# Fluctuating Interfaces and Paths in Disordered and Non-Equilibrium Systems

by

Sherry Chu

B.Sc., McGill University (2014)

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019



Author .....
Department of Physics
May 24, 2019

Certified by .....
Mehran Kardar
Professor of Physics
Thesis Supervisor

Accepted by .....
Nergis Mavalvala
Associate Department Head of Physics

# Fluctuating Interfaces and Paths in Disordered and Non-Equilibrium Systems

by

Sherry Chu

Submitted to the Department of Physics
on May 24, 2019, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

## Abstract

In this thesis, we study the statistics of fluctuating paths and interfaces in the presence of disorder. Specifically, we consider systems in the Kardar-Parisi-Zhang universality class for stochastic interface growth, from the perspectives of both fundamental statistical mechanics and applications to real world problems. We show numerically that the probability distribution associated with directed polymers in random media, a lattice model in this universality class, interpolates between Tracy-Widom and Gaussian distributions when spatial correlations are added to the random energy landscape. As a possible application, we examine the statistics of optimal paths on actual road networks as given by GPS routing, exploring connections and distinctions to directed polymers. We investigate also the effects of roughness in the growth front of a bacterial range expansion. There, we find that such roughness can account for the experimentally observed super-diffusivity, and leads to a rapid loss of genetic diversity. Finally, we explore the complete eigenvalue spectrum of products of random transfer matrices, as relevant to a finite density of non-intersecting directed polymers. We identify a correspondence in distribution to eigenvalues of Gaussian random matrices, and show that the density of states near the edge of the spectrum is altered by the presence of disorder.

Thesis Supervisor: Mehran Kardar
Title: Professor of Physics

# Acknowledgments

When I began my PhD, I had only a vague idea of what to expect. My advisor, Prof. Mehran Kardar, deserves my deepest gratitude, for without his guidance and direction, I could not have found my own path through the disordered landscape of graduate school. His patience, kindness, and generosity over the years are just a few of the reasons I feel incredibly fortunate to have him as a mentor. He has taught me to approach problems with an equal measure of intuition and rigour, and his dedication to excellence in both research and teaching is a constant source of inspiration.

I was fortunate as well in the collaborators I have had. I am especially thankful to Profs. David Nelson and Daniel Beller for introducing me to the exciting world of biophysics, as our work together marked the first time I dipped my toes into those bacteria-filled waters. Furthermore, I have benefited tremendously from working with Prof. Guy Bunin, Dr. Alexandre Solon, and Dr. Ramis Movassagh, each of whom brought their own expertise to our projects, and taught me to appreciate statistical physics in a new light.

What drew me to MIT when I first visited was its vibrant academic environment, and in the years since, it has never disappointed. I am grateful to the members of my thesis committee, Profs. Senthil Todadri and Riccardo Comin, for taking the time to review my thesis and provide their valuable feedback. Special thanks as well to Dr. Darshita Shah, Dr. Lilen Uchima and Greg Torrales, for the opportunity to develop my teaching skills as a graduate student. In addition, I would like to thank my friends and colleagues, including Hanrong Chen, Pavel Chvykov, Hridesh Kedia, and Benjamin Landon, for the many engaging conversations about both research and academic life in general.

I have learned many things during my time at MIT, the most unexpected of which has to be hockey. Growing up in Canada, I somehow missed the day at school when everyone learned to stickhandle, but I have found a more than welcoming hockey community here in Boston. To everyone I have skated with over the past five years, thanks for keeping me sane through my PhD.

Finally, this thesis is dedicated to my parents, Guimin and Xuehong, my sister Jessica, and my husband Eric. Without them, I would not be the person I am today.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Stochastic processes which give rise to fluctuations in paths and interfaces are ubiquitous in everyday life. Consider, for instance, a tear through a sheet of paper (example of a path) or the front of a spreading wildfire (example of an interface). The resulting shape is determined by the inherent noise in the system, and would differ slightly each time the process is repeated. By examining the fluctuations over a large number of realizations, statistical properties which are universal to many systems become evident. We introduce below this rich universality class for stochastic interface growth, and in the chapters which follow, present results from the perspectives of both statistical mechanics theory and applications to diverse physical problems.

## 1.1 Kardar-Parisi-Zhang equation

The time evolution of an interface undergoing stochastic growth is described by the Kardar-Parisi-Zhang (KPZ) equation [1, 2],

$$\frac{\partial h(\mathbf{x},t)}{\partial t} = v + \nu\nabla^2 h(\mathbf{x},t) + \frac{\lambda}{2}[\nabla h(\mathbf{x},t)]^2 + \eta(\mathbf{x},t), \tag{1.1}$$

where $h(\mathbf{x},t)$ is the interface height at spatial position $\mathbf{x}$ and time $t$. A sample realization of such an interface is shown in Fig. 1-1. The first term $v$ gives the average vertical growth velocity. It is often dropped by setting $v = 0$, since it does not

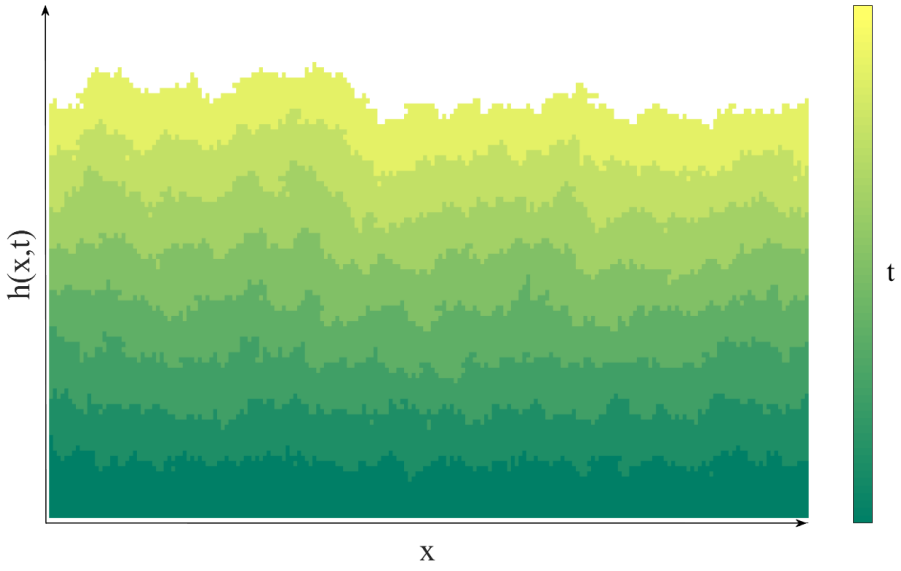


Figure 1-1: Growth of a KPZ interface in $d = 1 + 1$ dimensions. The interface $h(x, t)$ roughens over time as the height fluctuations grow with characteristic scaling exponents.

contribute to the shape of the interface, which is driven by disorder. The second term $\nu\nabla^2 h(\mathbf{x}, t)$ provides the familiar smoothing mechanism. The third term $\lambda/2[\nabla h(\mathbf{x}, t)]^2$ is of particular interest; it describes growth along the local normal to the interface. This leads to the intuitive picture that a bump in the interface will grow not only "upwards", but "outwards", as shown in Fig. 1-2. The strength of these mechanisms are governed by the parameters $\nu$ and $\lambda$. Finally, the fourth term $\eta(\mathbf{x}, t)$ is the stochastic noise, and the source of disorder in the system. It is typically taken to be uncorrelated white noise, with

$$\langle \eta(\mathbf{x}, t)\eta(\mathbf{x}', t')\rangle = 2D\delta(\mathbf{x} - \mathbf{x}')\delta(t - t'), \tag{1.2}$$

where $D$ determines the strength of the disorder.

As the interface grows, the fluctuations also grow with characteristic scaling exponents, in a process known as interface roughening. The statistical properties of these fluctuations are of great interest, as they form the foundation for the so-called *KPZ universality class*. The scaling exponents, as well as the limiting distribution of the fluctuations, are distinctive, and can be used to identify processes described by the

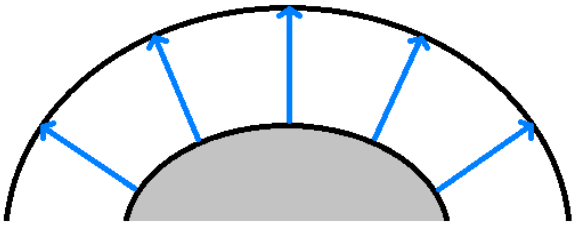

Figure 1-2: Illustration of how a protrusion on a KPZ interface propagates. The growth is "outwards" in the direction of the local normal, as would be intuitive for a spreading fire front or bacteria colony.

KPZ equation even when they differ drastically in the microscopic details.

### 1.1.1 Universal scaling form

At its core, the KPZ universality class is characterized by the scaling form for the interface height fluctuations [1],

$$\Delta h = \langle (h - \langle h \rangle)^2 \rangle^{1/2} \sim |\mathbf{x}|^{\chi} f\left(\frac{t}{|\mathbf{x}|^z}\right) \sim \begin{cases} t^{\beta}, & t \ll |\mathbf{x}|^z, \\ |\mathbf{x}|^{\chi}, & t \gg |\mathbf{x}|^z. \end{cases} \tag{1.3}$$

where the angular brackets represent averaging over many realizations of the disorder, and $f$ is a scaling function. In Eq. 1.3, we have introduced the *growth exponent* $\beta$, the *roughening exponent* $\chi$, and the *dynamic exponent* $z$. We will discuss the significance of each exponent in turn.

The dynamic exponent $z$ quantifies the scaling relationship between the spatial and temporal directions. In particular, it is often convenient to define the *wandering exponent* $\zeta = 1/z$, such that the transverse fluctuations in the system scale as $|\mathbf{x}| \sim t^{\zeta}$. The growth exponent $\beta$ and the roughening exponent $\chi$ then characterize the scaling of $\Delta h$ in the asymptotic regimes $t \ll |\mathbf{x}|^z$ and $t \gg |\mathbf{x}|^z$, respectively. From the definitions in Eq. 1.3, we immediately see that the exponents must be related as follows,

$$\beta = \chi / z. \tag{1.4}$$

In systems where the disorder does not have temporal correlations, a second exponent identity arises from Galilean invariance of the KPZ equation [1, 3],

$$\chi + z = 2. \tag{1.5}$$

The exponent identity in Eq. 1.5 can be derived from the symmetry of the equation under an infinitesimal tilt transformation by an angle $\boldsymbol{\epsilon}$ (see Ref. [1] for details),

$$h \to h + \boldsymbol{\epsilon} \cdot \mathbf{x}, \qquad \mathbf{x} \to \mathbf{x} + \lambda \boldsymbol{\epsilon} t. \tag{1.6}$$

Together, these exponent identities reduce the number of independent scaling exponents to one, and allow us to focus on, for instance, $\beta$ alone.

In the chapters which follow, we focus on KPZ growth in $d = 1 + 1$ dimensions (i.e. one spatial dimension $x \in \mathbb{R}$ and one time dimension $t \in \mathbb{R}^+$). In this case, the exact growth exponent is known to be $\beta = 1/3$ [4–7]. Of course, this also gives $z = 3/2$ and $\chi = 1/2$, via the exponent identities in Eqs. 1.4 and 1.5. Analogous results in $d = 2 + 1$ dimensions can be found in Ref. [6]; there, numerical studies place the growth exponent at $\beta \simeq 0.24$ for various models in the universality class, although the exact value is not yet known.

### 1.1.2 Tracy-Widom distribution

A second statistical property which characterizes the KPZ universality class is the probability distribution of the interface fluctuations. The height function for a KPZ interface can be written in the form

$$h(x, t) = v_\infty t + (\Gamma t)^{1/3} \xi \tag{1.7}$$

where $v_\infty$ is the asymptotic average growth velocity, $\Gamma$ is a system-dependent parameter which can be estimated using the Krug-Meakin toolbox [8], and $\xi$ is an $\mathcal{O}(1)$ random variable [9].

In $d = 1 + 1$ dimensions, the limiting distribution for this random variable $\xi$ has

| $\beta$ | Ensemble | Mean | Variance | Skewness | Kurtosis |
|---|---|---|---|---|---|
| 1 | GOE | -1.2065335745 | 1.6077810345 | 0.2934645240 | 0.1652429384 |
| 2 | GUE | -1.7710868074 | 0.8131947928 | 0.2240842036 | 0.0934480876 |
| 4 | GSE | -2.3068848932 | 0.5177237208 | 0.1655094943 | 0.0491951565 |

Table 1-1: Statistics of Tracy-Widom distributions for different ensembles of Gaussian random matrices.

been shown analytically to be Tracy-Widom (TW), with a parameter which depends on the initial conditions. These distributions were originally introduced in connection with the extremal eigenvalues of random matrices [10, 11]. They have since acquired iconic status [12] due to applications ranging from bioinformatic sequence alignments [13] to aircraft fault detection [14].

In the random matrix context, the TW distribution with parameter $\beta$ is defined as the limit

$$F_\beta(s) = \lim_{n\to\infty} \mathbb{P}_\beta\left(\sqrt{2}n^{1/6}(\lambda_{\max} - \sqrt{2n}) < s\right), \qquad \beta = 1, 2, 4, \tag{1.8}$$

where $\lambda_{\max}$ is the largest eigenvalue of an $n \times n$ random matrix in the Gaussian ensemble associated with parameter $\beta$ [10, 11]. The case $\beta = 1$ corresponds to the Gaussian Orthogonal Ensemble (GOE) of real symmetric matrices, while the case $\beta = 2$ corresponds to the Gaussian Unitary Ensemble (GUE) of hermitian matrices. The relevant distribution for a given KPZ growth process depends on the details of the initial condition: Growth in a flat geometry is described by TW-GOE, whereas growth in a radial geometry (i.e. from a droplet) is described by TW-GUE [9]. These limiting distributions have not only been derived analytically using various KPZ lattice models [5, 15–17], but have also been observed experimentally in a system of liquid crystals [18–20]. The last case $\beta = 4$ corresponds to the Gaussian Symplectic Ensemble (GSE). While TW-GSE arises theoretically in KPZ growth for some specific geometries, such as in half-space problems with an absorbing boundary at the origin [21], or with different growth rates on either side of the origin [22], it is not as physically relevant as the other ensembles, and will not be discussed further here. More details on the definition and properties of the GSE can be found in Ref. [23].

The GUE distribution can be written explicitly in the following integral form,

$$F_2(s) = \det\left(I - K_{\mathrm{Ai}}\right) = \exp\left(-\int_s^\infty (x-s)q^2(x)\,dx\right), \tag{1.9}$$

where $K_{\mathrm{Ai}}$ is the Airy kernel acting on $L^2(s,\infty)$,

$$K_{\mathrm{Ai}} = \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\mathrm{Ai}(y)}{x-y} \tag{1.10}$$

and $q(t)$ is the solution to the Painlevé II equation

$$q'' = xq + 2q^3 \tag{1.11}$$

with the boundary condition $q(x \to +\infty) \sim \mathrm{Ai}(x)$. $\mathrm{Ai}(x)$ is the Airy function. The GOE and GSE distributions can then be written in terms of the above, as

$$F_1(s) = \exp\left(-\frac{1}{2}\int_s^\infty q(x)\,dx\right)\left(F_2(s)\right)^{1/2},$$

$$F_4\left(\frac{s}{\sqrt{2}}\right) = \cosh\left(\frac{1}{2}\int_s^\infty q(x)dx\right)\left(F_2(s)\right)^{1/2}. \tag{1.12}$$

The TW distributions for the different ensembles are plotted in Fig. 1-3, and their characteristic statistics are summarized in Table 1-1. Numerical evaluation of the distributions has been achieved through approximation methods for both the Painlevé equation and the Fredholm determinant forms [24, 25].

## 1.2 Lattice models

The KPZ universality class comprises a diverse range of lattice models. Some are direct models of growth processes with a KPZ interface, while others can be mapped to the KPZ equation through transformations. The advantage of having access to such a rich universality class is that while each system may be susceptible to different analytical and numerical methods, the individual results can reveal insights about the universality class as a whole. Here, we introduce two such lattice models in

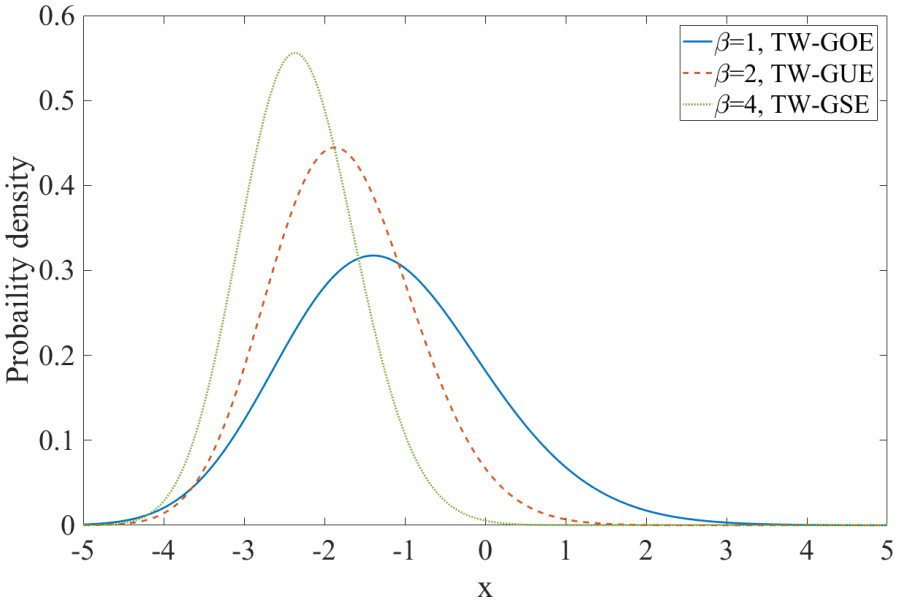


Figure 1-3: Tracy-Widom distributions associated with the largest eigenvalue of different ensembles of Gaussian random matrices.

detail: directed polymers in random media (DPRM) and the Eden model. Other models which are closely related, but will not be discussed here include the restricted solid-on-solid model (RSOS) [26,27], ballistic deposition (BD) [28,29], and the totally asymmetric simple exclusion process (TASEP) [30,31].

### 1.2.1 Directed polymers in random media

The DPRM problem considers configurations of a directed path (no overhangs) traversing a random energy landscape (see Fig. 1-4 for an illustration) [32–34]. Fluctuations in the transverse direction are permitted in order to explore the random energy landscape, and optimize the sum of energies along the path. The optimal path energy (or the free energy at finite temperature) exhibits sample to sample fluctuations, which scale with the path length $t$, as $t^{\beta}$. In $d = 1 + 1$ dimensions, and for uncorrelated random energies, the scaled probability of these fluctuations satisfies the TW distribution [9, 35]. As one of the simplest random processes described by the KPZ equation, DPRM has been extensively studied over the past three decades [36–38], with renewed recent interest [6, 7, 39, 40] due to its connection to TW. Extensive re-

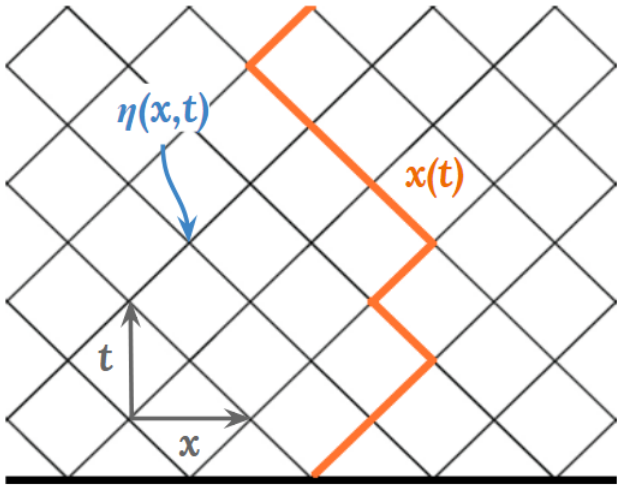


Figure 1-4: Schematic of DPRM on a square lattice with on-site random energies $\eta(x,t)$. The directed polymer $x(t)$ is oriented in the $t$-direction, but is allowed to fluctuate in the $x$-direction in order to minimize the sum of random energies along the path.

views from both statistical physics [33, 34] and mathematical [9] perspectives provide an excellent background on the subject.

Similar to other lattice models in the KPZ universality class, DPRM does not directly model interface growth. To see its connection to the KPZ equation, we introduce the Cole-Hopf transformation (in $d = 1 + 1$ dimensions for simplicity),

$$W(x,t) = \exp\left[\frac{\lambda}{2\nu} h(x,t)\right], \tag{1.13}$$

where $h(x,t)$ is the interface height and $\lambda$ and $\nu$ are the parameters from Eq. 1.1. Substituting Eq. 1.13 into Eq. 1.1 yields the stochastic heat equation,

$$\frac{\partial W}{\partial t} = \nu \nabla^2 W(x,t) + \eta(x,t) W(x,t). \tag{1.14}$$

This transformation removes the nonlinear term in the KPZ equation, thus removing a level of complexity, but exchanges the additive stochastic noise term for a multiplicative one.

Eq. 1.14 may look familiar, since it has the same form as the Schrödinger equation

but with imaginary time. We can therefore write down a formal path integral solution,

$$W(x,t) = \int_{(0,0)}^{(x,t)} \mathcal{D}x'(t') \exp\left\{ -\int_0^t \mathrm{d}t' \left[ \frac{1}{4\nu} \left( \frac{\mathrm{d}x'}{\mathrm{d}t'} \right)^2 - \frac{\lambda}{2\nu} \eta(x',t') \right] \right\}. \tag{1.15}$$

$W(x,t)$ in the above form can be interpreted as the partition function of a directed polymer in the continuum limit, with length $t$ and terminating at a point $x \in \mathbb{R}$. The $(dx'/dt')^2$ term corresponds to an elastic energy ($\nu$ is related to the polymer line tension), and the $\eta(x,t)$ term corresponds to the random potential energy at $(x,t)$. From the above transformation, we expect the free energy of a directed polymer,

$$f = -\ln W, \tag{1.16}$$

to be described by the same statistics as the height of a KPZ interface.

The tractability of DPRM in simulations makes it an especially attractive model with which to study the KPZ universality class. Unlike the traveling salesman problem (which allows overhangs and loops) [41], the optimization problem for DPRM can be solved in polynomial time using a transfer matrix formalism [32–34]. In the rotated square lattice geometry depicted in Fig. 1-4, the optimal path to a point $(x,t)$ depends only on the optimal paths at the previous time step, to the neighbouring point from the left, $(x-1,t-1)$, and to the neighbouring point from the right, $(x+1,t-1)$. The partition function for this discrete version of the directed polymer can be written recursively as

$$W(x,t) = e^{-\mu(x,t)} \left[ W(x-1,t-1) + W(x+1,t-1) \right] = \sum_{x'} \langle x|\mathbf{T}|x'\rangle W(x',t-1). \tag{1.17}$$

where $\mu(x,t) = -(\lambda/2\nu)\eta(x,t)$ is the rescaled random energy, and $\mathbf{T}$ is a transfer matrix which describes the time evolution. At zero temperature, the free energy is simply the energy of the optimal path,

$$E(x,t) = \mu(x,t) + \min\{E(x-1,t-1), E(x+1,t-1)\}. \tag{1.18}$$

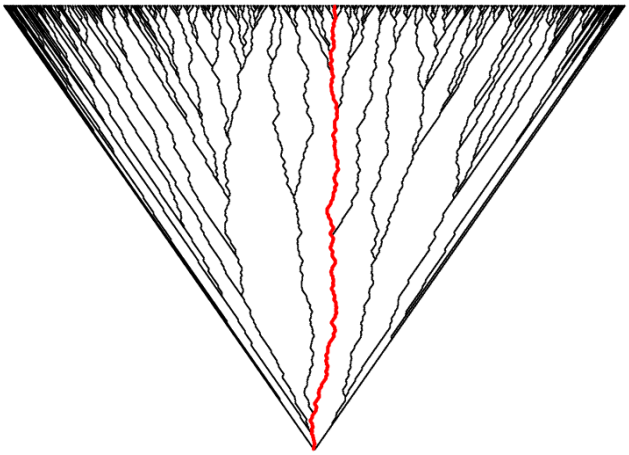

Figure 1-5: Illustration of DPRM paths. Each black line represents an optimal path in the pt-pt geometry, from the origin (the bottom of the wedge) to a specific point along the top. The bolded red line represents the optimal path in the pt-line geometry, from the origin to *any* point along the top.

Simulations of DPRM can be performed in two different geometries: The pt-pt model optimizes over paths from the origin to $(x, t)$ for given $x$ and $t$, and obeys TW-GUE statistics associated with radial interface growth. The pt-line model, on the other hand, optimizes over paths from the origin to *any* $x$ and a given $t$, and obeys TW-GOE statistics associated with flat interface growth [15–17, 39, 42, 43]. Sample paths for the two geometries are shown in Fig. 1-5.

### 1.2.2 Eden model

A more direct model for stochastic interface growth is the Eden model [44, 45], first introduced in the context of cancer cell growth. This simple algorithm takes as initial condition a subset of sites on a square lattice which are "filled". At each time step, the empty sites with at least one filled neighbour are identified as edge sites, and one of these edge sites is filled at random, thus introducing stochasticity (see Fig. 1-6). As this process is repeated and the Eden cluster grows, the roughening of the edge is described by KPZ statistics. Despite its apparent simplicity, however, analytical results have proven elusive for the Eden model in comparison to other KPZ lattice models.

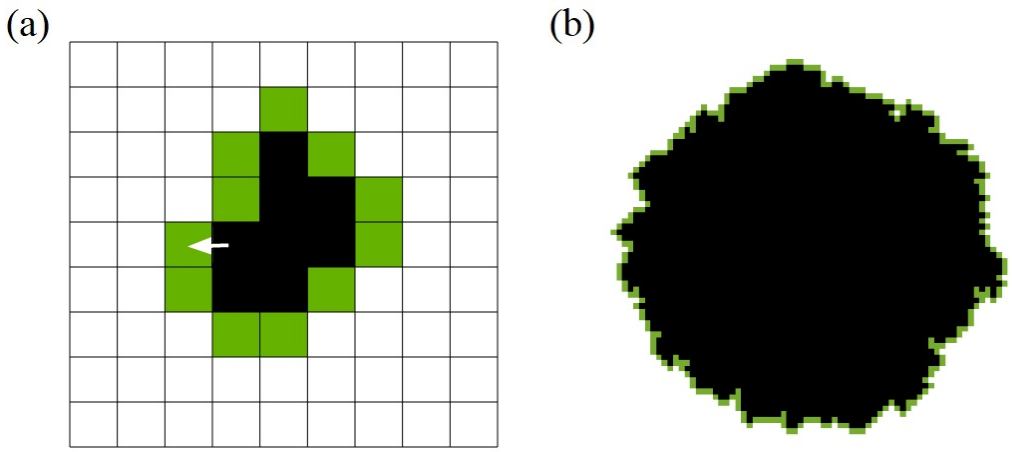


Figure 1-6: Schematic of the Eden model in the radial geometry. (a) An Eden cluster at time $t$ is shown in black, and the edge sites (empty sites which neighbour the growth front) are shown in green. At time $t+1$, the cluster expands into one of these edge sites at random, as indicated by the white arrow. (b) As the cluster grows over time, the edge develops increasing fluctuations which obey KPZ statistics.

## 1.3 Outline

The thesis is organized as follows.

In Chapter 2, we examine the robustness of the KPZ universality class in the presence of spatial correlations in the noise. Using the DPRM model, we investigate to what extent the scaling exponents and limiting distributions deviate from those introduced in Chapter 1.1, and verify the exponent identities.

In Chapter 3, we analyze the statistics of optimal paths given by GPS routing. We measure the distribution in the stochasticity of the underlying road network for different regions around the world, and model the paths as directed polymers in a correlated energy landscape.

In Chapter 4, we study the impact of fluctuations in the growth front of bacterial range expansions on the genetic diversity of the population. The statistics of genetic lineages can be characterized using coalescence events, which mark when two individuals share a common ancestor in the past. We also explore the effects of environmental heterogeneities, around which the bacteria must propagate.

In Chapter 5, motivated by the problem of non-intersecting DPRMs, we search

for connections between the transfer matrix formulation of DPRM and the Gaussian random matrices from which the TW distributions are derived. It is known that the DPRM free energy, related to the largest eigenvalue of a product of transfer matrices, shares the same limiting distribution. We ask the question why, and look to establish other similarities and distinctions between the two classes of matrices.

Finally, Chapter 6 provides an outlook on future directions.

# Chapter 2

# Probability distributions for directed polymers in random media with correlated noise

The results in this chapter are reproduced and adapted from Ref. [46].

The disorder in the KPZ equation is typically taken to be stochastic white noise, with no spatial or temporal correlations. However, this is not always the case in physical systems. Consider for instance a spreading forest fire, whose speed of propagation is affected locally by the changing wind speeds and availability of dry foliage to fuel it. In contrast to the sharp peaks and valleys associated with white noise, these fluctuations tend to be continuous and slowly varying over a larger length scale. They can be visualized as hills and valleys in the energy landscape over which the fire front propagates, causing the fire to spread faster in some regions, and slower in others. To incorporate this kind of randomness into our models for stochastic interface growth, it is natural to consider noise with correlations which decay slowly as a function distance (in either space or time).

In this chapter, we test the robustness of the KPZ universality class in the presence of long-range spatial correlations. Previous renormalization group (RG) calculations have predicted that the scaling exponents will deviate from their values in the uncorrelated case [1,3]. We perform numerical simulations of the DPRM model introduced in

Chapter 1.2.1, and compare the results to RG predictions. The energy fluctuations are well described by the growth exponent $\beta$ and the dynamic exponent $z$, extracted by standard data collapse. We find that the KPZ exponent identities reflecting Galilean invariance hold as expected, with some deviations which could be attributed to systematic errors. The exponent $\beta$ is qualitatively similar to RG predictions, but varies smoothly with the correlation exponent, in contrast to these predictions. As the correlations increase, the scaled probability distributions of the directed polymer energy are no longer well described by the TW form, instead becoming more symmetric. We thus find a class of distributions that continuously interpolates between TW and Gaussian forms.

## 2.1 Spatially correlated noise

Recall that the KPZ universality class is characterized by the scaling relation defined in Eq. 1.3, which, in the absence of temporal correlations in the noise, can be reduced to a single independent exponent $\beta$. It is known that this growth exponent $\beta$ can be modified by introducing noise that is fat-tailed [i.e. noise $\eta$ with probability distribution $P(\eta) \sim 1/\eta^{1+\mu}$ as $\eta \to -\infty$] [47–49], or long-range correlated [1]. The former was considered in Ref. [40], concluding that for $0 < \mu < 5$, both the scaling exponent and the end-point distributions are inconsistent with the KPZ/TW universality class described in Chapter 1.1. They did not, however, focus on the nature of the modified distributions. Here, we consider the latter, expanding on earlier work in Ref. [39]. Specifically, we examine both the scaling exponents and limiting probability distributions in the presence of long-range spatially correlated noise.

We generalize the KPZ equation as in Ref. [1] to include noise with spatial correlations which decay as a power law,

$$\frac{\partial h(x,t)}{\partial t} = v + \nu\nabla^2 h(x,t) + \frac{\lambda}{2}[\nabla h(x,t)]^2 + \eta(x,t),$$

$$\langle \eta(x,t)\eta(x',t')\rangle \sim |x-x'|^{2\rho-1}\delta(t-t'). \tag{2.1}$$

A one-loop dynamical renormalization group (RG) calculation [1,3] predicts

$$\beta(\rho) = \begin{cases} 1/3, & 0 < \rho < 1/4, \\ (1+2\rho)/(5-2\rho), & 1/4 < \rho < 1. \end{cases} \tag{2.2}$$

Eq. 2.2 was also obtained in the field-theoretic works of Ref. [50] and Ref. [51], using a stochastic Cole-Hopf transformation and a nonperturbative RG approach, respectively. The exponent $\rho$ is used as a measure for the degree of correlation in the system. On dimensional ground, the case of uncorrelated noise [$\delta(x) \sim 1/|x|$] corresponds to $\rho = 0$, in the regime where the RG result coincides with the exact value of $\beta = 1/3$. The limit $\rho = 1$ corresponds to the interface of a two-dimensional Ising model in random fields. The case of $\rho = 1/2$ is of particular interest: The DPRM problem is trivial if the noise does not depend on $x$, in which case the addition of random variables at different $t$ would lead to a Gaussian distribution whose width grows with $\beta = 1/2$ (as predicted by the above). However, as we shall elaborate in Appendix A, the numerical procedure used generates non-trivial correlations for $\rho = 1/2$ which vary logarithmically with $|x - x'|$.

Qualitatively, the change in scaling behaviour is best visualized by plotting the optimal paths on a correlated energy landscape. In Fig. 2-1, reproduced from Ref. [1], we see a comparison of optimal paths (in the pt-line geometry) for different values of $\rho$. As $\rho$ increases, large segments of the paths begin to resemble ballistic trajectories, which heuristically, leads to the expectation that the wandering exponent $\zeta = 1/z = (\beta + 1)/2$ increases, consistent with Eq. 2.2.

To study the effects of spatial correlations in more detail, we simulate the discrete pt-line DPRM on a square lattice, with random energies on each site $\eta(x, t)$. The path is directed along the diagonal, such that the minimal energy is calculated recursively according to the transfer matrix relation in Eq. 1.18. The square lattice is wrapped around a cylinder of size $L$, corresponding to periodic boundary conditions along the $x$-direction. For $\rho < 1$, random energies correlated as in Eq. 2.1 are generated using the Fourier transform method proposed in Ref. [52], the details of which can found in Appendix A. (A similar method for generating correlated noise was developed in

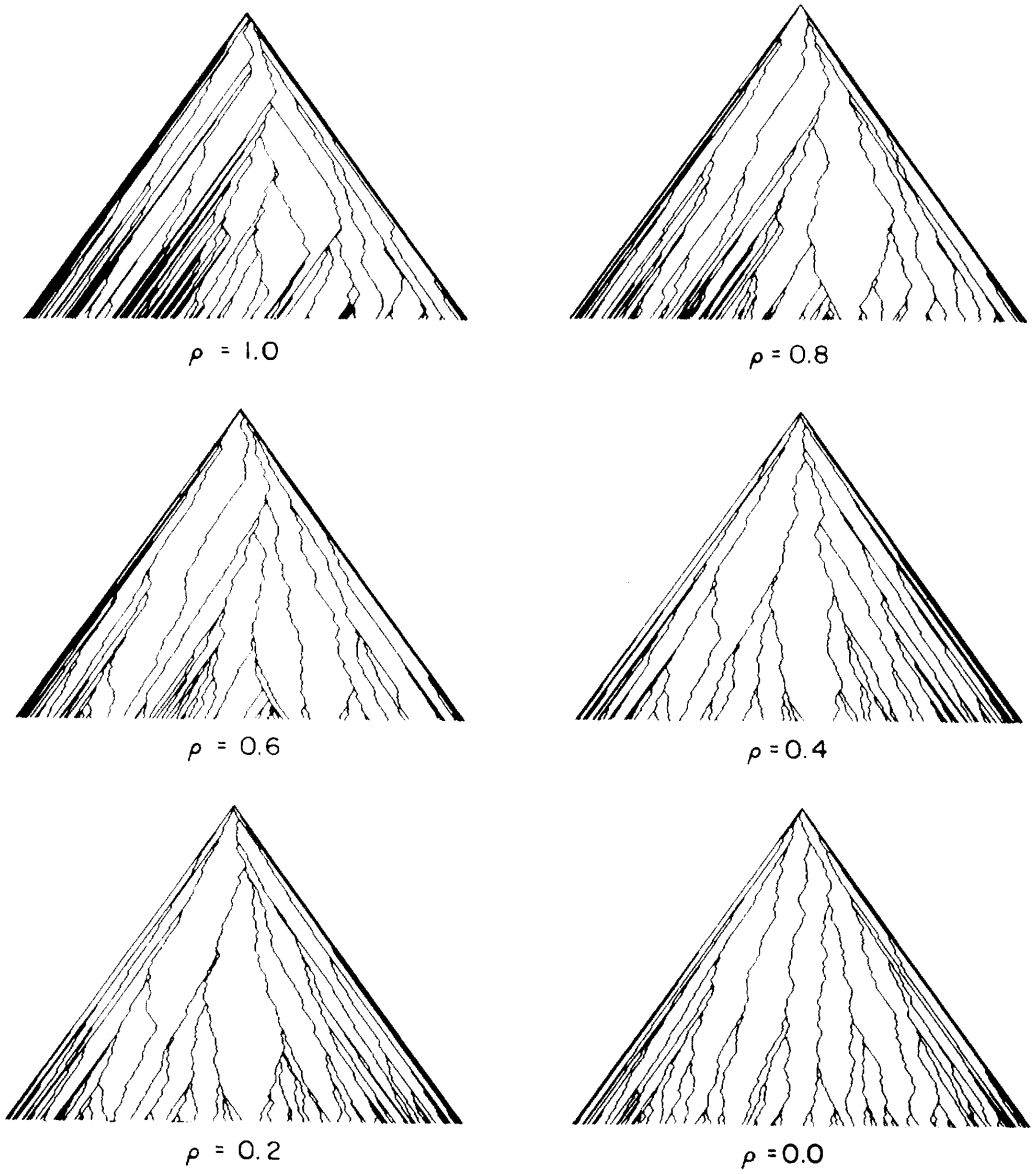


Figure 2-1: Optimal paths of 500 bonds along diagonals of a square connecting the apex to various points on the base. The bonds are all random variables, with correlations in horizontal strips, decaying with an exponent $\rho$. Different figures correspond to different values of $\rho$. Figure and caption are reproduced from Ref. [1].

Ref. [53].) For $\rho = 1$, the noise is constructed as a Brownian bridge, shifted to have zero mean. The simulated system size is $L = 10^6$, evolved over $t = 10^4$ time steps, and averaged over $10^2$ realizations.

## 2.2 Scaling exponents

In analogy to the KPZ height function described by Eq. 1.3, the fluctuations in the optimal DPRM energy, with finite system width $L$, is expected to satisfy the scaling form,

$$\Delta E = \langle (E - \langle E \rangle)^2 \rangle^{1/2} \sim L^{\chi} f\left(\frac{t}{L^z}\right) \sim \begin{cases} t^{\beta}, & t \ll L^z, \\ L^{\chi}, & t \gg L^z \end{cases}, \tag{2.3}$$

where the angular brackets denote averaging over different realizations (or independent segments in the same realization) of random energies.

We extract the growth exponent $\beta$ and the dynamic exponent $z$ from the collapse of the curves of $\Delta E/t^{\beta}$ vs. $L/t^{\zeta}$ for different times $t$. Alternatively, $\beta$ and $z$ can be deduced respectively from the scaling of the energy fluctuations $\Delta E \sim t^{\beta}$, and the transverse fluctuations $\Delta x = \langle (x - x_0)^2 \rangle^{1/2} \sim t^{\zeta}$, where $x_0$ is the origin of the directed polymer. This method yields exponents which are in good agreement with the data collapse approach for small $\rho$, where finite size effects are less important.

As presented in Fig. 2-2, the data is very well collapsed according to Eq. 2.3, although somewhat less so for larger values of $\rho$. In particular we note the excellent collapse at $\rho = 1/4$ which according to the RG result of Eq. 2.2 is the limiting point for which $\beta$ sticks to $1/3$. However, we find $\beta = 0.375 \pm 0.005$, (and $\zeta = 0.68 \pm 0.01$) in contradiction to RG, but consistent with previous results in Ref. [53] of $\beta = 0.364 \pm 0.005$ and $\zeta = 0.692 \pm 0.005$. Indeed, as depicted in Fig. 2-3, the exponent $\beta$ appears to vary continuously with $\rho$, in contradiction to Eq. 2.2. As in Ref. [53], we extend the simulations to $\rho \leq 0$, and throughout this regime obtain $\beta = 1/3$ consistent with uncorrelated noise. (We also find $\zeta = 2/3$ in this regime through data collapse as in Fig. 2-2). For larger values of $\rho$, the agreement with RG improves, and the expected random field Ising exponents of $\beta = \zeta = 1$ are recovered for $\rho = 1$. The continuous

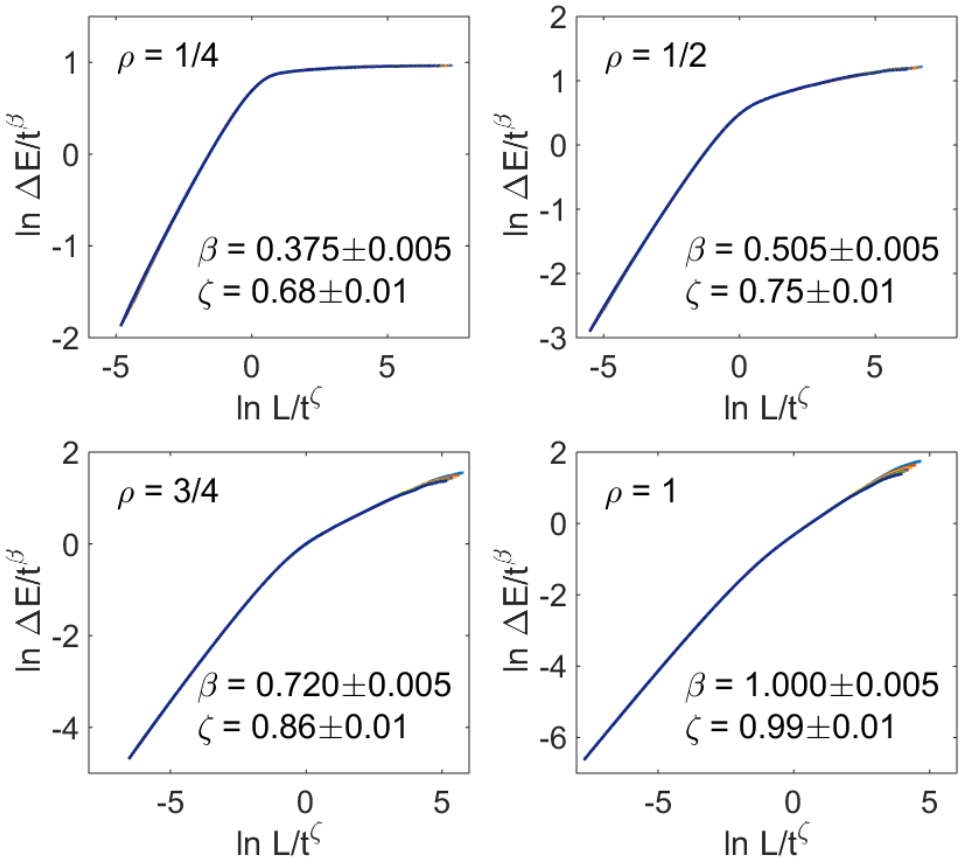


Figure 2-2: Collapse of energy fluctuations of DPRM with a spatially correlated landscape. The data corresponds to system size $L = 10^6$, evolved to time $t = 10^4$. The error bars on the exponents reflect statistical errors in the fits, neglecting potentially larger systematic errors.

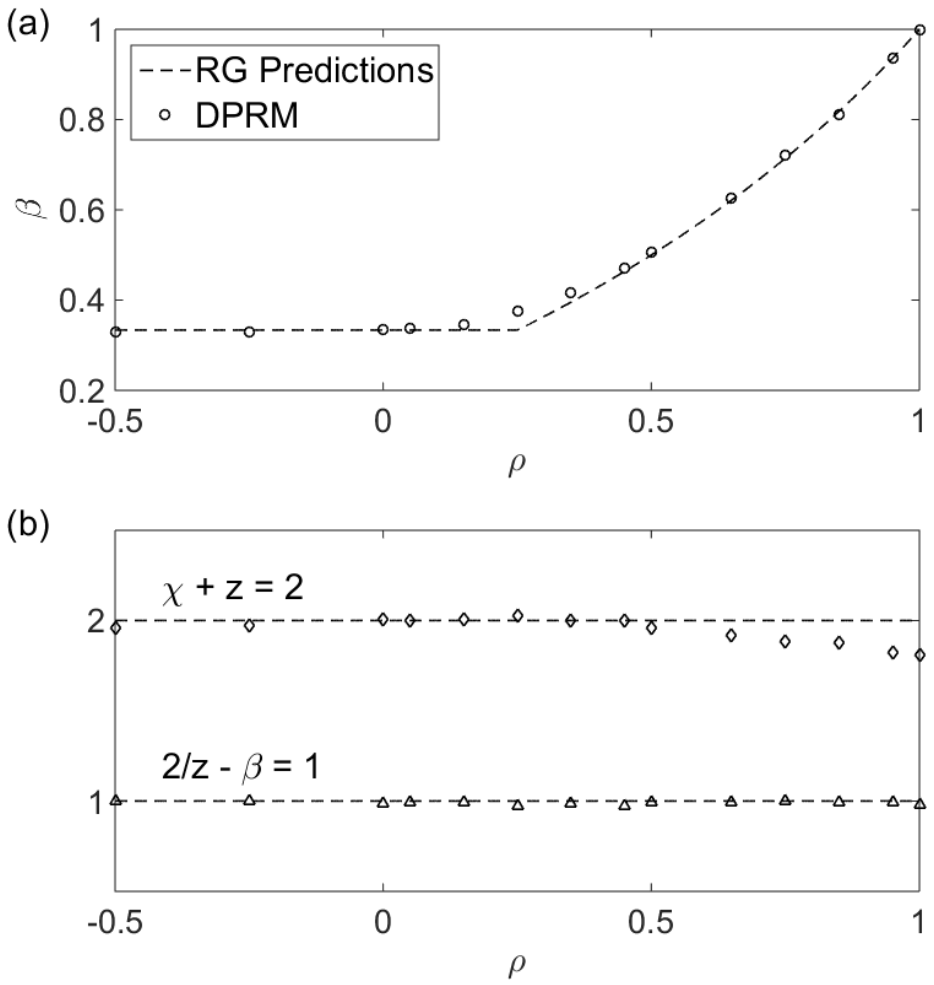


Figure 2-3: (a) Variation of $\beta$ with the exponent $\rho$ of spatially correlated energies. There is a small, but clear deviation from the predicted RG exponents (dashed line). (b) Validity of the exponent identities predicted by Galilean invariance; the discrepancies are likely a measure of systematic errors.

variation of $\beta$ for $\rho \leq 1/2$ is similar to observations in previous simulations of DPRM, restricted solid-on-solid (RSOS), and ballistic deposition (BD) models [53–56]. We note that the RG exponents are constrained to be exact for uncorrelated noise due to a fluctuation-dissipation condition. The exponents in Eq. 2.2, however, follow from a non-renormalization of correlated noise amplitude, which in view of the numerics is perhaps questionable.

In principle, the scaling relation, Eq. 2.3, involves two exponents ($\beta$ and $\zeta$, or $\chi$ and $z$). We estimated the roughening exponent $\chi$ from the slope of the collapsed curve in the regime $t \gg L^z$. A hallmark of the KPZ equation (even with spatially

correlated noise) is Galilean invariance [2, 3], which implies the exponent identity in Eq. 1.5, $\chi + z = 2$. The explicit check of this identity presented in Fig. 2-3 appears to indicate its breakdown for $\rho > 1/2$. However, simply dividing this identity by $z$, and using the relation in Eq. 1.4, $\beta = \chi/z$, we obtain a second form $2/z - \beta = 1$, which is excellently obeyed by the data! The discrepancy between these two identities is an indication of the systematic errors afflicting the fits, such as the small but systematic curvature in the initial rise of the collapsed curves in Fig. 2-2, whose slope is used to obtain the exponent $\chi$.

## 2.3 Probability distributions

We introduced in Chapter 1.1.2 the TW distributions which characterize systems in the KPZ universality class. Like the Gumbel and Gaussian distributions, TW is universal in being independent of various underlying (microscopic) details. However, whereas it is known how the addition of fat-tailed random variables modifies a Gaussian to a Lévy distribution, corresponding limitations for TW are not known. We take up this question in the DPRM context, and inquire if and how the TW form changes in spatially correlated energy landscapes.

The end-point energy probability distributions for DPRM are obtained from time $t = 10^3$ to $t = 10^4$, in increments of $\Delta t = 10^3$, and are normalized to have mean 0 and variance 1. The full distributions presented in Fig. 2-4 are qualitatively similar to the TW-GOE form for $\rho \leq 0$, but shift smoothly towards Gaussian as $\rho$ increases to $\rho = 1/2$, the borderline point between correlations growing or decaying with separation. Beyond $\rho = 1/2$, it is unclear whether the distribution remains Gaussian.

The skewness $s$ and kurtosis $k$, plotted in Fig. 2-5, are obtained by averaging results over the above snapshots in $t$. In the uncorrelated case, it is possible to estimate the true asymptotic values of $s$ and $k$ using methods developed in Ref. [8]. In the correlated case, however, we run into problems as the uncertainties grow rapidly with correlation. For $\rho \leq 0$, the skewness and kurtosis approach those of the TW-GOE, the limiting distribution for pt-line DPRM with uncorrelated noise. As $\rho$

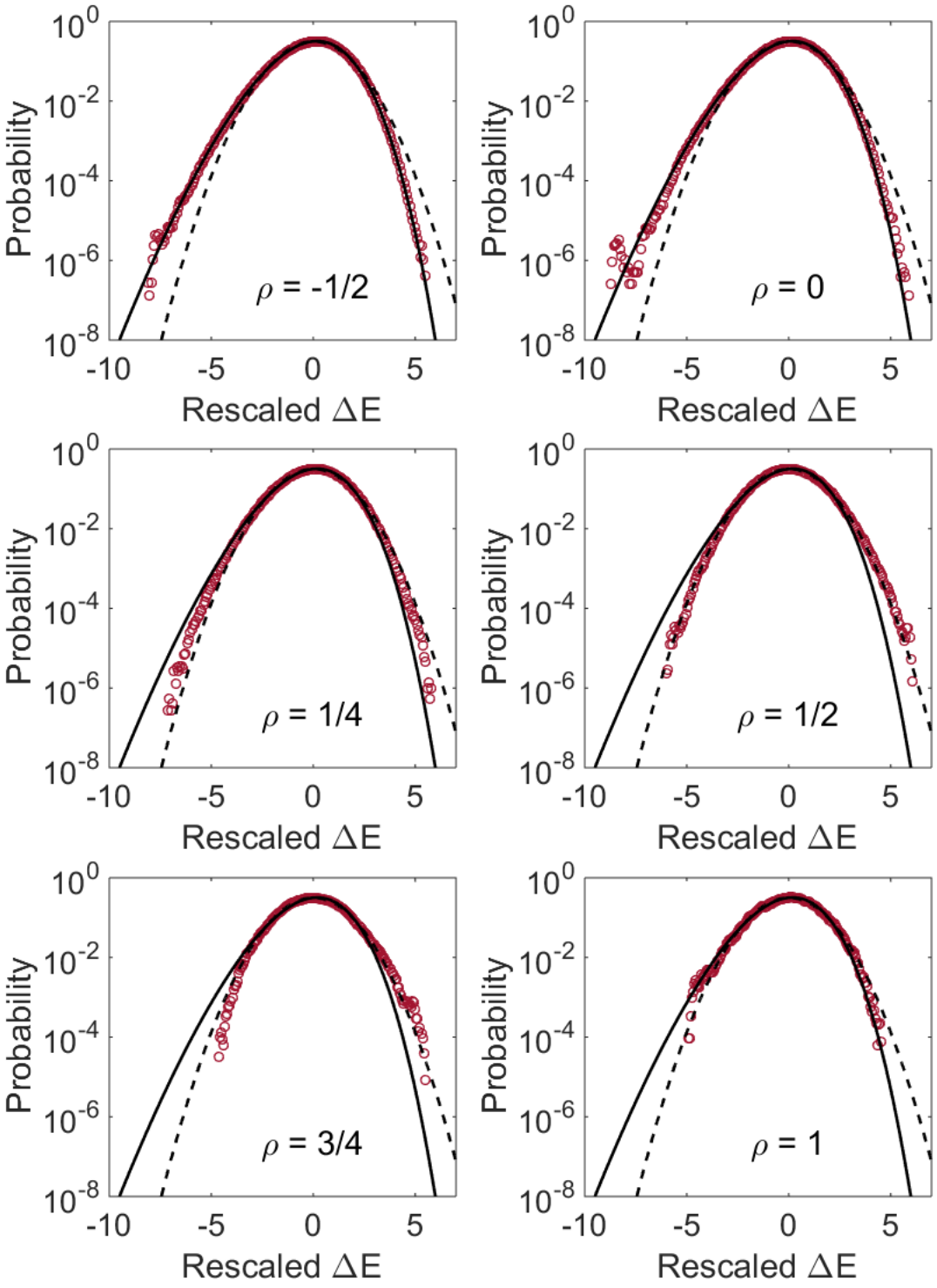


Figure 2-4: Probability distributions for the optimal energy of DPRM for different correlation exponents $\rho$. The data corresponds to system size $L = 10^6$ at time $t = 5 \times 10^3$, rescaled to have mean 0 and variance 1. Results are consistent with the TW-GOE form (solid line) for $\rho \leq 0$, and shift smoothly towards Gaussian (dashed line) at $\rho = 1/2$.

increases towards 1/2, both $s$ and $k$ decrease to 0, and the distribution becomes more symmetric. In particular, the data suggests that $s, k \to 0$ as $\rho \to 1/2$, consistent with a Gaussian distribution. This would be expected if $\rho = 1/2$ corresponded to random energies fully correlated in the $x$-direction, but randomly changing along the $t$-direction. The energy of the DPRM would then be a sum of random variables, thus $\beta = 1/2$, while the path executes a random walk with $z = 2$. The latter is not correct, as the Galilean exponent identity at $\beta = 1/2$ leads to the numerically observed exponent of $z = 4/3$. We note also that the Fourier transform procedure for generating spatially correlated noise, devised in Ref. [52] and used here, actually produces correlations which vary logarithmically at $\rho = 1/2$, as $\langle \eta(x,t)\eta(x',t')\rangle \sim (a - b\ln|x-x'|)\delta(t-t')$ (see Appendix A). Interestingly the RG result (also based on Fourier transformed noise) predicts the observed $\beta = 1/2$ as well. Finally, for $\rho > 1/2$, the uncertainty in $s$ and $k$ grow rapidly due to the increased correlations in random energies, and we cannot conclusively state whether the distribution is Gaussian or not. It is of interest to note that another class of distributions, interpolating between TW-GUE and Gaussian, was found in Ref. [57] in the context of random matrix theory, and the convergence of TW distributions to the Gumbel distribution was studied in Ref. [58].

For an uncorrelated landscape, the optimal energy of DPRM behaves analogously to Eq. 1.7, as

$$E = f_\infty t + (\Gamma t)^{1/3}\xi, \tag{2.4}$$

where $f_\infty$ and $\Gamma$ are non-universal, system-dependent parameters, and $\xi$ is a $\mathcal{O}(1)$ random variable obeying TW-GOE statistics. There is currently no analytical prediction for the limiting distribution in the case of correlated noise. From the overall scaling, we can propose a generalized form,

$$E = f_\infty t + (\Gamma t)^{\beta(\rho)}\xi(\rho), \tag{2.5}$$

where $\beta(\rho)$ is the modified growth exponent plotted in Fig. 2-3a. The random variable $\xi(\rho)$ is distributed according to TW-GOE statistics for $\rho \leq 0$. A priori one could have

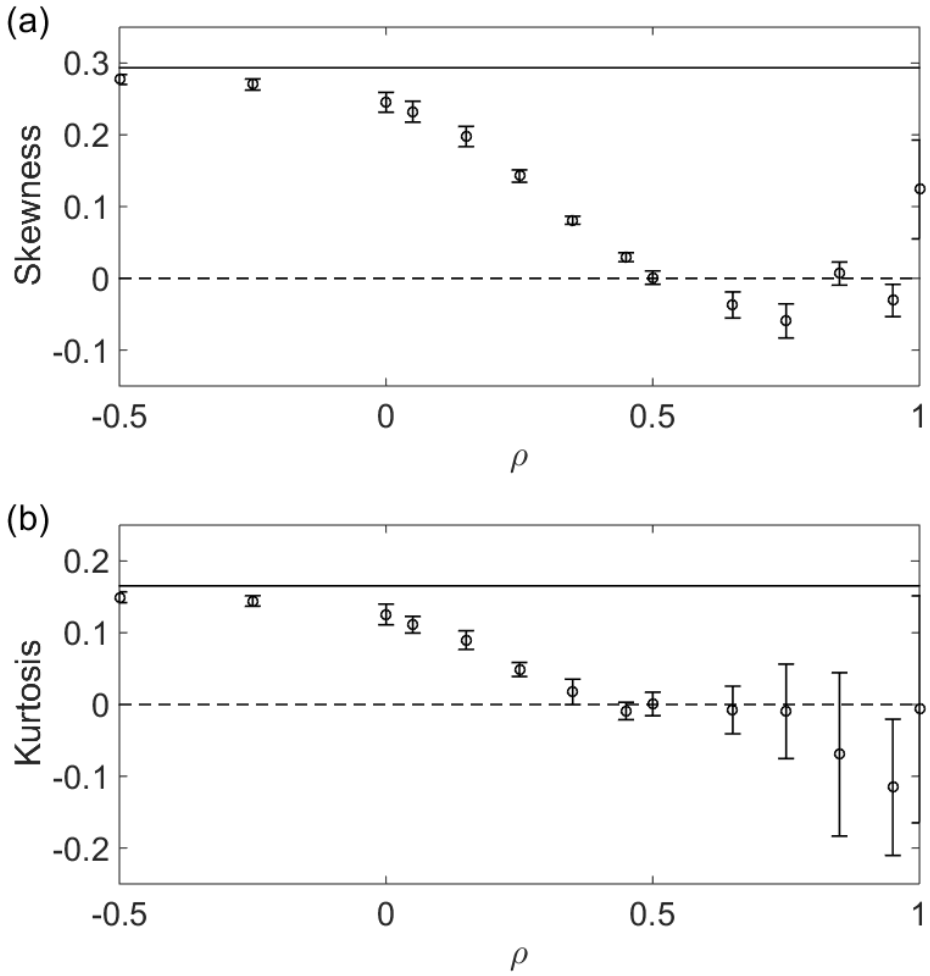


Figure 2-5: (a) Skewness and (b) kurtosis for DPRM with spatially correlated noise, compared to the TW-GOE values (solid lines), $s = 0.293$ and $k = 0.165$, respectively: Both approach the TW-GOE values for $\rho \leq 0$, and decrease to 0 as $\rho$ increases to $1/2$. Beyond $\rho = 1/2$, the uncertainties are too large to rule out $s = k = 0$.

imagined that the distribution retains the TW form in general, or that it discontinuously transitions to a different distribution for $\rho > 0$. Instead, we observe a smooth shift as $\rho$ increases, to a Gaussian form at $\rho = 1/2$. For $\rho > 1/2$, the uncertainty in skewness and kurtosis become too significant to conclude whether the distribution remains Gaussian. We thus find a class of distributions, interpolating between TW and Gaussian, which governs the statistics of DPRM with spatially correlated noise.

# Chapter 3

# Optimal paths on road networks as directed polymers

The results in this chapter are reproduced and adapted from Ref. [59].

Complex networks of nodes and links can be used to model a wide array of systems. Examples range from biological networks such as those formed by neurons and synapses in the brain or chemical reactions inside a cell, to social or transportation networks and the World Wide Web. Their topology in the abstract space of edges and vertices has been much studied, allowing one to identify widespread properties such as "small-world" effects, scale-free connectivity, and a high degree of clustering, which can be captured by simple physical models [60–64]. Comparatively, less is understood about the spatial organization of complex networks embedded in a Euclidean space, a very active area of research (see Ref. [65] for a review). The effect of geometry is especially relevant when the network is strongly constrained by the environment or when the "cost" to maintain edges increases significantly with their length (*e.g.* rivers [66], railways [67] or vascular networks [68]). The spatial structure of streets is another example that has been particularly studied to gain insight into the structure of cities and their development [69–71].

Much information about the geometry of a network can be obtained by studying the shortest paths between the nodes of the network. In many cases, it is also a problem of practical importance to characterize the paths that optimize a given cost

function. For example, in transportation networks, one would like to understand the properties of the paths that minimize the travel time, the distance, or the monetary cost to travel between two points. An obvious application is in the development of efficient global positioning system (GPS) routing algorithms which could use prior information on optimal paths to perform better [72]. The shortest paths between two generating nodes on the power grid are also important to predict the overloading of electric lines [14]. Understanding the properties of these optimal paths appears challenging since they are expected to depend strongly on the geometry of the network which can be shaped by various factors, from natural obstacles to historical development or differences in policy.

In this chapter, we analyze the statistics of the shortest and fastest paths on the road network between randomly sampled end points. We find that, to a good approximation, these optimal paths can be described as directed polymers from the KPZ universality class. Comparing the scaling behaviour of our data with simulations of DPRM and previous theoretical results, we are able to point out the few characteristics of the road network that are relevant to the large-scale statistics of optimal paths. Indeed, we show that the local structure is akin to a disordered environment with a power-law distribution which become less important at large scales where long-ranged correlations in the network control the scaling behaviour of the optimal paths.

## 3.1 Connection to directed polymers

The DPRM model introduced in Chapter 1.1.2 explores a physically distinct but mathematically related problem to that of road networks. It is concerned with the statistics of a path stretched between two points that minimizes its energy in a random environment modeled by a fluctuating potential. Configurations of DPRM paths such as in Fig. 1-5 bear qualitative resemblance to myriad natural transportation systems, from river deltas to vascular networks; the wealth of data on road networks provides the opportunity for a quantitative comparison.

Here, we study the statistics of optimal paths on the road network in light of

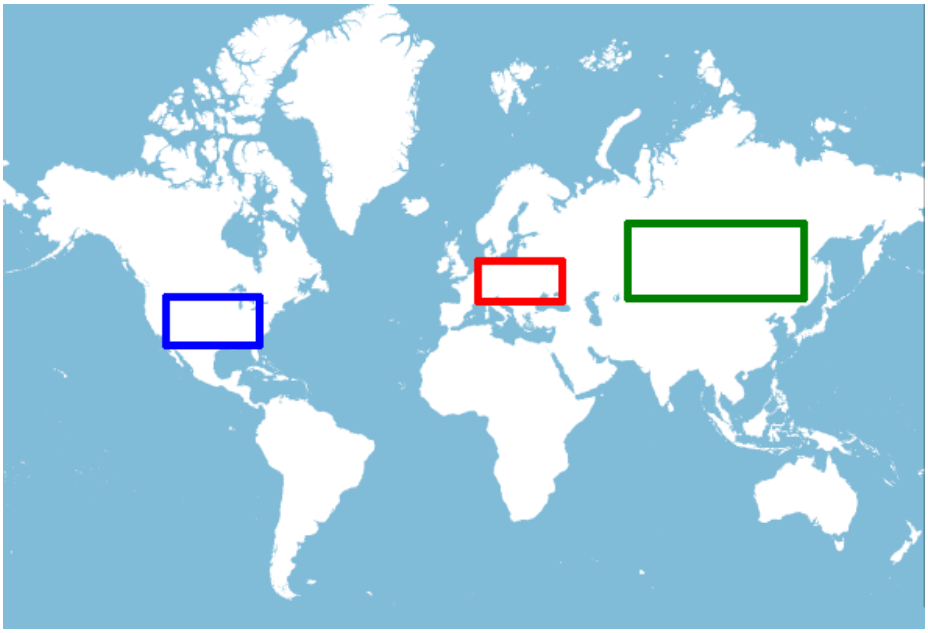

Figure 3-1: The location of the three regions considered. For simplicity and efficiency of our algorithm, they are chosen to be large rectangular areas (in latitude-longitude coordinates) without sea or ocean.

known statistics for DPRM. Gathering large data sets of millions of paths on three continents, we compute the probability distribution of path length and travel time as a function of the distance between the end points. As in the case of DPRM, appropriately scaled fluctuations can be collapsed approximately to a single curve, suggesting that details of the local structure of road networks are irrelevant to the statistics on larger scales. The local environment can be modeled by a power-law distributed noise with, remarkably, a universal decay exponent. Furthermore, we show that long-range correlations in the environment, on the scale of hundreds of kilometers, affect the scaling exponents and are thus relevant to the statistics of optimal paths. The transverse wandering of the paths is also found to be consistent with our modeling as a directed polymer.

Let us first make precise the DPRM model that is relevant to road networks, and summarize the relevant results. The appropriate model to consider here is the pt-pt DPRM, where the directed polymer is pinned at its ends, and sufficiently stretched to prevent overhangs. Its wanderings can thus be described by a function $h(x)$, where $x$ is a coordinate along the axis between the end points and $h$ the distance from

this axis. (Note that these variables are distinct from those introduced in Eq. 1.1 for interface growth. For instance, $x$ is traditionally denoted $t$ as a time direction, but we stick here to the spatial notation to avoid confusion with travel times.) The energy of a configuration of is then given by

$$E[h(x)] = \int_0^d dx \left[ \frac{\gamma}{2} \left( \frac{dh}{dx} \right)^2 + V(x,h) \right] , \tag{3.1}$$

where $d$ is the distance between the end points, $\gamma$ is related to the line tension of the directed polymer, and $V$ is a random potential modeling a disordered environment. In the zero-temperature limit, relevant to our problem, the free energy is simply the energy of the optimal path $E[h^*]$. Two exponents govern the scaling of the energy fluctuations $\langle (E - \langle E \rangle)^2 \rangle \sim d^{2\beta}$ (where the brackets denote an average over realizations of the disorder $V$) and the transverse wandering of the optimal chains, $\langle h^*(x)^2 \rangle \sim x^{2\zeta}$ [32]. If $V$ has only short-ranged correlations, the exponents $\beta = 1/3$ and $\zeta = 2/3$ are known exactly [3], and the limiting distribution is TW-GUE [15]. On the contrary, long-range correlated disorder leads to larger scaling exponents and different energy distributions [1, 46, 51, 55, 73].

## 3.2 Statistics of optimal paths

In light of these theoretical results, we now analyze the statistics of two types of optimal paths (the shortest and the fastest) on the road network. We compute the paths using the Open Source Routing Machine (OSRM) [74] operating on OpenStreetMap data, a collaborative effort to provide an open source map of the world. The fastest paths are determined using the default configuration of OSRM which takes into account speed limitations for cars and road types but no information on traffic. We gather six data sets for the two types of optimal paths in the three regions indicated in Fig. 3-1, sampling the end points of the paths uniformly on the network.

In Fig. 3-2, we show examples of optimal paths drawn from an arbitrary center point (near Munich, Germany) to uniformly sampled points at a 300 km distance.

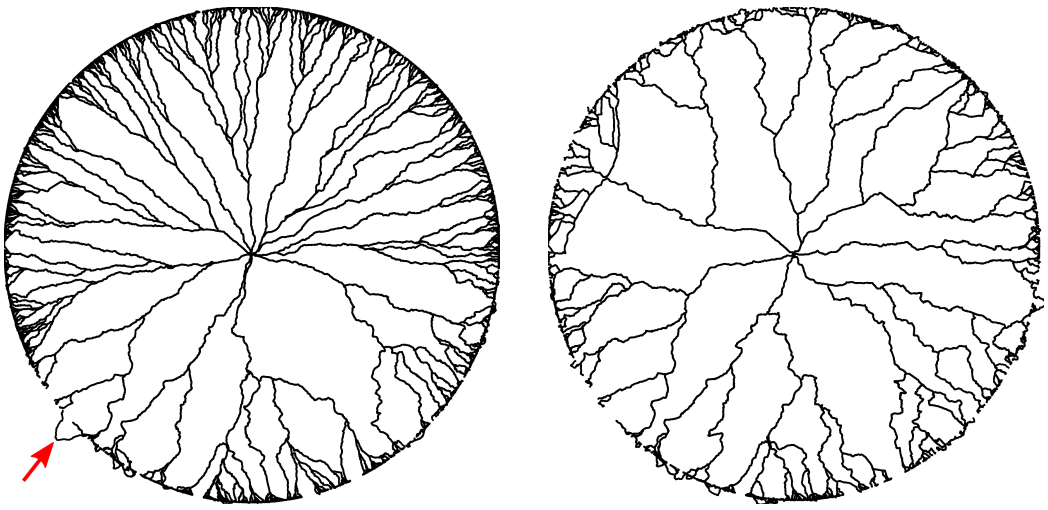

Figure 3-2: Shortest (left) and fastest (right) paths from a central point (near Munich, Germany) to $10^4$ randomly chosen points at a distance of 300 km. The arrow points to the most prominent overhang in the paths.

Both sets of optimal paths display a fractal branching pattern strongly resembling what is observed in directed polymer models [32]. However, these routes are not perfectly directed. This is especially visible near the end points where the local structure of the road network may impose overhangs (the most prominent is indicated by a red arrow in Fig. 3-2). This aligns well with the everyday experience of commuting, where depending on the neighbourhood, the optimal path may require backtracking and looping around side streets in order to reach the main street or highway. Nevertheless, overhangs are mostly avoided by the optimal paths on the rest of the trajectory, as quantified in Fig. 3-3 where we plot the average length of the paths $\langle L \rangle(d)$ and the part $\langle L_h \rangle(d)$ corresponding to overhangs (see Appendix B for a precise definition of overhangs). The average path length $\langle L \rangle$ is found to increase linearly with $d$ at large distances while $\langle L_h \rangle$ increases slower. Overhangs thus become less relevant at larger distances for which we expect a better comparison between road paths and directed polymers. In the following, we divide accordingly our study between short paths that are strongly constrained by the network, and longer paths which result from optimization.

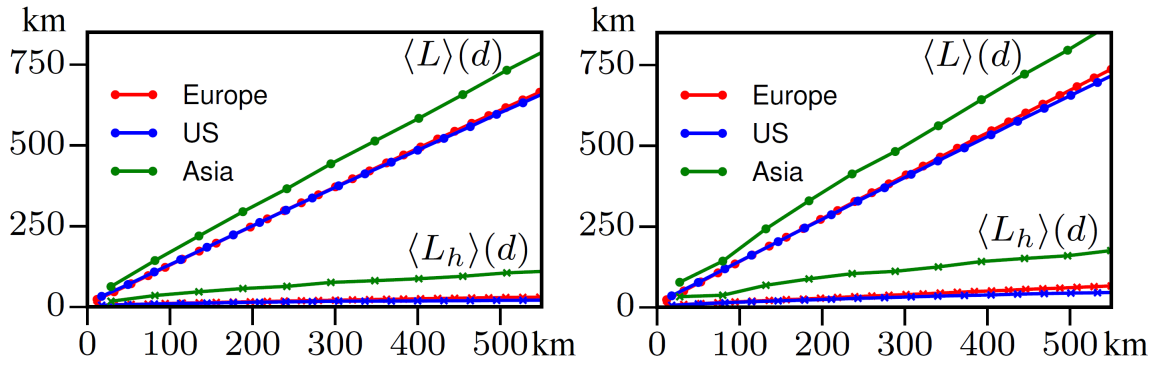


Figure 3-3: Average length of the optimal paths $\langle L \rangle$ and length of overhangs $\langle L_h \rangle$ as a function of the distance $d$ between the end points. All lengths are measured in km. Left: Shortest paths. Right: Fastest paths; $10^6$ points for each curve. Lines are guides to the eye.

### 3.2.1 Paths over short distances

We first look in Fig. 3-4 at the distribution of the length $L$ of the shortest paths (respectively the travel time $T$ on the fastest paths) between points at a short distance $d = 1$ km. We are interested in $L$ and $T$ as the quantities that are minimized and thus, in our interpretation, akin to the energy of a directed chain. The distributions display clear power-law tails at large $L$ and $T$ over more than three orders of magnitude. The tails correspond to situations where the path has to go around an obstacle to reach a nearby point, *e.g.*, reach the next bridge to cross a river. They thus characterize the overhangs described previously. Most remarkably, the decay exponent $P(L) \sim L^{-\alpha}$ [and $P(T) \sim T^{-\alpha}$] seems to be universal across continents with $\alpha \approx 3$ (the best-fit coefficients for the six curves are all found within $[2.89, 3.10]$). This appears surprising since we expect the paths at small $d$ to reflect the local structure of the road network which is *a priori* very different in the three regions considered. Although we lack an explanation for the value of the exponent, it can be compared to exponents derived for the same distribution in different environments. The shortest path between nearby points on the backbone of a percolation cluster has been numerically found to exhibit the same $\alpha = 3$ [75] at small distances, while for self-avoiding random walks, the probability of forming a loop of length $\ell$ in a 2D chain scales as $\ell^{-\alpha}$ with (exact) exponent $\alpha = 2.68$ [76, 77]. The latter superficially resembles the configuration of a

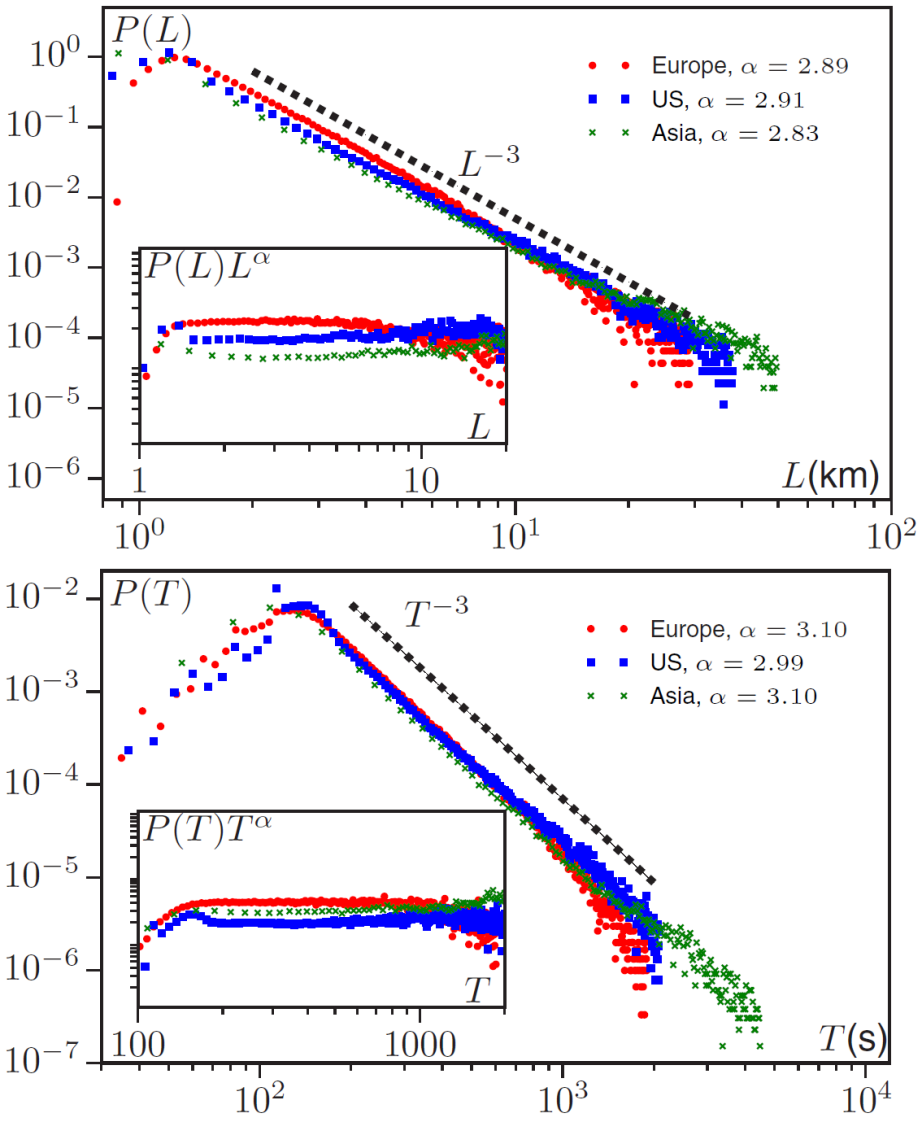


Figure 3-4: Optimal paths between points at a distance $d = 1$ km. Top: Probability distribution of the length $L$ of the shortest paths. Bottom: Probability distribution of the travel time $T$ of the fastest paths; $N = 5 \times 10^5$ paths for each curve. Insets: Scaled plots with the best-fit exponent $\alpha$ indicated in the legend.

road between nearby points that loops around to avoid intervening obstacles, while not intersecting other roads arriving and/or departing the two points.

### 3.2.2 Paths over long distances

Because of the fat tails in the distributions of Fig. 3-4, the variance of $L$ and $T$ is not defined. We thus cannot estimate the exponent $\beta$ characterizing energy fluctuations by simply looking at the scaling of $\langle L^2 \rangle_c(d)$ and $\langle T^2 \rangle_c(d)$, and need to look instead at the full probability distributions $P(L|d)$ and $P(T|d)$ for increasing distance $d$ between the end points. To compare these distributions, we superimpose their maxima and rescale their width by a factor $d^\beta$ where $\beta$ is adjusted so that the distributions converge at large d. The results are shown in Fig. 3-5 (top) for the shortest paths in Europe and in Appendix C for the five other data sets, which show similar behaviour. We find that the exponent $\beta$ can be adjusted such that the left tail of the distribution converges rapidly to a limit distribution well fitted by the TW distribution expected for directed polymers. On the contrary, the right tail converges slower and remains heavy at the largest $d$ attainable (larger $d$, comparable to the total size of the region, show strong finite-size effects). It is thus not clear if the right tail also converges to TW behaviour or to a different distribution, as was observed numerically for a directed polymer model with long-ranged correlations in the environment [46].

For comparison, we simulated the well-established DPRM model on a square lattice with random independent energies on each site [32,36,78]. The paths are directed in the diagonal of the lattice, parametrized by $d$. As before, the distance (in number of sites) from the diagonal is denoted by $h$. The energy of the optimal path is computed recursively as

$$E(d,h) = \eta(d,h) + \min\{E(d-1,h), E(d-1,h-1)\}. \tag{3.2}$$

After $d$ iterations, $E(d,h)$ is then the energy of the optimal path between the point $(d,h)$ and the line $d = 0$. As opposed to previous studies that considered Gaussian noise, we draw the noise $\eta(d,h)$ in a power-law distribution $P(\eta) = 2\eta^{-3}$ with

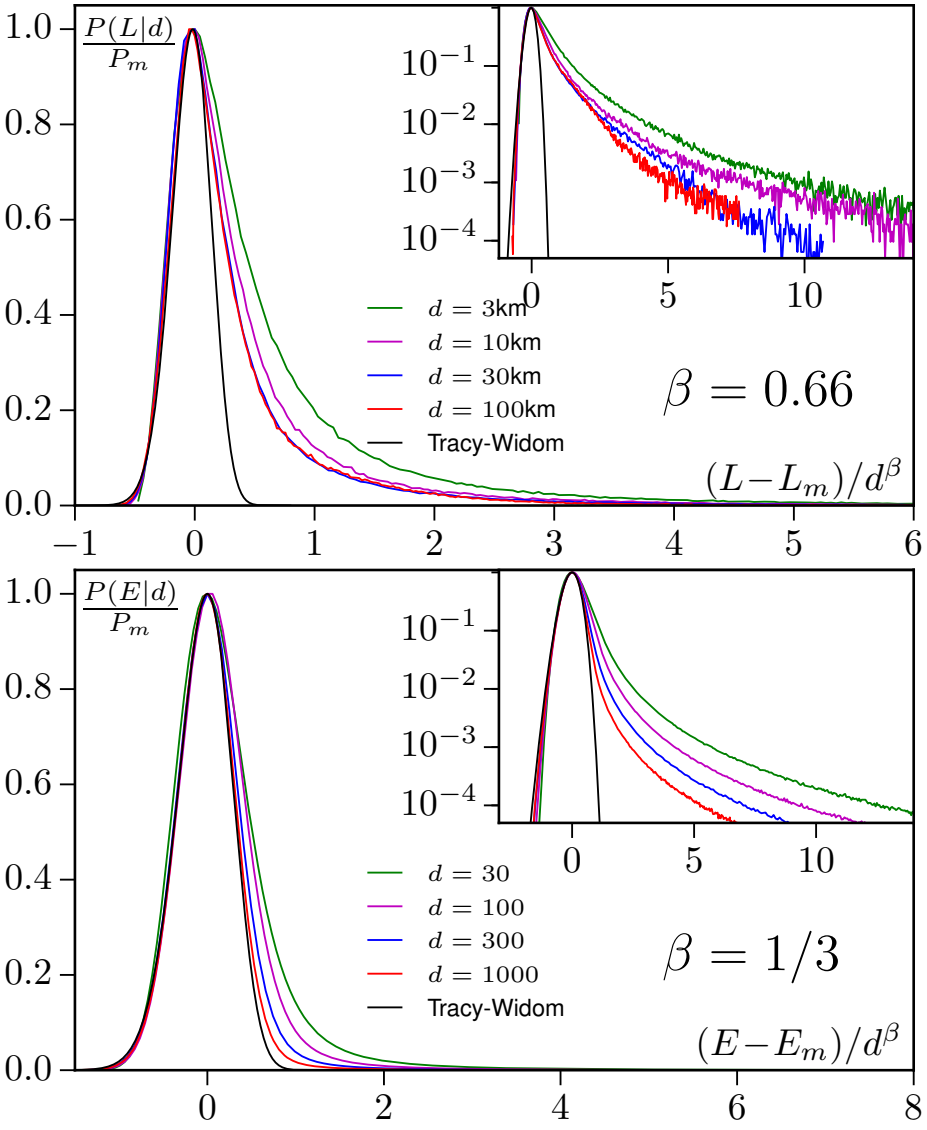


Figure 3-5: Top: Probability distribution of the length $L$ of the shortest paths in Europe rescaled with $\beta = 0.66$; $5 \times 10^5$ paths for each curve. Bottom: Probability distribution of the energy for the DPRM model with power-law noise rescaled with $\beta = 1/3$; $10^7$ paths for each curve. Lattice size of $10^7$ in the transverse direction with periodic boundary conditions. The insets show the same data with a logarithmic $y$-axis. $P_m$ denotes the maximum of the distribution, found at $L = L_m$ or $E = E_m$.

$\eta \in [1, \infty)$ to match qualitatively the short-scale distributions in Fig. 3-4. We then analyze the results as in the experimental case: We shift the energy distributions $P(E|d)$ to superimpose their maxima and rescale their width by $d^\beta$ (Fig. 3-5, bottom). We observe that, as with Gaussian noise [36], the distribution converges to a TW distribution with the KPZ exponent $\beta = 1/3$. Indeed, only a fat tail in the noise at *negative* energy, $P(\eta) \sim \eta^{-a}$ as $\eta \to -\infty$ would change the scaling exponents [40, 79]. Interestingly, the convergence when increasing $d$ happens in a similar manner in the model and the experimental data, with the right tails converging much slower. This also lends credit to our measure of $\beta$ as the exponent rescaling the left tail of the distributions.

## 3.3 Long-range correlations in the road network

One salient difference remains between the paths on the road and the directed polymer model: The measured $\beta$ exponents are found between 0.58 and 0.9 (with an error estimated around 15%) and are thus much larger than $\beta = 1/3$ in the KPZ universality class. We now argue that this can be explained by the presence of long-range correlations in the road network. To show this, we first discretize the full map of each region in squares of size 100m $\times$ 100m and assign the value $\rho(\mathbf{r}) = 1$ if a road is found inside the square and 0 otherwise. We then compute the correlation function $C(r) = \langle \rho(\mathbf{r})\rho(\mathbf{r}+\mathbf{x})\rangle - \langle \rho(\mathbf{x})\rangle^2$ where the average is taken over $\mathbf{x}$ and orientations of $\mathbf{r}$. As shown in Fig. 3-6, $C(r)$ decreases slowly [slower than $C(r) \sim r^{-0.5}$], remaining non-negligible on the scale of hundreds of kilometers. These long-range correlations reflect the fact that the road network is shaped by many factors acting at every scale, from different administrative divisions to natural obstacles. They were also shown to be important in modeling the development of cities [80], obviously related to that of the road network.

As we discussed in Chapter 2, such a slow decay of correlations has been proven to be relevant to the large-scale behaviour, both in numerical simulations [46, 73, 81] and analytic calculations [1, 51, 82, 83]. For Gaussian noise with isotropic correlations

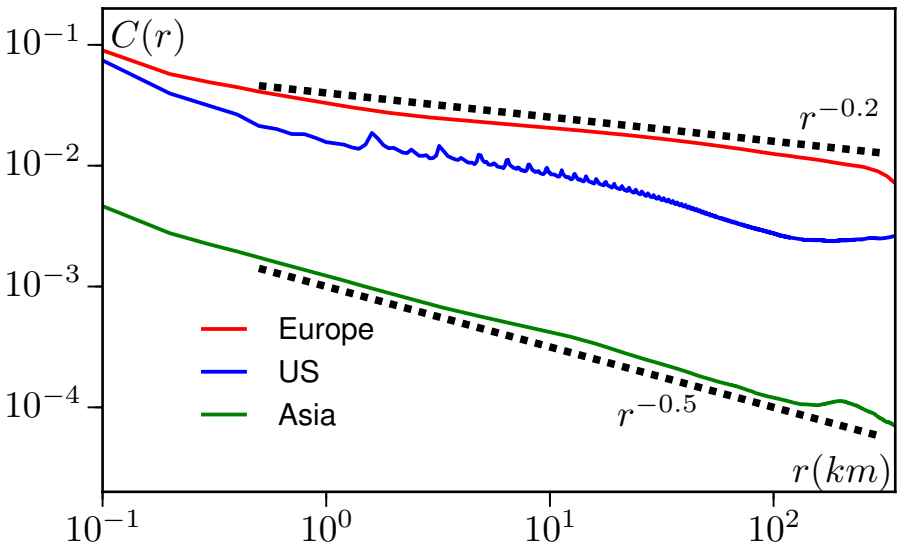


Figure 3-6: Autocorrelation functions of the road density as defined in the text. The oscillations in the curve for the US are not an artifact. The peaks are located every mile (with sub-peaks at half-miles) and correspond to large regions (up to 60 miles) of grid-like road network.

decaying as a power law with exponents between $-0.5$ and $-0.2$ (as measured for the road density correlations in Fig. 3-6), $\beta$ was measured between 0.5 and 0.7 [73]. Given experimental uncertainties, these values are in relatively good agreement with our measurements for the road network. Long-range correlations are thus likely to be at the origin of the large exponents observed.

## 3.4 Transverse fluctuations

Finally, we look at the wandering of the optimal paths in the transverse direction. The routing algorithm returns a list of points along each path (on average every 50m) that we use to construct the function $h(x)$, the distance to the end-to-end axis parametrized by $x$. We do so by discretizing $x$ in bins of size $dx = 100$m and averaging points falling in the same bin. This discards any overhang and thus produces a directed path approximating the real path. The leading behaviour is expected to be scale invariant, $\Delta h(x) = \sqrt{\langle h^2(x) \rangle} \sim x^{\zeta}$. However, because of overhangs near the end points, $h(0) \neq 0$, so that $\Delta h(0) \neq 0$, inducing large corrections to the

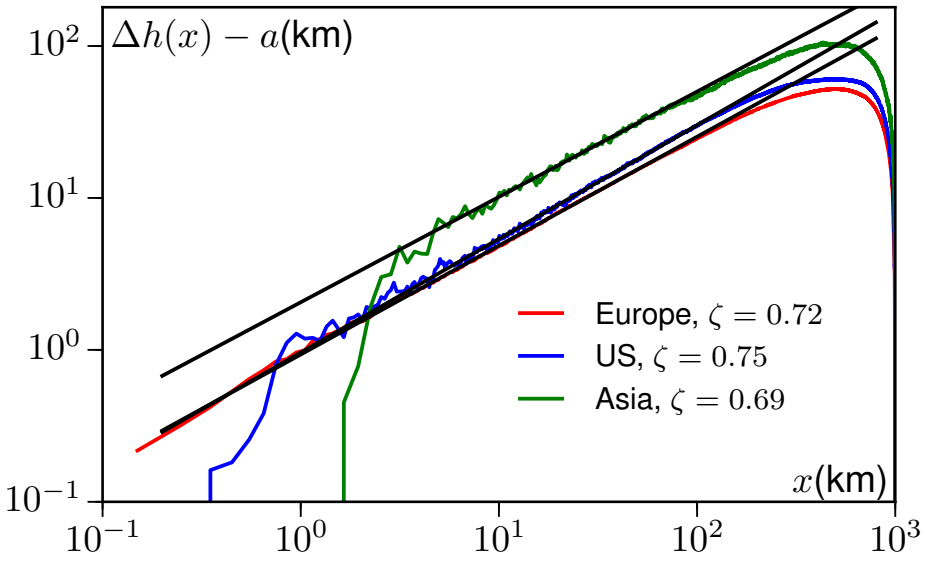


Figure 3-7: Transverse wanderings for the shortest paths as a function of the coordinate $x$ on the axis between the end points. Averages are over $5 \times 10^5$ paths between points at distance $d = 1000$ km.

putative scaling. Thus, as a first approximation, we estimate the $\zeta$ exponent by fitting $\Delta h(x) = a + bx^{\zeta}$ with free parameters $a$, $b$ and $\zeta$. The resulting functions $\Delta h(x) - a$ show scaling behaviour over two orders of magnitude with exponents $\zeta \in [0.69, 0.75]$ (see Fig. 3-7). Once again these values are larger than the KPZ exponent $\zeta = 2/3$, in qualitative agreement with the presence of long-range correlations that are expected to increase the value of $\zeta$. For comparison, isotropic long-range correlations with a decay exponent in the range of Fig. 3-6 give $\zeta \in [0.75, 0.85]$ [73] while correlations only in the transverse direction yield $\zeta \in [0.67, 0.72]$ [46].

# Chapter 4

# Evolution in range expansions with competition at rough boundaries

The results in this chapter are reproduced and adapted from Ref. [84].

When a biological population expands into new territory, genetic drift develops an enormous influence on evolution at the propagating front, and this front resembles the class of stochastic growing interfaces described by the KPZ equation. In such range expansion processes, fluctuations in allele frequencies occur through stochastic *spatial* wandering of both genetic lineages and the boundaries between genetically segregated sectors. Laboratory experiments on microbial range expansions have shown that this stochastic wandering, transverse to the front, is superdiffusive due to the front's growing roughness, implying much faster loss of genetic diversity than predicted by simple flat front diffusive models. We study the evolutionary consequences of this superdiffusive wandering using two complementary numerical models of range expansions: the stepping stone model, and a new interpretation of the DPRM model, in the context of a roughening population front. Through these approaches we compute statistics for the times since common ancestry for pairs of individuals with a given spatial separation at the front, and we explore how environmental heterogeneities can locally suppress these superdiffusive fluctuations.

## 4.1 Evolutionary dynamics in expanding populations

In evolutionary biology, changes in an allele's frequency in a population are driven not only by Darwinian selection but also by random fluctuations, the phenomenon of genetic drift. Selectively neutral or even deleterious alleles can rise to prominence purely by chance. In many scenarios an individual competes directly only with a small subset of the population, *e.g.* due to spatial proximity, and this small effective population size increases the influence of genetic drift [85].

Range expansions provide an important example: When a population expands spatially into new territory, as during species invasion or following environmental changes, the new territory is dominated by the descendants of a few ancestors at the expansion front. Genetic drift is amplified by the small effective population size at the front [85] – the founder effect – and by the related phenomenon of gene "surfing", in which alleles that happen to be present at the front spread to high frequency in the newly occupied space, despite being selectively neutral or even deleterious [86, 87].

Genetic drift in range expansions strongly ties fluctuations in allele frequencies to spatial fluctuations. In laboratory experiments, Hallatschek et al. [86] have shown that microbial range expansions develop, after a short demixing time, genetic sectors containing almost exclusively the descendants of a single individual. Thereafter, genetic drift occurs through *spatial* fluctuations of the sector boundaries, with a sector lost from the front each time two sector boundaries intersect. Similarly, the genealogical ancestry tree traced backward in time from the front becomes a tree of space curves that fluctuate transversely to the front propagation direction and coalesce upon intersection [88]. (See Fig. 4-2.)

The reverse-time coalescence of lineages is of central importance in population genetics, particularly in the approach known as coalescent theory [89, 90]. One of the key estimates of interest in coalescent theory is the expected number of pairwise site differences $\Pi$ between two sampled genomes, which is proportional to the expected time since common ancestry of the two sampled individuals, $T_2$, under the assumption that neutral mutations have accumulated in the (very long) genome at a

constant rate since the two lineages diverged. The relation $\Pi \propto T_2$ allows inferences to be made about the population's recent evolutionary past from measured genomic differences in the present, given reliable models of genealogy. The *structured* coalescent, which extends coalescent theory to populations with spatial structure (as opposed to well-mixed populations) [91], typically assumes migration rules that produce diffusive dynamics for gene flow. Theoretical studies of the genealogical structure of range expansions have similarly assumed diffusive spatial fluctuations of genetic boundaries (as would be appropriate to a flat front range expansion model; see below) in the interests of analytical tractability [85]. Flat front models are equivalent to conventional stepping stone models [92] and many exact results are available [93].

However, there is strong evidence that evolutionary dynamics in range expansions are often driven by *superdiffusive* spatial wandering of both genetic sector boundaries and lineages. Hallatschek et al. [86] measured the mean-square transverse displacement of sector boundaries in *E. coli* growing across hard agar Petri dishes, and found it to scale with the expansion distance $y$ as $y^{2\zeta}$ with wandering exponent $\zeta = 0.65 \pm 0.05$, greater than the value of $\zeta = 1/2$ characterizing diffusive wandering. In both *E. coli* and the yeast species *Saccharomyces cerevisiae*, genetic lineages similarly fluctuate with wandering exponent $\zeta \approx 2/3$ [88]. The same superdiffusive wandering exponent was found numerically for genetic lineages in an off-lattice model of microbial colony growth [88] and for sector boundaries in a two-species Eden model [85, 94]. Consequently, the number of distinct sectors decreases as $y^{-\zeta}$, with $\zeta$ measured to be $\approx 0.67$ [94], a dramatically faster loss of genetic diversity than the $y^{-1/2}$ scaling that would result from diffusive dynamics [85]; see Fig. 4-2, where genetically neutral strains are competing.

The underlying cause of this superdiffusive behaviour is that the population front profile has a characteristic roughness that increases with time. Because the range expansion causes the front to advance along its local normal direction, stochastically generated protrusions in the front are self-amplifying, and the lineages and genetic sector boundaries moving with these protrusions experience a faster-than-diffusive average lateral motion. Such roughening fronts are precisely characterized by the

KPZ equation [1, 2]. Recall that in $d = 1 + 1$ dimensions, the growth exponent $\beta = 1/3$ and the wandering exponent $\zeta = 2/3$ are known analytically [4, 5]; this value of $\zeta$ nicely matches the measured value from experiments and simulations of the microorganism range expansions discussed above.

While there is a wealth of literature on the KPZ equation and its rich universality class [9, 34, 95], there does not yet exist a similar understanding of the statistics of coalescing space curves – here, lineages and genetic sector boundaries – whose superdiffusive wandering is driven by such KPZ roughening. We term these curves "KPZ walkers" in contrast to diffusive random walkers. In developing a quantitative understanding of neutral evolution in a biological range expansion, we are thus led to new questions in statistical physics.

## 4.2 Models for microbial growth

We employ a complementary pair of simulation approaches: The first, a lattice-based stepping stone model, introduces front roughness through stochasticity in replication time. In our second approach, we reinterpret the problem of DPRM [32], introduced in Chapter 1.2.1, as a model for range expansions with stochastic variation in organism size. The DPRM approach can be simulated at large scales with much less computational expense than our stochastic stepping stone model. We also apply analytical results from the DPRM problem to rationalize the measured asymptotic coalescence behaviours. Finally, we study numerically how environmental heterogeneities temporarily suppress the wandering of KPZ walkers, an effect observed recently in experiment [96].

### 4.2.1 Stepping stone model

The stepping stone model [92] imagines a biological population arranged on a spatial lattice of individually well-mixed subpopulations called "demes", each containing $N$ individuals, with exchange of individuals between neighbouring demes. We implement the stepping stone model on a triangular lattice with $N = 1$ individual per deme,

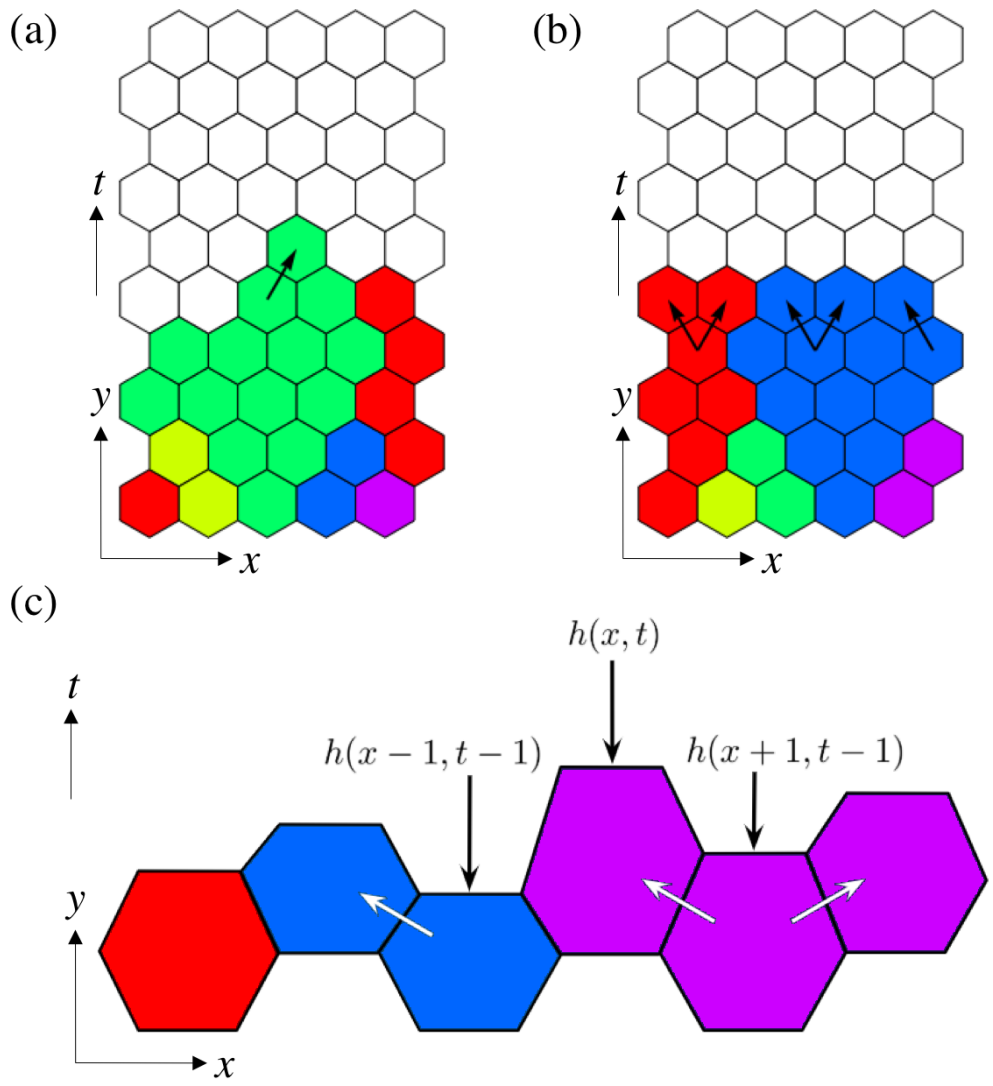


Figure 4-1: Illustrations of the the update rules in our numerical models of range expansions. (a,b) The stepping stone model with deme size $N = 1$ on a triangular lattice, using (a) rough front and (b) flat front update rules. We visualize each individual on the initial line and its descendants with a distinct colour. (c) DPRM model of range expansion. At horizontal position $x$, the height of the front in the $y$-direction, $h(x,t)$, is increased by a quantity that depends on the two adjacent heights, namely $\max\{h(x-t,t-1)+\eta, h(x+1,t-1)+\eta'\}$, where $\eta$, $\eta'$ are zero-mean stochastic noise terms that cause front roughness. The nearest neighbour cell which maximizes the above relation is chosen to reproduce, and passes on its allele label (denoted by the colour), as represented by white arrows in the illustration.

which models cases in which local fixation of one allele occurs rapidly compared to spatial diffusion [85].

As an initial condition, we take the lattice of demes in two dimensions to be unpopulated except for a linear inoculation "homeland". Once a deme is populated, its allele remains unchanged thereafter, as in the microbial experiments on agar plates, where cell divisions occur only near the frontier, so that the spatial pattern of alleles is effectively frozen behind the front [86]. We choose as our update rule that of the Eden model for two-dimensional growth processes [45] (see Chapter 1.2.2): One site is chosen at random from among all occupied sites with some empty neighbour site, and the allele is copied from the chosen occupied site into a randomly chosen empty neighbour (Fig. 4-1a). The key difference from the traditional Eden model, however, is that we track not only whether a site is occupied or empty, but also the allele label and genetic ancestry of occupied sites. The simulations use a system width of $L = 2000$ sites, and are evolved until the front has advanced a height $h = 1000$ sites. Results are taken from ensembles of 5000 realizations. Periodic boundary conditions are used in the direction transverse to the mean front propagation. In order to avoid finite size effects, we keep the system width $L$ at least twice as large as the maximum time $t_{\max}$, so that no lineage or sector boundary can wind completely (or even halfway) across the system.

By introducing stochasticity in the replication time, this procedure generates an irregular interface between the occupied and empty regions (see Fig. 4-2a), simulating a rough front range expansion. By contrast, the expansion front remains flat if the update rule fills an entire row in parallel (Fig. 4-1b), with each newly filled site inheriting the allele marker of one of its two filled neighbours below, chosen randomly with equal probability. The dynamics in Fig. 4-1b is equivalent to a one-dimensional stepping stone model in discrete time with deme size $N = 1$. The $N = 1$ stepping stone model is also studied as the voter model with $L$ different opinions [97]. We note that accelerated coarsening brought about by superdiffusive wandering has been studied for the voter model [98], but with opinions spreading by Lévy flights of algebraically distributed distances, in contrast to the purely nearest-neighbour microscopic dynamics

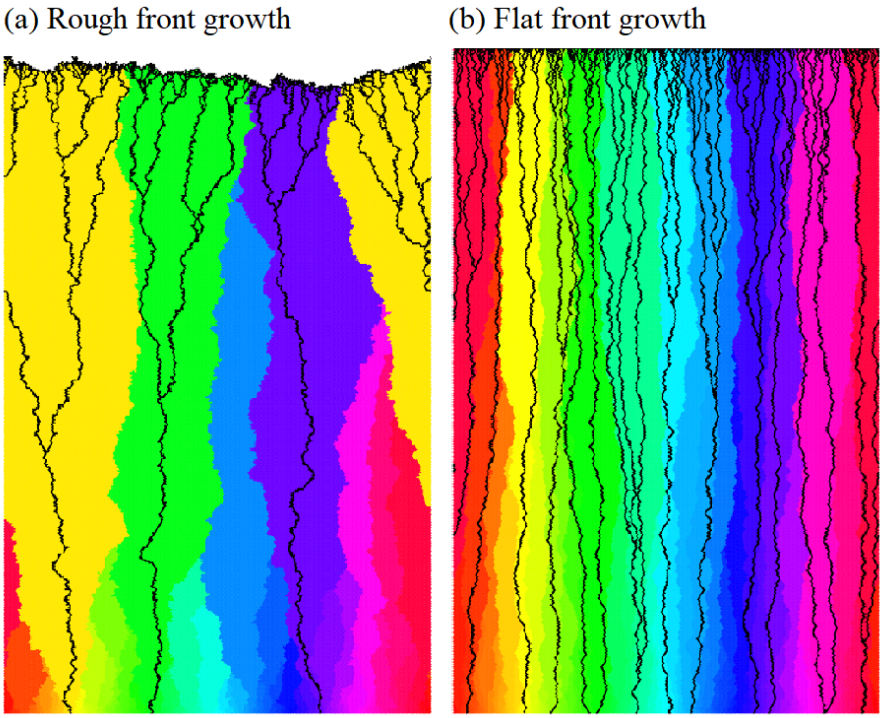


Figure 4-2: Range expansions generated by the stepping stone model, using the (a) rough front and (b) flat front update rules, with periodic boundary conditions in the horizontal direction. The colours represent allele labels, while the black lines mark the genetic lineages. Time runs upward in both cases. Note that there are fewer *sectors* at the top (genetic coarsening), but fewer *lineages* at the bottom (lineage coalescence). Typical coalescence rates are much larger in (a) than in (b).

employed in this work.

### 4.2.2 Geometric interpretation of directed paths

The second model is a reinterpretation of DPRM, where the accumulated "energy" of the directed path, characterized by the KPZ equation, can be mapped to the height of a range expansion front (see Fig. 4-1c). In this mapping, the stochastic noise $\eta$ corresponds to fluctuations in the lengths of individual microbes in the direction of average propagation $y$, about a mean length $\ell$.

We simulate DPRM on a square lattice rotated at 45° to the $x$, $t$ axes (see Fig. 1-4), and optimize over paths from the origin to any site $(x, t)$ using the transfer matrix method [32]. An allele label is added to each site, as in the stepping stone model.

The height of the front $h(x,t)$ is updated according to

$$h(x,t) = \ell + \max\{h(x-t,t-1)+\eta, h(x+1,t-1)+\eta'\}, \tag{4.1}$$

where $\eta$, $\eta'$ are zero-mean, independent and identically distributed random variables. Each site at time $t$ is then filled by the offspring of one of its nearest neighbours from time $t-1$, and inherits the corresponding allele label. The choice of competing mother cells is taken to be the cell that optimizes the relation in Eq. 4.1. Each DPRM directed path is interpreted as a single lineage, and the set of optimal directed paths to all available endpoints forms the lineage tree. The simulated system has width along the $x$-direction $L = 2^{16}$, is evolved over $t_{\max} = 10^4$ time steps, and is averaged over $2^{10}$ realizations. We use periodic boundary conditions in the $x$ direction transverse to the front propagation.

Thus, while replication time is constant in this model, front roughness is generated by stochasticity in cell size, with larger size favoured for propagation. While we assume that the mean cell size at time of division for the microbe in question has already evolved to a fitness maximum, variance in the cell size leads to front roughness and accelerated loss of genetic diversity (Fig. 4-3a).

Note that if we fix $\eta$ to have zero variance, and instead choose the mother cell at random between the left- and right-neighbours, we recover a flat front range expansion with diffusive dynamics associated with lineages and genetic boundaries (Fig. 4-3b). Also, if we reduce the system width to a single organism, the front height $h(x,t)$ performs a random walk about the deterministic value $\ell t$, the variance growing linearly in $t$ with slope given by the variance in $\eta$. A dramatic experimental realization of such a scenario in *E. coli* was demonstrated by the "mother machine" of Wang et al. [99]: Bacteria growing and dividing in narrow channels, quasi-one-dimensionally, show stability in growth rate over hundreds of generations.

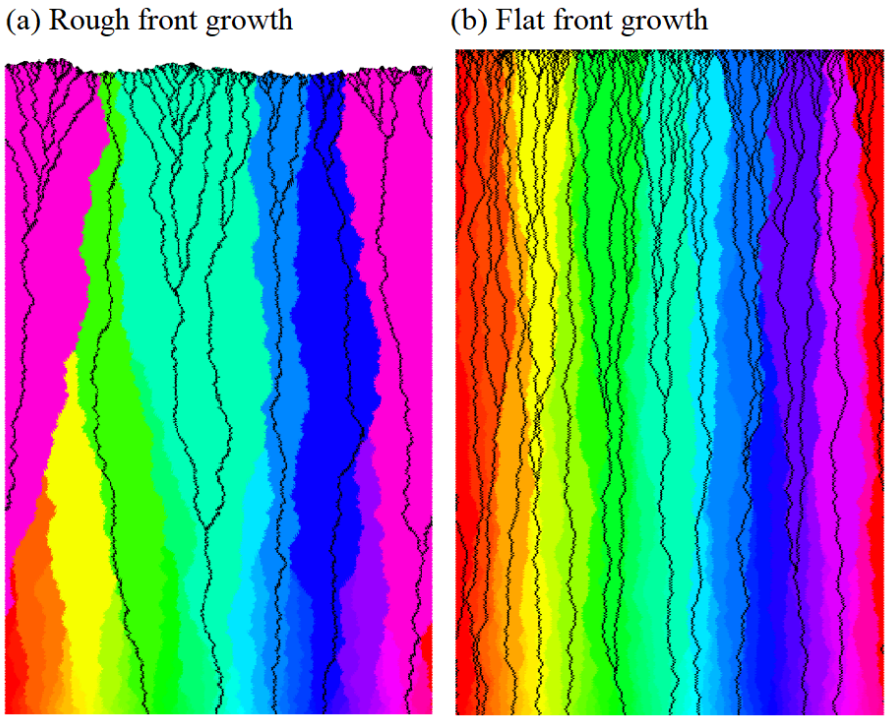


Figure 4-3: Range expansions generated by the DPRM model, with periodic boundary conditions in the horizontal direction, as in Fig. 4-2. The colours represent allele labels, while the black lines mark the genetic lineages. In contrast to the flat front case (b), the rough front case (a) with the same number of generations shows a significantly faster decrease in genetic diversity, and much larger coalescence rates, similar to Fig. 4-2.

### 4.2.3 Superdiffusivity of transverse fluctuations

In both the rough front stepping stone model and the DPRM model, lineages and sector boundaries have superdiffusive lateral fluctuations with wandering exponent $\zeta = 2/3$ [4, 5, 32, 85, 94]. For DPRM models, this behaviour is well-known as the transverse fluctuations of the minimal-energy directed path. In contrast, for the flat front stepping stone model and the zero-noise variant of DPRM, the lateral fluctuations of lineages and sector boundaries are merely diffusive, $\zeta = 1/2$.

This superdiffusive behaviour has stark consequences for the genetic structure of the population. Comparing the flat front and rough front realizations for the stepping stone model in Fig. 4-2 and for the DPRM model in Fig. 4-3, we see striking differences in both the coalescing lineage trees and the decay in the number of surviving monoclonal sectors. Genetic diversity is lost much more rapidly in the rough front case, and nearby individuals at the front are much more likely to have a common ancestor in the recent past, reflecting much larger coalescence rates.

## 4.3 Coalescence of lineages

We investigate numerically the genealogical structure of populations with superdiffusive migration of the KPZ walker type, driven by roughening fronts. We are chiefly interested in how the expected time since common ancestry $T_2$ for a pair of individuals depends on spatial separation $\Delta x_0$ at the front, as well as in the probability per unit time $J(\tau|\Delta x_0)$ of lineage coalescence at time $\tau$ in the past, whose first moment $\int_0^\infty d\tau\, \tau J(\tau|\Delta x_0)$ equals $T_2(\Delta x_0)$. As a first approach to this problem, our work focuses on neutral evolution from a linear inoculation, avoiding effects such as selection, mutualism/antagonism, and geometrical inflation [100], interesting topics of future study.

### 4.3.1 Rate of coalescence

For two lineages separated by $\Delta x_0$ at the front, $J(\tau|\Delta x_0)$ is the probability per unit time for them to coalesce in a common ancestor at reverse time $\tau$. In the diffusive case, on an infinite line, this is the well-known coalescence rate for two diffusive random walkers with diffusion constant $D$ [101]:

$$J_{\text{diff}}(\tau|\Delta x_0) = \frac{1}{\sqrt{8\pi}}\frac{1}{\tau}\left(\frac{\Delta x_0^2}{D\tau}\right)^{1/2}\exp\left[-\frac{1}{8}\left(\frac{\Delta x_0^2}{D\tau}\right)\right]. \tag{4.2}$$

As a function of the dimensionless ratio $\Delta x_0^2/(D\tau)$, this rate behaves as a power law in the limit of large reverse time or small separations at the front, and as an exponential decay in the opposite limit.

Results such as Eq. 4.2, valid here for flat front models, will serve as a useful guide to our investigations of more complex coalescent phenomena at rough frontiers. In population genetics, systems analogous to our flat front models also arise in the continuum limit of one-dimensional Kimura-Weiss stepping stone models [92]. As reviewed in Ref. [85], many exact results for quantities such as the heterozygosity correlation function and coalescent times are available [102–105]. The $x$-coordinate of stepping stone models represents the horizontal axis of flat front simulations such as those displayed in Fig. 4-2b and 4-3b, while its time coordinate maps on to the $y$-axis. Nullmeier and Hallatschek have used a stepping stone model to study how coalescent times change in 1-dimensional populations when one boundary of a habitable domain moves in a linear fashion due to, say, a changing climate [106]. Results from this later investigation could thus be reinterpreted as applicable to a two-dimensional range expansion in a trapezoidal domain, in the flat front approximation with diffusive genetic boundaries.

For superdiffusive lineages, however, the full expression for $J(\tau|\Delta x_0)$ is not known. We focus instead on its asymptotic behaviours using predictions from DPRM and intuition gained from the diffusive case. For lattice models like those in Fig. 4-1, it will be convenient to measure distances $\Delta x_0$ in units of the space-like direction $x$, and $\tau$ in units of the fundamental step in the time-like direction, which amounts to

scaling out the analog of the diffusion constant in Eq. 4.2. We expect on theoretical grounds that $J$ depends on $\Delta x_0$ only through the combination $\Delta x_0/\tau^\zeta$, with exponent $\zeta = 2/3$ as opposed to $\zeta = 1/2$ in the diffusive case. (The coefficient making this combination dimensionless, analogous to $D$, will be system-specific and is suppressed in our notation.)

First, we consider the regime $\tau/\Delta x_0^{3/2} \ll 1$, representing rare coalescence events where lineages located far apart at the front can be traced back to a recent common ancestor. For the analogous regime of $\tau/\Delta x_0^2 \ll 1$ in the diffusive case, the coalescence rate behaves as $J_{\text{diff}}(\tau|\Delta x_0) \sim \exp[-(\Delta x_0/\tau^{1/2})^2]$. We hypothesize a similar decay for the superdiffusive case, as

$$J(\tau|\Delta x_0) \sim \exp\left(-\left(\frac{\Delta x_0}{\tau^{2/3}}\right)^{\gamma'}\right) = \exp\left(-\left(\frac{\tau}{\Delta x_0^{3/2}}\right)^{\gamma}\right) \tag{4.3}$$

for some exponent $\gamma = -\frac{2}{3}\gamma'$. In Fig. 4-4, we plot $-\ln[\Delta x_0^{3/2} J(\tau|\Delta x_0)]$ vs. $\tau/\Delta x_0^{3/2}$ for both the stepping stone model and DPRM on a log-log scale, so that Eq. 4.3 predicts a linear plot with slope $\gamma$. At small $\tau/\Delta x_0^{3/2}$, both sets of data appear linear, confirming the above hypothesized form. The slopes in the linear regime provide estimates of $\gamma = -1.96 \pm 0.03$ for DPRM and $-1.93 \pm 0.02$ for the stepping stone model.

In fact, we can analytically derive this exponential form, including the value of $\gamma$, using the known distribution of directed path endpoints in DPRM [107]. The calculation, given in Appendix D, shows that

$$J(\tau|\Delta x_0) \sim \frac{1}{\tau}\left(\frac{\Delta x_0}{\tau^{2/3}}\right)^{1/2} \exp\left(-\frac{c}{4}\left(\frac{\Delta x_0}{\tau^{2/3}}\right)^3\right), \tag{4.4}$$

where $c$ is a constant of order unity. For $\tau/\Delta x_0^{3/2} \ll 1$, the leading asymptotic behaviour of $J(\tau|\Delta x_0) \sim \exp(-\frac{1}{4}c(\Delta x_0/\tau^{2/3})^3)$ thus corresponds to $\gamma' = 3$, $\gamma = -2$. From the numerical results in Fig. 4-4, we see from DPRM that $\gamma \approx -1.96 \pm 0.03$, and from the rough front stepping stone model we compute $\gamma \approx -1.93 \pm 0.02$. Both numerical results are in good agreement with the analytically derived prediction.

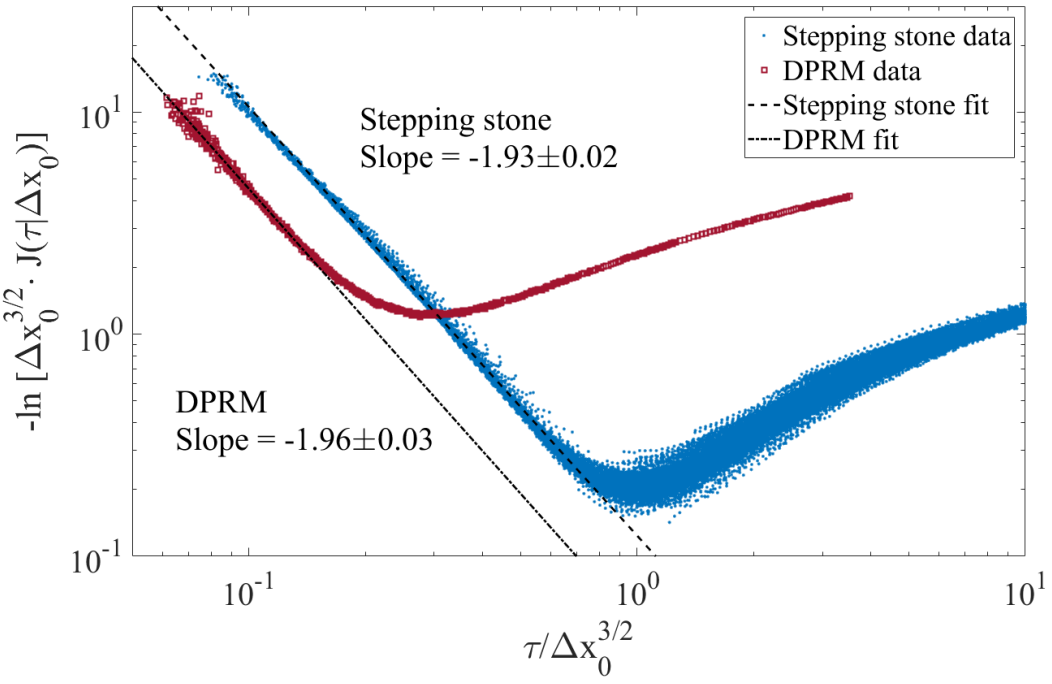


Figure 4-4: Log-log plot of $-\ln[\Delta x_0^{3/2} J(\tau|\Delta x_0)]$ vs. the KPZ-rescaled variable $\tau/\Delta x_0^{3/2}$ for lineages in the stepping stone model and for DPRM. Here, we focus on the regime $\Delta x_0 \ll L$, to avoid finite size effects associated with periodic boundary conditions. Asymptotically for $\tau/\Delta x_0^{3/2} \ll 1$, the relationship is linear, indicating an exponential form for $J(\tau|x_0)$. The fitted slopes are $-1.93 \pm 0.02$ for stepping stone, and $-1.96 \pm 0.03$ for DPRM, providing measurements of $\gamma$ as defined in Eq. 4.3. (For comparison, the DPRM theory predicts a slope of $-2$.)

In the opposite regime of $\tau/\Delta x_0^{3/2} \gg 1$, we can again hypothesize a form for $J$ in analogy with the diffusive case, for which Eq. 4.2 shows $J_{\text{diff}}(\tau|\Delta x_0) \sim \tau^{-1}(\Delta x_0/\tau^{1/2})$. For KPZ walkers, the analogous form is

$$J(\tau|\Delta x_0) \sim \frac{1}{\tau}\left(\frac{\Delta x_0}{\tau^{2/3}}\right)^{\alpha'} = \frac{1}{\Delta x_0^{3/2}}\left(\frac{\tau}{\Delta x_0^{3/2}}\right)^{\alpha}, \tag{4.5}$$

for some exponent $\alpha = -\left(1 + \frac{2}{3}\alpha'\right)$. Although the expression in Eq. 4.4 is consistent with this form, that result is obtained by assuming the two KPZ walkers to be independent (valid at small $\tau/\Delta x_0^{3/2}$), so there is no reason to expect the apparent value of $\alpha' = 1/2$, $\alpha = -4/3$ to hold for $\tau/\Delta x_0^{3/2} \gg 1$.

The rate of coalescence for the two computational approaches in this regime is plotted in Fig. 4-5. The asymptotic behaviour is consistent with the hypothesized power-law decay. The exponent $\alpha$ is determined numerically to be $\alpha = -1.62 \pm 0.03$ for the stepping stone model, and $\alpha = -1.65 \pm 0.01$ for DPRM, giving good agreement between the two models. Furthermore, these values do not rule out the possibility that $\alpha = -5/3$, $\alpha' = 1$, which would give the noteworthy conclusion that $J(\tau|\Delta x_0)$ is linear in the separation $\Delta x_0$, just as in the diffusive case.

We have shown, through both DPRM calculations and a stepping stone model with rough fronts, that the superdiffusive "KPZ walkers" describing genetic lineages have coalescence statistics whose limiting behaviours are qualitatively, but not at all quantitatively, similar to those of coalescing diffusive random walkers. In the limit of large separation or small time in the past, the coalescence rate for KPZ walkers decays as $J \sim \exp[-(\tau/\Delta x_0^{3/2})^{-2}]$, in contrast to the scaling $J_{\text{diff}} \sim \exp[-(\tau/\Delta x_0^2)^{-1}]$ for the diffusive case in the same limit. In the opposite limit of small separation or large time in the past, we find that the coalescence rate varies algebraically as $J \sim \tau^{-1}(\Delta x_0/\tau^{2/3})^{\alpha'}$ with $\alpha' \approx 1$, whereas the corresponding form for diffusive random walkers is $J_{\text{diff}} \sim \tau^{-1}(\Delta x_0/\tau^{1/2})$.

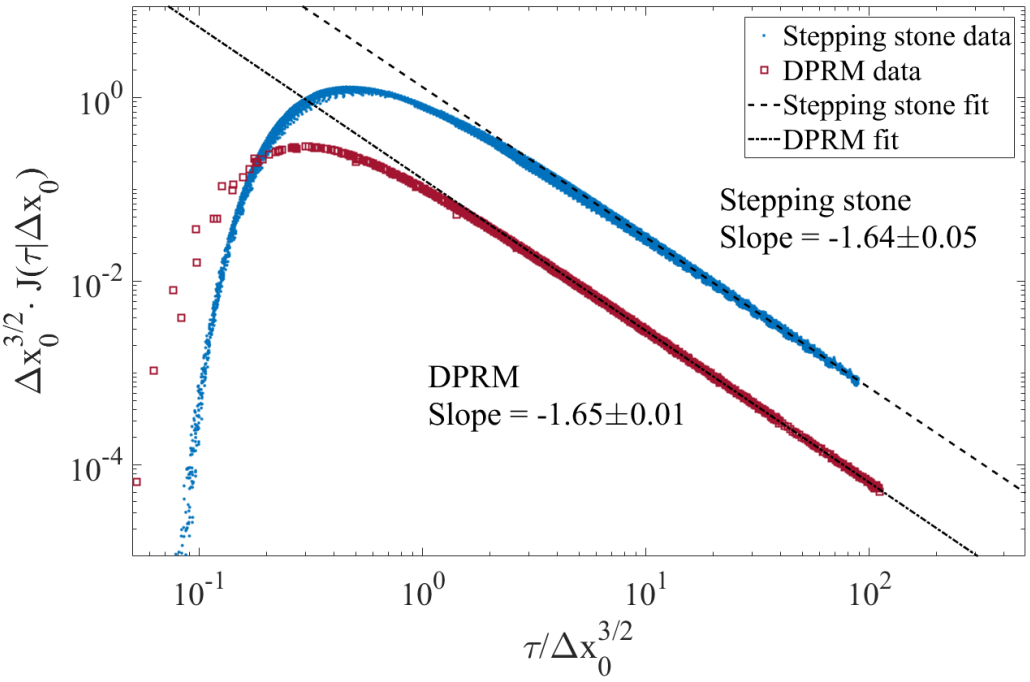


Figure 4-5: Log-log plot of $\Delta x_0^{3/2} J(\tau|\Delta x_0)$ vs. the KPZ-rescaled variable $\tau/\Delta x_0^{3/2}$ for lineages in the stepping stone model and for DPRM. For $\tau/\Delta x_0^{3/2} \gg 1$, the exponent of the power-law decay (Eq. 4.5) is extracted from a linear fit to the numerical data, yielding $\alpha = -1.62 \pm 0.03$ for stepping stone, and $\alpha = -1.65 \pm 0.01$ for DPRM. As in Fig. 4-4, we work in the limit $\Delta x_0 \ll L$ to avoid effects due to periodic boundary conditions.

### 4.3.2 Expected time to coalescence

For a range expansion that has proceeded for a time $t_{\max}$ after a linear inoculation, if two lineages separated by $\Delta x_0$ share a common ancestor on the initial line, we can calculate their expected time to coalescence (time since common ancestry) as

$$T_2(\Delta x_0, t_{\max}) \equiv \frac{\int_0^{t_{\max}} d\tau \; \tau J(\tau|\Delta x_0)}{\int_0^{t_{\max}} d\tau \; J(\tau|\Delta x_0)}. \tag{4.6}$$

Note that the denominator represents normalization by the probability that the two lineages do indeed coalesce.

In the case of diffusive lineages, Eq. 4.2 leads to an analytic expression for $T_2$,

$$\frac{T_{2,\text{diff}}(\Delta x_0, t_{\max})}{t_{\max}} = \left(\frac{\Delta x_0^2}{8Dt_{\max}}\right) \frac{\Gamma\left[-1/2, \Delta x_0^2/8Dt_{\max}\right]}{\Gamma\left[1/2, \Delta x_0^2/8Dt_{\max}\right]}, \tag{4.7}$$

where $\Gamma(x, y)$ is the incomplete gamma function. In Fig. 4-6 we compare the numerical $T_2$ data for KPZ walkers in the rough front stepping stone model with the analytical prediction from the diffusive case under the same conditions. For large $\Delta x_0$, in principle $T_2$ approaches $t_{\max}$; our data do not show this saturation because lineage coalescence events at $\tau \approx t_{\max}$ are so rare that the statistics become poor as $\Delta x_0$ approaches $t_{\max}$. The behaviour for small $\Delta x_0$ is controlled by the scaling in Eq. 4.5: an approximately linear scaling leading to $T_2 \sim \Delta x_0 t_{\max}^{1-\zeta}$. We see that lineages with the same separation $\Delta x_0$ coalesce much faster on average when they behave as KPZ walkers, and that this difference becomes more pronounced for large $t_{\max}$, as is evident qualitatively from Figs. 4-2 and 4-3. The scaling of $T_2$ for KPZ walkers can be written in a form analogous to Eq. 4.7, and reflects the KPZ transverse scalings inherent in the system (see Appendix E).

In biological terms, common ancestry is expected to be more recent with rough front dynamics than under diffusive dynamics. As a result, assuming a constant rate of neutral mutations, the number of differences $\Pi(\Delta x_0)$ between pairs of two sampled genomes at the front is expected to increase more slowly with separation $\Delta x_0$ along the front. This anomaly arises because we expect the habitat to be populated by

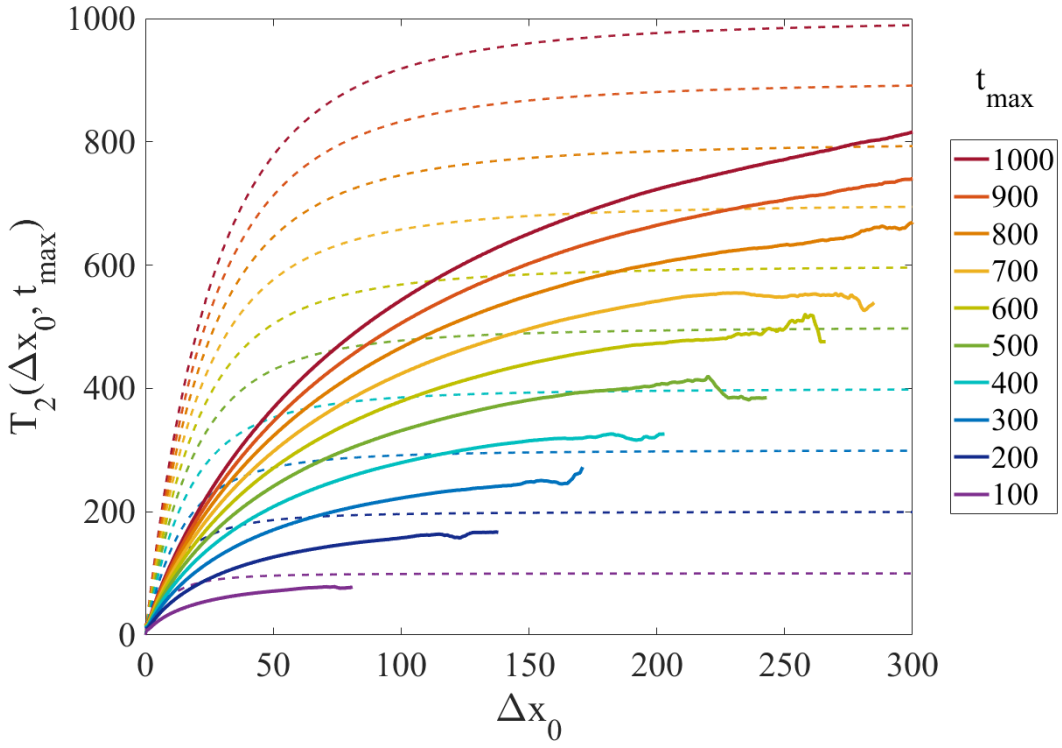


Figure 4-6: Average time $T_2$ since common ancestry for pairs of individuals with some common ancestor and with separation $\Delta x_0 \ll L$ at the front, and for a range of system expansion times $t_{\max}$. Solid lines represent numerical data for KPZ walkers in the stepping stone model, and dashed lines represent analytical predictions for diffusive walkers with the same parameters. The plateau values are simply $t_{\max}$.

the offspring of a small number of common ancestors, which decays as $t^{-2/3}$ for KPZ walkers, rather than the $t^{-1/2}$ decay characterizing diffusive random walkers, where $t$ is the time since the initial inoculation. We have thus shown that the superdiffusive wandering of lineages suppresses $T_2$ significantly compared to estimates based on diffusive dynamics. Our results go beyond the known scaling difference between diffusive and KPZ lineages and genetic boundaries, and provide quantitative information about how front roughness leads to more recent, and fewer, common ancestors for the "pioneers" comprising the front.

## 4.4 Environmental Heterogeneities

The presence of environmental heterogeneities in the habitat can have a significant impact on a range expansion, including on the front shape and propagation speed, and on the genetic diversity at the front. A prototypical example of environmental heterogeneity is the obstacle, a nutrient-depleted zone, that the population must grow around rather than through. As we show here, two different types of KPZ fluctuations come into play when an obstacle is present.

Range expansions around an obstacle were studied experimentally and via simple geometrical optics ideas by Möbius et al. [96] (see also [108]). A notable feature of the experimental (and numerical) results from Ref. [96] is that the sector boundary which forms at the apex of the obstacle shows suppressed transverse fluctuations compared to all other sector boundaries. As the front propagates past the obstacle, a component of its velocity is directed inward from both sides. This in effect pins the sector boundary to the middle, at a kink in the front, and suppresses its fluctuations.

While we have considered only fluctuations of lineages until now, the fluctuations of sector boundaries are inextricably related, as a lineage necessarily remains inside a single sector. Since the lineage fluctuations grow in reverse time as $\tau^{\zeta}$, their coalescence causes the number of distinct lineages to decay as $\tau^{-\zeta}$. Thus for a front at time $t$, the number of roots that the lineage tree has in the initial population decays as $t^{-\zeta}$. As this number of roots equals the number of sectors, the sector boundaries must fluctuate in forward time as $t^{\zeta}$.

Here, we study the suppression of sector boundary fluctuations by obstacles in greater detail using the stepping stone model with a rough front. A gap of width $w_{\mathrm{gap}}$ of unoccupied sites is left in the initially populated line, providing a simplified representation of a range expansion past an obstacle of such width, or the result of an environmental trauma (Fig. 4-7a). By considering only two "alleles" (colours) and using hard-wall boundary conditions, we can track the wandering of the single sector boundary that forms approximately above the center of the obstacle. We examine only times sufficiently early that the system's finite width cannot affect the sector

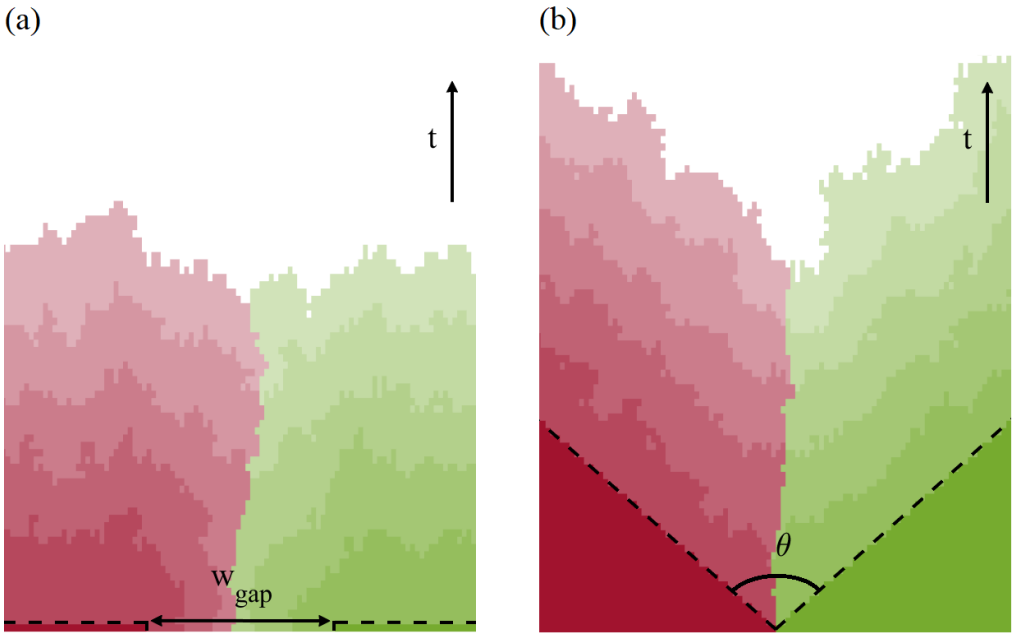


Figure 4-7: Geometries of the sector boundary between two alleles (labeled red and green). The initial inoculations are marked by dashed lines. (a) Illustration of the gap geometry: A segment of width $w_{\text{gap}}$ is left unpopulated initially, separating the two alleles which grow from an otherwise flat initial condition. The width $w_{\text{gap}}$ could represent, say, the width of a square obstacle that terminates at time $t = 0$, or the size of an interval along the horizontal $x$-direction where all organisms are removed by an environmental trauma. (b) Illustration of the wedge geometry: The initial population occupies two triangular regions whose growth fronts meet at a wedge angle $\theta$. In both systems, the two alleles meet at a single sector boundary, along which fluctuations are suppressed. The front of the range expansion is illustrated for a series of equally spaced time values $t$, with lighter shades representing later times.

boundary. As shown in Fig. 4-8a, the effective wandering exponent $\zeta$ is suppressed from the usual value of 2/3, to $\zeta \approx 1/3$ for times $vt \lesssim w_{\text{gap}}$, where $v$ is the average front velocity. At later times, as the kink in the front heals and the average front normals return to the vertical, $\zeta$ recovers the expected value of 2/3 for KPZ genetic boundaries. Notably, the effective $\zeta$ appears to exceed 2/3 in an intermediate transitory regime when $vt \approx w_{\text{gap}}$.

To gain further insight into this changing wandering exponent, we modify the numerical experiment to a wedge geometry (Fig. 4-7b). This allows us to fix the kink angle $\theta$ to be a constant value, as opposed to the gap geometry where the kink heals from some initial $\theta_0$ toward $\pi$ with increasing time.

Now, the stepping stone model with deme size of 1 is, in essence, identical to the Eden model on a triangular lattice, with the added complication of tracking different genotypes. The boundary between two Eden clusters meeting at an angle $\theta$ has previously been studied, [109]. The transverse fluctuations scale as $t^{\zeta}$, where $t$ is the simulation time, and the wandering exponent $\zeta$ was conjectured to be

$$\zeta(\theta) = \begin{cases} 1/3, & \theta < \pi, \\ 2/3, & \theta = \pi, \\ 1, & \theta > \pi. \end{cases} \tag{4.8}$$

The value $\theta = \pi$ corresponds to two Eden clusters growing side by side with flat initial conditions, in which case one recovers the KPZ value of $\zeta = 2/3$ as expected.

The regime $\theta < \pi$ is of relevance to range expansions with obstacles. Heuristically, the sector boundary becomes pinned by the two Eden clusters growing into each other, and the usual KPZ transverse fluctuations are suppressed. Instead, the fluctuations which dominate are those of the propagating fronts themselves, which scale with the KPZ growth exponent $\beta = 1/3$ rather than the wandering exponent $\zeta = 2/3$.

The original simulations which led to the estimates in Eq. 4.8 sampled only 3 points in the range $\theta < \pi$, namely $\theta = \pi/3$, $\pi/2$, and $2\pi/3$ [109]. We expand on this previous work by fitting to an an effective $\zeta(\theta)$ for many more values of $\theta$. The results plotted in Fig. 4-8b indicate a smooth crossover between $\zeta = 1/3$ and $\zeta = 2/3$ as $\theta$ increases from 0 to $\pi$. A heuristic explanation for this change in $\zeta$ is given in Appendix F. The results from the wedge geometry are qualitatively consistent with the $\zeta$ values measured from the gap geometry. As the range expansion propagates around an obstacle, the fronts from either side meet at some angle $\theta_0 < \pi$, which can be predicted by a deterministic model of constant speed propagation for wavefronts in the same geometry, inspired by geometrical optics [96]. The incident angle increases up to $\theta = \pi$ as the kink in the front heals. Therefore, for the sector boundary formed after the obstacle, we expect the wandering exponent to initially take some value $\zeta < 2/3$, and then slowly recover to $\zeta = 2/3$. The kink has healed when the fluctuations of the front (perpendicular to the direction of propagation) are

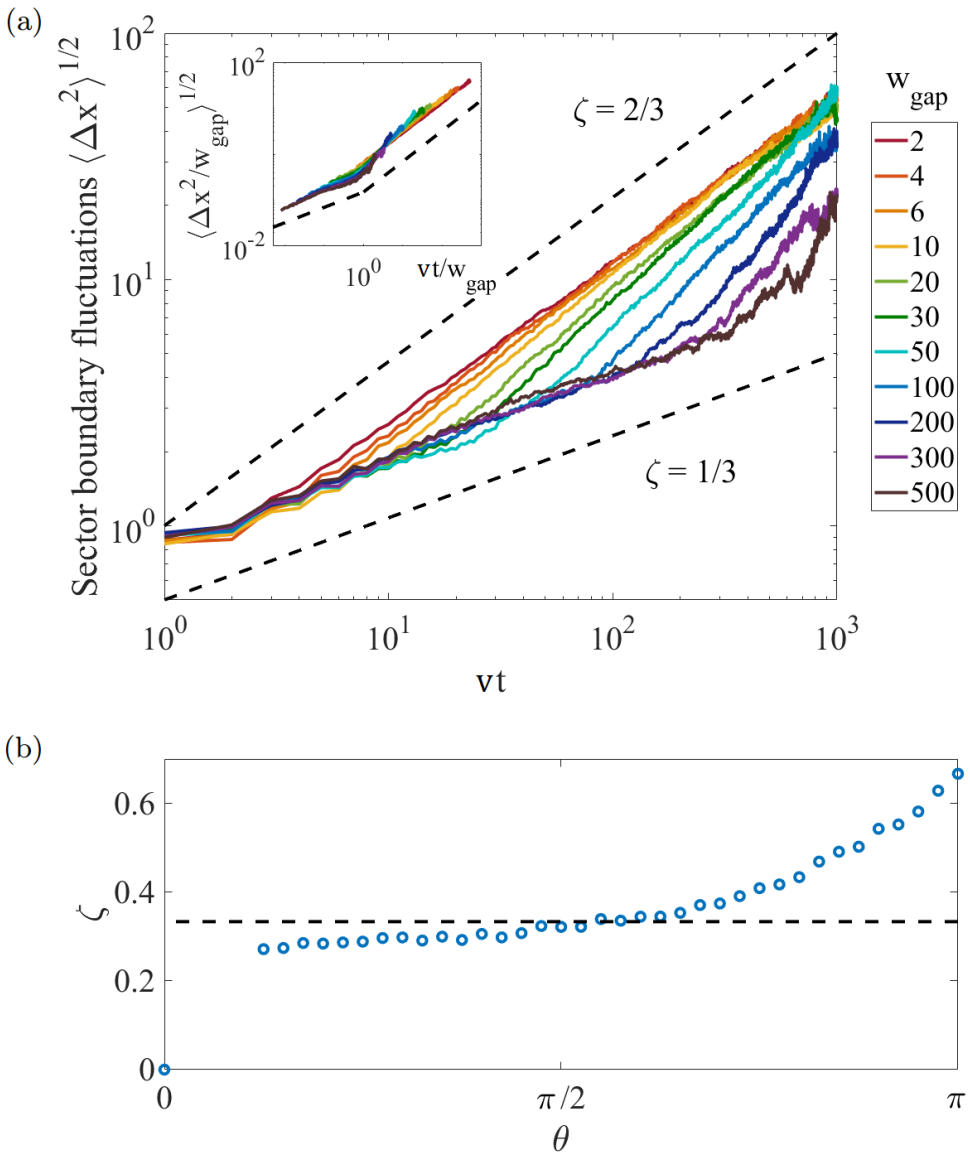


Figure 4-8: (a) Log-log plot of fluctuations of the sector boundary $\langle \Delta x^2 \rangle^{1/2}$ vs. vertical distance along the sector boundary $vt$ in the gap geometry for a range of gap sizes $w_{\text{gap}}$. Fits to a power law scaling form $\langle \Delta x^2 \rangle^{1/2} \sim t^{\zeta}$ yield exponents varying from $\zeta \approx 1/3$ to $\zeta \approx 2/3$, with a crossover region in between. Inset: Data collapse after rescaling with respect to $w_{\text{gap}}$. By geometrical arguments, $vt/w_{\text{gap}}$, where $v$ is the average front speed, is a measure of the angle of incidence of the fronts as determined by a constant speed or "geometrical optics" model. We see a reasonably good collapse across many different gap sizes, with $\zeta \approx 1/3$ for $vt/w_{\text{gap}} < 1$, and $\zeta \approx 2/3$ for $vt/w_{\text{gap}} > 1$. (b) Wandering exponent $\zeta$ as a function of the angle of incidence $\theta$ in the wedge geometry. As $\theta$ increases from 0 to $\pi$, the wandering exponent increases smoothly from approximately $\zeta = 1/3$ (marked by the dashed line) to the KPZ value of $\zeta = 2/3$.

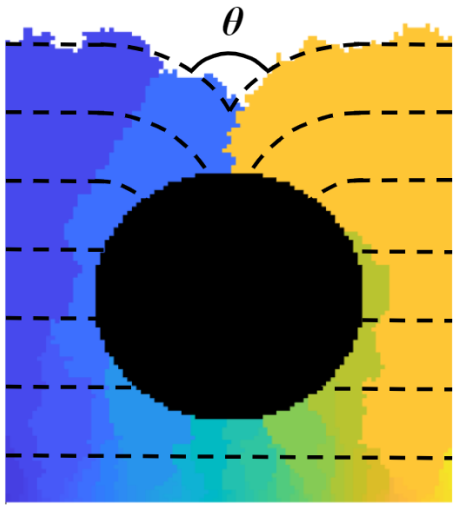


Figure 4-9: Illustration of a range expansion propagating around a spatial obstacle, depicted by the black circle, which prohibits growth. The colours represent allele labels, and the dashed lines mark the average front shape at equal time intervals, as predicted by geometrical optics. The growth fronts around either side of the obstacle meet at some angle $\theta < \pi$, forming a kink which heals over time.

comparable to the size of the dip. See Fig. 4-9 for an illustration.

Our results explain the suppressed fluctuations of genetic sector boundaries behind an obstacle observed in experimental work, and connect them with prior numerical work on Eden model growth. The effect of obstacles can be viewed as a competition between the usual roughening of the front, which favours the KPZ wandering exponent $\zeta = 2/3$, and the collision of two segments of the front propagating around either side of the obstacle, which suppresses $\zeta$ toward the value of the front roughness exponent $\beta = 1/3$.

# Chapter 5

# Eigenvalue spectrum of the product of transfer matrices for directed polymers

The work in this chapter is part of a collaboration with Dr. Ramis Movassagh.

In Chapter 1, we saw that the KPZ universality class is characterized by fluctuations which obey TW statistics. In particular, this result has been derived analytically for the free energy in the DPRM model using a replica Bethe ansatz method [15–17]. We discussed also that the TW probability distributions originate from the mathematical study of Gaussian random matrices, and more specifically, from the study of their eigenvalue spectra. In this chapter, we turn to a mathematical analysis of the transfer matrix formulation of DPRM in order to explore this connection on the level of matrices. Similar questions have been considered in the context of disordered elastic networks, where each transfer matrix contains information about the propagation of the displacement field [110]. Our interest in the eigenvalue spectrum of the product of DPRM transfer matrices is physically motivated by systems of non-intersecting paths, a problem which has been studied extensively in the context of the commensurate-incommensurate transition [4, 111].

To simplify the analysis, we construct a set of transfer matrices in such a way that ensures all eigenvalues of the product matrix are real and positive. We compute the

fluctuations in the spectrum, and find similarities in distribution to Gaussian random matrices for all eigenvalues. The spacing between eigenvalues is also relevant for a finite density of non-intersecting DPRMs, as it determines the cost associated with adding more such directed polymers in a grand canonical setting. Compared to the pure system, the presence of disorder changes the scaling of the spacing near the largest eigenvalue from $1/N^2$ to $1/N$, where $N$ is the system size, thus changing the density of states.

## 5.1 Connection to random matrix theory

We first examine more closely the relationship between the DPRM free energy and eigenvalues of Gaussian random matrices. For a DPRM system of size $N$, let $\tilde{Z}(x_0, x, t)$ denote the partition function of a directed polymer originating from $(x_0, 0)$ and terminating at $(x, t)$, with $x_0, x \in [1, \ldots, N]$ and $t > 0$. Using the transfer matrix formulation, $\tilde{Z}(x_0, x, t)$ can be written recursively in terms of the partition functions at time $t - 1$,

$$\tilde{Z}(x_0, x, t) = \sum_{x'} \langle x|\mathbf{T}(t)|x'\rangle \tilde{Z}(x_0, x', t-1). \tag{5.1}$$

If we define the product of transfer matrices,

$$\mathbf{W}(t) = \prod_{t'=1}^{t} \mathbf{T}(t'). \tag{5.2}$$

we can rewrite the partition function as

$$\tilde{Z}(x_0, x, t) = \langle x|\mathbf{W}(t)|x_0\rangle. \tag{5.3}$$

We consider the ensemble of directed polymers whose endpoints are fixed to be at the same spatial position $x$. Summing over all such paths then gives the partition function

$$Z(t) = \sum_{x=1}^{N} \tilde{Z}(x, x, t) = \mathrm{tr}\mathbf{W}(t). \tag{5.4}$$

The free energy is

$$f = -\frac{\ln Z}{t} = -\frac{\ln[\mathrm{tr}\mathbf{W}(t)]}{t} \simeq -\frac{\ln \lambda_1(t)}{t}, \tag{5.5}$$

where $\lambda_1(t)$ is the largest eigenvalue of the product matrix $\mathbf{W}(t)$ which dominates the trace. This motivates the definition of the quantity of interest,

$$\epsilon_i(t) = \frac{\ln \lambda_i(t)}{t}, \tag{5.6}$$

where $\lambda_i(t)$ is the $i$th largest eigenvalue of the product matrix $\mathbf{W}(t)$.

It has been well-established both indirectly through the Cole-Hopf transformation, [33] and directly through replica Bethe ansatz solutions, [15–17] that the DPRM free energy obeys TW statistics, with the details dependent on the geometry. In the case of Eq. 5.4, however, the geometry is not strictly pt-pt or pt-line (defined in Chapter 1.2.1). Rather, $Z(t)$ is the sum of an ensemble of pt-pt paths. Thus we expect the limiting distribution to be very similar to TW, but perhaps not precisely the GOE or GUE form. In other words, we can write

$$\epsilon_1(t) = c_0 + c_1 t^{-2/3}\xi, \tag{5.7}$$

where $\xi$ is an $\mathcal{O}(1)$ random variable whose distribution is TW-like in the limit of large $t$, and $c_0, c_1$ are system-specific constants. It is important to note that the regime relevant to DPRM requires $1 \ll t \ll N^{3/2}$, where $N$ is the system size. This scaling constraint stems from the dynamic exponent which governs the KPZ universality class, and ensures that the scalings are not affected by finite size.

In comparison, we consider an $n \times n$ GOE matrix $\mathbf{L}$ with i.i.d. elements $L_{ii} \sim \mathcal{N}(0, 2/n)$, $L_{ij} = L_{ji} \sim \mathcal{N}(0, 1/n)$ [or respectively, GUE matrix $\mathbf{M}$ with i.i.d. elements $M_{ii} \sim \mathcal{N}(0, 1/n)$, $M_{ij} = \overline{M_{ji}} \sim \mathcal{N}(0, 1/2n) + i\mathcal{N}(0, 1/2n)$]. The largest eigenvalue then has the following scaling form [10, 11],

$$\lambda_1^{\mathrm{TW}}(n) = 2 + n^{-2/3}\xi, \tag{5.8}$$

where $\xi$ is a TW-GOE (TW-GUE) random variable in the limit of large $n$.

The similarities between $\epsilon_1(t)$ in Eq. 5.7 and $\lambda_1^{\rm TW}(n)$ in Eq. 5.8 are immediately evident. The time $t$ [also number of transfer matrices in the product matrix $\mathbf{W}(t)$] in the DPRM context appears to play an analogous role to the matrix size $n$ in the GOE and GUE context. We compare numerics for the DPRM product matrix with known results from random matrix theory in Chapter 5.4, and explore whether the connection extends beyond the scaling forms of the largest eigenvalue.

## 5.2 Non-intersecting paths

We motivate the study of the quantities $\epsilon_i$, related to the eigenvalue spectrum of the DPRM product matrix, by examining their role in a system of non-intersecting DPRMs [4]. Physical examples include magnetic domain walls in Ising models [112, 113], and pinned flux lines in superconductors [114]. In such systems, if it is favourable to create one domain wall or flux line, it is natural to ask why an infinite number is not created. As we explain below, this is a consequence of the non-crossing condition. The more general problem of non-intersecting paths is also found in the adsorption of an atomic monolayer on a crystalline surface [111, 115–117], and the equilibrium shapes of crystals [118, 119]. The statistical behaviour is exemplified by the commensurate-incommensurate transition, a topic which has sparked much theoretical interest [111, 120–122].

In the pure case (without disorder), the grand canonical free energy $\mathcal{F}$ is obtained by minimizing the following expression over the density $r$ of non-intersecting paths, [111]

$$\frac{\mathcal{F}(r)}{Nt} = f_1 r + b r^3. \tag{5.9}$$

Using the language of domain walls, $N$ and $t$ are the system dimensions, $f_1$ is the free energy of a single wall, and $b > 0$ is a constant. The first term, proportional to the density, is intuitive. The second term represents an effective repulsion due to the non-crossing restriction. Performing the minimization over $r$, we see that for $f_1 > 0$, no domain walls are formed ($r = 0$). On the other hand, for $f_1 < 0$, a finite density

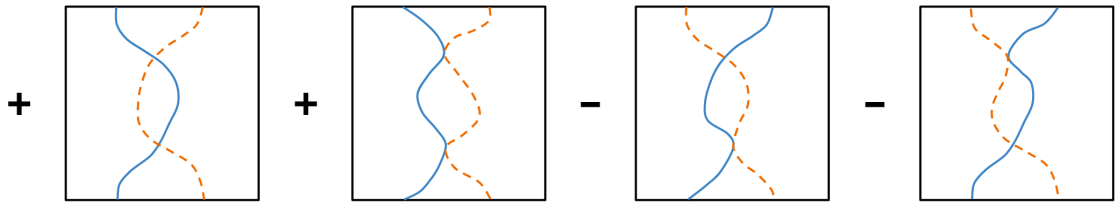


Figure 5-1: Illustration of configurations of two DPRMs (solid and dashed lines) which do not satisfy the non-crossing condition. Introducing the factor $(-1)^{\#\text{ of crossings}}$ leads all such terms to cancel.

($r > 0$) of domain walls are added in such a way that the free energy gain is balanced by the entropy loss from imposing the non-crossing condition. These walls can also be interpreted as world-lines of 1D fermions, an approach taken by Pokrovsky and Talapov in Ref. [111].

If we now consider the presence of quenched impurities, the domain walls can be represented by a system of non-intersecting DPRMs. Using the replica Bethe ansatz, the quantity to be minimized over in Eq. 5.9 becomes [4]

$$\frac{\langle \mathcal{F}(r) \rangle}{Nt} = \langle f_1 \rangle r + br^2, \tag{5.10}$$

where the angular brackets denote averaging over the disorder, and again, $b > 0$. The effective repulsion term in this case is proportional to $r^2$ rather than $r^3$. For the pure system, this repulsion relates to the typical distance between "collisions" of domain walls due to transverse wanderings, whereas for the disordered system, it relates to effects of confinement on a finite density of optimal paths.

In writing down the grand canonical description of DPRM, the key difficulty lies in the implementation of the non-crossing condition. We use the same trick which proved powerful in the Ising problem, of removing the non-crossing condition, and introducing a factor of $(-1)^{\#\text{ of crossings}}$ for each term in the partition function. The terms describing intersecting paths or shared bonds then cancel, as shown in Fig. 5-1, leaving only contributions from non-crossing configurations.

The grand canonical partition function $\mathcal{Z}$ can then be written in terms of the

canonical partition functions $Z_n$ for $n$ DPRMs.

$$\begin{aligned}\mathcal{Z} &= \exp\left\{\sum_{n=1}^{N}\frac{(-1)^{n+1}}{n}Z_n\right\}\\ &= \exp\left\{\sum_{n=1}^{N}\frac{(-1)^{n+1}}{n}\mathrm{tr}\mathbf{W}^n\right\}\\ &= \exp\left\{\sum_{i=1}^{N}[\ln(1+\lambda_i)]\right\}\\ &= \exp\left\{\sum_{i=1}^{N}\left[\ln\left(1+e^{\epsilon_i t}\right)\right]\right\}. \end{aligned} \tag{5.11}$$

(One could introduce a chemical potential $\nu$ such that $Z_n \to e^{\nu t n} Z_n$ counts the number of added steps. However, without loss of generality, $\nu$ can be absorbed into the parametrization of the energies.) In the limit of large $t$, only terms with $\epsilon_i > 0$ in Eq. 5.11 will contribute to the free energy,

$$\mathcal{F} = -\frac{\ln \mathcal{Z}}{t} \xrightarrow{t\to\infty} -\sum_{\epsilon_i>0}\epsilon_i. \tag{5.12}$$

The condition on $\epsilon_i$ reflects a constraint on the strength of the average disorder compared to the hopping energy, in order for it to be energetically favourable to create more directed polymers. The value of $\epsilon_1$ determines whether a single DPRM is favourable; after that, the difference $\epsilon_i - \epsilon_{i+1}$ becomes relevant for adding subsequent DPRMs. We can interpret this as filling levels $-\epsilon_i$ in an energy band, starting from $-\epsilon_1$, up to the Fermi energy. The resistance to adding more non-intersecting DPRMs is therefore related to the density of states near the edge of the spectrum.

## 5.3 Definition of the transfer matrix

We define the DPRM transfer matrix at time $t$ as

$$\mathbf{T}(t) = \begin{pmatrix} \eta_1(t) & \gamma & 0 & \cdots & 0 \\ \gamma & \eta_2(t) & \gamma & \cdots & 0 \\ 0 & \gamma & \eta_3(t) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \eta_N(t) \end{pmatrix}, \tag{5.13}$$

where $\eta_i(t) = \exp[-\varepsilon_i(t)]$ are i.i.d. random variables on the main diagonal, with constant elements $\gamma > 0$ on the off-diagonals. We choose $\varepsilon_i(t) \in \mathcal{U}(\mu, \sigma^2)$, uniformly distributed with mean $\mu$ and variance $\sigma^2$. For mathematical reasons explained in Chapter 5.5, the parameters are constrained by the relation

$$\mu + \sqrt{3}\sigma < -\ln(2\gamma), \tag{5.14}$$

which ensures $\eta_i(t) > 2\gamma$, $\forall i$ and $\forall t$.

The above transfer matrix describes DPRM on a square lattice with closed boundary conditions. At each time $t$, the path at position $i$ may propagate forward, picking up a random energy $\varepsilon_i(t)$, or deviate to the left or right, picking up a hopping energy $K = -\ln\gamma$ (see Fig. 5-2). This differs slightly from the geometry introduced in previous chapters, but crucially, does not affect the universality class of the model.

The details of the model are chosen to ensure that all eigenvalues $\lambda_i$ of the product matrix $\mathbf{W}$ are real and positive (see Chapter 5.5 for a detailed proof). This allows us to study the objects of interest, $\epsilon_i = \ln\lambda_i/t$, which would otherwise be ill-defined. Note that the condition $\lambda_i \in \mathbb{R}^+$ is not a trivial one. Although an individual transfer matrix $\mathbf{T}$ is real, positive, and symmetric, there is a different realization of randomness for each time $t$. Thus the product matrix $\mathbf{W}$ is in general, real and positive, but not symmetric. The Perron-Fröbenius theorem guarantees that the largest eigenvalue is unique and real (positive), but *a priori*, all other eigenvalues need not be real. Indeed,

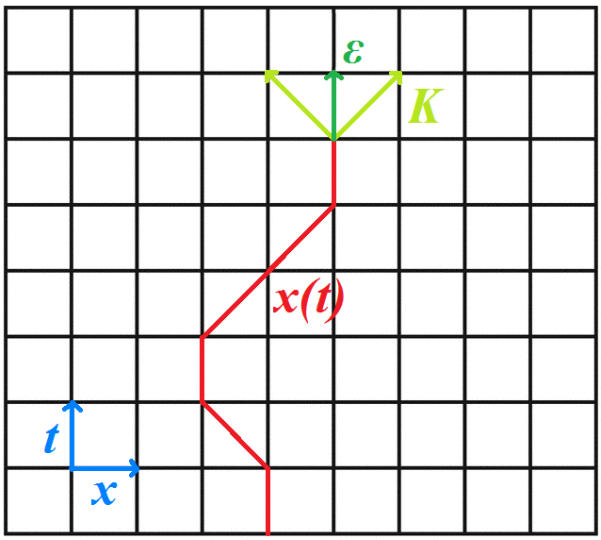


Figure 5-2: Illustration of the DPRM geometry described by the transfer matrix in Eq. 5.13. For a directed polymer $x(t)$, propagating forwards is associated with a random energy $\varepsilon$, while deviating to the left or right is associated with a hopping energy $K$.

for other geometries of DPRM, the spectrum is in general composed of many complex conjugate pairs of eigenvalues, the physical significance of which is unclear.

## 5.4 Statistics of the eigenvalue spectrum

Armed with the above definition, we analyze the complete spectrum of the product of transfer matrices. We compute numerically the eigenvalues $\lambda_i$ of $\mathbf{W}$, with parameters $t = 32$ and $N = 32$, over $2^{25}$ realizations (the scaling of fluctuations in Chapter 5.4.2 is computed over $2^{22}$ realizations instead). We choose $\mu = -2$ and $\sigma^2 = 1/12$ for the random energies, and $\gamma = 1$ for the hopping energies. The mean values of $\epsilon_i = \ln \lambda_i / t$ are plotted in Fig. 5-3, with errorbars indicating the respective standard deviations.

### 5.4.1 Probability distributions

We focus on the distributions of $\epsilon_i$ after appropriate rescaling of the the mean and variance. From Eq. 5.7, we expect $\epsilon_1$, corresponding to the largest eigenvalue of $\mathbf{W}$, as well as the DPRM free energy, to obey TW-like statistics. This is indeed confirmed in

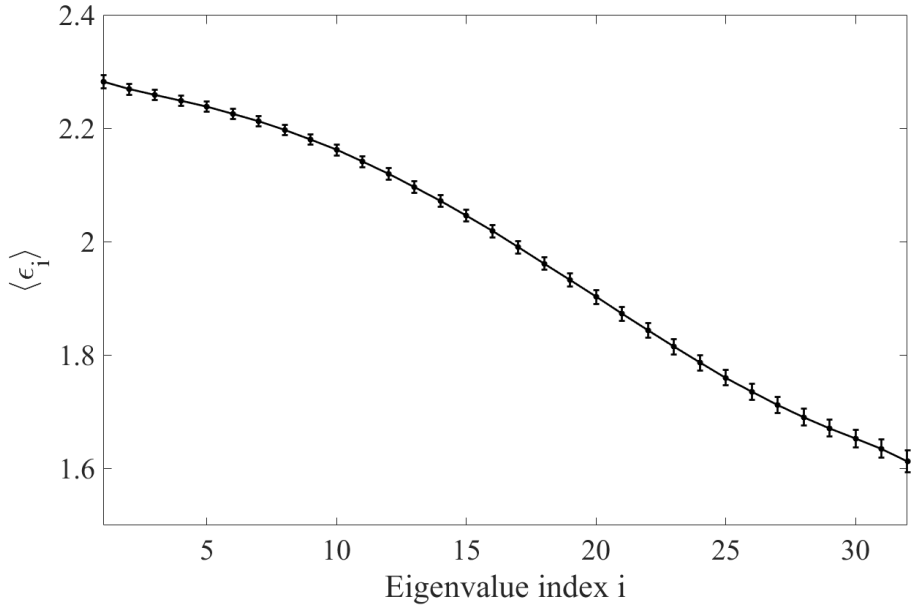


Figure 5-3: Average eigenvalue spectrum for the product of DPRM transfer matrices $\mathbf{W}$ with parameters $t = 32$ and $N = 32$. The data points indicate the mean of $\epsilon_i = \ln \lambda_i / t$, where $\lambda_i$ is the $i$th largest eigenvalue of $\mathbf{W}$, and the errorbars indicate the standard deviation.

Fig. 5-4a. Rather surprisingly however, we find that $\epsilon_N$, corresponding to the smallest eigenvalue of $\mathbf{W}$, also has the same TW-like distribution. This is especially remarkable given the discrepancy in the variances (see Fig. 5-3). Nevertheless, it is reminiscent of the symmetry known to exist in Gaussian random matrices, between the pairs of eigenvalues $\lambda_i^{\mathrm{TW}}$ and $\lambda_{N+1-i}^{\mathrm{TW}}$. There, the behaviour is dictated by the symmetry of the Wigner semicircle distribution $[\rho(\lambda) = \sqrt{4-\lambda^2}/2\pi]$ for the overall spectrum [123]. The study of non-intersecting DPRMs in Chapter 5.2 provides some physical intuition for the correspondence observed between pairs $\epsilon_i$ and $\epsilon_{N+1-i}$. Since the quantity $\epsilon_i$ is related to creating $i$ directed paths, it could, conversely, be interpreted as creating $N - i + 1$ empty "paths".

Numerically, this pairwise correspondence persists beyond the extremal eigenvalues for the DPRM product matrix as well. In fact, for any $i$, the pair $\epsilon_i$ and $\epsilon_{N+1-i}$ shares the same distribution not only with each other, but also with the analogous eigenvalue pair for Gaussian random matrices. For instance, in Fig. 5-4b, we see that $\epsilon_{N/2}$ and $\epsilon_{N/2+1}$ are Gaussian in distribution, mirroring the expectation for bulk

eigenvalues in GOE and GUE matrices [124, 125]. The distribution of the remaining eigenvalues interpolate between TW near the edge of the spectrum, and Gaussian in the bulk.

### 5.4.2 Scaling of fluctuations

We compute also the scaling exponents for fluctuations of individual eigenvalues. For Gaussian random matrices, the following rigidity estimate exists [126, 127],

$$\mathrm{Var}[\lambda_i^{\mathrm{TW}}(n)] \sim n^{-4/3}[\min\{i, n+1-i\}]^{-2/3}. \tag{5.15}$$

For the extremal eigenvalues $\lambda_1^{\mathrm{TW}}$ and $\lambda_n^{\mathrm{TW}}$, the scaling of fluctuations reduces to $n^{-2/3}$, consistent with Eq. 5.8; for the bulk eigenvalues, however, this scaling is $n^{-1}$. The term rigidity refers to an effective repulsion between consecutive eigenvalues, specifically in comparison to i.i.d. random variables drawn from the same Wigner semicircle distribution. In the latter case, order statistics yield typical fluctuations of order $n^{-1/2}$ instead [127]. We plot the analogous scaling exponents for the DPRM product matrix in Fig. 5-5. For $\epsilon_1$ related to the DPRM free energy, we find $[\mathrm{Var}(\epsilon_1)]^{1/2} \sim t^{-2/3}$, as expected from Eq. 5.7. However, for $\epsilon_i$ in the bulk, the scaling is approximately $t^{-1/2}$, consistent with ordered i.i.d random variables rather than GOE or GUE eigenvalues. The correspondence between pairs of eigenvalues observed earlier in the probability distributions is also present in the scaling of fluctuations, although to a lesser extent.

### 5.4.3 Density of states

We now compare the eigenvalue spacings for the DPRM product matrix in disordered and pure systems, focusing our attention on the density of states near $\epsilon_1$. In the pure system, the transfer matrix is time-independent, and the eigenvalues are fixed. We

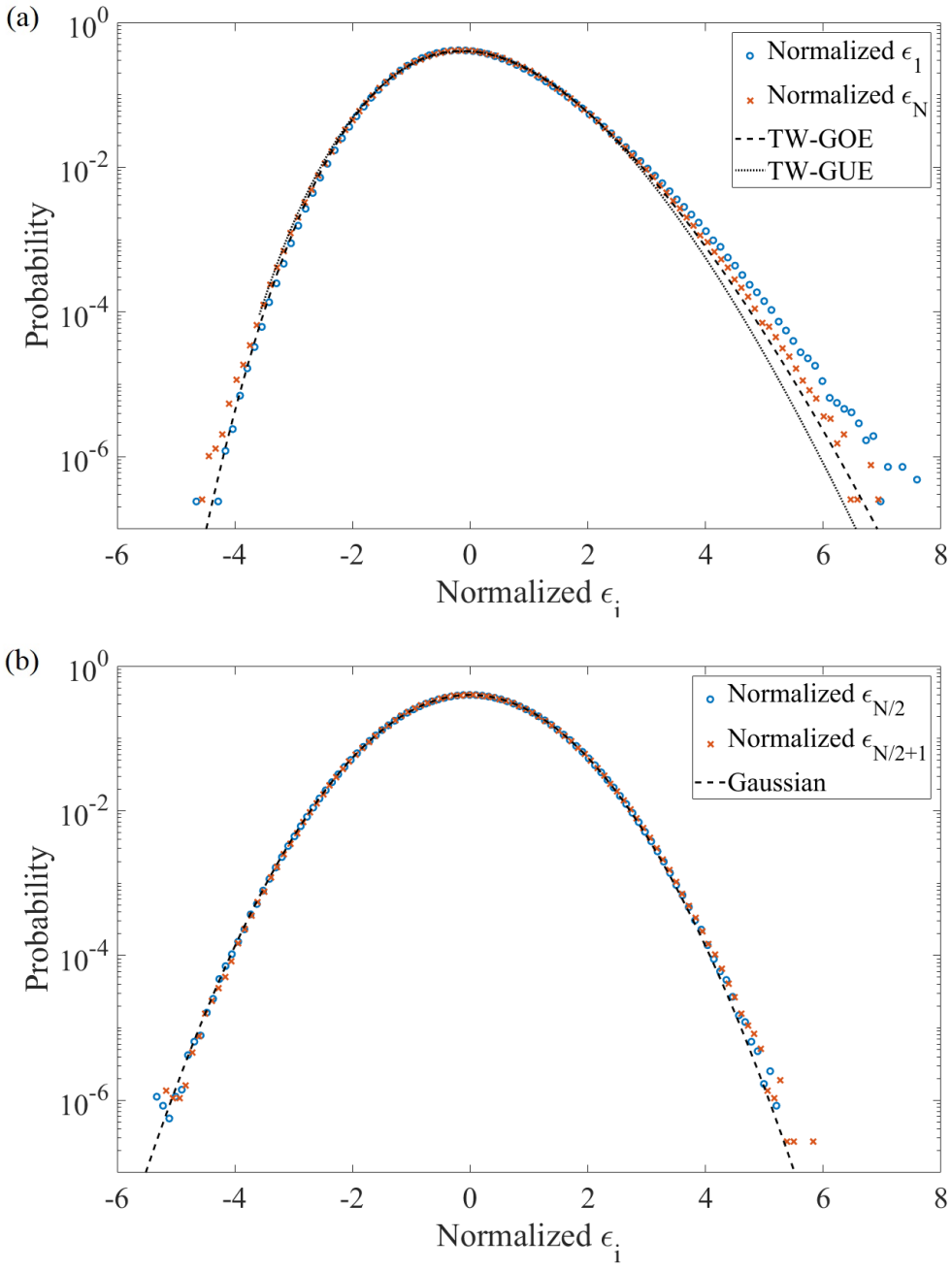


Figure 5-4: Probability distribution of $\epsilon_i = \ln \lambda_i / t$ for the product of DPRM transfer matrices. All distributions are normalized to have mean 0 and variance 1. (a) On the edges of the spectrum, $\epsilon_1$ and $\epsilon_N$ have asymmetric distributions consistent with the TW forms. (b) In contrast, $\epsilon_i$ in the bulk are Gaussian distributed, similar to the bulk eigenvalues of GOE and GUE matrices.

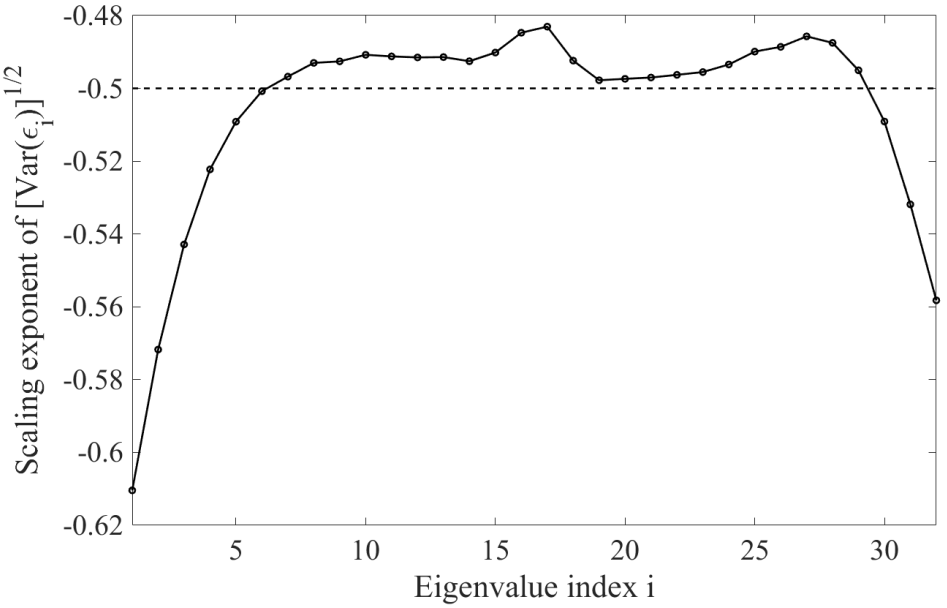


Figure 5-5: Scaling exponents for the variance of $\epsilon_i = \ln \lambda_i / t$ for the product of DPRM transfer matrices. For $\epsilon_1$, the scaling of fluctuations $[\mathrm{Var}(\epsilon_i)]^{1/2}$ is consistent with $t^{-2/3}$, expected for the DPRM free energy. For bulk $\epsilon_i$, however, the scaling is approximately $t^{-1/2}$, similar to that of i.i.d. random variables drawn from a Wigner semicircle distribution.

take $\mathbf{T}_{\mathrm{pure}}$ to be of the following form,

$$\mathbf{T}_{\mathrm{pure}} = \begin{pmatrix} \bar{\eta} & \gamma & 0 & \cdots & \gamma \\ \gamma & \bar{\eta} & \gamma & \cdots & 0 \\ 0 & \gamma & \bar{\eta} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma & 0 & 0 & \cdots & \bar{\eta} \end{pmatrix}, \tag{5.16}$$

where $\bar{\eta}$ is the average strength of disorder $\eta_i(t)$ defined in Eq. 5.13 and 5.14. We have introduced additional hopping terms $\gamma$ in the $(1, N)$ and $(N, 1)$ positions, which correspond to periodic boundary conditions. The effect on the spectrum is negligible for large system sizes, and the advantage of this choice is that Eq. 5.16 is a circulant matrix whose spectrum is known analytically,

$$\epsilon_k^{\mathrm{pure}} = \frac{\ln \lambda_k^{\mathrm{pure}}}{t} = \bar{\eta} + 2\gamma \cos\left[\frac{2\pi k}{N}\right], \qquad k = 0, 1, \ldots, N-1. \tag{5.17}$$

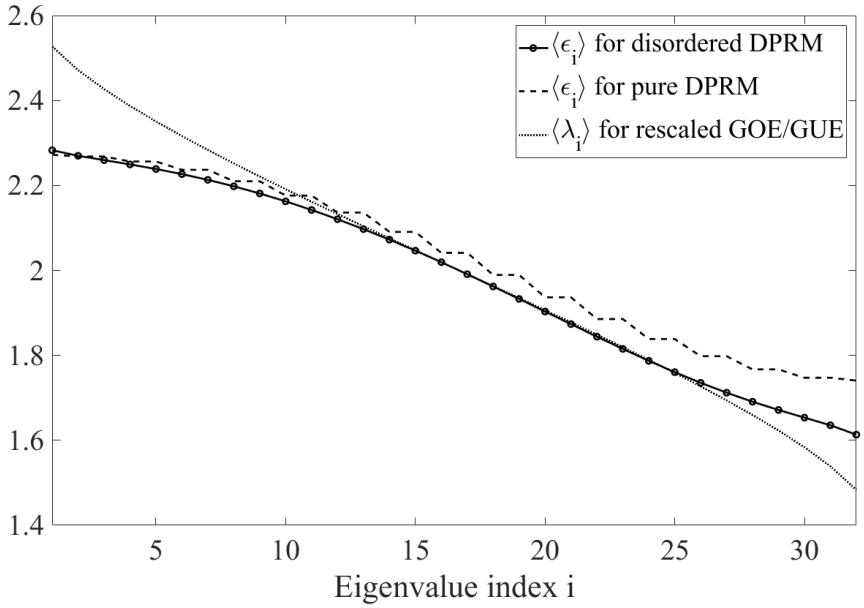


Figure 5-6: $\epsilon_i$ for disordered DPRM (solid line), plotted against the analogous quantity for the pure system (dashed line), and the rescaled mean eigenvalues of Gaussian random matrices (dotted line). The "steps" in the dashed line arise from degeneracy. Near $i = 1$, the curve is linear for disordered DPRM, in contrast to quadratic for pure DPRM.

[Note that the eigenvalues in the above form are not ordered, and that the bulk eigenvalues (i.e. not the maximum or minimum) are degenerate.]

We plot $\epsilon_i$ for the disordered and pure DPRM systems in Fig. 5-6. We see that in the disordered case, the "energy band" is linear near $\epsilon_1$, rather than quadratic. It is therefore more energetically costly to add non-intersecting directed polymers into the system. We plot also the mean eigenvalues for rescaled Gaussian random matrices for reference. (The curves for GOE and GUE matrices are indistinguishable, so only one is plotted.) There, the density of states follows the Wigner semicircle distribution, which vanishes continuously at the edge of the spectrum as $n \to \infty$.

## 5.5 Totally positive matrices

We devote this section to proving the claim that the product of DPRM transfer matrices, as defined in Eq. 5.2 and 5.13, has eigenvalues which are all real and positive.

To see this, we turn to a class of matrices known as totally positive matrices [128,129].

We begin with some definitions. An $N \times N$ matrix $\mathbf{A} = \{a_{ij}\}_{i,j=1}^{N}$ is *totally positive* if the determinant of any square submatrix (obtained by omitting $N - k$ rows and columns) is positive. More precisely, let $\{i_l\}_{l=1}^{k}$ and $\{j_m\}_{m=1}^{k}$ be increasing subsequences of $\{1, \ldots, N\}$, with length $k < N$. Then $\mathbf{A}' = \{a_{i_l, j_m}\}_{l,m=1}^{k}$ is a square submatrix of $\mathbf{A}$. As a special case, a *minor* $M_{ij}$ of $\mathbf{A}$ is the determinant of the submatrix obtained by removing the $i$th row and the $j$th column. Furthermore, if $i = j$, $M_{ij}$ is called a *principal minor*. A consequence of this definition is that a totally positive matrix $\mathbf{A}$ is necessarily also *positive* (has all positive entries) and *positive-definite* (has all positive eigenvalues).

The key property we exploit is that total positivity is preserved under matrix multiplication. We begin by proving this closure property. We then verify that each individual transfer matrix $\mathbf{T}$ defined in Eq. 5.13 is totally positive. These results combine to show that the product matrix $\mathbf{W}$ is totally positive, with real and positive eigenvalues.

### 5.5.1 Closure under matrix multiplication

Consider two $N \times N$ totally positive matrices $\mathbf{A}$ and $\mathbf{B}$. By the Cauchy-Binet theorem, the determinant of a $k \times k$ submatrix of the product $\mathbf{AB}$ can be written as

$$\det[(\mathbf{AB})_{\mathcal{IJ}}] = \sum_{\mathcal{H}} \det(\mathbf{A}_{\mathcal{IH}}) \det(\mathbf{B}_{\mathcal{HJ}}) \tag{5.18}$$

where $\mathcal{I}$, $\mathcal{J}$, and $\mathcal{H}$ are increasing subsequences of $\{1, \ldots, N\}$, with length $k$. The sum is over all possible such subsequences $\mathcal{H}$. Since $\mathbf{A}$ and $\mathbf{B}$ are totally positive, $\det(\mathbf{A}_{\mathcal{IH}}) > 0$ and $\det(\mathbf{B}_{\mathcal{HJ}}) > 0$ for any $\mathcal{H}$. This immediately gives $\det[(\mathbf{AB})_{\mathcal{IJ}}] > 0$ for any minor of the product $\mathbf{AB}$. Thus $\mathbf{AB}$ must also be totally positive. The argument trivially generalizes to the product of $n > 2$ matrices. Therefore, totally positive matrices are closed under matrix multiplication.

### 5.5.2 Total positivity of the transfer matrix

It remains to show that the DPRM transfer matrix introduced in Eq. 5.13 is indeed totally positive. We rewrite Eq. 5.13 as

$$\mathbf{T}(t) = \mathbf{T}_0 + \mathbf{E}(t),$$

$$\mathbf{T}_0 = \gamma \begin{pmatrix} 2 & 1 & 0 & \cdots & 0 \\ 1 & 2 & 1 & \cdots & 0 \\ 0 & 1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{pmatrix},$$

$$\mathbf{E}(t) = \begin{pmatrix} \eta_1(t) - 2\gamma & 0 & 0 & \cdots & 0 \\ 0 & \eta_2(t) - 2\gamma & 0 & \cdots & 0 \\ 0 & 0 & \eta_3(t) - 2\gamma & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \eta_N(t) - 2\gamma \end{pmatrix}. \tag{5.19}$$

$\mathbf{T}_0$ is now a time-independent Jacobi (tri-diagonal) matrix, while $\mathbf{E}(t)$ is a positive diagonal matrix due to the constraints we placed on the noise $\eta_i(t)$ in Eq. 5.14. To proceed, we make use of Theorem 2.3 and Corollary 2.4 from Ref. [129].

**Theorem.** (Ando, 1987) Let $\mathbf{A}$ be an $N$-square Jacobi matrix. If $\mathbf{A}$ is positive, and all principal minors are positive, then $\mathbf{A}$ is totally positive. Furthermore, for any $s_i > 0, i = 1, 2, \ldots, N$,

$$\det[\mathbf{A} + \mathrm{diag}(s_1, \ldots, s_N)] \geq \det \mathbf{A} + \sum_i s_i, \tag{5.20}$$

and it follows that $\mathbf{A} + \mathrm{diag}(s_1, \ldots, s_N)$ is also totally positive.

If $\mathbf{T}_0$ is totally positive, then by the above theorem, $\mathbf{T}(t) = \mathbf{T}_0 + \mathbf{E}(t)$ is totally positive $\forall t$. It is trivial that $\mathbf{T}_0$ is positive. We need only show that all principal minors are positive as well. However, if we denote the Jacobi matrix $\mathbf{T}_0$ of size $N$ as

$\mathbf{T}_0^N$, the principal minor $M_{ii}$ of $\mathbf{T}_0^N$ can be written in terms of a block matrix,

$$M_{ii} = \det \begin{pmatrix} \mathbf{T}_0^{i-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_0^{N-i} \end{pmatrix} = \det(\mathbf{T}_0^{i-1}) \det(\mathbf{T}_0^{N-i}) \tag{5.21}$$

Thus the proof reduces to showing $\det(\mathbf{T}_0^k) > 0$, $\forall k < N$.

We write $\mathbf{T}_0^k$ as a block matrix in the following form,

$$\mathbf{T}_0^k/\gamma = \begin{pmatrix} 2 & \mathbf{C} \\ \mathbf{C}^T & \mathbf{T}_0^{k-1}/\gamma \end{pmatrix}, \qquad \mathbf{C} = \underbrace{(1, 0, \ldots, 0)}_{k-1 \text{ terms}}. \tag{5.22}$$

Then by the Schur determinant identity,

$$\begin{aligned} \det(\mathbf{T}_0^k/\gamma) &= \det(2) \det(\mathbf{T}_0^{k-1}/\gamma - \mathbf{C}^T 2^{-1} \mathbf{C}) \\ &= 2 \det \left[ \mathbf{T}_0^{k-1}/\gamma - \operatorname{diag}\left(\frac{1}{2}, 0, \ldots, 0\right) \right] \\ &= 2 \left(2 - \frac{1}{2}\right) \det \left[ \mathbf{T}_0^{k-2}/\gamma - \operatorname{diag}\left(\frac{1}{2 - \frac{1}{2}}, 0, \ldots, 0\right) \right] \\ &= 2 \left(2 - \frac{1}{2}\right) \cdots \underbrace{\left(2 - \cfrac{1}{2 - \cfrac{1}{2 - \ddots}}\right)}_{k \text{ levels}}. \end{aligned} \tag{5.23}$$

In the second line, we can again write the matrix in square brackets as a block matrix, as in Eq. 5.22. Recursively applying the determinant identity $k$ times gives the final expression in Eq. 5.23. It is not difficult to see that the limit of the continued fraction is

$$\lim_{m \to \infty} \underbrace{\left(2 - \cfrac{1}{2 - \cfrac{1}{2 - \ddots}}\right)}_{m \text{ levels}} = 1^+ \tag{5.24}$$

and more importantly, for any $m$, the continued fraction is positive. Thus $\det(\mathbf{T}_0^k) > 0$, $\forall k$, and the proof is complete.

# Chapter 6

# Outlook

We have considered in this thesis the statistics of paths and interfaces in the presence of stochastic noise in a random landscape, as captured by the KPZ universality class. We examined the scaling and distribution of fluctuations from a theoretical perspective, and, using the lattice models at our disposal, studied the consequences of these fluctuations for diverse experimental systems. Here, we summarize our work, and discuss possible future directions.

In Chapter 2, we showed that the probability distribution for the DPRM free energy interpolates between TW and Gaussian forms in the presence of spatially correlated noise. It is worthwhile to extend this study of distributions to energy landscapes with temporal [1, 82, 130] or isotropic [73] correlations, as these are often relevant in physical systems. One could also develop an analysis similar to the Krug-Meakin toolbox in the uncorrelated case [8] in order to characterize higher order moments of the distributions.

In Chapter 3, we effectively modeled optimal paths on the road network as DPRMs. In doing so, we replaced the complex road structure by a homogeneous noise featuring only the relevant properties to account for the observed statistics of optimal paths. A more realistic model may be constructed by introducing isotropic correlations in the noise, in place of correlations which are purely transverse to the direction of travel. It would be interesting as well to see if this approach can be extended to other transportation networks or different environments, for example, to the

study of shortest paths on a critical percolation cluster [131], which have important practical applications.

In Chapter 4, we connected the population genetics of propagating range expansions with new calculations in models from the KPZ universality class. Going forward, our calculations of the rate and expected time to coalescence for KPZ walkers in a totally uniform environment will be valuable as a standard against which deviations can be measured, to reveal the effects of various realistic complications. These complications include end effects from habitat boundaries [93, 106], selectively advantageous or deleterious mutations, mutualism or antagonism between subpopulations [132], geometrical inflationary effects in radial expansions [100], and more complex heterogeneities in the environment [96]. With minor modifications, the interface height interpretation of DPRM can be extended to model growth in a radial geometry. The transfer matrix formulation provides a significant numerical advantage over other off-lattice models, and, with an appropriate choice of discretization, would also allow us to circumvent difficulties related to lattice artifacts [100]. On the topic of environmental heterogeneities, we have made headway here by studying a simplified representation of an obstacle, which already illustrates the subtle issue of locally suppressed fluctuations. It will be interesting to extend this analysis of Eden model growth to situations with multiple obstacles, and with other types of heterogeneities such as nutrient "hotspots" [108] and uneven topography [133]. The dynamics can also be made more sophisticated by increasing the number of organisms per deme above $N = 1$, and reintroducing aspects of the original stepping stone model's migration dynamics between neighbouring demes [92]. Finally, our results have drawn upon two quite different processes in the KPZ universality class, the rough front stepping stone model and DPRM, to obtain quantitative insights about biological experiments that can be realized in the laboratory. We hope that this work will inspire future investigations to seek other useful links between disparate model systems that shed light on the evolutionary dynamics of rough front range expansions, a problem with much fertile territory.

In Chapter 5, we investigated the eigenvalue spectrum of the product of DPRM

transfer matrices in relation to a finite density of non-intersecting DPRMs. The presence of disorder was found to modify the density of states near the edge of the spectrum in an important way. In defining the transfer matrix, however, we imposed mathematical constraints on the boundary conditions and the relative strengths of random energies and hopping energies in order to ensure all eigenvalues of the product matrix are real and positive. The physical significance of this is unclear, as transfer matrices which do not satisfy these constraints still describe valid physical systems.

# Appendix A

# Algorithm for generating correlated noise

We present here the method used for generating correlated noise developed in Ref. [52]. Note that a similar method was developed in Ref. [53] to study DPRM and BD models with correlated noise.

In $d = 1$ dimension, the goal is to use a sequence of i.i.d. Gaussian random variables $\{u_i\}_{i=1,\ldots,L}$ to generate a sequence $\{\eta_i\}_{i=1,\ldots,L}$ with correlation function [1]

$$C(j) = \langle \eta_i \eta_{i+j} \rangle \sim j^{2\rho-1}, \qquad j \to \infty. \tag{A.1}$$

Taking the Fourier transform gives the spectral density function $S(q)$, which has the asymptotic form

$$S(q) = \langle \eta_q \eta_{-q} \rangle \sim q^{-2\rho}, \qquad q \to 0. \tag{A.2}$$

The correlated random variables can then be obtained from the relation

$$\eta_q = [S(q)]^{1/2} u_q, \tag{A.3}$$

where $\{\eta_q\}$ and $\{u_q\}$ are the Fourier transform coefficients of $\{\eta_i\}$ and $\{u_i\}$, respectively.

As in Ref. [52], we define the correlation function to be

$$C(j) \equiv (1 + j^2)^{\rho - 1/2}, \tag{A.4}$$

on the interval $j \in [-L/2, L/2]$. This then has the same asymptotic power law decay as in Eq. (A.1). The spectral density can be calculated analytically as the discrete Fourier transform,

$$S(q) = \frac{2\pi^{1/2}}{\Gamma(-\rho + 1)} \left(\frac{q}{2}\right)^{-\rho} K_\rho(q), \tag{A.5}$$

where $q = 2\pi n/L$ with $n = -L/2, \ldots, L/2$, and $K_\rho$ is the modified Bessel function of the second kind of order $\rho$. We define $S(q = 0) = 0$ to avoid any divergences. For $\rho > 0$, the modified Bessel function has asymptotic form

$$K_\rho(q) \sim \begin{cases} \dfrac{\Gamma(\rho)}{2} \left(\dfrac{2}{q}\right)^{\rho}, & q \ll 1, \\ \sqrt{\dfrac{\pi}{2q}} e^{-q}, & q \gg 1. \end{cases} \tag{A.6}$$

As $q \to 0$, we recover the asymptotic form in Eq. (A.2).

The process for generating the correlated noise used to study DPRM can then be summarized in the following steps.

1. Generate i.i.d. Gaussian random variables $\{u_i\}$, and calculate $\{u_q\}$ using the fast Fourier transform.

2. Calculate the spectral density function $S(q)$ using Eqs. (A.5) and (A.6).

3. Calculate $\{\eta_q\}$ using Eq. (A.3), and calculate $\{\eta_i\}$ using the inverse Fourier transform.

For the DPRM model simulated in Chapter 2, the system size is $L = 10^6$, evolved over $t = 10^4$ time steps. We check the correlation of the noise generated using the above method by averaging over $10^2$ realizations. The results for $\rho < 1/2$ and $\rho > 1/2$ are plotted in Fig. A-1 and A-2, respectively, and we see that the data is in good agreement with the expected values of $\rho$ up to a separation of $j = 10^3$. In the special

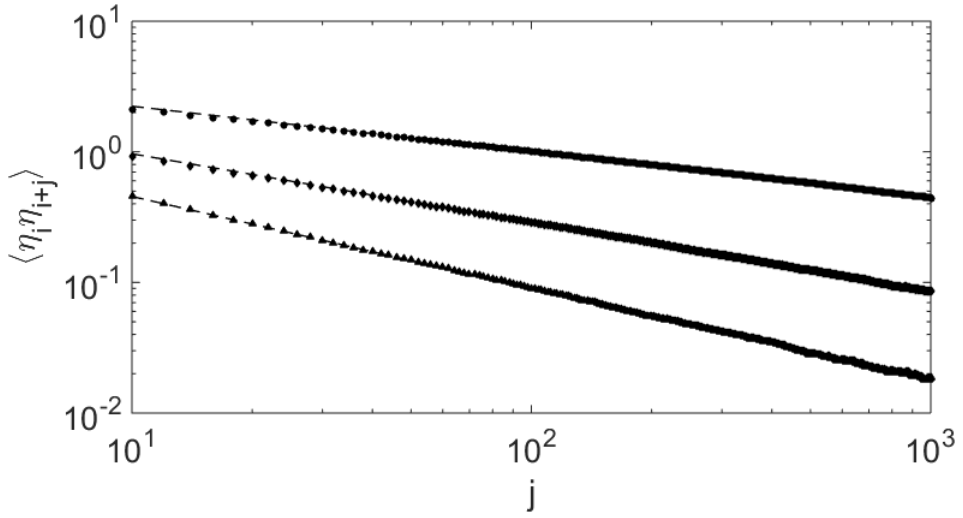


Figure A-1: Log-log plot of correlation of generated noise with $\rho < 1/2$ for system size $L = 10^6$. The data for $\rho = 0.35$, $\rho = 0.25$, and $\rho = 0.15$ are plotted (from top to bottom). The best fit lines (dashed) give $\rho = 0.33 \pm 0.02$, $\rho = 0.24 \pm 0.02$, and $\rho = 0.15 \pm 0.01$ respectively.

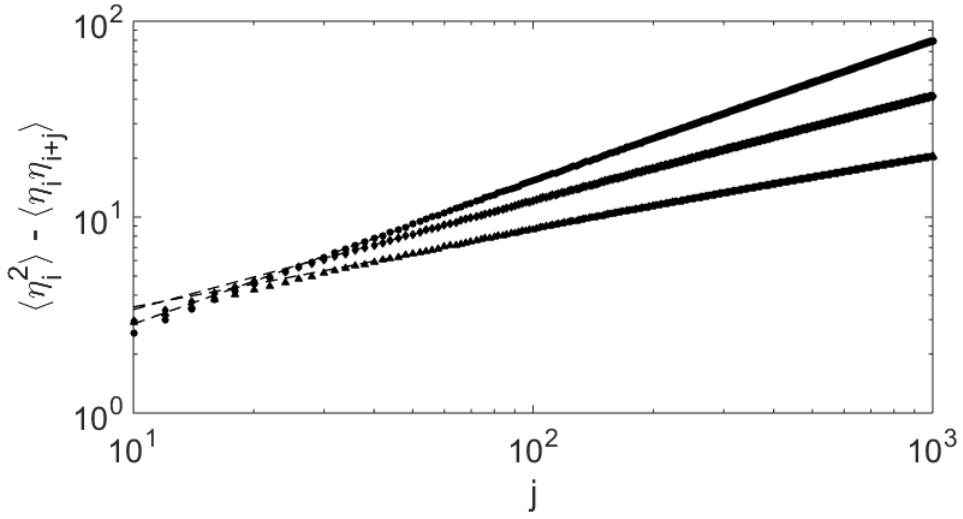


Figure A-2: Log-log plot of correlation of generated noise with $\rho > 1/2$ for system size $L = 10^6$. The data for $\rho = 0.85$, $\rho = 0.75$, and $\rho = 0.65$ are plotted (from top to bottom). The best fit lines (dashed) give $\rho = 0.86 \pm 0.03$, $\rho = 0.77 \pm 0.04$, and $\rho = 0.69 \pm 0.04$ respectively.

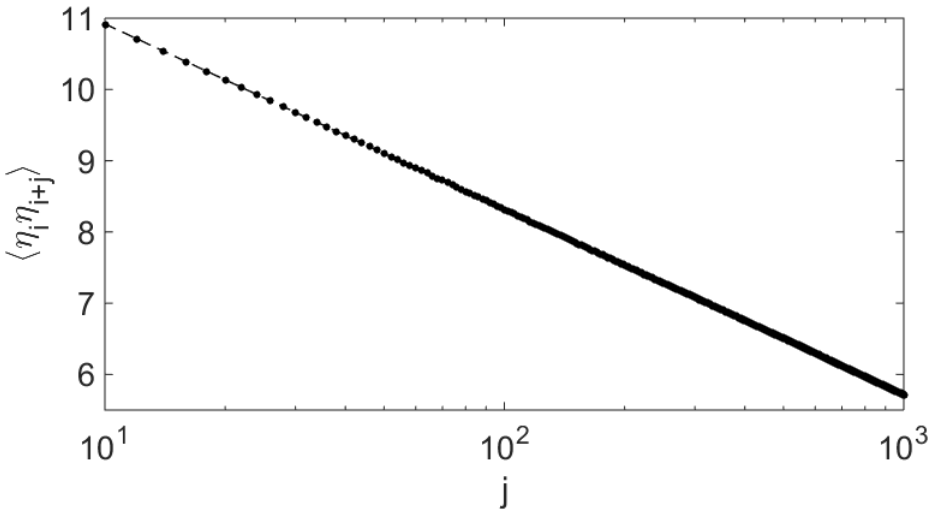


Figure A-3: Semilog plot of correlation of generated noise with $\rho = 1/2$ for system size $L = 10^6$. The correlation decays logarithmically with separation, and the best fit line (dashed) gives $C(j) = -1.13 \ln |j| + 13.52$.

case $\rho = 1/2$, we find that the correlation decays logarithmically with separation, as shown in Fig. A-3.

# Appendix B

# Overhangs in optimal paths on road networks

We define here properly the overhang length $L_h$ displayed in Fig. 3-3. Our algorithm returns a list of points $\{P_i\}$ along the path. For each of them we compute its coordinate $x_i$ on the axis running between the end points of the path. We then consider a point as part of an overhang if we have $x_i < x_j$ for some $j < i$. If the point falls outside the end points ($x < 0$ or $x > d$) it is also considered as part of an overhang. This definition is illustrated in the sketch below where the red portion of the path are defined as overhangs. (The overhangs are exaggerated in Fig. B-1; in reality, for an optimal path of sufficient length, their contribution to the overall trajectory is negligible.)

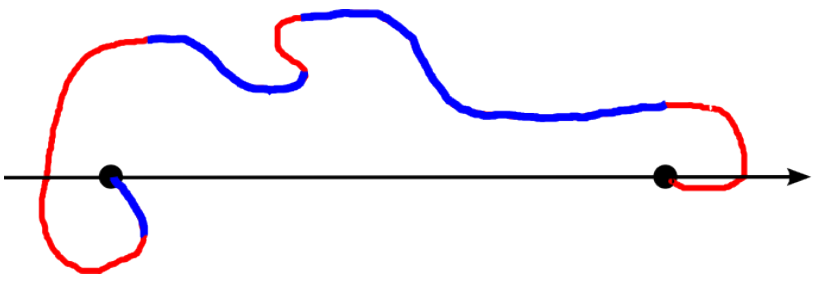

Figure B-1: Illustration of overhangs in an optimal path from GPS routing. The overhang length $L_h$ is defined as the length of the portion in red.

# Appendix C

# Convergence of distributions for optimal paths in different regions

We show in Fig. C-1 the distributions $P(L|d)$ (as in Fig. 3-4 top) for the shortest paths in the US and Asia. Fig. C-2 shows the equivalent curves $P(T|d)$ for the fastest paths in the three regions. Rescaling happens in a similar way as for the shortest paths except for a deviation in the Europe data set at $d = 100$km (Fig. C-2 top left). In Fig. C-2 (bottom left), we plot the end points of the paths corresponding to the left bump of the distribution. They appear concentrated along the German highway network which has no speed limitations on large portions and thus allows for faster routes.

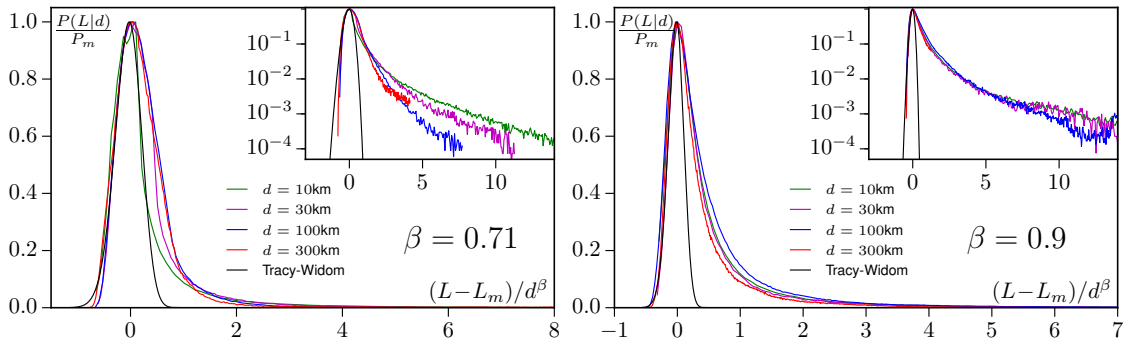


Figure C-1: Distribution of the length of the shortest path in the US (left) and Asia (right) as in Fig. 3-4 top.

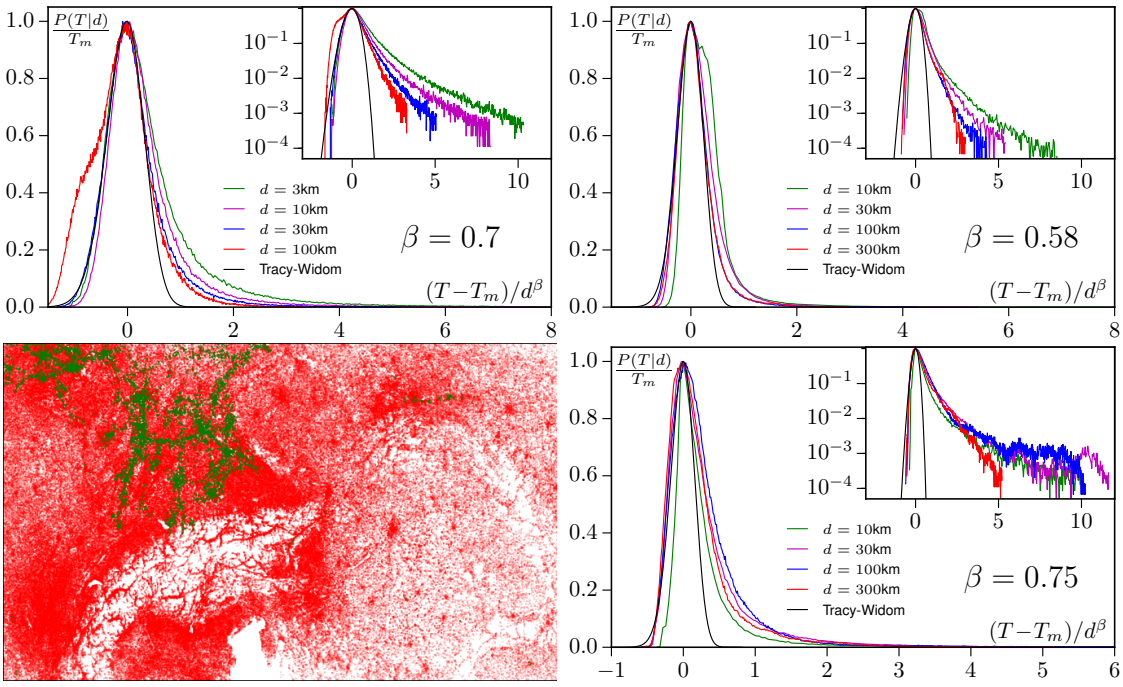


Figure C-2: Distribution of the travel time $P(T|d)$ on the fastest paths in Europe (top left), the US (top right) and Asia (bottom right). The bottom left corner shows the end points of the paths between points at distance $d = 100$km sampled in the top left figure. Green points correspond to the fast paths $T < 4500$ that make up the left bump in the distribution. Red points correspond to all slower paths.

# Appendix D

# Analytical derivation of the rate of coalescence for DPRM

Here we derive the form of the lineage coalescence rate in rough front range expansions, Eq. 4.4, using the DPRM endpoint distribution obtained in Ref. [107].

Consider two directed paths $x_1(\tau)$ and $x_2(\tau)$ starting from $x_1(0) = 0$ and $x_2(0) = \Delta x_0 > 0$ at $\tau = 0$. At a later time $\tau$, for $\tau/\Delta x_0^{3/2} \ll 1$, the spatial fluctuations for each path are small compared to their initial separation $\Delta x_0$, and we can consider the two paths to be independent. More specifically, setting $\tilde{x} = x/\tau^{2/3}$, we can take the rescaled $\tilde{x}_1$ and $\tilde{x}_2$ to be i.i.d. random variables drawn from the asymptotic DPRM endpoint distribution $f_{\text{end}}$ obtained in [107]. The probability distribution $f_{21}$ for the random variable $\tilde{x} = \tilde{x}_2 - \tilde{x}_1$ is then obtained from the convolution of the individual endpoint distributions, as

$$f_{21}(\tilde{x}) = \int_{-\infty}^{\infty} f_{\text{end}}(\tilde{y}) f_{\text{end}}(\tilde{y} - (\Delta\tilde{x}_0 - \tilde{x}))\, d\tilde{y}. \tag{D.1}$$

For $\Delta\tilde{x}_0 \gg 1$, we are interested in the tails of the $f_{\text{end}}$ distribution, which are known to decay as $f_{\text{end}}(z) \sim \exp(-cz^3)$ with $c$ a system-specific constant [107]. This allows us to estimate the integral in Eq. D.1 using the saddle point method. The maximum of the exponent $g(\tilde{y}) = c|\tilde{y}|^3 + c|\tilde{y} - (\Delta\tilde{x}_0 - \tilde{x})|^3$ occurs at $\tilde{y}_* = (\Delta\tilde{x}_0 - \tilde{x})/2$,

yielding

$$f_{21}(\tilde{x}) \sim \frac{\exp(-g(\tilde{y}_*))}{\sqrt{g''(\tilde{y}_*)}} \sim \frac{1}{\sqrt{\tilde{x}_0 - \tilde{\tilde{x}}}} \exp\left(-\frac{c}{4}(\Delta\tilde{x}_0 - \tilde{x})^3\right). \tag{D.2}$$

The coalescence events are represented by $\tilde{x} < 0$, resulting in the cumulative coalescence probability

$$C(\Delta\tilde{x}_0) = \int_{-\infty}^{0} f_{21}(\tilde{x}) d\tilde{x} \sim \Gamma\left(\frac{1}{6}, \frac{c\Delta\tilde{x}_0^3}{4}\right), \tag{D.3}$$

where $\Gamma(x, y)$ is the incomplete gamma function. After properly normalizing and differentiating with respect to $\tau$, we obtain the rate of coalescence displayed in Eq. 4.4,

$$J(\tau|\Delta x_0) \sim \frac{1}{\tau}\left(\frac{\Delta x_0^{3/2}}{\tau}\right)^{1/3} \exp\left(-\frac{c\Delta x_0^3}{4\tau^2}\right). \tag{D.4}$$

# Appendix E

# Scaling of the expected time to coalescence for KPZ walkers

Analogous to the diffusive case given by Eq. 4.7, the expected time to coalescence $T_2$ for KPZ walkers can be written in the form

$$\frac{T_{2,\mathrm{KPZ}}(\Delta x_0, t_{\max})}{t_{\max}} \propto f\left(\frac{\Delta x_0^{3/2}}{t_{\max}}\right), \tag{E.1}$$

where $f$ is some scaling function which depends only on the combination $\Delta x_0^{3/2}/t$, thus reflecting the KPZ wandering. To make this scaling relation evident, we plot a high quality collapse of the data from Fig. 4-6 in Fig. E-1.

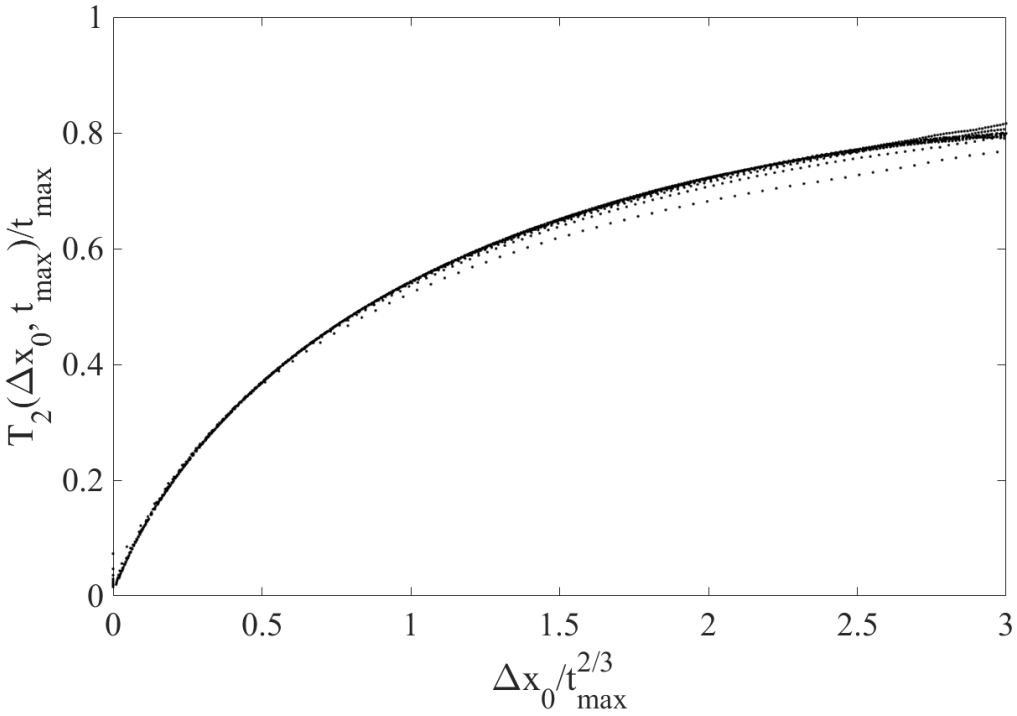


Figure E-1: Expected time to coalesce $T_2$ for KPZ lineages with initial separation $\Delta x_0$, collapsed with respect to the transverse scaling $\Delta x_0 \sim t_{\max}^{2/3}$. The lineages are taken from rough front stepping stone simulations of size $t_{\max} = 100$ to 1000.

# Appendix F

# Boundary fluctuations in the stepping stone model with wedge geometry

We present here a heuristic justification for the smooth increase in the wandering exponent $\zeta$ from $1/3$ to $2/3$ in the wedge geometry stepping stone model, as the wedge angle $\theta$ is increased from 0 to $\pi$.

Consider a wedge of opening angle $\theta$, with two distinct genotypes inoculated at its edges. In the case of *flat front growth* with velocity $u$, the advancing fronts meet at a tip which zips away from the initial apex as $y(t) = ut/\sin(\theta/2)$. With *rough front growth* the sector boundary is no longer straight but meanders as the intersection of the advancing fronts is no longer deterministic. At a time $t$, fluctuations of the front position are governed by KPZ scaling, growing as $t^{1/3}$. While on average the time for the tip to move a distance $y$ behaves as $y\sin(\theta/2)/u$, the fluctuations in this time scale as $[y\sin(\theta/2)/u]^{1/3}$.

The geometry is sketched in Fig. F-1. Height fluctuations $\delta h_L$, $\delta h_R$ push the advancing tip of the sector boundary – the intersection of the black dashed lines lines – away from $x = 0$, which is the zero-noise result illustrated by the intersection of the fainter blue dotted lines. From Fig. F-1, we can solve for the intersection point

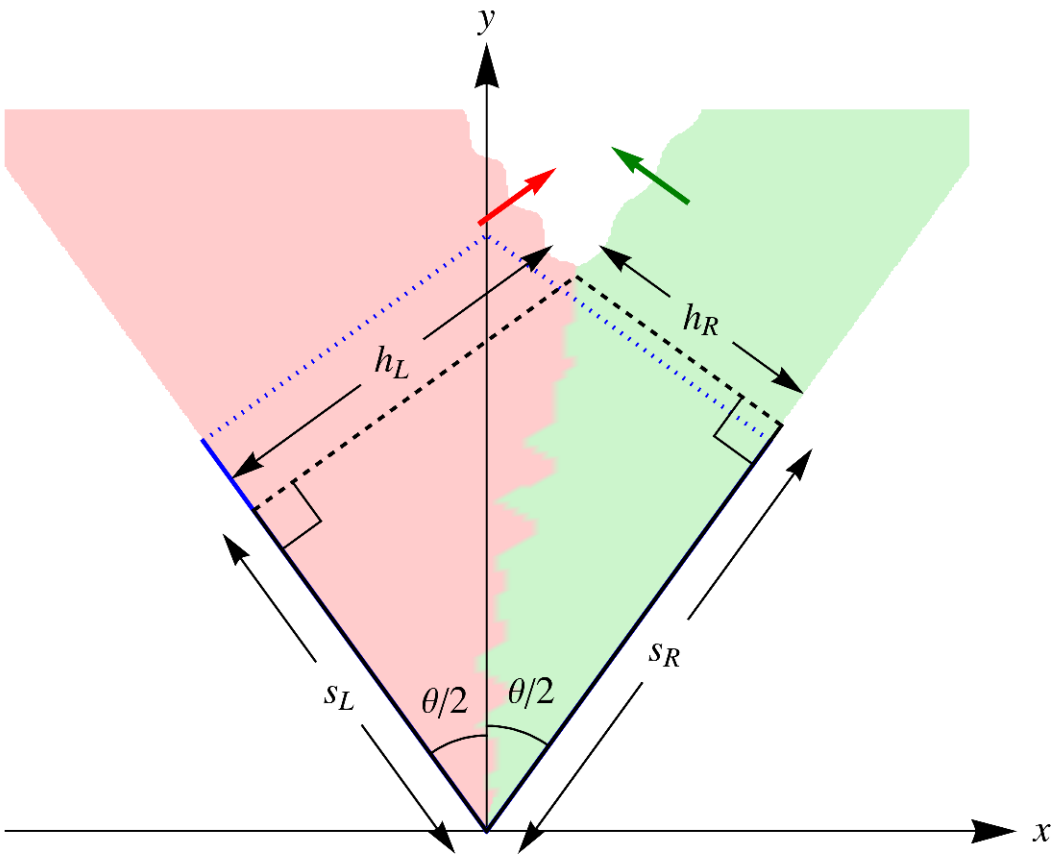


Figure F-1: Illustration of fluctuations in the wedge geometry with opening angle $\theta$. The red (left) and green (right) sectors meet at a sector boundary whose advancing tip, the intersection of the two dashed black lines, is pushed away from $x = 0$ by fluctuations in the front propagation heights $h_L$, $h_R$, which grow as $t^{1/3}$. The fainter blue dotted lines illustrate the zero-noise case (flat front). Coordinates $s_L$ and $s_R$ are defined to be orthogonal to $h_L$ and $h_R$, respectively.

$(x(t), y(t))$ representing the advancing tip:

$$x(t) = -s_L \sin(\theta/2) + h_L \cos(\theta/2) = s_R \sin(\theta/2) - h_R \cos(\theta/2),$$
$$y(t) = s_L \cos(\theta/2) + h_L \sin(\theta/2) = s_R \cos(\theta/2) + h_R \sin(\theta/2). \tag{F.1}$$

The height fluctuations $\delta h_L$, $\delta h_R$ can thus be expressed in terms of the resulting displacements $\delta x$, $\delta y$ of the tip, as

$$\delta h_L = \delta x \cos(\theta/2) + \delta y \sin(\theta/2),$$
$$\delta h_R = -\delta x \cos(\theta/2) + \delta y \sin(\theta/2), \tag{F.2}$$

from which we obtain

$$\delta x = \frac{\delta h_L - \delta h_R}{2\cos(\theta/2)}. \tag{F.3}$$

Both $\delta h_L$ and $\delta h_R$ scale as $ut^{1/3}$, which at a given $y$ value is $u[y\sin(\theta/2)/u]^{1/3}$. Therefore, the fluctuations in $x(t)$ for a given $y$-value of the tip vary as

$$\delta x \propto \frac{u}{\cos(\theta/2)} \left( \frac{y \sin(\theta/2)}{u} \right)^{1/3}. \tag{F.4}$$

While the meandering exponent remains as $\zeta = 1/3$, the overall amplitude increases with $\theta$, diverging as the wedge opens up to a single flat edge for $\theta \to \pi$. In that limit, the transverse fluctuations $\delta x$ scale as $t^{2/3}$.

# Bibliography


[1] E. Medina, T. Hwa, M. Kardar, and Y.-C. Zhang, Phys. Rev. A **39**, 3053 (1989).

[2] M. Kardar, G. Parisi, and Y.-C. Zhang, Phys. Rev. Lett. **56**, 889 (1986).

[3] D. Forster, D. R. Nelson, and M. J. Stephen, Phys. Rev. A **16**, 732 (1977).

[4] M. Kardar, Nucl. Phys. B **290** [FS20], 582 (1987).

[5] T. Sasamoto, and H. Spohn, Phys. Rev. Lett. **104**, 230602 (2010).

[6] T. Halpin-Healy, Phys. Rev. Lett. **109**, 170602 (2012).

[7] T. Halpin-Healy, Phys. Rev. E **88**, 042118 (2013).

[8] J. Krug and P. Meakin, J. Phys. A: Math. Gen. **23**, L987 (1990).

[9] J. Quastel and H. Spohn, J. Stat. Phys. **160**, 965 (2015).

[10] C. A. Tracy and H. Widom, Commun. Math. Phys. **177**, 727 (1996).

[11] C. A. Tracy and H. Widom, ICM **I**, 587 (2002).

[12] N. Wolchover, Quanta Magazine, 10/27 (2014).

[13] S. N. Majumdar and S. Nechaev, Phys. Rev. E **72**, 020901(R) (2005).

[14] C. Hajiyev, ISA Transactions **51**, 189 (2012).

[15] V. Dotsenko, Europhys. Lett. **90**, 20003 (2010).

[16] P. Calabrese, P. Le Doussal, and A. Rosso, Europhys. Lett. **90**, 20002 (2011).

[17] P. Calabrese and P. Le Doussal, Phys. Rev. Lett. **106**, 250603 (2011).

[18] K. A. Takeuchi and M. Sano, Phys. Rev. Lett. **104**, 230601 (2010).

[19] K. A. Takeuchi, M. Sano, T. Sasamoto, and H. Spohn, Sci. Rep. **1** 34 (2011).

[20] K. A. Takeuchi and M. Sano, J. Stat. Phys. **147**, 853 (2012).

[21] T. Gueudré and P. Le Doussal, Europhys. Lett. **100**, 26006 (2012).

[22] Y. Ito and K. A. Takeuchi, Phys. Rev. E **97**, 040103 (2018).

[23] M. L. Mehta, *Random Matrices*, 2nd ed. (Academic Press, New York, 1991).

[24] A. Edelman and P.-O. Persson, arXiv:math-ph/0501068, (2005).

[25] F. Bornemann, arXiv:0904.1581, (2009).

[26] J. M. Kim and J. M. Kosterlitz, Phys. Rev. Lett. **62**, 2289 (1989).

[27] J. M. Kim, J. M. Kosterlitz, and T. Ala-Nissila, J. Phys. A: Math. Gen. **24** 5569 (1991).

[28] M. J. Vold, J. Colloid Sci. **14**, 168 (1959).

[29] P. Meakin, J. Phys. A: Math. Gen. **20**, L1113 (1987).

[30] T. M. Liggett, *Interacting Particle Systems* (Springer-Verlag, New York, 1985).

[31] A. Borodin, P. L. Ferrari, M. Prähofer, and T. Sasamoto, J. Stat. Phys. **129**, 1055 (2007).

[32] M. Kardar and Y.-C. Zhang, Phys. Rev. Lett. **58**, 2087 (1987).

[33] T. Halpin-Healy and Y.-C. Zhang, Phys. Rep. **254**, 215 (1995).

[34] T. Halpin-Healy and K. A. Takeuchi, J. Stat. Phys. **160**, 794 (2015).

[35] I. Corwin, Random Matrices: Theory Appl. **1**, 1130001 (2012).

[36] J. M. Kim, M. A. Moore, and A. J. Bray, Phys. Rev. A **44**, 2345 (1991).

[37] J. M. Kim, M. A. Moore, and A. J. Bray, Phys. Rev. A **44**, R4782 (1991).

[38] T. Halpin-Healy, Phys. Rev. A **44**, R3415 (1991).

[39] T. Halpin-Healy and Y. Lin, Phys. Rev. E **89**, 010103(R) (2014).

[40] T. Gueudré, P. Le Doussal, J.-P. Bouchaud, and A. Rosso, Phys. Rev. E **91**, 062110 (2015).

[41] E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, and D. B. Shmoys, *The traveling salesman problem: a guided tour of combinatorial optimization*, Vol. 3 (Wiley, New York, 1985).

[42] K. Johansson, Commun. Math. Phys. **209**, 437 (2000).

[43] M. Prähofer and H. Spohn, Phys. Rev. Lett. **84**, 4882 (2000).

[44] M. Eden, *Symposium on Information Theory in Biology* (Pergamon Press, New York, 1958).

[45] M. Eden, *Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability* (University of California Press, 1961).

[46] S. Chu and M. Kardar, Phys. Rev. E **94**, 010101(R) (2016).

[47] Y.-C. Zhang, Physics A: Statistical Mechanics and its Applications, **170**, 1 (1990).

[48] J. Krug, J. de Phys. I, **1**, 9 (1991).

[49] G. Biroli, J.-P. Bouchaud, and M. Potters, J. Stat. Mech.: Theory and Experiment (2007), P07019.

[50] E. Frey, U. C. Täuber, and H. K. Janssen, Europhys. Lett. **47**, 14 (1999).

[51] T. Kloss, L. Canet, B. Delamotte, and N. Wschebor, Phys. Rev. E **89**, 022108 (2014).

[52] H. A. Makse, S. Havlin, M. Schwartz, H. E. Stanley, Phys. Rev. E **53**, 5445 (1996).

[53] N.-N. Pang, Y.-K. Yu, and T. Halpin-Healy, Phys. Rev. E **52**, 3224 (1995).

[54] J. G. Amar, P.-M. Lam, and F. Family, Phys. Rev. A **43**, 4548 (1991).

[55] C.-K. Peng, S. Havlin, M. Schwartz, and H. E. Stanley, Phys. Rev. A **44**, R2239 (1991).

[56] M. Wu and K. Y. R. Billah, Phys. Rev. E **51**, 995 (1995).

[57] K. Johansson, Probab. Theory Related Fields **138**, 75 (2007).

[58] R. Allez and L. Dumaz, J. Stat. Phys. **156**, 1146 (2014).

[59] A. P. Solon, G. Bunin, S. Chu, and M. Kardar, Phys. Rev. E **96**, 050301(R) (2017).

[60] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).

[61] A.-L. Barabàsi and R. Albert, Science **286**, 509 (1999).

[62] R. Albert and A.-L. Barabàsi, Rev. Mod. Phys. **74**, 47 (2002).

[63] E. Ravasz and A.-L. Barabàsi, Phys. Rev. E **67**, 026112 (2003).

[64] M. E. Newman, SIAM Rev. **45**, 167 (2003).

[65] M. Barthélemy, Phys. Rep. **499**, 1 (2011).

[66] P. S. Dodds and D. H. Rothman, Phys. Rev. E **63**, 016115 (2000).

[67] P. Sen, S. Dasgupta, A. Chatterjee, P. Sreeram, G. Mukherjee, and S. Manna, Phys. Rev. E **67**, 036106 (2003).

[68] D. Hunt and V. M. Savage, Phys. Rev. E **93**, 062305 (2016).

[69] A. Cardillo, S. Scellato, V. Latora, and S. Porta, Phys. Rev. E **73**, 066107 (2006).

[70] S. Lämmer, B. Gehlsen, and D. Helbing, Physica A Stat. Mech. Appl. **363**, 89 (2006).

[71] M. Barthélemy and A. Flammini, Phys. Rev. Lett. **100**, 138702 (2008).

[72] R. Geisberger, P. Sanders, D. Schultes, and D. Delling, in *Experimental Algorithms: 7th International Workshop*, WEA 2008 Provincetown, MA, USA, May 30-June 1, 2008, edited by C. C. McGeoch (Springer, Berlin, 2008), pp. 319-333.

[73] R. Schorr and H. Rieger, Eur. Phys. J. B **33**, 347 (2003).

[74] D. Luxen and C. Vetter, in *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (ACM, New York, 2011), pp. 513-516.

[75] M. Porto, S. Havlin, H. E. Roman, and A. Bunde, Phys. Rev. E **58**, R5205 (1998).

[76] B. Duplantier, Phys. Rev. B **35**, 5290 (1987).

[77] R. Metzler, A. Hanke, P. G. Dommersnes, Y. Kantor, and M. Kardar, Phys. Rev. Lett. **88**, 188101 (2002).

[78] D. A. Huse, C. L. Henley, and D. S. Fisher, Phys. Rev. Lett. **55**, 2924 (1985).

[79] N.-N. Pang and T. Halpin-Healy, Phys. Rev. E **47**, R784 (1993).

[80] H. A. Makse, J. S. Andrade, M. Batty, S. Havlin, and H. E. Stanley, Phys. Rev. E **58**, 7054 (1998).

[81] T. Song and H. Xia, J. Stat. Mech. Theory Exp. (2016) 113206.

[82] E. Katzav and M. Schwartz, Phys. Rev. E **70**, 011601 (2004).

[83] A. A. Fedorenko, Phys. Rev. B **77**, 094203 (2008).

[84] S. Chu, M. Kardar, D. R. Nelson, and D. A. Beller, submitted: J. Theor. Biol. (2019).

[85] K. Korolev, M. Avlund, O. Hallatschek, and D. R. Nelson, Rev. Mod. Phys **82**, 1691 (2010).

[86] O. Hallatschek, P. Hersen, S. Ramanathan, and D. R. Nelson, Proc. Nat. Acad. Sci. **104**, 19926 (2007).

[87] L. Excoffier and N. Ray, Trends in Ecology & Evolution **23**, 347 (2008).

[88] M. Gralka, F. Stiewe, F. Farrell, W. Möbius, B. Waclaw, and O. Hallatschek, Ecol. Lett. **19**, 889 (2016).

[89] J. F. Kingman, J. Appl. Prob. **19**, 27 (1982).

[90] J. Wakeley, *Coalescent theory: an introduction* (Roberts & Co., 2009). .

[91] H. M. Wilkinson-Herbots, J. Math. Biol. **37**, 535 (1998).

[92] M. Kimura and G. H. Weiss, Genetics **49**, 561 (1964).

[93] J. F. Wilkins and J. Wakeley, Genetics **161**, 873 (2002).

[94] Y. Saito and H. Müller-Krumbhaar, Phys. Rev. Lett. **74**, 4325 (1995).

[95] T. Halpin-Healy and Y.-C. Zhang, Phys. Rep. **254**, 215 (1995).

[96] W. Möbius, A. W. Murray, and D. R. Nelson, PLoS Comput. Biol. **11**, e1004615 (2015).

[97] G. Ódor, Rev. Mod. Phys. **76**, 663 (2004).

[98] H. Hinrichsen and M. Howard, Eur. Phys. J. B **7**, 635 (1999).

[99] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun, Current Biology **20**, 1099 (2010).

[100] M. O. Lavrentovich, K. S. Korolev, and D. R. Nelson, Phys. Rev. E **87** 012103 (2013).

[101] S. Redner, *A guide to first-passage processes* (Cambridge University Press, 2001).

[102] N. F. Barton, F. Depaulis, and A. Etheridge, Theor Popul. Biol. **61**, 31 (2002).

[103] G. Malécot, Theor Popul. Biol. **8**, 212 (1975).

[104] T. Nagylaki, Proc. Natl. Acad. Sci. U.S.A. **71**, 2932 (1974).

[105] H. Wilkinson-Herbots, J. Math. Biol. **37**, 535 (1998).

[106] J. Nullmeier and O. Hallatschek, Evolution **67**, 1307 (2013).

[107] G. M. Flores, J. Quastel, and D. Remenik, Commun. Math. Phys. **317**, 363 (2013).

[108] F. Tesser, Ph.D. thesis, Technische Universiteit Eindhoven (2016).

[109] B. Derrida and R. Dickman, J. Phys. A: Math. Gen. **24**, L191 (1991).

[110] L. Yan, J.-P. Bouchaud, and M. Wyart, Soft Matter **13**, 5795 (2017).

[111] V. L. Pokrovski and A. L. Talapov, Phys. Rev. Lett. **42**, 65 (1979).

[112] D. A. Huse and C. L. Henley, Phys. Rev. Lett. **54**, 2708 (1985).

[113] D. S. Fisher, Phys. Rev. Lett. **56**, 1964 (1986).

[114] G. Blatter, M. V. Feigel'man, V. B. Geshkenbein, A. I. Larkin, and V. M. Vinokur, Rev. Mod. Phys. **66**, 1125 (1994).

[115] M. Jaubert, A. Glachant, M. Bienfait, and G. Boato, Phys. Rev. Lett. **46**, 1679 (1981).

[116] K. Kern, R. David, P. Zeppenfeld, R. Palmer, and G. Comsa, Solid State Commun. **62**, 391 (1987).

[117] A. R. Kortan, A. Erbil, R. J. Birgeneau, and M. S. Dresselhaus, Phys. Rev. Lett. **49**, 1427 (1982).

[118] C. Rottman and M. Wortis, Phys. Rep. **103**, 59 (1984).

[119] C. Rottman, M. Wortis, J. C. Heyraud, and J. J. Métois, Phys. Rev. Lett. **52**, 1009 (1984).

[120] S. N. Coppersmith, D. S. Fisher, B. I. Halperin, P. A. Lee, and W. F. Brinkman, Phys. Rev. B **25**, 349 (1982).

[121] M. E. Fisher and D. S. Fisher, Phys. Rev. B **25**, 3192 (1982).

[122] J. Villain, Journal de Physique Lettres, **43**, 551 (1982).

[123] E. P. Wigner, Ann. Math. **67**, 325 (1958).

[124] J. Gustavsson, Annales de l'I.H.P. Probabilités et statistiques **41**, 151 (2005).

[125] S. O'Rourke, J. Stat. Phys. **138**, 1045 (2010).

[126] L. Erdős, H.-T. Yau, and J. Yin, Adv. Math. **229**, 1435 (2012).

[127] F. Benaych-Georges and A. Knowles, arXiv:1601.04055 (2016).

[128] C. Loewner, Mathematische Zeitschrift **63**, 338 (1955).

[129] T. Ando, Linear Algebra Appl. **90**, 165 (1987).

[130] C.-H. Lam, L. M. Sander, and D. E. Wolf, Phys. Rev. A **46**, R6128 (1992).

[131] M. Barma and P. Ray, Phys. Rev. B **34**, 3403 (1986).

[132] M O. Lavrentovich and D.R. Nelson, Phys. Rev. Lett. **112**, 138102 (2014).

[133] D. A. Beller, K. M. Alards, F. Tesser, R. A. Mosna, F. Toschi, and W. Möbius, Europhys. Lett. **123**, 58005 (2018).