

Published in final edited form as:

Phys Rev E Stat Nonlin Soft Matter Phys. 2014 March ; 89(3): 032142.

Information pathways in a disordered lattice

Lawrence R. Frank^{1,2,*} and Vitaly L. Galinsky^{1,3,†}

¹Center for Scientific Computation in Imaging, University of California at San Diego, La Jolla, California 92037-0854, USA

²Center for Functional MRI, University of California at San Diego, La Jolla, California 92037-0677, USA

³Department of ECE, University of California, San Diego, La Jolla, California 92093-0407, USA

Abstract

The maximum entropy random walk in a disordered lattice is obtained as a consequence of the principle of maximum entropy for a particular type of prior information without restriction on the number of steps. This novel result demonstrates that transition probabilities defining the random walk represent a general characterization of information on a defective lattice and does not necessarily reflect a physical process. The localization phenomenon is shown to be a consequence of solution of the Laplacian on the lattice—hence it contradicts the previous interpretation as a spherical Lifshitz state—and naturally generalizes to multiple modes, whose order reflects the significance of information. The dynamics of information flow on the microscale is related to the macroscopic structure of the lattice through a Fokker-Planck formalism. This newly derived theoretical framework is opening doors for a wide range of applications in analysis of (information) flow in disordered systems. That includes potentially breakthrough resolution of the outstanding problem of inferring connectivity from discrete imaging (i.e., neural) data.

I. INTRODUCTION

Theories concerning the diffusive motion of particles have a long and storied history in physics [1], and have been applied to a wide range of physical systems. Of particular interest is diffusion in nonuniform media where particle pathways are obstructed by physical barriers or influenced by distributed sources of attractive or repulsive forces, which we will refer to collectively as “disordered” media. In such cases, characterization of preferable diffusion pathways arises as a problem of both theoretical and practical interest. Recently, this problem has been addressed for the specific case of a random walk on a regular lattice containing random defects (i.e., vertices inaccessible to the particles) by positing that, unlike the standard or generic random walk (GRW) (either uniform or Gaussian), all paths of equal length, regardless of the path details, should be equally probable [2]. It was shown that a transition matrix defining such a process can be constructed in the limit of a large number of steps whose resulting path has maximum entropy per unit step (i.e., maximum entropy

production rate). This transition probability is thus said to define the maximum entropy random walk, or MERW, [2]. The MERW was shown to exhibit an interesting localization phenomenon in regions free of defects, an effect explained by analogy with the potential theory of Lifshitz [3], and it was concluded that the localization occurs in the largest Lifshitz sphere in the lattice, i.e., the largest spherical region in the lattice free of defects.

In this paper, we revisit this problem with three specific goals in mind: (1) derive the MERW transition probability from first principles without the assumption of a large number of steps; (2) characterize the localization phenomenon; and (3) develop a theoretical model for the relationship between the microscopic dynamics of the particles trajectories and the macroscopic structure of the lattice.

The main claim of this paper is that the MERW solution can be viewed as a specific manifestation of a more general result concerning inference on a lattice which has nothing necessarily to do with diffusion, or any other physical process. The general approach results in a new theoretical framework suitable for application to a wide range of problems involved with analysis of disordered lattice systems. As a byproduct, we show that the previous interpretation of the localization phenomenon by the Lifshitz sphere argument is not true in general. Moreover, the correct general solution we derive elucidates the source of the localization phenomenon, demonstrates that it actually occurs on multiple scales, and paves the way for the classification of optimal pathways of information in a lattice. With these considerations in mind, we call this theory *entropy spectrum pathways*, or ESP.

II. THE DEFECTIVE LATTICE

In this paper we will consider the random walk on a defective lattice, a schematic of which is shown in Fig. 1. A random walk is defined by the simple rule that at each time step a particle at location (i, j) can move one step in either the i or j direction. In a regular lattice all lattice sites are available, as illustrated in Fig. 1(a). In a defective lattice, certain sites are inaccessible, as shown in Fig. 1(b).

III. RANDOM WALK TRANSITION PROBABILITY

Consider a two-dimensional (2D) Cartesian grid of equally spaced points at R spatial locations $(x_1, \dots, x_R)$ where each point can take on any one of s available values $\{v_1, \dots, v_s\}$. To simplify the notation, we will equate the spatial path with the sequence of values and speak of the trajectory as the sequence of values along the spatial path

$\gamma_{ba}^n = \{v_{x_0}, v_{x_1}, \dots, v_{x_n}\}$, and compute the probability of this sequence $p(\gamma_{ba}^n | I) = p(v_{x_0}, v_{x_1}, \dots, v_{x_n} | I)$ given our prior information I .

The logical procedure for determining the path probabilities is the principle of maximum entropy [4] in which the Shannon information entropy [5]

$$S(\gamma) = - \sum_{\{\gamma\}} p(\gamma) \ln p(\gamma) \quad (1)$$

is maximized subject to the constraints imposed by the the basic rules of probability theory and by the s data points u_k for $k = 1, \dots, s$:

$$p(\gamma) \geq 0, \quad (2a)$$

$$\sum_{\{\gamma\}} p(\gamma) = 1, \quad (2b)$$

$$\sum_{\{\gamma\}} p(\gamma) u_k(\gamma) = \langle u_k \rangle \equiv U_k, \quad (2c)$$

where the U_k are the expected values of the data along the set of all possible paths $\{\gamma\}$. This is a variational problem solved by the method of Lagrange multipliers and has the general solution [4]

$$p(\gamma) = Z^{-1} \exp[-\Lambda(\gamma)], \quad (3)$$

where

$$\Lambda(\gamma) = \sum_{k=1}^s \lambda_k u_k(\gamma), \quad (4)$$

and the partition function is

$$Z(\lambda_1, \dots, \lambda_s) = \sum_{\{\gamma\}} \exp[-\Lambda(\gamma)]. \quad (5)$$

The Lagrange multipliers $\{\lambda_k\}$ are determined from the data by

$$\langle u_k \rangle = - \frac{\partial}{\partial \lambda_k} \ln Z(\lambda_1, \dots, \lambda_s), \quad (6)$$

for $k = 1, \dots, s$, and the fluctuations are determined from

$$\langle u_k u_l \rangle - \langle u_k \rangle \langle u_l \rangle = - \frac{\partial^2}{\partial \lambda_k \partial \lambda_l} \ln Z(\lambda_1, \dots, \lambda_s). \quad (7)$$

The entropy Eq. (1) of the maximum entropy distribution Eq. (3) is

$$S = \ln Z + \sum_k \lambda_k U_k. \quad (8)$$

The reformulation of the path probabilities in terms of the maximum entropy formalism, as expressed by Eqs. (3)–(8), allows the construction of path probabilities consistent with given prior information. We now consider two different sets of prior information and show how these lead to the GRW and MERW, respectively. We stress here the fact that the dependence of the derived distribution $p(\gamma)$ on the prior information means that it is an expression of our processing of information, rather than of a physical effect. Moreover, while the specific

examples addressed in this paper are confined to prior information about nearest neighbor coupling, the formalism is very general and can incorporate prior information about more complex couplings.

A. Prior information about node degrees

Suppose that the known values v_i represent the node degrees d and that the prior information consists of the frequency f_i with which each value d_i occurs. If

N_i = number of times d_i appears along γ ,

f_i = expected frequencies of d_i ,

then

$$\langle N_i \rangle = \sum_{\{\gamma\}} N_i(\gamma) p(\gamma) = N f_i, \quad i=1, \dots, s, \quad (9)$$

where $N = \sum_i N_i$ is the total number of sites visited. The path probability $p(\gamma)$ is then found by maximizing Eq. (1) subject to Eqs. 2(a)–2(c) [in the form of Eq. (9)]. The solution to this is

$$p(\gamma) = \frac{1}{Z} \exp \left[- \sum_{i=1}^s \lambda_i N_i(\gamma) \right], \quad (10)$$

where the partition function is

$$Z(\lambda_i) = \sum_{\{\gamma\}} \exp \left[- \sum_{i=1}^s \lambda_i N_i(\gamma) \right] = z^N \quad (11)$$

and

$$z = \sum_{i=1}^s e^{-\lambda_i}. \quad (12)$$

From Eqs. (6) and (9),

$$\lambda_i = -\ln(z f_i), \quad 1 \leq i \leq s, \quad (13)$$

which, when substituted into Eq. (3) and properly normalized, gives the multinomial distribution [6]

$$p(\gamma) = \prod_i \frac{1}{d_{x_i}} = \prod_{j=1}^s \left(\frac{1}{d_j} \right)^{N_j} \quad (14)$$

For a 2D Cartesian lattice $s = 4$. The number of different paths for specified N_i is $N!/(N_1! \cdots N_s!)$.

Equation (14) says that the probability of any path only depends on how many times the values $\{v_i\}$ appear along the path, but not on the order in which they appear. Thus the GRW can be viewed as the maximum entropy solution when the prior information is limited to the frequency of occurrences of the defects.

B. Prior information about the coupling of the available values $\{v\}$

Suppose now that our prior information consists of the frequency f_{ij} with which the pairs of value $v_i v_j$ occur together. At the end, we will consider the special case in which this information is reduced to whether or not location i and j are connected, so that the prior information is just the adjacency matrix. Now we consider the more general case where

N_{ij} = number of times $v_i v_j$ appears along γ ,

f_{ij} = expected frequencies of the pairs $v_i v_j$,

and the f_j are known (they are again the prior information), then

$$\langle N_{ij} \rangle = \sum_{\{\gamma\}} N_{ij}(\gamma) p(\gamma) = (n-1) f_{ij}, \quad \{i, j\} = 1, \dots, s, \quad (15)$$

where $n = \sum_{ij} N_{ij}$ is the total number of jumps between sites, and thus the trajectory length, and again $\{\gamma\}$ denotes the set of all possible paths γ . In the path γ the number of times the pair $x_i x_j$ appears is

$$N_{ij}(\gamma) = \sum_{k=1}^{n-1} \delta_{i,k} \delta_{j,k+1} \quad (16)$$

where δ represents the Dirac delta function: $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ for $i \neq j$.

This problem is logically identical to the problem of digram frequencies in communication theory addressed by Jaynes [7]. The path probability $p(\gamma)$ that has maximum entropy subject to the constraint Eq. (15) has the solution

$$p(\gamma) = \frac{1}{Z} \exp \left[- \sum_{i,j=1}^s \lambda_{ij} N_{ij}(\gamma) \right], \quad (17)$$

where the partition function is

$$Z(\lambda_{ij}) = \sum_{\{\gamma\}} \exp \left[- \sum_{i,j=1}^s \lambda_{ij} N_{ij}(\gamma) \right]. \quad (18)$$

This complicated sum over all the different paths γ is simplified by noting that this partition function can be rewritten in terms of a matrix product

$$Z(\lambda_{ij}) = \sum_{i,j=1}^s [Q^{(n-1)}]_{ij}, \quad (19)$$

where the matrix Q is defined as

$$Q_{ij} = e^{-\lambda_{ij}}. \quad (20)$$

This matrix defines the interactions between locations on the lattice and so will be called the *coupling matrix*. As we show later, the Lagrange multiplier λ_{ij} that define the interactions can be seen as local potentials that depend on some function of the spatial locations x_{ij} on the lattice. We will suppress the more complete notation $\lambda_{ij}(x_{ij})$, and thus $Q(x_{ij})$, for clarity.

A useful trick to simplify the computation of the partition function [7] is to add the step (x_n, x_1) to the pathway, which adds another $\exp(-\lambda_{ij})$ to the partition function Eq. (18) and creates periodic boundary conditions. This modifies Eq. (19) to

$$Z(\gamma_{ij}) = \sum_{i,j=1}^s [Q^n]_{ij} = \text{Tr}(Q^n) = \sum_{k=1}^s q_k^n, \quad (21)$$

where $\{q_k\}$ are the roots of $|Q_{ij} - q\delta_{ij}|$. This trick is justified in the limit of long trajectories $n \rightarrow \infty$. The probability of the entire path, Eq. (17) can be written using Eqs. (16) and (20),

$$p(\gamma_{ab}|I) = Z^{-1} Q_{x_1, x_2} Q_{x_2, x_3} \cdots Q_{x_{n-1}, x_n}, \quad (22)$$

where the periodic boundary conditions trick has been invoked.

While Eq. (17) is formally the solution of the path probability, we would like to determine the transition probability. In order to do this, we can consider the problem of how our estimates change as we move along a path. In other words, if we have moved part way along a path, what does this tell us about the remainder of the path? This is analogous to the partial message problem [7]. To address this question, imagine that we break the path γ_{ab} from an initial point a to a final point b into two segments (Fig. 3) defined by some intermediate point $c = x_{m-1}$ (i.e., $a \leq c \leq b$), so that the first segment is of length $m - 1$ and the second is of length $n - m + 1$:

$$\gamma_{ac}^{(m-1)} = v_{x_1} v_{x_2} \cdots v_{x_{m-1}}, \quad (23a)$$

$$\gamma_{cb}^{(n-m+1)} = v_{x_m} v_{x_{m+1}} \cdots v_{x_n}. \quad (23b)$$

The probability of the entire path is just the joint probability of the two path segments $\{\gamma_{ac}, \gamma_{cb}\}$ and from the basic rules of probability theory is equal to Eq. (17):

$$p(\gamma_{ab}|I) = p(\gamma_{ac} \gamma_{cb}|I) = p(\gamma_{cb}|\gamma_{ac}, I) p(\gamma_{ac}|I), \quad (24)$$

so the conditional probability of γ_{cb} , given γ_{ac} , is

$$p(\gamma_{cb}|\gamma_{ac}, I) = \frac{p(\gamma_{ac} \gamma_{cb}|I)}{p(\gamma_{ac}|I)} = \frac{p(\gamma_{ab}|I)}{p(\gamma_{ac}|I)}. \quad (25)$$

The marginal distribution of initial part of the path, $p(\gamma_{ac}|I)$, is

$$p(\gamma_{ac}|I) = \sum_{\gamma_{bc}} p(\gamma_{ab}|I) = \sum_{x_m=1}^s \cdots \sum_{x_n=1}^s p(x_1 \cdots x_n|I) \quad (26)$$

which, from Eq. (22), is

$$p(\gamma_{ac}|I) = R \sum_{x_m=1}^s \cdots \sum_{x_n=1}^s Q_{x_{m-1}, x_m} \cdots Q_{x_{n-1}, x_n}, \quad (27)$$

where

$$R = Z^{-1} Q_{x_1, x_2} \cdots Q_{x_{m-2}, x_{m-1}}. \quad (28)$$

Define the transition point from the initial path to the second path as ij , where $i = x_{m-1}$ is the last point in the first path and $j = x_m$ is the first point in the second path. Just as in the step from Eqs. (18) to (19), the sum over paths in Eq. (27) can be written as a matrix product:

$$p(\gamma_{ac}|I) = R \sum_{k,l=1}^s Q_{ik} [Q^{(n-m)}]_{kl} = R \sum_k Q_{ik} T_k, \quad (29)$$

where

$$T_k = \sum_{l=1}^s [Q^{(n-m)}]_{kl}. \quad (30)$$

The conditional probability distribution of the second part of the path, given the first part [Eq. (25)] is then, from Eqs. (22) and (29),

$$p(\gamma_{cb}|\gamma_{ac}, I) = \frac{Q_{ij} Q_{x_m, x_{m+1}} \cdots Q_{x_{n-1}, x_n}}{\sum_{k=1}^s Q_{ik} T_k} \quad (31)$$

since the common factor R cancels. This distribution represents a Markov chain because the probability for the second path $\{x_m \cdots x_n\}$ depends only on the previous location x_{m-1} and not on any of the details of the path the particle took to get to that point. From Eq. (31) we can determine the probability that the path switches from the first path at $i = x_{m-1}$ to the second path at $j = x_m$. This is called the *transition probability* and is found from the basic rules of probability by summing Eq. (31) over the locations that are not of interest:

$$\begin{aligned} p_{ij} &\equiv p(x_m|x_{m-1}, I) = \sum_{x_{m+1}} \cdots \sum_{x_n} p(\gamma_{cb}|\gamma_{ac}, I) \quad (32) \\ &= \frac{Q_{ij}}{\sum_{k=1}^s Q_{ik} T_k} \left(\sum_{x_{m+1}} \cdots \sum_{x_n} Q_{j, x_{m+1}} \cdots Q_{x_{n-1}, x_n} \right), \quad (33) \end{aligned}$$

where the term in parentheses is just T_j of Eq. (30). Thus the transition probability is

$$p_{ij}^{(n,m)} = \frac{Q_{ij} T_j}{\sum_{k=1}^s Q_{ik} T_k}, \quad (34)$$

where the superscript notation is to remind us of the dependency on both n and m . This result was previously derived in the context of communication theory.¹ This represents the maximum entropy transition probability between location $i = x_{m-1}$ and location $j = x_m$ for a path of length n .

Having derived the general case Eq. (34), the limiting case for $n \rightarrow \infty$ can be determined [7]. The term containing both m and n is T_k [Eq. (30)], so we look at that first. The matrix Q can be reduced to block diagonal $\bar{Q}$ form as there exists a nonsingular matrix B for which $\bar{Q} = B^{-1} Q B$ so that the powers of the matrix Q can be expressed as

$$Q^n = B \bar{Q}^n B^{-1}, \quad (35)$$

so as $n \rightarrow \infty$ the element(s) q_1^n of Q^n dominate all others. In general, the roots $\{q_1, q_2, \dots, q_r\}$ (assumed to be arranged in the order $|q_1| \geq |q_2| \geq \dots \geq |q_r|$) of the characteristic equation $D(q) = \det(Q_{ij} - q \delta_{ij})$ can be degenerate and complex. However, if q_1 is nondegenerate and real, then from Eqs. (30) and (35)

$$\lim_{n \rightarrow \infty} T_j = q_1^{(n-m)} \psi_{1j} \sum_{k=1}^s (B^{-1})_{1k}. \quad (36)$$

$\psi_1 \equiv B_1$ is an eigenvector of Q (the one with the largest eigenvalue) and ψ_{1i} is the i th component of ψ_1 . The denominator of Eq. (34) then contains a term

$$\sum_{k=1}^s Q_{ik} T_k = \sum_{k=1}^s Q_{ik} \psi_{1k} = q_1 \psi_{1i}. \quad (37)$$

Using this and canceling common factors, the transition probability Eq. (34) in the limit of large n becomes [7]

$$p_{ij}^{(\infty)} = \frac{Q_{ij} \psi_{1j}}{q_1 \psi_{1i}} \quad (38)$$

where q_1 is the maximum eigenvalue of Q and ψ_{1i} is the i th element of the eigenvector ψ_1 of Q associated with the maximum eigenvalue.

It is useful at this junction to recall the parameter dependencies in Eq. (38). As noted above, the coupling matrix depends on the spatial locations $Q(x)$ through the Lagrange multipliers. Thus, so do the eigenvectors $\psi = \psi(x)$ and the associated eigenvalues $q(x)$. In the examples shown below, the spatial dependence of these quantities is what produces the distribution maps directly from the eigenstructure of the lattice.

¹In Ref. [7], note typographical error in Eq. 20–61 (22.74) for the transition probability.

Equation (38) looks similar to the expression for the transition matrix *a priori* introduced in Ref. [2] [Eq. (5)] but with several important differences. First of all, rather than being postulated, it was obtained as a limit $n \rightarrow \infty$ from a more general expression, Eq. (34). The derivation of Eq. (34) itself is general and depends on the sequence length n and the transition point x_m , both of which may be of arbitrary length (provided $m < n$). Moreover, it is not required that Q represent an adjacency matrix. If the Lagrange multipliers take the form

$$\lambda_{ij} = \begin{cases} 0 & \text{connected} \\ \infty & \text{not connected} \end{cases} \Rightarrow Q_{ij} = e^{-\lambda_{ij}} = \begin{cases} 1 \\ 0 \end{cases} \quad (39)$$

then Q becomes an adjacency matrix and Eq. (38) is identical to the expression [Eq. (5)] in Ref. [2]. Taking the Lagrange multipliers as “potentials” [8], Eq. (39) can be viewed as representing local potentials that are either completely attractive ($\lambda = 0$) or completely repulsive ($\lambda = \infty$).

The entropy of the maximum entropy distribution Eq. (8), in the limit $n \rightarrow \infty$ can be obtained using the expression for the partition function [7],

$$Z = \sum_{i,j=1}^s [Q^n]_{ij} = \text{Tr}(Q^n) = \sum_{k=1}^s q_k^n, \quad (40)$$

where $\{q_k\}$ are the roots of $|Q_{ij} - q\delta_{ij}|$. Taking $\lim_{n \rightarrow \infty} Z = q_1^n$ and using Eq. (15), the entropy per step becomes, from Eq. (8),

$$\frac{S}{n} = \ln q_1 + \sum_{ij} \lambda_{ij} f_{ij}. \quad (41)$$

From Eq. (39), for the connected components $\lambda_{ij} = 0$, in which case Eq. (41) becomes $S/n = \ln q_1$, which is the same as the limit given by Burda *et al.* But we see that, in general, Eq. (41) is the correct limit.

IV. LOCALIZATION

In the special case that Q reduces to the adjacency matrix A [Eq. (39)], an interesting property of the transition probability $p_{ij}^{(\infty)}(\gamma)$ in the limit $n \rightarrow \infty$ (the equilibrium transition probability) noted by Burda [2] is that it localizes in what appears to be the largest accessible region of a defective lattice. They explained this effect by reformulating the problem in terms of a Hamiltonian equation, then making the analogy with Lifshitz spheres [9], defined as the largest spherical region of the lattice that is free of defects [3]. We show here that this view is not correct in general.

A. Potential theory

It has been noted that the spatial distribution of the equilibrium probability density is described by the eigenvector $\psi^{(1)}$ associated with the maximum eigenvalue q_1 , and thus localization can be investigated by looking at the structure of $\psi^{(1)}$ [2]. While it is possible to

work directly in the eigencoordinates of the adjacency matrix $\mathbf{A}\psi^{(k)} = q_k\psi^{(k)}$, it is useful and common to recast this in the form of a differential equation by noting that the adjacency matrix is related to the graph Laplacian by $\mathbf{L} = \mathbf{D} - \mathbf{A}$. The elements of the diagonal degree matrix are $D_{ii} = d_i$, where d_i is the vertex degree, and thus

$$\mathbf{L}\psi^{(k)} - \mathbf{D}\psi^{(k)} = -q_k\psi^{(k)}. \quad (42)$$

In an undirected graph where edges have no orientation (which is all we will consider here) the degree is the number of edges incident to the vertex [10]. For graphs that in the absence of defects are regular every vertex has the same degree $d_{\max}$. Then vertices with defects have $d = 0$ and those with $d < d_{\max}$ are adjacent to defects. Adding $d_{\max}\psi^{(k)}$ to each side of Eq. (42), and noting that the graph Laplacian is the negative of the Laplacian operator Δ for the Dirichlet boundary conditions considered here [11], yields the differential equation

$$-\Delta\psi + V\psi = E_k\psi, \quad (43)$$

where $V_j = d_{\max} - d_j$ is the *potential* and $E_k = d_{\max} - q_k$ is the *energy*. The potential V is a vector of length $n = \text{length}(\psi)$, and $V\psi$ in Eq. (43) is an n -dimensional vector whose j th element is $V_j\psi_j$. Spatial variations in the potential are thus encoded through the components V_j . Equation (43) has the familiar form of a Hamiltonian equation $\mathcal{H}\psi = E_k\psi$ where $\mathcal{H} = -\Delta + V$. The addition of $d_{\max}\psi$ to both sides of Eq. (42) allows the interpretation of $\psi^{(1)}$ in Eq. (43) as the ground state wave function [2], since $E_1 = d_{\max} - q_1$ is the lowest energy because q_1 is the largest eigenvalue.

B. Graph eigenstructure

While the spatial distribution of the equilibrium MERW probability is encoded in ψ_1 , the higher order MERW eigenfunctions convey important information, as we shall demonstrate. Thus while it is possible to examine Eq. (43) in the context of Lifshitz potentials ([2]), it is perhaps more illuminating to recognize that this equation expresses the fact that the eigenvectors of the adjacency matrix are the different energy modes of the Laplacian with boundary conditions determined by the potentials. This viewpoint permits a clear understanding of the localization phenomenon, and will further inform our understanding of the dynamics.

To illustrate this view, we revisit the localization examples presented in Ref. [6] of the disk and the ellipse. Interestingly, for a lattice containing multiple well separated connected regions, as in Fig. 4(a), the eigenvectors of the adjacency matrix are the eigenvectors of the individual connected regions, ranked according to eigenvalue. This is shown in Fig. 4 where we have specifically chosen the eigenvalues of the spherical region to be $\lambda = 10$ pixels, and those of the elliptical region to be $\{\lambda_1, \lambda_2\} = \{9, 14\}$ so that the largest spherical region is the disk, but the largest eigenvalue of the adjacency matrix belongs to the lowest mode of the elliptical region. This contradicts the claim by Burda that the maximum entropy solution is the one in the largest spherical region [2]. The first eigenvalue [Fig. 4(b)] determines the maximum entropy solution p^∞ , which is evidently not determined by the largest spherical region [2] but rather the largest eigenvalue of the Laplacian.

Thus there are in fact *multiple* localization regions within the lattice, ranked according to the corresponding eigenvalues. It is therefore the *spectrum* of the maximum entropy eigenvectors, in descending order of the associated eigenvalues, that describes the information flow in the lattice. This flow occurs via a multitude of paths over multiple spatial scales of the lattice. We call this characterization *entropy spectrum pathways*, or ESP. In practical applications, the lattice can be described in terms of m pathways constructed from the first m eigenvectors of the adjacency matrix (in decreasing order of the eigenvalues) which, from Eq. (38), is

$$p_{ij}^{(\infty)} = \sum_{k=1}^m p_{ijk}^{(\infty)}, \quad (44a)$$

$$\text{where } p_{ijk}^{(\infty)} = \frac{Q_{ij} \psi_j^{(k)}}{q_k \psi_i^{(k)}}. \quad (44b)$$

For each transition matrix Eq. (44) there is a unique stationary distribution associated with each path k ,

$$\mu^{(k)} = [\psi^{(k)}]^2, \quad (45)$$

that satisfies

$$\mu_i^{(k)} = \sum_j \mu_j^{(k)} p_{ijk}^{(\infty)}, \quad (46)$$

the first of which, $\mu_i^{(k)}$, corresponds to the maximum entropy stationary distribution [2].

The localization phenomenon in a random lattice can now be made clear by combining a random lattice with the perfect disk and ellipse of Fig. 4(a), as shown in Fig. 5(a). The eigenvectors, shown in Fig. 5, are then distorted versions of the idealized lattice in Fig. 4 caused by the alteration in the boundary conditions.

The transition probabilities Eqs. (38) and (44) determine the dynamics towards the equilibrium distribution p_t^∞ through the update formula

$$p_{t+1} = p_{ij} p_t. \quad (47)$$

In the lattice Fig. 6(a) [with associated lattice degrees, Fig. 6(b)], a point distribution [Fig. 6(c)] for the GRW [Figs. 6(e) and 6(f)] evolves into a relatively uniform spatial distribution, disrupted only locally by the defects. The MERW [Figs. 6(g) and 6(h)], on the other hand, “flows” to the equilibrium distribution. We emphasize that *both* sets of Figs. 6(e)–6(h) are generated by maximum entropy distributions, but with different prior information. The marked difference in these dynamics is purely a consequence of different prior information and, as such, can be viewed as a model of *information flow*, rather than the realization of a physical process. Moreover, while the dynamics in Fig. 6 were generated only from the

primary eigenvector [Fig. 7(a)], in lieu of Eq. (44b), there exist multiple levels of information and associated pathways, as demonstrated in Figs. 7(b)–7(d).

V. DYNAMICS

A. Dynamics of the most probable pathways

While Eq. (47) provides a method to compute the information flow in the lattice, it provides little insight into how the macroscopic structure of the pathways is related to the microscopic dynamics of the information flow. One possibility to introduce this relation (due to Jaynes albeit in a highly abbreviated form) is “bubble dynamics” [12], in which the spatial-temporal characteristics of a probability density $P(x_1 \cdots x_m; t)$ of a set of macroscopic variables $x_i, i = 1, \cdots, m$ is characterized by the conservation of probability

$$\partial_t P + \nabla \cdot J_I = 0, \quad (48)$$

where the information flux J_I is the sum of a diffusive component $J_d = -D \nabla P$ and a convective component $J_c = -L P \nabla S$,

$$J_I = J_d + J_c, \quad (49)$$

and $S(x) = k \ln W(x)$ is the entropy in which $W(x)$ is the density of states, D is the diffusion coefficient (or, more generally, the diffusion tensor), and $L = \kappa D$ (where $\kappa \equiv k^{-1}$ is the Onsager coefficient [13]). Here x refers to spatial coordinates, so k , which is Boltzmann’s constant in thermodynamics, just scales the entropy to the macroscopic variable space. Onsager coefficients are thus diffusion coefficients scaled to the spatial coordinates. Substitution of Eq. (49) into Eq. (48) gives

$$\partial_t P + L \nabla \cdot (P \nabla S) = D \nabla^2 P. \quad (50)$$

This is the Fokker-Planck equation with the potential equal to the entropy: $V = S$, and connects the global structure of the probability with the local structure of the lattice through the local structure of the entropy. Equation (50) was previously derived (in a slightly different form) in Ref. [14]. It can be shown that Eq. (50) accurately describes the dynamics of the ESP, such as that illustrated in Fig. 6, reproducing not only the accurate final ESP distribution but the flow of information through the lattice from an initial point distribution. This formulation can show that information flow occurs not only over different *spatial* scales, but over different *temporal* scales as well.

B. Construction of path entropy

In order to investigate the dynamics via Eq. (50) we need to construct the entropy S . Having determined the maximum entropy transition matrix p_{ij}^∞ [Eq. (38)] between an initial point i and a final point j on the lattice, we want to construct the entropy map by calculating the entropy for *every* path x_{ij} between these two points. This amounts to calculating the matrix

$$S_{ij} = - \sum_{\{x_{ij}\}} p(x_{ij}) \ln p(x_{ij}). \quad (51)$$

We then utilize a theorem by Ekroot [15] to construct the entropy map for all paths between a specified initial and final lattice locations. This theorem demonstrates that the matrix Eq. (51) can be computed directly from the transition matrix p_{ij} and the equilibrium distribution μ from the expression

$$S = K - \tilde{K} + S_{\Delta}, \quad (52)$$

where $K = (I - P + B)^{-1}(S^* - S_{\Delta})$ in which I is the identity matrix, P is the transition matrix, and $\tilde{K}_{ij} = K_{jj}$, $B_{ij} = \mu_j$, $S_{ij}^* = S(p_{ij})$, and

$$(S_{\Delta})_{ij} = \begin{cases} h/\mu_i & i=j, \\ 0 & i \neq j, \end{cases} \quad (53)$$

where h is the entropy per step in the limit $n \rightarrow \infty$,

$$h = - \sum_{i,j} \mu_i p_{ij} \ln p_{ij}, \quad (54)$$

and $S(p_i)$ is the entropy of the first step of a trajectory initially at location i , given by

$$S(p_i) = - \sum_j p_{ij} \ln p_{ij}. \quad (55)$$

The columns of S correspond to spatial maps of maximal entropy pathways from each point in the image to the target points and thus reveal preferred pathways throughout the image volume. This procedure can be done for any other of the k modes using p_{ijk}^{∞} [Eq. (44)].

Using this construction, the path entropy $S(x, y)$ [Eq. (51)] from the initial distribution location, Fig. 6(c), to every other location can be determined [Fig. 8(a)] and from this can be determined the first and second spatial derivatives [Eqs. (8b) and (8c)].

The time-varying distribution $P(x, y, t)$ for the path entropy in Fig. 8 is shown in Fig. 9. The starting distribution follows the maximum entropy path shown in Eq. (6). The initially localized distribution moves and spreads in accordance with the local entropy field structure, then stalls and tightens at the maximum entropy location [the dark red region in Fig. 8(a)], and the location of the highest probability concentration of p_t^{∞} [Fig. 6(d)]. Further details of this formulation will be presented elsewhere.

VI. APPLICABILITY

The presented formalism can be used for finding static relations and for assessing dynamical information flow in many real world situations. With the ever-increasing number of applications in which connectivity plays a critical role (social networks, brain function and structure, etc.), methods for quantitative assessment of connectivity measures will play an increasingly significant role in a wide range of applications.

As an illustration of possible applications, we have included one practical example of ESP processing of magnetic resonance diffusion tensor imaging (MR-DTI) data. DTI data is

often used for neural fiber tractography in the studies of brain connectivity. This is a complex and severely ill-posed problem. Within an imaging volume, local (voxel) DTI data measurements are used to reconstruct a (possibly high dimensional) tensor in each voxel that is able to capture some broad aspects of the underlying tissue microstructure, but on a scale much greater than the fibers themselves. From these tensor estimates are reconstructed the purported pathways of neural fiber bundles throughout the brain that produced the underlying variations in the diffusion signal. Imaging resolution is never (currently) fine enough to resolve individual fibers, and thus individual voxel measurements are degraded by averaging over fiber bundles, possibly at different orientations, and other tissue compartments. Given the great complexity of the neural structure of the human brain, reconstruction of the macroscopic neural pathways from large volumes of noisy, highly multidimensional tensors derived from measurements of microscopic signal variations poses a significant theoretical and computational challenge.

The reconstruction of the macroscopic neural fiber pathways from the microscopic measurements of the local diffusion from DTI data is precisely the type of problem suited for the ESP formalism. The goal is to determine the most probable global pathways (neural fibers) consistent with measured values (diffusion tensors) based upon the available prior information. The ESP formalism provides a general method for the incorporation of prior information regarding the relationship between voxels. For the current paper, we limit the demonstration to the nearest neighborhood coupling discussed in detail above, though we stress that this is but one possible realization of the method. For the nearest neighbor coupling, the local potential can be derived from the interaction of the tensors, which is chosen here to be their inner product.

A complete details of implementation, including computation of diffusion tensors, generation of fractional anisotropy ($F A$) map, assignment of the potential matrix [Eq. (20)] with an appropriate choice of coefficients and thresholds will be deferred to a more specialized publication. We include here only a short comparison of the final trajectory generated between two chosen points by ESP [Fig. 10(c)] and GRW [Fig. 10(b)] (using the same number of time steps $n_t = 500$). A composite map of $F A$ overlayed with the principal eigenvectors is shown for a single slice in Fig. 10(a).

The presented example clearly shows a “global” nature of the ESP method, in the sense that it probes the most probable of all possible paths between the two points and the optimization is based on the entropy of the entire path, which depends upon all of the possible connections in all of the possible paths. One important advantage as demonstrated here is that the neighborhood of the path is explicitly taken into account. Different coupling schemes can produce different trajectory calibers. The method is quite general and can incorporate more sophisticated models of both intervoxel diffusion anisotropy, such as high angular resolution reconstruction [20], and intravoxel coupling schemes, such as long range correlations. An initial implementation of this method was presented in Ref. [21] and a more detailed paper is forthcoming.

VII. CONCLUSION

In this paper we have demonstrated that logical inference concerning the spatial-temporal characteristics of probabilities on a defective lattice provides a theoretical justification for both the uniform random walk and the so-called maximum entropy random walk [2,6]. Within this framework, both are maximum entropy distributions, but with different prior information.

We derived from first principles the transition probability for probabilities on a defective lattice with a known coupling matrix for the general case of any partial path length m within a path of total length n . This was shown to be expressed in terms of a potential theory through the coupling matrix Q . It was then shown in the limit of a large number of step and a binary coupling scheme, where Q becomes the adjacency matrix, to reduce to the so-called maximum entropy random walk of Burda *et al.* [2,6].

We then demonstrated that the localization phenomena can be understood in terms of the graph Laplacian and the eigenstructure of the adjacency matrix, and depends upon the ranking of the eigenvalues of the adjacency matrix, and thus does not (necessarily) localize in the largest Lifshitz sphere, as previously claimed [2]. Moreover, the complexity of the graph eigenstructure in a lattice with random defects may possess several close localization modes.

Finally, we sketched a theoretical framework within which to understand the relationship between the microscopic (local) transition probabilities and the macroscopic prediction of the lattice structure, the probability concentration, and dynamics governing the infiltration and flow of information through the lattice. The general nature of the results may prove useful for the investigation of a variety of experimental situations where data can be modeled as a grid of discrete measurements from which one seeks to reconstruct an underlying continuous structure or understand how information flows within complex networks.

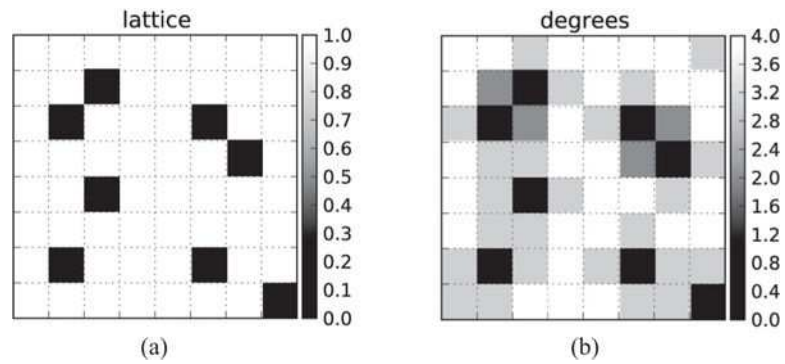
The presented framework has potential application in a wide area of problems dealing with analysis of disordered systems. In particular, it is ideally suited for addressing two challenging problems related to understanding the form and function of the human brain: inferring neural structural connectivity from the local (voxel) data provided by diffusion tensor (DT) MRI data, and inferring functionally connected regions from spatial-temporal activations from functional MRI (fMRI) data. Both techniques offer the potential for providing unique data to elucidate the structure and function of the human brain, but are hampered by the great difficulty in inferring global connectivity from local measurements in such a highly complex structural and functional system. The numerical implementation of the theory presented here provides a means by which to infer such connections and thus offers the possibility of providing unique insights into the structure and function of the human brain. These applications are currently being pursued in our laboratory.

Acknowledgments

We thank D. Meyer and D. Arovos for useful discussions. L.R.F. and V.L.G. were supported by NSF Grants No. DBI-1143389, No. DBI-1147260, No. EF-0850369, and No. PHY-1201238, and by NIH Grant No. R01 MH096100.

References

1. Einstein A. Ann Phys (NY). 1905; 322:549. Ann Phys (NY). 1906; 324:289.
2. Burda Z, Duda J, Luck JM, Waclaw B. Phys Rev Lett. 2009; 102:160602. [PubMed: 19518691]
3. Lifshitz I. Adv Phys. 1964; 13:483. Sov Phys Usp. 1965; 7:549.
4. Jaynes E. Phys Rev. 1957; 106:620. Phys Rev. 1957; 108:171.
5. Shannon C. Bell Syst Tech J. 1948; 27:379.
6. Burda Z, Duda J, Luck JM, Waclaw B. Acta Phys Pol B. 2010; 41:949.
7. Jaynes, E. Probability Theory with Applications in Science and Engineering. Washington University Press; St Louis: 1974. Probability Theory: The Logic of Science. Cambridge University Press; New York: 2003.
8. Jaynes, E. In: Smith, C.; Grandy, W., editors. Proceedings of the Second Workshop on Maximum Entropy and Bayesian Methods; Laramie. 1982; Holland: D Reidel; 1982.
9. Friedberg R, Luttinger JM. Phys Rev B. 1975; 12:4460. Kirsch W, Martinelli F. Commun Math Phys. 1983; 89:27.
10. Diestel, R. Graph Theory. 3. Springer-Verlag; Berlin: 2005.
11. Ding, C.; Simon, HD.; Jin, R.; Li, T. Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; San Jose. 2007; New York: ACM; 2007. p. 260-269.
12. Jaynes, E. Complex Systems - Operational Approaches in Neurobiology, Physics, and Computers. Haken, H., editor. Springer-Verlag; Berlin: 1985. p. 254-269.
13. Onsager L. Ann NY Acad Sci. 1945; 46:241. [PubMed: 21024247]
14. Grabert H, Green MS. Phys Rev A. 1979; 19:1747.
15. Ekroot L, Cover TM. IEEE Trans Inf Theory. 1993; 39:1418.
16. Reese T, Heid O, Weisskoff R, Wedeen V. Magn Reson Med. 2003; 49:177. [PubMed: 12509835]
17. Jones D, Horsfield M, Simmons A. Magn Reson Med. 1999; 42:515. [PubMed: 10467296]
18. Jezzard P, Balaban RS. Magn Reson Med. 1995; 34:65. [PubMed: 7674900]
19. Andersson J, Skare S. NeuroImage. 2002; 16:177. [PubMed: 11969328]
20. Frank L. Magn Reson Med. 2002; 47:1083. [PubMed: 12111955]
21. Frank, L.; Meyer, D. Proceedings of the 20th Annual Meeting. Society of Magnetic Resonance in Medicine; Melbourne: 2012. p. 1909

**FIG. 1.**

A defective two-dimensional lattice. White squares are accessible, black squares (defects) are not (a) A defective lattice. (b) The lattice degree d at each site.

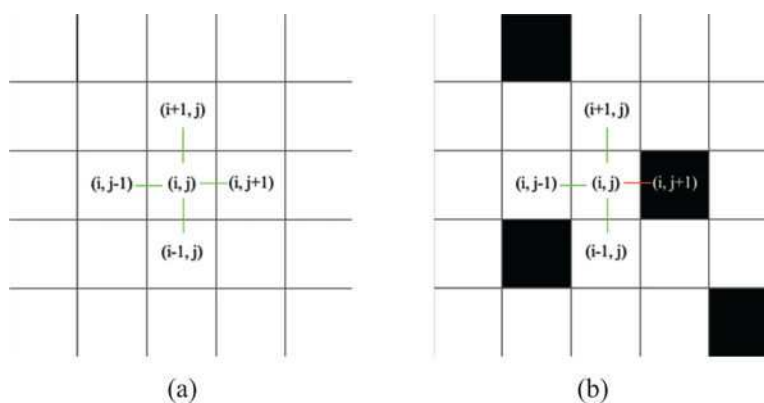


FIG. 2. (Color online) Random walk on a lattice. Allowed steps from the location (i, j) . (a) Generic random walk on a regular lattice. All adjacent locations are accessible. (b) Random walk on a defective lattice. Allowable steps are shown in green (from white to white square), disallowed steps are in red (from white to black square).

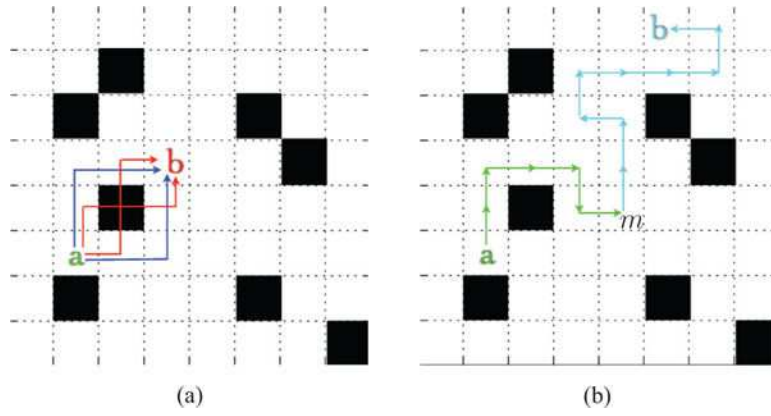
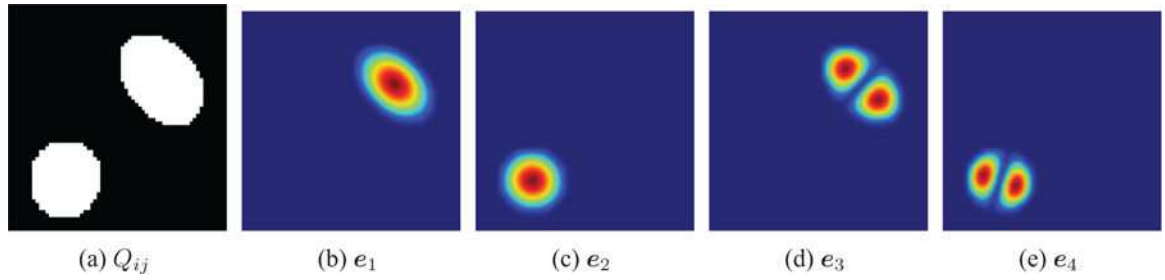
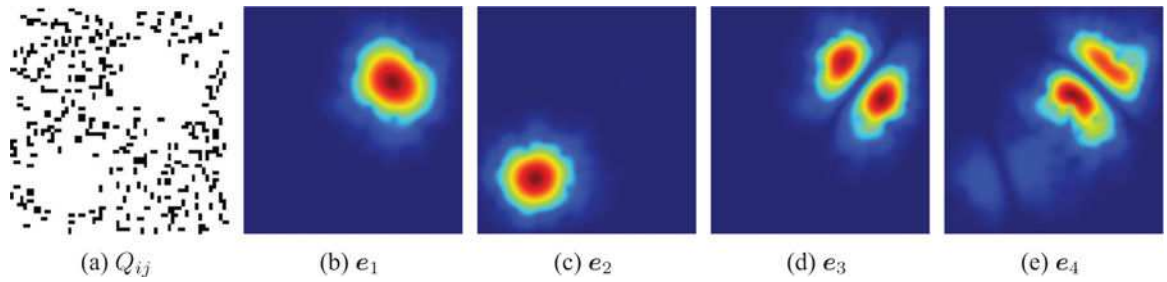


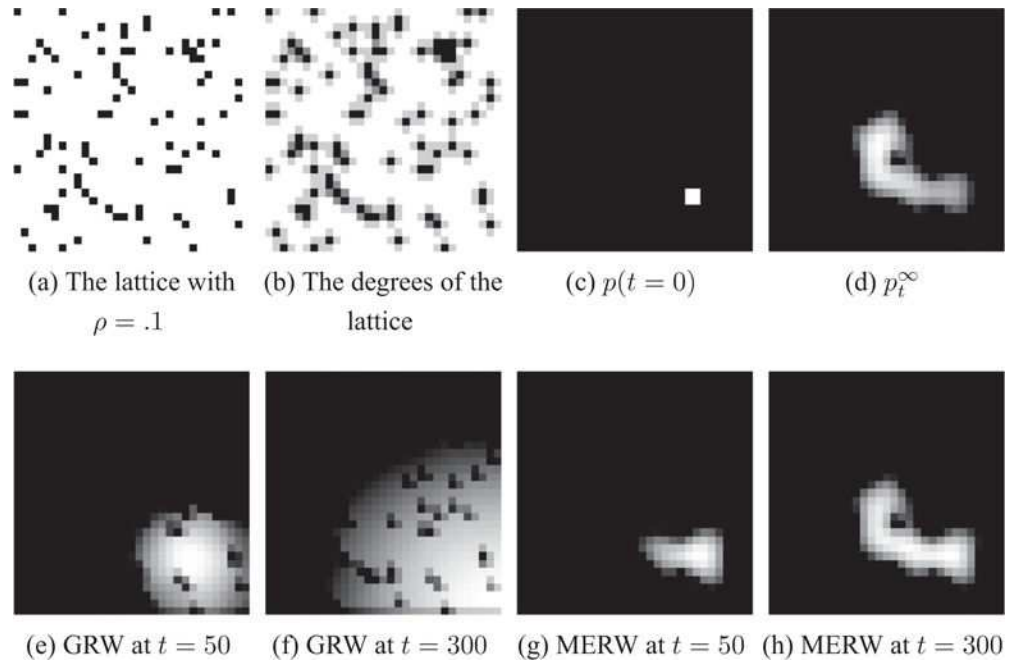
FIG. 3. (Color online) Paths on a defective lattice. (a) How many ways are there to get from a to b in 4 steps? Blue paths (visiting only white squares) are allowed, red (visiting at least one black square) are not. (b) The transition probability can be found by splitting the path into two segments, $\gamma_{ac}^{(m-a)}$ and $\gamma_{cb}^{(n-m+1)}$ at the point m (Eq. 23).

**FIG. 4.**

(Color online) The adjacency matrix Q_{ij} (a) and eigenvectors e_i (b)–(e) for a periodic square lattice of size $L \times L$ ($L = 64$) containing both a disk and an ellipse arranged in decreasing order of their eigenvalues λ_i . The eigenvectors distinguish separately the two regions and rank their relative modes according to their eigenvalues, and are the eigenvectors for the individual shapes. The first eigenvalue (b) determines the maximum entropy solution p^∞ , which is evidently not determined by the largest spherical region [2] but rather the largest eigenvalue of the Laplacian for these bounded regions (i.e., the ellipse). White and black regions denote accessible and inaccessible sites, respectively.

**FIG. 5.**

(Color online) The adjacency matrix Q_{ij} (a) and eigenvectors e_i (b)–(e) for a periodic square lattice of size $L \times L$ ($L = 64$) containing both a disk and an ellipse and random defects at a density of $\rho = 0.05$, arranged in decreasing order of their eigenvalues λ_i .

**FIG. 6.**

Dynamic evolution of the generic random walk [GRW: (e) and (f)] and the maximum entropy random walk [MERW: (g) and (h)] for the lattice shown in (a) with a starting distribution shown in (c). The lattice “defects” are, unlike in Ref. [2], inaccessible regions.

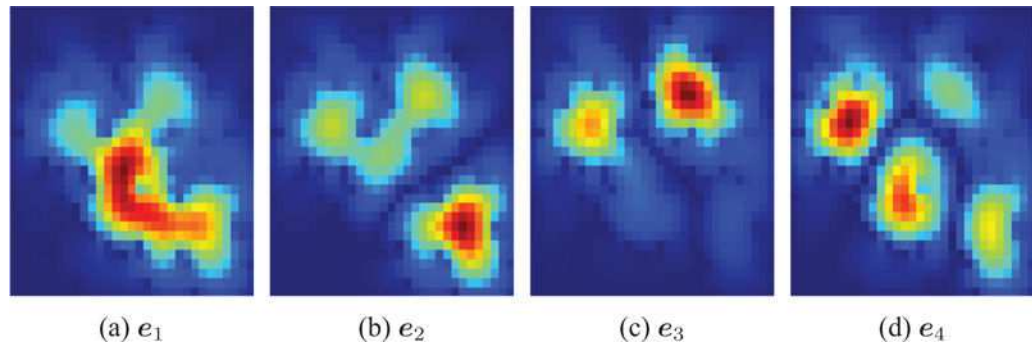
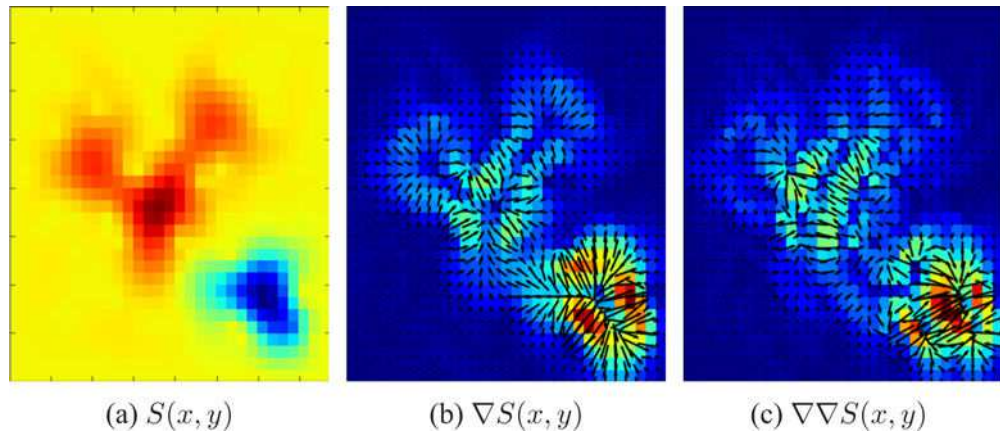
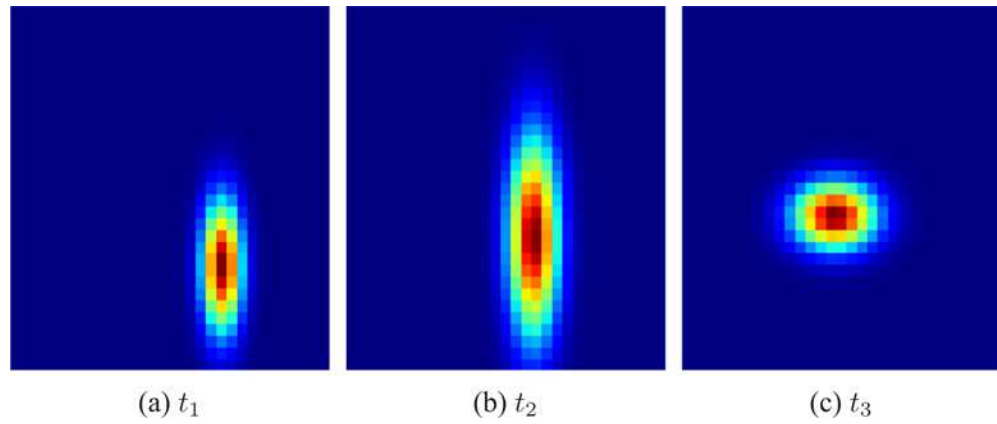


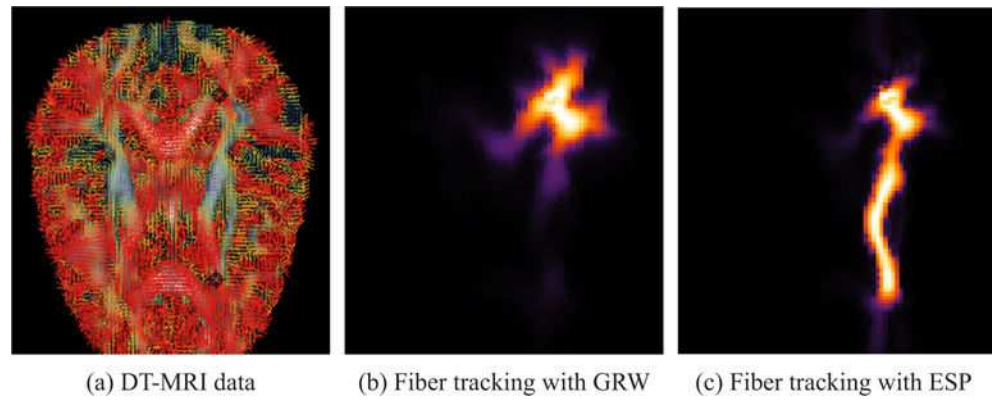
FIG. 7.
 (Color online) Entropy spectrum pathways (ESP): the first four eigenvectors of the lattice in Fig. 6(a).

**FIG. 8.**

(Color online) The path entropy $S(x, y)$, Eq. (51), from the initial distribution location Fig. 6(c) to every other location and its spatial first and second derivatives. Note that the map has the characteristics of a source (at the low entropy blue starting region) and a sink (at the high entropy red destination region).

**FIG. 9.**

(Color online) The time-varying distribution $P(x, y, t)$ for the path entropy Fig. 8 at three successive time points. The starting distribution follows the maximum entropy path shown in Eq. (6). The initially localized distribution moves and spreads in accordance with the local entropy field structure, then stalls and tightens at the maximum entropy location [the dark red region in Fig. 8(a)], and the location of the highest probability concentration of p_t^∞ [Fig. 6(d)].

**FIG. 10.**

(Color online) The application of ESP to neural fiber tractography using diffusion tensor magnetic resonance imaging (DT-MRI) and comparison with the generic uniform random walk (GRW). Data were collected on a normal human subject on a 3T GE Excite MR system with an eight-channel phase-array head coil using a spin echo echo-planar acquisition optimized for minimum echo time and the reduction of eddy current artifacts [16]. Diffusion weighted images were collected along 61 gradient directions distributed according to the electrostatic repulsion model [17] at a b value of $b = 1500 \text{ s/mm}^2$. The acquisition parameters were $TE/TR = 93/10,900 \text{ ms}$, $FOV = 240 \text{ mm}$, $NEX = 1$, matrix = 128×128 with 34 contiguous 3 mm slices. Two field maps were collected for unwarping to correct for signal loss and geometric distortion due to B_0 field inhomogeneities [18,19]. Total scan time including field maps was approximately 16 minutes.