\boldsymbol{x}, t) - m(\boldsymbol{x}, t)m(\boldsymbol{x} + \boldsymbol{e}_1, t) - m(\boldsymbol{x}, t)m(\boldsymbol{x} + \boldsymbol{e}_2, t) \\ &+ m(\boldsymbol{x}, t)m(\boldsymbol{x} + \boldsymbol{e}_1, t)m(\boldsymbol{x} + \boldsymbol{e}_2, t)m(\boldsymbol{x} + \boldsymbol{e}_1 + \boldsymbol{e}_2, t)]. \end{aligned} \tag{3.2}$$

This expression is at the basis of all calculations concerning islands on the infinite lattice. Finite lattice calculations require a further remark, which is made below with equation (4.44).

3.2. *Other observables*

We are now interested in other variables whose values can be obtained by the 'window' method. To see the general form of these variables, let A be a finite subset of lattice vectors. The set $\{\boldsymbol{x} + \boldsymbol{a} | \boldsymbol{a} \in A\}$, obtained by translating A by a vector $\boldsymbol{x}$, will be written as $\boldsymbol{x} + A$. The variable

$$m_{\boldsymbol{x}+A}(t) = \prod_{\boldsymbol{a} \in A} m(\boldsymbol{x} + \boldsymbol{a}, t) \tag{3.3}$$

is equal to unity if at time t all the sites of this set are white (unvisited), and is zero otherwise. The sum variable

$$M_A(t) = \sum_{\boldsymbol{x}} m_{\boldsymbol{x}+A}(t) \tag{3.4}$$

counts the total number of wholly white sets in the lattice that can be obtained from A by a translation. In the remainder we shall also wish to take for A the empty set $\emptyset$. In that case

Table 1. The coefficients μ_A for the four observables S, E, I and I_1 defined in the text; entries not shown are zero. Symmetry under rotations over $\pi/2$ has been exploited to reduce the length of the table. The coefficients in each column add up to zero. The parameter g_A is undefined for the empty set $\emptyset$. A is defined only up to a translation.

	μ_A				
A	S	E	I	I_1	g_A
$\emptyset$	1				–
$\{\mathbf{0}\}$	-1	4	1	1	0
$\{\mathbf{0}, \boldsymbol{e}_1\}$		-4	-2	-4	$1/2$
$\{\mathbf{0}, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_1+\boldsymbol{e}_2\}$			1		$(\pi+2)/2\pi$
$\{\mathbf{0}, \boldsymbol{e}_1, \boldsymbol{e}_2\}$				4	$\pi/2(\pi-1)$
$\{\mathbf{0}, \boldsymbol{e}_1, 2\boldsymbol{e}_1\}$				2	$\pi/4$
$\{\mathbf{0}, \boldsymbol{e}_1, -\boldsymbol{e}_1, \boldsymbol{e}_2\}$				-4	$\pi(\pi-6)/2(\pi^2-6\pi+4)$
$\{\boldsymbol{e}_1, \boldsymbol{e}_2, -\boldsymbol{e}_1, -\boldsymbol{e}_2\}$				1	1

the right-hand side of equation (3.3) should be assigned the value unity and equation (3.4) shows that $M_\emptyset(t)$ is the total number of lattice sites.

In this work we consider the class of observables

$$M(t) = \sum_A \mu_A M_A(t) \tag{3.5}$$

with a finite number of nonvanishing coefficients, μ_A. Whereas an $M(t)$ is not in general a unique linear combination of pattern numbers, the representation (3.5) is unique. The number of islands, $I(t)$, for example, was initially found in section 3.1 as a linear combination of pattern numbers, and, after a rearrangement of terms, became equal to expression (3.2), which has the form (3.5). Various other quantities of potential interest can be expressed this way. Some of these, with their coefficients μ_A, have been listed in table 1. The best known example is the total number, $S(t)$, of sites in the support,

$$S(t) = \sum_{\boldsymbol{x}} [1 - m(\boldsymbol{x}, t)] \tag{3.6}$$

which has $\mu_A = \pm 1$ for $A = \emptyset$ and $A = \{\mathbf{0}\}$, respectively, and $\mu_A = 0$ otherwise. Another example is the total boundary length $E(t)$ between the visited and unvisited lattice sites (that is, the total number of pairs of neighbouring sites of which one is white and one black). It can be expressed as

$$\begin{aligned} E(t) = \sum_{\boldsymbol{x}} [& m(\boldsymbol{x}, t)(1 - m(\boldsymbol{x}+\boldsymbol{e}_1, t) + (1 - m(\boldsymbol{x}, t))m(\boldsymbol{x}+\boldsymbol{e}_1, t) \\ & + m(\boldsymbol{x}, t)(1 - m(\boldsymbol{x}+\boldsymbol{e}_2, t) + (1 - m(\boldsymbol{x}, t))m(\boldsymbol{x}+\boldsymbol{e}_2, t)] \\ = \sum_{\boldsymbol{x}} [& 4m(\boldsymbol{x}, t) - 2m(\boldsymbol{x}, t)m(\boldsymbol{x}+\boldsymbol{e}_1, t) - 2m(\boldsymbol{x}, t)m(\boldsymbol{x}+\boldsymbol{e}_2, t)]. \end{aligned} \tag{3.7}$$

We note that one has the relation

$$\sum_A \mu_A = 0 \tag{3.8}$$

for the three observables S, E, and I, but that

$$\sum_{A \neq \emptyset} \mu_A = 0 \tag{3.9}$$

only for E and I, but not for S. Property (3.8) is required if the sum on $\boldsymbol{x}$ in equation (3.4) is to have a finite limit when the lattice size, N, tends to infinity. The property (3.9) is easily traced back to the fact that for E and I the window selects patterns consisting of both white and black sites, whereas for S it selects only black sites.

4. Averages of observables

4.1. Relation to first passage time probabilities

We shall now calculate in a unified way averages and fluctuations of quantities $M(t)$ of type (3.5). The approach of this subsection will serve as the basis for the developments to follow. Using equations (3.3)–(3.5) we see that the averages $\overline{M}(t)$ are linear combinations of expressions of the type

$$\begin{aligned}\overline{m_{\boldsymbol{x}+A}(t)} &= \overline{\prod_{\boldsymbol{a}\in A} m(\boldsymbol{x}+\boldsymbol{a},t)} \\ &= 1-\sum_{\tau=0}^{t} f_{\boldsymbol{x}+A}(\tau) \\ &= 1-\sum_{\tau=0}^{t}\sum_{\boldsymbol{a}\in A} f_{\boldsymbol{x}+A}(\boldsymbol{x}+\boldsymbol{a},\tau)\end{aligned} \tag{4.1}$$

in which the last two transformations make sense only when $A \neq \emptyset$; $f_{\boldsymbol{x}+A}(\tau)$ is the probability that the walker's first visit to any site of the set $\boldsymbol{x}+A$ takes place at time τ; and $f_{\boldsymbol{x}+A}(\boldsymbol{x}+\boldsymbol{a},\tau)$ is the probability that it takes place at time τ *and* that it concerns the specific site $\boldsymbol{x}+\boldsymbol{a}$. Upon averaging equation (3.4), using equation (4.1), and passing to generating functions we find, for all nonempty A,

$$\hat{\overline{M}}_A(z) = -\frac{1}{1-z}\sum_{\boldsymbol{x}}\Big[\sum_{\boldsymbol{a}\in A}\hat{f}_{\boldsymbol{x}+A}(\boldsymbol{x}+\boldsymbol{a},z)-1\Big]. \tag{4.2}$$

For the empty set one derives directly that

$$\hat{\overline{M}}_{\emptyset}(z) = N/(1-z). \tag{4.3}$$

We now sum the $\hat{\overline{M}}_A(z)$ given by equations (4.2) and (4.3) on all A with coefficients μ_A. Using equation (3.8) and introducing for all $A \neq \emptyset$

$$\hat{F}_A(\boldsymbol{a},z) = \sum_{\boldsymbol{x}}\hat{f}_{\boldsymbol{x}+A}(\boldsymbol{x}+\boldsymbol{a},z) \tag{4.4}$$

yields

$$\hat{\overline{M}}(z) = -\frac{1}{1-z}\sum_{A\neq\emptyset}\mu_A\sum_{\boldsymbol{a}\in A}\hat{F}_A(\boldsymbol{a},z). \tag{4.5}$$

With this formula we have reduced the generating function of the average of interest, $\overline{M}(t)$, to the quantities $\hat{F}_A$ which are closely related to first passage times, but still unknown. We shall now proceed to determine the $\hat{F}_A$.

4.2. Solving the first passage time probabilities

Standard random walk theory [27, 2] relates the first passage probabilities $\hat{f}$ to the Green's function $\hat{G}$ by

$$\hat{G}(\boldsymbol{x}+\boldsymbol{a},z) = \sum_{\boldsymbol{a}'\in A}\hat{f}_{\boldsymbol{x}+A}(\boldsymbol{x}+\boldsymbol{a}',z)\hat{G}(\boldsymbol{a}-\boldsymbol{a}',z) \tag{4.6}$$

for all $\boldsymbol{a} \in A$. With the aid of equation (4.4) we find for $\hat{F}_A$ the equation

$$\sum_{\boldsymbol{a}' \in A} \hat{F}_A(\boldsymbol{a}', z)\hat{G}(\boldsymbol{a} - \boldsymbol{a}', z) = \frac{1}{1-z} \tag{4.7}$$

for all $\boldsymbol{a} \in A$. This is a matrix equation for the $\hat{F}_A$ whose dimension is the number $|A|$ of sites in the set A. This equation possesses special properties which are best exhibited by converting it to the shorthand notation

$$\begin{aligned} \gamma_{\boldsymbol{a}\boldsymbol{a}'}(z) &= g(\boldsymbol{a} - \boldsymbol{a}', z)/G_0(z) \\ \mathcal{F}_{\boldsymbol{a}} &= (1-z)G_0(z)\hat{F}_A(\boldsymbol{a}, z). \end{aligned} \tag{4.8}$$

Equation (4.7) then becomes

$$\sum_{\boldsymbol{a}' \in A} (1 - \gamma_{\boldsymbol{a}\boldsymbol{a}'}(z))\mathcal{F}_{\boldsymbol{a}'} = 1 \tag{4.9}$$

for all $\boldsymbol{a} \in A$. If we denote by $\gamma^{(A)}$ the matrix of elements $\gamma_{\boldsymbol{a},\boldsymbol{a}'}$ with $\boldsymbol{a}$, $\boldsymbol{a}' \in A$, then in matrix notation

$$(J - \gamma^{(A)}(z))\mathcal{F} = \boldsymbol{j} \tag{4.10}$$

where J and $\boldsymbol{j}$ are the matrix and vector, respectively, of dimension $|A|$, whose elements all equal 1. As shown in equation (4.5), we only need the sum of the components of $\mathcal{F}$. Formal inversion gives

$$\sum_{\boldsymbol{a} \in A} \mathcal{F}_{\boldsymbol{a}} = \sum_{\boldsymbol{a}, \boldsymbol{a}' \in A} [(J - \gamma^{(A)}(z))^{-1}]_{\boldsymbol{a}\boldsymbol{a}'}. \tag{4.11}$$

In a later stage we shall wish to take the limit $z \to 1$. In view of equation (4.8) and the known behaviour of $G_0(z)$ this implies that $\gamma^{(A)}(z) \to 0$, so that, except when $|A| = 1$, the matrix inverse $(J - \gamma^{(A)}(z))^{-1}$ in equation (4.11) ceases to exist. We therefore convert that equation to a form more suitable for taking that limit. In the appendix it is shown that

$$\sum_{\boldsymbol{a} \in A} \mathcal{F}_{\boldsymbol{a}} = \frac{1}{1 - \gamma_A} \tag{4.12}$$

where

$$\gamma_A^{-1}(z) \equiv \sum_{\boldsymbol{a}, \boldsymbol{a}' \in A} [(\gamma^{(A)}(z))^{-1}]_{\boldsymbol{a}\boldsymbol{a}'}. \tag{4.13}$$

Upon coming back to the original notation, but now with the abbreviation

$$g_A(z) \equiv \gamma_A(z)G_0(z) \tag{4.14}$$

we find from equation (4.12) the solution of equation (4.7) in the form

$$\sum_{\boldsymbol{a} \in A} \hat{F}_A(\boldsymbol{a}, z) = \frac{1}{(1-z)[G_0(z) - g_A(z)]}. \tag{4.15}$$

When $|A| = 1$, in which case we may take $A = \{\mathbf{0}\}$, one deduces directly from (4.7) that (4.15) holds with $g_{\{\mathbf{0}\}}(z) = 0$. For A of diameter not too large, as is the case in many examples of interest, the quantity $g_A(z)$ is easily expressed explicitly in terms of the $g(\boldsymbol{a} - \boldsymbol{a}', z)$. If after substitution of expression (4.15) in equation (4.5) we sum on A, use equation (3.8), and transform back to the time domain, we get

$$\overline{M}(t) = -\frac{1}{2\pi i} \oint \frac{\mathrm{d}z}{z^{t+1}} \frac{1}{(1-z)^2} \sum_{A \neq \emptyset} \mu_A \frac{1}{G_0(z) - g_A(z)} \tag{4.16}$$

where the integral runs counterclockwise around the origin. This result is still fully exact and applies to both finite and infinite lattices.

4.3. Long-time behaviour of averages. Infinite lattice

The starting point for the analysis of this section is equation (4.16). On an infinite lattice the asymptotic behaviour of $\overline{M}(t)$ as $t \to \infty$ is determined by the $z \to 1$ behaviour of the integrand. This behaviour follows from expression (2.8) for $G_0(z)$ and from equations (4.14) and (4.13) which together determine $g_A(z)$. We shall satisfy ourselves with retaining the leading $z \to 1$ behaviour *and* corrections that are of the relative order of negative powers of $\log(1-z)$. The terms neglected are of the relative order of $1-z$, apart from logarithmic factors. This means that in equation (4.16) we may replace $G_0(z)$ with $\pi^{-1}\log(8/(1-z))$ and the $g_A(z)$ with their values at $z = 1$, which for brevity we shall denote by g_A. The coefficient g_A appears in potential theory and is the two-dimensional lattice analogue of the electrostatic capacity of the set A; its properties have been reviewed by Spitzer [27].

We expand the summand in equation (4.16) in inverse powers of $\log(8/(1-z))$. Using the above results we so obtain, asymptotically for $t \to \infty$,

$$\overline{M}(t) \simeq \sum_{n=0}^{\infty} m_n \mathcal{J}_{n+1,2}(t) \tag{4.17}$$

where the coefficients, m_n, are determined by the observable, M, according to

$$m_n = -\sum_{A \neq \emptyset} \mu_A g_A^n \tag{4.18}$$

for $n = 0, 1, 2, \ldots$ and where

$$\mathcal{J}_{n\ell}(t) = \frac{1}{2\pi\mathrm{i}} \oint \frac{\mathrm{d}z}{z^{t+1}} \frac{1}{(1-z)^{\ell}} \frac{\pi^n}{\log^n \frac{8}{1-z}}. \tag{4.19}$$

The $t \to \infty$ behaviour of the integrals (4.19) has been studied, in particular, by Henyey and Seshadri [28] for the case $\ell = 2$. A generalization of their result is

$$\mathcal{J}_{n\ell}(t) \simeq \frac{\pi^n t^{\ell-1}}{\log^n 8t} \sum_{m=0}^{\infty} (-1)^m \binom{m+n-1}{m} \frac{b_{m\ell}}{\log^m 8t} \tag{4.20}$$

with coefficients

$$b_{m\ell} = \frac{\mathrm{d}^m}{\mathrm{d}x^m} \frac{1}{\Gamma(x+\ell)}\bigg|_{x=0}. \tag{4.21}$$

We have $b_{0\ell} = 1/(\ell-1)!$ and shall also need explicitly below

$$\begin{aligned} b_{12} &= -1 + C = -0.422\ldots \\ b_{13} &= -\tfrac{3}{4} + \tfrac{1}{2}C = -0.461\ldots \\ b_{22} &= 2 - \tfrac{1}{6}\pi^2 - 2C + C^2 = -0.466\ldots \\ b_{23} &= \tfrac{7}{4} - \tfrac{1}{12}\pi^2 - \tfrac{3}{2}C + \tfrac{1}{2}C^2 = 0.227\ldots \end{aligned} \tag{4.22}$$

where $C = 0.577\,215\ldots$ denotes Euler's constant. Combining equations (4.17) and (4.20) we find for $\overline{M}(t)$ an asymptotic expansion in inverse powers of $\log t$,

$$\begin{aligned} \overline{M}(t) \simeq \frac{\pi t}{\log 8t} m_0 &+ \frac{\pi t}{\log^2 8t}[-b_{12}m_0 + \pi m_1] \\ &+ \frac{\pi t}{\log^3 8t}[b_{22}m_0 - 2\pi b_{12}m_1 + \pi^2 m_2] \\ &+ \frac{\pi t}{\log^4 8t}[-b_{32}m_0 + 3\pi b_{22}m_1 - 3\pi^2 b_{12}m_2 + \pi^3 m_3] + \ldots \end{aligned} \tag{4.23}$$

where we have set $b_{02} = 1$. We recall that, whereas the $b_{m\ell}$ are numerical coefficients, the m_n defined by equation (4.18) are specific for the observable M. equation (4.23) shows that the leading asymptotic behaviour is $\sim t/\log 8t$ for observables that have $m_0 \neq 0$, and $\sim t/\log^2 8t$ for those that have $m_0 = 0$ but $m_1 \neq 0$. In view of the definition of m_0 and the discussion at the end of section 3.2, observables with $m_0 \neq 0$ involve only black patterns and will be called, for short, 'black' observables, whereas those with $m_0 = 0$ will be called 'black-and-white' observables.

4.4. Examples

It is now easy to derive results for many examples of interest by applying the formulae of section 4.3.

Example 1.1. We take for M the total number, S, of sites in the support. The first column of coefficients μ_A in table 1 shows that the only nonzero term in the sum (4.18) is due to the set $A = \{\mathbf{0}\}$. Since this set has $\mu_A = -1$ and $g_A = 0$, the only nonzero coefficient produced by equation (4.18) is $m_0 = 1$. From equation (4.23) we then have

$$\overline{S}(t) \simeq \frac{\pi t}{\log 8t}\left[1 - \frac{b_{12}}{\log 8t} + \frac{b_{22}}{\log^2 8t} - \frac{b_{32}}{\log^3 8t} + \cdots\right]. \tag{4.24}$$

The numerical values of the coefficients of the first two subleading terms are given in equation (4.22) and agree with those of Torney [11].

Example 1.2. Next we take for M the total boundary length, E, between the white and black areas. The second column of coefficients μ_A in table 1 shows that only two sets, A, enter, with the pair (μ_A, g_A) equal to $(4, 0)$ and $(-4, \frac{1}{2})$. Equation (4.18) then leads to $m_0 = 0$ and $m_n = 2^{2-n}$ for $n = 1, 2, \ldots$, after which equation (4.23) gives

$$\overline{E}(t) \simeq \frac{2\pi^2 t}{\log^2 8t}\left[1 + \frac{\varepsilon_1}{\log 8t} + \frac{\varepsilon_2}{\log^2 8t} + \cdots\right] \tag{4.25}$$

with the coefficients

$$\begin{aligned} \varepsilon_1 &= \tfrac{1}{2}\pi - 2b_{12} = 2.416\,36\ldots \\ \varepsilon_2 &= \tfrac{1}{4}\pi^2 - \tfrac{3}{2}\pi b_{12} + 3b_{22} = 3.061\,16\ldots. \end{aligned} \tag{4.26}$$

Example 1.3. Now take for M the total number, I, of islands. Using the third column of coefficients μ_A in table 1 as input in equation (4.18) we find that $m_0 = 0$ and

$$m_n = 2^{1-n} - [(\pi + 2)/2\pi]^n \tag{4.27}$$

for $n = 1, 2, \ldots$. Substituting as in the previous examples we find

$$\overline{I}(t) \simeq \tfrac{1}{2}\pi(\pi - 2)\frac{t}{\log^2 8t}\left[1 + \frac{\iota_1}{\log 8t} + \frac{\iota_2}{\log^2 8t} + \cdots\right]. \tag{4.28}$$

Analytic expressions for the coefficients ι_i are easily found with the aid of the preceding formulae but are of little interest. The numerical values of the first two of them are

$$\begin{aligned} \iota_1 &= 12.087\ldots \\ \iota_2 &= -21.304\ldots. \end{aligned} \tag{4.29}$$

In the leading asymptotic behaviour of both $\overline{E}(t)$ and $\overline{I}(t)$ an extra factor $1/\log 8t$ appears compared with that of $\overline{S}(t)$ as a consequence of m_0 being zero, i.e. of $E(t)$ and $I(t)$ being black-and-white observables. In section 4.5 we shall come back to $\overline{I}(t)$ and also make a comparison with the numerical simulations by Coutinho *et al* [4].

Example 1.4. Let β be a specific type of island, where *type* indicates shape, size and orientation. Let $I_\beta(t)$ be the observable that counts the total number of islands of that type. According to the preceding discussion we must have that

$$\overline{I}_\beta(t) \simeq \tfrac{1}{2}\pi(\pi-2) f_\beta \frac{t}{\log^2 8t} \qquad \text{as } t \to \infty \tag{4.30}$$

for some proportionality constant, f_β, even though we cannot expect the approach to this asymptotic behaviour to be uniform in β. By summing equation (4.30) on all β and comparing to equation (4.28) we conclude that the average number of islands of type β represents, as $t \to \infty$, a fixed *fraction*, f_β, of the average total number of islands. We have calculated the fraction, f_1, of islands that are single isolated sites and the fraction, f_2, of '*dimer*' islands, consisting of two neighbouring sites, with the result

$$\begin{aligned} f_1 &= -\frac{\pi(\pi^3 - 7\pi^2 + 14\pi - 4)}{(\pi-1)(\pi^2 - 6\pi + 4)} = 0.560\,079\ldots \\ f_2 &= 0.073\,557\ldots. \end{aligned} \tag{4.31}$$

The main effort goes into the calculation of the necessary coefficients g_A. Those needed for I_1 are listed in table 1. We do not present the 22 coefficients needed for I_2, nor the final analytic expression for f_2. Since the dimers may have two orientations, the islands of sizes 1 and 2 represent a fraction $f_1 + 2f_2 = 0.707\,193\ldots$ of all islands.

With these results in hand we return to the total boundary length considered in example 1.2. On an infinite lattice one can write $E(t) = E_{\text{ext}}(t) + E_{\text{int}}(t)$, where $E_{\text{ext}}(t)$ is the external perimeter of the support and $E_{\text{int}}(t)$ the total perimeter of the islands enclosed by it. In example 1.2. we determined only the average of their sum; determining $\overline{E}_{\text{ext}}(t)$ and $\overline{E}_{\text{int}}(t)$ separately is a much more difficult problem that we have not seen how to solve by the present method. A rigorous lower bound for $\overline{E}_{\text{int}}(t)$ is nevertheless easily obtained by adding up the perimeters of the single-site and dimer islands, and taking into account that all other islands have a perimeter at least equal to 8. This gives $\overline{E}_{\text{int}}(t) > 0.4965\ldots\overline{E}(t)$. By arguments different from ours Lawler [22] shows that in fact for $t \to \infty$ the external perimeter increases only as $\overline{E}_{\text{ext}}(t) \sim t^{2/3}$, so that $\overline{E}_{\text{int}}(t) \simeq \overline{E}(t)$.

Example 1.5. Finally let $M(t)$ be equal to the pattern number $N_\alpha(t)$ obtained by summing the expression (1.3) over all sites $\boldsymbol{x}$. If neither A_1 nor A_2 is empty, then the pattern α is composed of both white and black sites and has $m_0 = 0$. Hence the average total number of these patterns is obtained by setting $m_0 = 0$ in equation (4.23) and increases with t as

$$\overline{N}_\alpha(t) \simeq m_1 \frac{\pi^2 t}{\log^2 8t} \tag{4.32}$$

where m_1 is α dependent. The black-and-white patterns considered here necessarily lie on the boundary of the support; comparison of equations (4.32) and (4.25) shows that the *number per unit of boundary length* of such patterns tends to a fixed value as $t \to \infty$.

4.5. Long-time behaviour of averages. Finite lattice

For a finite lattice of N sites, after an initial increase with time identical to what happens on the infinite lattice, we must expect deviations from the infinite lattice behaviour to appear on a characteristic N dependent time scale $\tau(N)$ that will tend to infinity when N does. For larger times all black observables will level off and tend to a constant times N, and all black-and-white observables will pass through a maximum value, then bend down and asymptotically approach zero. The expansion that we shall look for in the case of a finite lattice will therefore involve a determination of this time scale.

We shall not attempt in this case a full asymptotic expansion as for the infinite lattice, but only determine the leading asymptotic behaviour. Our starting point is again equation (4.16), in which, when $z \to 1$, using equations (2.5) and (2.6), we may substitute

$$G_0(z) - g_A(z) = \frac{1}{N(1-z)} + \frac{1}{\pi}\log cN - g_A + \cdots \tag{4.33}$$

where as before g_A stands for $g_A(1)$ and the dots denote higher-order terms. As for the infinite lattice, our procedure will be to bring the summation on A outside the integration on z and shift the integration path around the poles of the integrand. Therefore we should now discuss these poles. All required knowledge about the behaviour of the various quantities involved has been collected in section 2.2. First, it follows from equation (4.33) and the definition (4.18) of m_0 that for $z \to 1$

$$\sum_{A\neq\emptyset} \mu_A \frac{1}{G_0(z) - g_A(z)} \simeq -Nm_0(1-z). \tag{4.34}$$

Hence the integrand of equation (4.16) has a simple pole at $z = 1$ with residue Nm_0. Secondly, as is clear from the discussion in section 2.2, this integrand has special simple poles for $z = z_A$, where

$$z_A = 1 + \frac{\pi}{N\log cN}\left[1 + \frac{\pi g_A}{\log cN} + \mathcal{O}\left(\frac{1}{\log^2 N}\right)\right]. \tag{4.35}$$

These are the only poles at a distance of $\mathcal{O}(1/N\log N)$ from $z = 1$, all the other ones being at least at distances of $\mathcal{O}(1/N)$. Therefore, on time scales that are at least of $\mathcal{O}(N\log N)$, the other poles will contribute vanishingly to the result.

Carrying the integral out but retaining only the contribution of the poles at $z = 1$ and $z = z_A$ leads to

$$\overline{M}(t) \simeq Nm_0 + N\sum_{A\neq\emptyset} \mu_A \exp\left[-\frac{\pi t}{N\log cN}\left(1 + \frac{\pi g_A}{\log cN} + \mathcal{O}\left(\frac{1}{\log^2 N}\right)\right)\right]. \tag{4.36}$$

The first term in this equation is not present for the black-and-white observables, which have $m_0 = 0$. We shall discuss these observables first. Two different time scales are of interest.

(1) *The main regime*, in which an observable takes values of the same order as its maximum value. To focus on this regime we scale time as

$$\tau = \frac{\pi t}{N\log cN} \tag{4.37}$$

and take the limit $N \to \infty$, $t \to \infty$ at $\tau > 0$ fixed. In this 'τ-*limit*' equation (4.36) leads directly to

$$\overline{M}(t) \simeq \frac{\pi N}{\log cN} m_1 \tau \mathrm{e}^{-\tau} \tag{4.38}$$

with m_1 defined by equation (4.18).

(2) *The long-time regime,* in which the observable approaches its final value. It is now appropriate to scale time as

$$\sigma = \frac{\pi t}{N \log^2 cN} \tag{4.39}$$

with $\sigma > 0$. In this 'σ*-limit*' equation (4.36) gives

$$\overline{M}(t) \simeq N(cN)^{-\sigma} \sum_{A \neq \emptyset} \mu_A \mathrm{e}^{-\pi g_A \sigma}. \tag{4.40}$$

One may notice that the result (4.38) is recovered from equation (4.40) if one sets $\sigma = \tau/\log cN$ and expands the terms inside the sum on A to first order in $1/\log cN$.

For black observables equation (4.36), with the scaling of equation (4.37), leads directly to

$$\overline{M}(t) \simeq N m_0 (1 - \mathrm{e}^{-\tau}). \tag{4.41}$$

This decay law depends on the type of the black pattern only through the prefactor m_0.

4.6. Examples

We consider the same examples as in section 4.4 but now on the finite lattice.

Example 2.1. Let $M = S$. Substituting $m_0 = 1$ in equation (4.36) we see that on a finite lattice of N sites the average number of unvisited sites decays as

$$N - \overline{S}(t) \simeq N\mathrm{e}^{-\tau} \tag{4.42}$$

a result first obtained by Weiss *et al* [24]. It is also valid in the long-time regime, where it can be written as $N(cN)^{-\sigma}$. In this regime, since $\sigma > 0$, the unvisited sites constitute an infinitesimally small fraction of all lattice sites.

Example 2.2. Let now $M = E$. Using in equation (4.36) the appropriate coefficients μ_A from table 1 we obtain

$$\begin{aligned} \overline{E}(t) &\simeq \frac{2\pi N}{\log cN} \tau \mathrm{e}^{-\tau} \\ \overline{E}(t) &\simeq 4N(cN)^{-\sigma}(1 - \mathrm{e}^{-\pi\sigma/2}) \end{aligned} \tag{4.43}$$

in the main and long-time regimes, respectively. The prefactor $N(cN)^{-\sigma}$ in the second one of these equations is the average number of unvisited sites found in the preceding example. If these sites were randomly distributed, then since they are infinitely dilute, each of them would have four visited neighbours and the result for $\overline{E}(t)$ would be only $4N(cN)^{-\sigma}$. Hence the factor $1 - \mathrm{e}^{-\pi\sigma/2}$ represents nontrivial correlations due to unvisited sites clustering together.

Example 2.3. Let $M = I$. Using in equation (4.36) the coefficients μ_A from table 1 appropriate to this case, we find

$$\begin{aligned} \overline{I}(t) &\simeq \frac{(\pi/2 - 1)N}{\log cN} \tau \mathrm{e}^{-\tau} \\ \overline{I}(t) &\simeq N(cN)^{-\sigma}(1 - 2\mathrm{e}^{-\pi\sigma/2} + \mathrm{e}^{-(\pi+2)\sigma/2}) \end{aligned} \tag{4.44}$$

for the main regime and the long-time regime, respectively. Again, the factor in parentheses in the last equation is due to correlations in the positions of the unvisited sites.

This example requires the following remark. The expression of equation (3.2) for the number of islands, $I(t)$, is correct only for the infinite lattice. For the finite lattice with periodic boundary conditions in both directions one should add -1 when the support closes onto itself around the torus in one of the directions and -2 when it does so in both directions. In the latter case the first term in equation (3.2) is absent and one obtains the expression that we used to derive equations (4.28) and (4.44). The extra term -2 is of course of no importance in the main regime, but it needs to be taken into account in the long-time regime to ensure that $\overline{I}(t)$ vanishes when t tends to infinity.

If we neglect the difference between averages of ratios and ratios of averages, then we have from equations (4.42), (4.43), and (4.44) that the quantities

$$\begin{aligned} &[N-\overline{S}(t)]/\overline{I}(t) \simeq 1/(1-2\mathrm{e}^{-\pi\sigma/2}+\mathrm{e}^{-(\pi+2)\sigma/2}) \\ &\overline{E}(t)/\overline{I}(t) \simeq 4(1-\mathrm{e}^{-\pi\sigma/2})/(1-2\mathrm{e}^{-\pi\sigma/2}+\mathrm{e}^{-(\pi+2)\sigma/2}) \end{aligned} \tag{4.45}$$

represent the average area and the average perimeter, respectively, of an island. In the main time regime the expression for the average area simplifies to

$$[N-\overline{S}(t)]/\overline{I}(t) \simeq \frac{2\log cN}{\pi-2}\tau^{-1}. \tag{4.46}$$

It is now possible to make a comparison with the simulations by Coutinho *et al* [4], carried out on finite square lattices of up to $N = 1200^2$ sites with periodic boundary conditions. These authors were interested in the 'fragmentation' of the finite lattice into islands and the way the average number and size of the islands eventually tend to zero. The comparison leads to the following conclusions.

(i) In an *early time regime*, which for a lattice of 600^2 sites corresponds to t less than $\approx 0.5\times 10^6$, the finite lattice size still plays no role and our equation (4.28) for $\overline{I}(t)$ agrees within error bars with the simulation data shown in [4] for $N = 600^2$. These data are for $t > 0.05\times 10^6$; for the agreement to be of this quality, not only the leading-order term but also the two correction terms in equation (4.28) have to be taken into account.

(ii) In the early time regime and in the main regime, where $\overline{I}(t)$ passes through its maximum, our long-time expansion equation (4.44) overestimates the numerical values of $\overline{I}(t)$ by up to 50%; in the long-time regime, when $\overline{I}(t)$ starts to decay, the agreement with the simulation data becomes rapidly better and stays good all the way up to $t \approx 19\times 10^6$, where with a large probability, no unvisited sites are left.

(iii) Coutinho *et al* [4] in their simulation find the time-dependent average island size on a lattice of N sites to be a function only of the scaling variable σ. Our expression (4.45) confirms this result. The scaling function is not, however, the power law that it was thought to be in [4], but the inverse of a sum of exponentials given in equation (4.45). According to our equation (4.46) a power law appears only in the main regime and has the exponent -1. In the long-time regime equation (4.45) leads to an apparent exponent with larger absolute value.

(iv) In the preceding discussion we have compared the numerical data for the average island size $\overline{[N-S(t)]/I(t)}$ to the theoretical result (4.45) for $[N-\overline{S}(t)]/\overline{I}(t)$ that we were able to calculate. We have not been able to directly calculate the average island size. Nor have we been able to calculate still another quantity determined in the simulation of [4], namely the time-dependent 'diversity' of the island sizes, defined as the number of different sizes that occur at any given time.

5. Fluctuations and correlations

5.1. Relation to a first passage time problem

The original determination by Dvoretzky and Erdös [6] of the average number $\overline{S}(t)$ of lattice sites in the support of a random walk was followed only much later [10, 11] by a calculation of the root-mean-square deviation of this quantity from its average. Yet that calculation was important, because the result, exhibited in our equation (1.2), shows that in the limit $t \to \infty$ the probability distribution of $S(t)$ becomes infinitely narrow, even though only logarithmically slowly with t.

It is now natural to ask if the more general observables whose averages we studied in section 4 also have infinitely narrow distributions for $t \to \infty$. Without much extra effort it will be possible to also calculate the cross-correlations. We therefore consider two observables of the form (3.5), namely

$$M(t) = \sum_A \mu_A M_A(t) \qquad M'(t) = \sum_B \mu'_B M_B(t) \tag{5.1}$$

which have $\sum_A \mu_A = \sum_B \mu'_B = 0$, and focus on

$$\overline{\Delta M(t) \Delta M'(t)} \equiv \overline{M(t) M'(t)} - \overline{M}(t)\overline{M'}(t). \tag{5.2}$$

In this section we confine our analysis to the infinite lattice.

The first step will be to find an expression for the generating function $\hat{C}_{MM'}$ defined by

$$\hat{C}_{MM'}(z) = \sum_{t=0}^{\infty} z^t \overline{M(t) M'(t)}. \tag{5.3}$$

When working out the right-hand side of equation (5.3) with the aid of equations (5.1) and (3.4) we encounter averages of products $m_{\boldsymbol{x}+A}(t) m_{\boldsymbol{y}+B}(t)$ where $\boldsymbol{x}$ and $\boldsymbol{y}$ are arbitrary lattice vectors. It is then convenient to write $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{r}$ and to define $U(\boldsymbol{r})$ as the union of A and $\boldsymbol{r} + B$.

With this notation we get

$$\overline{M(t) M'(t)} = \sum_{A,B} \mu_A \mu'_B \sum_{\boldsymbol{r}} \overline{M}_{U(\boldsymbol{r})}(t). \tag{5.4}$$

The observable $M_{U(\boldsymbol{r})}$ is of the form (3.4), with A replaced by $U(\boldsymbol{r})$. We can therefore directly apply the results of sections 4.1 and 4.2. After separating the terms with either $A = \emptyset$ or $B = \emptyset$ from the others one finds

$$\hat{C}_{MM'}(z) = -\frac{1}{1-z} \sum_{A,B \neq \emptyset} \mu_A \mu'_B \sum_{\boldsymbol{r}} \left[\sum_{\boldsymbol{u} \in U(\boldsymbol{r})} \hat{F}_{U(\boldsymbol{r})}(\boldsymbol{u}, z) - \sum_{\boldsymbol{a} \in A} \hat{F}_A(\boldsymbol{a}, z) - \sum_{\boldsymbol{b} \in B} \hat{F}_B(\boldsymbol{b}, z) \right] \tag{5.5}$$

which is exactly of the form (4.5). The first sum in the square brackets, in particular, is just given by equation (4.15) with A replaced by $U(\boldsymbol{r})$, and is therefore equal to $(1-z)^{-1}[G_0(z) - g_{U(\boldsymbol{r})}(z)]^{-1}$.

5.2. Scaling limit and long-time behaviour

Further analysis of equation (5.5) requires that we render the $\boldsymbol{r}$ dependence of $g_{U(\boldsymbol{r})}$ more transparent. This we shall be able to do only in the scaling limit $z \to 1$, $r \to \infty$, with $\xi^2 \equiv 4r^2(1-z)$ fixed. But since $z \to 1$ is exactly the limit of interest, this suffices provided the sum on $\boldsymbol{r}$ in equation (5.5) is dominated by values $r \sim (1-z)^{-1/2}$. In the notation of section 4.2 the quantity $g_{U(\boldsymbol{r})}$ is defined with the aid of the matrix $g^{(U(\boldsymbol{r}))}$. In the scaling limit we need to consider only the case where the constituents A and $\boldsymbol{r}+B$ of $U(\boldsymbol{r})$ are disjoint sets. We may then use the scaling form equation (2.14) for the $\hat{G}(\boldsymbol{r}+\boldsymbol{b}-\boldsymbol{a},z)$ that appear in the definition of $g^{(U(\boldsymbol{r}))}$. These scaling expressions are to leading order in $1-z$ independent of $\boldsymbol{a}$ and $\boldsymbol{b}$, which is at the origin of the resulting simplification. In the scaling limit $g^{(U(\boldsymbol{r}))}$ has two diagonal blocks, $g^{(A)}$ and $g^{(B)}$, and two off-diagonal blocks all whose elements are equal to $1-\lambda$ with

$$\lambda(\xi,z)=2\pi^{-1}K_0(\xi)/G_0(z). \tag{5.6}$$

In the appendix it is shown that

$$\frac{1}{1-\gamma_{U(\boldsymbol{r})}}\simeq\frac{2-2\lambda-\gamma_A-\gamma_B}{1-\lambda^2-\gamma_A-\gamma_B+\gamma_A\gamma_B}. \tag{5.7}$$

Upon using the preceding results in equation (5.5) and replacing the sum on $\boldsymbol{r}$ by an integral on ξ we find

$$\hat{C}_{MM'}(z)\simeq-\frac{\pi}{2(1-z)^3G_0(z)}\sum_{A,B\neq\emptyset}\mu_A\mu'_B\int_0^\infty\mathrm{d}\xi\,\xi I(\xi,z) \tag{5.8}$$

in which the function $I(\xi,z)$ is given by

$$I=\frac{2-2\lambda-\gamma_A-\gamma_B}{1-\lambda^2-\gamma_A-\gamma_B+\gamma_A\gamma_B}-\frac{1}{1-\gamma_A}-\frac{1}{1-\gamma_B}. \tag{5.9}$$

It is useful to observe that the three quantities $\gamma_A(z)$, $\gamma_B(z)$, and $\lambda(\xi,z)$ are all of the order of $G_0^{-1}(z)$.

The steps that follow are again analogous to the procedure of section 4.3. We wish to expand $I(\xi,z)$ in inverse powers of $G_0(z)$ and substitute the result in equation (5.8). The μ'_B that characterize the observable M' define coefficients m'_n analogous to the m_n of equation (4.18). Furthermore in the expansion we encounter the coefficients

$$a_n=\int_0^\infty\mathrm{d}\xi\,\xi K_0^n(\xi) \tag{5.10}$$

of which we shall need the explicit values $a_1=1$ and $a_2=\frac{1}{2}$, as well as [10, 11]

$$a_3=-\frac{1}{2}\int_0^1\mathrm{d}\xi\,\frac{\log\xi}{1-\xi+\xi^2}=0.585\,97\ldots. \tag{5.11}$$

Finally, before working out this expansion it is useful to classify the observables, M, according to their order. We shall say that M is of the *order* of k if

$$m_0=m_1=\cdots=m_{k-1}=0\qquad\text{and}\qquad m_k\neq0. \tag{5.12}$$

Observables of the order of $k=0$ and $k=1$ have occurred in the preceding sections, and physically interesting examples with $k\geqslant2$ perhaps exist. Now let k and k' be the orders of M and M', respectively. We anticipate—as will be confirmed by the calculation—that we have to expand $\hat{C}_{MM'}(z)$ in equation (5.8) to the order of $1/G_0^{k+k'+4}(z)$. In view of equations (5.12) and (4.18) only those terms in the expansion will survive the summation on A and B that contain at least a factor γ_A^k and a factor $\gamma_B^{k'}$. Since there is one factor $1/G_0(z)$ outside the sum in equation (5.8), this leaves room for at most three factors λ. We have therefore found it convenient to begin by expanding $I=I'\lambda+I''\lambda^2+I'''\lambda^3+\cdots$ and then to determine the first three coefficients of this series in terms of γ_A and γ_B. Then the $1/G_0(z)$ expansion of $\hat{C}_{MM'}(z)$ leads to

$$\begin{aligned}\hat{C}_{MM'}(z) \simeq \frac{1}{(1-z)^3 G_0^{k+k'+2}(z)}&[2a_1 m_k m'_{k'} - G_0^{-1}(z)(\pi^{-1}(k+k'+2)a_2 m_k m'_{k'}\\&-2a_1(m_k m'_{k'+1} + m_{k+1}m'_{k'})) + G_0^{-2}(z)(8\pi^{-2}(k+1)(k'+1)a_3 m_k m'_{k'}\\&-2\pi^{-1}(k+k'+3)a_2(m_k m'_{k'+1} + m_{k+1}m'_{k'})\\&+2a_1(m_k m'_{k'+2} + m_{k+1}m'_{k'+1} + m_{k+2}m'_{k'})) + \cdots].\end{aligned} \tag{5.13}$$

The next step is to invert equation (5.3). After integrating on z with the aid of equation (4.20) one finds

$$\begin{aligned}\overline{M(t)M'(t)} \simeq \frac{\pi^{k+k'+2}t^2}{\log^{k+k'+2} 8t}\Bigg[&a_1 m_k m'_{k'} + \frac{1}{\log 8t}(-(k+k'+2)(2a_1 b_{13} + a_2)m_k m'_{k'}\\&+\pi a_1(m_k m'_{k'+1} + m_{k+1}m'_{k'})) + \frac{1}{\log^2 8t}(\{(k+k'+2)(k+k'+3)\\&\times(a_1 b_{23} + 2a_2 b_{13}) + 4(k+1)(k'+1)a_3)\}m_k m'_{k'}\\&-(k+k'+3)\pi(a_2 + 2a_1 b_{13})(m_k m'_{k'+1} + m_{k+1}m'_{k'})\\&+\pi^2 a_1(m_k m'_{k'+2} + m_{k+1}m'_{k'+1} + m_{k+2}m'_{k'})) + \cdots\Bigg].\end{aligned} \tag{5.14}$$

From equation (4.23) we deduce that when M is of the order of k its average is given by

$$\begin{aligned}\overline{M}(t) \simeq \frac{\pi^{k+1}t}{\log^{k+1} 8t}\Bigg[&m_k + \frac{1}{\log 8t}(-(k+1)b_{12}m_k + \pi m_{k+1}) + \frac{1}{\log^2 8t}(\tfrac{1}{2}(k+1)(k+2)b_{22}m_k\\&-\pi(k+2)b_{12}m_{k+1} + \pi^2 m_{k+2}) + \cdots\Bigg].\end{aligned} \tag{5.15}$$

Equation (5.15), its counterpart for $\overline{M'}(t)$, and equation (5.14) now have to be combined in equation (5.2). Using the explicit expressions for the coefficients a_1, a_2, b_{12}, b_{13}, b_{22}, and b_{23} leads to the desired correlation function. For $t \to \infty$ the two leading orders cancel in the subtraction in equation (5.2). The final result is

$$\frac{1}{(k+1)(k'+1)}\frac{\overline{\Delta M(t)\Delta M'(t)}}{\overline{M}(t)\overline{M'}(t)} \simeq \frac{\mathcal{A}^2}{\log^2 8t} + \cdots \tag{5.16}$$

where the dots indicate terms of higher-order in $1/\log 8t$ and

$$\mathcal{A}^2 = 4a_3 + 1 - \tfrac{1}{6}\pi^2 = 1.698\,97\ldots. \tag{5.17}$$

Equation (5.16) contains as a special case the well known result of equation (1.2) for the variance of $S(t)$, originally due to Jain and Pruitt [10], and rederived with the aid of a method more similar to ours by Torney [11].

5.3. Conclusions

It is remarkable that the ratio on the right-hand side of equation (5.16) is universal. It is independent of the choice of the observables M and M'. But we shall now see that equation (5.16) has consequences that reach far beyond this simple fact.

For $k = 0, 1, 2, \ldots$ we define for each observable M of order k its normalized deviation from average, η_M, by

$$\eta_M(t) = \frac{\log 8t}{(k+1)\mathcal{A}}\frac{\Delta M(t)}{\overline{M}(t)}. \tag{5.18}$$

These variables satisfy to the leading order

$$\overline{\eta_M(t)} = 0 \qquad \overline{\eta_M(t)\eta_{M'}(t)} = 1 \qquad \text{for all } M, M'. \tag{5.19}$$

It follows that for any two M and M' the difference $\eta_M(t) - \eta_{M'}(t)$ has zero variance, and therefore the $\eta_M(t)$ *are all equal to a single random variable* that we shall call $\eta(t)$. As a consequence we can relate the deviation from average $\Delta M(t)$ of any observable $M(t)$ of order k to $\eta(t)$ by

$$\Delta M(t) \simeq (k+1)\frac{\mathcal{A}}{\log 8t}\overline{M}(t)\eta(t). \tag{5.20}$$

Upon writing down this equation for the special case $M = S$, using the explicit expression (4.24) for $\overline{S}$, and eliminating $\eta(t)$, one finds

$$\Delta M(t) \simeq (k+1)\frac{\pi^k}{\log^k 8t}\, m_k \Delta S(t). \tag{5.21}$$

This last equation embodies one of the main conclusions of this work. *All the observables, M, fluctuate around their averages in strict proportionality with the fluctuation of total number of sites, $S(t)$, in the support.* This conclusion applies, in particular, to the pattern numbers, $N_\alpha(t)$, the total perimeter length, $E(t)$, of the support, the total number, $I(t)$, of islands enclosed by it, and the total number, $I_\beta(t)$, of islands of a specific type β. We are not aware of any computer simulations that confirm equation (5.21), although they would be easy to carry out.

There is another different and instructive way to formulate this conclusion. Let M be an observable of order k and ρ a quantity that remains of $\mathcal{O}(1)$ when $t \to \infty$. equation (5.15) can now be used to establish the asymptotic expansion in powers of $1/\log 8t$ of $\rho^{-1}\overline{M}(\rho t)$. Upon comparison with equation (5.20) and choosing $\log \rho = -\mathcal{A}\eta(t)$ one finds that all fluctuating observables, $M(t)$, can be written to the second order in the form

$$M(t) \simeq \mathrm{e}^{\mathcal{A}\eta(t)}\overline{M}(\mathrm{e}^{-\mathcal{A}\eta(t)}t). \tag{5.22}$$

In the mathematical literature on Brownian motion (for a review see Le Gall [18]) the quantity $-\mathcal{A}\eta/(2\pi)$ has appeared in the study of the asymptotic behaviour of the volume of the Wiener sausage (where it is commonly denoted by the symbol γ) and is known as the *renormalized local time of self-intersections*, a concept introduced by Varadhan in an appendix to an article by Symanzik [29].

6. Discussion

In a preceding letter [5] a more general approach was presented to the much more restricted problem of how to calculate the average $\overline{I}(t)$. The total number of islands was written as $I(t) = C(t) - D(t)$, where the increments $\Delta C(t) = C(t) - C(t-1)$ and $\Delta D(t) = D(t) - D(t-1)$ are the numbers of islands created and destroyed, respectively, in the tth step (either $\Delta C(t)$ or $\Delta D(t)$ vanishes). Hence $C(t)$ and $D(t)$ are the total number of islands created and destroyed, respectively, up until time t. It was shown [5] that asymptotically for $t \to \infty$ to leading order

$$\overline{C}(t) \simeq \overline{D}(t) \simeq A\frac{\pi t}{\log 8t} \tag{6.1}$$

where $A = 0.1017\ldots$. In the difference $\overline{I}(t) = \overline{C}(t) - \overline{D}(t)$ the leading order (6.1) cancels and the result (4.28) appears. It does not seem possible to express $C(t)$ and $D(t)$ as observables of the type $M(t)$. The additional determination of $\overline{C}(t)$ and $\overline{D}(t)$

makes the calculation of [5] more involved. Nevertheless, the generating function found there (equation (5) of [5]) for $\mathrm{d}^2\overline{I}(t)/\mathrm{d}t^2$ is equivalent to the one for $\overline{I}(t)$ implied by equation (4.16) of this work.

The success of the study presented here is due to the generating function method, whose potential is fully exploited. We also run into what may be the limitations of this method. There are several quantities that appear naturally but whose averages and variances we do not see a way to determine. These include the total external perimeter discussed in section 4.4 and the area of the islands enclosed by the support.

Many of the quantities discussed in the preceding chapters have close analogues in planar Brownian motion. We shall denote these analogues by the superscript B. The Brownian motion analogue of the support of the lattice random walk is the set of points $\mathcal{S}_b^B(t) \subset \mathbb{R}^2$ that has been swept out in the time interval $[0, t]$ by a disk of radius b performing Brownian motion with diffusion constant, D. The set, $\mathcal{S}_b^B(t)$, is commonly called the *Wiener sausage* associated with the Brownian motion trajectory. The total area, $S_b^B(t)$, of this set is analogueous to the number of sites, $S(t)$, in the support of the lattice random walk.

The diffusion constant, D, is defined by the requirement that the mean square displacement equal $4Dt$. It can be scaled away, but we shall keep it here to facilitate comparisons between results from different sources. In the mathematical literature one customarily sets $D = \frac{1}{2}$, whereas the long-time, large-distance limit of the random walk of this work yields $D = \frac{1}{4}$.

The asymptotic behaviour of $\overline{S_b^B(t)}$ was first determined by Leontovitsh and Kolmogorov [16]. Berezhkovskii *et al* [19] give the complete asymptotic expansion

$$\overline{S_b^B(t)} \simeq \frac{4\pi Dt}{\log \kappa Dtb^{-2}} \sum_{m=0}^{\infty} \frac{(-1)^m b_{m2}}{\log^m \kappa Dtb^{-2}} \tag{6.2}$$

where $\kappa \equiv 4\mathrm{e}^{-2C}$ and the b_{m2} are as defined by equation (4.21). The difference $\Delta S_b^B(t)$ was shown [18] to be a random variable such that the distribution of $t^{-1}(\log t)^2 \Delta S_b^B(t)$ converges for $t \to \infty$ to a limit distribution *identical* to the one of $t^{-1}(\log t)^2 \Delta S(t)$.

Two remarks can be made about the relations obtained by differentiating equation (6.2) with respect to b. First, let $E_b^B(t)$ be the total boundary length of the Wiener sausage $\mathcal{S}_b^B(t)$. Assuming that $\partial\mathcal{S}_b^B$ is sufficiently regular we have

$$E_b^B(t) = \frac{\mathrm{d}S_b^B(t)}{\mathrm{d}b}. \tag{6.3}$$

Upon averaging, using equation (6.2) and setting $b_{02} = 1$, we find from equation (6.3)

$$\overline{E_b^B}(t) \simeq \frac{8\pi Dtb^{-1}}{\log^2 \kappa Dtb^{-2}}[1 + \cdots]. \tag{6.4}$$

This expression has the same asymptotic time dependence as equation (4.25) for $\overline{E}(t)$, but the coefficients do not coincide. Secondly, differentiate once more and consider the dimensionless number

$$J_b^B(t) \equiv -\frac{\mathrm{d}^2 S_b^B(t)}{\mathrm{d}b^2}. \tag{6.5}$$

Its average behaves in the large t limit as

$$\overline{J_b^B}(t) \simeq \frac{8\pi Dtb^{-2}}{\log^2 \kappa Dtb^{-2}}. \tag{6.6}$$

It is easy to show that *if* the boundary $\partial\mathcal{S}_b^B(t)$ is sufficiently regular (having at least a tangent vector in each point), then the quantity (6.5) is equal to $2\pi(I_b^B(t)-2)$, with $I_b^B(t)$ the number of islands. However, these regularity conditions are not satisfied here since the boundary has cusp points. Nevertheless, comparison of equations (6.6) and (4.28) shows that $\overline{J_b^B}(t)$ has the same asymptotic time dependence as $\overline{I}(t)$ (but with a different coefficient).

We now discuss the relation of our work to results that have appeared in the mathematical literature. Throughout the comparison it should be borne in mind that whereas those results are rigorous, the ones of this paper have been obtained by the usual methods of mathematical physics.

Mountford [20] was the first to study the connected components of the complement of the Brownian path of a point (i.e. the case $b=0$). For all times $t>0$ the number of these components is infinite due to the presence of many small ones. However, one can ask, for example, what the number $C_\varepsilon(t)$ is of connected components with an area larger than a prescribed value $\pi\varepsilon^2$. Le Gall, strengthening the results due to Mountford, has shown that for almost all Brownian motion trajectories (with $D=\frac{1}{2}$) in a fixed time interval

$$\lim_{\varepsilon\to 0}\frac{\varepsilon^2\log^2 Dt\varepsilon^{-2}}{4Dt}C_\varepsilon(t)=2 \tag{6.7}$$

where we have used dimensional analysis to restore the variables D and t. If one now assumes that $C_\varepsilon(t)$ is of the same order of magnitude as the *total* number $I_\varepsilon^B(t)$ of islands in the Wiener sausage associated with the same Brownian motion trajectory executed by a disk of radius ε, then equation (6.7) has an asymptotic time dependence that agrees with the result of our equation (4.28).

Werner [21] considers the *shape* of the connected component containing a prescribed point not on the trajectory, and shows that for $t\to\infty$ the probability distribution of this quantity tends to a well-defined limit law. This result is possible only because of the scale invariance of Brownian motion and it would be cumbersome to formulate its lattice counterpart, even in the long-time, large-distance limit. The statement of our equation (4.30) about the number of islands, $I_\beta(t)$, of given type β, and the associated result about its fluctuation, $\Delta I_\beta(t)$, implied by the discussion at the end of section 5, constitute the point of closest approach between this paper and Werner's.

Finally, we summarize the new results of this paper. On the one hand, we give the explicit asymptotic behaviour as $t\to\infty$ for several new observables associated with the support of the lattice random walk. These include the total perimeter length, the total number of islands, and the total number of single-site and dimer islands. On an infinite lattice the asymptotic behaviours all consist of a leading term multiplied by a series in inverse powers of $1/\log 8t$. The perimeter and the number of islands are also considered on a finite lattice, where scaling laws are obtained in terms of the time, t, and the lattice size, N, and a comparison with computer simulations by Coutinho *et al* [4] is possible.

On the other hand, there is the important general result of section 5.3, stating that the pattern numbers all fluctuate in strict proportionality with one another and with the total number, $S(t)$, of sites in the support. The fundamental fluctuating variable, called η in this work, is the renormalized local time of self-intersections. This result strongly contributes to shape our picture of the support of the two-dimensional random walk. At any fixed time, t, a large class of detailed properties is determined by the value of the single random variable, η.

One of the open questions that can now be formulated is connected exactly with this random variable, which appears in our work as the time dependent variable, $\eta(t)$. Existing results seem to concern exclusively its stationary distribution, $\mathcal{P}(\eta)$, which prevails in the limit $t \to \infty$. Our investigations point towards the interest of also studying the autocorrelation function, $\overline{\eta(t)\eta(t')}$, and possibly other time-dependent properties. A second question that naturally comes up is: how does the picture in higher dimensions differ from the one found here in $d = 2$? It should certainly be expected that, in sufficiently high dimension, the pattern numbers lose their rigid proportionality and that independently fluctuating random variables appear. The mechanism by which this independence comes about seems worthy of further elucidation. We leave these and other questions for future work.

Acknowledgments

HJH acknowledges a discussion with Professor M Coutinho-Filho that laid the germ out of which part of this work grew, and thanks him for correspondence on the simulation results for finite lattices. FvW and HJH thank Dr M Yor, Dr W Werner, and Professor G Lawler for discussions helping them to interpret their results in the light of the mathematical literature.

Appendix

In this appendix we prove the matrix algebra results we use to derive averages and correlations.

The following result is used in section 4.2. Let A denote an invertible matrix of dimension $\ell \times \ell$, with $\ell \geqslant 2$. In what follows $J^{[m,n]}$ stands for the $m \times n$ matrix whose elements are $J_{ij}^{[m,n]} = 1$. Define furthermore

$$g_A^{-1} \equiv \sum_{i,j} (A^{-1})_{ij} = \mathrm{Tr}(J^{[\ell,\ell]} A^{-1}). \tag{A.1}$$

We wish to express

$$\Gamma_A \equiv \sum_{i,j} [(J^{[\ell,\ell]} + A)^{-1}]_{ij} \tag{A.2}$$

in terms of g_A. To this end we rewrite equation (A.2) as

$$\Gamma_A = \mathrm{Tr}[J^{[\ell,\ell]} A^{-1} (\mathbf{1} + J^{[\ell,\ell]} A^{-1})^{-1}] \tag{A.3}$$

where $\mathbf{1}$ is the unit matrix. An intermediate step of the demonstration consists in noting that

$$J^{[k,l]} C J^{[\ell,m]} = g_{C^{-1}}^{-1} J^{[k,m]} \tag{A.4}$$

where C is a square matrix of dimension $\ell \times \ell$. In equation (A.3) we now expand the argument of the trace in powers of $J^{[\ell,\ell]} A^{-1}$. Iteration of equation (A.4) and use of equation (A.1) lead to

$$\mathrm{Tr}[(J^{[\ell,\ell]} A^{-1})^n] = g_A^{-n} \tag{A.5}$$

whence

$$\Gamma_A = \frac{1}{1 + g_A}. \tag{A.6}$$

When $\ell = 1$ formula (A.6) remains valid for all $A \in \mathbb{R}$ (and in particular for $A = 0$) if for that case one supplements equation (A.1) with $g_A = A$.

A generalized version of the above result is needed in section 5. It involves two invertible matrices A and B of dimensions $\ell \times \ell$ and $m \times m$, respectively, with $\ell, m \geqslant 2$. We now wish to express

$$\Gamma_{AB}(\lambda) \equiv \sum_{i,j} \left[\begin{pmatrix} -A & (1-\lambda)J^{[\ell,m]} \\ (1-\lambda)J^{[m,\ell]} & -B \end{pmatrix}^{-1} \right]_{ij} \tag{A.7}$$

in terms of g_A and g_B. The matrix that appears in equation (A.7) can be decomposed as the sum of two terms,

$$(1-\lambda)\left[J^{[\ell+m,\ell+m]} - \begin{pmatrix} J^{[\ell,\ell]} + \frac{1}{1-\lambda}A & 0 \\ 0 & J^{[m,m]} + \frac{1}{1-\lambda}B \end{pmatrix} \right]. \tag{A.8}$$

After applying equation (A.6) first to the block-diagonal matrix appearing in this decomposition and then to both of its blocks, one obtains

$$\Gamma_{AB}(\lambda) = \frac{2 - 2\lambda + g_A + g_B}{1 - \lambda^2 + g_A + g_B + g_A g_B}. \tag{A.9}$$

By redoing the calculation with ℓ or $m = 1$ one finds that this expression for Γ_{AB} remains valid for those special cases.

The transformed expressions of equations (A.6) and (A.9) are well-suited for taking the limit $A \to 0$ and $B \to 0$ with all their matrix elements proportional to a vanishing parameter, since it is easily seen that then g_A and g_B are also proportional to that parameter.

References

[1] Polya G 1921 *Math. Ann.* **84** 149

[2] Weiss G H 1994 *Aspects and Applications of the Random Walk* (Amsterdam: North-Holland)

[3] Hughes B D 1996 *Random Walks and Random Environments: Random Walks* vol 1 (Oxford: Clarendon Press)

[4] Coutinho K R, Coutinho-Filho M D, Gomes M A F and Nemirovsky A M 1994 *Phys. Rev. Lett.* **72** 3745

[5] Caser S and Hilhorst H J 1996 *Phys. Rev. Lett.* **77** 992

[6] Dvoretzky A and Erdös E 1951 *2nd Berkeley Symp. Math. Stat. Prob.* (Berkeley: University of California Press) p 33

[7] Vineyard G H 1963 *J. Math. Phys.* **4** 1191

[8] Montroll E W 1964 *Proc. Symp. Appl. Math. Am. Math. Soc.* **16** 193

[9] Montroll E W and Weiss G H 1965 *J. Math. Phys.* **6** 167

[10] Jain N C and Pruitt W E 1970 *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **16** 279

[11] Torney D C 1986 *J. Stat. Phys.* **44** 49

[12] Rudnick J and Gaspari G 1986 *J. Phys. A: Math. Gen.* **19** L191

[13] Sciutto S J 1995 *J. Phys. A: Math. Gen.* **28** 3667

[14] Weiss G H and Havlin S 1987 *Phil. Mag.* B **56** 941

[15] Kac M and Luttinger J M 1974 *J. Math. Phys.* **15** 183

[16] Leontovitsh M A and Kolmogorov A N 1933 *Phys. Z. Sowjet.* **4** 1

[17] Le Gall J-F 1988 *Ann. Prob.* **16** 991

[18] Le Gall J-F 1992 *Ecole d'été de probabilités de Saint-Flour (I. Lecture Notes in Mathematics)* vol 1527, ed P L Hennequin (Berlin: Springer)

[19] Berezhkovskii A M, Makhnovskii Yu A and Suris R A 1989 *J. Stat. Phys.* **57** 333

[20] Mountford T S 1989 *Stochastics* **28** 177

[21] Werner W 1994 *Prob. Theor. Relat. Fields* **98** 307

[22] Lawler G 1996 *Elect. Comm. Prob.* **1** 29

[23] den Hollander W Th F and Kasteleyn P W 1982 *Physica* **112A** 523

[24] Weiss G H, Havlin S and Bunde A 1985 *J. Stat. Phys.* **40** 191

[25] Brummelhuis M J A M and Hilhorst H J 1991 *Physica* **176A** 387
Brummelhuis M J A M and Hilhorst H J 1992 *Physica* **185** 35

[26] Zumofen G and Blumen A 1982 *J. Chem. Phys.* **76** 3713

[27] Spitzer F 1964 *Principles of Random Walk* (Princeton NJ: Van Nostrand)
[28] Henyey F S and Seshadri V 1982 *J. Chem. Phys.* **76** 5530
[29] Symanzik K 1969 *Euclidean Quantum Field Theory* ed R Jost (New York: Academic)