Lecture Notes in Mathematics 2144

Anton Bovier · David Brydges
Amin Coja-Oghlan · Dmitry Ioffe
Gregory F. Lawler

Random Walks, Random Fields, and Disordered Systems

Marek Biskup · Jiří Černý
Roman Kotecký *Editors*

Springer

Lecture Notes in Mathematics 2144

More information about this series at http://www.springer.com/series/304

Anton Bovier • David Brydges •
Amin Coja-Oghlan • Dmitry Ioffe •
Gregory F. Lawler

Random Walks, Random Fields, and Disordered Systems

Marek Biskup • Jiří Černý • Roman Kotecký
Editors

Springer

Authors
Anton Bovier
Inst. for Applied Mathematics
University of Bonn
Bonn, Germany

David Brydges
Department of Mathematics
University of British Columbia
Vancouver, BC, Canada

Amin Coja-Oghlan
Mathematics Institute
Goethe University
Frankfurt, Germany

Dmitry Ioffe
Fac. of Industrial Engineering & Ma
Technion - Israel Institute of Technology
Haifa, Israel

Gregory F. Lawler
Department of Mathematics
University of Chicago
Chicago, IL, USA

Editors
Marek Biskup
Department of Mathematics
University of California in Los Angeles
Los Angeles, CA, USA

Jiří Černý
Faculty of Mathematics
University of Vienna
Vienna, Austria

Roman Kotecký
Department of Mathematics
University of Warwick
Warwick, United Kingdom

ISSN 0075-8434 ISSN 1617-9692 (electronic)
Lecture Notes in Mathematics
ISBN 978-3-319-19338-0 ISBN 978-3-319-19339-7 (eBook)
DOI 10.1007/978-3-319-19339-7

Library of Congress Control Number: 2015949243

Mathematics Subject Classification (2010): 00-02, 60-02, 60-06, 60J85, 60K35, 60K05, 82D30, 82B28, 82B41, 05D40, 82B26

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Foreword

These Lecture Notes contain detailed expositions of five (out of six total) lecture series delivered at the 2013 Prague School on Mathematical Statistical Mechanics. The lecture series aimed to address some of the currently most rapidly developing subjects at the interface of probability theory, statistical physics, combinatorics, and computer science. One common thread running through all lectures is the prominent role played by random walks and random fields; these appear both as primary objects of study or as tools to represent other quantities of interest. Another important common aspect is the role of underlying disorder that is responsible for many spectacular effects found in the behavior of these systems. We will now describe the content of the forthcoming chapters in more detail, highlighting the important features and further connections.

The opening chapter, contributed by Anton Bovier, summarizes recent advances in the understanding of the Branching Brownian Motion. This is a process of particles that perform independent Brownian motions until they branch, at random times, to form more Brownian particles that henceforth move (and branch) independently. Such models are naturally interesting for population dynamics, although other uses can be found in the literature as well. The present notes address the Branching Brownian Motion from the perspective of statistical mechanics of disordered systems, particularly, spin glasses with Gaussian disorder. What connects these together is the reliance on extreme order statistics, a method for systematic study of the maximum and, in fact, all near-maximal values of a large number of random variables. There is also a close link to the subject of random fields as the extrema of the Branching Brownian Motion turn out to behave quite similarly to those of the two-dimensional Gaussian Free Field. The latter is a topic that plays an important role in several other chapters of the Lecture Notes as well.

The second chapter, written by David Brydges, explains how the behavior of the four-dimensional weakly self-avoiding walk can be studied using the methods of field theory and renormalization group. A representation of a weakly self-avoiding

random walk by means of an interacting random field (specifically, the ϕ^4-theory) goes back several decades, but it was only recently that this connection could be exploited to obtain rigorous conclusions about the random walks as well. This is particularly challenging in four spatial dimensions, where the renormalization group has to deal with (what are called) marginal terms. As a result, logarithmic corrections appear next to the ordinary "diffusive" or "free-field like" scaling of the two-point correlation function known to occur above the upper critical dimension. Brydges's notes explain all necessary background for entering this field and, in particular, his own extensive papers on the subject that have been circulated recently.

The third contribution to these Lecture Notes, written by Amin Coja-Oglan, takes us to the world of discrete structures exhibiting phase transitions as the parameters of the problem vary. The "interaction" in these systems is usually of a rather simple kind; what makes the problem hard is the disorder that is built into the system. The main focus of the notes is a specific tool, called the cavity method, whose name fittingly describes the basic idea: evaluate the change of the physical properties of the system when a "cavity" is created by removing a single constituent. The cavity method is commonly used by physicists to study spin-glass systems; unfortunately, little of it has been put on rigorous footing until recently. Coja-Oghlan's notes describe some of the mathematical advances using salient examples of the Random-Graph Coloring and the Random Satisfiability Problem. Various other concepts that arise in this subject area are also explained and used in practice, e.g., the replica symmetry breaking, belief propagation, or quiet planting.

The fourth chapter of these notes, contributed by Dmitry Ioffe, is a survey of recent advances in the area of polymer models with disorder. The focus is on the ballistic regime, i.e., the situation when the polymer spreads linearly in space. The polymer itself is subject to self-interactions as well as interactions with a random environment. A rigorous version of the Ornstein-Zernike theory is reviewed, which gives full control of the limiting distribution of the polymer endpoint in the ballisticity regime. The role of a strong disorder in low spatial dimensions and a weak disorder in high spatial dimensions is discussed. Renormalization methods are employed to make a link to effective models that are easier to study.

The final chapter of the notes, written jointly by Greg Lawler and Jacob Perlman, summarizes Lawler's lectures on random walk loop-soup models. Here, a loop-soup is a family of random closed paths on a graph or in the Euclidean space that are subject to an interaction that depends on (what Lawler and Perlman call) an acceptable weight. Several two-dimensional systems of interest have a natural representation by way of a loop-soup measure, the self-avoiding walk being one natural example. The notes give an introduction to the approach based on complex weights and the relation between loop soups, the loop-erased random walk and uniform spanning trees. There are also natural connections to the Gaussian Free Field as well as other topics arising elsewhere in these Lecture Notes.

These Lecture Notes are primarily aimed at early graduate students with only a modest background in probability and mathematical physics. Notwithstanding, the notes will be enjoyed also by seasoned researchers as well as general audiences interested in learning about recent advances in the above fields.

Los Angeles, CA, USA — Marek Biskup
Vienna, Austria — Jiří Černý
Warwick, UK — Roman Kotecký
December 2014

Contents

Contributors

Anton Bovier Institut für Angewandte Mathematik, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany

David Brydges Department of Mathematics, The University of British Columbia, Vancouver, BC, Canada

Amin Coja-Oghlan Mathematics Institute, Goethe University, Frankfurt, Germany

Dmitry Ioffe Faculty of IE & M, Technion, Haifa, Israel

Gregory F. Lawler Department of Mathematics, University of Chicago, Chicago, IL, USA

Jacob Perlman Department of Mathematics, University of Chicago, Chicago, IL, USA

From Spin Glasses to Branching Brownian Motion—and Back?

Anton Bovier

1 Introduction

These lecture notes are mainly about work that I have done over the last 5 years with Louis-Pierre Arguin and Nicola Kistler on branching Brownian motion (BBM) [3–7]. Louis-Pierre and Nicola should be considered co-authors of these notes (but not responsible for any errors or nonsense I may introduce).

BBM was studied over the last 50 years as a subject of interest in its own right, with seminal contributions by McKean [44], Bramson [18, 19], Lalley and Sellke [40], Chauvin and Rouault [22, 23], and others. Recently, the field has experienced a revival with many remarkable contributions and repercussions into other areas. My personal motivation comes from the connection to spin glass theory, and specifically the so-called Generalised Random Energy models (GREM) introduced by Gardner and Derrida [30–32]. These models can be seen as special examples of Gaussian processes on trees, and so can BBM. One question of central interest in spin glass theory is to understand the structure of *ground states*, which translates into the analysis of the extremal process in the languor of extreme value theory of stochastic processes.

In work I did with Irina Kurkova [15, 16] it emerged that BBM is a particularly interesting example, lying right at the borderline where correlations begin to influence the behaviour of the extremes of the process. This made the study of the *extremal process* of BBM an obvious target.

This goal was reached in our work with Arguin and Kistler [6] as well as, in parallel, in that of Aïdékon, Beresticky, Brunet, and Shi [2]. A review of this work is at the centre of these notes, in Sect. 4. To begin, however, I will, in Sect. 2 explain

A. Bovier (✉)
Institut für Angewandte Mathematik, Rheinische Friedrich-Wilhelms-Universität Bonn,
Endenicher Allee 60, 53115 Bonn, Germany
e-mail: bovier@uni-bonn.de

M. Biskup et al. (eds.), *Random Walks, Random Fields, and Disordered Systems*,
Lecture Notes in Mathematics 2144, DOI 10.1007/978-3-319-19339-7_1

the connection to spin glasses and the models of Derrida. In Sect. 3 I review some fundamental background work on BBM. The final Sect. 5 takes a swing back to Derrida's model and describes some recent results on so-called variable speed BBM that is motivated by these models.

Many results on BBM were extended to branching random walk, see e.g. Aïdékon [1]. Other developments concern the connection to extremes of the free Gaussian random field in $d = 2$ by Bramson and Zeitouni [20], Ding [26], Bramson, Ding, and Zeitouni [21], and Biskup and Louidor [11].

The recent activities in and around BBM have triggered a number of lecture notes and reviews, that hopefully are complementary to this one. I mention the review by Gouéré [33] that presents and compares the approaches of Arguin et al. and Aïdékon et al. Ofer Zeitouni has lecture notes on his homepage that deal with branching random walk and the Gaussian free field [51]. Nicola Kistler just wrote a survey linking REM and GREM to other correlated Gaussian processes [38].

2 Spin Glasses

The motivation for the results I present in these lectures comes from spin glass theory. There is really not enough time to go into this at any depth and I will only sketch a few key concepts. Those interested in more should read a book on the subject, e.g. [12, 50] or [47], or the less technical introduction by Newman and Stein [45].

2.1 Setting and Examples

Spin glasses are spin systems with competing, random interactions. To remain close to the main frame of these lectures, let us remain in the realm of mean-field spin glasses of the type proposed by Sherrington and Kirkpatrick.

Here we have a state space, $\mathscr{S}_N \equiv \{-1, 1\}^N$. On this space we define a Gaussian process $H_N : \mathscr{S}_N \to \mathbb{R}$, characterised by its mean $\mathbb{E}H_N(\sigma) = 0$, for all $\sigma \in \mathscr{S}_N$, and covariance

$$\mathbb{E}H_N(\sigma)H_N(\sigma') \equiv Ng(\sigma, \sigma'), \qquad \sigma, \sigma' \in \mathscr{S}_N, \tag{1}$$

where g is a positive definite quadratic from. The random functions H_N are called *Hamiltonians*. In physics, they represent the energy of a configuration of spins, σ. In the examples that are most popular, g is assumed to depend only on some *distance* between σ and σ'. There are two popular distances:

- The Hamming distance,

$$d_{ham}(\sigma,\sigma') \equiv \sum_{i=1}^{N} \mathbb{1}_{\sigma_i \neq \sigma'_i} = \frac{1}{2}\left(N - \sum_{i=1}^{N} \sigma_i \sigma'_i\right). \tag{2}$$

  In this case one choses

$$g(\sigma,\sigma') = p\left(1 - 2d_{ham}(\sigma,\sigma')/N\right), \tag{3}$$

  with p a polynomial with non-negative coefficients. The most popular case is $p(x) = x^2$. One denotes the quantity

$$1 - 2d_{ham}(\sigma,\sigma')/N \equiv R_N(\sigma,\sigma') \tag{4}$$

  as the *overlap* between σ and σ'. The resulting class of models are called *Sherrington-Kirkpatrick* models.
- A second popular distance is the lexicographic distance

$$d_{lex}(\sigma,\sigma') = N + 1 - \min(i : \sigma_i \neq \sigma'_i), \tag{5}$$

  Note the d_{lex} is an *ultrametric*. In this case,

$$g(\sigma,\sigma') = A(1 - N^{-1} d_{lex}(\sigma,\sigma')). \tag{6}$$

  $A : [0,1] \to [0,1]$ can here be chosen as any non-decreasing function. It will be convenient to denote $q_N(\sigma,\sigma') \equiv N^{-1}(\min(i : \sigma_i \neq \sigma'_i) - 1)$ and call it the ultrametric overlap. The models obtained in this way were introduced by Derrida and Gardener [30, 31] and are called *generalised random energy models* (GREM).

The main objects one studies in statistical mechanics are the *Gibbs measures* associated to these Hamiltonians. They are probability measures on $\mathscr{S}_N$ that assign to $\sigma \in \mathscr{S}_N$ the probability

$$\mu_{\beta,N}(\sigma) \equiv \frac{\mathrm{e}^{-\beta H_N(\sigma)}}{Z_{\beta,N}}, \tag{7}$$

where the normalising factor $Z_{\beta,N}$ is called the *partition function*. The parameter β is called the *inverse temperature*. Note that these measures are *random variables* on the underlying probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ on which the Gaussian processes H_N are defined.

The objective of statistical mechanics is to understand the geometry of these measures for very large N. This problem is usually interesting in the case when β is large, where the Gibbs measures will feel the geometry of the random process

H_N. In particular, one should expect that for large enough β (and finite N), $\mu_{\beta,N}$ will assign mass only to configurations of almost minimal energy.

One way to acquire information on the Gibbs measures its thus to first analyse the structure of the minima of H_N.

2.2 Classical Extreme Value Theory, Aka the REM

In order to illustrate what can go on in spin glasses, Derrida had the bright idea to introduce a simple toy model, the *random energy model* (REM). In the REM, the values $H_N(\sigma)$ of the Hamiltonian are simply independent Gaussian random variables with mean zero and variance N. One might think that this should be *too* simple, but, remarkably, quite a lot can be learned from this example.

Understanding the structure of the ground states in this model turns into the classical problem of extreme value theory for iid random variables, which is of course extremely well understood (there are numerous textbooks of which I like [41] and [48] best). We will set $H_N(\sigma) \equiv -X_\sigma$.

The first question is for the value of the maximum of the $X_\sigma, \sigma \in \mathscr{S}_N$.

2.3 Rough Estimates, the Second Moment Method

To get a first feeling we ask for the right order to the maximum. More precisely, we ask for the right choice of functions $u_N : \mathbb{R} \to \mathbb{R}$, such that the maximum is of that order $u_N(x)$ with positive, x-dependent probability. To do so, introduce the counting function

$$M_N(x) \equiv \sum_{\sigma \in \mathscr{S}_N} \mathbb{1}_{X_\sigma > u_N(x)}, \tag{8}$$

which is the number of σ's such that X_σ exceeds $u_N(x)$. Then

$$\mathbb{P}\left(\max_{\sigma \in \mathscr{S}_N} X_\sigma > u_N(x)\right) \leq \mathbb{E}M_N(x) = 2^N \mathbb{P}(X_\sigma > u_N(x)). \tag{9}$$

This may be called a *first moment estimate*. So a good choice of $u_N(x)$ would be such that $2^N \mathbb{P}(X_\sigma > u_N(x)) = O(1)$. For later use, we even choose

$$2^N \mathbb{P}(X_\sigma > u_N(x)) \sim \mathrm{e}^{-x}. \tag{10}$$

It is easy to see that this can be achieved with

$$u_N(x) = N\sqrt{2\ln 2} - \frac{1}{2}\frac{\ln(N\ln 2) + \ln(4\pi)}{\sqrt{2\ln 2}} + \frac{x}{\sqrt{2\ln 2}}. \tag{11}$$

The fact that $\mathbb{E}M_N(x) = O(1)$ precludes that there are too many exceedances of the level $u_N(x)$, by Chebyshev's inequality. But it may still be true that the probability that $M_N(x) > u_N(x)$ tends to zero. To show that this in not the case, one may want to control, e.g., the variance of $M_N(x)$. Namely, using the Cauchy-Schwartz inequality, we have that

$$(\mathbb{E}M_N(x))^2 = \left(\mathbb{E}(M_N(x)\mathbb{1}_{M_N(x)\geq 1})\right)^2 \leq \mathbb{E}\left(M_N(x)^2\right)\mathbb{P}\left(M_N(x)\geq 1\right), \tag{12}$$

so that

$$\mathbb{P}\left(M_N(x)\geq 1\right) \geq \frac{(\mathbb{E}M_N(x))^2}{\mathbb{E}\left(M_N(x)^2\right)}. \tag{13}$$

Now it is easy to compute

$$\begin{aligned}\mathbb{E}\left(M_N(x)^2\right) &= 2^N\mathbb{P}\left(X_\sigma > u_N(x)\right) + 2^N(2^N-1)\mathbb{P}\left(X_\sigma > u_N(x)\right)^2 \\ &= (\mathbb{E}M_N(x))^2\left(1-2^{-N}\right) + \mathbb{E}M_N(x).\end{aligned} \tag{14}$$

Hence

$$\mathbb{P}\left(M_N(x)\geq 1\right) \geq \frac{1}{1-2^{-N}+1/\mathbb{E}M_N(x)} \sim \frac{1}{1+\mathrm{e}^x}, \tag{15}$$

for N large. Together with the Chebyshev upper bound, this gives already good control on the probability to exceed $u_N(x)$, at least for large x:

$$\frac{\mathrm{e}^{-x}}{1+\mathrm{e}^{-x}} \leq \mathbb{P}\left(\max_{\sigma\in\mathscr{S}_N} X_\sigma > u_N(x)\right) \leq \mathrm{e}^{-x}. \tag{16}$$

In particular,

$$\lim_{x\uparrow\infty}\mathrm{e}^x \lim_{N\uparrow\infty}\mathbb{P}\left(\max_{\sigma\in\mathscr{S}_N} X_\sigma > u_N(x)\right) = 1, \tag{17}$$

which is a bound on the upper tail of the distribution of the maximum. This is called the *second moment estimate*. First and second moment estimates are usually easy to use and give the desired control on maxima, if they work. Unfortunately, they do not always work as nicely as here.

2.3.1 Law of the Maximum and the Extremal Process

In this simple case, we can of course do better. Just use that, by independence,

$$\begin{aligned}\mathbb{P}\left(\max_{\sigma\in\mathscr{S}_N} X_\sigma \le u_N(x)\right) &= \left(\mathbb{P}\left(X_\sigma \le u_N(x)\right)\right)^{2^N} \\ &= \left(1-\mathbb{P}\left(X_\sigma > u_N(x)\right)\right)^{2^N} \\ &\sim \left(1-2^{-N}\mathrm{e}^{-x}\right)^{2^N} \to \mathrm{e}^{-\mathrm{e}^{-x}},\end{aligned} \tag{18}$$

as $N \uparrow \infty$. This is the classic result that the distribution of the rescaled maximum of iid Gaussian random variables is Gumbel distributed [41].

The same simple argument yields a lot more: let

$$\mathscr{E}_N \equiv \sum_{\sigma\in\mathscr{S}_N} \delta_{u_N^{-1}(X_\sigma)} \tag{19}$$

denote the point process (see the Appendix for a brief review on the theory of point processes) of the rescaled X_σ. Then we can look at the so-called Laplace functional of $\mathscr{E}_N$: For $\phi : \mathbb{R} \to \mathbb{R}_+$ a continuous, non-negative function of compact support, we set

$$\Psi_N(\phi) \equiv \mathbb{E}\left[\exp\left(-\int \phi(x)\mathscr{E}_N(\mathrm{d}x)\right)\right] = \mathbb{E}\left[\exp\left(-\sum_{\sigma\in\mathscr{S}_N}\phi\left(u_N^{-1}(X_\sigma)\right)\right)\right]. \tag{20}$$

It is well-known that the Laplace functional determines the law of the process $\mathscr{E}_N$, and that convergence to a limit implies convergence in law of $\mathscr{E}_N$ to some point process (see the Appendix). The computations in the iid case are very simple. By independence,

$$\begin{aligned}\mathbb{E}\left[\exp\left(-\sum_{\sigma\in\mathscr{S}_N}\phi\left(u_N^{-1}(X_\sigma)\right)\right)\right] &= \prod_{\sigma\in\mathscr{S}_N}\mathbb{E}\left[\exp\left(-\phi\left(u_N^{-1}(X_\sigma)\right)\right)\right] \\ &= \left(\mathbb{E}\left[\exp\left(-\phi\left(u_N^{-1}(X_\sigma)\right)\right)\right]\right)^{2^N}.\end{aligned} \tag{21}$$

We know that the probability for $u_N^{-1}(X_\sigma)$ to be in the support of ϕ is small of order 2^{-N}. Thus if we write

$$\mathbb{E}\left[\mathrm{e}^{-\phi\left(u_N^{-1}(X_\sigma)\right)}\right] = 1 + \mathbb{E}\left[\mathrm{e}^{-\phi\left(u_N^{-1}(X_\sigma)\right)} - 1\right], \tag{22}$$

the second term will be of order 2^{-N}. Therefore,

$$\left(\mathbb{E}\left[\mathrm{e}^{-\phi\left(u_N^{-1}(X_\sigma)\right)}\right]\right)^{2^N} \sim \exp\left(2^N \mathbb{E}\left[\mathrm{e}^{-\phi\left(u_N^{-1}(X_\sigma)\right)} - 1\right]\right) \tag{23}$$

Finally,

$$\lim_{N\uparrow\infty} 2^N \mathbb{E}\left[\mathrm{e}^{-\phi\left(u_N^{-1}(X_\sigma)\right)} - 1\right] = \int \mathrm{e}^{-x}\left(\mathrm{e}^{-\phi(x)} - 1\right)\mathrm{d}x. \tag{24}$$

To show the latter, note first that $u_N^{-1}(x) = \sqrt{2\ln 2}\,(x - u_N(0))$. Now assume that ϕ is supported on the interval $[a\sqrt{2\ln 2}, b\sqrt{2\ln 2}]$. Then

$$\begin{aligned}
\mathbb{E}\left[\mathrm{e}^{-\phi\left(u_N^{-1}(X_\sigma)\right)} - 1\right] &= \frac{1}{\sqrt{2\pi N}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{z^2}{2N}} \left(\mathrm{e}^{-\phi\left(\sqrt{2\ln 2}(z - u_N(0))\right)} - 1\right)\mathrm{d}z \qquad (25)\\
&= \frac{1}{\sqrt{2\pi N}} \int_{a+u_N(0)}^{b+u_N(0)} \mathrm{e}^{-\frac{z^2}{2N}} \left(\mathrm{e}^{-\phi\left(\sqrt{2\ln 2}(z - u_N(0))\right)} - 1\right)\mathrm{d}z\\
&= \frac{1}{\sqrt{2\pi N}} \int_{a}^{b} \mathrm{e}^{-\frac{(z+u_N(0))^2}{2N}} \left(\mathrm{e}^{-\phi\left(\sqrt{2\ln 2}z\right)} - 1\right)\mathrm{d}z
\end{aligned}$$

Now

$$\mathrm{e}^{-(z+u_N(0))^2/2N} = \mathrm{e}^{-u_N(0)^2/2N}\mathrm{e}^{-zu_N(0)/N}\mathrm{e}^{-z^2/2N}. \tag{26}$$

By definition of u_N, as $N \uparrow \infty$,

$$2^N \frac{\mathrm{e}^{-u_N(0)^2/2N}}{\sqrt{2\pi N}} \to \sqrt{2\ln 2}, \tag{27}$$

$$zu_N(0)/N \to \sqrt{2\ln 2}z, \tag{28}$$

and $\mathrm{e}^{-z^2/2N} \to 1$, uniformly on the domain of integration. This shows that

$$\lim_{N\uparrow\infty} \mathbb{E}\left(\mathrm{e}^{-\phi\left(u_N^{-1}(X_\sigma)\right)} - 1\right) = \int_{a\sqrt{2\ln 2}}^{b\sqrt{2\ln 2}} \mathrm{e}^{-z}\left(\mathrm{e}^{-\phi(z)} - 1\right)\mathrm{d}z. \tag{29}$$

Since the integrand vanishes beyond the integration domain, we can extend the integration range to all $\mathbb{R}$ which yields (24). Putting all things together, we have shown:

Theorem 2.1 *In the REM, for any positive continuous function ϕ of bounded support,*

$$\lim_{N\uparrow\infty} \Psi_N(\phi) = \exp\left(\int \left(\mathrm{e}^{-\phi(z)} - 1\right) \mathrm{e}^{-z} \mathrm{d}z\right). \tag{30}$$

Now the right-hand side of (30) is recognised as the Laplace functional of a Poisson point process with intensity measure $\mathrm{e}^{-z}\mathrm{d}z$. Therefore we have as a corollary the classical result on Poisson convergence of the extremal process [41]:

Corollary 2.2 *In the REM, the sequence of point processes $\mathscr{E}_N$ converge in law to the Poisson point process with intensity measure $\mathrm{e}^{-z}\mathrm{d}z$ on $\mathbb{R}$.*

2.4 The GREM, Two Levels

To understand what happens if correlation is introduced into the game, we consider the simplest version of the generalised random energy model with just two hierarchies. That is, for $\sigma^1 \in \mathscr{S}_{N/2}$ and $\sigma^2 \in \mathscr{S}_{N/2}$, we define the Hamiltonian

$$H_N(\sigma^1\sigma^2) \equiv a_1 X_{\sigma^1} + a_2 X_{\sigma^1\sigma^2}. \tag{31}$$

where all X are iid centred Gaussian with variance N, and $a_1^2 + a_2^2 = 1$.

Note that this corresponds to the Gaussian process with covariance

$$\mathbb{E}H_N(\sigma)H_N(\sigma') = NA(q_N(\sigma,\sigma')), \tag{32}$$

where

$$A(x) = \begin{cases} 0, & \text{if } x < 1/2 \\ a_1^2, & \text{if } 1/2 \le x < 1, \\ 1, & \text{if } x = 1. \end{cases} \tag{33}$$

2.4.1 Second Moment Method

We may be tempted to retry the second moment method that worked so nicely in the REM. Of course we get

$$\mathbb{E}M_N(x) = 2^N \mathbb{P}(H_N(\sigma) > u_N(x)), \tag{34}$$

as in the REM. But this quantity does not see any correlation, so this should make us suspicious. But we know how to check whether this is significant: compute the

second moment. This is easily done:

$$\begin{aligned}\mathbb{E}\left[M_N(x)^2\right] &= \sum_{\sigma^1,\sigma^2,\tau^1,\tau^2} \mathbb{E}\left[\mathbb{1}_{H_N(\sigma^1\sigma^2)>u_N(x)}\mathbb{1}_{H_N(\tau^1\tau^2)>u_N(x)}\right] \\ &= \sum_{\sigma^1,\sigma^2} \mathbb{E}\left[\mathbb{1}_{H_N(\sigma^1\sigma^2)>u_N(x)}\right] \\ &\quad + \sum_{\sigma^1}\sum_{\sigma^2\neq\tau^2} \mathbb{E}\left[\mathbb{1}_{H_N(\sigma^1\sigma^2)>u_N(x)}\mathbb{1}_{H_N(\sigma^1\tau^2)>u_N(x)}\right] \\ &\quad + \sum_{\sigma^1\neq\tau^1}\sum_{\sigma^2,\tau^2} \mathbb{E}\left[\mathbb{1}_{H_N(\sigma^1\sigma^2)>u_N(x)}\right]\mathbb{E}\left[\mathbb{1}_{H_N(\tau^1\tau^2)>u_N(x)}\right].\end{aligned} \tag{35}$$

The first terms yield $2^N\mathbb{P}(H_N(\sigma) > u_N(x))$, the last give $(2^N-1)^2\mathbb{P}(H_N(\sigma) > u_N(x))^2$, so we already know that these are ok, i.e. of order one. But the middle terms are different. In fact, a straightforward Gaussian computation shows that

$$\mathbb{E}\left[\mathbb{1}_{H_N(\sigma^1\sigma^2)>u_N(x)}\mathbb{1}_{H_N(\sigma^1\tau^2)>u_N(x)}\right] \sim \exp\left(-\frac{u_N(x)^2}{N(1+a_1^2)}\right). \tag{36}$$

Thus, with $u_N(x)$ as in the REM, the middle term gives a contribution of order

$$2^{3N/2}2^{-2N/(1+a_1^2)}. \tag{37}$$

This does not explode only if $a_1^2 \leq \frac{1}{3}$. What is going on here?

To see this, look in detail into the computations:

$$\mathbb{P}\left(H_N(\sigma^1\sigma^2) \sim u \wedge H_N(\sigma^1\tau^2) \sim u\right) \sim \int \exp\left(-\frac{x^2}{2a_1^2N} - \frac{(u-x)^2}{(1-a_1^2)N}\right)\mathrm{d}x. \tag{38}$$

Now this integral, for large u gets its main contribution from values of x that maximise the exponent, which is easily checked to be reached at $x_c = \frac{2ua_1^2}{1+a_1^2}$. For $u = u_N(x)$, this yields indeed

$$\mathbb{P}\left(H_N(\sigma^1\sigma^2) \sim u_N(x) \wedge H_N(\sigma^1\tau^2) \sim u_N(x)\right) \sim \exp\left(-\frac{u_N(x)^2}{N(1+a_1^2)}\right). \tag{39}$$

However, in this case

$$x_c \sim N\frac{2a_1^2\sqrt{2\ln 2}}{1+a_1^2} = \frac{2\sqrt{2}a_1}{1+a_1^2}a_1N\sqrt{\ln 2} \tag{40}$$

Now $a_1 N\sqrt{\ln 2}$ is the maximum that any of the $2^{N/2}$ variables $a_1 X_{\sigma^1}$ can reach, and the value x_c is larger than that as soon as $a_1 > \sqrt{2}-1$. This indicates that this moment computation is non-sensical above that value of a_1. On the other hand, when we compute $\mathbb{P}(H_N(\sigma^1\sigma^2) \sim u_N)$ in the same way, we find that the corresponding critical value of the first variable is $x_c = u_N a_1^2 = a_1\sqrt{2}a_1 N\sqrt{\ln 2} = a_1\sqrt{2}\max_{\sigma_1}(a_1 X_{\sigma_1})$. Thus, here a problem occurs only for $a_1 > 1/\sqrt{2}$. In that latter case we must expect a change in the value of the maximum, but for smaller values of a_1 we should just be more clever.

2.4.2 Truncated Second Moments

We see that the problem comes with the excessively large values of the first component contributing to the second moments. A natural idea is to cut these off. Thus we introduce

$$\hat{M}_N(x) \equiv \sum_{\sigma^1,\sigma^2} \mathbb{1}_{H_N(\sigma^1\sigma^2)>u_N(x)} \mathbb{1}_{a_1 X_{\sigma^1} \le a_1 N\sqrt{\ln 2}}. \tag{41}$$

A simple computation shows that

$$\mathbb{E}\hat{M}_N(x) = \mathbb{E}M_N(x)(1+o(1)), \tag{42}$$

as long as $a_1^2 < 1/2$. On the other hand, when we now compute $\mathbb{E}\left[\hat{M}_N(x)^2\right]$, the previously annoying term becomes

$$2^{3N/2}\mathbb{E}\left[\mathbb{1}_{H_N(\sigma^1\sigma^2)>u_N(x)}\mathbb{1}_{H_N(\sigma^1\tau^2)>u_N(x)}\mathbb{1}_{X_{\sigma^1}\le N\sqrt{\ln 2}}\right] \sim 2^{-N\frac{(1-\sqrt{2}a_1)^2}{1-a_1^2}}, \tag{43}$$

which is insignificant for $a_1^2 < 1/2$. Since $M_N(x) \ge \hat{M}_N(x)$, it follows also that

$$\mathbb{P}\left(M_N(x) \ge 1\right) \ge \frac{\left(\mathbb{E}\hat{M}_N(x)\right)^2}{\mathbb{E}\left[\hat{M}_N(x)^2\right]}. \tag{44}$$

So the result can be used to show that as long as $a_1^2 < 1/2$ to show that the maximum is of the same order as in the REM. If $a_1^2 = 1/2$, the bound on (43) is order one. Thus we still get the same behaviour for the maximum. If a_1 is even larger, then the behaviour of the maximum changes. Namely, one cannot reach the value $\sqrt{2\ln 2}N$ any more, and the maximum will be achieved by adding the maxima of the first hierarchy to the maximum in one of the branches of the second hierarchy. This yield for the leading order $(a_1+a_2)\sqrt{2}N$.

2.4.3 The Finer Computations

As in the REM, we do of course want to compute things more precisely. This means we want to compute the limit of

- $\mathbb{P}\left(\max_{(\sigma^1,\sigma^2)} H_N(\sigma^1\sigma^2) \le u_N(x)\right)$, or, better,
- compute the limit of the Laplace functionals of the extremal process,

$$\mathbb{E}\exp\left(-\sum_{\sigma^1,\sigma^2\in\mathscr{S}_{N/2}} \phi\left(u_N^{-1}(H_N(\sigma^1\sigma^2))\right)\right). \tag{45}$$

Technically, there is rather little difference in the two cases, Let us for simplicity compute the Laplace functional for the case $\phi(x) = \lambda\mathbb{1}_{(u_N(x),\infty)}$. Then

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(-\lambda\sum_{\sigma^1,\sigma^2\in\mathscr{S}_{N/2}} \mathbb{1}_{H_N(\sigma^1\sigma^2)>u_N(x)}\right)\right] \qquad (46)\\
&= \prod_{\sigma^1\in\mathscr{S}_{N/2}} \mathbb{E}\left[\exp\left(-\lambda\sum_{\sigma^2\in\mathscr{S}_{N/2}} \mathbb{1}_{H_N(\sigma^1\sigma^2)>u_N(x)}\right)\right]\\
&= \left[\frac{1}{\sqrt{2\pi N}}\int \mathrm{e}^{-\frac{t^2}{2N}}\left(\mathbb{E}\left[\mathrm{e}^{-\lambda\mathbb{1}_{a_2X>u_N(x)-a_1t}}\right]\right)^{2^{n/2}}\mathrm{d}t\right]^{2^{N/2}}\\
&= \left[\frac{1}{\sqrt{2\pi N}}\int \mathrm{e}^{-\frac{t^2}{2N}}\left(1+(\mathrm{e}^{-\lambda}-1)\mathbb{P}(a_2X>u_N(x)-a_1t)\right)^{2^{N/2}}\mathrm{d}t\right]^{2^{N/2}}.
\end{aligned}$$

Here X is a centred Gaussian random variable with variance N.

Basically, to get something nontrivial, the probability in the last expression should be of order 2^{-N}. We define t_c by

$$\mathbb{P}(a_2X > u_N(x) - a_1t_c) = 2^{-N}. \tag{47}$$

By standard Gaussian asymptotics, this yields

$$t_c = \frac{u_N(x)}{a_1} - \frac{a_2N\sqrt{\ln 2}}{a_1}. \tag{48}$$

Let us anticipate that the contribution for $t > t_c$ will be negligible. Then we get that the integral in (46) is essentially

$$\frac{1}{\sqrt{2\pi N}}\int_{-\infty}^{t_c} \mathrm{e}^{-\frac{t^2}{2N}}\left(1+2^{N/2}(\mathrm{e}^{-\lambda}-1)\mathbb{P}(a_2X>u_N(x)-a_1t)\right)\mathrm{d}t. \tag{49}$$

Inserting the standard Gaussian asymptotics and reorganising things, this is asymptotically equal to

$$1+\left(\mathrm{e}^{-\lambda}-1\right)2^{N/2}\frac{\mathrm{e}^{-\frac{u_N(x)^2}{2N}}}{\sqrt{2\pi}u_N(x)/\sqrt{N}}\int_{-\infty}^{t_c-u_N(x)\alpha_1}\frac{\mathrm{e}^{-\frac{s^2}{2a_2^2N}}}{\sqrt{2\pi N}a_2}\mathrm{d}s \tag{50}$$

$$\sim 1+2^{-N/2}\mathrm{e}^{-x}\left(\mathrm{e}^{-\lambda}-1\right)\int_{-\infty}^{t_c-u_N(x)\alpha_1}\frac{\mathrm{e}^{-\frac{s^2}{2a_2^2N}}}{\sqrt{2\pi N}a_2}\mathrm{d}s.$$

The integral is essentially one, if $t_c-u_N(x)a_1>0$, which is the case when $a_1^2<1/2$. In the case $a_1^2=1/2$, the integral is $1/2$. Inserting this fact, we get in both cases

$$\lim_{N\uparrow\infty}\mathbb{E}\exp\left(-\lambda\sum_{\sigma^1,\sigma^2\in\mathscr{S}_{N/2}}\mathbb{1}_{H_N(\sigma^1\sigma^2)>u_N(x)}\right) \tag{51}$$

$$=\exp\left(\mathrm{e}^{-x}\left(\mathrm{e}^{-\lambda}-1\right)K\right),$$

which is the Laplace functional for a Poisson process with intensity $K\mathrm{e}^{-x}\mathrm{d}x$, K being 1 or $1/2$, depending on whether $a_1^2<1/2$ or $a_1^2=1/2$, respectively.

In [15] the full picture is explored when the number of levels is arbitrary (but finite). The general result is that the extremal process remains Poisson with intensity measure $\mathrm{e}^{-x}\mathrm{d}x$ whenever the function A is strictly below the straight line $A(x)=x$, and it is Poisson with intensity $K\mathrm{e}^{-x}\mathrm{d}x$ when $A(x)$ touches the straight line finitely many times. The value of the constant $K<1$ can be expressed in terms of the probability that a Brownian bridge stays below 0 in the points where $A(x)=x$.

If $a_1^2>1/2$, the entire picture changes. In that case, the maximal values of the process are achieved by adding up the maxima of the two hierarchies, and this leads to a lower value even on the leading scale N. The extremal process also changes, but it is simply a concatenation of Poisson processes. This has all been fully explored in [15].

2.4.4 The General Model and Relation to Branching Brownian Motion

We have seen that the general case corresponds to a Gaussian process indexed by $\mathscr{S}_N$ with covariance $\mathbb{E}X_sX_{\sigma'}=NA(q_N(\sigma,\sigma'))$, for a non-decreasing function A. Now we are invited to think of $\mathscr{S}_N$ as the leaves of a tree with binary branching and N levels. Then it makes sense to extend the Gaussian process from the leaves of the tree to the entire tree. If we think of the edges of the tree of being of length one, this process should be indexed by $t\in[0,N]$, with covariance

$$\mathbb{E}\left[X_\sigma(t)X_{\sigma'}(s)\right]\equiv NA((t\wedge s)/N\wedge q_N(\sigma,\sigma')). \tag{52}$$

These models were introduced by Derrida and Gardner [30, 31] and further investigated in [16]. In the case $A(x) = x$, one sees easily that this can be realised as follows. Start at time 0 with a Brownian motion that splits into two independent Brownian motions at time $t = 1$. Each of these splits again into two independent copies at time $t = 2$, and so on. The case A non-decreasing just corresponds to a time change of this process. The case studied above, when A is a step function, corresponds to singular time changes where for some time the Brownian motions stop to move and just split and then recover their appropriate size instantaneously.

This observation strongly links the GREM to the process we will study next, branching Brownian motion.

3 Branching Brownian Motion

In this section we start to look at branching Brownian motion as a continuous time version of the GREM. We collect some basic facts that will be needed later.

3.1 Definition and Basics

The simplest way to describe branching Brownian motion is as follows. At time zero, a single particle $x_1(0)$ starting at the origin, say, begins to perform Brownian motion in $\mathbb{R}$. After an exponential time, τ, of parameter one, the particle splits into two identical, independent copies of itself that start Brownian motion at $x_1(\tau)$. This process is repeated ad infinitum, producing a collection of $n(t)$ particles $x_k(t), 1 \leq k \leq n(t)$.

Remark 3.1 In these notes we consider for notational simplicity only the case of binary branching described above. Everything extends to the case of more general offspring distributions where particles split in $k \geq 1$ particles with probabilities p_k such that $\sum_{k\geq 1} p_k = 1$, $\sum_{k\geq 2} kp_k = 2$ and $\sum_{k\geq 1} k(k-1)p_k < \infty$.

Another way to see this process is to start with a continuous time Galton-Watson (resp. Yule, in the binary splitting case) process. Here particles are just numbered but are not immersed into space. At time t, there are $n(t)$ particles $p_k(t), k \leq n(t)$. The point is that the collection of particles is endowed with a genealogical structure. Namely for each pair of particles at time t, $p_k(t), p_\ell(t)$, there is a unique time $0 \leq s \leq t$ when their most recent common ancestor split into two. We call this time $d(p_k(t), p_\ell(t))$.

Branching Brownian motion (at time t) can now be interpreted as a Gaussian process $x_k(t)$, indexed by the leaves of a GW-process, such that $\mathbb{E}x_k(t) = 0$ and, given a realisation of the Galton Watson process,

$$\mathbb{E}\left[x_k(t)x_\ell(t)\right] = d(p_k(t), p_\ell(t)). \tag{53}$$

This observation makes BBM perfectly analogous to the CREM [16] with covariance function $A(x) = x$, which we have seen to be the critical covariance.

To get into the mood of this process, let us show that

Lemma 3.2

$$\mathbb{E}n(t) = \mathrm{e}^t. \tag{54}$$

Proof The idea is to use the recursive structure of the process. Clearly, if $t < \tau$, then $n(t) = 1$, Otherwise, it is $n'(t-\tau) + n''(t-\tau)$, where n' and n'' are the number of particles in the two independent offspring processes. This reasoning leads to

$$\begin{aligned}\mathbb{E}n(t) &= \mathbb{P}(\tau > t) + \int_0^t \mathbb{P}(\tau \in \mathrm{d}s) 2\mathbb{E}n(t-s) \\ &= \mathrm{e}^{-t} + 2\int_0^t \mathrm{e}^{-s}\mathbb{E}n(t-s)\mathrm{d}s.\end{aligned} \tag{55}$$

Differentiating this equation yields

$$\frac{d}{dt}\mathbb{E}n(t) = -\mathrm{e}^{-t} + 2\mathbb{E}n(t) - 2\int_0^t \mathrm{e}^{-s}\mathbb{E}n(t-s)\mathrm{d}s = \mathbb{E}n(t), \tag{56}$$

where we used integration by parts. The assertion follows by solving this differential equation with $\mathbb{E}n(0) = 1$. □

We can also show the following classical result (see the standard textbook by Athreya and Ney [8] for this and many further results on branching processes):

Lemma 3.3 *If $n(t)$ is the number of particles of BBM at time t, then*

$$M(t) \equiv \mathrm{e}^{-t}n(t) \tag{57}$$

is a martingale. Moreover, $M(t)$ converges, a.s. and in L^1, to an exponential random variable of parameter 1.

Proof The verification that $M(t)$ is a martingale is elementary. Since $M(t)$ is positive with mean one, it is bounded in L^1 and hence, by Doob's martingale convergence theorem, $M(t)$ converges a.s. to a random variable M. To show that the martingale is uniformly integrable and hence converges in L^1, we show that

$$\phi(t) \equiv \mathbb{E}\left[M(t)^2\right] = 2 - \mathrm{e}^{-t}. \tag{58}$$

This can be done by noting that $\mathbb{E}M(t)^2$ satisfies the recursion

$$\phi(t) = \mathrm{e}^{-3t} + 2\int_0^t \mathrm{e}^{-3s}\phi(t-s)\mathrm{d}s + \frac{2}{3}\left(1 - \mathrm{e}^{-3t}\right). \tag{59}$$

Differentiating yields the differential equation

$$\phi'(t) = 2 - \phi(t), \tag{60}$$

of which (58) is the unique solution with $\phi(0) = 1$. Convergence to the exponential distribution can be proven similarly. □

3.2 The F-KPP Equation

The fundamental link between BBM and the F-KPP equation [29, 39] is generally attributed to McKean [44], but I learned from an anonymous referee that it had first been discovered by Ikeda, Nagasawa, and Watanabe [34–36].

Lemma 3.4 ([44]) *Let $f : \mathbb{R} \to [0, 1]$ and $\{x_k(t) : k \le n(t)\}$ a branching Brownian motion starting at 0. The function*

$$u(t, x) = \mathbb{E}\left[\prod_{k=1}^{n(t)} f(x - x_k(t))\right] \tag{61}$$

is the solution of the F-KPP equation (62)

$$\partial_t u = \frac{1}{2}\partial_x^2 u + u^2 - u, \tag{62}$$

with initial condition $u(0, x) = f(x)$.

Proof The derivation of the F-KPP equation is quite similar to the arguments used in the previous section. Let $f : \mathbb{R} \to [0, 1]$. Define $u(t, x)$ by (61). Then, distinguishing the cases when the first branching occurs before or after t, we get

$$\begin{aligned} u(t, x) = \mathrm{e}^{-t} \int \frac{\mathrm{e}^{-\frac{z^2}{2t}}}{\sqrt{2\pi t}} f(x - z)\mathrm{d}z \\ + \int_0^t \mathrm{e}^{-s} \int \frac{\mathrm{e}^{-\frac{z^2}{2s}}}{\sqrt{2\pi s}} u(t - s, x - z)^2 \mathrm{d}z\, \mathrm{d}s. \end{aligned} \tag{63}$$

Differentiating with respect to t, using integration by parts as before together with the fact that the heat kernel satisfies the heat equation, we find that

$$\partial_t u = \frac{1}{2}\partial_x^2 u + u^2 - u. \tag{64}$$

Obviously, $u(0, x) = f(x)$. □

The first example is obtained by choosing $f(x) = \mathbb{1}_{x\geq 0}$. Then

$$u(t,x) = \mathbb{E}\left[\prod_{k=1}^{n(t)} \mathbb{1}_{(x-x_k(t)\geq 0)}\right] = \mathbb{P}\left(\max_{k\leq n(t)} x_k(t) \leq x\right). \tag{65}$$

A second example we will exploit is obtained by choosing $f(x) = \exp(-\phi(x))$ for ϕ a non-negative continuous function with compact support. Then

$$u(t,x) = \mathbb{E}\left[\prod_{k=1}^{n(t)} \mathrm{e}^{-\phi(x-x_k(t))}\right] = \mathbb{E}\left[\exp\left(-\int \phi(x-z)\mathscr{E}_t(\mathrm{d}z)\right)\right], \tag{66}$$

where $\mathscr{E}_t \equiv \sum_{k=1}^{n(t)} \delta_{x_k(t)}$ is the point process associated to BBM at time t. We see that in this case u is the *Laplace functional* of the point process $\mathscr{E}_t$.

3.3 The Travelling Wave

Bramson [19] studied convergence of solutions to a large class of KPP equations

$$\partial_t u = \frac{1}{2}\partial_x^2 u + F(u), \tag{67}$$

for a large class of functions F that include the those that arise from BBM. In particular, he established under what conditions on the initial data solutions converge to travelling waves.

The following theorem slightly specialises Bramson's Theorems A and B from [19] to the cases we need here. Bramson states the result actually not in the form that applies to the equation in the form (62), but for the equation

$$\partial_t u = \frac{1}{2}\partial_x^2 u + u - u^2. \tag{68}$$

Clearly, of u solves (62) with initial condition $u(0,x) = f(x)$, then $v = 1-u$ solves Eq. (68) with initial condition $v(0,x) = 1-f(x)$. I am afraid that there will be many switches between the two equations and that this is a constant source of confusion.

Lemma 3.5 ([39]) *The equation*

$$\frac{1}{2}\omega'' + \sqrt{2}\omega' - \omega^2 + \omega = 0. \tag{69}$$

has a unique solution satisfying $0 < \omega(x) < 1$, $\omega(x) \to 0$, *as* $x \to +\infty$, *and* $\omega(x) \to 1$, *as* $x \to -\infty$, *up to translation, i.e. if* ω, ω' *are two solutions, then there exists* $a \in \mathbb{R}$ *s.t.* $\omega'(x) = \omega(x+a)$.

Theorem 3.6 ([19]) *Let u be solution of the F-KPP equation (68) with $0 \leq u(0,x) \leq 1$. Then there is a function $m(t)$ such that*

$$u(t, x + m(t)) \to \omega(x), \tag{70}$$

uniformly in x, where ω is one of the solutions of (69) from Lemma 3.5, as $t \uparrow \infty$, if and only if

(i) for some $h > 0$, $\limsup_{t\to\infty} \frac{1}{t} \ln \int_t^{t(1+h)} u(0,y)\mathrm{d}y \leq -\sqrt{2}$, and
(ii) for some $\nu > 0$, $M > 0$, $N > 0$, $\int_x^{x+N} u(0,y)\mathrm{d}y > \nu$ for all $x \leq -M$.

Moreover, if $\lim_{x\to\infty} \mathrm{e}^{bx} u(0,x) = 0$, for some $b > \sqrt{2}$, then one may choose

$$m(t) = \sqrt{2}t - \frac{3}{2\sqrt{2}} \ln t. \tag{71}$$

Of course, we can get an equivalent statement for solutions to (62) by simple substitution $u(0,x) \to 1 - u(0,x)$. The equation for the limit is then

$$\frac{1}{2}\omega'' + \sqrt{2}\omega' + \omega^2 - \omega = 0, \tag{72}$$

Theorem 3.6 is one of the core results of Bramson's work on Brownian motion and most of the material in his monograph [19] is essentially needed for its proof. Naturally, we cannot reproduce the proof here, but we limit ourselves to a few comments.

First, this result implies that

$$\mathbb{P}\left(\max_{k\leq n(t)} x_k(t) - m(t) \leq x\right) \to \omega(x). \tag{73}$$

It should be noted that it follows already from the results of Kolmogorov et al. [39] that (73) must hold for some function $m(t)$ with $m(t)/t \to \sqrt{2}$. But only Bramson's precise evaluation of $m(t)$ shows that the shift is chafed from the iid case where it would have to be

$$m'(t) = \sqrt{2}t - \frac{1}{2\sqrt{2}} \ln t, \tag{74}$$

which is larger than $m(t)$. Moreover, $\omega(x)$ would have to be the Gumbel distribution, which does not solve (69).

The rough heuristics of the correct rescaling can be obtained from second moment calculations. From the experience with the GREM, we would expect that the entire ancestral path of BBM will stay below the straight line with slope $\sqrt{2}$,

with probability essentially one. More precisely, we should consider the events

$$\left\{\exists_{k\le n(t)}\,\{x_k(t)-m(t)>x\}\wedge\left\{x_k(s)\le\sqrt{2}s\right\},\forall s\in(r,t-r)\right\}, \tag{75}$$

for r fixed. Here we write, for given t, and $s\le t$, $x_k(s)$ for the ancestor at time s of the particle $x_k(t)$. One of the crucial observations of Bramson [18] is that the event in (75) has the same probability as the event

$$\left\{\exists_{k\le n(t)}x_k(t)-m(t)>x\right\}, \tag{76}$$

when first $t\uparrow\infty$ and then $r\uparrow\infty$. This is not very easy, but also not too hard.

It is clear that the ancestral path from $x_k(t)$ to zero is just a Brownian bridge from $x_k(t)$ to zero. That is, we can write

$$x_k(t-s)=\frac{t-s}{t}x_k(t)+\mathfrak{z}_k(s), \tag{77}$$

where $\mathfrak{z}_k(s), s\in[0,t]$, is a Brownian bridge form 0 to 0, independent of $x_k(t)$. The event in (75) can then be expressed as

$$\left\{\exists_{k\le n(t)}\,\{x_k(t)-m(t)>x\}\wedge\left\{\mathfrak{z}_k(s)\le(1-s/t)(\sqrt{2}t-x_k(t))\right\},\forall s\in(r,t-r)\right\}, \tag{78}$$

where of course the Brownian bridges are not independent. The point, however, is, that as far as the maximum is concerned, they might as well be independent.

We need to bound the probability of a Brownian bridge to stay below a straight line. The following lemma is taken from Bramson [18], but is of course classic.

Lemma 3.7 *Let $y_1,y_2>0$. Let $\mathfrak{z}$ be a Brownian bridge from 0 to 0 in time t. Then*

$$\mathbb{P}\left(\mathfrak{z}(s)\le\tfrac{s}{t}y_1+\tfrac{t-s}{t}y_2,\forall_{0\le s\le t}\right)=1-\exp(-y_1y_2/t). \tag{79}$$

Proof The proof is a nice exercise using the reflection principle. First we can rotate everything to see that the probability in question is the same as the probability that Brownian motion, conditioned to end in y_2-y_1, stays below y_2,

$$\mathbb{P}\left(\mathfrak{z}(s)\le\tfrac{s}{t}y_1+\tfrac{t-s}{t}y_2,\forall_{0\le s\le t}\right)=\mathbb{P}\left(B_s\le y_2,\forall_{0\le s\le t}\big|\sigma(B_t)\right)(y_2-y_1). \tag{80}$$

Now use the reflection principle. All paths that do not stay below y_2 have a reflected counterpart that ends in y_2+y_1 in instead in y_2-y_1. Expressing the conditional probability in terms of heat kernels $p(t,x)$, we get

$$\mathbb{P}\left(B_s\le y_2,\forall_{0\le s\le t}\big|\sigma(B_t)\right)(y_2-y_1)=1-\frac{p(t,y_2+y_1)}{p(t,y_2-y_1)}=1-\exp(-y_1y_2/t), \tag{81}$$

and the proof is completed. □

The first moment estimate shows (the calculations are a bit intricate and constitute a good part of Bramsons first paper [18])

$$\mathbb{P}\left(\exists_{k\le n(t)}\{x_k(t)-m(t)>x\}\wedge\left\{\mathfrak{z}_k(s)\le(1-s/t)(\sqrt{2}t-x_k(t))\right\},\forall s\in(r,t-r)\right)$$
$$\sim \mathrm{e}^t\frac{\mathrm{e}^{-\frac{(m(t)+x)^2}{2t}}}{\sqrt{2\pi}m(t)/\sqrt{t}}\frac{Cx^2}{t}\sim C'\mathrm{e}^t\frac{\mathrm{e}^{-\frac{m(t)^2}{2t}}}{t^{3/2}}x^2\mathrm{e}^{-\sqrt{2}x}. \tag{82}$$

The t-dependent term is equal to one if

$$m(t)=\sqrt{2}t-\frac{3}{2\sqrt{2}}\ln t. \tag{83}$$

To justify that this upper bound is not bad, one has to simply perform a second moment computation in which one retains the condition that at the splitting point, the particle is not above the straight line $\sqrt{2}s$. This is not so easy here and actually yields a lower bound there the x^2 is not present. Indeed, it also turns out that the upper bound is not optimal, and that the x^2 should be replaced by x. We will show that in the next section. A refinement of the truncated second moment method was used more recently by Roberts [49].

One can actually localise the ancestral paths of extremal particles much more precisely, as was shown in [3]. Namely, define

$$f_{t,\gamma}(s)\equiv\begin{cases}s^\gamma, & 0\le s\le t/2\\ (t-s)^\gamma, & t/2\le s\le t.\end{cases} \tag{84}$$

Then the following holds:

Theorem 3.8 ([3]) *Let $D=[d_l,d_u]\in\mathbb{R}$ be a bounded interval. Then for any $0<\alpha<\frac{1}{2}<\beta$, and any $\epsilon>0$, there are $r_0\equiv r_0(D,\epsilon,\alpha,\beta)$, such that for all $r>r_0$ and $t>3r$,*

$$\mathbb{P}\Big(\forall_{k\le n(t)}\text{such that } x_k(t)-m(t)\in D,\text{ it holds that} \tag{85}$$
$$\forall_{s\in[r,t-r]},\ \frac{s}{t}x_k(t)+f_{t,\alpha}(s)\le x_k(s)\le\frac{s}{t}x_k(t)+f_{t,\beta}(s)\Big)\ge 1-\epsilon.$$

Basically this theorem says that the paths of maximal particles lie below the straight line to their endpoint along the arc $f_{t,1/2}(s)$. This is just due to a property of Brownian bridges: in principle, a Brownian bridge wants to oscillate (a little bit more) than $\sqrt{t}$ at its middle. But if it is not allowed to cross zero, than it also will not get close to zero, because each time it does so, it is unlikely to not do the crossing. This phenomenon is known as *entropic repulsion* in statistical mechanics.

Note also that for the ancestral paths of particles that end up much lower than $\sqrt{2}t$, say around at with $a < \sqrt{2}$, there is no such effect present. The ancestral paths of these particles are just Brownian bridges from zero to their end position which are localised around the straight line by a positive upper and a negative lower envelope $f_{t,\gamma}$, with any $\gamma > \frac{1}{2}$.

3.4 The Derivative Martingale

While Bramson analyses the asymptotic behaviour of the functions $\omega(x)$, he does not give an explicit form of these solutions, and in particular he does not provide a link to the Gumbel distribution from classical extreme value theory. This was achieved a few years later by Lalley and Sellke. They wrote a short and very insightful paper [40] that provides, after Bramson's work, the most significant insights into BBM that we have. It's main achievement is to give a probabilistic interpretation for the limiting distribution of the maximum of BBM. But this is not all.

For $1 - f$ satisfying the hypothesis of Theorem 3.6 let u be given by (61). Now define

$$\hat{u}(t,x) \equiv \mathbb{E}\left[\prod_{i=1}^{n(t)} f(\sqrt{2}t + x - x_i(t))\right] = u(t, \sqrt{2}t + x). \tag{86}$$

One checks that $\hat{u}$ solves the equation [note that here we are in the realm of the form (62)]

$$\partial_t \hat{u} = \frac{1}{2}\partial_x^2 \hat{u} + \sqrt{2}\partial_x \hat{u} - \hat{u}(1 - \hat{u}). \tag{87}$$

Let ω solve Eq. (72). Then $\hat{u}(t,x) \equiv \omega(x)$ is a stationary solution to (87) with initial condition $\hat{u}(0,x) = \omega(x)$. Therefore, we have the stochastic representation

$$\omega(x) = \mathbb{E}\left[\prod_{i=1}^{n(t)} \omega(\sqrt{2}t + x - x_i(t))\right]. \tag{88}$$

Now probability enters the game.

Lemma 3.9 *For every $x \in \mathbb{R}$, the function $W(t,x) \equiv \prod_{i=1}^{n(t)} \omega(\sqrt{2}t + x - x_i(t))$ is a martingale with respect to the natural filtration $\mathfrak{F}_t$ of BBM. Moreover, $W(t,x)$ converges almost surely and in L^1 to a non-trivial limit, $W^*(x)$. Moreover,*

$$\mathbb{E}W^*(x) = \omega(x). \tag{89}$$

Proof The proof is straightforward. It helps to introduce the notation $x_k^y(t)$ for BBM started in y. Clearly $W(t,x)$ is integrable. By the Markovian nature of BBM,

$$W(t+s) = \prod_{i=1}^{n(t)} \prod_{j=1}^{n_i(s)} \omega\left(\sqrt{2}(t+s) - x_i^{-x}(t) - \left(x_{i,j}^0(s)\right)\right). \tag{90}$$

where the $x_{i,j}$ are independent BBM's. Now

$$\begin{aligned}
\mathbb{E}\left[W(t+s,x)\middle|\mathfrak{F}_t\right] &= \prod_{i=1}^{n(t)} \mathbb{E}\left[\prod_{j=1}^{n_i(s)} \omega\left(\sqrt{2}(t+s) - x_i^{-x}(t) - \left(x_{i,j}^0(s)\right)\right)\middle|\mathfrak{F}_t\right] \qquad (91)\\
&= \prod_{i=1}^{n(t)} \mathbb{E}\left[\prod_{j=1}^{n_i(s)} \omega\left(\sqrt{2}s - \left(x_{i,j}^{x_i^{-x}(t)-\sqrt{2}t}(s)\right)\right)\middle|\mathfrak{F}_t\right]\\
&= \prod_{i=1}^{n(t)} \omega\left(\sqrt{2}t - x_i^{-x}(t)\right) = W(t,x).
\end{aligned}$$

Now $W(t,x)$ is bounded and therefore converges almost surely and in L^1 to a limit, $W^*(x)$. □

There are more martingales. First, by a trivial computation,

$$Y(t) \equiv \sum_{i=1}^{n(t)} \mathrm{e}^{\sqrt{2}x_i(t)-2t} \tag{92}$$

is a martingale. Since it is positive and $\mathbb{E}Y(t) = 1$, it converges almost surely to a finite non-negative limit. But this means that the exponents in the sum must all go to minus infinity, as otherwise no convergence is possible. This means that

$$\min_{i\le n(t)} \left(\sqrt{2}t - x_i(t)\right) \uparrow +\infty, \qquad \text{a.s.} \tag{93}$$

(this argument is convincing but a bit fishy, see the remark below).

One of Bramson's results is that (we will see the proof of this in the next section, see Corollary 4.3)

$$1 - \omega(x) \sim Cx\mathrm{e}^{-\sqrt{2}x}, \quad \text{as} \quad x \uparrow \infty. \tag{94}$$

Hence

$$W(t,x) = \exp\left(\sum_{i=1}^{n(t)} \ln \omega(\sqrt{2}t + x - x_i(t))\right) \tag{95}$$
$$\sim \exp\left(-C\sum_{i=1}^{n(t)} (\sqrt{2}t + x - x_i(t))\mathrm{e}^{-\sqrt{2}(\sqrt{2}t+x-x_i(t))}\right)$$
$$= \exp\left(-Cx\mathrm{e}^{-\sqrt{2}x}Y(t) - C\mathrm{e}^{-\sqrt{2}x}Z(t)\right),$$

where

$$Z(t) \equiv \sum_{i=1}^{n(t)} \left(\sqrt{2}t - x_i(t)\right)\mathrm{e}^{-\sqrt{2}(\sqrt{2}t-x_i(t))}. \tag{96}$$

$Z(t)$ is also a martingale, called the *derivative martingale*, with $\mathbb{E}Z_t = 0$. The fact that $Z(t)$ is a martingale can be verified by explicit computation, but it will actually not be very important for us. In any case, $Z(t)$ is not even bounded in L^1 and therefore it is a priori not cleat that $Z(t)$ converges. By the observation (93), $Z(t)$ is much bigger than $Y(t)$, which implies that unless $Y(t) \to 0$, a.s., it must be true that $\liminf Z(t) = +\infty$, which is impossible since this would imply that $W(t,x) \to 0$, which we know to be false. Hence $Y(t) \to 0$, and thus $Z(t) \to Z$, a.s., where Z is finite and positive. It follows that

$$\lim_{t\uparrow\infty} W(t,x) = \exp\left(-CZ\mathrm{e}^{-\sqrt{2}x}\right). \tag{97}$$

Remark 3.10 While the argument of Lalley and Sellke for (93) may not convince everybody,[1] the following gives an alternative proof for the fact that $Y_t \to 0$. By a simple Gaussian computation,

$$\mathbb{P}\left(\exists_{k\le n(t)} : \sqrt{2}t - x_k(t) < K\right) \le \frac{\mathrm{e}^{\sqrt{2}K}}{\sqrt{4\pi t}}. \tag{98}$$

But this implies that

$$\mathbb{P}\left(\frac{Z(t)}{Y(t)} < K\right) \le ct^{-1/2}. \tag{99}$$

[1] Thanks to Marek Biskup for voicing some doubt about this claim.

Now assume that $Y_t \to a > 0$, a.s. Then, for large enough t,

$$\mathbb{P}\left(Z(t) < 2Ka\right) \leq ct^{-1/2}. \tag{100}$$

But this implies that

$$\mathbb{P}\left(\liminf_{n\uparrow\infty} Z(2^n) < 2aK\right) = 0. \tag{101}$$

and hence

$$\mathbb{P}\left(\limsup_{t\uparrow\infty} Z(t) < 2aK\right) = 0. \tag{102}$$

But this implies that

$$\liminf_{t\uparrow\infty} W(t,x) = 0, \quad \text{a.s.}, \tag{103}$$

and since we know that $W(t,x)$ converges almost surely, it must hold that the limit is zero. But this is in contradiction with the fact that the limit is a non-negative random variable with positive expectation and that convergence holds in L^1. Hence it must be the case that $Y_t \to 0$, almost surely.

Remark 3.11 One may interpret $Y(t)$ as a *partition function*. Namely, if we set

$$Z_\beta(t) \equiv \sum_{i=1}^{n(t)} \mathrm{e}^{\beta x_i(t)}, \tag{104}$$

then

$$Y(t) = \frac{Z_{\sqrt{2}}(t)}{\mathbb{E}Z_{\sqrt{2}}(t)}. \tag{105}$$

$\sqrt{2}$ has the natural interpretation of the *critical inverse temperature* for this model, which can be interpreted as the value where the "law of large numbers" starts to fail in a strong sense, namely that $Z_\beta(t)/\mathbb{E}Z_\beta(t)$ does no longer converge to a non-trivial limit. In the REM, the critical value is $\sqrt{2\ln 2}$, and it was shown in [17] that in this case, at the critical value, this ratio converges to $1/2$. For BBM, it was used and shown in [13], that

$$\frac{Z_\beta(t)}{\mathbb{E}Z_\beta(t)}. \tag{106}$$

is a uniformly integrable martingale for all values $\beta < 1$ that converges a.s. and in L^1 to a positive random variable of mean 1.

Finally, we return to the law of the maximum. We have that

$$\begin{aligned}
&\lim_{t\uparrow\infty} \mathbb{P}\left(\max_{k\le n(t)} x_k(t) - m(t) \le x\right)\\
&\quad = \lim_{t\uparrow\infty} \mathbb{P}\left(\max_{k\le n(t+s)} x_k(t+s) - m(t+s) \le x\right)\\
&\quad = \lim_{t\uparrow\infty} \mathbb{E}\left[\mathbb{P}\left(\max_{k\le n(t+s)} x_k(t+s) - m(t+s) \le x|\mathfrak{F}_s\right)\right]\\
&\quad = \lim_{t\uparrow\infty} \mathbb{E}\left[\prod_{i=1}^{n(s)} \mathbb{P}\left(\max_{k\le n_i(t)} x_{i,k}(t) - m(t+s) - x_i(s) \le x|\mathfrak{F}_s\right)\right]\\
&\quad = \lim_{t\uparrow\infty} \mathbb{E}\left[\prod_{i=1}^{n(s)} u(t, x + m(t+s) - x_i(s))\right].
\end{aligned} \tag{107}$$

where u is the solution of the F-KPP equation with Heavyside initial condition. Next we use that

$$m(t+s) - m(t) - \sqrt{2}s \to 0, \quad \text{as } t \uparrow \infty, \tag{108}$$

for fixed s. This shows that

$$\begin{aligned}
&\lim_{t\uparrow\infty} \mathbb{P}\left(\max_{k\le n(t)} x_k(t) - m(t) \le x\right)\\
&\quad = \lim_{t\uparrow\infty} \mathbb{E}\left[\prod_{i=1}^{n(s)} u(t, x + m(t) - \sqrt{2}s - x_i(s))\right]\\
&\quad = \mathbb{E}\left[\prod_{i=1}^{n(s)} \omega(x - \sqrt{2}s - x_i(s))\right]\\
&\quad = \mathbb{E}W(s, x).
\end{aligned} \tag{109}$$

As now $s \uparrow \infty$,

$$W(s, x) \to \mathrm{e}^{-CZ\mathrm{e}^{-\sqrt{2}x}}, \text{ a.s.} \tag{110}$$

which proves the main theorem of Lalley and Sellke [40]:

Theorem 3.12 ([40]) *For BBM,*

$$\lim_{t\uparrow\infty} \mathbb{P}\left(\max_{k\le n(t)} x_k(t) - m(t) \le x\right) = \mathbb{E}e^{-CZe^{-\sqrt{2}x}}. \tag{111}$$

Moreover

$$\lim_{s\uparrow\infty}\lim_{t\uparrow\infty} \mathbb{P}\left(\max_{k\le n(t)} x_k(t) - m(t) \le x|\mathfrak{F}_s\right) = e^{-CZe^{-\sqrt{2}x}}, \text{ a.s.} \tag{112}$$

Remark 3.13 Of course, the argument above shows that any solution of (69) satisfying the conditions of Lemma 3.5 has a representation of the form $1 - e^{-CZe^{-\sqrt{2}x}}$, with only different constants C.

Remark 3.14 Lalley and Sellke conjectured in [40] that

$$\lim_{T\uparrow\infty} \frac{1}{T}\int_0^T \mathbb{1}_{\max_{k\le n(t)} x_k(t)-m(t)\le x} dt = e^{-CZe^{-\sqrt{2}x}}, \text{ a.s.} \tag{113}$$

This was proven to be true in [5].

4 The Extremal Process of BBM

In this section we discuss the construction of the extremal process of branching Brownian motion following the paper [6]. This construction has two parts. The first is the proof of convergence of Laplace functionals of the point process of BBM as $t \uparrow \infty$. This will turn out to be doable using estimates on the asymptotics of solutions of the F-KPP equation that are the basis of Bramson's work. The ensuing representation for Laplace functionals is rather indirect. In the second part, one gives an explicit description of the point process that has this Laplace functional as a cluster point process.

4.1 Controlling Solutions of the F-KPP Equation

In what follows, $\{x_k(t),\ k \le n(t)\}$ will always denote a branching Brownian motion of length t started in zero. We will usually consider the F-KPP equation in the form (68).

4.1.1 Feynman-Kac Representation

Bramson's analysis of the asymptotics of solutions of the KPP equation in [19] relies on a Feynman-Kac representation. Namely, if $u(t,x)$ is the solution of the

linear equation

$$\partial_t u = \frac{1}{2}\partial_x^2 u + k(t,x)u, \tag{114}$$

with initial conditions $u(0,x)$, then the standard Feynman-Kac formula yields the representation

$$u(t,x) = \mathbb{E}_x\left(\exp\left(\int_0^t k(t-s,B_s)\mathrm{d}s\right)u(0,B_t)\right). \tag{115}$$

where the expectation is with respect to ordinary Brownian motion started at x. To get back to the full equation (68), we use this with

$$k(t,x) = (1-u(t,x)), \tag{116}$$

where u itself is the solution of the F-KPP equation. Still, the ensuing representation is very useful as it allows to process a priori information on the solution into finer estimates. Bramson's key idea here was to use the Feynman-Kac representation not all the way down to the initial condition, but to go back to some arbitrary time $0 \le r < t$. This shows that a solution of the F-KPP equation must satisfy[2]

$$u(t,x) = \mathbb{E}_x\left(\exp\left(\int_0^{t-r} k(t-s,B_s)\mathrm{d}s\right)u(r,B_{t-r})\right). \tag{117}$$

This can be conveniently be rewritten in terms of Brownian bridges. Denote for $T>0$ and $0<\alpha<\beta<\infty$, by $\{\mathfrak{z}^T_{\alpha,\beta}(s), 0\le s\le T\}$ the Brownian bridge of length T starting in α and ending in β. Then

$$u(t,x) = \int_{-\infty}^{\infty}\mathrm{d}y\, u(r,y)\frac{\mathrm{e}^{-(x-y)^2/(2(t-r))}}{\sqrt{2\pi(t-r)}}\exp\left(\int_0^{t-r} k(t-s,\mathfrak{z}^{t-r}_{x,y}(s))\mathrm{d}s\right). \tag{118}$$

Note that $k(t,x) = 1-u(t,x)$ lies between zero and one and tends to one as $x\uparrow\infty$. Bramson's basic idea to exploit this formula is to prove a priori estimates on the function $k(s,x)$. In particular, he shows that $k(s,x)\sim 1$ if x is a above a certain curve, $\overline{\mathscr{M}}$. On the other hand, he shows that the probability that the Brownian bridge descends below a curve $\underline{\mathscr{M}}$, is negligibly small. The following proposition is an application of this strategy and essentially contained in Bramson's monograph [19, Proposition 8.3].

[2]One should appreciate the beauty of Bramson's construction: start with a probabilistic model (BBM), derive a pde whose solutions represent quantities of interest, and then use a different probabilistic representation of the solution (in terms of Brownian motion) to analyse these solutions...

Proposition 4.1 *Let u be a solution to the F-KPP equation (68) with initial condition satisfying*

$$\int_0^\infty y\mathrm{e}^{\sqrt{2}y}u(0,y)\mathrm{d}y < \infty\,, \tag{119}$$

and such that $u(t,\cdot+m(t))$ converges. Define

$$\psi(r,t,x+\sqrt{2}t) \equiv \frac{\mathrm{e}^{-\sqrt{2}x}}{\sqrt{2\pi(t-r)}}\int_0^\infty u(r,y+\sqrt{2}r)\mathrm{e}^{\sqrt{2}y}\mathrm{e}^{-\frac{(y-x)^2}{2(t-r)}}(1-\mathrm{e}^{-2y\frac{(x+\frac{3}{2\sqrt{2}}\ln t)}{t-r}})\mathrm{d}y \tag{120}$$

Then for r large enough, $t \geq 8r$, and $x \geq 8r - \frac{3}{2\sqrt{2}}\ln t$,

$$\gamma^{-1}(r)\psi(r,t,x+\sqrt{2}t) \leq u(t,x+\sqrt{2}t) \leq \gamma(r)\psi(r,t,x+\sqrt{2}t), \tag{121}$$

where $\gamma(r)\downarrow 1$, as $r\to\infty$.

We see that ψ basically controls u, but of course ψ still involves u. The fact that u is bounded by u may seem strange, but we shall see that this is very useful.

Proof Bramson (see [19, Proposition 8.3]) showed that for u satisfying the assumptions in the Proposition 4.1, the following holds:

1. For r large enough, $t \geq 8r$ and $x \geq m(t) + 8r$

$$\begin{aligned} u(t,x) &\geq \psi_1(r,t,x) \qquad (122)\\ &\equiv C_1(r)\mathrm{e}^{t-r}\int_{-\infty}^\infty u(r,y)\frac{e^{-\frac{(x-y)^2}{2(t-r)}}}{\sqrt{2\pi(t-r)}}\mathbb{P}\left(\mathfrak{z}_{x,y}^{t-r}(s) > \overline{\mathscr{M}}_{r,t}^x(t-s), s\in[0,t-r]\right)\mathrm{d}y \end{aligned}$$

and

$$\begin{aligned} u(t,x) &\leq \psi_2(r,t,x) \qquad (123)\\ &\equiv C_2(r)\mathrm{e}^{t-r}\int_{-\infty}^\infty u(r,y)\frac{\mathrm{e}^{-\frac{(x-y)^2}{2(t-r)}}}{\sqrt{2\pi(t-r)}}\mathbb{P}\left(\mathfrak{z}_{x,y}^{t-r}(s) > \underline{\mathscr{M}}'_{r,t}(t-s), s\in[0,t-r]\right)\mathrm{d}y, \end{aligned}$$

where the functions $\overline{\mathscr{M}}_{r,t}^x(t-s)$, $\underline{\mathscr{M}}'_{r,t}(t-s)$ satisfy

$$\underline{\mathscr{M}}'_{r,t}(t-s) \leq n_{r,t}(t-s) \leq \overline{\mathscr{M}}_{r,t}^x(t-s). \tag{124}$$

Here

$$n_r(s) \equiv \sqrt{2}r + \frac{(s-r)}{t-r}(m(t)-\sqrt{2}r). \tag{125}$$

Moreover, $C_1(r) \uparrow 1$, $C_2(r) \downarrow 1$ as $r \uparrow \infty$.

2. The bounds $\psi_1(r,t,x)$ and $\psi_2(r,t,x)$ satisfy

$$1 \le \frac{\psi_2(r,t,x)}{\psi_1(r,t,x)} \le \gamma(r), \tag{126}$$

where $\gamma(r) \downarrow 1$ as $r \uparrow \infty$.

Hence, if we denote by

$$\psi(r,t,x) = \mathrm{e}^{t-r} \int_{-\infty}^{\infty} u(r,y) \frac{\mathrm{e}^{-\frac{(x-y)^2}{2(t-r)}}}{\sqrt{2\pi(t-r)}} \mathbb{P}\left(\mathfrak{z}_{x,y}^{t-r}(s) > n_{r,t}(t-s), s \in [0,t-r]\right) \mathrm{d}y \tag{127}$$

we have by domination $\psi_1 \le \psi \le \psi_2$. Therefore, for r, t and x large enough

$$\frac{u(t,x)}{\psi(r,t,x)} \le \frac{\psi_2(r,t,x)}{\psi(r,t,x)} \le \frac{\psi_2(r,t,x)}{\psi_1(r,t,x)} \le \gamma(r)\,, \tag{128}$$

and

$$\frac{u(t,x)}{\psi(r,t,x)} \ge \frac{1}{\gamma(r)}\,. \tag{129}$$

Combining (128) and (129) we get

$$\gamma^{-1}(r)\psi(r,t,x) \le u(t,x) \le \gamma(r)\psi(r,t,x). \tag{130}$$

For $x \ge 8r - \frac{3}{2\sqrt{2}} \ln t$ we get from (130) that

$$\gamma^{-1}(r)\psi(r,t,x+\sqrt{2}t) \le u(t,x+\sqrt{2}t) \le \gamma(r)\psi(r,t,x+\sqrt{2}t). \tag{131}$$

The nice thing is that the probability involving the Brownian bridge in the definition of ψ can be explicitly computed, see Lemma 3.7.

Since our bridge goes from $x+\sqrt{2}t$ to y and the straight line is between $\sqrt{2}r$ and $m(t)$, and it stays above this line, we have to adjust this to length $t-r$, $y_1 = \sqrt{2}t + x - m(t) = x + \frac{3}{2\sqrt{2}} \ln t > 0$ (for $t > 1$) and $y_2 = y - \sqrt{2}r$.

Of course $\mathbb{P}\left(\mathfrak{z}_{x,y}^{t-r}(s) > n_{r,t}(t-s), s \in [0,t-r]\right) = 0$ for $y \le \sqrt{2}r$. For $y > \sqrt{2}r$, from Lemma 3.7 we have

$$\mathbb{P}\left(\mathfrak{z}_{x,y}^{t-r}(s) > n_{r,t}(t-s), s \in [0,t-r]\right) = 1 - \exp\left(-\frac{\left(x + \frac{3}{2\sqrt{2}} \ln t\right)(y-\sqrt{2}r)}{t-r}\right). \tag{132}$$

Changing variables to $y' = y - \sqrt{2}r$ in the integral appearing in the definition of ψ, we get (120). This, together with the formula (130), concludes the proof of the proposition. □

4.1.2 Limit Theorems for Solutions

The bounds in (121) have been used by Chauvin and Rouault to compute the probability of deviations of the maximum of BBM, see Lemma 2 in [23]. We will use their arguments in this slightly more general setting. Our aim is to control limits when t and later r tend to infinity.

We will need to analyse solutions $u(t, x + \sqrt{2}t)$ where $x = x(t)$ depends on t in different ways. First we look at x fixed. This corresponds to particles that are just $\frac{3}{2\sqrt{2}} \ln t$ ahead of the natural front of BBM. We start with the asymptotics of ψ.

Proposition 4.2 *Let the assumptions of Proposition 4.1 be satisfied, and assume in addition that* $y_0 \equiv \sup\{y : u(0, y) > 0\}$ *is finite. Then, for any* $x \in \mathbb{R}$,

$$\lim_{t\uparrow\infty} \mathrm{e}^{x\sqrt{2}} \frac{t^{3/2}}{\frac{3}{2\sqrt{2}} \ln t} \psi(r, t, x + \sqrt{2}t) = \sqrt{\frac{2}{\pi}} \int_0^\infty y \mathrm{e}^{y\sqrt{2}} u(r, y + \sqrt{2}r) \mathrm{d}y \equiv C(r). \tag{133}$$

Moreover, $\lim_{r\uparrow\infty} C(r) \equiv C$ *exists and it is positive and finite. Finally,*

$$\lim_{t\uparrow\infty} \mathrm{e}^{x\sqrt{2}} \frac{t^{3/2}}{\frac{3}{2\sqrt{2}} \ln t} u(t, x + \sqrt{2}t) = C. \tag{134}$$

Proof Consider first the limit $t \uparrow \infty$. For fixed x, the integrand converges pointwise to $\frac{3}{2\sqrt{\pi}} y \mathrm{e}^{y\sqrt{2}} u(r, y + \sqrt{2}r)$, as can be seen by elementary calculations. To deduce convergence of the integral, we need to show that we can use dominated convergence. It is clear that the integrand in (120) after multiplication with the prefactor in (133) is bounded from above by

$$const. y \mathrm{e}^{\sqrt{2}y} u(r, y + \sqrt{2}r). \tag{135}$$

Thus we need an a priori bound on u. Bramson observed that this can be obtained in terms of a solution, $u^{(2)}$, of the *linearised* F-KPP equation,

$$\partial_t u^{(2)} = \frac{1}{2} \partial_{xx} u^{(2)} - u^{(2)} \tag{136}$$

with the same initial conditions $u^{(2)}(0,x) = u(0,x)$. Using the Feynman-Kac formula for this linear equation yields

$$u^{(2)}(t,x) = \mathbb{E}_x\left(\exp\left(\int_0^t \mathrm{d}s\right) u(0,B_t)\right) \leq \mathrm{e}^t \int_{-\infty}^{y_0} \frac{1}{\sqrt{2\pi t}} \mathrm{e}^{-\frac{(y-x)^2}{2t}} \mathrm{d}y, \tag{137}$$

where we used that $u(0,x) \leq \mathbb{1}_{x\leq y_0}$. Bramson also established [19, Corollary 1, p.29] a maximum principle for nonlinear parabolic pde's, which implies that

$$u(t,x) \leq u^{(2)}(t,x). \tag{138}$$

Using this estimate, we can bound the integrand in (120) by (ignoring constants)

$$\mathrm{e}^{-\sqrt{2}y}\mathbb{1}_{y\leq y_0-\sqrt{2}r} + \mathrm{e}^{\sqrt{2}y}\mathrm{e}^{-(y+\sqrt{2}r-y_0)^2/2r}\mathbb{1}_{y\geq y_0-\sqrt{2}r}, \tag{139}$$

which is integrable in y uniformly in t. Hence Lebesgue's dominated convergence theorem applies and the first part of the proposition follows.

The more interesting task is to show now that $C(r)$ converges to a non-trivial limit, as $r \uparrow \infty$. By Proposition 4.1, for r large enough,

$$\limsup_{t\uparrow\infty} \mathrm{e}^{x\sqrt{2}} \frac{t^{3/2}}{\frac{3}{2\sqrt{2}}\ln t} u(t, x+\sqrt{2t}) \leq \gamma(r) \lim_{t\uparrow\infty} \mathrm{e}^{x\sqrt{2}} \frac{t^{3/2}}{\ln t} \psi(r,t,x+\sqrt{2t}) = C(r)\gamma(r)\,, \tag{140}$$

and

$$\liminf_{t\uparrow\infty} \mathrm{e}^{x\sqrt{2}} \frac{t^{3/2}}{\frac{3}{2\sqrt{2}}\ln t} u(t, x+\sqrt{2t}) \geq \frac{1}{\gamma(r)} \lim_{t\uparrow\infty} \mathrm{e}^{x\sqrt{2}} \frac{t^{3/2}}{\ln t} \psi(r,t,x+\sqrt{2t}) = \frac{C(r)}{\gamma(r)}\,, \tag{141}$$

These bounds hold for all r, and since $\gamma(r) \to 1$, we can conclude that the left-hand sides is bounded from above by $\liminf_{r\uparrow\infty} C(r)$ and from below by and $\limsup_{r\uparrow\infty} C(r)$. So the lim sup $\leq$ lim inf and hence $\lim_{r\uparrow\infty} C(r) = C$ exists. As a byproduct, we see that also the left-hand sides of (140) and (141) must agree, and thus we obtain (134). A very important aspect is that the bounds (140) and (141) hold for *any* r. But for large enough finite r, the constants $C(r)$ are strictly positive and finite by the representation (133), and this shows that C is strictly positive and finite. □

We will make a slight stop on the way and show how the sharp asymptotics for the upper tail of solutions follows from what we just did.

Corollary 4.3 *Let u be as in the preceding proposition. Then*

$$\lim_{x\uparrow\infty}\lim_{t\uparrow\infty}\frac{\mathrm{e}^{x\sqrt{2}}}{x}u(t,x+m(t))=C, \tag{142}$$

for some $0<C<\infty$.

Proof Note that

$$\psi(r,t,x+m(t))=\psi\left(r,t,\sqrt{2}t+x-\tfrac{3}{2\sqrt{2}}\ln t\right). \tag{143}$$

Then, by the same arguments as in the proof of Proposition 4.2, we get that

$$\lim_{t\uparrow\infty}\frac{\mathrm{e}^{x\sqrt{2}}}{x}\psi(r,t,x+\sqrt{2t})=\sqrt{\frac{2}{\pi}}\int_0^\infty y\mathrm{e}^{y\sqrt{2}}u(r,y+\sqrt{2}r)\mathrm{d}y=C(r). \tag{144}$$

Here $C(r)$ the same constant as before. Now we can choose $x>8r$ in order to be able to apply Proposition 4.1 and then let $r\uparrow\infty$. This yields (142). □

The next lemma is a variant of the previous proposition for the case when $x\sim\sqrt{t}$.

Lemma 4.4 *Let u be a solution to the F-KPP equation (68) with initial condition satisfying the assumptions of Theorem 3.6 and*

$$y_0\equiv\sup\{y:u(0,y)>0\}<\infty\,. \tag{145}$$

Then, for $a>0$, and $y\in\mathbb{R}$,

$$\lim_{t\to\infty}\frac{\mathrm{e}^{\sqrt{2}a\sqrt{t}}t^{3/2}}{a\sqrt{t}}\psi(r,t,y+a\sqrt{t}+\sqrt{2}t)=C(r)\mathrm{e}^{-\sqrt{2}y}\mathrm{e}^{-a^2/2} \tag{146}$$

where $C(r)$ is the constant from above. Moreover, the convergence is uniform for a in compact sets.

Proof The structure of the proof is the same as in the proposition. Compute the pointwise limit and produce an integrable majorant from a bound using the linearised F-KPP equation. □

It will be very important to know that as $r\uparrow\infty$, in the integral representing $C(r)$, only the y's that are of order $\sqrt{r}$ give a non-vanishing contribution. The precise version of this statement is the following lemma.

Lemma 4.5 *Let u be a solution to the F-KPP equation (68) with initial condition satisfying the assumptions of Theorem 3.6 and*

$$y_0\equiv\sup\{y:u(0,y)>0\}<\infty\,. \tag{147}$$

Then for any $x \in \mathbb{R}$

$$\begin{aligned}\lim_{A_1\downarrow 0}\limsup_{r\uparrow\infty}\int_0^{A_1\sqrt{r}} u(r,x+y+\sqrt{2}r)\, y\mathrm{e}^{y\sqrt{2}}\mathrm{d}y = 0\\ \lim_{A_2\uparrow\infty}\limsup_{r\uparrow\infty}\int_{A_2\sqrt{r}}^{\infty} u(r,x+y+\sqrt{2}r)\, y\mathrm{e}^{y\sqrt{2}}\mathrm{d}y = 0\,.\end{aligned} \tag{148}$$

Proof Again it is clear that we need some a priori information on the behaviour of u. This time this will be obtained by comparing to a solution of the F-KPP equation with Heaviside initial conditions, which we know are probabilities for the maximum of BBM. Namely, by assumption, $u(0,y) \leq \mathbb{1}_{\{y<y_0+1\}}$. It follows from the representation of u as an expectation (61) the u is monotone in the initial conditions and thus

$$u(r,x+y+\sqrt{2}r) \leq \mathbb{P}\left(\max_{k\leq n(t)} x_k(r) - \sqrt{2}r > y_0 + 1 + x + y\right). \tag{149}$$

For the probabilities we have a sharp estimate on the tail that follows from an estimate due to Bramson [19, Proposition 8.2]:

Lemma 4.6 ([4, Cor. 10]) *There exists a numerical constants,* $\rho, t_0 < 0$*, such that for all* $x > 1$*, and all* $t \geq t_0$*,*

$$\mathbb{P}\left(\max_{k\leq n(t)} x_k(t) - m(t) \geq x\right) \leq \rho x \exp\left(-\sqrt{2}x - \frac{x^2}{2t} + \frac{3x}{2\sqrt{2}}\frac{\ln t}{t}\right). \tag{150}$$

.

Proof The proof follows from the exploitation of the Feynman-Kac formula with $r = 0$. This gives, for the special case that u is the Heaviside initial condition, a bound

$$u(t,x+m(t)) \leq \rho \mathrm{e}^t \int_{-1}^{0} \frac{\mathrm{e}^{-(x+m(t)-y)^2/2t}}{\sqrt{2\pi t}}\left(1-\mathrm{e}^{-2(y+1)x/t}\right)\mathrm{d}y. \tag{151}$$

Writing out what $m(t)$ is the result follows. □

We will use this for $x \sim \sqrt{t}$, so the x^2/t term is relevant. The last term involving $\ln t/t$ can, however, be safely ignored. The lemma follows in a rather straightforward and purely computational way once we insert this bound in integrals. The details can be found in [6]. Note that the precise bound (150) is needed and the trivial bound comparing to iid variables would not suffice. □

The following lemma shows how $C = \lim_{r\uparrow\infty}\sqrt{\frac{2}{\pi}}\int_0^\infty u(r,y+\sqrt{2}r)\, y\mathrm{e}^{y\sqrt{2}}\mathrm{d}y$ behaves when the spatial argument of u is shifted.

Lemma 4.7 *Let u be a solution to the F-KPP equation (68) with initial condition satisfying the assumptions of Theorem 3.6 and*

$$y_0 \equiv \sup\{y : u(0, y) > 0\} < \infty . \tag{152}$$

Then for any $x \in \mathbb{R}$:

$$\begin{aligned} &\lim_{r\uparrow\infty} \sqrt{\frac{2}{\pi}} \int_0^\infty y\mathrm{e}^{y\sqrt{2}} u(r, x + y + \sqrt{2}r)\mathrm{d}y \\ &\quad = \mathrm{e}^{-\sqrt{2}x} \lim_{r\uparrow\infty} \sqrt{\frac{2}{\pi}} \int_0^\infty y\mathrm{e}^{y\sqrt{2}} u(r, y + \sqrt{2}r)\mathrm{d}y = C\mathrm{e}^{-\sqrt{2}x}. \end{aligned} \tag{153}$$

Proof The proof is obvious from the fact proven in Lemma 4.5. Namely, changing variables,

$$\int_0^\infty y\mathrm{e}^{y\sqrt{2}} u(r, x+y+\sqrt{2}r)\mathrm{d}y = \mathrm{e}^{-\sqrt{2}x} \int_x^\infty (y-x)\mathrm{e}^{y\sqrt{2}} u(r, y+\sqrt{2}r)\mathrm{d}y \tag{154}$$

Now the part of the integral from x to zero cannot contribute, as it does not involve y of the order of $\sqrt{r}$, and for the same reason replacing y by $y - x$ changes nothing in the limit. □

The following is a slight generalisation of Lemma 4.7:

Lemma 4.8 *Let $h(x)$ be continuous function that is bounded and is zero for x small enough. Then*

$$\begin{aligned} &\lim_{t\uparrow\infty} \sqrt{\frac{2}{\pi}} \int_{-\infty}^0 \mathbb{E}\left[h(y + \max_i x_i(t) - \sqrt{2}t)\right] (-y)\mathrm{e}^{-\sqrt{2}y}\mathrm{d}y \\ &\quad = \int_{\mathbb{R}} h(z)\sqrt{2}C\mathrm{e}^{-\sqrt{2}z}\mathrm{d}z, \end{aligned} \tag{155}$$

where C is the constant appearing in the law of the maximum, Theorem 3.12.

Proof Consider first the case when $h(x) = \mathbb{1}_{[b,\infty)}(x)$, for $b \in \mathbb{R}$.

$$\begin{aligned} &\lim_{t\uparrow\infty} \sqrt{\frac{2}{\pi}} \int_{-\infty}^0 \mathbb{E}\left[h(y + \max_i x_i(t) - \sqrt{2}t)\right] (-y)\mathrm{e}^{-\sqrt{2}y}\mathrm{d}y \\ &\quad \lim_{t\uparrow\infty} \int_{-\infty}^0 \mathbb{P}\left(y + \max_i x_i(t) - \sqrt{2}t > b\right) \sqrt{\tfrac{2}{\pi}}(-y)\mathrm{e}^{-\sqrt{2}y}\mathrm{d}y \\ &\quad = \lim_{t\uparrow\infty} \int_0^\infty \mathbb{P}\left(\max_i x_i(t) > b + y + \sqrt{2}t\right) \sqrt{\tfrac{2}{\pi}} y\mathrm{e}^{\sqrt{2}y}\mathrm{d}y \\ &\quad = \lim_{t\uparrow\infty} \int_0^\infty u(t, b + y + \sqrt{2}t)\sqrt{\tfrac{2}{\pi}} y\mathrm{e}^{\sqrt{2}y}\mathrm{d}y, \end{aligned} \tag{156}$$

where u solves the F-KPP equation (68) with initial condition $u(0,x) = \mathbb{1}_{x<0}$. Hence the right-hand side of (156) converges, by Lemma 4.7, to

$$Ce^{-\sqrt{2}b} = C\int_{\mathbb{R}} h(z)e^{-\sqrt{2}z}\sqrt{2}dz. \tag{157}$$

To pass to the general case (155), note that by linearity, (157) holds for linear combinations of indicator functions of intervals. The general case when follows from a monotone class type argument. □

4.2 *Existence of a Limiting Process*

We now look at the construction of the limit of the point processes

$$\mathscr{E}_t \equiv \sum_{k\le n(t)} \delta_{x_k(t)-m(t)}. \tag{158}$$

The results presented here were achieved in a paper with Arguin and Kistler [6], following two preparatory papers [3, 4]. An alternative proof was given by Aïdékon, Brunet, Beresticky, and Shi [2]. A comparison of the two approaches is given in the notes by Gouéré [33], and is also discussed in the published versions of [2, 6].

4.2.1 Convergence of the Laplace Functional

To prove the existence of the limiting extremal process we show the convergence of Laplace functionals

$$\psi_t(\phi) \equiv \mathbb{E}\left[\exp\left(-\int \phi(y)\mathscr{E}_t(dy)\right)\right], \tag{159}$$

for $\phi \in \mathscr{C}_c(\mathbb{R})$ non-negative. Clearly, the Laplace functional related to a solution u of the F-KPP equation with initial conditions $u(0,x) = e^{-\phi(x)}$ via

$$\psi_t(\phi) = u(t, m(t)). \tag{160}$$

Theorem 4.9 *The point process $\mathscr{E}_t = \sum_{k\le n(t)} \delta_{x_k(t)-m(t)}$ converges in law to a point process $\mathscr{E}$.*

Proof It suffices to show that, for $\phi \in \mathscr{C}_c(\mathbb{R})$ positive, the Laplace transform $\psi_t(\phi)$, defined in (159), of the extremal process of branching Brownian motion converges. Notice first that this limit cannot be 0, since in the case of BBM it can be checked (see [3]) that for any bounded interval, $B \subset \mathbb{R}$,

$$\lim_{N\uparrow\infty} \lim_{t\uparrow\infty} \mathbb{P}\left(\mathscr{E}_t(B) > N\right) = 0, \tag{161}$$

hence the limiting point process must be locally finite. (Basically, if (161) failed then the maximum of BBM would have to be much larger than it is know to be).

For convenience, we define $\max \mathscr{E}_t \equiv \max_{k\le n(t)} x_k(t) - m(t)$. By Theorem 3.6 we know already that

$$\lim_{t\uparrow\infty} \mathbb{P}(\max \mathscr{E}_t > \delta) = 1 - \omega(\delta), \tag{162}$$

and

$$\lim_{\delta\uparrow\infty} (1 - \omega(\delta)) = 0\,. \tag{163}$$

We write the Laplace transform as a sum of two terms, corresponding to whether $\max \mathscr{E}_t \le \delta$, or $\max \mathscr{E}_t > \delta$, for some $\delta > 0$:

$$\mathbb{E}\left[\exp\left(-\int \phi \, \mathrm{d}\mathscr{E}_t\right)\right] = \mathbb{E}\left[\exp\left(-\int \phi \, \mathrm{d}\mathscr{E}_t\right) \mathbb{1}_{\{\max \mathscr{E}_t \le \delta\}}\right] \tag{164}$$

$$+\mathbb{E}\left[\exp\left(-\int \phi \, \mathrm{d}\mathscr{E}_t\right) \mathbb{1}_{\{\max \mathscr{E}_t > \delta\}}\right] \tag{165}$$

$$\equiv \psi_t^\delta(\phi) + \Phi_t^\delta(\phi).$$

Note that, by (163),

$$\limsup_{\delta\uparrow\infty} \limsup_{t\uparrow\infty} \Phi_t^\delta(\phi) \le \limsup_{\delta\uparrow\infty} \limsup_{t\uparrow\infty} \mathbb{P}(\max \mathscr{E}_t > \delta) = 0\,. \tag{166}$$

We will show that

$$\lim_{\delta\uparrow\infty} \lim_{t\uparrow\infty} \psi_t^\delta(\phi) \equiv \psi(\phi) \tag{167}$$

exists, and is strictly smaller than one. To see this, set $g_\delta(x) \equiv \mathrm{e}^{-\phi(-x)}\mathbb{1}_{\{-x\le\delta\}}$, and

$$u_\delta(t,x) \equiv \mathbb{E}\left[\prod_{k\le n(t)} g_\delta(x - x_k(t))\right]. \tag{168}$$

By Lemma 3.4, u_δ is then solution to the F- KPP equation (62) with $u_\delta(0,x) = g_\delta(x)$. Also $g_\delta(x) = 1$ for x large enough, and $g_\delta(x) = 0$ for $-x$ large enough. Therefore, $1-u_\delta$ solves the F-KPP equation in the form (68) with initial condition that satisfies the hypothesis of Bramson's theorem (3.6). Note that this would not be the case without the presence of the cutoff. Therefore

$$u_\delta(t, x + m(t)) = \mathbb{E}\left[\prod_{k=1}^{n(t)} g_\delta(x - x_k(t) + m(t))\right] \tag{169}$$

converges, as $t \uparrow \infty$, uniformly in x to a solution of (72). But

$$\begin{aligned}\psi_t^\delta(\phi) &= \mathbb{E}\left[\exp\left(-\int \phi \,\mathrm{d}\mathscr{E}_t\right)\mathbb{1}_{\{\max \mathscr{E}_t \le \delta\}}\right] \qquad (170)\\ &= \mathbb{E}\left[\prod_{k\le n(t)} \exp\left(-\phi\left(x_k(t) - m(t)\right)\right)\mathbb{1}_{\{x_k(t)-m(t)\le\delta\}}\right]\\ &= \mathbb{E}\left[\prod_{k\le n(t)} g_\delta(x_k(t) - m(t))\right] = u_\delta(t, 0 + m(t)),\end{aligned}$$

and therefore $\lim_{t\uparrow\infty}\psi_t^\delta(\phi) \equiv \psi^\delta(\phi)$ exists. But the function $\delta \mapsto \psi^\delta(\phi)$ is increasing and bounded by construction. Therefore, $\lim_{\delta\uparrow\infty}\psi^\delta(\phi) = \psi(\phi)$ exists. Moreover, there is the obvious bound (uniformly in δ)

$$\mathbb{E}\left[\exp\left(-\int \phi \,\mathrm{d}\mathscr{E}_t\right)\mathbb{1}_{\{\max \mathscr{E}_t \le \delta\}}\right] \le \mathbb{E}\left[\exp\left(-\phi(\max \mathscr{E}_t)\right)\right] < 1, \tag{171}$$

since $\max \mathscr{E}_t$ is an atom of $\mathscr{E}_t$ and ϕ is non-negative. The limit $t \uparrow \infty$ of the right side exists and is strictly smaller than 1 by the convergence in law of the re-centred maximum to $\omega(x)$. (Note that the support of $\omega(x)$ is $\mathbb{R}$). Therefore $\psi(\phi) = \lim_{\delta\uparrow\infty}\lim_{t\uparrow\infty}\psi_t^\delta(\phi) < 1$ which proves (167). □

4.3 Flash Back to the Derivative Martingale

Having established the tail asymptotics for solutions of the KPP equation, one might be tempted to compute the law of the maximum (or Laplace functionals) by recursion at an intermediate time r, e.g.

$$\begin{aligned}\mathbb{E}\left[\exp\left(-\int \phi(-x+y)\mathscr{E}_t(\mathrm{d}y)\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[\exp\left(-\int \phi(-x+y)\mathscr{E}_t(\mathrm{d}y)\right)|\mathfrak{F}_r\right]\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[\prod_{i=1}^{n(r)}\exp\left(-\int \phi(-x+x_i(r)-m(t)+m(t-r)+y)\mathscr{E}^{(i)}_{t-r}(\mathrm{d}y)\right)|\mathfrak{F}_r\right]\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[\prod_{i=1}^{n(r)}\left(1-v(t-r,-x+x_i(r)-\sqrt{2}r)\right)|\mathfrak{F}_r\right]\right], \qquad (172)\end{aligned}$$

where $\mathscr{E}^{(i)}_s$ are iid copies on BBM and

$$v(s,x-x_i(r)+\sqrt{2}r+m(s)) \equiv 1-\exp\left(-\int \phi(-x+x_i(r)-\sqrt{2}r+y)\mathscr{E}^{(i)}_s(\mathrm{d}y)\right). \qquad (173)$$

We have used that $m(t)-m(r) = \sqrt{2}r+o(1)$. We want to use the fact (see Theorem 3.8) that the particles that will show up in the support of ϕ at time t must come from particles in an interval $(\sqrt{2}r-c_1\sqrt{r}, \sqrt{2}r-c_2\sqrt{r})$, if r is large enough and $c_1 > c_2 > 0$ become large. Hence we may assume that the conditional expectations that appear in (172) can be replaced by their tail asymptotics, i.e.

$$v(t-r,x-x_i(r)+\sqrt{2}r+m(t-r)) \sim C(-x-x_i(r)+\sqrt{2}r)\mathrm{e}^{-\sqrt{2}(x-x_i(r)+\sqrt{2}r)}. \qquad (174)$$

Hence we would get,

$$\begin{aligned}&\lim_{t\uparrow\infty}\mathbb{E}\left[\exp\left(-\int \phi(x+y)\mathscr{E}_t(\mathrm{d}y)\right)\right]\\ &\quad= \lim_{r\uparrow\infty}\mathbb{E}\left[\mathrm{e}^{-C\sum_{i=1}^{n(r)}(x-x_i(r)-\sqrt{2}r)\mathrm{e}^{-\sqrt{2}(x-x_i(r)+\sqrt{2}r)}}\right]\\ &\quad= \lim_{r\uparrow\infty}\mathbb{E}\left[\mathrm{e}^{-CZ_re^{-\sqrt{2}x}-CY_r x\mathrm{e}^{-\sqrt{2}x}}\right], \qquad (175)\end{aligned}$$

with Z_t and Y_t the martingales from Sect. 3.4.

Now one could argue like Lalley and Sellke that Y_r must converge to a non-negative limit, implying that $\sqrt{2}r-x_i(r) \uparrow \infty$. Then again this implies that $Y_r \to 0$,

a.s., and hence

$$\lim_{t\uparrow\infty}\mathbb{E}\left[\exp\left(-\int\phi(x+y)\mathscr{E}_t(\mathrm{d}y)\right)\right]=\lim_{r\uparrow\infty}\mathbb{E}\left[\mathrm{e}^{-CZ_re^{-\sqrt{2}x}}\right]. \tag{176}$$

The argument above is not quite complete. In particular, the function v does not satisfy the hypothesis needed to apply the convergence results obtained above. In the next subsection we will show that the final result is nonetheless true.

4.4 A Representation for the Laplace Functional

In Sect. 4.2 we have shown convergence of the Laplace functional. The following proposition exhibits the general form of the Laplace functional of the limiting process.

Proposition 4.10 *Let $\mathscr{E}_t$ be the process (158). For $\phi\in\mathscr{C}_c(\mathbb{R})$ non-negative and any $x\in\mathbb{R}$,*

$$\lim_{t\uparrow\infty}\mathbb{E}\left[\exp\left(-\int\phi(y+x)\mathscr{E}_t(\mathrm{d}y)\right)\right]=\mathbb{E}\left[\exp\left(-C(\phi)Z\mathrm{e}^{-\sqrt{2}x}\right)\right] \tag{177}$$

where, for $v(t,y)$ solution of F-KPP (62) with initial condition $v(0,y)=\mathrm{e}^{-\phi(y)}$,

$$C(\phi)=\lim_{t\uparrow\infty}\sqrt{\frac{2}{\pi}}\int_0^\infty\left(1-v(t,y+\sqrt{2}t)\right)y\mathrm{e}^{\sqrt{2}y}\mathrm{d}y \tag{178}$$

is a strictly positive constant depending on ϕ only, and Z is the derivative martingale.

Proof We first establish an integral representation for the Laplace functionals of the extremal process of BBM which are truncated; later we show that the truncation can be removed.

Lemma 4.11 *Set*

$$u_\delta(t,x)\equiv 1-\mathbb{E}\left[\exp\left(-\sum_{k\le n(t)}\phi(-x+x_k(t))\right)\mathbb{1}_{\{\max_{k\le n(t)}(-x+x_k(t))\le\delta\}}\right]. \tag{179}$$

Then $u_\delta(t,x)$ is the solution of the F-KPP equation (68) with initial condition $u_\delta(0,x)=1-\mathrm{e}^{-\phi(-x)}\mathbb{1}_{\{-x\le\delta\}}$. Moreover,

$$C(\phi,\delta)\equiv\lim_{t\uparrow\infty}\sqrt{\frac{2}{\pi}}\int_0^\infty u_\delta(t,y+\sqrt{2}t)y\mathrm{e}^{\sqrt{2}y}\mathrm{d}y\,, \tag{180}$$

exists and

$$\lim_{t\uparrow\infty} u_\delta(t, x + m(t)) = 1 - \mathbb{E}\left[\exp\left(-C(\phi,\delta)Ze^{-\sqrt{2}x}\right)\right]. \tag{181}$$

Proof Due to the presence of the truncation, the function $u_\delta(0, x)$ satisfies the hypothesis that were required in the previous results on sharp approximation. In particular the truncation ensures that $\lim_{x\downarrow-\infty} u_\delta(0, x) = 1$ (and not zero, as it would without the cutoff).

The existence of the constant $C(\phi, \delta)$ follows from Proposition 4.2.

On the other hand, it follows from Theorem 3.6 and the Lalley-Sellke representation, that

$$\lim_{t\uparrow\infty} u_\delta(t, x + m(t)) = 1 - \mathbb{E}\left[\exp\left(-CZe^{-\sqrt{2}x}\right)\right], \tag{182}$$

for some constant C (recall that the relevant solution of (69) is unique up to shifts (see Lemma 3.5), which translates into uniqueness of the representation (181) up to the choice of the constant C).

Thus all what is left is to identify this constant with $C(\phi, \delta)$. This is essentially done by rerunning Proposition 4.2 with $\sqrt{2}t$ replaced by the more precise $m(t)$.

For x, r so that $x \geq m(t) + 8r$, Proposition 4.1 yields the bounds

$$\frac{1}{\gamma(r)}\psi(r, t, x + m(t)) \leq u_\delta(t, x + m(t)) \leq \gamma(r)\psi(r, t, x + m(t)) \tag{183}$$

where

$$\psi(r, t, x + m(t)) \tag{184}$$
$$= \frac{t^{3/2}e^{-\sqrt{2}x}}{\sqrt{2\pi(t-r)}} \int_0^\infty u_\delta(r, y + \sqrt{2}r)e^{y\sqrt{2}}e^{-\frac{\left(y-x+\frac{3}{2\sqrt{2}\ln t}\right)^2}{2(t-r)}}(1 - e^{-2\frac{yx}{t-r}})dy.$$

Using dominated convergence as in Proposition 4.2, one gets

$$\lim_{t\uparrow\infty} \psi(r, t, x + m(t)) = \frac{2xe^{-\sqrt{2}x}}{\sqrt{2\pi}} \int_0^\infty u_\delta(r, y + \sqrt{2}r)ye^{y'\sqrt{2}}dy. \tag{185}$$

Putting this back in (183),

$$\frac{1}{\gamma(r)}C(r) \leq \lim_{t\uparrow\infty} \frac{u_\delta(t, x + m(t))}{xe^{-\sqrt{2}x}} \leq \gamma(r)C(r), \tag{186}$$

for $C(r) \equiv \sqrt{\frac{2}{\pi}} \int_0^\infty u_\delta(r, y + \sqrt{2}r)ye^{y\sqrt{2}}dy$, and $x > 8r$. We know that $\lim_{r\uparrow\infty} C(r) \equiv C > 0$ exists by Proposition 4.2. Thus taking $x = 9r$, letting

$r \uparrow \infty$ in (186), and using that $\gamma(r) \downarrow 1$, one has

$$\lim_{x\uparrow\infty}\lim_{t\uparrow\infty}\frac{\mathrm{e}^{\sqrt{2}x}}{x}u_\delta(t,x+m(t)) = \lim_{r\uparrow\infty}\sqrt{\frac{2}{\pi}}\int_0^\infty u_\delta(r,y+\sqrt{2}r)y\mathrm{e}^{y\sqrt{2}}\mathrm{d}y = C(\phi,\delta). \tag{187}$$

But the constant C appearing in the representation (182) was just the constant form the large x asymptotics of $\lim_{t\uparrow\infty} u(t,x+m(t)$, which is $C(\phi,\delta)$. This proves (181). □

We now want to show that the cutoff δ can be taken to infinity.

Lemma 4.12 *Let $u(t,x), u_\delta(t,x)$ be solutions of the F-KPP equation (68) with initial condition $u(0,x) = 1 - \mathrm{e}^{-\phi(-x)}$ and $u_\delta(0,x) = 1 - \mathrm{e}^{-\phi(-x)}\mathbb{1}_{\{-x\le\delta\}}$, respectively. Then $\lim_{\delta\uparrow\infty} C(\phi,\delta)$ exists, and*

$$C(\phi) \equiv \lim_{\delta\uparrow\infty} C(\phi,\delta) = \lim_{t\uparrow\infty}\sqrt{\frac{2}{\pi}}\int_0^\infty u(t,y+\sqrt{2}t)y\mathrm{e}^{\sqrt{2}y}\mathrm{d}y\,. \tag{188}$$

Moreover

$$\lim_{t\uparrow\infty} u(t,x+m(t)) = 1 - \mathbb{E}\left[\exp -C(\phi)Z\mathrm{e}^{-\sqrt{2}x}\right]\,. \tag{189}$$

Proof Since ϕ is non-negative, it is easy to see from Lemma 3.4 that

$$0 \le u_\delta(t,x) - u(t,x) \le \mathbb{P}(\max x_k(t) > \delta + x)\,, \tag{190}$$

which implies that

$$\begin{aligned}\int_0^\infty u_\delta(t,x+\sqrt{2}t)x\mathrm{e}^{\sqrt{2}x}\mathrm{d}x - \int_0^\infty \mathbb{P}\left(\max x_k(t) - \sqrt{2}t > \delta + x\right)x\mathrm{e}^{\sqrt{2}x}\mathrm{d}x \\ \le \int_0^\infty u(t,x+\sqrt{2}t)x\mathrm{e}^{\sqrt{2}x}\mathrm{d}x \le \int_0^\infty u_\delta(t,x+\sqrt{2}t)x\mathrm{e}^{\sqrt{2}x}\mathrm{d}x.\end{aligned} \tag{191}$$

Define

$$F(t,\delta) \equiv \int_0^\infty u_\delta(t,x+\sqrt{2}t)x\mathrm{e}^{\sqrt{2}x}\mathrm{d}x, \tag{192}$$

$$U(t) \equiv \int_0^\infty u(t,x+\sqrt{2}t)x\mathrm{e}^{\sqrt{2}x}\mathrm{d}x, \tag{193}$$

and

$$M(t,\delta) \equiv \int_0^\infty \mathbb{P}\left(\max x_k(t) - \sqrt{2}t > \delta + x\right)x\mathrm{e}^{\sqrt{2}x}\mathrm{d}x. \tag{194}$$

The inequalities in (191) then read

$$F(t,\delta) - M(t,\delta) \le U(t) \le F(t,\delta). \tag{195}$$

By Lemma 4.7, it holds

$$\lim_{\delta\uparrow\infty} \lim_{t\uparrow\infty} M(t,\delta) = 0. \tag{196}$$

Note also that Proposition 4.2 implies that for each δ, $\lim_{t\uparrow\infty} F(t,\delta) \equiv F(\delta)$ exists and is strictly positive. We thus deduce from (195) that

$$\liminf_{\delta\uparrow\infty} F(\delta) \le \liminf_{t\uparrow\infty} U(t) \le \limsup_{t\uparrow\infty} U(t) \le \limsup_{\delta\uparrow\infty} F(\delta). \tag{197}$$

We must show that $\lim_{\delta\uparrow\infty} F(\delta)$ exists, is strictly positive, and is finite. To see this, observe that the function $\delta \to F(\delta)$ is by construction decreasing and positive. Therefore the limit $\lim_{\delta\uparrow\infty} F(\delta)$ exists. Strict positivity is follows from the convergence of the maximum. Namely, if $\lim_{\delta\uparrow\infty} F(\delta) = 0$, then also $\lim_{t\uparrow\infty} U(t) = 0$. But this would imply that $\lim_{\delta\uparrow\infty} C(\phi,\delta) = 0$, and therefore

$$\lim_{\delta\uparrow\infty} \lim_{t\uparrow\infty} u_\delta(t, x + m(t)) = 0, \tag{198}$$

which is impossible due to the convergence of the maximum and the same argument as used in (171). □

Proposition 4.10 follows directly from Lemma 4.12. □

4.5 Interpretation as Cluster Point Process

In the preceding section we have given a full construction of the Laplace functional of the limiting extremal process of BBM and have given a representation of it in Proposition 4.10. Note that all the information on the limiting process is contained in the way how the constant $C(\phi)$ depends on the function ϕ. The characterisation of this dependence via a solution of the F-KPP equation with initial condition given in terms of ϕ does not appear very revealing at first sight. In the following we will remedy this by giving explicit probabilistic descriptions of the underlying point process.

4.5.1 Interpretation via an Auxiliary Process

We will construct an auxiliary class of point processes that a priory have nothing to do with the real process of BBM. To make the distinction transparent, we denote

probabilities and expectation for this process by P and E. Let $(\eta_i; i \in \mathbb{N})$ denote the atoms of a Poisson point process on $(-\infty, 0)$ with intensity measure

$$\sqrt{\frac{2}{\pi}}(-x)\mathrm{e}^{-\sqrt{2}x}\mathrm{d}x\,. \tag{199}$$

For each $i \in \mathbb{N}$, consider independent BBM's $\{x_k^{(i)}(t), k \le n^{(i)}(t)\}$. Note that, for each $i \in \mathbb{N}$,

$$\max_{k\le n^{(i)}(t)} x_k^{(i)}(t) - \sqrt{2}t \to -\infty, \text{ a.s.} \tag{200}$$

The auxiliary point process of interest is constructed from these ingredients as

$$\Pi_t \equiv \sum_{i,k} \delta_{\frac{1}{\sqrt{2}}\ln Z+\eta_i+x_k^{(i)}(t)-\sqrt{2}t}, \tag{201}$$

where Z has the same law as limit of the derivative martingale. The existence and non-triviality of the process in the limit $t \uparrow \infty$ is not obvious, especially in view of (200). We will show that not only it exists, but it has the same law as the limit of the extremal process of BBM.

Theorem 4.13 *Let $\mathscr{E}_t$ be the extremal process (158) of BBM. Then*

$$\lim_{t\uparrow\infty} \mathscr{E}_t \overset{law}{=} \lim_{t\uparrow\infty} \Pi_t\,. \tag{202}$$

Proof The proof of this result just requires the computation of the Laplace transform, which we are already quite skilled in.

The Laplace functional of Π_t using the form of the Laplace functional of a Poisson process reads

$$\begin{aligned}
&E\left[\exp\left(-\int \phi(x)\Pi_t(\mathrm{d}x)\right)\right]\\
&\quad= E\Bigg[\exp\Big(-\int_{-\infty}^{0}\Big\{1-E\Big[\exp\big(-\sum_{k\le n(t)}\phi(x+x_k(t)-\sqrt{2}t+\frac{1}{\sqrt{2}}\ln Z)\big)\Big]\Big\}\\
&\qquad\times\sqrt{\frac{2}{\pi}}(-x)\mathrm{e}^{-\sqrt{2}x}\mathrm{d}x\Big)\Bigg]\\
&\quad= E\left[\exp\left(-\sqrt{\frac{2}{\pi}}\int_0^\infty u(t,x+\sqrt{2}t-\frac{1}{\sqrt{2}}\ln Z)x\mathrm{e}^{\sqrt{2}x}\mathrm{d}x\right)\right],
\end{aligned} \tag{203}$$

with

$$u(t,x) = 1 - E\left[\exp\left(-\sum_{k=1}^{n(t)} \phi(-x + x_k(t))\right)\right]. \tag{204}$$

By (153), the exponent in the last line of (203) converges to $-ZC(\phi)$, to wit

$$\begin{aligned}\lim_{t\uparrow\infty} \sqrt{\frac{2}{\pi}} \int_0^\infty u(t, x + \sqrt{2}t - \frac{1}{\sqrt{2}} \ln Z) x \mathrm{e}^{\sqrt{2}x} \mathrm{d}x \qquad (205)\\ = Z\sqrt{\frac{2}{\pi}} \lim_{t\uparrow\infty} \int_0^\infty u(t, x + \sqrt{2}t) x \mathrm{e}^{\sqrt{2}x} \mathrm{d}x = ZC(\phi),\end{aligned}$$

where the last equality follows by Lemma 4.12. This implies that the Laplace functionals of $\lim_{t\uparrow\infty} \Pi_t$ and of the extremal process of BBM are equal and proves the proposition. □

4.5.2 The Poisson Process of Cluster Extremes

We will now give another interpretation of the extremal process. In the paper [4], the following result was shown: consider the ordered enumeration of the particles of BBM at time t,

$$x_1(t) \geq x_2(t) \geq \cdots \geq x_{n(t)}(t). \tag{206}$$

Fix $r > 0$ and construct the equivalence classes of these particles such that each class consists of particles that have a common ancestor at time $s > t - r$. For each of these classes, we can select as a representative its largest member. This can be constructed recursively as follows:

$$\begin{aligned} i_1 &= 1, \\ i_k &= \min(j > i_{k-1} : q(i_\ell, j) < t - r, \forall \ell \leq k - 1). \end{aligned} \tag{207}$$

This is repeated to exhaustion and provides the desired decomposition. We denote the resulting number of classes by $n^*(t)$. We can think of the $x_{i_k}(t)$ as the heads of families at time t. We then construct the point processes

$$\Theta_t^r \equiv \sum_{k=1}^{n^*(t)} \delta_{x_{i_k}(t) - m(t)}. \tag{208}$$

In [4] the following result was proven.

Theorem 4.14 *Let Θ_t^r be defined above. Then*

$$\lim_{r\uparrow\infty}\lim_{t\uparrow\infty}\Theta_t^r = \Theta, \tag{209}$$

where convergence is in law and Θ is the random Poisson process (Cox process) with intensity measure $CZe^{-\sqrt{2}x}dx$, Z being the derivative martingale, and C the constant from Theorem 3.12.

Remark 4.15 The thinning-out of the extremal process removes enough correlation to Poissonise the process. This is rather common in the theory of extremal processes, see e.g. [9, 10].

I will not prove this theorem here. Rather, I will show how this result links up to the representation of the extremal process of BBM given above.

Proposition 4.16 *Let*

$$\Pi_t^{ext} \equiv \sum_i \delta_{\frac{1}{\sqrt{2}}\ln Z+\eta_i+M_i(t)-\sqrt{2}t} \tag{210}$$

where $M^{(i)}(t) \equiv \max_k x_k^{(i)}(t)$, i.e. be the point process obtained by retaining from Π_t the maximal particles of each of the BBM's. Then

$$\lim_{t\uparrow\infty}\Pi_t^{ext} \stackrel{law}{=} \mathscr{P}_Z = \mathrm{PPP}\left(Z\sqrt{2}Ce^{-\sqrt{2}x}dx\right) \tag{211}$$

as a point process on $\mathbb{R}$, where C is the same constant appearing the law of the maximum. In particular, the maximum of $\lim_{t\uparrow\infty}\Pi_t^{ext}$ has the same law as the limit law of the maximum of BBM.

Proof Consider now the Laplace functional of our thinned out auxiliary process,

$$E\left[\exp\left(-\sum_i \phi(\eta_i + M^{(i)}(t) - \sqrt{2}t)\right)\right] \tag{212}$$

Since the $M^{(i)}$'s are i.i.d., and denoting by $\mathfrak{F}_\eta$ the σ-algebra generated by the Poisson point process of the η's, we get that

$$\begin{aligned} &E\left[\exp\left(-\sum_i \phi(\eta_i + M^{(i)}(t) - \sqrt{2}t)\right)\right] \\ &= E\left[\prod_i E\left(\exp\left(-\phi(\eta_i + M^{(i)}(t) - \sqrt{2}t)\right)|\mathfrak{F}_\eta\right)\right] \\ &= E\exp\left(-\sum_i g(\eta_i)\right), \end{aligned} \tag{213}$$

where

$$g(z) \equiv -\ln\left(E\left[\exp\left(-\phi(z+M(t)-\sqrt{2}t)\right)\big|\mathfrak{F}_\eta\right]\right), \tag{214}$$

$M(t)$ has the same law as $M_i(t)$. Using now the form of a Laplace functional for Poisson processes, (213) is equal to

$$\exp\left(-\int_{-\infty}^{0}\left(1-\mathrm{e}^{-g(y)}\right)\sqrt{\frac{2}{\pi}}(-y)\mathrm{e}^{-\sqrt{2}y}\mathrm{d}y\right). \tag{215}$$

Finally,

$$\begin{aligned}&\int_{-\infty}^{0}\left[1-\mathrm{e}^{-g(y)}\right]\sqrt{\frac{2}{\pi}}(-y)\mathrm{e}^{-\sqrt{2}y}\mathrm{d}y\\ &\quad=\int_{-\infty}^{0}E\left(1-\mathrm{e}^{-\phi(y+M(t)-\sqrt{2}t)}\right)\sqrt{\frac{2}{\pi}}(-y)\mathrm{e}^{-\sqrt{2}y}\mathrm{d}y.\end{aligned} \tag{216}$$

Using Lemma 4.8 with $h(x) = 1-\mathrm{e}^{-\phi(x)}$, we see that

$$\begin{aligned}&\lim_{t\uparrow\infty}\int_{-\infty}^{0}E\left(1-\mathrm{e}^{-\phi(y+M(t)-\sqrt{2}t)}\right)\sqrt{\frac{2}{\pi}}(-y)\mathrm{e}^{-\sqrt{2}y}\mathrm{d}y\\ &\quad=\int_{\mathbb{R}}\left(1-\mathrm{e}^{-\phi(a)}\right)C\sqrt{2}\mathrm{e}^{-\sqrt{2}a}\mathrm{d}a,\end{aligned} \tag{217}$$

which is the Laplace functional of the PPP with intensity $\sqrt{2}C\mathrm{e}^{-\sqrt{2}a}\mathrm{d}a$ on $\mathbb{R}$. Adding the random shift $\frac{1}{\sqrt{2}}\ln Z$, the assertion of Proposition 4.16 follows. □

So this is nice. The process of the maxima of the BBM's in the auxiliary process is the same Poisson process as the limiting process of the heads of families. This gives a clear interpretation of the BBM's in the auxiliary process.

4.5.3 More on the Law of the Clusters

Let us continue to look at the auxiliary process. We know that it is very hard for any of the $M^{(i)}(t)$ to exceed $\sqrt{2}t$ and thus to end up on above a level a as $t \uparrow \infty$. Therefore, many of the atoms η_i of the Poisson process will not have relevant offspring.

The following proposition states that there is a narrow window deep below zero from which all relevant particles come. This is analogous to the observation in Lemma 4.5:

Proposition 4.17 *For any $y \in \mathbb{R}$ and $\varepsilon > 0$ there exist $0 < A_1 < A_2 < \infty$ and t_0 depending only on y and ε, such that*

$$\sup_{t \geq t_0} P\left(\exists_{i,k} : \eta_i + x_k^{(i)}(t) - \sqrt{2}t \geq z, \wedge \eta_i \notin \left[-A_1\sqrt{t}, -A_2\sqrt{t}\right]\right) < \varepsilon. \tag{218}$$

Proof Throughout the proof, the probabilities are considered conditional on Z. Clearly we have

$$\begin{aligned} &P\left(\exists_{i,k} : \eta_i + x_k^{(i)}(t) - \sqrt{2}t \geq z, \text{ but } \eta_i \geq -A_1\sqrt{t}\right) \qquad (219)\\ &\leq E\left[\sum_i \mathbb{1}_{\eta_i \in [-A_1\sqrt{t},0]} P\left(M^{(i)}(t) - \sqrt{2}t + \eta_i \geq y \middle| \mathfrak{F}_\eta\right)\right],\\ &= \int_{-A_1\sqrt{t}}^{0} P\left(M^{(i)}(t) - \sqrt{2}t + \eta_i \geq z\right) C(-y)\mathrm{e}^{-\sqrt{2}y}\mathrm{d}y\\ &= \int_0^{A_1\sqrt{t}} P\left(M^{(i)}(t) - \sqrt{2}t \geq z + y + \sqrt{2}t\right) C y \mathrm{e}^{\sqrt{2}y}\mathrm{d}y. \end{aligned}$$

Similarly, we have that

$$\begin{aligned} &P\left(\exists_{i,k} : \eta_i + x_k^{(i)}(t) - \sqrt{2}t \geq z, \text{ but } \eta_i \leq -A_2\sqrt{t}\right) \qquad (220)\\ &\leq \int_{A_2\sqrt{t}}^{\infty} P\left(M^{(i)}(t) - \sqrt{2}t \geq z + y + \sqrt{2}t\right) C y \mathrm{e}^{\sqrt{2}y}\mathrm{d}y. \end{aligned}$$

Both terms can be made smaller than $\varepsilon/2$ (as $t \uparrow \infty$) by choosing A_1 small enough and A_2 large enough, due to Lemma 4.5 by taking A_1 small enough. □

We see that there is a close analogy to Lemma 4.5. How should we interpret the auxiliary process? Think of a very large time t, and go back to time $t - r$. A that time, there is a certain distribution of particles which all lie well below $\sqrt{2}(t - r)$ by an order of $\sqrt{r}$. From those a small fraction will have offspring that reach the excessive size $\sqrt{2}r + \sqrt{r}$ and thus contribute to the extremes. The selected particles form the Poisson process, while their offspring, which are now BBMs conditioned to be extra large, form the clusters.

4.5.4 The Extremal Process Seen from the Cluster Extremes

We finally come to yet another description of the extremal process. Here we start with the Poisson process of Proposition 4.16 and look at the law of the clusters "that made it up to there". Now we know that the point in the auxiliary process all came from the $\sqrt{t}$ window around $-\sqrt{2}t$. Thus we may expect the clusters to look like

BBM that was bigger than $\sqrt{2}t$. Therefore we define the process

$$\overline{\mathscr{E}}_t = \sum_{k \le n(t)} \delta_{x_k(t)-\sqrt{2}t}\,. \tag{221}$$

Obviously, the limit of such a process must be trivial, since the probability that the maximum of BBM shifted by $-\sqrt{2}t$ does not drift to $-\infty$ is vanishing. However, *conditionally* on the event $\{\max_k x_k(t) - \sqrt{2}t \ge 0\}$, the process $\overline{\mathscr{E}}_t$ does converge to a well defined point process $\overline{\mathscr{E}} = \sum_j \delta_{\xi_j}$ as $t \uparrow \infty$.

We may then define the *point process of the gaps*

$$\mathscr{D}_t \equiv \sum_k \delta_{x_k(t)-\max_{j\le n(t)} x_j(t)}, \tag{222}$$

and

$$\mathscr{D} = \sum_j \delta_{\Delta_j}, \quad \Delta_j \equiv \xi_j - \max_i \xi_i, \tag{223}$$

where ξ_i are the atoms of the limiting process $\overline{\mathscr{E}}$. Note that $\mathscr{D}$ is a point process on $(\infty, 0]$ with an atom at 0.

Theorem 4.18 *Let $\mathscr{P}_Z$ be as in (211) and let $\{\mathscr{D}^{(i)}, i \in \mathbb{N}\}$ be of independent copies of the gap- process (223). Then the point processes $\mathscr{E}_t = \sum_{k\le n(t)} \delta_{x_k(t)-m(t)}$ converges in law as $t \uparrow \infty$ to a Poisson cluster point process $\mathscr{E}$ given by ν*

$$\mathscr{E} \equiv \lim_{t\uparrow\infty} \mathscr{E}_t \overset{law}{=} \sum_{i,j} \delta_{p_i+\Delta_j^{(i)}}. \tag{224}$$

This theorem looks quite reasonable. The only part that may be surprising is that all the Δ_j have the same law, since a priori we only know that they come from a $\sqrt{t}$ neighbourhood of $\sqrt{2}t$. This is the content of the next theorem.

Theorem 4.19 *Let $x \equiv -a\sqrt{t} + b$ for some $a > 0, b \in \mathbb{R}$. The point process*

$$\sum_{k\le n(t)} \delta_{x+x_k(t)-\sqrt{2}t} \tag{225}$$

converges in law under $\mathbb{P}\left(\cdot \middle| \{x + \max_k x_k(t) - \sqrt{2}t > 0\}\right)$, as $t \uparrow \infty$ to a well-defined point process $\overline{\mathscr{E}}$. The limit does not depend on a and b, and the maximum of $\overline{\mathscr{E}}$ shifted by x has the law of an exponential random variable of parameter $\sqrt{2}$.

Proof Set $\max \overline{\mathscr{E}}_t \equiv \max_i x_i(t) - \sqrt{2}t$. We first show that

$$\lim_{t\uparrow\infty} \mathbb{P}\left(x + \max \overline{\mathscr{E}}_t > X \,\middle|\, x + \max \overline{\mathscr{E}}_t > 0\right) = \mathrm{e}^{-\sqrt{2}X}, \tag{226}$$

for $X > 0$. This follows from the fact that the conditional probability is just

$$\frac{\mathbb{P}\left(x + \max \overline{\mathscr{E}}_t > X\right)}{\mathbb{P}\left(x + \max \overline{\mathscr{E}}_t > 0\right)}, \tag{227}$$

and the numerator and denominator can be well approximated by the functions $\psi(r, t, X - x + \sqrt{2}t)$ and $\psi(r, t, -x + \sqrt{2}t)$, respectively. Using Lemma 4.4 for these, we get the assertion (226).

Second, we show that for any function ϕ that is continuous with compact support, the limit of

$$\mathbb{E}\left[\exp - \int \phi(x + z)\overline{\mathscr{E}}_t(\mathrm{d}z) \middle| x + \max \overline{\mathscr{E}}_t > 0\right] \tag{228}$$

exists and is independent of x. The proof is just a tick more complicated than before, but relies again on the properties of the functions ψ. I will skip the details. □

We now can proof Theorem 4.18.

Proof (of Theorem 4.18) We show that for $\phi : \mathbb{R} \to \mathbb{R}_+$ continuous with compact support, the Laplace functional $\Psi_t(\phi)$ of the extremal process $\mathscr{E}_t$ satisfies

$$\lim_{t\uparrow\infty} \Psi_t(\phi) = E\left[\exp\left(-CZ \int_{\mathbb{R}} \mathbb{E}[1 - \mathrm{e}^{-\int \phi(y+z)\mathscr{D}(\mathrm{d}z)}]\sqrt{2}\mathrm{e}^{-\sqrt{2}y}\mathrm{d}y\right)\right] \tag{229}$$

for the point process $\mathscr{D}$ defined in (223).

By Theorem 4.13,

$$\lim_{t\uparrow\infty} \Psi_t(\phi) = \lim_{t\uparrow\infty} E\left[\exp\left(-\sum_{i,k} \phi(\eta_i + \tfrac{1}{\sqrt{2}} \ln Z + x_k^{(i)}(t) - \sqrt{2}t)\right)\right]. \tag{230}$$

Using the form for the Laplace transform of a Poisson process we have for the right side

$$\begin{aligned} &\lim_{t\uparrow\infty} E\left[\exp\left(-\sum_{i,k} \phi(\eta_i + \tfrac{1}{\sqrt{2}} \ln Z + x_k^{(i)}(t) - \sqrt{2}t)\right)\right] \\ &= E\left[\exp\left(-Z \lim_{t\uparrow\infty} \int_{-\infty}^{0} E\left[1 - \exp\left(-\int \phi(x + y)\overline{\mathscr{E}}_t(\mathrm{d}x)\right)\right] \sqrt{\frac{2}{\pi}}(-y)\mathrm{e}^{-\sqrt{2}y}\mathrm{d}y\right)\right]. \end{aligned} \tag{231}$$

Define

$$\mathscr{D}_t \equiv \sum_{i \le n(t)} \delta_{x_i(t) - \max_{j \le n(t)} x_j(t)}. \tag{232}$$

The integral on the right-hand side equals

$$\lim_{t\uparrow\infty}\int_{-\infty}^{0} E\left[f\left(\int \phi(z+y+\overline{\mathscr{E}}_t)\mathscr{D}_t(\mathrm{d}z)\right)\right]\sqrt{\frac{2}{\pi}}(-y)\mathrm{e}^{-\sqrt{2}y}\mathrm{d}y \tag{233}$$

with $f(x)\equiv 1-\mathrm{e}^{-x}$. By Proposition 4.17, there exist A_1 and A_2 such that

$$\begin{aligned}&\int_{-\infty}^{0} E\left[f\left(\int \phi(zy+\max\overline{\mathscr{E}}_t)\mathscr{D}_t(\mathrm{d}z)\right)\right]\sqrt{\frac{2}{\pi}}(-y)\mathrm{e}^{-\sqrt{2}y}\mathrm{d}y \qquad (234)\\ &\quad=\Omega_t(A_1,A_2)\\ &\qquad+\int_{-A_2\sqrt{t}}^{-A_1\sqrt{t}} E\left[f\left(\int \phi(z+y+\max\overline{\mathscr{E}}_t)\mathscr{D}_t(\mathrm{d}z)\right)\right]\sqrt{\frac{2}{\pi}}(-y)\mathrm{e}^{-\sqrt{2}y}\mathrm{d}y, \qquad (235)\end{aligned}$$

where the error term satisfies $\lim_{A_1\downarrow 0,A_2\uparrow\infty}\sup_{t\geq t_0}\Omega_t(A_1,A_2)=0$. Let m_ϕ be the minimum of the support of ϕ. Note that

$$f\left(\int \phi(z+y+\max\overline{\mathscr{E}}_t)\mathscr{D}_t(\mathrm{d}z)\right) \tag{236}$$

is zero when $y+\max\overline{\mathscr{E}}_t<m_\phi$, and that the event $\{y+\max\overline{\mathscr{E}}_t=m_\phi\}$ has probability zero. Therefore,

$$\begin{aligned}&E\left[f\left(\int \phi(z+y+\max\overline{\mathscr{E}}_t)\mathscr{D}_t(\mathrm{d}z)\right)\right]\\ &\quad=E\left[f\left(\int \phi(z+y+\max\overline{\mathscr{E}}_t)\mathscr{D}_t(\mathrm{d}z)\right)\mathbb{1}_{\{y+\max\overline{\mathscr{E}}_t>m_\phi\}}\right]\\ &\quad=E\left[f\left(\int \phi(z+y+\max\overline{\mathscr{E}}_t)\mathscr{D}_t(\mathrm{d}z)\right)\Big|y+\max\overline{\mathscr{E}}_t>m_\phi\right]P\left(y+\max\overline{\mathscr{E}}_t>m_\phi\right).\end{aligned} \tag{237}$$

One can show (see Corollary 4.12 in [6]) that the conditional law of the pair $(\mathscr{D}_t, y+\max\overline{\mathscr{E}}_t)$ *given* $\{y+\max\overline{\mathscr{E}}_t>m_\phi\}$ converges as $t\uparrow\infty$. Moreover, the convergence is uniform in $y\in[-A_1\sqrt{t},-A_2\sqrt{t}]$. This implies the convergence of the random variable $\int\phi(z+y+\max\overline{\mathscr{E}}_t)\mathscr{D}_t(\mathrm{d}z)$. Therefore,

$$\begin{aligned}&\lim_{t\uparrow\infty}E\left[f\left(\int \phi(z+y+\max\overline{\mathscr{E}}_t)\mathscr{D}_t(\mathrm{d}z)\right)\Big|y+\max\overline{\mathscr{E}}_t>m_\phi\right]\\ &\quad=\mathrm{e}^{\sqrt{2}m_\phi}\int_{m_\phi}^{\infty}E\left[f\left(\int\phi(z+y)\mathscr{D}(\mathrm{d}z)\right)\right]\sqrt{2}\mathrm{e}^{-\sqrt{2}y}\mathrm{d}y.\end{aligned} \tag{238}$$

On the other hand,

$$\int_{-A_2\sqrt{t}}^{-A_1\sqrt{t}} P\left(y+\max\overline{\mathscr{E}}_t > m_\phi\right)\sqrt{\frac{2}{\pi}}(-y)\mathrm{e}^{-\sqrt{2}y}\mathrm{d}y = C\mathrm{e}^{-\sqrt{2}m_\phi} + \Omega_t(A_1,A_2) \tag{239}$$

by Lemma 4.8, using the same approximation as in (234).

Combining (239), (238) and (237), one sees that (231) converges to

$$E\left[\exp\left(-CZ\int_{\mathbb{R}} E[1-\mathrm{e}^{-\int\phi(y+z)\mathscr{D}(\mathrm{d}z)}]\sqrt{2}\mathrm{e}^{-\sqrt{2}y}\mathrm{d}y\right)\right], \tag{240}$$

which is by (230) the limiting Laplace transform of the extremal process of branching Brownian motion: this shows (229) and concludes the proof of Theorem 4.18. □

The properties of BBM conditioned to exceed their natural threshold were already described in detail by Chauvin and Rouault [23, Theorem 5]. There will be one branch (the *spine*) that exceeds the level $\sqrt{2}t$ by an exponential random variable of parameter $\sqrt{2}$ (see the preceding proposition). The spine seen from its endpoint is a Brownian motion with drift $-\sqrt{2}$ (i.e. very close to a straight line of slope $\sqrt{2}$). This spine is decorated with size biased BBM's conditions to stay below it. In the work of Aïdékon et al [2], the description of the limiting point process is obtained directly in this form (see their Theorem 2.3).

From this spine ordinary BBM's branch off at Poissonian times. Clearly, all the branches that split off at times later than r before the end-time, will reach at most the level

$$\sqrt{2}(t-r)+\sqrt{2}r-\frac{3}{2\sqrt{2}}\ln r = \sqrt{2}t-\frac{3}{2\sqrt{2}}\ln r. \tag{241}$$

Seen from the top, i.e. $\sqrt{2}t$, this tends to $-\infty$ as $r\uparrow\infty$. Thus only branches that are created "a finite time" before the end-time t remain visible in the extremal process. This does, of course, correspond perfectly to the observation in Theorem 4.14 that implies that all the points visible in the extremal process have a common ancestor at a *finite* time before the end. As a matter of fact, from the observation above one can prove Theorem 4.14 easily.

5 Variable Speed BBM

We have seen that BBM is somewhat related to the GREM where the covariance is a linear function of the distance. It is natural to introduce versions of BBM that have more general covariance functions $A(x)$. This can be achieved by changing the speed (= variance) of the Brownian motion with time. Variable speed BBM was

introduced by Derrida and Spohn [25] and has recently been studied recently by Fang and Zeitouni [27, 28] and others [42, 43, 46].

The general model can be constructed as follows. Let $A : [0,1] \to [0,1]$ be a right-continuous increasing function. Fix a time horizon t and let

$$\Sigma^2(x) = tA(x/t). \tag{242}$$

Note that Σ^2 is almost everywhere differentiable and denote by $\sigma^2(x)$ its derivative wherever it exists. We define Brownian motion with speed function Σ^2 as time change of ordinary Brownian motion on the interval $[0,t]$ as

$$B_x^\Sigma = B_{\Sigma^2(x)}. \tag{243}$$

Now branching Brownian motion with speed function Σ^2 is constructed like ordinary Brownian motion except that if a particle splits at some time $s < t$, then the offspring particles perform variable speed Brownian motions with speed function Σ^2, i.e. their laws are independent copies $\{B_r^\Sigma - B_s^\Sigma\}_{t \geq r \geq s}$, all starting at the position of the parent particle at time s.

We denote by $n(s)$ the number of particles at time s and by $\{x_i(s); 1 \leq i \leq n(s)\}$ the positions of the particles at time s. If we denote by $d(x_k(t), x_\ell(t))$ the time of the most recent common ancestor of the particles i and k, then a simple computation shows that

$$\mathbb{E}x_k(s)x_\ell(s) = \Sigma^2\left(d(x_k(s), x_\ell(s))\right). \tag{244}$$

Moreover, for different times, we have

$$\mathbb{E}x_k(s)x_\ell(r) = \Sigma^2\left(d(x_k(t), x_\ell(t) \wedge s \wedge r)\right). \tag{245}$$

Remark 5.1 Strictly speaking, we are not talking about a single stochastic process, but about a family $\{x_k(s), k \leq n(s)\}_{s \leq t}^{t \in \mathbb{R}_+}$ of processes with finite time horizon, indexed by that horizon, t.

The case when A is a step function with finitely many steps corresponds to Derrida's GREMs, with the only difference that the binary tree is replaced by a Galton-Watson tree. The case we discuss here corresponds to A being a piecewise linear function. The case when A is arbitrary has been dubbed CREM in [16] (and treated for binary regular trees). In that case the leading order of the maximum was obtained; this analysis carries over mutando mutandis to the BBM situations. Fang and Zeitouni [28] have obtained the order of the correction (namely $t^{1/3}$) in the case when A is strictly concave and continuous, but there are no results on the extremal process or the law of the maximum. This result has very recently strengthened by Maillard and Zeitouni [42] who proved convergence of the law of the maximum to some travelling wave and computed the next order of the correction (which is logarithmic).

Understanding the piecewise linear case seems to be a prerequisite to getting the full picture. Fang and Zeitouni [28] have in this case obtained the correct order of the corrections, which is completely analogous to the GREM situation, without further information on the law of the maximum and the extremal process.

The simplest case corresponds to choosing a speed σ_1^2 up to a time $t_1 = tb$ and a speed σ_2^2 after time t_1. This case was fully analysed in a recent paper with Lisa Hartung [13], we provide the construction of the extremal processes. Interestingly, this provides a few surprises.

In this case, Fang and Zeitouni [27] showed that

$$\max_{k\le n(t)} x_k(t) = \begin{cases} \sqrt{2}t - \frac{1}{2\sqrt{2}}\ln t + O(1), & \text{if } \sigma_1 < \sigma_2, \\ \sqrt{2}t(\sigma_1 b + \sigma_2(1-b)) & \\ -\frac{3}{2\sqrt{2}}(\sigma_1+\sigma_2)\ln t + O(1), & \text{if } \sigma_1 > \sigma_2. \end{cases} \tag{246}$$

The second case has a simple interpretation: the maximum is achieved by adding to the maxima of BBM at time tb the maxima of their offspring at time $t(1-b)$ later, just as in the analog case of the GREM. The second case suggests that we are in the simple REM case, as would be the case of the GREM. But here things are a lot more subtle.

The main result of [13] is the following.

Theorem 5.2 ([13]) *Let $x_k(t)$ be branching Brownian motion with two speeds as described above. Assume that $\sigma_1 < \sigma_2$ and $b \in (0,1)$. Then*

(i)

$$\lim_{t\uparrow\infty} \mathbb{P}\left(\max_{k\le n(t)} x_k(t) - \tilde{m}(t) \le x\right) = \mathbb{E}e^{-C'Ye^{-\sqrt{2}x}}, \tag{247}$$

where $\tilde{m}(t) = \sqrt{2}t - \frac{1}{2\sqrt{2}}\ln t$, C' is a constant and Y is a random variable that is the limit of a martingale (but different from Z!).

(ii) The point process

$$\tilde{\mathscr{E}}_t \equiv \sum_{k\le n(t)} \delta_{x_k(t)-\tilde{m}(t)} \to \tilde{\mathscr{E}}, \tag{248}$$

as $t \uparrow \infty$, in law, where

$$\tilde{\mathscr{E}} = \sum_{k,j} \delta_{\eta_k + \sigma_2 \Lambda_j^{(k)}}, \tag{249}$$

where η_k is the kth atom of a mixture of Poisson point process with intensity measure $C'Ye^{-\sqrt{2}x}\mathrm{d}x$, with C' and Y as in (i), and $\Lambda_i^{(k)}$ are the atoms of independent and identically distributed point processes $\Lambda^{(k)}$, which are the

limits in law of

$$\sum_{j\le n(t)} \delta_{x_i(t)-\max_{j\le n(t)} x_j(t)}, \tag{250}$$

where $x(t)$ is BBM of speed 1 conditioned on $\max_{j\le n(t)} x_j(t) \ge \sqrt{2}\sigma_2 t$.

The picture is completed by the limiting extremal process in the case $\sigma_1 > \sigma_2$. This result is much simpler and could be guessed form the known fact on the GREM.

Theorem 5.3 ([13]) *Let $x_k(t)$ be as in Theorem 5.2, but $\sigma_2 < \sigma_1$. Again $b \in (0,1)$. Let $\mathscr{E} \equiv \mathscr{E}^0$ and $\mathscr{E}^{(i)}, i \in \mathbb{N}$ be independent copies of the extremal process of standard branching Brownian motion. Let*

$$m(t) \equiv \sqrt{2}t(b\sigma_1 + (1-b)\sigma_2) - \frac{3}{2\sqrt{2}}(\sigma_1+\sigma_2)\ln t - \frac{3}{2\sqrt{2}}(\sigma_1 \ln b + \sigma_2 \ln(1-b)), \tag{251}$$

and set

$$\tilde{\mathscr{E}}_t \equiv \sum_{k\le n(t)} \delta_{x_k(t)-m(t)}. \tag{252}$$

Then

$$\lim_{t\uparrow\infty} \tilde{\mathscr{E}}_t = \tilde{\mathscr{E}}, \tag{253}$$

exists, and

$$\tilde{\mathscr{E}} = \sum_{i,j} \delta_{\sigma_1 e_i + \sigma_2 e_j^{(i)}}, \tag{254}$$

where $e_i, e_j^{(i)}$ are the atoms of the point processes $\mathscr{E}$ and $\mathscr{E}^{(i)}$, respectively.

We just comment on the main steps in the proof of Theorem 5.2. The first step is a localisation of the particles that will eventually reach the top at the time of the speed change. The following proposition says that these are in a $\sqrt{t}$ neighbourhood of $\sigma_1^2 t\sqrt{2}b$, which is much smaller then the position $\sigma_1 t\sqrt{2}b$ of the leading particles at time tb.

Proposition 5.4 *Let $\sigma_1 < \sigma_2$. For any $d \in \mathbb{R}$ and any $\epsilon > 0$, there exists a constant $A > 0$ such that for all t large enough*

$$\mathbb{P}\left(\exists_{j\le n(t)} : x_j(t) > \tilde{m}(t) - d \wedge |x_j(bt) - \sqrt{2}\sigma_1^2 bt| > A\sqrt{t}\right) \le \epsilon. \tag{255}$$

Thus, the faster particles in the second half-time must make up for this. It is not very hard to know everything about the particles after time tb. The main problem

one is faced with is to control their initial distribution at this time. Fortunately, this can be done with the help of a martingale. Define

$$Y_s = \sum_{i=1}^{n(s)} e^{-s(1+\sigma_1^2)+\sqrt{2}x_i(s)}. \tag{256}$$

When $\sigma_1 < 1$, one can show that Y is a uniformly integrable positive martingale with mean value one. This yields by the usual martingale convergence theorems the following theorem.

Theorem 5.5 *The limit* $\lim_{s\to\infty} Y_s$ *exists almost surely and in* L^1*, is finite and strictly positive.*

The martingale Y_s will take over the role of the derivative martingale in the standard case. The proof of the convergence of the maximal law and the Laplace functionals now follows the strategy indicated in Section 4.3. That is, we use explicitly the convergence of the martingale Y_t together with the asymptotics of the post-bt BBMs to prove convergence without relying on analytic results on this convergence.

The remainder of the proofs is fairly similar to the proof in the standard case and will not be given here.

Remark 5.6 We see that in the case $\sigma_1 < \sigma_2$ only the variances enter in the description of the limiting process. The McKean martingale depends only on σ_1 and the process of the clusters depends only on σ_2. One may in fact obtain exactly the same result if $b = b(t) \downarrow 0$, provided $b(t)t \uparrow \infty$. It is interesting to look at the extreme cases. If b becomes small, nothing seems to change. There is always enough time for the McKean martingale to converge. If $\sigma_1 = 0$, then $Y_s = e^{-s}n(s)$, which is known to converge to an exponential random variable. If b tends to one, then σ_2 has to tend to infinity. It is not very hard to convince oneself that in that case, the processes $\Lambda^{(k)}$ of the clusters converge to the trivial process δ_0.

Remark 5.7 In the course of the proof one finds that in all cases, the Laplace functional of the limiting process has the form

$$\lim_{t\uparrow\infty} \Psi_t(\phi(\cdot + x)) = \mathbb{E}\exp\left(-C(\phi)M\mathrm{e}^{-\sqrt{2}x}\right), \tag{257}$$

where M is a martingale limit (either Y of Z) and $C(\phi)$ is a constant whose dependence on ϕ contains all the information on the limiting process. This is in agreement with the finding in [42] in the case where the speed is a concave function of s/t. The universal form (257) is thus misleading and without knowledge of the specific form of $C(\phi)$, (257) contains almost no information.

Remark 5.8 After these lectures were given, the results of Theorem 5.3 were extended in [14]. The result is the following theorem.

Theorem 5.9 *Assume that $A : [0,1] \to [0,1]$ satisfies some very mild regularity assumptions. Let $A'(0) = \sigma_b^2 < 1$ and $A'(1) = \sigma_e^2 > 1$. Let $\tilde{m}(t) = \sqrt{2}t - \frac{1}{2\sqrt{2}} \log t$. Then there is a constant $\tilde{C}(\sigma_e)$ depending only on σ_e and a random variable Y_{σ_b} depending only on σ_b such that*

(i)

$$\lim_{t\uparrow\infty} \mathbb{P}\left(\max_{1\le i\le n(t)} x_i(t) - \tilde{m}(t) \le x \right) = \mathbb{E}\left[e^{-\tilde{C}(\sigma_e) Y_{\sigma_e} e^{-\sqrt{2}x}} \right]. \tag{258}$$

(ii) The point process

$$\sum_{k\le n(t)} \delta_{x_k(t)-\tilde{m}(t)} \to \mathscr{E}_{\sigma_b,\sigma_e} = \sum_{i,j} \delta_{p_i+\sigma_e \Lambda_j^{(i)}}, \tag{259}$$

as $t \uparrow \infty$, in law, where the p_i are the atoms of a Poisson point process on $\mathbb{R}$ with intensity measure $\tilde{C}(\sigma_e) Y_{\sigma_b} e^{-\sqrt{2}x} \mathrm{d}x$, and the $\Lambda^{(i)}$ are the limits of the processes as in (250), but conditioned on the event $\{\max_k \tilde{x}_k(t) \ge \sqrt{2}\sigma_e t\}$.

(iii) If $A'(1) = \infty$, then $C(\infty) = 1/\sqrt{4\pi}$, and $\Lambda^{(i)} = \delta_0$, i.e. the limiting process is a Cox process.

The proof of this theorem is based on Gaussian comparison techniques.

Appendix: Point Processes

Here we provide some basic background on point processes and in particular Poisson point processes. For more details, see [24, 37, 48].

Definition and Basic Properties

Point processes are designed to describe the probabilistic structure of point sets in some metric space. The key idea is to associate to a collection of points a *point measure*.

Let us first consider a single point x. We consider the usual Borel-sigma algebra, $\mathfrak{B} \equiv \mathfrak{B}(\mathbb{R}^d)$, of $\mathbb{R}^d$, that is generated by the open sets in the Euclidean topology of $\mathbb{R}^d$. Given $x \in \mathbb{R}^d$, we define the Dirac measure, δ_x, such that, for any Borel set $A \in \mathfrak{B}$,

$$\delta_x(A) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A. \end{cases} \tag{260}$$

A point measure is now a measure, μ, on $\mathbb{R}^d$, such that there exists a countable collection of points, $\{x_i \in \mathbb{R}^d, i \in \mathbb{N}\}$, such that

$$\mu = \sum_{i=1}^{\infty} \delta_{x_i} \tag{261}$$

and, if K is compact, then $\mu(K) < \infty$.

Note that the points x_i need not be all distinct. The set $S_\mu \equiv \{x \in \mathbb{R}^d : \mu(x) \neq 0\}$ is called the support of μ. A point measure such that for all $x \in \mathbb{R}^d$, $\mu(x) \leq 1$ is called simple.

We denote by $M_p(\mathbb{R}^d)$ the set of all point measures on $\mathbb{R}^d$. We equip this set with the sigma-algebra $\mathscr{M}_p(\mathbb{R}^d)$, the smallest sigma algebra that contains all subsets of $M_p(\mathbb{R}^d)$ of the form $\{\mu \in M_m(\mathbb{R}^d) : \mu(F) \in B\}$, where $F \in \mathfrak{B}(\mathbb{R}^d)$ and $B \in \mathfrak{B}([0,\infty))$. $\mathscr{M}_p(\mathbb{R}^d)$ is also characterised by saying that it is the smallest sigma-algebra that makes the evaluation maps, $\mu \to \mu(F)$, measurable for all Borel sets $F \in \mathfrak{B}(\mathbb{R}^d)$.

A *point process*, N, is a random variable taking values in $M_p(\mathbb{R}^d)$, i.e. a measurable map, $N : (\Omega, \mathfrak{F}; \mathbb{P}) \to M_p(\mathbb{R}^d)$, from a probability space to the space of point measures.

This looks very fancy, but in reality things are quite down-to-earth:

Proposition 6.1 *N is a point process, if and only if the map $N(\cdot, F) : \omega \to N(\omega, F)$, is measurable from $(\Omega, \mathfrak{F}) \to \mathfrak{B}([0,\infty))$, for any Borel set F, i.e. if $N(F)$ is a real random variable.*

Proof Let us first prove necessity, which should be obvious. In fact, since $\omega \to N(\omega, \cdot)$ is measurable into $(M_p(\mathbb{R}^d), \mathscr{M}_p(\mathbb{R}^p))$, and $\mu \to \mu(F)$ is measurable from this space into $(\mathbb{R}_+, \mathfrak{B}(\mathbb{R}_+))$, the composition of these maps is also measurable.

Next we prove sufficiency. Define the set

$$\mathfrak{G} \equiv \{A \in \mathscr{M}_p(\mathbb{R}^d) : N^{-1}A \in \mathfrak{F}\} \tag{262}$$

This set is a sigma-algebra and N is measurable from $(\Omega, \mathfrak{F}) \to (M_p(\mathbb{R}^d), \mathfrak{G})$ by definition. But $\mathfrak{G}$ contains all sets of the form $\{\mu \in M_p(\mathbb{R}^d) : \mu(F) \in B\}$, since

$$N^{-1}\{\mu \in M_p(\mathbb{R}^d) : \mu(F) \in B\} = \{\omega \in \Omega : N(\omega, F) \in B\} \in \mathfrak{F}, \tag{263}$$

since $N(\cdot, F)$ is measurable. Thus $\mathfrak{G} \supset \mathscr{M}_p(\mathbb{R}^d)$, and N is measurable a fortiori as a map from the smaller sigma-algebra. □

We will have need to find criteria for convergence of point processes. For this we recall some notions of measure theory. If $\mathfrak{B}$ is a Borel-sigma algebra, of a metric space E, then $\mathscr{T} \subset \mathfrak{B}$ is called a Π-system, if $\mathscr{T}$ is closed under finite intersections; $\mathfrak{G} \subset \mathfrak{B}$ is called a λ-system, or a sigma-additive class, if

(i) $E \in \mathfrak{G}$,

(ii) If $A, B \in \mathfrak{G}$, and $A \supset B$, then $A \setminus B \in \mathfrak{G}$,
(iii) If $A_n \in \mathfrak{G}$ and $A_n \subset A_{n+1}$, then $\lim_{n\uparrow\infty} A_n \in \mathfrak{G}$.

The following useful observation is called Dynkin's theorem.

Theorem 6.2 *If $\mathscr{T}$ is a Π-system and $\mathfrak{G}$ is a λ-system, then $\mathfrak{G} \supset \mathscr{T}$ implies that $\mathfrak{G}$ contains the smallest sigma-algebra containing $\mathscr{T}$.*

The most useful application of Dynkin's theorem is the observation that, if two probability measures are equal on a Π-system that generates the sigma-algebra, then they are equal on the sigma-algebra (since the set on which the two measures coincide forms a λ-system containing $\mathscr{T}$).

As a consequence we can further restrict the criteria to be verified for N to be a Point process. In particular, we can restrict the class of F's for which $N(\cdot, F)$ need to be measurable to bounded rectangles.

Proposition 6.3 *Suppose that $\mathscr{T}$ are relatively compact sets in $\mathfrak{B}$ satisfying*

(i) $\mathscr{T}$ is a Π-system,
(ii) The smallest sigma-algebra containing $\mathscr{T}$ is $\mathfrak{B}$,
(iii) Either, there exists $E_n \in \mathscr{T}$, such that $E_n \uparrow E$, or there exists a partition, $\{E_n\}$, of E with $\cup_n E_n = E$, with $E_n \subset \mathscr{T}$.

Then N is a point process on $(\Omega, \mathfrak{F})$ in $(E, \mathfrak{B})$, if and only if the map $N(\cdot, I) : \omega \to N(\omega, I)$ is measurable for any $I \in \mathscr{T}$.

Corollary 6.4 *Let $\mathscr{T}$ satisfy the hypothesis of Proposition 6.3 and set*

$$\mathfrak{G} \equiv \left\{ \{\mu : \mu(I_j) = n_j, 1 \le j \le k\}, k \in \mathbb{N}, I_j \in \mathscr{T}, n_j \ge 0 \right\}. \tag{264}$$

Then the smallest sigma-algebra containing $\mathfrak{G}$ is $\mathscr{M}_p(\mathbb{R}^d)$ and $\mathfrak{G}$ is a Π-system.

Next we show that the law, P_N, of a point process is determined by the law of the collections of random variables $N(F_n)$, $F_n \in \mathfrak{B}(\mathbb{R}^d)$.

Proposition 6.5 *Let N be a point process in $(\mathbb{R}^d, \mathfrak{B}(\mathbb{R}^d)$ and suppose that $\mathscr{T}$ is as in Proposition 6.3. Define the mass functions*

$$P_{I_1,\ldots,I_k}(n_1, \ldots, n_k) \equiv \mathbb{P}\left[N(I_j) = n_j, \forall 1 \le j \le k\right] \tag{265}$$

for $I_j \in \mathscr{T}$, $n_j \ge 0$. Then P_N is uniquely determined by the collection

$$\{P_{I_1,\ldots,I_k}, k \in \mathbb{N}, I_j \in \mathscr{T}\} \tag{266}$$

We need some further notions. First, if N_1, N_2 are point processes, we say that they are independent, if and only if, for any collection $F_j \in \mathfrak{B}$, $G_j \in \mathfrak{B}$, the vectors

$$(N_1(F_j), 1 \le j \le k) \quad \text{and} \quad (N_2(G_j), 1 \le j \le \ell) \tag{267}$$

are independent random vectors.

The *intensity measure*, λ, of a point process N is defined as

$$\lambda(F) \equiv \mathbb{E}N(F) = \int_{M_p(\mathbb{R}^d)} \mu(F) P_N(\mathrm{d}\mu) \tag{268}$$

for $F \in \mathfrak{B}$.

For measurable functions $f : \mathbb{R}^d \to \mathbb{R}_+$, we define

$$N(\omega, f) \equiv \int_{\mathbb{R}^d} f(x) N(\omega, \mathrm{d}x) \tag{269}$$

Then $N(\cdot, f)$ is a random variable. We have that

$$\mathbb{E}N(f) = \lambda(f) = \int_{\mathbb{R}^d} f(x)\lambda(\mathrm{d}x). \tag{270}$$

Laplace Functionals

If Q is a probability measure on $(M_p, \mathscr{M}_p)$, the *Laplace transform* of Q is a map, ψ from non-negative Borel functions on $\mathbb{R}^d$ to $\mathbb{R}_+$, defined as

$$\psi(f) \equiv \int_{M_p} \exp\left(-\int_{R^d} f(x)\mu(\mathrm{d}x)\right) Q(\mathrm{d}\mu). \tag{271}$$

If N is a point process, the *Laplace functional* of N is

$$\begin{aligned} \psi_N(f) \equiv \mathbb{E}\mathrm{e}^{-N(f)} &= \int \mathrm{e}^{-N(\omega,f)} \mathbb{P}(\mathrm{d}\omega) \\ &= \int_{M_p} \exp\left(-\int_{R^d} f(x)\mu(\mathrm{d}x)\right) P_N(\mathrm{d}\mu) \end{aligned} \tag{272}$$

Proposition 6.6 *The Laplace functional, ψ_N, of a point process, N, determines N uniquely.*

Proof For $k \geq 1$, and $F_1, \ldots, F_k \in \mathfrak{B}$, $c_1, \ldots, c_k \geq 0$, let $f = \sum_{i=1}^k c_i \mathbb{1}_{F_i}(x)$. Then

$$N(\omega, f) = \sum_{i=1}^k c_i N(\omega, F_i) \tag{273}$$

and

$$\psi_N(f) = \mathbb{E}\exp\left(-\sum_{i=1}^k c_i N(F_i)\right). \tag{274}$$

This is the Laplace transform of the vector $(N(F_i), 1 \leq i \leq k)$, that determines uniquely its law. Hence the proposition follows from Proposition 6.5 □

Poisson Point Processes

The most important class of point processes for our purposes will be Poisson point processes.

Definition 6.7 Let λ be a σ-finite, positive measure on $\mathbb{R}^d$. Then a point process, N, is called a Poisson point process with intensity measure λ ($PPP(\lambda)$), if

(i) For any $F \in \mathfrak{B}(\mathbb{R}^d)$, and $k \in \mathbb{N}$,

$$\mathbb{P}[N(F) = k] = \begin{cases} e^{-\lambda(F)} \frac{(\lambda(F))^k}{k!}, & \text{if } \lambda(F) < \infty \\ 0, & \text{if } \lambda(F) = \infty, \end{cases} \tag{275}$$

(ii) If $F, G \in \mathfrak{B}$ are disjoint sets, then $N(F)$ and $N(G)$ are independent random variables.

In the next theorem we will assert the existence of a Poisson point process with any desired intensity measure. In the proof we will give an explicit construction of such a process.

Proposition 6.8

(i) $PPP(\lambda)$ exists and its law is uniquely determined by the requirements of the definition.
(ii) The Laplace functional of $PPP(\lambda)$ is given, for $f \geq 0$, by

$$\Psi_N(f) = \exp\left(\int_{\mathbb{R}^d} (e^{-f(x)} - 1)\lambda(dx)\right). \tag{276}$$

Proof Since we know that the Laplace functional determines a point process, in order to prove that the conditions of the definition uniquely determine the $PPP(\lambda)$, we show that they determine the form (276) of the Laplace functional. Thus suppose that N is a $PPP(\lambda)$. Let $f = c\mathbb{1}_F$. Then

$$\begin{aligned} \Psi_N(f) &= \mathbb{E}\exp(-N(f)) = \mathbb{E}\exp(-cN(F)) \\ &= \sum_{k=0}^{\infty} e^{-ck} e^{-\lambda(F)} \frac{(\lambda(F))^k}{k!} = e^{(e^{-c}-1)\lambda(F)} \\ &= \exp\left(\int (e^{-f(x)} -)\lambda(dx)\right), \end{aligned} \tag{277}$$

which is the desired form. Next, if F_i are disjoint, and $f = \sum_{i=1}^k c_i \mathbb{1}_{F-i}$, it is straightforward to see that

$$\Psi_N(f) = \mathbb{E}\exp\left(-\sum_{i=1}^k c_i N(F_i)\right) = \prod_{i=^1}^k \mathbb{E}\exp\left(-c_i N(F_i)\right) \tag{278}$$

due to the independence assumption (ii); a simple calculations shows that this yields again the desired form. Finally, for general f, we can choose a sequence, f_n, of the form considered, such that $f_n \uparrow f$. By monotone convergence then $N(f_n) \uparrow N(f)$. On the other hand, since $\mathrm{e}^{-N(g)} \leq 1$, we get from dominated convergence that

$$\Psi_N(f_n) = \mathbb{E}\mathrm{e}^{-N(f_n)} \to \mathbb{E}\mathrm{e}^{-N(f)} = \Psi_N(f). \tag{279}$$

But, since $1 - \mathrm{e}^{-f_n(x)} \uparrow 1 - \mathrm{e}^{-f(x)}$, and monotone convergence gives once more

$$\Psi_N(f_n) = \exp\left(\int (1 - \mathrm{e}^{-f_n(x)})\lambda(\mathrm{d}x)\right) \uparrow \exp\left(\int (1 - \mathrm{e}^{-f(x)})\lambda(\mathrm{d}x)\right) \tag{280}$$

On the other hand, given the form of the Laplace functional, it is trivial to verify that the conditions of the definition hold, by choosing suitable functions f.

Finally we turn to the *construction* of $PPP(\lambda)$. Let us first consider the case $\lambda(\mathbb{R}^d) < \infty$. Then construct

(i) A Poisson random variable, τ, of parameter $\lambda(\mathbb{R}^d)$.
(ii) A family, X_i, $i \in \mathbb{N}$, of independent, $\mathbb{R}^d$- valued random variables with common distribution λ. This family is independent of τ.

Then set

$$N^* \equiv \sum_{i=1}^{\tau} \delta_{X_i} \tag{281}$$

It is not very hard to verify that N^* is a $PPP(\lambda)$.

To deal with the case when $\lambda(\mathbb{R}^d)$ is infinite, decompose λ into a countable sum of finite measures, λ_k, that are just the restriction of λ to a finite set F_k, where the F_k form a partition of $\mathbb{R}^d$. Then N^* is just the sum of independent $PPP(\lambda_k)$ N_k^*. □

Convergence of Point Processes

Before we turn to applications to extremal processes, we still have to discuss the notion of convergence of point processes. As point processes are probability distributions on the space of point measures, we will naturally think about *weak convergence*. This means that we will say that a sequence of point processes, N_n,

converges weakly to a point process, N, if for all continuous functions, f, on the space of point measures,

$$\mathbb{E}f(N_n) \to \mathbb{E}f(N). \tag{282}$$

However, to understand what this means, we must discuss what continuous functions on the space of point measures are, i.e. we must introduce a topology on the set of point measures. The appropriate topology for our purposes will be that of *vague convergence*.

Vague Convergence

We consider the space $\mathbb{R}^d$ equipped with its natural Euclidean metric. Clearly $\mathbb{R}^d$ is a complete, separable metric space. We will denote by $C_0(\mathbb{R}^d)$ the set of continuous real-valued functions on $\mathbb{R}^d$ that have compact support; $C_0^+(\mathbb{R}^n)$ denotes the subset of non-negative functions. We consider $\mathscr{M}_+(\mathbb{R}^d)$ the set of all σ-finite, positive measures on $(\mathbb{R}^d, \mathfrak{B}(\mathbb{R}^d))$. We denote by $\mathscr{M}_+(\mathbb{R}^d)$ the smallest sigma-algebra of subsets of $M_+(\mathbb{R}^d)$ that makes the maps $m \to m(f)$ measurable for all $f \in C_0^+(\mathbb{R}^d)$.

We will say that a sequence of measures, $\mu_n \in M_+(\mathbb{R}^d)$ converges *vaguely* to a measure $\mu \in M_+(\mathbb{R}^d)$, if, for all $f \in C_0^+(\mathbb{R}^d)$,

$$\mu_n(f) \to \mu(f) \tag{283}$$

Note that for this topology, typical open neighbourhoods are of the form

$$B_{f_1,\ldots,f_k,\epsilon}(\mu) \equiv \{\nu \in M_+(\mathbb{R}^d) : \forall_{i=1}^k |\nu(f_i) - \mu(f_i)| < \epsilon\}, \tag{284}$$

i.e. to test the closeness of two measures, we test it on their expectations on finite collections of continuous, positive functions with compact support. Given this topology, on can of course define the corresponding Borel sigma algebra, $\mathfrak{B}(M_+(\mathbb{R}^d))$, which (fortunately) turns out to coincide with the sigma algebra $\mathscr{M}_+(\mathbb{R}^d)$ introduced before.

The following properties of vague convergence are useful.

Proposition 6.9 *Let μ_n, $n \in \mathbb{N}$ be in $M_+(\mathbb{R}^d)$. Then the following statements are equivalent:*

(i) μ_n *converges vaguely to* μ, $\mu_n \xrightarrow{v} \mu$.
(ii) $\mu_n(B) \to \mu(B)$ *for all relatively compact sets, B, such that* $\mu(\partial B) = 0$.
(iii) $\limsup_{n\uparrow\infty} \mu_n(K) \leq \mu(K)$ *and* $\limsup_{n\uparrow\infty} \mu_n(G) \geq \mu(G)$, *for all compact K, and all open, relatively compact G.*

Weak Convergence

Having established the space of σ-finite measures as a complete, separable metric space, we can think of weak convergence of probability measures on this space just as if we were working on an Euclidean space.

Definition 6.10 A sequence of point processes N_n converges *weakly* with respect to the vague topology to a point process, N, iff for all functions $F : \mathscr{M}_+(\mathbb{R}^d) \to \mathbb{R}$ that are continuous with respect to the vague topology,

$$\lim_{n\uparrow\infty} \mathbb{E}F(N_n) = \mathbb{E}F(N). \tag{285}$$

An important convergence criterion is given by the following theorem (see [24, Chap. 11]).

Theorem 6.11 *A sequence of point processes N_n converges weakly to a point process N, if for all positive continuous functions, ϕ, with compact support,*

$$\lim_{n\uparrow\infty} \Psi_{N_n}(\phi) = \Psi_N(\phi). \tag{286}$$

Acknowledgements I am deeply grateful to my collaborators on the matters of these lectures, Louis-Pierre Arguin, Nicola Kistler, Irina Kurkova, and, most recently, Lisa Hartung. They did the bulk of the work, and without them none this would have been achieved. I also thank Jiří Černý, Lisa Hartung, and an anonymous referee for pointing out various mistakes and suggesting improvement. I thank Marek Biskup, Jiří Černý, and Roman Kotecký for organising the excellent school in Prague and for the opportunity to present these lectures there.

This work is partially supported through the German Research Foundation in the Collaborative Research Center 1060 "The Mathematics of Emergent Effects", the Priority Programme 1590 "Probabilistic Structures in Evolution", the Hausdorff Center for Mathematics (HCM), and the Cluster of Excellence "ImmunoSensation" at Bonn University.

References

1. E. Aïdékon, Convergence in law of the minimum of a branching random walk. Ann. Probab. **41**, 1362–1426 (2013)
2. E. Aïdékon, J. Berestycki, E. Brunet, Z. Shi, Branching Brownian motion seen from its tip. Probab. Theory Relat. Fields **157**, 405–451 (2013)
3. L.-P. Arguin, A. Bovier, N. Kistler, Genealogy of extremal particles of branching Brownian motion. Commun. Pure Appl. Math. **64**(12), 1647–1676 (2011)
4. L.-P. Arguin, A. Bovier, N. Kistler, Poissonian statistics in the extremal process of branching Brownian motion. Ann. Appl. Probab. **22**(4), 1693–1711 (2012)
5. L.-P. Arguin, A. Bovier, N. Kistler, An ergodic theorem for the frontier of branching Brownian motion. Electron. J. Probab. **18**(53), 25 (2013)
6. L.-P. Arguin, A. Bovier, N. Kistler, The extremal process of branching Brownian motion. Probab. Theory Relat. Fields **157**, 535–574 (2013)
7. L.-P. Arguin, A. Bovier, N. Kistler, An ergodic theorem for the extremal process of branching Brownian motion. Ann. Inst. Henri Poincaré Probab. Stat. **51**, 557–569 (2015)

8. K.B. Athreya, P.E. Ney, *Branching Processes*. Die Grundlehren der mathematischen Wissenschaften, Band 196 (Springer, New York, 1972)
9. G. Ben Arous, A. Kuptsov, REM universality for random Hamiltonians, in *Spin Glasses: Statics and Dynamics*. Progress in Probability, vol. 62 (Birkhäuser, Basel, 2009), pp. 45–84
10. G. Ben Arous, V. Gayrard, A. Kuptsov, A new REM conjecture, in *In and Out of Equilibrium. 2*. Progress in Probability, vol. 60 (Birkhäuser, Basel, 2008), pp. 59–96
11. M. Biskup, O. Louidor, Extreme local extrema of two-dimensional discrete Gaussian free field. ArXiv e-prints (June 2013)
12. A. Bovier, *Statistical Mechanics of Disordered Systems*. Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press, Cambridge, 2006)
13. A. Bovier, L. Hartung, The extremal process of two-speed branching Brownian motion. Electron. J. Probab. **19**(18), 1–28 (2014)
14. A. Bovier, L. Hartung, Variable speed branching Brownian motion 1. Extremal processes in the weak correlation regime. ALEA, Lat. Am. J. Probab. Math. Stat. **12**, 261–291 (2015)
15. A. Bovier, I. Kurkova, Derrida's generalised random energy models. I. Models with finitely many hierarchies. Ann. Inst. Henri Poincaré Probab. Stat. **40**(4), 439–480 (2004)
16. A. Bovier, I. Kurkova, Derrida's generalized random energy models. II. Models with continuous hierarchies. Ann. Inst. Henri Poincaré Probab. Stat. **40**(4), 481–495 (2004)
17. A. Bovier, I. Kurkova, M. Löwe, Fluctuations of the free energy in the REM and the p-spin SK models. Ann. Probab. **30**(2), 605–651 (2002)
18. M.D. Bramson, Maximal displacement of branching Brownian motion. Commun. Pure Appl. Math. **31**(5), 531–581 (1978)
19. M. Bramson, Convergence of solutions of the Kolmogorov equation to travelling waves. Mem. Am. Math. Soc. **44**(285), iv+190 (1983)
20. M. Bramson, O. Zeitouni, Tightness of the recentered maximum of the two-dimensional discrete Gaussian free field. Commun. Pure Appl. Math. **65**(1), 1–20 (2012)
21. M. Bramson, J. Ding, O. Zeitouni, Convergence in law of the maximum of the two-dimensional discrete Gaussian free field (Jan 2013). http://arXiv.org/abs/1301.6669
22. B. Chauvin, A. Rouault, KPP equation and supercritical branching Brownian motion in the subcritical speed area. Application to spatial trees. Probab. Theory Relat. Fields **80**(2), 299–314 (1988)
23. B. Chauvin, A. Rouault, Supercritical branching Brownian motion and K-P-P equation in the critical speed-area. Math. Nachr. **149**, 41–59 (1990)
24. D.J. Daley, D. Vere-Jones, *An Introduction to the Theory of Point Processes*. Springer Series in Statistics (Springer, New York, 1988)
25. B. Derrida, H. Spohn, Polymers on disordered trees, spin glasses, and traveling waves. J. Stat. Phys. **51**(5–6), 817–840 (1988). New directions in statistical mechanics (Santa Barbara, CA, 1987)
26. J. Ding, Exponential and double exponential tails for maximum of two-dimensional discrete gaussian free field. Probab. Theory Relat. Fields **157**, 285–299 (2013)
27. M. Fang, O. Zeitouni, Branching random walks in time inhomogeneous environments. Electron. J. Probab. **17**(67), 18 (2012)
28. M. Fang, O. Zeitouni, Slowdown for time inhomogeneous branching Brownian motion. J. Stat. Phys. **149**(1), 1–9 (2012)
29. R. Fisher, The wave of advance of advantageous genes. Ann. Eugen. **7**, 355–369 (1937)
30. E. Gardner, B. Derrida, Magnetic properties and function $q(x)$ of the generalised random energy model. J. Phys. C **19**, 5783–5798 (1986)
31. E. Gardner, B. Derrida, Solution of the generalised random energy model. J. Phys. C **19**, 2253–2274 (1986)
32. E. Gardner, B. Derrida, The probability distribution of the partition function of the random energy model. J. Phys. A **22**(12), 1975–1981 (1989)
33. J.-B. Gouéré, Branching Brownian motion seen from its left-most particle. ArXiv e-prints (May 2013)

34. N. Ikeda, M. Nagasawa, S. Watanabe, Markov branching processes I. J. Math. Kyoto Univ. **8**, 233–278 (1968)
35. N. Ikeda, M. Nagasawa, S. Watanabe, Markov branching processes II. J. Math. Kyoto Univ. **8**, 365–410 (1968)
36. N. Ikeda, M. Nagasawa, S. Watanabe, Markov branching processes III. J. Math. Kyoto Univ. **9**, 95–160 (1969)
37. O. Kallenberg, *Random Measures* (Akademie, Berlin, 1983)
38. N. Kistler, Derrida's random energy models from spin glasses to the extremes of correlated random fields, in *Correlated Random Systems: Five Different Methods*, V. Gayrard, N. Kistler, editors. Lecture Notes in Mathematics, vol. 2143 (Springer, Berlin, 2015), pp. 71–120
39. A. Kolmogorov, I. Petrovsky, N. Piscounov, Etude de l' équation de la diffusion avec croissance de la quantité de matière et son application à un problème biologique. Moscou Universitet Bull. Math. **1**, 1–25 (1937)
40. S.P. Lalley, T. Sellke, A conditional limit theorem for the frontier of a branching Brownian motion. Ann. Probab. **15**(3), 1052–1061 (1987)
41. M. Leadbetter, G. Lindgren, H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*. Springer Series in Statistics (Springer, New York, 1983)
42. P. Maillard, O. Zeitouni, Slowdown in branching Brownian motion with inhomogeneous variance. ArXiv e-prints (July 2013)
43. B. Mallein, Maximal displacement of a branching random walk in time-inhomogeneous environment. ArXiv e-prints (July 2013)
44. H.P. McKean, Application of Brownian motion to the equation of Kolmogorov-Petrovskii-Piskunov. Commun. Pure Appl. Math. **28**(3), 323–331 (1975)
45. C. Newman, D. Stein, *Spin Glasses and Complexity* (Princeton University Press, Princeton, 2013)
46. J. Nolen, J.-M. Roquejoffre, L. Ryzhik, Power-like delay in time inhomogeneous fisher-kpp equations. Commun. Partial Differ. Equ. **40**, 475–505 (2015)
47. D. Panchenko, *The Sherrington-Kirkpatrick Model* (Springer, New York, 2013)
48. S. Resnick, *Extreme Values, Regular Variation, and Point Processes*. Applied Probability. A Series of the Applied Probability Trust, vol. 4 (Springer, New York, 1987)
49. M.I. Roberts, A simple path to asymptotics for the frontier of a branching Brownian motion. Ann. Probab. **41**(5), 3518–3541 (2013)
50. M. Talagrand, *Spin Glasses: A Challenge for Mathematicians*. Ergebnisse der Mathematik und ihrer Grenzgebiete. (3) [Results in Mathematics and Related Areas (3)], vol. 46 (Springer, Berlin, 2003)
51. O. Zeitouni, Branching random walks and Gaussian free fields. Lecture notes (2013)

The Renormalization Group and Self-avoiding Walk

David Brydges

1 Introduction

I am grateful and honoured to be given this opportunity to give an introduction to the renormalisation group. It is based on recent work with Gordon Slade, and Roland Bauerschmidt. These notes are intended to be interesting for mathematicians. No knowledge of physics is assumed, but our topic began life in physics and so I begin with a review of this background. This review contains references that cannot be understood without a background in physics. I include them for historical reasons, not because they are required reading. However it would make me very happy if the mathematical developments in these notes help someone in my audience understand physics better.

The renormalisation group (RG) first appeared in quantum electrodynamics with work by Stueckelberg and Petermann [45] and Gell-Mann and Low [30]. Quantum electrodynamics is the quantum theory that extends the classical theory of electromagnetism. Classical electromagnetism is the combination of Maxwell's partial differential equations for the electromagnetic field and the Lorentz equation that describes the force experienced by a charged particle moving in an electromagnetic field. The force is proportional to the charge and charge is an example of a *coupling constant*. Maxwell's equations are invariant under a common rescaling of space and time and the charge does not change under such rescaling. However scale invariance of charge as a coupling constant does not survive in quantum electrodynamics. These founding papers were statements about how the charge coupling constant transforms under rescaling. Later the insight of these papers was put in a more useful form, called the Callan-Symanzik equations, [23, 48, 49]. These equations

D. Brydges (✉)
Department of Mathematics, The University of British Columbia, 1984 Mathematics Road, Vancouver, BC, Canada V6T 1Z2
e-mail: db5d@math.ubc.ca

M. Biskup et al. (eds.), *Random Walks, Random Fields, and Disordered Systems*, Lecture Notes in Mathematics 2144, DOI 10.1007/978-3-319-19339-7_2

are a set of coupled differential equations for the flow of the coupling constants under scaling. In these lectures we will see how it can be that coupling constants depend on scale.

A quite different line of thought started with Symanzik [46] who began the study of quantum field theory on $\mathbb{R}^d$ with the euclidean metric instead of the Minkowski metric prescribed by Nature. He realised that these, despite the name "quantum fields", are random fields in the standard sense of probability. By formulating this idea also for fields on a lattice such as $\mathbb{Z}^d$ these theories were connected with the scaling limits of lattice spin systems such as the Ising model. This was exploited in a deep way in [31, 32, 42, 43] where euclidean quantum field theory is studied as the limit of classical spin systems and, in particular, correlation inequalities are used. In the other direction, from euclidean field theory to lattice spin systems, Wilson and Fisher [53] and Wilson [51] showed that the renormalisation group and the Feynman expansion of quantum field theory can systematically calculate critical exponents in statistical mechanics. In these lectures we will be explaining some parts of this method.

Other authors [40, 57] expressed the same ideas for calculating critical exponents in terms of the Callan-Symanzik equations. In particular, using an idea of de Gennes [24], Brezin et al. [13] and Duplantier [26] used the Callan-Symanzik equations to calculate exponents for self-avoiding walk in three and four dimensions. These authors are assuming that the scaling limit of a lattice spin system in four dimensions exists as a continuum random field with the scaling properties prescribed by the Callan-Symanzik equations and that the coefficients in these equations are given by perturbation theory. They give very efficient calculations of critical exponents based on these assumptions.

In our work, [10] and the five papers that it references, we avoid such assumptions by following the work of Ken Wilson, who invented a more detailed form of the renormalisation group that applies to statistical mechanical models on lattices. It is in fact a program to prove the existence of scaling limits as quantum field theories. We only make part of that program mathematically complete, just enough to prove that the susceptibility of a walk with weak self-repulsion has a $\log^{\frac{1}{4}}$ correction. If you are seeing this log correction for the first time then it will not seem very interesting but think of it as a signal: Wilson's program is systematic, in the same sense that calculus is systematic, and it is a calculus for critical statistical mechanics and scaling limits. It is more complicated than you will like, but most profound ideas have not been easy at first. I regard the influential lecture course [54] as still one of the best places to appreciate the ideas and scope of this program. Perhaps even better and also amusing in places, is his Nobel prize lecture [52].

Wilson and Kogut start with a random field ϕ defined on a lattice $\mathbb{Z}^d$. They "integrate out" fluctuations in the random field up to some chosen length scale L. One intuitively attractive way to do this is called the *block spin* method and it is defined as follows. The lattice is partitioned into disjoint cubes of side L and the random field is conditioned on its empirical averages over the cubes. The centres of the cubes are a new coarser lattice of spacing L and the empirical averages are a new

random field on this coarser lattice. The coarse lattice is scaled so that it becomes the unit lattice $\mathbb{Z}^d$. Thus the combination of conditioning followed by rescaling is a map RG taking the space $\mathscr{M}$ of all probability laws for a lattice random field into itself. Thus RG can be iterated and can be analysed as a dynamical system.

Suppose RG has a fixed point. Then one can define the stable set S to be the set of all points in $\mathscr{M}$ that have orbits under RG that converge to the fixed point. This S is invariant under RG. The Wilson picture is that points on S are probability distributions for random fields that are critical, for example an Ising model at its critical temperature. The set S is the universality class of all random fields whose scaling limit is the probability distribution represented by the fixed point. This special probability distribution is expected to have more symmetries than the distributions represented by points on S. For example it will be scale invariant. In fact in some cases it is conjectured to satisfy the axioms of euclidean quantum field theory. To be clearer, the fixed point is a probability distribution for a lattice random field, but in fact there will be a continuum generalised random field whose empirical averages over cubes centred on the lattice points are the random variables for the fixed point model. It is this continuum field that will be scale invariant and may satisfy the axioms of euclidean quantum field theory. One of the most important such fixed point distributions is called the massless free field. Wilson argued that in more than two dimensions the stable set for this fixed point has finite codimension: you have to carefully select values of finitely many parameters to be a critical model on the stable set for this fixed point. In four and more dimensions, and for even measures, Wilson asserts that the codimension is 2. One of the parameters is a parameter of the massless free field analogous to the variance of the normal distribution. As in the central limit theorem you have to normalise the empirical averages correctly in order that the scaling limit be a standard massless free field. Selecting the other parameter corresponds, for the Ising model, to choosing the temperature to be critical. In these lectures we will see this codimensionality of 2 for self-avoiding walk in four dimensions.

We implement the Wilson RG a little differently. Our definition of RG relies on the fact that the gaussian free field ϕ on the lattice $\mathbb{Z}^d$ is equal in distribution to a sum $\phi = \sum_j \zeta_j$ over scales j of independent gaussian fields ζ_j with a strong independence property: $\zeta_{j,x}$ and $\zeta_{j,y}$ are independent when the spatial points x and y are separated by at least $L^j/2$. This suggests an interesting open problem: characterise the gaussian random fields ϕ which in distribution can be written as such a sum. There is a nice proof that the lattice gaussian free field has such a decomposition in [7] along with references to the original constructions of such representations.

The first rigorous control of RG was achieved by Gawedzki and Kupiainen [28, 29]. Hara and Tasaki [33] used their methods to prove the existence of log corrections in the ϕ^4 lattice field theory. There is a different program with similar outcomes called the phase cell expansion described in [2]. Using the phase cell expansion Iagolnitzer and Magnen [34] considered a model that is roughly speaking an Edwards model in the continuum. They determine the asymptotics of the Greens function as one spatial argument tends to infinity. A nice feature of their approach is that it deals quite directly with walks whereas our method transforms self-avoiding

walk into massless free field language. Balaban has made the most far reaching accomplishments in the rigorous renormalisation group, for example in his series of papers on the classical lattice Heisenberg models that starts with [6].

At this time the scaling limit of self-avoiding walk in three dimensions is, for mathematicians, a complete mystery. It is believed that the scaling limit is a non-gaussian fixed point for RG and we hardly understand these at all. However in [57] you can read the remarkable non rigorous progress that started with ideas of Wilson and Fisher. With $\epsilon := 4-d$, critical exponents are obtained as expansions in powers of ϵ. The first few terms of these expansions can be calculated and results in accurate agreement with simulations of three dimensional self-avoiding walk are obtained by setting $\epsilon = 1$. A step in this direction of varying a dimension-like parameter has been made in [41]. Related to this is [1] which, for a ϕ^4 theory in three dimensions, constructs a *complete* renormalization group trajectory that at one end converges to a gaussian and at the other end to a non-gaussian fixed point.

2 The Lattice Edwards Model

We start with a continuous time random walk $\{X_t : t \geq 0\}$ on the euclidean lattice $\mathbb{Z}^d$ or on the torus $\Lambda \equiv (\mathbb{Z}/R\mathbb{Z})^d$ where R is a positive integer. The side or period R of Λ is chosen from the geometric sequence $L^N, N = 1, 2, \ldots$, where $L \geq 2$ is an integer. The (negative) generator of the walk is the lattice Laplacian $-\Delta$, acting on bounded functions f defined on $\mathbb{Z}^d$ or on Λ. It is given by

$$(-\Delta f)_x := \sum_{y \sim x} (f_x - f_y), \tag{1}$$

where $y \sim x$ means that y is a nearest neighbour to x. On the torus Λ every lattice point has $2d$ neighbours because there is no boundary. We use $\mathbb{P}_a$ to denote the law of the random walk starting at the vertex $a \in \Lambda$ and E_a is the corresponding expectation. In addition, we consider the random walk that gets killed with a rate $m^2 \geq 0$, i.e. the process on either $\mathbb{Z}^d \cup \star$ or $\Lambda \cup \star$ whose (negative) generator is given by the following block matrix

$$\begin{array}{c} \\ \Lambda \\ \star \end{array} \begin{array}{c} \begin{array}{cc} \Lambda & \star \end{array} \\ \begin{pmatrix} m^2 - \Delta, & -m^2 \\ 0, & 0 \end{pmatrix} \end{array}$$

The state $*$ is called the *cemetery*. The dynamics is easy to describe: the walk waits for a random time with distribution $\mathrm{Exp}(m^2 + 2d)$ and then jumps to a nearest neighbour with probability $1/(m^2 + 2d)$ or to the state $\star$ with probability $m^2/(m^2 + 2d)$. Further, let us denote by $\zeta := \inf\{t \geq 0 : X_t = \star\}$ the killing time, i.e. the first

hitting time of $\star$. For any $x \in \Lambda$ we define the local time spent at x by

$$L_x := \int \mathbb{1}_{\{X_s = x\}} \, \mathrm{d}s, \tag{2}$$

where the integral is over $[0, \infty)$. The time spent in self-intersection is defined to be

$$\iint \mathbb{1}_{\{X_s = X_t \neq *\}} \, \mathrm{d}s \, \mathrm{d}t \tag{3}$$

and by (2),

$$\iint \mathbb{1}_{\{X_s = X_t \neq *\}} \, \mathrm{d}s \, \mathrm{d}t \; = \; \sum_{x \in \Lambda} L_x^2. \tag{4}$$

We use $\mathbb{P}_a^{(m)}$ for the law of random walk starting from a and with killing rate m^2 and $\mathrm{E}_a^{(m)}$ for the corresponding expectation. Notice that for any function F of the local times we have

$$\mathrm{E}_a^{(m)}\big[F(L)\big] \; = \; m^2 \int \mathrm{E}_a^{(0)}\Big[F(L^T)\Big] \, \mathrm{e}^{-m^2 T} \, \mathrm{d}T \tag{5}$$

with

$$L_x^T := \int_{[0,T]} \mathbb{1}_{\{X_s = x\}} \, \mathrm{d}s. \tag{6}$$

Definition 2.1 (Susceptibility) For $g \geq 0$ and $\nu \in \mathbb{R}$ and random walk on $\mathbb{Z}^d$ starting in state a, define

$$\chi(g, \nu) := \int \mathrm{E}_a^{(0)}\Big[\mathrm{e}^{-g \sum_x (L_x^T)^2}\Big] \, \mathrm{e}^{-\nu T} \, \mathrm{d}T \tag{7}$$

with values in $(0, \infty]$. This is called the infinite volume susceptibility of the lattice Edwards model. We define the finite volume susceptibility $\chi_\Lambda(g, \nu) \in (0, \infty)$ by replacing $\mathbb{Z}^d$ by Λ so that the random walk is on Λ. In the infinite volume case we define the critical value of ν by

$$\nu_{\mathrm{c}} := -\inf\big\{\nu \in \mathbb{R} \colon \chi(g, \nu) < \infty\big\}. \tag{8}$$

In [10, Lemma A.1] we use standard subadditivity arguments to prove that $\chi(g, \nu)$ is finite if and only if ν is strictly greater than ν^{c}. The exponential term suppresses self-intersection. As g is taken larger the suppression becomes stronger, but at the same time the speed of the walk increases because the time it spends between jumps is also being suppressed. The various large g limits are discussed in

[22]. For $g = 0$ we have simple random walk for which $\nu_c = 0$. We will find that ν^c is negative for $g > 0$.

The finite volume susceptibility χ_Λ increases to the infinite volume susceptibility χ as Λ increases through tori of side-length $R = L^N$ with $N \to \infty$. This is proved in [10] by an argument based on wrapping an infinite volume walk onto a torus and noticing that this increases the number of self-intersections. Furthermore, all derivatives of χ_Λ with respect to ν converge to the corresponding derivatives of χ because this is a property of the one-sided Laplace transform and the susceptibility is a one-sided Laplace transform, with ν dual to time. This good property of the Laplace transform comes about because the one-sided Laplace transform is an analytic function of ν for ν in the half-plane $\Re\nu < \nu_c$. By the Vitali theorem pointwise convergence for real values of ν implies uniform convergence on compact subsets of the half-plane and from this it follows that derivatives converge.

The next theorem on the Edwards model is the focus for these notes. To state it we define $a(\epsilon) \sim b(\epsilon)$ to mean that $\lim_{\epsilon\downarrow 0} \frac{a(\epsilon)}{b(\epsilon)} = 1$. Let

$$\gamma = \frac{1}{4}. \tag{9}$$

We make this definition in order to be able to track the origin of the exponent in the logarithmic correction through these notes.

Theorem 2.2 (Part of Theorem 1.1 in [10]) *Let $d := 4$. For $g > 0$ sufficiently small, there exists $A_g > 0$ such that*

$$\chi\big(g, \nu_c(g) + \epsilon\big) \sim \frac{A_g}{\epsilon} \log^\gamma(1/\epsilon), \qquad \epsilon \downarrow 0. \tag{10}$$

For simple random walk, $\chi(0, \nu_c + \epsilon) = 1/\epsilon$ for each $\epsilon > 0$ (in all dimensions) so the log is an effect of $g > 0$. The conclusion of this theorem can be rewritten as

$$\int e^{-\nu_c T} \underbrace{E_a^{(0)}\Big[e^{-g\sum_x (L_x^T)^2}\Big]}_{c_T} \epsilon\, e^{-\epsilon T} dT \underset{\epsilon\downarrow 0}{\sim} A_g \log^{1/4}(1/\epsilon). \tag{11}$$

Thanks to the fact that $\epsilon\, e^{-\epsilon T} dT$ is a probability on $(0, \infty)$, this may be interpreted as a weak version of the (conjectural) statement that the quantity c_T marked above obeys

$$c_T \underset{T\to\infty}{\sim} A_g e^{\nu_c T} (\log T)^{1/4}. \tag{12}$$

Remark 2.3 For dimensions $d > 4$ there will not be a log correction. I do not know if exactly this result has been proved for the Edwards model but such results are proved for the standard discrete time self-avoiding walk by the lace expansion [8, 38]. I am not sure if anyone has considered lace expansions for continuous time

models like the Edwards model. It should be possible and they might even be neater than the standard discrete case.

3 The Free Field and Local Time

The free field is usually introduced as a real gaussian field. However we will consider a complex valued free field because the next section requires it and because the connections we are about to describe between gaussian fields and continuous time random walk are more general when complex fields are used. We only define gaussian fields on the torus Λ. Therefore we consider $\phi \in \mathbb{C}^{\Lambda}$, i.e. $\phi = \{\phi_x : x \in \Lambda\}$. We can also write $\phi_x = u_x + \mathrm{i}v_x$, where u and v are gaussian fields. A gaussian *complex valued* measure has the form

$$\mathrm{e}^{-(\phi, A\bar{\phi})} \det(A) \prod_{x \in \Lambda} \frac{\mathrm{d}u_x \mathrm{d}v_x}{\pi}, \tag{13}$$

where we are using the notation

$$(\phi, A\bar{\phi}) \ := \ \sum_{x,y \in \Lambda} \phi_x A_{xy} \bar{\phi}_y. \tag{14}$$

We say that A is *dissipative* if $\Re\mathfrak{e}\,(\phi, A\bar{\phi}) \geq m^2 \sum \phi_x \bar{\phi}_x$ for some $m^2 > 0$. The above gaussian complex measure exists for any dissipative matrix A. If $A_{xy} = A_{yx}$ and A is real, then the gaussian measure in a genuine probability distribution, in other words it is not complex valued. Furthermore, the gaussian fields u, v mentioned above are independent. We will mostly be working with a real symmetric A, but for the moment let us not make this assumption. Even though there is no probability measure we will still denote integration with respect to the complex measure by the symbol $\mathbb{E}$. The following formulas may surprise you a little. The first is an exercise. The second follows from the invariance of the gaussian density under the change of variables $\phi_x \mapsto e^{i\alpha}\phi_x$ and $\bar{\phi}_x \mapsto e^{-i\alpha}\bar{\phi}_x$ for all $x \in \Lambda$ with any real α. Recall that A is real. This symmetry is called gauge invariance.

$$\mathbb{E}\big[\bar{\phi}_x \phi_y\big] \ = \ (A^{-1})_{xy} \tag{15}$$

while

$$\mathbb{E}\big[\bar{\phi}_x \bar{\phi}_y\big] \ = \ 0 \ = \ \mathbb{E}\big[\phi_x \phi_y\big]. \tag{16}$$

Definition 3.1 If $A := m^2\,\mathrm{id} - \Delta$, then ϕ is called the (complex) free field with mass m.

The real field version of the following theorem has been called the *Dynkin Isomorphism* since [27], but Dynkin references two earlier papers. One of them is [19], where this isomorphism appears in Theorem 2.2, but stated in terms of the skeleton walk defined by a continuous time walk, and the second is [47], which is the first time a connection between local time of random walk and the square of a gaussian field was found. Since we and Symanzik are dropping out of sight in the probability literature on this theorem I am going to come back and haunt everyone who does not remember us! See [50] where the recent history of this theorem is discussed. It includes the relation between $\phi\bar{\phi}$ and the local time of loop soup discovered first in [47], but admittedly not stated very precisely.

A version [14, Proposition 3.1] of the following theorem is valid when A is a real matrix such that (1) all row sums are strictly positive, (2) A is dissipative. This is interesting because this allows A to be the generator of a nonsymmetric walk, but I do not want to pursue this here and will from now on assume that

$$A = m^2 \,\mathrm{id} - \Delta \tag{17}$$

for some $m^2 > 0$.

Theorem 3.2 *For bounded continuous* $F\colon \mathbb{R}_+^{\Lambda} \to \mathbb{R}$, $m^2 > 0$ *and* $a, b \in \Lambda$,

$$\mathbb{E}\big[F(\phi\bar{\phi})\,\bar{\phi}_a\phi_b\big] \;=\; m^{-2}\;\mathbb{E}\otimes \mathrm{E}_a\Big[F(\phi\bar{\phi}+L)\mathbb{1}_{\{X_\zeta-=b\}}\Big]. \tag{18}$$

Proof By the monotone class theorem and the linearity of both sides in the function F it suffices to check that both sides are equal for functions of the form $F(t) := \exp\{-\sum_{x\in\Lambda} w_x t_x\}$ where $w_x \geq 0$. Let W be the $\Lambda \times \Lambda$ diagonal matrix whose diagonal entries are $w_x, x \in \Lambda$. Then

$$F(\phi\bar{\phi}) = \exp\{-\sum_{x\in\Lambda} w_x\phi_x\bar{\phi}_x\} = \exp\{-(\phi, W\bar{\phi})\}. \tag{19}$$

Since this is gaussian the left hand side of (18), up to a normalisation, is the covariance of a gaussian measure whose density is the exponential of $-\big(\phi, (A + W)\bar{\phi}\big)$, where $A = m^2\,\mathrm{id} - \Delta$. Therefore the left hand side of (18) equals

$$(A + W)^{-1}_{ab}\,\mathbb{E}\big[F(\phi\bar{\phi})\big]. \tag{20}$$

Now consider the right hand side of (18): our special choice of F is such that $F(\phi\bar{\phi}+L) = F(\phi\bar{\phi})F(L)$ so the right hand side is

$$m^{-2}\;\mathbb{E}\Big[F(\phi\bar{\phi})\Big]\,\mathrm{E}_a\Big[F(L)\,\mathbb{1}_{\{X_\zeta-=b\}}\Big] \tag{21}$$

and we are reduced to checking that

$$(A+W)^{-1}_{ab} = m^{-2}\ \mathrm{E}_a\Big[F(L)\mathbb{1}_{\{X_\zeta - = b\}}\Big]. \tag{22}$$

The continuous time walk X defines the sequence $Y_0, \ldots, Y_\eta$ of lattice sites visited by X before it arrives at $*$. Let $\mathscr{W}_{n;ab}$ be the set of all nearest neighbour walks $y = (y_0, y_1, \ldots, y_n)$ such that $y_0 = a$ and $y_n = b$. For $y \in \mathscr{W}_{n;ab}$ the probability of the event $Y = y$ is the probability that the continuous time walk X will make the independent transitions $y_0 \to y_1$, $y_1 \to y_2$ and so on ending with y_n to $*$. Therefore

$$\mathrm{E}_a\Big[\mathbb{1}_{Y=y}\Big] = \left(\frac{1}{m^2+2d}\right)^n \frac{m^2}{m^2+2d}. \tag{23}$$

On the event $Y = y$, for $i = 0, \ldots, n$ the time that X waits at y_i is exponential with parameter $m^2 + 2d$ and all these exponential times are independent. Therefore, recalling that $F(L) = e^{-\sum w_x L_x}$,

$$\begin{aligned}\mathrm{E}_a\Big[F(L)\mathbb{1}_{Y=y}\Big] &= \frac{m^2+2d}{m^2+2d+w_{y_0}} \cdots \frac{m^2+2d}{m^2+2d+w_{y_n}}\ \mathrm{E}_a\Big[\mathbb{1}_{Y=y}\Big] \\ &= \frac{1}{m^2+2d+w_{y_0}} \cdots \frac{1}{m^2+2d+w_{y_n}} m^2. \end{aligned} \tag{24}$$

By summing this last formula over $y \in \mathscr{W}_{n;ab}$ and over n,

$$m^{-2}\ \mathrm{E}_a\Big[F(L)\mathbb{1}_{\{X_\zeta - = b\}}\Big] = \sum_{n\geq 0} \sum_{y\in\mathscr{W}_{n;ab}} \frac{1}{m^2+2d+w_{y_0}} \cdots \frac{1}{m^2+2d+w_{y_n}}. \tag{25}$$

To complete the proof we need the right hand side of this equation to be equal to $(A+W)^{-1}_{ab}$. To check this write $A+W$ as a diagonal matrix D minus an off-diagonal matrix J. By the expansion $(D-J)^{-1} = D^{-1} + D^{-1}JD^{-1} + D^{-1}JD^{-1}JD^{-1} + \ldots$ with matrix products written out in terms of sums over indices,

$$(A+W)^{-1}_{ab} = \sum_{n\geq 0} \sum_{y\in\mathscr{W}_{n;ab}} \frac{1}{m^2+2d+w_{y_0}} \cdots \frac{1}{m^2+2d+w_{y_n}}, \tag{26}$$

as desired. This expansion is a sum of positive terms because J has nonnegative entries and D has positive diagonal entries. The expansion is convergent because the number of terms in $\mathscr{W}_{n;ab}$ is $(2d)^n$ whereas the summand is smaller than $(2d + m^2)^{n+1}$.

Example 3.3 In Theorem 3.2 set

$$F(t) \ := \ \mathrm{e}^{-g\sum_x t_x^2}. \tag{27}$$

Then from Theorem 3.2,

$$\mathbb{E}\Big[\mathrm{e}^{-\sum_x g\,(\bar\phi_x\phi_x)^2}\,\bar\phi_a\,\phi_b\Big] \;=\; m^{-2}\int_0^\infty \mathbb{E}\otimes \mathrm{E}_a\Big[\mathrm{e}^{-\sum_x g\,(\bar\phi_x\phi_x+L_x)^2}\Big]\,\mathrm{e}^{-m^2T}\,\mathrm{d}T. \tag{28}$$

The left hand side is known as the *lattice* $|\phi|^4$ *quantum field theory*. According to, for example, Theorem 2 and the last item on page 22 of [35], in

$$\big(\bar\phi_x\phi_x + L_x\big)^2 \;=\; \big(\bar\phi_x\phi_x\big)^2 + L_x^2 + 2\,\bar\phi_x\phi_x L_x \tag{29}$$

the random variables $(\phi_x\bar\phi_x, x\in\Lambda)$ have the same distribution as the local times of loop soup plus a field of independent Γ variables that represent the local times of trivial loops that stay at one point. Thus the first term in (29) represents an interaction between the loops of loop soup and also a self-interaction for each loop. This interaction suppresses all mutual and self-intersections in the loop soup, much like the factor in the Edwards model suppresses self-intersections of a continuous time random walk. In fact, in the second term of (29) we see that as well as the loops there is a random walk with its own Edwards interaction. The third term suppresses all intersections between the loops and the walk. Thus the lattice $|\phi|^4$ quantum field theory is a model of many polymers which mutually repel. This is what Symanzik discovered in [47]. He was advocating this as a way to understand euclidean quantum field theory, but it might also be a good way to study polymers because the correlation inequalities of euclidean quantum field theory are interesting statements for the polymer model.

4 The Free Field, Local Time and Differential Forms

The main result in this section is Proposition 4.3 which is a variant of Theorem 3.2 that will be used to express the susceptibility of the Edwards model in terms of the massless free field. However it requires a conceptual extension of the massless free field and we first prepare the way with a review of differential forms.

4.1 Review of Differential Forms

A good reference for (differential) forms is [5]. As a motivational example let f be a smooth real function on $\mathbb{R}^2$. Given a point (u,v) in $\mathbb{R}^2$ let $(\dot u,\dot v)$ be also a point in $\mathbb{R}^2$, but think of it as a direction one can travel in, starting at (u,v), and call it a *tangent vector* at (u,v). Then define $df = df_{(u,v)}$, as a linear function on the vector

space of tangent vectors at (u, v), by

$$df : (\dot{u}, \dot{v}) \mapsto f_u \dot{u} + f_v \dot{v}, \tag{30}$$

where $f_u = f_u(u, v)$ and $f_v = f_v(u, v)$ are the partial derivatives of f at (u, v). With this definition with f replaced first by the function $\hat{u} : (u, v) \mapsto u$ and second by the function $\hat{v} : (u, v) \mapsto v$ the reader can verify that $df = f_u d\hat{u} + f_v d\hat{v}$. It is usual to leave off the hats as soon as we have understood that u and v are being used in two senses. Thus, by defining the symbols df, du, dv as linear functions on tangent spaces, we gain a precise meaning for

$$df = f_u du + f_v dv. \tag{31}$$

The space of forms at (u, v) *of degree one* is the vector space dual to the vector space of tangent vectors at (u, v). When we omit the phrase "at (u, v)" and say simply that "ω is a form of degree one" then we mean that for each point (u, v) in $\mathbb{R}^2$ $\omega_{(u,v)}$ is a form of degree one at (u, v) and that $(u, v) \mapsto \omega_{(u,v)}$ is smooth. This statement is clearly true for du and dv and it follows that it is true for df since $f_u(u, v)$ and $f_v(u, v)$ are smooth functions of (u, v). It is obvious how to generalise this discussion to $\mathbb{R}^n$.

Given two forms ω and ω' of degree one at a point in $\mathbb{R}^n$, we create a bilinear form $\omega \otimes \omega'$ by setting, for any two tangent vectors e and e',

$$(\omega \otimes \omega')(e, e') = \omega(e)\omega'(e') \tag{32}$$

but the bilinear form with geometrical significance is the antisymmetric tensor product defined by

$$(\omega \wedge \omega')(e, e') \ := \ \omega(e)\omega'(e') - \omega'(e)\omega(e). \tag{33}$$

For example

$$(\mathrm{d}u \wedge \mathrm{d}v)(e, e') \ := \ \mathrm{d}u(e)\mathrm{d}v(e') - \mathrm{d}u(e')\mathrm{d}v(e) \tag{34}$$

is the (signed) area of the parallelogram generated by e and e'. A form of degree two on $\mathbb{R}^n$ is, by definition, an antisymmetric bilinear function of directions assigned to points smoothly. All such objects can be written as $\sum a_{ij} \mathrm{d}u_i \wedge \mathrm{d}u_j$ where a_{ij} are smooth functions on $\mathbb{R}^n$. Similarly forms of degree p on $\mathbb{R}^n$ are antisymmetric p-linear functions of directions assigned smoothly to points. Notice that the degree is at most $p = n$ because there are no antisymmetric functions of higher degree. By definition forms of degree 0 are functions on $\mathbb{R}^n$.

Now the textbooks do something that looks strange but works out well: given two forms ω and ω' of degree $p \neq p'$ we define the direct sum $\omega \oplus \omega'$, which we can do because the set of forms of degree p is a vector space $\Omega^{(p)}$ and so $\omega \oplus \omega'$ is an element of the vector space $\Omega^{(p)} \oplus \Omega^{(p')}$. This amounts to saying that when $\omega \oplus \omega'$ is evaluated on p directions $\dot{u}_1, \ldots, \dot{u}_p$ it equals the evaluation of ω on these

p directions and likewise when $\omega \oplus \omega'$ is evaluated on p' directions it equals the evaluation of ω' on these p' directions, and evaluation on q directions is zero for q not p or p'. Let $\Omega^* = \bigoplus \Omega^{(p)}$. This is the vector space of all forms and we write the addition in this space using $+$ in place of $\oplus$. After defining $\mathrm{d}u_{i_1} \wedge \cdots \wedge \mathrm{d}u_{i_p}$ by antisymmetrising the tensor product the general form $\omega \in \Omega^*$ on $\mathbb{R}^n$ can be written as

$$\omega = \sum_p \sum_{i_1,\dots,i_p} a_{i_1,\dots,i_p} \mathrm{d}u_{i_1} \wedge \cdots \wedge \mathrm{d}u_{i_p} \tag{35}$$

where the coefficients are smooth functions on $\mathbb{R}^n$. The important fact is that the $\wedge$ product is *associative*, as well as distributive, over $+$. When a is a form of degree zero, that is, a smooth function on $\mathbb{R}^n$, and ω is a general form, then by definition

$$a\omega = a \sum_p \sum_{i_1,\dots,i_p} a_{i_1,\dots,i_p} \mathrm{d}u_{i_1} \wedge \cdots \wedge \mathrm{d}u_{i_p} = \sum_p \sum_{i_1,\dots,i_p} a a_{i_1,\dots,i_p} \mathrm{d}u_{i_1} \wedge \cdots \wedge \mathrm{d}u_{i_p} \tag{36}$$

which is the pointwise scalar product of the vector space Ω^*. In this case we omit the wedge. Note that

$$a(\omega \wedge \omega') = (a\omega) \wedge \omega' = \omega \wedge (a\omega'). \tag{37}$$

From now on we rarely need to know that a form is a linear combination of antisymmetric multilinear functions of tangent vectors. Most of the time we only use the fact that Ω^* is an algebra with n generators du_i, $i = 1, \dots, n$, that satisfy the relations $\mathrm{d}u_i \wedge \mathrm{d}u_i = 0$ for $i = 1, \dots, n$. An algebra whose generators satisfy such relations is called a *Grassmann algebra*. These relations imply that $\mathrm{d}u_i \wedge \mathrm{d}u_j = -\mathrm{d}u_j \wedge \mathrm{d}u_i$ for all $i, j = 1, \dots, n$. This means that multiplication is commutative for forms of even degree and anticommutative for forms of odd degree. In our case we have a Grassmann algebra *over* $\mathscr{C}^\infty(\mathbb{R}^n)$ which means that for f in $\mathscr{C}^\infty(\mathbb{R}^n)$ and forms ω and $\omega', f(\omega \wedge \omega') = (f\omega) \wedge \omega' = \omega \wedge (f\omega')$. The Grassmann algebra generated by linear functions on a vector space V is called the *exterior algebra* of V.

Given a form ω as in (35) with integrable coefficients we define

$$\int_{\mathbb{R}^n} \omega = \int_{\mathbb{R}^n} a_{1,2,\dots,n} \mathrm{d}u_1 \dots \mathrm{d}u_n. \tag{38}$$

At first it seems strange that this definition of the integral of a form ignores all coefficients except the one in the top degree form, but it is consistent with our earlier remarks about the meaning of addition in the exterior algebra and the idea that $\mathrm{d}u_1 \wedge \cdots \wedge \mathrm{d}u_n$ is a multilinear function that assigns volume to parallelopipeds of dimension n. The important point about integration of forms is that the value of the integral is independent of the choice of coordinates for $\mathbb{R}^n$. To understand this try as an exercise case $n = 2$. Let f be a smooth orientation preserving bijection from $\mathbb{R}^2$ to itself. Rewrite $\omega = \sum a_{ij} \mathrm{d}u_i \wedge \mathrm{d}u_j$ in terms of (u_1', u_2') where $u_i = f_i(u_1', u_2')$ by

substituting $du = \sum f_i du'_i$ into ω. Check that $\int \omega$ is the same regardless of whether it is expressed in terms of u or u'. This invariance property of forms is what motivated Cartan to introduce this formalism.

4.2 Gaussian Integrals in Terms of Forms

We make the natural extension of the previous constructions for $\mathbb{R}^n$ to the complex space $\mathbb{C}^\Lambda$. A point in this space is given by $\phi = u + \mathrm{i}v$ where $u = (u_x)_{x\in\Lambda}$ and $v = (v_x)_{x\in\Lambda}$ are in $\mathbb{R}^\Lambda$. For each $x \in \Lambda$ we have a degree one form $\mathrm{d}\phi_x := \mathrm{d}u_x + \mathrm{i}\mathrm{d}v_x$ and

$$\mathrm{d}\phi_x \wedge \mathrm{d}\bar{\phi}_x \;=\; (\mathrm{d}u_x + \mathrm{i}\mathrm{d}v_x) \wedge (\mathrm{d}u_x - \mathrm{i}\mathrm{d}v_x) \;=\; -2\,\mathrm{i}\,(\mathrm{d}u_x \wedge \mathrm{d}v_x). \tag{39}$$

Let $\mathscr{N}(\Lambda)$ denote the exterior algebra over $C^\infty\big(\mathbb{R}^{2\Lambda}\big)$ generated by $\{\mathrm{d}u_x, \mathrm{d}v_x : x \in \Lambda\}$ or, alternatively, $\{\mathrm{d}\phi_x, \mathrm{d}\bar{\phi}_x : x \in \Lambda\}$.

Now let's go back to the definition of the gaussian (complex) measure (13). For F a random variable (form of degree zero)

$$\mathbb{E}\big[F\big] \;=\; \int_{R^{2\Lambda}} \mathrm{e}^{-(\phi, A\bar{\phi})} F \det(A) \prod_{x\in\Lambda} \frac{\mathrm{d}u_x \mathrm{d}v_x}{\pi}. \tag{40}$$

We claim that for any dissipative real matrix A,

$$\mathbb{E}\big[F\big] \;=\; \int_{R^{2\Lambda}} \mathrm{e}^{-(\phi, A\bar{\phi}) - \frac{1}{2\pi\mathrm{i}}(\mathrm{d}\phi\wedge, A\mathrm{d}\bar{\phi})}\; F \tag{41}$$

where

$$\big(\mathrm{d}\phi\wedge, A\,\mathrm{d}\bar{\phi}\big) \;:=\; \sum_{x,y\in\Lambda} \mathrm{d}\phi_x \wedge A_{xy} \mathrm{d}\bar{\phi}_y \tag{42}$$

and we really are claiming that there is no constant of normalisation in (41). For α a function and β a form the exponential $e^{\alpha+\beta}$ is the element of the algebra $\mathscr{N}(\Lambda)$ defined by

$$\mathrm{e}^{\alpha+\beta} \;:=\; \mathrm{e}^{\alpha} \sum_{p\geq 0} \frac{1}{p!}\, \beta^{\wedge p}. \tag{43}$$

The sum is finite because all terms of degree more than $2|\Lambda|$ are zero. For our case

$$\beta = -\frac{1}{2\pi\mathrm{i}}(\mathrm{d}\phi\wedge, A\mathrm{d}\bar{\phi}). \tag{44}$$

So an expansion of the second term in the exponent results in a linear combination of forms of various degrees. Recall that the rule of integration (38) of forms is that the integral of all but the highest-degree form vanishes. The highest degree form is the term $p = |\Lambda|$ in the sum in (43). As an exercise show that this term equals

$$\beta^{2p} = (2\pi i)^{-p} \det A \prod_{x\in\Lambda} (-\mathrm{d}\phi_x \wedge \mathrm{d}\bar\phi_x) \tag{45}$$

and complete the proof of (41) using (39).

Let $c = \frac{1}{2\pi i}$ and define

$$\begin{aligned}(\tau_\Delta)_x := \frac{1}{2}\Big(&\phi_x\,(-\Delta\bar\phi)_x + c\,\mathrm{d}\phi_x \wedge (-\Delta \mathrm{d}\bar\phi)_x \\ &+(-\Delta\phi)_x\,\bar\phi_x + c\,(-\Delta \mathrm{d}\phi)_x \wedge \mathrm{d}\bar\phi_x\Big)\end{aligned} \tag{46}$$

and

$$\tau_x \;:=\; \phi_x\bar\phi_x + c\,\mathrm{d}\phi_x \wedge \mathrm{d}\bar\phi_x. \tag{47}$$

When $(\tau_\Delta)_x$ appears under a sum over x in Λ the second two terms in $(\tau_\Delta)_x$ make the same contribution as the first two terms, which cancels the $1/2$. Let

$$\tau_\Delta(\Lambda) = \sum_{x\in\Lambda} (\tau_\Delta)_x, \qquad \tau(\Lambda) = \sum_{x\in\Lambda} \tau_x. \tag{48}$$

Then, for $A = m^2 - \Delta$, the exponent in (41) is the same as $m^2\tau(\Lambda) + \tau_\Delta(\Lambda)$, as claimed in the following definition.

Definition 4.1 (Super-Expectation) The super-expectation for massive free field is defined for $m^2 > 0$ and for bounded forms F in $\mathscr{N}(\Lambda)$ by

$$\mathbb{E}^{(m)}[F] \;:=\; \int_{\mathbb{R}^{2\Lambda}} \mathrm{e}^{-(\phi, A\bar\phi) - c(\mathrm{d}\phi, A\mathrm{d}\bar\phi)} F \;=\; \int_{\mathbb{R}^{2\Lambda}} \mathrm{e}^{-(m^2\tau + \tau_\Delta)(\Lambda)}\, F \tag{49}$$

where

$$A := m^2 - \Delta, \qquad c = \frac{1}{2\pi \mathrm{i}}. \tag{50}$$

A bounded form is a form whose coefficients are bounded. Here and from now on we are omitting the $\wedge$ between the exponential and F.

The point of this definition is that the right hand side makes sense if F is a form in $\mathscr{N}(\Lambda)$ but when we evaluate the super-expectation of a form of degree 0, in other words a random variable, by (41) the super-expectation is the same as the expectation, so we are defining an extension of the standard expectation to the

algebra of integrable forms. Of course not all properties we are used to remain valid when we are taking the super-expectation of a form. For example we do not have a Jensen inequality unless the form under $\mathbb{E}^{(m)}$ is of degree zero.

4.3 The Local Time Isomorphism and Forms

Recall the definition of τ_x from (47) and let τ denote the sequence of forms (τ_x), where x ranges over Λ. Likewise let $\phi\bar{\phi}$ denote the sequence $(\phi_x\bar{\phi}_x)$ and $c\,\mathrm{d}\phi \wedge \mathrm{d}\bar{\phi}$ denote the sequence $(c\,\mathrm{d}\phi_x \wedge \mathrm{d}\bar{\phi}_x)$. For a smooth function F defined on $\mathbb{R}^\Lambda$, there is a multivariable Taylor expansion

$$F(t+\dot{t}) \sim \sum_p \frac{1}{\alpha!} F^{(\alpha)}(t)\dot{t}^\alpha \tag{51}$$

about the point t in $\mathbb{R}^\Lambda$ in powers $\dot{t}^\alpha = \prod_{x\in\Lambda} \dot{t}_x^{\alpha_x}$ of the components (t_x) of t. Let $t = \phi\bar{\phi}$ and $\dot{t} = c\,\mathrm{d}\phi \wedge \mathrm{d}\bar{\phi}$ in this Taylor series, replacing $\prod_{x\in\Lambda} \dot{t}_x^{\alpha_x}$ by the wedge product $\wedge_{x\in\Lambda}\dot{t}_x^{\alpha_x}$. This product is well defined regardless of the order with which x ranges over Λ because the forms $c\,\mathrm{d}\phi_x \wedge \mathrm{d}\bar{\phi}_x$ are even. Also, the series terminates after finitely many terms because each term is a form of degree $2|\alpha|_1$ and forms of degree larger than $2|\Lambda|$ vanish. Therefore the Taylor expansion with these substitutions defines a form. We denote this form by $F(\tau)$. It is a good notation because $F \mapsto F(\tau)$ is an algebra homomorphism from the algebra of smooth functions into $\mathscr{N}(\Lambda)$. Also the map respects composition $f(F(\tau)) = (f \circ F)(\tau)$. We will not prove these claims, but they are consequences of the uniqueness of the Taylor expansion.

Lemma 4.2 *For any bounded smooth function $F\colon \mathbb{R}^\Lambda \to \mathbb{R}$ with bounded derivatives and for $m^2 > 0$,*

$$\mathbb{E}^{(m)}\big[F(\tau)\big] \;=\; F(0). \tag{52}$$

This also holds for any dissipative matrix A in place of A as defined in (50)*.*

Proof A complete proof is given in [21] but it is instructive to check the claim for the special case $F(\tau) := \exp\{-\sum_x w_x\tau_x\}$ with $w_x \geq 0$. We exhibit dependence on the matrix A by writing $\mathbb{E}_A$. Let W be the diagonal matrix with w_x on the diagonal. Then with this special F we have from Definition 4.1 that $\mathbb{E}_A[F] = \mathbb{E}_{A+W}[1]$ and $\mathbb{E}_{A+W}[1] = 1 = F(0)$ as desired because the super-expectation equals the expectation on forms of degree 0.

Recall that when we proved Theorem 3.2 we also checked the special case of an exponential F and this sufficed to prove the general case because both sides are linear in F. In this algebra of forms context we no longer have monotone class theorems to extend from linear combinations of exponentials to the general case. The idea in [21] is instead to write a general F as the fourier transform of its Fourier

transform to see that it is a limit of linear combinations of exponentials. The same idea is at work in the next isomorphism theorem, which I will call the "$\tau = L$" theorem.

Proposition 4.3 *For a bounded smooth function $F:\mathbb{R}^{\Lambda} \to \mathbb{R}$ with bounded derivatives,*

$$\mathbb{E}^{(m)}\big[F(\tau)\,\bar\phi_a\phi_b\big] \;=\; m^{-2}\,\mathrm{E}_a\big[F(L)\,\mathbb{1}_{\{X_\zeta - = b\}}\big]. \tag{53}$$

Proof For a complete proof see [21, Proposition 2.4]. It is again sufficient to prove it is true for the special case $F(\tau) := \exp\{-\sum_x w_x\tau_x\}$ with $w_x \ge 0$. As in the last proof,

$$\mathbb{E}^{(m)}\big[F(\phi\bar\phi)\,\bar\phi_a\phi_b\big] = \mathbb{E}_A\big[F(\phi\bar\phi)\,\bar\phi_a\phi_b\big] = \mathbb{E}_{A+W}\big[\bar\phi_a\phi_b\big]. \tag{54}$$

The super-expectation on the right coincides with the expectation so the right hand side equals $(A+W)^{-1}_{ab}$ and in the proof of Theorem 3.2 we proved that this equals the right hand side of Proposition 4.3, as desired.

5 Susceptibility as a Gaussian Integral

For the parameters g, ν that appeared in the Edwards model and a new one called z and $X \subset \Lambda$ define

$$\begin{aligned} V_{g,\nu,z;x} &:= g\,\tau_x^2 + \nu\,\tau_x + z(\tau_\Delta)_x, \\ V_{g,\nu,z}(X) &= \sum_{x\in X} V_{g,\nu,z;x}, \\ \phi(X) &= \sum_{x\in X}\phi_x. \end{aligned} \tag{55}$$

Notice that we are starting to omit $\wedge$, for example τ_x^2 is really $\tau_x \wedge \tau_x$. By choosing $F(\tau) = \exp\big[-g\,\tau_x^2(\Lambda) - (\nu - m^2)\,\tau(\Lambda)\big]$ in Proposition 4.3 and using the definition of the super-expectation we find that

$$\chi_\Lambda(g,\nu) \;=\; \int_{\mathbb{R}^{2\Lambda}} \mathrm{e}^{-V_{g,\nu,1}(\Lambda)}\;\bar\phi_a\,\phi(\Lambda). \tag{56}$$

Thus the susceptibility is represented as a $2|\Lambda|$ dimensional integral of forms. Perhaps this does not seem like a very pleasant reward for so much work, but let us see.

These differential form representations came from [39, 44] and particularly [37] in the physics literature, where they are instances of supersymmetry. The results in these papers are expressed in terms of *anticommuting numbers* which

are also known as *ghosts*. Anticommuting numbers are another name for elements of a Grassmann algebra as defined above. The definition of a *Grassmann* integral as the coefficient of the highest degree monomial in a Grassmann algebra is called the Berezin integral after the standard reference [12]. I first encountered the identification of anticommuting numbers with differential forms in [36, 55]. I discussed the isomomorphism between local time and the gaussian field for the complex case with and without Berezin integration in [14].

Lemma 4.2 for the exponential is a special case of the remarkable Duistermaat-Heckman theorem [25, Theorem 4.1 on p. 267]. We shall not need this theorem but to see why it is a more general statement note that it applies to even dimensional spheres. Since we are considering integrals over $\mathbb{R}^{2|\Lambda|}$ of functions that decay at infinity we can add a point at infinity and replace $\mathbb{R}^{2|\Lambda|}$ by an even dimensional sphere. The mathematical literature on the Duistermaat-Heckman theorem makes unfounded assumptions about my education in topology and I found the more informal Sect. 2.2.2 and the standard example in Appendix A of [56] helpful.

5.1 The Most General Split into Gaussian Plus Perturbation

At the end of the last section we found that the susceptibility has the representation

$$\chi_\Lambda(g,\nu) \;=\; \int_{\mathbb{R}^{2\Lambda}} \mathrm{e}^{-V_{g,\nu,1}(\Lambda)}\, \bar{\phi}_a\, \phi(\Lambda). \tag{57}$$

where

$$\begin{aligned} V_{g,\nu,1}(\Lambda) &= \sum_{x\in\Lambda} V_{g,\nu,1;x}, \\ V_{g,\nu,1;x} &= \left(g\tau^2 + \nu\tau + \tau_\Delta\right)_x. \end{aligned} \tag{58}$$

We are now going to try to regard this as an almost gaussian integral because in Theorem 2.2 the hypothesis was that g is positive, but small. Since $V_{g,\nu,1}$ has two quadratic terms a naive attempt is to use them to define the gaussian measure. However, recall that Theorem 2.2 concerns the case where ν is just a little larger than the critical value ν^{c} given by Definition 2.1. Furthermore ν^{c} will turn out to be negative. Therefore we cannot make this naive choice of gaussian measure because $A = \nu^{\mathrm{c}}\mathrm{id} - \Delta$ is not dissipative, which means that the gaussian is not integrable.

Actually we want to choose the gaussian part to be whatever best approximates the long distance behaviour of the model. As an analogy recall that when we want to approximate a sum of n identically distributed centred independent random variables by a gaussian as in the central limit theorem, we have to know (1) to normalise the sum by $\sqrt{n}$, (2) what the variance of the gaussian will be. (1) and (2) are both determined by computing the variance of the sum and this can be done

because the variables are independent. In the present case we cannot guess how to scale ϕ or what the best gaussian will be because there is no obvious independence. Instead, we consider a general split that is parameterised by two parameters called m^2 and z_0 and wait patiently for the renormalisation group to tell us what the values of the parameters should be.

The most general way to split

$$V_{g,\nu,1}(\Lambda) = \big(g\tau^2 + \nu\tau + \tau_\Delta\big)(\Lambda) \tag{59}$$

into a quadratic (gaussian) part and a perturbation can be parameterised by two parameters as follows. First, we introduce a parameter $z_0 > -1$ and split the coefficient 1 implicit in front of τ_Δ as

$$1 = \frac{1}{1+z_0} + \frac{z_0}{1+z_0}. \tag{60}$$

Then we introduce another parameter $m^2 > 0$ to split

$$\nu = \frac{m^2}{1+z_0} + \left(\nu - \frac{m^2}{1+z_0}\right). \tag{61}$$

The reason that $1 + z_0$ is written in the denominators is so that we can get rid of it by rescaling. Introducing the rescaled field $\hat\phi = (1+z_0)^{-1/2}\phi$, we have

$$V_{g,\nu,1}(\phi) = V_{0,m^2,1}(\hat\phi) + V_{g_0,\nu_0,z_0}(\hat\phi), \tag{62}$$

where

$$g_0 = g(1+z_0)^2, \quad \nu_0 = \nu(1+z_0) - m^2. \tag{63}$$

I will call the parameters z_0, m^2 *splitting parameters*.

From the representation (57) of χ_Λ and noticing that the exponent in the super-expectation in Definition 4.1 is $V_{0,m^2,1}(\Lambda)$, we have

$$\chi_\Lambda(g,\nu) \;=\; (1+z_0)\,\mathbb{E}^{(m)}\Big[\mathrm{e}^{-V_{g_0,\nu_0,z_0}(\Lambda)}\,\bar\phi_a\,\phi(\Lambda)\Big]. \tag{64}$$

We obtained this by rewriting the integral in terms of the scaled variable $\hat\phi$ and then renaming $\hat\phi$ back to ϕ. The change of variable does not give a Jacobian factor because it is a form integral. We define

$$\hat\chi_\Lambda(m^2, g_0, \nu_0, z_0) = \mathbb{E}^{(m)}\Big[\mathrm{e}^{-V_{g_0,\nu_0,z_0}(\Lambda)}\,\bar\phi_a\,\phi(\Lambda)\Big], \tag{65}$$

$$\hat\chi(m^2, g_0, \nu_0, z_0) = \lim_{\Lambda\to\mathbb{Z}^d} \hat\chi_\Lambda(m^2, g_0, \nu_0, z_0) \tag{66}$$

As explained in Sect. 2 the limit as Λ increases to $\mathbb{Z}^d$ exists for χ_Λ and partial derivatives with respect to coupling constants can be taken under this limit. Therefore we have proved the following splitting lemma

Lemma 5.1 *Given $\nu \in \mathbb{R}$ and $g > 0$, and given splitting parameters $z_0 > -1$ and $m^2 > 0$, let*

$$g_0 \;=\; g\,(1+z_0)^2 \quad \textit{and} \quad \nu_0 \;=\; \nu\,(1+z_0) - m^2 \tag{67}$$

then

$$\chi_\Lambda(g,\nu) \;=\; (1+z_0)\;\mathbb{E}^{(m)}\Big[\mathrm{e}^{-V_{g_0,\nu_0,z_0}(\Lambda)}\;\bar{\phi}_a\,\phi(\Lambda)\Big]. \tag{68}$$

The infinite volume limit exists and is given by

$$\chi(g,\nu) \;=\; (1+z_0)\,\hat{\chi}(m^2,g_0,\nu_0,z_0). \tag{69}$$

Moreover, for $\nu_0 > \nu^{\mathrm{c}}$,

$$\frac{\partial}{\partial\nu}\chi(g,\nu) \;=\; (1+z_0)^2\,\frac{\partial}{\partial\nu_0}\hat{\chi}(m^2,g_0,\nu_0,z_0). \tag{70}$$

5.2 The Proof of Theorem 2.2

The Edwards model that we are studying contains two parameters ν and g. Our general problem is: given ν, g calculate the susceptibility for ν slightly larger than the critical value $\nu_{\mathrm{c}}(g)$. In the last section we introduced a strategy: show that this model has a gaussian approximation and calculate the susceptibility of this approximation. This strategy was started by splitting the model into (a scaling of) a free field with mass m^2 and a perturbation described by parameters g_0, ν_0, z_0. The four new parameters m^2, ν_0, g_0, z_0, are linked by two relations (67) so we expect to need two more relations to completely specify all of them in terms of the given ν, g. Theorem 5.2 in this section provides these two relations in the form $\nu_0 = \nu_0^{\mathrm{c}}(m^2, g_0)$ and $z_0 = z_0^{\mathrm{c}}(m^2, g_0)$ and it gives enough information to prove our main Theorem 2.2.

In this Theorem 5.2 appears the expected time

$$\mathsf{B}_{m^2} = 8\iint P\big(X(t)=Y(s)\big)e^{-m^2t}e^{-m^2s}\,dt\,ds \tag{71}$$

that two independent simple random walks with killing spend intersecting each other. In $d = 4$ dimensions it can be shown that as m^2 tends to zero,

$$\mathsf{B}_{m^2} \sim 8\,\log m^{-2} \tag{72}$$

with $\mathsf{b} = 1/(2\pi^2)$. In more than four dimensions there is no divergence as $m^2 \to 0$.

Theorem 5.2 (Theorem 4.1 in [10]) *Let $d = 4$, and let $\delta > 0$ be sufficiently small. There are continuous real-valued functions $\nu_0^{\mathrm{c}}, z_0^{\mathrm{c}}$, defined for $(m^2, g_0) \in [0, \delta)^2$ and continuously differentiable in g_0, and there is a continuous function $c(g_0) = 1 + O(g_0)$, such that for all $m^2, g_0, \hat{g}_0 \in (0, \delta)$,*

$$\hat{\chi}\left(m^2, g_0, \nu_0^{\mathrm{c}}(m^2, g_0), z_0^{\mathrm{c}}(m^2, g_0)\right) = \frac{1}{m^2}, \tag{73}$$

$$\frac{\partial \hat{\chi}}{\partial \nu_0}\left(m^2, g_0, \nu_0^{\mathrm{c}}(m^2, g_0), z_0^{\mathrm{c}}(m^2, g_0)\right) \sim -\frac{1}{m^4}\frac{c(\hat{g}_0)}{(\hat{g}_0 B_{m^2})^{\gamma}} \text{ as } (m^2, g_0) \to (0, \hat{g}_0). \tag{74}$$

The functions $\nu_0^{\mathrm{c}}, z_0^{\mathrm{c}}$ obey

$$\begin{aligned} \nu_0^{\mathrm{c}}(m^2, 0) =& z_0^{\mathrm{c}}(m^2, 0) = 0, \\ \frac{\partial \nu_0^{\mathrm{c}}}{\partial g_0}(m^2, g_0) =& O(1), \\ \frac{\partial z_0^{\mathrm{c}}}{\partial g_0}(m^2, g_0) =& O(1), \end{aligned} \tag{75}$$

where $O(1)$ means that these derivatives are bounded on their whole domain by constants uniform in (m^2, g_0).

Remark 5.3 In the standard theory of renormalisation (73) is not a theorem, but merely the definition of m^2. In our work m^2 has been defined as the mass in a free field and we will instead use the renormalisation group to prove that $\nu_0^{\mathrm{c}}, z_0^{\mathrm{c}}$ exist such that at large scales this system becomes this free field.

Proof (of Theorem 2.2) Define the map

$$A : (m^2, g_0) \mapsto \left(m^2, g_0, \nu_0^{\mathrm{c}}(m^2, g_0), z_0^{\mathrm{c}}(m^2, g_0)\right) \tag{76}$$

with the domain $(m^2, g_0) \in [0, \delta)^2$ specified in the theorem for ν_0^{c} and z_0^{c}. We eliminate m^{-2} in (74) using (73). The elimination includes the m^2 in B_{m^2} using (72). We obtain

$$\left[\frac{\partial \hat{\chi}}{\partial \nu_0}\right] \circ A \sim -(\hat{\chi} \circ A)^2 \frac{c(\hat{g}_0)}{\left(\hat{g}_0 8\,\mathsf{b} \log(\hat{\chi} \circ A)\right)^{\gamma}}, \tag{77}$$

We define another map $B : (m^2, g_0, \nu_0, z_0) \mapsto (g, \nu)$ by solving (67) explicitly. Let $C = B \circ A$. By Lemma 5.1, we obtain the following equation for $\chi = \chi(g, \nu)$.

$$\begin{aligned} \left[\frac{\partial \chi}{\partial \nu}\right] \circ C &\sim -(\chi \circ C)^2 \frac{c(\hat{g}_0)}{\left(8\mathsf{b} g_0 \log(\chi \circ C) - 8\mathsf{b} g_0 \log(1 + z_0^{\mathrm{c}})\right)^{\gamma}} \\ &\sim -(\chi \circ C)^2 \frac{c(\hat{g}_0)}{\left(8\mathsf{b} g_0 \log(\chi \circ C)\right)^{\gamma}}. \end{aligned} \tag{78}$$

The $\sim$ allows us to omit the term involving $\log(1 + z_0^{\mathrm{c}})$ term because (75) implies it is bounded as $(m^2, g_0) \to (0, \hat{g}_0)$, whereas $\chi \circ C$ diverges. The divergence of $\chi \circ C$ follows from Lemma 5.1 and (73), which together assert that

$$\chi \circ C = \frac{1 + z_0^{\mathrm{c}}}{m^2}, \tag{79}$$

noting that (75), and decreasing δ if necessary, implies that $1 + z_0^{\mathrm{c}}$ does not vanish as $m^2 \downarrow 0$.

Proposition 4.2 (ii) in [10] states that the inverse $C^{-1} : (g, \nu) \mapsto (m^2, g_0)$ exists, is right-continuous in ν for g fixed and is defined on a domain $0 < g < \delta_1$, $\nu_{\mathrm{c}}(g) \leq \nu < \nu_{\mathrm{c}}(g) + \delta_1$. In the notation of Bauerschmidt et al. [10], the two components of C^{-1} are called $\tilde{m}^2, \tilde{g}$ and they are written as functions of (g, ϵ) where $\epsilon = \nu - \nu_{\mathrm{c}}(g)$. By the definition of ν_{c} in Definition 2.1 and the comment below this definition, $\chi(g, \nu)$ is finite for $\nu \in (\nu_{\mathrm{c}}(g), \infty)$ and diverges as $\nu \downarrow \nu_{\mathrm{c}}(g)$. Therefore, by (79), $m^2 = 0$ when $\nu = \nu_{\mathrm{c}}(g)$. By the right continuity of C^{-1}, $\nu \downarrow \nu_0^{\mathrm{c}}$ with g fixed implies $(m^2, g_0) \to (0, \hat{g}_0)$. Therefore (78) simplifies to

$$\frac{\partial \chi}{\partial \nu} \sim -\chi^2 \frac{c(\hat{g}_0)}{\big(8\mathsf{b} g_0 \log \chi\big)^{\gamma}}, \quad \nu \downarrow \nu_0^{\mathrm{c}} \text{ with } g \text{ fixed}, \tag{80}$$

where $\hat{g}_0$ is the g component of $C^{-1}(g, 0)$.

We fix $g < \delta_1$ and define $F(\epsilon) = \frac{1}{\chi(g,\nu)}$ for $\nu = \nu_{\mathrm{c}}(g) + \epsilon$ with $\epsilon \in (0, \delta_1)$ and we set $F(0) = 0$. By dividing (80) by χ^2 (80) becomes, for $\epsilon \downarrow 0$,

$$\frac{dF}{d\epsilon} \sim \frac{1}{A_g \big(\log F^{-1}\big)^{\gamma}}, \quad A_g = \frac{(\tilde{g}_0(g, 0)\mathsf{b})^{\gamma}}{c_0(g)} \tag{81}$$

and this differential relation can be easily integrated [10, Lemma 4.3] to show that

$$F(\nu_c + \varepsilon) \sim A_g^{-1} \varepsilon (-\log \varepsilon)^{-\gamma}. \tag{82}$$

Recalling that $F = \chi^{-1}$ this is the claim in Theorem 2.2.

5.3 *The Susceptibility in Terms of Super-Convolution*

Our main result Theorem 2.2 has been reduced to Theorem 5.2. To prepare for the proof of Theorem 5.2, we put the susceptibility into a form that suits the renormalisation group which will be introduced in the next sections. The conclusion of this section is Proposition 5.5. The proof of this Proposition contains the important idea of using a translation to approximately evaluate a generating function. This is part of the "evaluation as if gaussian" strategy that started in

Sect. 5.1. The use of "tilting" in the theory of large deviations is also an instance of this strategy.

This section and the renormalisation group use super-convolution. In order to motivate the definition of super-convolution, recall that the convolution of a function f with a probability measure P is the function $x \mapsto \int f(x+z)\,P(dz)$. In order to think of this as an operation that transforms f into a new function g we define a homomorphism θ from the algebra of functions of one variable to the algebra of functions of two variables by: for $x \mapsto f(x)$ let $\theta f : (x,z) \mapsto f(x+z)$. Then the convolution can be written as $f \mapsto \mathbb{E}^{(z)}\theta f$.

Definition 5.4 Define the algebra homomorphism $\theta\colon \mathscr{N}(\Lambda) \to \mathscr{N}(\Lambda \sqcup \Lambda)$ to be the map that replaces ϕ by $\phi + \zeta$ and $\mathrm{d}\phi$ by $\mathrm{d}\phi + \mathrm{d}\zeta$. Then the *super-convolution* of a form $F \in \mathscr{N}(\Lambda)$ by the super-expectation $\mathbb{E}^{(m)}$ is given by

$$F \mapsto \mathbb{E}^{(m,\zeta)}\big[\theta F\big], \tag{83}$$

where $\mathbb{E}^{(m,\zeta)}$ acts only on ζ; (Therefore, in Definition 4.1 rename ϕ to ζ and then replace F by θF.)

Let F be an element of $\mathscr{N}(\Lambda)$. Recall that this means that F is a form whose coefficients are functions of ϕ and $\phi \in \mathbb{C}^\Lambda$. The directional derivative of F with respect to ϕ in the direction $f \in \mathbb{C}^\Lambda$ is defined by replacing ϕ by $\phi + zf$, $\bar\phi$ by $\bar\phi + \bar z\bar f$ and evaluating $(\partial/\partial z)F$ at $z = 0$ by the rules $\partial\bar z/\partial z = 0$ and $\partial z/\partial z = 1$. These rules follow from the definition $\partial/\partial z = \frac{1}{2}(\partial/\partial x - \mathrm{i}\partial/\partial y)$ when $z = x + \mathrm{i}y$. The directional derivative of a differential is zero because the replacement of ϕ by $\phi + zf$ is a change of variable and f does not depend on ϕ so $d(\phi + zf) = d\phi$. Similarly there is also the directional derivative of F with respect to $\bar\phi$ in the direction $\bar f$.

Let 1 denote the function in $\mathbb{C}^\Lambda$ which is the constant function $1_x = 1$ for all $x \in \Lambda$. In the next Proposition $D^2F(0;1,1)$ denotes the result of taking two directional derivative of F with respect to ϕ in the direction 1 and then setting $\phi = 0$ including $d\phi = 0$. (In [10] we used the notation $D^2F(0,0;1,1)$ since ϕ and $d\phi$ are both set to zero.)

Recall that the finite volume susceptibility $\hat\chi_\Lambda = \hat\chi_\Lambda(m^2, g_0, \nu_0, z_0)$ was defined in (65). We repeat the definition here in order to introduce some new notation,

$$\hat\chi_\Lambda \;=\; \mathbb{E}^{(m)}\Big[Z_0\,\bar\phi_a\,\phi(\Lambda)\Big], \quad Z_0 = \mathrm{e}^{-V_0(\Lambda)}, \quad V_0(\Lambda) = V_{g_0,\nu_0,z_0}(\Lambda). \tag{84}$$

Proposition 5.5 *Let m^2, g_0, ν_0, z_0 be real numbers with g_0, m^2 positive and $z_0 > -1$. Then*

$$\hat\chi_\Lambda(m^2, g_0, \nu_0, z_0) \;=\; \frac{1}{m^2} + \frac{1}{m^4\,|\Lambda|}\,D^2F(0;1,1), \tag{85}$$

where $F = \mathbb{E}^{(m,\zeta)}\big[\theta Z_0\big]$.

To prepare for the proof of this result we first discuss generating functions in this context. Given an *external field* $J : \Lambda \to \mathbb{C}$, we write

$$(J, \bar{\phi}) = \sum_{x \in \Lambda} J_x \bar{\phi}_x, \quad (\bar{J}, \phi) = \sum_{x \in \Lambda} \bar{J}_x \phi_x. \tag{86}$$

Recall that 1 denotes the function in $\mathbb{C}^\Lambda$ that is identically one. By translation invariance,

$$\hat{\chi}_\Lambda = |\Lambda|^{-1} \mathbb{E}^{(m)} \Big[(1, \bar{\phi})(1, \phi) Z_0 \Big]. \tag{87}$$

We define the *generating function* $\Sigma : \mathbb{C}^\Lambda \to \mathbb{C}$ by

$$\Sigma(J, \bar{J}) = \mathbb{E}^{(m)} \Big[e^{(J,\bar{\phi}) + (\phi, \bar{J})} Z_0 \Big]. \tag{88}$$

By taking two directional derivatives, one with respect J in the direction 1, the second with respect to $\bar{J}$ in the direction 1 and setting $J = 0$ we generate a factor $(1, \bar{\phi})(1, \phi)$ and so we have

$$\hat{\chi}_\Lambda = |\Lambda|^{-1} D^2_{\bar{J},J} \Sigma(0; 1, 1), \tag{89}$$

where $D^2_{\bar{J},J}$ indicates two directional derivatives with respect to J and $\bar{J}$, the argument 0 means the derivative is at $J = \bar{J} = 0$ and the two arguments 1 indicate the directions. The evaluation of $\hat{\chi}_\Lambda$ now becomes reduced to the evaluation of $D^2_{\bar{J},J} \Sigma$ on the right-hand side of (89). Here is where the above mentioned strategy of evaluation as if gaussian ($V_0 = 0$) comes into play. Recall that gaussian integrals are evaluated by change of variables by an optimal translation that centres the gaussian. By using such a translation $\phi = \zeta + H$, where ζ is the new integration variable, we obtain

$$\Sigma(J, \bar{J}) = e^{(J, C\bar{J})} \Big[\mathbb{E}^{(m,\zeta)} \big[\theta Z_0 \big] \Big]_{|_{\phi = CJ, d\phi = 0}}, \tag{90}$$

where $C = (-\Delta + m^2)^{-1}$. In more detail, with $A = -\Delta + m^2 = C^{-1}$ and referring to Definition 4.1,

$$\begin{aligned} &\sum_{x \in \Lambda} \big(\tau_{\Delta,x} + m^2 \tau_x \big) - (J, \bar{\phi}) - (\phi, \bar{J}) \\ &\quad = (\zeta, A\bar{\zeta}) + c(d\zeta, \wedge A d\bar{\zeta}) \\ &\qquad - (J, \bar{\zeta}) - (\bar{J}, \zeta) + (H, A\bar{\zeta}) + (\zeta, A\bar{H}) \\ &\qquad\quad + (H, A\bar{H}) - (J, \bar{H}) - (\bar{J}, H) \\ &\quad = (\zeta, A\bar{\zeta}) + c(d\zeta, \wedge A d\bar{\zeta}) - (J, C\bar{J}), \end{aligned} \tag{91}$$

where the last line is obtained by choosing H to make the terms in the second line sum to zero. This happens when $H = CJ$, $\bar{H} = C\bar{J}$. Since H does not depend on ϕ, $d\phi = d(\zeta + H) = d\zeta$ and so the form part is as written. The formula (90) follows immediately. Notice that the translation also changes ϕ in Z_0 to $\zeta + CJ$ and $d\phi$ to $d\zeta$. This is implemented in (90) by θ which changes ϕ to $\phi + \zeta$ followed by evaluation at $\phi = CJ$ and $d\phi = 0$ after taking the super-expectation over ζ. Since our strategy was based on the hope that our splitting into gaussian and interaction is such that $V_0(\Lambda)$ can be neglected, $Z_0 = e^{-V_0(\Lambda)}$ should not be very dependent on this translation and so this calculation should be a good way to "almost" evaluate the generating function $\Sigma(J, \bar{J})$.

Proof (of Proposition 5.5) We use (89) followed by (90) to obtain

$$\hat{\chi}_\Lambda(m^2, g_0, \nu_0, z_0) = |\Lambda|^{-1} D^2_{\bar{J},J} \left(e^{(J,C\bar{J})} \Big[\mathbb{E}^{(m,\zeta)} \big[\theta Z_0 \big] \Big]_{|_{\phi = CJ, d\phi = 0}} \right) (0; 1, 1). \tag{92}$$

The desired result is obtained by evaluating the directional derivatives in the direction 1 noting that $C1 = (m^2 - \Delta)^{-1} 1 = m^{-2}$.

6 The Renormalisation Group

The renormalisation group is a method to evaluate $D^2 F(0; 1, 1)$ in the right hand side of Proposition 5.5. From this point on these notes become a selection of topics from the six papers that collectively comprise the proof of Theorem 2.2. There are many references to these papers, but I suggest that anyone who wishes to continue reading ignore these references until some of the general ideas emphasised by these notes start to come into focus. Many of the important ideas are also discussed in great detail in [15] for much simpler problems.

As discussed in Sect. 1 the renormalisation group can be defined in different ways, which are different interpretations of what the phrase "integrating out fluctuations" should mean. In our case we are going to write the gaussian field ϕ as a sum $\zeta_1 + \cdots + \zeta_N$ of independent gaussian fields ζ_j and then integrate over ζ_1, followed by ζ_2, and so on. First we will discuss the representation in distribution of ϕ as $\zeta_1 + \cdots + \zeta_N$. This depends on the following theorem [11, Sect. 6.1] about the inverse $(m^2 - \Delta)^{-1}$ where $\Delta = \Delta_\Lambda$ is the finite difference Laplacian for the torus Λ defined in (1). Recall that Λ has period L^N.

Theorem 6.1 (Finite Range Decomposition) *For $m^2 > 0$ let $C = (m^2 - \Delta_\Lambda)^{-1}$, regarded as a $\Lambda \times \Lambda$ matrix. There exist positive-definite $\Lambda \times \Lambda$ matrices $C_j = C_j(m^2)$ defined for $j = 1, 2, \ldots, N-1$ and $m^2 \geq 0$, and there exists $C_{N,N} =$*

$C_{N,N}(m^2)$ defined for $m^2 > 0$, such that

1. $(m^2 - \Delta_\Lambda)^{-1} = \sum_{j=1}^{N-1} C_j + C_{N,N}$,
2. *For* $j = 1, \dots, N-1$, $C_{j;x,y} = 0$, *if* $|x-y| \geq \frac{1}{2}L^j$,
3. $|C_{j;x,y}| \leq c\big(1 + m^2 L^{2j}\big)^{-k} L^{-2(j-1)}$, $j = 1, \dots, N-1$.

Finite difference derivatives up to any fixed order p are also bounded, according to

$$|\nabla_x^\alpha \nabla_y^\beta C_{j;x,y}| \leq c(1 + m^2 L^{2(j-1)})^{-k} L^{-(j-1)(2+(|\alpha|_1+|\beta|_1))},$$

where $c = c(p, k, \bar{m}^2)$ is independent of m^2, j, L.

The matrix entries $C_{j;x,y}$ are Λ independent functions of $z \in \mathbb{Z}^d$ where $(x, y) \mapsto z$ is defined for x, y in Λ by setting z equal to the minimal $\mathbb{Z}^d$ representative of $x - y$ in Λ. Part (3) also holds for $j = N$ provided $m^2 L^{2N}$ is bounded away from zero.

Since $C_{j;x,y} = 0$ when the representative z has $|z| \geq L^j/2$ it does not matter how the minimal representative is chosen when there is more than one. If a covariance G can be written as a sum of positive-definite matrices as in (1, 2) then we say that G has a *finite range decomposition*. Quite a large class of covariances are shown to have finite range decompositions in [4, 7, 18, 20], where the proof in [7] is particularly economical. Exactly what class of covariances have finite range decompositions is an open question of great interest to us.

Property (3) is special to the covariance $(m^2 - \Delta)^{-1}$. It expresses in a crude way that the covariances C_j are approximately scalings of each other for j such that $m^2 L^j \ll 1$. When j is too large for this to hold the covariances become small because these covariances have to be consistent with the exponential decay that C_{xy} is known to have for $|x - y| \gg m^{-1}$. In fact in the finite range decomposition discovered in [20] they are double exponentially small in $j \leq N$ such that $m^2 L^j \gg 1$.

Theorem 6.2 (ϕ as Sum of Increments) *For each covariance C_j of Theorem 6.1, let $\mathbb{E}_j$ be the super-expectation given by Definition 4.1, but with $A = C_j^{-1}$, and denote by ζ_j the associated gaussian field. Let $Z : \phi \mapsto Z(\phi)$ be a bounded function defined on $\mathbb{C}^\Lambda$. Then*

$$\mathbb{E}^{(m)}\big[Z(\phi)\big] = \mathbb{E}_N \mathbb{E}_{N-1} \dots \mathbb{E}_1\big[Z(\zeta_1 + \dots + \zeta_N)\big]. \tag{93}$$

Furthermore, this also holds for smooth bounded forms $Z : (\phi, d\phi) \mapsto Z(\phi, d\phi)$ in $\mathcal{N}(\Lambda)$, with the understanding that $d\phi$ becomes $d\zeta_1 + \dots + d\zeta_N$ on the right hand side.

For the case where Z is a function on $\mathbb{C}^\Lambda$, we have a form of degree zero which is just another name for a random variable. We have seen that in this case the super-expectation is the usual expectation of probability; the theorem is just a restatement of the standard fact that the distribution of a sum of independent gaussian random variables $\zeta_1, \dots, \zeta_N$ is also gaussian with covariance equal to the

sum of the covariances of the ζ's. Thus the new content in this theorem is in the case where Z is a form. For a proof of an equivalent result see [16, Proposition 2.6].

6.1 Progressive Integration

Now we re-organise Theorem 6.2 into an iterative procedure where the fields $\zeta_1, \dots, \zeta_N$ are successively integrated out. The process is: for each $j = 1, \dots, N$, replace ϕ by $\phi + \zeta_j$, fix ϕ and integrate over ζ_j with respect to the gaussian probability distribution with covariance C_j. The accumulation of all these replacements is the same as replacing ϕ by $\phi + \zeta_1 + \cdots + \zeta_N$. Setting $\phi = 0$ and taking the $\mathbb{E}_j$ expectations reproduces the right hand side of Theorem 6.2. Since in this procedure we only deal with one fluctuation field ζ_j at a time we often write ζ in place of ζ_j, using the subscript j on $\mathbb{E}_j$ to show which field is being integrated out.

Recall the discussion around Definition 5.4 where we have defined the algebra homomorphism $\theta : \mathscr{N}(\Lambda) \to \mathscr{N}(\Lambda \sqcup \Lambda)$ to be the map that replaces ϕ by $\phi + \zeta$ and $\mathrm{d}\phi$ by $\mathrm{d}\phi + \mathrm{d}\zeta$. Returning to the context of Proposition 5.5, we define a sequence of forms in $\mathscr{N}(\Lambda)$ by the recursion

$$\begin{aligned} Z_0 &= \mathrm{e}^{-V_0(\Lambda)}, \\ Z_{j+1} &= \mathbb{E}_{j+1}\big[\theta Z_j\big], \quad j = 0, 1, \dots, N-1. \end{aligned} \tag{94}$$

We will keep referring back to this important sequence because our goal is to calculate Z_N which is the function F in Proposition 5.5 that determines the susceptibility $\hat{\chi}_\Lambda$ that appears in Theorem 5.2.

6.2 First Order Perturbation Theory

The recursion (94) defines forms Z_j. We will now be interested in how the functional dependence of Z_j on arguments $\phi, d\phi$ changes under the map $Z_j \mapsto Z_{j+1}$. In this section we examine this in a preliminary way using perturbation theory. Perturbation theory is purely algebraic: we fix an order p and regard Z_j as a power series in V_j where

$$V_j = g_j \tau^2 + L^{-2j} \mu_j \tau + z_j \tau_\Delta. \tag{95}$$

Then we calculate modulo V_j^{p+1}, or equivalently, modulo monomials of degree $p+1$ in g_j, z_j, μ_j. This is called *pth order perturbation theory*. Notice that we have started to write V_j in terms of μ_j instead of ν_j, where

$$\nu_j = L^{2j} \mu_j. \tag{96}$$

Also we are writing the parameters in the order g_j, z_j, μ_j instead of the former order g_j, μ_j, z_j because of a triangular property of some forthcoming equations for these coupling constants.

The proof of Theorem 5.2 requires second order perturbation theory, that is $p = 3$, but first order is easier and shows why coupling constants are scale dependent, so we do the first order calculation in detail and then summarise the conclusions of second order calculations. For first order perturbation theory, let g_j, μ_j, z_j be given real numbers, subject to the usual integrability constraints $g_j > 0, z_j > -1$. If $Z_j = \mathrm{e}^{-V_j(\Lambda)}$ modulo V_j^2, then I claim that, modulo V_j^2 (or V_{j+1}^2),

$$Z_{j+1} = \mathrm{e}^{-V_{j+1}(\Lambda)}, \tag{97}$$

where V_{j+1} is defined with coupling constants

$$\begin{aligned} g_{j+1} &= g_j, \\ \mu_{j+1} &= L^2\mu_j + 2L^{2j+2}C_{j+1,0,0}\, g_j, \\ z_{j+1} &= z_j. \end{aligned} \tag{98}$$

Here we see a very important idea: there is a scale dependent coupling constant $j \mapsto \mu_j$ determined by the above recursion. According to part (3) of Theorem 6.1, for $m^2 = 0$, we have $L^{2j+2}C_{j+1,0,0} = O(1)$. In fact in Proposition 6.1 of [11] we prove that $L^{2j+2}C_{j+1,0,0}$ has a positive limit as $j \to \infty$ so this recursion of μ_j becomes independent of j for j large.

Proof (of Claim) Modulo V^2 we have

$$\begin{aligned} \mathbb{E}_{j+1}\big[\theta Z_j\big] &= \mathbb{E}_{j+1}\Big[\big(1 - V_j(\Lambda)\big)\Big] \\ &= 1 - \mathbb{E}_{j+1}\big[V_j(\Lambda)\big] = \mathrm{e}^{-\mathbb{E}_{j+1}[V_j(\Lambda)]}. \end{aligned} \tag{99}$$

Therefore it is sufficient to prove that

$$\mathbb{E}_{j+1}\big[\theta V_j(\Lambda)\big] \;=\; V_{j+1}(\Lambda). \tag{100}$$

Let $Q := \mathbb{E}_{j+1}\theta\tau_x^2$. We start by calculating Q. By the definition (47) of τ,

$$\tau_x^2 = (\phi_x\bar{\phi}_x)^2 + 2\,\phi_x\bar{\phi}_x\, c\, d\phi_x \wedge d\bar{\phi}_x. \tag{101}$$

Applying θ replaces ϕ by $\phi + \zeta$ in this expression. Therefore the terms $Q(0)$ in Q that do not depend on ϕ are

$$Q(0) = \mathbb{E}_{j+1}\big((\zeta_x\bar{\zeta}_x)^2 + 2\,\zeta_x\bar{\zeta}_x\, c\, d\zeta_x \wedge d\bar{\zeta}_x\big) \tag{102}$$

which is the ζ expectation of τ_x^2 with ϕ replaced by ζ. By the last line in Lemma 4.2 applied with $F(\tau) = \tau_x^2$ this equals zero. Next we calculate the terms $Q^{(2)}$ in Q that are of degree 2 in $\phi, d\phi$. From θ applied to (101) and using gauge invariance as in (16) to escape writing some terms,

$$\begin{aligned} Q^{(2)} &= \mathbb{E}_{j+1}\big(4\phi_x\bar\phi_x\zeta_x\bar\zeta_x + 2\,\phi_x\bar\phi_x\, c\, d\zeta_x \wedge d\bar\zeta_x + 2\,\zeta_x\bar\zeta_x\, c\, d\phi_x \wedge d\bar\phi_x\big) \\ &= \mathbb{E}_{j+1}\big(2\phi_x\bar\phi_x\zeta_x\bar\zeta_x + 2\phi_x\bar\phi_x[\zeta_x\bar\zeta_x + \, c\, d\zeta_x \wedge d\bar\zeta_x] + 2\,\zeta_x\bar\zeta_x\, c\, d\phi_x \wedge d\bar\phi_x\big) \\ &= \mathbb{E}_{j+1}\big(2\phi_x\bar\phi_x\zeta_x\bar\zeta_x + 2\,\zeta_x\bar\zeta_x\, c\, d\phi_x \wedge d\bar\phi_x\big) = 2\tau_x\,\mathbb{E}_{j+1}\big(\zeta_x\bar\zeta_x\big)\,. \end{aligned} \tag{103}$$

In the third equality we used the fact that the expectation of the terms in square brackets is zero by the last line in Lemma 4.2 applied with $F(\tau) = \tau_x$. The last equality holds by the definition (47) of τ. Since the terms $Q^{(4)}$ of degree 4 in $\phi, d\phi$ in Q are τ_x^2 and since $Q = Q(0) + Q^{(2)} + Q^{(4)}$ these formulas for $Q(0)$, $Q^{(2)}$ and $Q^{(4)}$ imply that

$$\mathbb{E}_{j+1}\big[\theta\tau_x^2\big] = \tau_x^2 + 2\,\tau_x\,\mathbb{E}_{j+1}\zeta_x\bar\zeta_x. \tag{104}$$

By a much shorter version of the same calculation $\mathbb{E}_{j+1}\theta\tau_x = \tau_x$. By summing these formulas over x in Λ and using the definition $C_{j+1,x,x} = \mathbb{E}_{j+1}\zeta_x\bar\zeta_x$ of the covariance and translation invariance of the covariance we have proved (100).

6.3 Second Order Perturbation Theory

A much more efficient method for calculating in perturbation theory is used in [11]. It is based on [16, Proposition 2.6] and [11, Lemma 5.2]. In second order perturbation theory, where one calculates modulo V^3, the function $V \mapsto e^{-V(\Lambda)}$ does not retain its form under the map $\mathbb{E}_{j+1}\theta$, and we have modify it so that it is form invariant. In [11, (3.21)] we construct a sequence of explicit functions $V \mapsto W_j(V, \Lambda)$ that are quadratic in V and take values in $\mathscr{N}(\Lambda)$. There is a map [11, (4.11)]

$$\varphi^{(0)}_{\mathrm{pt},j} : \mathbb{R}^3 \to \mathbb{R}^3 \tag{105}$$

such that the function

$$I_j \,:\, (V, \Lambda) \mapsto e^{-V(\Lambda)}(1 + W_j(V, \Lambda)) \tag{106}$$

satisfies, modulo V^3,

$$\mathbb{E}_{j+1}\big[\theta I_j(V, \Lambda)\big] = I_{j+1}\big(V \circ \varphi^{(0)}_{\mathrm{pt},j}, \Lambda\big), \tag{107}$$

where we regard V as a function of its coupling constants so that the composition $V \circ \varphi^{(0)}_{\mathrm{pt},j}$ makes sense. Equation (107) says that integrating out the fluctuation field ζ_{j+1} in $\theta I_j(V, \Lambda)$ is equivalent to changing the three coupling constants g, μ, z in V according to the map $\varphi^{(0)}_{\mathrm{pt},j}$. This is what we mean when we say that I_j retains its form at second order under the map $\mathbb{E}_{j+1}\theta$.

We will not need the explicit formula [11, (4.11)] for the map $\varphi^{(0)}_{\mathrm{pt},j}$ because it is conjugate to a simpler map. According to [11, Proposition 4.3], for each scale $j = 0, \dots, N-1$ there is an explicit quadratic map $T_j : \mathbb{R}^3 \to \mathbb{R}^3$, which is almost the identity map,

$$T_0(V) = V, \quad T_j(V) = V + O(\|V\|^2). \tag{108}$$

Therefore it is invertible near the origin. Furthermore, modulo terms that only contribute to the discarded $O(V^3)$,

$$\varphi^{(0)}_{\mathrm{pt},j} = T_{j+1}^{-1} \circ \bar{\varphi}_j \circ T_j, \tag{109}$$

where $\bar{\varphi}_j : \mathbb{R}^3 \to \mathbb{R}^3$ defines, by $(\bar{g}_{j+1}, \bar{z}_{j+1}, \bar{\mu}_{j+1}) = \bar{\varphi}_j(\bar{g}_j, \bar{z}_j, \bar{\mu}_j)$, the recursion

$$\bar{g}_{j+1} = \bar{g}_j - \beta_j \bar{g}_j^2, \tag{110}$$

$$\bar{z}_{j+1} = \bar{z}_j - \theta_j \bar{g}_j^2, \tag{111}$$

$$\bar{\mu}_{j+1} = L^2 \bar{\mu}_j (1 - \gamma \beta_j \bar{g}_j) + \eta_j \bar{g}_j - \xi_j \bar{g}_j^2 - \pi_j \bar{g}_j \bar{z}_j, \tag{112}$$

for scale dependent coupling constants.

The coefficients $\beta_j, \theta_j, \eta_j, \xi_j, \pi_j$ are real coefficients defined precisely in [11, (3.24), (3.27), (3.28)]. These coefficients, and also those of the transformations T_j, are independent of the side L^N of the torus. This means that all the recursions obtained as N varies are consistent and it is not necessary to know what N is so long as N is larger than whatever scale j one is considering. Thus there is a formal infinite volume recursion where j runs over all integers $j \geq 0$, not just $j < N$. To put it another way, we can use this infinite volume recursion for $j < N$ since it coincides with the N recursion until $j = N$.

The sequence $(\beta_j)_{0 \leq j < \infty}$ plays a key role in the analysis. These are positive numbers that have a non zero limit when $m^2 = 0$ and the dimension is 4. For $m^2 > 0$ they decay faster than exponentially to zero, but the decay does not set in until the scale j is large enough that $L^{2j} m^2$ is roughly 1. We make the following definitions [9, (1.7)]. Given $\Omega > 1$, we define a scale j_Ω as the first scale where the exponential decay sets in,

$$j_\Omega = \inf\{k \geq 0 : |\beta_j| \leq \Omega^{-(j-k)_+} \beta_{\max} \text{ for all } j \geq 0\}, \tag{113}$$

where $\beta_{\max} = \max_j \beta_j$. The infimum of the empty set is defined to equal ∞, e.g., if $\beta_j = b$ for all j then $j_\Omega = \infty$. The choice of Ω is arbitrary; let $\Omega = 10$. We also

define

$$\chi_j = \Omega^{-(j-j_\Omega)_+}. \tag{114}$$

Thus the sequence χ_j is a normalised version of β_j which equals one for $j \le \Omega$ and decays exponentially for $j > \Omega$.

Unlike the recursion defined by $\varphi^{(0)}_{\mathrm{pt},j}$ the $\bar\varphi$ recursion (110)–(112) is triangular: the $\bar g$-equation does not depend on $\bar z$ or $\bar\mu$, the $\bar z$-equation depends only on $\bar g$, and the $\bar\mu$-equation depends both on $\bar g$ and $\bar z$. Therefore we can solve the recursion one equation at a time and thereby prove the following Proposition, which follows from [9, Proposition 1.2]. In its statement, z_∞ denotes the limit $z_\infty = \lim_{j\to\infty} z_j$, and similarly for μ_∞.

Proposition 6.3 *If $\bar g_0 > 0$ is sufficiently small, then there exists a unique global solution $(\bar g_j, \bar z_j, \bar\mu_j)_{j\in\mathbb{N}_0}$ to the recursion (110)–(112) with initial condition $\bar g_0$ and final condition $(\bar z_\infty, \bar\mu_\infty) = (0,0)$. This flow satisfies, for any real $p \in [1,\infty)$,*

$$\chi_j \bar g_j^p = O\left(\frac{\bar g_0}{1+\bar g_0 j}\right)^p, \quad \bar z_j = O(\chi_j \bar g_j), \quad \bar\mu_j = O(\chi_j \bar g_j), \tag{115}$$

with constants independent of j_Ω and $\bar g_0$, and dependent on p in the first bound. Furthermore, $(\bar g_j, \bar z_j, \bar\mu_j)$ is continuously differentiable in the initial condition $\bar g_0$ and continuous in the mass parameter $m^2 \ge 0$, for every $j \in \mathbb{N}_0$.

The proposition shows that when $m^2 = 0$ the coupling constants $\bar g_j$ tends to zero like j^{-1}. This is called *asymptotic freedom*. It suggests that perturbation theory becomes more accurate at larger scales and the theory becomes more gaussian, as needed for the "evaluation as if gaussian" strategy that started in Sect. 5.1.

6.4 The Error Coordinate

The perturbation theory of the preceding sections suggests that we approximate Z_j defined in (94) by $I_j(V_j, \Lambda)$ defined in (106) and then, by (107), Z_{j+1} should be approximately equal to $I_{j+1}(V_{j+1}, \Lambda)$ where the coupling constants in V_{j+1} are obtained by applying the map (109) to the coupling constants in V_j. However our calculations have given us formulas which hold in a very weak algebraic sense since we have worked modulo $O(V^3)$. For example there is no uniformity in Λ. Now we need a way to include the $O(V^3)$ error terms so that all our formulas hold in the usual sense of equality. For scale j in the range $0, \dots, N$ we will write Z_j as a function of (V_j, K_j) where V_j specifies the I_j approximation by perturbation theory and the *error coordinate* K_j specifies the difference between Z_j and perturbation theory. Our main result for this section is Lemma 6.6 where we show that Z_{j+1} is the scale equivalent function of V_{j+1} and the error coordinate K_{j+1} on the next scale.

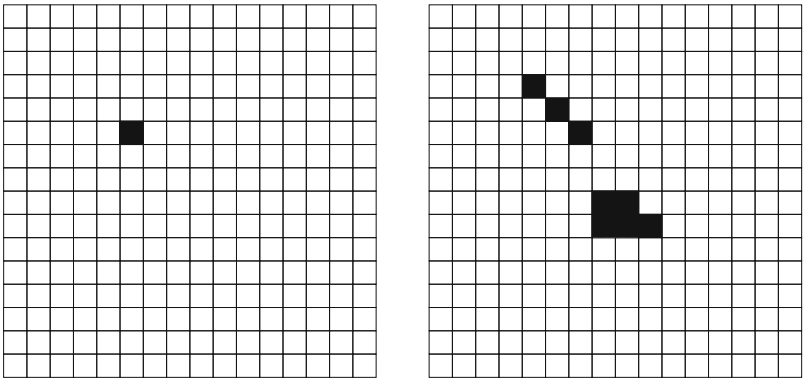

Fig. 1 Illustration of both blocks B with side length L^j and a polymer at scale j

Lemma 6.6 is an important step in our definition of the renormalisation group map, but it is not a complete description of our formalism. Firstly, we only give the version for first order perturbation theory, and secondly it is but one of a family of ways to define K_{j+1} such that Z_{j+1} is represented by (V_{j+1}, K_{j+1}). This many-to-one aspect is important for overcoming a problem that we will discuss in Sect. 7.2.

Geometry Let B be a block as in Fig. 1 with side length L^j and denote by $\mathscr{B}_j(\Lambda)$ the set of all such j-blocks or *blocks at scale j*. By definition a *j-polymer* or *polymer at scale j* is a union of j-blocks. A polymer can be the empty set. If X is a j-polymer, then $\mathscr{B}_j(X)$ denotes the set of j-blocks contained in X. Notice that polymers at scale j are also polymers at scales smaller than j because we have chosen the sides of blocks to be powers of L. Λ is a polymer at all scales and is a single block at scale N. We denote by $\mathscr{P}_j(\Lambda)$ the set of j-polymers in Λ, and for a j-polymer X let $\mathscr{P}_j(X)$ denote the set of all j-polymers contained in X. Two polymers X, Y are said to *touch* if there is a point x in X and a point y in Y such that $|x - y|_\infty = 1$. A polymer X is said to be *connected* if it is not empty and if whenever Y and Y' are nonempty polymers such that $X = Y \cup Y'$ then Y, Y' touch. For each scale j and j-polymer X let

$$X \mapsto X^{\square} \tag{116}$$

be an assignment of a polymer $X^{\square}$ that contains X. We think of this as an assignment of a neighbourhood of X to X. The assignment must be translation invariant and satisfy $(X \cup Y)^{\square} = X^{\square} \cup Y^{\square}$. Our specific choice is

$$X^{\square} = \bigcup \{Y | Y \supset X, Y \in \mathscr{S}\}, \tag{117}$$

where $\mathscr{S}$ is an important class of sets specified in

Definition 6.4 For any scale j we say that a polymer X in $\mathscr{P}_j(\Lambda)$ is a small polymer if X is connected and is a union of at most 2^d blocks. Let $\mathscr{S}_j$ be the set of small scale j polymers.

Thus the neighbourhood assigned to a block on scale j has diameter $O(L^j)$ and, for L large enough, when j-polymers X and X' are separated by a distance L^{j+1},

$$\operatorname{dist}(X^\square, X'^\square) > \frac{1}{2}L^{j+1} \tag{118}$$

holds for all scales j. We will always assume that L is at least this large.

For $X \subset \Lambda$ let $\mathscr{N}(X)$ be the set of forms in $\mathscr{N}(\Lambda)$ that only depend on $\phi_x, d\phi_x$ (and their conjugates for $x \in X$). These sets $\mathscr{N}(X)$ are the form analogues of σ-algebras of random variables. For example, when j-polymers X and X' are separated by a distance L^{j+1}, $A \in \mathscr{N}(X^\square)$ and $B \in \mathscr{N}(X'^\square)$, then θA and θB are ζ independent,

$$\mathbb{E}_{j+1}\theta\Big[AB\Big] = \mathbb{E}_{j+1}\Big[(\theta A)(\theta B)\Big] = \mathbb{E}_{j+1}\big[\theta A\big]\,\mathbb{E}_{j+1}\big[\theta B\big], \tag{119}$$

by item 2 of Theorems 6.1 and (118).

Let $B \mapsto I_j(B)$ assign to each block in $\mathscr{B}_j(\Lambda)$ a form $I_j(B) \in \mathscr{N}(B^\square)$. For example our discussion of first order perturbation theory suggests

$$I_j(B) \;=\; \mathrm{e}^{-V_j(B)}, \tag{120}$$

and in this case the condition $I_j(B) \in \mathscr{N}(B^\square)$ holds since the term $z_j\tau_\Delta(B)$ in $V_j(B)$ depends only on fields ϕ_x for x in B or x a nearest neighbour to some site in B. The other terms in $V_j(B)$ only depend on fields in B. For second order perturbation theory as in (106) the formula [11, (3.21)] for W_j also satisfies $I_j(B) \in \mathscr{N}(B^\square)$. For a polymer$X \in \mathscr{P}_j$ let

$$I^X = \prod_{B\in\mathscr{B}(X)} I(B). \tag{121}$$

Definition 6.5 Let $\mathscr{K}_j$ be the set of maps $K : \mathscr{P}_j(\Lambda) \to \mathscr{N}(\Lambda)$ such that

1. $K(\varnothing) = 1$,
2. For $X \in \mathscr{P}_j$, $K(X) \in \mathscr{N}(X^\square)$,
3. If X and Y are polymers that do not touch then $K(X \cup Y) \;=\; K(X)K(Y)$,
4. Symmetry properties.

The symmetry properties express invariance under lattice automorphism and supersymmetry. See [17, Definition 1.7] for details, but they are not needed here.

At scale $j = 0$, let

$$I_0(B) \;=\; \mathrm{e}^{-V_0(B)} \quad \text{and} \quad K_0(X) = \begin{cases} 1, & X = \varnothing \\ 0, & \text{else} \end{cases} \tag{122}$$

then

$$Z_0 = \mathrm{e}^{-V_0(\Lambda)} = \sum_{X \in \mathscr{P}_0(\Lambda)} I_0^{\Lambda \setminus X} K_0(X). \tag{123}$$

More generally, for any elements A, B of $\mathscr{K}_j$ we define a new element $A \circ B$ of $\mathscr{K}_j$ by

$$A \circ B(Y) = \sum_{X \in \mathscr{P}_j(Y)} A(X)\, B(Y \setminus X). \tag{124}$$

This product is easily verified to be commutative and associative with identity

$$\mathbb{1}_{\varnothing}(X) = \begin{cases} 1, & X = \varnothing \\ 0, & \text{else} \end{cases} \tag{125}$$

(which is the same function as K_0.) Then, after extending the domain of I_j from $\mathscr{B}_j(\Lambda)$ to $\mathscr{P}_j(\Lambda)$ by setting

$$I_0(X) = I_0^X, \quad X \in \mathscr{P}_j(\Lambda) \tag{126}$$

we can now write (123) in the shorter form

$$Z_0 = I_0 \circ K_0(\Lambda). \tag{127}$$

Notice also that the $\circ$ product depends on scale. For example at scale N, $\mathscr{B}_N(\Lambda_N)$ consists only of $\varnothing, \Lambda$, therefore for A, B that equal 1 when evaluated on the empty set $\varnothing$,

$$A \circ B(\Lambda) = A(\Lambda) + B(\Lambda). \tag{128}$$

The next result, which is the main result of this subsection, is stated for first order perturbation theory. We use a more complicated second order version in the proof of Theorem 2.2. We implicitly assume integrability. When we discuss norms there will be a property called the integration property that takes care of this issue.

Lemma 6.6 *Let j be in $\{0, 1, \ldots, N-1\}$ and for $k = j, j+1$ let $I_k = e^{-V_k}$. Given $K_j \in \mathscr{K}_j$ there exists $K_{j+1} \in \mathscr{K}_{j+1}$ such that*

$$\mathbb{E}_{j+1}\theta\Big[I_j \circ K_j(\Lambda)\Big] = I_{j+1} \circ K_{j+1}(\Lambda). \tag{129}$$

Thus if (V_j, K_j) *are such that* $Z_j = I_j \circ K_j(\Lambda)$*, then* $Z_{j+1} = I_{j+1} \circ K_{j+1}(\Lambda)$*. A formula for* K_{j+1} *is given in* (137)*.*

In particular, we can make any choice of $V_1, \dots, V_N$ and then by induction based on this Lemma and by (128) there exists K_N such that

$$Z_N = I_N \circ K_N(\Lambda) = I_N(\Lambda) + K_N(\Lambda). \tag{130}$$

By combining this result with Proposition 5.5 we have

$$\hat{\chi}_\Lambda(m^2, g_0, \nu_0, z_0) = \frac{1}{m^2} + \frac{1}{m^4\,|\Lambda|}\left(D^2 I_N(\Lambda; 0; 1, 1) + D^2 K_N(\Lambda; 0; 1, 1)\right), \tag{131}$$

Notation When we are concerned with a fixed scale and the transition to the next scale we clean up the equations by suppressing the subscript j, writing I and K in place of I_j and K_j, etc., and shorten the $j+1$ subscript to $+$ so that, for example, K_{j+1}, $\mathbb{E}_{j+1}$, $\mathscr{P}_{j+1}$ become K_+, $\mathbb{E}_+$, $\mathscr{P}_+$.

I have the following pictorial view of the proof of Lemma 6.6, but you lose nothing in the logical sense by skipping this. I give the actual proof below. The representation $I \circ K(\Lambda)$ is a sum over ways to partition Λ into a polymer $\Lambda \setminus X$ weighted by a product of I factors, one per block in $\Lambda \setminus X$, and a polymer X weighted by $K(X)$. The left hand side in Fig. 2 represents one term in this sum over partitions: the white region is the polymer $\Lambda \setminus X$ and each square in the white region represents a block B with a factor $I(B)$. The blue region is the polymer X for which there is a factor $K(X)$. If we apply the algebra homomorphism θ to $I \circ K(\Lambda)$ then the white region represents a product over blocks B of $\theta I(B)$ and the blue represents $\theta K(X)$. In the first step (133) of the proof below each $\theta I(B)$ is expanded into $I_+ + \delta$ (which defines an error term δ). Thus there arises a sum of ways to colour a subset of the white blocks red, each red block B denoting a $\delta(B)$. The remaining white blocks are

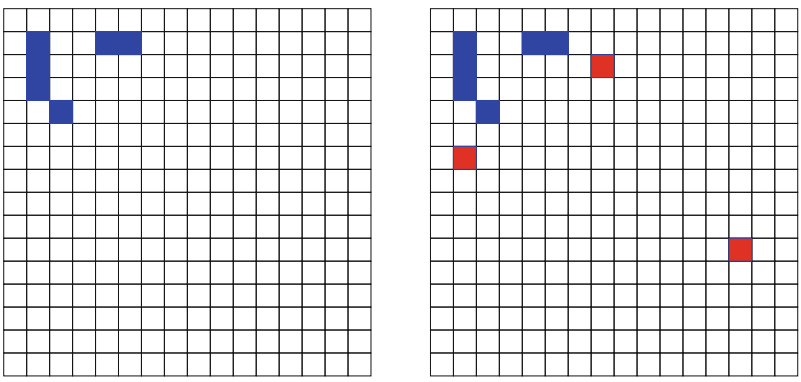

Fig. 2 Illustration of Eq. (135)

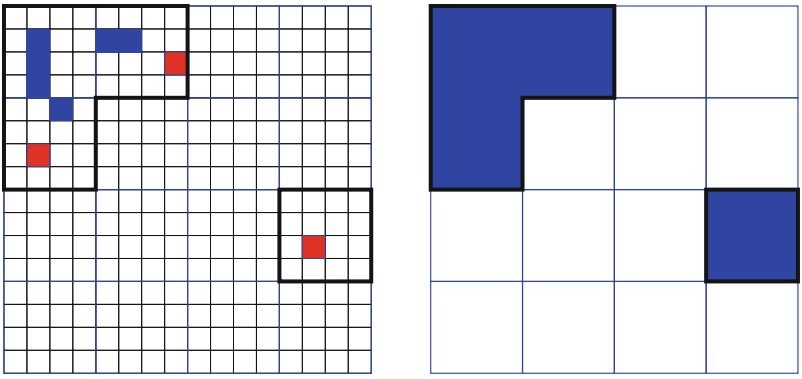

Fig. 3 Illustration of Eq. (138)

now I_+ factors as in the right hand side of Fig. 2, which stands for just one term in a sum over partitions into red, white and blue. The next step (138) in the proof is a passage to blocks on the next scale, accomplished by considering the smallest scale $j+1$ polymer Y that covers the red and blue region, as indicated in the left hand side of Fig. 3. The right hand side of Fig. 3 represents the sum $\Sigma(Y)$ over all red, white and blue partitions as in the left hand side that generate the same polymer Y. Since this right hand picture is a partition of Λ into a white polymer $\Lambda \setminus Y$ that represent products of I_+ blocks and a blue polymer Y, and since these are polymers on the next scale, this picture is a term in $I_+ \circ \Sigma$ and it tells us to define $K_+(Y) = \mathbb{E}_{j+1}\Sigma(Y)$ because then we match the desired right hand side of the Lemma.

Proof (of Lemma 6.6.) Define $B \mapsto \delta(B)$ on $\mathscr{B}(\Lambda)$ by

$$\delta = \theta I - I_+. \tag{132}$$

Notice that θI is evaluated on scale j blocks so we have to extend the domain of I_+ which is defined on scale $j+1$ block to include scale j blocks, but we can do this easily for first order perturbation theory since V_{j+1} can be summed over points in blocks of any scale: we set $I_{j+1}(X) = e^{-V_{j+1}(X)}$ for X any scale block. For later in the proof note that δ depends on ζ because of the θ, but I_+ is not dependent on ζ. For the next equation we also extend the domain of δ from blocks to polymers by writing $\delta(X) = \delta^X$ for a $X \in \mathscr{P}$. Then, for $X \in \mathscr{P}$,

$$\begin{aligned} \theta\, I^X = \big(I_+ + \delta\big)^X &= \sum_{X_\delta \in \mathscr{P}(X)} I_+^{X\setminus X_\delta}\, \delta^{X_\delta} \\ &= \sum_{X_\delta \in \mathscr{P}(X)} I_+^{X\setminus X_\delta}\, \delta(X) = I_+ \circ \delta(X). \end{aligned} \tag{133}$$

Let

$$H = \delta \circ \theta K. \tag{134}$$

Then

$$\begin{aligned}\theta(I \circ K)(\Lambda) &= \theta I \circ \theta K(\Lambda) \\ &= \big(I_+ \circ \delta\big) \circ \theta K(\Lambda) = I_+ \circ \big(\delta \circ \theta K\big)(\Lambda) \\ &= I_+ \circ H(\Lambda). \end{aligned} \tag{135}$$

For $X \in \mathscr{P}(\Lambda)$, define the closure $\overline{X}$ of X on the next scale to be smallest polymer in $\mathscr{P}_+(\Lambda)$ containing X. For Y in $\mathscr{P}_+$ let $\overline{\mathscr{P}}(Y)$ be the set of all polymers in $\mathscr{P}$ whose closure is Y, then

$$\begin{aligned} I_+ \circ H(\Lambda) &= \sum_{X \in \mathscr{P}(\Lambda)} I_+^{\Lambda\setminus X} H(X) \\ &= \sum_{Y \in \mathscr{P}_+(\Lambda)} \sum_{X \in \overline{\mathscr{P}}(Y)} I_+^{\Lambda\setminus X} H(X) \\ &= \sum_{Y \in \mathscr{P}_+(\Lambda)} I_+^{\Lambda\setminus Y} \sum_{X \in \overline{\mathscr{P}}(Y)} I_+^{Y\setminus X} H(X). \end{aligned} \tag{136}$$

For Y in $\mathscr{P}_+(\Lambda)$ let

$$\begin{aligned} K_+(Y) &= \mathbb{E}_+\Big[\sum_{X \in \overline{\mathscr{P}}(Y)} I_+^{Y\setminus X} H(X) \Big] \\ &= \sum_{X \in \overline{\mathscr{P}}(Y)} I_+^{Y\setminus X}\, \mathbb{E}_+\big[H(X)\big]. \end{aligned} \tag{137}$$

Then by (135)–(137)

$$\begin{aligned} \mathbb{E}_+\theta\big[I_j \circ K_j(\Lambda)\big] &= \sum_{Y \in \mathscr{P}_+(\Lambda)} I_+^{\Lambda\setminus Y}\, \mathbb{E}_+\Big[\sum_{X \in \overline{\mathscr{P}}(Y)} I_+^{Y\setminus X} H(X) \Big] \\ &= \sum_{Y \in \mathscr{P}_+(\Lambda)} I_+^{\Lambda\setminus Y} K_+(Y) \;=\; I_+ \circ K_+(\Lambda), \end{aligned} \tag{138}$$

which verifies the desired property with respect to $\mathbb{E}_+$ of K_+.

We now outline the proof that K_+ defined by (137) is in $\mathscr{K}_+$. This is where the finite range property of Theorem 6.1 plays its crucial role and part (2) of Definition 6.5 and the analogous property of I are used. If K_+ is evaluated on $X \cup X'$

where X and X' are polymers in $\mathscr{P}_+$ that do not touch, then the distance between X and X' is at least L^{j+1} because they are each unions of $j+1$ scale blocks which cannot touch. Therefore factors of δ evaluated on blocks in X are independent of factors of δ evaluated on blocks in X'. Furthermore, by part (3) of Definition 6.5 for K, when θK is evaluated on a subset of $X \cup X'$, it is a product of a factor of K evaluated on a subset of X and a factor of K evaluated on a subset of X'. These are also independent. Finally it is straightforward to check that the sums defining H and K_+ factor into sums of ways to partition X and a separate sum over ways to partition X'. Taken all together this implies that K_+ has property (3) of the definition of $\mathscr{K}_+$. The other properties of $\mathscr{K}_+$ are much easier to check.

7 The Norm of the Error Coordinate

So far all our analysis of the action of $\mathbb{E}_{j+1}\theta$ in terms of (V, K) has been algebraic with no hint on how the errors could be controlled. In this section we explain part of our formalism in [17] for the control of errors. In particular we will see that the error coordinate at scale j can be regarded as an element of a Banach space $\mathscr{F}_j$. The spaces $\mathscr{K}_j, j = 0, 1, \ldots, N$, for the error coordinate introduced in Definition 6.5 are not vector spaces because part (3) of Definition 6.5 is not linear, but the restrictions of elements of $\mathscr{K}_j$ to connected polymers form a vector space as in the following definition.

Definition 7.1 For $j = 0, 1, \ldots, N$, let $\mathscr{C}_j(\Lambda)$ be the set of connected polymers in $\mathscr{P}_j(\Lambda)$. Let $\mathscr{C}\mathscr{K}_j$ be the complex vector space under pointwise addition and scalar multiplication of maps $K : \mathscr{C}_j(\Lambda) \to \mathscr{N}(\Lambda)$ such that

1. For $X \in \mathscr{C}_j(\Lambda)$, $K(X) \in \mathscr{N}(X^{\square})$,
2. Symmetry properties,

where the symmetry properties are a repetition of part (4) of Definition 6.5.

Not only does an element of $\mathscr{K}_j$ determine an element of $\mathscr{C}\mathscr{K}_j$ by restriction, but, conversely, given an element K of $\mathscr{C}\mathscr{K}_j$ we can extend its domain from $\mathscr{C}_j(\Lambda)$ to $\mathscr{P}_j(\Lambda)$ by imposing property (3) of Definition 6.5: a polymer X in $\mathscr{P}_j(\Lambda)$ can be decomposed into its connected components $X_1, \ldots, X_n$ and then we define $K(X) = \prod_{i=1}^{n} K(X_i)$. If X is the empty set we define $K(X) = 1$.

7.1 The $\mathscr{F}$ Norm

7.1.1 Norm on $\mathscr{C}\mathscr{K}_j$

For each scale j the norm on $\mathscr{C}\mathscr{K}_j$ is constructed from a family of complete norms $\|F\|_j = \|F\|_{j,\mathscr{N}(X^{\square})}$, one for each space $\mathscr{N}(X^{\square})$ where X ranges over connected

j-polymers. Given such norms we define the norm $\|K\|_{\mathscr{F}_j}$ for $K \in \mathscr{C}\mathscr{K}$ by

$$\|K\|_{\mathscr{F}_j} = \sup_{X \in \mathscr{C}_j(\Lambda)} W(X) \|K(X)\|_{j,\mathscr{N}(X^\square)} \tag{139}$$

with a weight $X \mapsto W(X)$. In these notes we set

$$W(X) = \epsilon^{-|X|_j}, \tag{140}$$

where $|X|_j = |\mathscr{B}_j(X)|$ is the number of j-blocks in X and ϵ is a positive parameter smaller than one. This is not our choice in [17] but could have been and it is instructive.

7.1.2 The Norm on $\mathscr{N}(X^\square)$

The definitions are given in [16] and they take time to assimilate, so instead of repeating them, I list some desired properties and illustrate the role of these properties by using them axiomatically to prove a bound in the following section.

- *Product Property:* For all scales $j = 0, \dots, N-1$, for disjoint j-polymers X, Y, for forms A in $\mathscr{N}(X^\square)$, B in $\mathscr{N}(Y^\square)$ and AB in $\mathscr{N}((X \cup Y)^\square)$,

$$\|AB\|_j \ \le \ \|A\|_j \|B\|_j. \tag{141}$$

  See [16, Proposition 3.16]. All the spaces $\mathscr{N}(X)$, where X ranges over subsets of Λ, are subalgebras of $\mathscr{N}(\Lambda)$ so it makes sense to multiply A, B.
- *I_+ Bound:* There is a constant α_I and a coupling constant domain $\mathscr{D}_{j+1}$ for the coupling constants in V_{j+1} such that

$$\|e^{-V_{j+1}(B)}\|_{j+1} \ \le \ \alpha_I \tag{142}$$

  for $B \in \mathscr{B}_j(\Lambda)$.
- *Integration Property:* Recall the definition of δ from (132) and let $k = j$ or $j+1$. There is a constant $\alpha_{\mathbb{E}}$ such that for disjoint $X, Y \in \mathscr{P}_j$, $F(Y) \in \mathscr{N}(Y^\square)$ and V_k with coupling constants in $\mathscr{D}_k$,

$$\| \mathbb{E}_{j+1} \delta^X \theta F(Y)\|_{j+1} \ \le \ \alpha_{\mathbb{E}}^{|X|_j + |Y|_j} \epsilon_\delta^{|X|_j} \|F(Y)\|_j, \tag{143}$$

  where ϵ_δ tends to zero as the coupling constants in V_k tend to zero.

7.1.3 Comment on Norms

If we were not working with forms, but just functions of fields then the L^∞ norm would have these properties. In [16] we show how to construct norms that record information on derivatives by being equivalent to $\mathscr{C}^p$ norms, but which also have the product property. The construction is based on the idea that the Taylor expansion of a product AB is the product of the Taylor expansions for A and for B. As a corollary we can extend the norms to forms in $\mathscr{N}(X^\square)$ using the fact that coefficients of monomials in $d\phi$ and $d\bar{\phi}$ are analogous to coefficients in a Taylor expansion.

Another consideration is that at scale j an element F of $\mathscr{N}(X^\square)$ is not going to be evaluated on arbitrary fields $\phi \in \mathbb{C}^\Lambda$, but on fields that are typical for the distribution of $\phi = \sum_{k>j} \zeta_k$. Therefore, for each polymer X, we consider $K_j(X)$ as a smooth ($\mathscr{C}^p$) function defined on the vector space $\mathbb{C}^\Lambda$ with norm chosen so that the supremum of a unit norm field is $O(L^{-j})$. This factor is because part (3) of Theorem 6.1 says that the standard deviation of $\sum_{k>j} \zeta_k$ is $O(L^{-j})$, provided $m^2 \le O(L^{-2j})$. The norm on $K_j(X)$ measures directional derivatives of $K_j(X)$ as a function of fields in $\mathbb{C}^\Lambda$. For example,

$$DK_j(X;\phi;\dot{\phi}) = \frac{d}{dt}_{|_0} K_j(X;\phi + t\dot{\phi}) \tag{144}$$

is bounded in norm by the norm of $K_j(X)$ times a sup norm of $\dot{\phi}$ divided by L^{-j}. As an important example for us, let $\dot{\phi}$ be the constant test function 1, then the second directional derivative $D^2K_N(0;1,1)$ at $\phi = 0$ and $d\phi = 0$ satisfies

$$|D^2K_N(\Lambda;0;1,1)| \le \|K\|_{\mathscr{F}_N} O(L^{2N}). \tag{145}$$

7.2 The Irrelevant Parts of K_+

We will now illustrate the use of the properties postulated in Sect. 7.1.2 for the norm on $\mathscr{N}(X^\square)$. Lemma 6.6 has provided us with a formula (137) for a map $K_+ : (V_j, V_{j+1}, K_j) \mapsto K_{j+1}$ that exactly represents $Z_j \mapsto Z_{j+1}$ as in (94). The main result of this section is Proposition 7.2 which shows a very good property of K_+. It shows that K_+ is contractive in K_j provided K_j, which is a map from connected polymers to $\mathscr{N}$, restricts to be zero on the connected polymers in the class $\mathscr{S}$ of Definition 6.4.

For the rest of this section we omit j subscripts and abbreviate $j+1$ to $+$ as in the notation explained below (131). We also suppress the dependence of K_+ on V and V_+ because our estimates are pointwise and uniform for V, V_+ in domains $\mathscr{D}, \mathscr{D}_+$. Referring to the formula (137) for K_+ let $\mathscr{P}^{(2)}(Y)$ denote the set of pairs (X, X_K) such that $X \in \bar{\mathscr{P}}_j(Y)$ and $X_K \in \mathscr{P}_j(X)$. From (137) and (134),

$$K_+(Y) = \sum_{(X,X_K)\in\mathscr{P}^{(2)}(Y)} I_+^{Y\setminus X}\, \mathbb{E}_+\big[\delta^{X\setminus X_K} K(X_K)\big]. \tag{146}$$

(i) We subtract from $K_+(Y)$ the value of $K_+(Y)$ when $K = 0$ by omitting terms in the sum which do not depend on K, that is terms where $X_K = \varnothing$. (ii) We omit terms where $X = X_K \in \mathscr{S}$, where $\mathscr{S}$ is the class of small sets defined in Definition 6.4. To study the remaining part of K_+, let

$$K_+^{\mathscr{I}}(Y) = \sum_{(X,X_K)\in\mathscr{I}(Y)} I_+^{Y\setminus X}\,\mathbb{E}_+\big[\delta^{X\setminus X_K}K(X_K)\big]. \tag{147}$$

where

$$\mathscr{I}(Y) = \mathscr{P}^{(2)}(Y) \setminus \Big(\{X_K = \varnothing\} \cup \{X = X_K \in \mathscr{S}\}\Big). \tag{148}$$

Wilson called terms that contract *irrelevant* so we have used the letter $\mathscr{I}$ to label the set of terms that we can prove are collectively contracted. Wilson called terms that expand *relevant* and terms that stay the same size *marginal*. Thus we use $\mathscr{R}$ to label the complement of the set $\mathscr{I}$, but neither class is in precise correspondence with Wilson's classification because although we are capturing his intuition we are working outside his original context.

For the next section we define $K_+^{\mathscr{R}}$ be the part of K_+ that we subtracted out under item (ii). We could write it in exactly the same form as (147) with the set $\mathscr{I}(Y)$ replaced by

$$\mathscr{R}(Y) = \mathscr{P}^{(2)}(Y) \setminus \Big(\{X_K = \varnothing\} \cup \{X = X_K \notin \mathscr{S}\}\Big), \tag{149}$$

but it is easy to verify that this is the same as

$$K_+^{\mathscr{R}}(Y) = \sum_{X_K\in\mathscr{S}:\,\overline{X}_K = Y} I_+^{Y\setminus X_K}\,\mathbb{E}_+\big[K(X_K)\big]. \tag{150}$$

If the restriction of K to $\mathscr{S}$ is zero then $K_+^{\mathscr{R}} = 0$, but in general we have

$$K_+ = K_{+|_{K=0}} + K_+^{\mathscr{R}} + K_+^{\mathscr{I}}. \tag{151}$$

For the main result of this section let $\sigma = \frac{1}{2}(1+\eta)$ where $\eta > 1$ is defined by geometry in Lemma 7.3. It follows that $\eta > \sigma > 1$ and this is the only fact about σ that is important for us. Recall the definitions of the parameters $\alpha_I, \alpha_{\mathbb{E}}$ from Sect. 7.1.2 and let α be the maximum of α_I, $\alpha_{\mathbb{E}}$ and 1. We choose ϵ in (140) smaller than 1 and such that

$$\kappa < \frac{2}{3}, \quad \text{where} \quad \kappa = (3\alpha)^{L^d}\epsilon^{\sigma-1}. \tag{152}$$

We decrease, if necessary, the domain $\mathscr{D}_k$ of the coupling constants g_k, z_k, μ_k with $k = j, j+1$ so that, by the remark below (143),

$$\epsilon_\delta \leq \epsilon^\sigma \epsilon^{\sigma 2^{d+1}}. \tag{153}$$

Let $B_{\mathscr{F}}$ be the ball in $\mathscr{F}$ given by

$$\|K\|_{\mathscr{F}} \leq \left(\epsilon^{\sigma 2^{d+1}}\right)^2. \tag{154}$$

Proposition 7.2 *With the choices (148)–(154), the function $K \mapsto K_+^{\mathscr{J}}$ with domain $B_{\mathscr{F}}$ satisfies*

$$\|K_+^{\mathscr{J}}\|_{\mathscr{F}_+} \leq \kappa \|K\|_{\mathscr{F}}, \tag{155}$$

and, for K and K' in the smaller domain $\frac{1}{4}B_{\mathscr{F}}$,

$$\left\| K_+^{\mathscr{J}}(K') - K_+^{\mathscr{J}}(K) \right\|_{\mathscr{F}_+} \leq \frac{3}{2}\kappa \|K' - K\|_{\mathscr{F}}. \tag{156}$$

Thus $K_+^{\mathscr{J}}$ is a contractive map from $\frac{1}{4}B_{\mathscr{F}}$ to itself. If the restriction of K to $\mathscr{S}$ is zero then K_+ is contractive.

To prove this Proposition we need the following two geometrical estimates which show why small sets have an exceptional role. We measure the size of a polymer at scale j by counting the number of scale j blocks it contains. Similarly the size of a polymer at scale $j+1$ is measured by counting the number of scale $j+1$ blocks it contains. Does the closure map $X \mapsto \overline{X}$ make a polymer smaller? For example let X be a polymer that consists of a single block. Then $\overline{X}$ is also a polymer which consists of a single block so it is the same size as X. Lemma 7.3 says (1) that connected polymers that are not small always decrease in size under closure and (2) that no polymer increases in size under closure.

Lemma 7.3 ([15, Lemma 6.15]) *There is an $\eta = \eta(d) > 1$ such that for all $L \geq L_0(d) = 2^d + 1$ and for all connected scale j polymers that are not in $\mathscr{S}_j$,*

$$|X|_j \geq \eta |\overline{X}|_{j+1}. \tag{157}$$

In addition, (157) holds with η replaced by 1 for all $X \in \mathscr{P}_j$ (not necessarily connected, and possibly small).

Corollary 7.4 ([15, Lemma 6.16]) *Let $X \in \mathscr{P}_j$ and let n be the number of components of X. Then*

$$|X|_j \geq \frac{1}{2}(1+\eta)|\overline{X}|_{j+1} - \frac{1}{2}(1+\eta)2^{d+1}n. \tag{158}$$

Proof (of Proposition 7.2) The estimate (156) follows from (155) and [3, Lemma 1] with $\nu = \frac{1}{4}$. The final claim in Proposition 7.2 follows from (151) and the remark above this equation. Therefore we are reduced to proving (155).

By applying to (147), the product property (141), the bound (142) on I_+ and the integration property (143), for $Y \in \mathscr{C}(\Lambda)$,

$$\|K_+^{\mathscr{I}}(Y)\|_{j+1} \le \sum_{(X,X_K)\in\mathscr{I}(Y)} \alpha^{|Y|_j} \epsilon_\delta^{|X\setminus X_K|_j} \|K(X_K)\|_j. \tag{159}$$

The polymer X_K may not be connected. Suppose it has connected components $X_1, \dots, X_n$. We use property (3) of Definition 6.5 together with the product property (141) and the definition (139) of $\|K\|_{\mathscr{F}} = \|K\|_{\mathscr{F}_j}$ to obtain

$$\begin{aligned} \|K(X_K)\|_j &= \prod_{i=1,\dots,n} \|K(X_i)\|_j \\ &= \epsilon^{|X_K|_j} \prod_{i=1,\dots,n} \epsilon^{-|X_i|_j} \|K(X_i)\|_j \le \epsilon^{|X_K|_j} \|K\|_{\mathscr{F}}^n, \end{aligned} \tag{160}$$

where $n = n(X_K)$ is the number of components in X_K. We substitute (160) into (159) and estimate the sum by noting that it extends over less than $3^{|Y|_j}$ terms because this is the number of ways to partition Y into three disjoint subsets $X_K, X \setminus X_K, Y \setminus X$ that are each polymers at scale j. Also we can replace Y by $\overline{X}$ because one of the defining conditions for $\mathscr{I}(Y)$ is that they are equal. Therefore we have

$$\|K_+^{\mathscr{I}}(Y)\|_{j+1} \le \sup_{(X,X_K)\in\mathscr{I}(Y)} (3\alpha)^{|\overline{X}|_j} \epsilon_\delta^{|X\setminus X_K|_j} \epsilon^{|X_K|_j} \|K\|_{\mathscr{F}}^{n(X_K)}. \tag{161}$$

We multiply both sides by $\epsilon^{-|Y|_{j+1}}$ which equals $\epsilon^{-|\overline{X}|_{j+1}}$ and take the supremum over connected $j+1$ scale polymers in order to form the $\mathscr{F}_{j+1} = \mathscr{F}_+$ norm on the left hand side and we obtain

$$\begin{aligned} \|K_+^{\mathscr{I}}\|_{\mathscr{F}_+} &\le \sup_{(X,X_K)\in\mathscr{I}} F(X,X_K), \quad \text{where} \\ F(X,X_K) &= \epsilon^{-|\overline{X}|_{j+1}} (3\alpha)^{|\overline{X}|_j} \epsilon_\delta^{|X\setminus X_K|_j} \epsilon^{|X_K|_j} \|K\|_{\mathscr{F}}^{n(X_K)}, \end{aligned} \tag{162}$$

and $\mathscr{I} = \cup_{Y\in\mathscr{C}_+(\Lambda)} \mathscr{I}(Y)$. Recall that $Y \in \mathscr{C}_+(\Lambda)$ implies that Y is not empty. Now we cover $\mathscr{I}$ by three subsets and it suffices to prove that for (X, X_K) in each covering set we have $F(X, X_K) \le \kappa \|K\|_{\mathscr{F}}$. We give the proof in most detail for subset 1 because the proofs are similar for the other sets.

Subset 1 For $(X, X_K) \in \mathscr{I} \cap \{n(X_K) = 1, |X_K| > 2^d\}$, by using Lemma 7.3 twice, $\eta > \sigma$ and the hypothesis (153) on ϵ_δ and $\epsilon_\delta \le 1$,

$$\begin{aligned}\epsilon_\delta^{|X\setminus X_K|_j}\epsilon^{|X_K|_j} &\le \epsilon_\delta^{|\overline{X\setminus X_K}|_{j+1}}\epsilon^{\eta|\overline{X_K}|_{j+1}} \\ &\le \epsilon^{\sigma|\overline{X\setminus X_K}|_{j+1}}\epsilon^{\sigma|\overline{X_K}|_{j+1}} \le \epsilon^{\sigma|\overline{X}|_{j+1}}. \end{aligned} \tag{163}$$

The final inequality holds because $\overline{A} \cup \overline{B} \supset \overline{A \cup B}$ since the closure is the smallest next scale polymer cover. Putting this estimate into (162) and using the definition (152) of κ we have

$$\begin{aligned}F(X, X_K) &\le (3\alpha)^{|\overline{X}|_j}\epsilon^{(\sigma-1)|\overline{X}|_{j+1}}\|K\|_{\mathscr{F}} \\ &= \left((3\alpha)^{L^d}\epsilon^{\sigma-1}\right)^{|\overline{X}|_{j+1}}\|K\|_{\mathscr{F}} \le \kappa\|K\|_{\mathscr{F}}. \end{aligned} \tag{164}$$

Subset 2 For $(X, X_K) \in \mathscr{I} \cap \{n(X_K) = 1, |X \setminus X_K|_j \ge 1\}$, by Lemma 7.3 and Corollary 7.4 and the hypothesis (153) on ϵ_δ,

$$\epsilon_\delta^{|X\setminus X_K|_j}\epsilon^{|X_K|_j} \le \epsilon_\delta^{|\overline{X\setminus X_K}|_{j+1}}\epsilon^{\sigma|\overline{X_K}|_{j+1}}\epsilon^{-\sigma 2^{d+1}} \le \epsilon^{\sigma|\overline{X}|_{j+1}}. \tag{165}$$

Putting this estimate into (162) we have

$$\begin{aligned}F(X, X_K) &\le (3\alpha)^{|\overline{X}|_j}\epsilon^{(\sigma-1)|\overline{X}|_{j+1}}\|K\|_{\mathscr{F}} \\ &\le (3\alpha)^{L^d}\epsilon^{\sigma-1}\|K\|_{\mathscr{F}} \le \kappa\|K\|_{\mathscr{F}}. \end{aligned} \tag{166}$$

Subset 3 For $(X, X_K) \in \mathscr{I} \cap \{n(X_K) \ge 2\}$, by Lemma 7.3 and Corollary 7.4, the hypothesis (153) on ϵ_δ and the domain (154) for K,

$$\begin{aligned}\epsilon_\delta^{|X\setminus X_K|_j}\epsilon^{|X_K|_j}\|K\|_{\mathscr{F}}^{n(X_K)} &\le \epsilon_\delta^{|\overline{X\setminus X_K}|_{j+1}}\epsilon^{\sigma|\overline{X_K}|_{j+1}}\left(\epsilon^{-\sigma 2^{d+1}}\|K\|_{\mathscr{F}}\right)^{n(X_K)} \\ &\le \epsilon^{\sigma|\overline{X}|_{j+1}}\left(\epsilon^{-\sigma 2^{d+1}}\|K\|_{\mathscr{F}}\right)^2 \\ &\le \epsilon^{\sigma|\overline{X}|_{j+1}}\|K\|_{\mathscr{F}}. \end{aligned} \tag{167}$$

Putting this estimate into (162) we have

$$\begin{aligned}F(X, X_K) &\le (3\alpha)^{|\overline{X}|_j}\epsilon^{(\sigma-1)|\overline{X}|_{j+1}}\|K\|_{\mathscr{F}} \\ &\le \left((3\alpha)^{L^d}\epsilon^{\sigma-1}\right)^{|\overline{X}|_{j+1}}\|K\|_{\mathscr{F}} \le \kappa\|K\|_{\mathscr{F}}. \end{aligned} \tag{168}$$

7.3 The Complete Recursion

We have seen in the previous section that the map $(V_j, V_{j+1}, K_j) \mapsto K_+$ provided by Proposition 7.2 is contractive provided K_j vanishes when evaluated on polymers in the class $\mathscr{S}$ of small sets. In this section we will discuss a better choice for K_+ which is contractive. Part of the improvement comes from specifying V_{j+1} carefully, recalling that Proposition 7.2 put essentially no constraint on it. However we need also another idea which is a change of variable formula for K. The conclusions which play a role in the sequel are the recursion (170) and its conjugation (175).

Recall from (142) and (143) that $\mathscr{D}_j$ is a domain in $\mathbb{R}^3$ for the coupling constants g_j, z_j, μ_j. Let $B_{\mathscr{F}_j}$ denote a ball in a Banach space $\mathscr{F}_j$. In [17] we define, for all scales $= 0, 1, \dots, N-1$, domains $\mathscr{D}_j \times B_{\mathscr{F}_j}$ and functions

$$R_+ : \mathscr{D}_j \times B_{\mathscr{F}_j} \to \mathbb{R}^3, \quad K_+ : \mathscr{D}_j \times B_{\mathscr{F}_j} \to B_{\mathscr{F}_{j+1}}. \tag{169}$$

These functions, together with the second order perturbative map $\varphi^{(0)}_{\mathrm{pt},j}$ appearing in (105), build a recursion,

$$\begin{aligned}(g_{j+1}, z_{j+1}, \mu_{j+1}) &= \varphi^{(0)}_{\mathrm{pt},j}(g_j, z_j, \mu_j) + R_+(g_j, z_j, \mu_j; K), \\ K_{j+1} &= K_+(g_j, z_j, \mu_j; K_j),\end{aligned} \tag{170}$$

This recursion, with initial condition $(g_0, z_0, \mu_0; K_0)$, where $K_0 = \mathbb{1}_\varnothing$, has the following properties:

1. K_+ is contractive in K_j.
2. The recursion generates a sequence $(g_j, z_j, \mu_j; K_j)_{j=0,\dots,j_{\mathrm{exit},N}}$ which terminates at scale $j_{\mathrm{exit},N}$, which is the first scale $j \wedge N$ such that (g_j, z_j, μ_j) is not in $\mathscr{D}_j$.
3. For Z_j given by (94) and $I_j = I_j(V_j)$ given by (106) and (95), we have

$$Z_j = I_j \circ K_j(\Lambda), \quad j \le j_{\mathrm{exit},N}. \tag{171}$$

4. Volume Λ compatibility.

The existence of R and K_+ with the first and third properties are parts of Theorem 2.2 of [17]. The second property follows immediately from the domains of the functions R_+ and K_+. The fourth property is a statement analogous to the statements below (112). The essential idea is that the recursions for two tori of different sizes N, N' generate the "same" $g_j, z_j, \mu_j; K_j$ for j up to the scale before the smaller of N, N'. However, it takes time to formulate the meaning of "same" so we refer to [17, Sect. 1.8.3]. This compatibility enables the definition of the infinite volume limit of (170) which is a recursion that generates a sequence $(g_j, z_j, \mu_j; K_j)_{j=0,\dots,j_{\mathrm{exit},\infty}}$ where K_j is defined on polymers in $\mathbb{Z}^d$ and

$$j_{\mathrm{exit},\infty} = \limsup_N j_{\mathrm{exit},N}. \tag{172}$$

To put this into the context of Proposition 7.2 refer to Eq. (151). The term $K_+^{\mathscr{J}}$ is contractive. Indeed, the preamble for this Proposition shows that the contractivity constant κ can be made arbitrarily small by choosing ϵ in the norm small and the ball $B_{\mathscr{F}_j}$ small. But the term $K_+^{\mathscr{R}}$ in (151) cannot be contractive. To understand why, consider (150) when Y is a single block on the scale $j+1$. Then the sum over polymers X_K has a range that includes the L^d possible choices of a single block $X_K = B$ in $\mathscr{B}_j(Y)$. This factor of L^d prevents $K_+^{\mathscr{R}}$ from being contractive.

The function R_+ in (170) is part of the solution to this problem. By choosing $g_{j+1}, z_{j+1}, \mu_{j+1}$ to be different from $\varphi^{(0)}_{\mathrm{pt},j}(g_j, z_j, \mu_j)$ we generate terms in K_{j+1} that I will call *counterterms*. These counterterms cancel most of $K_j^{\mathscr{R}}$ in (151) which enables K_+ to be contractive. However, to achieve this cancellation, one must have some way of changing the allocation of the counterterms to the different small sets $X_K \in \mathscr{S}$, because Proposition 7.2 does not match the counterterms correctly with the small set parts of K_j. We solve this allocation problem by taking advantage of the fact that there are re-allocation changes of variable $K_j \mapsto K_j'$ such that

$$I_j \circ K_j'(\Lambda) = I_j \circ K_j(\Lambda). \tag{173}$$

These changes of variable are supplied by Brydges and Slade [17, Proposition 4.1]. Roughly speaking these changes of variable cancel some part $J(X)$ of $K_j(X)$ for small sets $X \in \mathscr{S}_j$ which are not single blocks and compensate by adding $\sum_{X \supset B} J(X)$ to $K_j(B)$ for each block B. Using this re-allocation we can prove that when R_+ is chosen correctly the small set part of the re-allocated K_j' is almost zero. Therefore when K_+ given by (146) is evaluated on the re-allocated K' the term $K_+^{\mathscr{J}}$ is almost zero and Proposition 7.2 proves that the composition is contractive. Thus the composition of (146) on the right with a re-allocation gives a contractive formula. In passing let me remark that there is another re-allocation problem that is solved by a further composition with a re-allocation on the left. This problem is that one cannot prove that the first term in (151) is $O(V_j^3)$ as expected from our choice of the second order formula for I_j unless the contributions from perturbation theory are re-allocated. This is happening in Lemmas 4.2 and 5.8 of [17].

Recall from (108) that T_j is a map that conjugates $\varphi^{(0)}_{\mathrm{pt},j}$ to the triangular map $\bar{\varphi}_j$. We rewrite the recursion (170) in terms of new variables defined by

$$(\check{g}_j, \check{z}_j, \check{\mu}_j; \check{K}_j) = \big(T_j(g_j, z_j, \mu_j); K_j\big). \tag{174}$$

By (109) the new recursion is

$$\begin{aligned}(\check{g}_{j+1}, \check{\mu}_{j+1}, \check{z}_{j+1}) &= \bar{\varphi}_j(\check{g}_j, \check{\mu}_j, \check{z}_j) + \check{R}_+(\check{g}_j, \check{\mu}_j, \check{z}_j; \check{K}),\\ \check{K}_{j+1} &= \check{K}_+(\check{g}_j, \check{\mu}_j, \check{z}_j; \check{K}_j),\end{aligned} \tag{175}$$

where $\check{R}_{j+1} = T_{j+1} \circ R_{j+1} \circ T_j^{-1}$ plus the $O(V^3)$ error in (109) and on the right hand side of (175) $\check{K}_{j+1} = K_{j+1} \circ T_j^{-1}$ where T_j^{-1} acts only on the $\mathbb{R}^3$ part.

As explained in Sect. 6.6 of Bauerschmidt et al. [10] these new functions have almost the same domains and satisfy essentially the same bounds as the old ones because the conjugations T_j are close (108) to the identity. In particular $\check{K}_{j+1}$ is still contractive.

8 Outline of Proof of Theorem 5.2

This is a survey of the complete argument which is in [10, Sect. 8.3]. I omit the proofs of (75). One point that I find remarkable is that there are exact formulas for χ and its derivative in terms of the recursion of the coupling constant part of (170). The error coordinate does not appear in these formulas; the only role it has is to slightly change, via R_+, the recursion of second order perturbation theory. I should mention that this feature has been strongly emphasised in physics, but it is nice to be able to verify it here.

8.1 Construction of z_0^{c}, ν_0^{c}

Recall from (172) the definition of $j_{\mathrm{exit},\infty}$. The first step is to prove that there exist functions z_0^{c} and ν_0^{c} of (m^2, g_0) such that $j_{\mathrm{exit},\infty} = \infty$ for the (infinite volume limit of the) recursion (170) with the initial condition

$$(g_0, z_0, \mu_0; K_0) \text{ such that } z_0 = z_0^{\mathrm{c}}(m^2, g_0), \quad \mu_0 = \nu_0^{\mathrm{c}}(m^2, g_0), \quad K_0 = \mathbb{1}_{\varnothing}. \tag{176}$$

Proposition 7.1 of Bauerschmidt et al. [10] shows that z_0^{c} and ν_0^{c} exist. This proposition is proved by showing that the conjugated recursion (175) generates an infinite sequence $(\check{g}_j, \check{z}_j, \check{\mu}_j; \check{K}_j)_{j\in\mathbb{N}_0}$ with the same initial condition (176). The initial condition is the same because, by (108), T_0 is the identity. The main ideas in the proof of Proposition 7.1 of Bauerschmidt et al. [10] are that (1) the existence of infinite sequences for the $\bar{\varphi}$ recursion (110)–(112) is given by Proposition 6.3. (2) Norm estimates on the functions $\check{R}_+$ and $\check{K}_+$ in the recursion (175) show that they are always small compared with $\bar{\varphi}$ on the infinite sequence supplied by Proposition 6.3. By the main result of Bauerschmidt et al. [9] there exists a unique infinite sequence generated by the recursion (175) that stays close to the infinite sequence supplied by Proposition 6.3.

8.2 Coupling Constants at Large Scales

The following results are needed for the next steps in the proof. Recall from Sect. 8.1 that the infinite sequence generated by the recursion (175) stays close to the infinite sequence supplied by Proposition 6.3. In particular, by Proposition 7.1 of Bauerschmidt et al. [10],

$$\|K_j\|_{\mathscr{W}_j(s_j,\Lambda_N)} \le O(\chi_j \bar{g}_j^3), \quad \check{z}_j = O(\chi_j \check{g}_j), \quad \check{\mu}_j = O(\chi_j \check{g}_j), \tag{177}$$

where $\mathscr{W}_j(s_j, \Lambda_N)$ is a weighted maximum of two different choices of $\mathscr{F}_j$ norms. These bounds are the same as the coupling constant bounds (115) and K_j is third order which manifests the idea that the recursion is staying close to the second order perturbative recursion.

The coupling constant $\check{g}_j$ has the same asymptotic behaviour as the solution to (110), but does not tend to zero unless $m^2 = 0$. By Lemma 8.5 of Bauerschmidt et al. [10] it tends to a limit $\check{g}_\infty$ and, for $\hat{g}_0$ small and positive, as $m^2 \downarrow 0$ and $g_0 \to \hat{g}_0$,

$$\check{g}_\infty \sim \frac{1}{\mathsf{B}_{m^2}}. \tag{178}$$

By the formulas for the conjugation T_j which are given in [11, (6.93)], there exist constants $a_j = O(1)$ such that $\check{\mu}_j = \mu_j + a_j \mu_j^2$. By (177) and the a-priori limitation of coupling constants to be in a small domain $\mathscr{D}_j$, this implies that

$$\mu_j = O(\chi_j \check{g}_j), \quad \mu_j' = \check{\mu}_j'(1 + O(\mu_j)) \sim \check{\mu}_j', \quad \text{as} \quad j \to \infty, \tag{179}$$

where the prime denotes the derivative of with respect to v_0. Note that $\mu_0 = \check{\mu}_0 = v_0$ by (108) and (96). Therefore

$$\check{\mu}_0' = \mu_0' = 1. \tag{180}$$

We will be applying these bounds with $j = N$ which is the scale where the finite volume recursion parts company with the infinite volume recursion; at the scale where the torus becomes a single block we are integrating out a field ζ_N with the Λ dependent covariance $C_{N,N}$ of Theorem 6.1. But, provided N is large such that $m^2 L^{2N} \ge 1$, the recursion (170) obeys the same bounds as the infinite volume recursion and we can set $j = N$ in (177). Since we studying the limits in the order $N \to \infty$ followed by $m^2 \downarrow 0$ we can assume that $m^2 L^{2N} \ge 1$.

8.3 *Proof of* (73)

By (131) and (106),

$$\begin{aligned}\hat{\chi}_\Lambda(m^2, g_0, \nu_0, z_0) =& \frac{1}{m^2} + \frac{1}{m^4\,|\Lambda|} D^2 I_N(\Lambda; 0; 1, 1) + \frac{1}{m^4\,|\Lambda|} D^2 K_N(\Lambda; 0; 1, 1) \\ =& \frac{1}{m^2} + \frac{1}{m^4|\Lambda|} D^2 e^{-V_N}(\Lambda; 0; 1, 1) + \frac{1}{m^4|\Lambda|} D^2 W_N(\Lambda; 0; 1, 1) \\ &+ \frac{1}{m^4\,|\Lambda|} D^2 K_N(\Lambda; 0; 1, 1). \end{aligned} \tag{181}$$

where cross-terms in $D^2 I_N$ are zero when $\phi = 0$ and $d\phi = 0$ because W_N defined in [11, Sect. 3.5] has no monomials of odd degree, in particular of degree one. The first term on the right-hand side of (181) can be evaluated by direct calculation, using (95), (46) and (47), to give

$$\begin{aligned} D^2 e^{-V_N}(\Lambda; 0; 1, 1) &= D^2(-V_N)(\Lambda; 0; 1, 1) \\ &= -\sum_{x,y} \mu_N L^{-2N} \delta_{xy} 1_x 1_y - \sum_{x,y} z_N(-\Delta_{xy}) 1_x 1_y \\ &= -\mu_N L^{-2N} |\Lambda|, \end{aligned} \tag{182}$$

since the quartic term τ^2 does not contribute, and $\Delta 1 = 0$. Therefore

$$\begin{aligned} \hat{\chi}_N = \frac{1}{m^2} - \frac{\mu_N L^{-2N}}{m^4} + \frac{1}{m^4} \frac{1}{|\Lambda|} D^2 W_N^0(0, 0; 1, 1) \\ + \frac{1}{m^4} \frac{1}{|\Lambda|} D^2 K_N^0(0, 0; 1, 1). \end{aligned} \tag{183}$$

For m^2 fixed the final term tends to zero as $N \to \infty$ like $O(\chi_N g_N^3) L^{-2N}$. This follows from (145) and

1. $\frac{1}{|\Lambda|} = L^{-4N}$.
2. $\|K_N\|_{\mathscr{F}_N} \le O(\chi_N g_N^3)$ by (177).

By (177) μ_N is bounded and therefore the second term tends to zero like L^{-2N}. The third term is estimated from the explicit formula for W_N and tends to zero as $O(\chi_N \bar{g}_N^2 L^{-2N})$ by Bauerschmidt et al. [10, (8.56)]. Thus we have proved Part (1) of Theorem 5.2.

8.4 *Proof of* (74)

By the remark above Theorem 2.2 we can interchange a derivative with respect to ν_0 with the infinite volume limit. Thus we take the derivative of both sides of (183) with respect to ν_0 and obtain

$$\frac{\partial \hat{\chi}}{\partial \nu_0} = -\frac{1}{m^4} \lim_{N\to\infty} L^{-2N} \frac{\partial \mu_N}{\partial \nu_0} = -\frac{1}{m^4} \lim_{N\to\infty} L^{-2N} \check{\mu}'_N, \tag{184}$$

where we used (179) to obtain the second equality. As in Sect. 8.3, there are no contributions to this derivative from the third and fourth terms in (183) because they decay to zero as $N \to \infty$. To calculate the derivative $\check{\mu}'_N$ we differentiate the recursion (175) with respect to ν_0 and obtain a recursion for the derivatives $(\check{g}'_j, \check{z}'_j, \check{\mu}'_j; \check{K}'_j)$. Since there are many terms and the details are given in Sect. 8.3 of Bauerschmidt et al. [10] we write only the terms that will turn out to be dominant. From (175) and (112) the $\check{\mu}$ equation is

$$\check{\mu}_{j+1} = L^2 \check{\mu}_j (1 - \gamma \beta_j \check{g}_j) + r_j, \tag{185}$$

where r_j is the sum of all the other terms. It changes from line to line in the next equations. Therefore

$$\begin{aligned} \check{\mu}'_{j+1} &= L^2 \check{\mu}'_j (1 - \gamma \beta_j \check{g}_j) + r'_j \\ &= L^2 \check{\mu}'_j (1 - \beta_j \check{g}_j)^\gamma + r'_j. \end{aligned} \tag{186}$$

From (175) and (110) and an estimate on the $\check{g}$ component of $\check{R}_+$,

$$\check{g}_{j+1} = \check{g}_j - \beta_j \check{g}_j^2 + O(\chi_j \check{g}_j^3) = \check{g}_j (1 - \beta_j \check{g}_j) + O(\chi_j \check{g}_j^3). \tag{187}$$

Using this to eliminate $(1 - \beta_j \check{g}_j)$ in (186) and dropping r'_j we obtain

$$\check{\mu}'_{j+1} = L^2 \check{\mu}'_j \left(\frac{\check{g}_{j+1}}{\check{g}_j} \right)^\gamma \left(1 + O(\chi_j \check{g}_j^2) \right)^\gamma. \tag{188}$$

By iterating this equality and recalling (180) we have

$$\lim_{N\to\infty} L^{-2N} \check{\mu}'_N = c(m^2, g_0) \check{\mu}'_0 \left(\frac{\check{g}_\infty}{\check{g}_0} \right)^\gamma = c(m^2, g_0) \left(\frac{\check{g}_\infty}{\check{g}_0} \right)^\gamma, \tag{189}$$

where $\check{g}_\infty = \lim \check{g}_N$ and $c(m^2, g_0)$ arises from the r'_j terms and the factors $1 + O(\chi_j \check{g}_j^2)$. The coefficient $c(m^2, g_0)$ has a limit as $m^2 \downarrow 0$. We insert (189) into (184)

and obtain

$$\frac{\partial \hat{\chi}}{\partial \nu_0} = -\frac{1}{m^4} c(m^2, g_0) \left(\frac{\check{g}_\infty}{\check{g}_0} \right)^\gamma . \tag{190}$$

By (178), we obtain the desired (74).

Acknowledgements This work was supported in part by NSERC of Canada. Marek Biskup, Roman Kotecký and Martin Slowik helped me write these notes, asking many questions that led to corrections and improvements, but I enjoy all the credit for errors. My colleagues Gordon Slade and Roland Bauerschmidt may not agree with every equation in these notes but none of them would be here without them. I thank my wife Betty Lu Brydges, for her amused tolerance and patience during all the long time this work has been in progress.

References

1. A. Abdesselam, A complete renormalization group trajectory between two fixed points. Commun. Math. Phys. **276**(3), 727–772 (2007)
2. A. Abdesselam, J. Magnen, V. Rivasseau, Bosonic monocluster expansion. Commun. Math. Phys. **229**(2), 183–207 (2002)
3. A. Abdesselam, A. Chandra, G. Guadagni, Rigorous quantum field theory functional integrals over the *p*-adics I: anomalous dimensions (2013 preprint)
4. S. Adams, R. Kotecký, S. Müller, Finite range decomposition for families of gradient Gaussian measures. J. Funct. Anal. **264**(1), 169–206 (2013)
5. V.I. Arnol'd, *Mathematical Methods of Classical Mechanics*. Graduate Texts in Mathematics, vol. 60, 2nd edn. (Springer, New York, 1989). Translated from the Russian by K. Vogtmann and A. Weinstein
6. T. Bałaban, A low temperature expansion for classical *N*-vector models. I. A renormalization group flow. Commun. Math. Phys. **167**(1), 103–154 (1995)
7. R. Bauerschmidt, A simple method for finite range decomposition of quadratic forms and Gaussian fields. Probab. Theory Related Fields, **157** (3–4), 817–845, (2011)
8. R. Bauerschmidt, H. Duminil-Copin, J. Goodman, G. Slade, Lectures on self-avoiding walks, in *Probability and Statistical Physics in Two and More Dimensions*, ed. by D. Ellwood, C. Newman, V. Sidoravicius, W. Werner. Clay Mathematics Proceedings, vol. 15 (American Mathematical Society, Providence, RI, 2012), pp. 395–467
9. R. Bauerschmidt, D.C. Brydges, G. Slade, Structural stability of a dynamical system near a non-hyperbolic fixed point. Ann. Henri Poincaré **16**(4), 1033–1065, (2013)
10. R. Bauerschmidt, D.C. Brydges, G. Slade, Logarithmic correction for the susceptibility of the 4-dimensional weakly self-avoiding walk: a renormalisation group analysis. Comm. Math. Phys. **337**(2), 817–877, (2014)
11. R. Bauerschmidt, D.C. Brydges, G. Slade, A Renormali- sation group method. III. Perturbative analysis. J. Stat. Phys. **159**(3), 492–529, (2014)
12. F.A. Berezin, *The Method of Second Quantization*. Translated from the Russian by Nobumichi Mugibayashi and Alan Jeffrey. Pure and Applied Physics, vol. 24 (Academic, New York, 1966)
13. E. Brezin, J.C. Le Guillou, J. Zinn-Justin, Approach to scaling in renormalized perturbation theory. Phys. Rev. D **8**, 2418–2430 (1973)
14. D.C. Brydges, Functional integrals and their applications. Lecture Notes Taken by Roberto Fernandez. Course in the "Troisième Cycle de la Physique en Suisse Romande", May 1992
15. D.C. Brydges, Lectures on the renormalisation group, in *Statistical Mechanics*, ed. by S. Sheffield, T. Spencer. IAS/Park City Mathematics Series, vol. 16 (American Mathematical Society, Providence, RI, 2009), pp. 7–93

16. D.C. Brydges, G. Slade, A renormalisation group method. I. Gaussian integration and normed algebras. J. Stat. Phys., **159**(3), 421–460 (2014)
17. D.C. Brydges, G. Slade, A renormalisation group method. V. A single renormalisation group step. J. Stat. Phys., **159**(3), 589–667, (2014)
18. D. Brydges, A. Talarczyk, Finite range decompositions of positive-definite functions. J. Funct. Anal. **236**, 682–711 (2006)
19. D.C. Brydges, J. Fröhlich, T. Spencer, The random walk representation of classical spin systems and correlation inequalities. Commun. Math. Phys. **83**, 123–150 (1982)
20. D. Brydges, G. Guadagni, P.K. Mitter, Finite range decomposition of Gaussian processes. J. Stat. Phys. **115**(1–2), 415–449 (2004)
21. D.C. Brydges, J.Z. Imbrie, G. Slade, Functional integral representations for self-avoiding walk. Probab. Surv. **6**, 34–61 (2009)
22. D.C. Brydges, A. Dahlqvist, G. Slade, The strong interaction limit of continuous-time weakly self-avoiding walk, in *Probability in Complex Physical Systems*, ed. by J.-D. Deuschel, B. Gentz, W. König, M. von Renesse, M. Scheutzow, U. Schmock. Springer Proceedings in Mathematics, vol. 11 (Springer, Berlin/Heidelberg, 2012), pp. 275–287
23. C.G. Callan, Broken scale invariance in scalar field theory. Phys. Rev. D **2**, 1541–1547 (1970)
24. P.G. de Gennes, Exponents for the excluded volume problem as derived by the Wilson method. Phys. Lett. **A38**, 339–340 (1972)
25. J.J. Duistermaat, G.J. Heckman, On the variation in the cohomology of the symplectic form of the reduced phase space. Invent. Math. **69**(2), 259–268 (1982)
26. B. Duplantier, Polymer chains in four dimensions. Nucl. Phys. B **275**(2), 319–355 (1986)
27. E.B. Dynkin, Markov processes as a tool in field theory. J. Funct. Anal. **50**, 167–187 (1983)
28. K. Gawedzki, A. Kupiainen, Block spin renormalization group for dipole gas and $(\nabla\phi)^4$. Ann. Phys. **147**, 198 (1983)
29. K. Gawedzki, A. Kupiainen, Massless lattice φ_4^4 theory: rigorous control of a renormalizable asymptotically free model. Commun. Math. Phys. **99**(2), 197–252 (1985)
30. M. Gell-Mann, F.E. Low, Quantum electrodynamics at small distances. Phys. Rev. (2) **95**, 1300–1312 (1954)
31. J. Glimm, A. Jaffe, *Quantum Physics: A Functional Integral Point of View*, 2nd edn. (Springer, Berlin, 1987)
32. F. Guerra, L. Rosen, B. Simon, The $\mathbf{P}(\phi)_2$ Euclidean quantum field theory as classical statistical mechanics. I, II. Ann. Math. (2) **101**, 111–189 (1975); Ann. Math. (2) **101**, 191–259 (1975)
33. T. Hara, H. Tasaki, A rigorous control of logarithmic corrections in four-dimensional ϕ^4 spin systems. II. Critical behavior of susceptibility and correlation length. J. Stat. Phys. **47**(1–2), 99–121 (1987)
34. D. Iagolnitzer, J. Magnen, Polymers in a weak random potential in dimension four: rigorous renormalization group analysis. Commun. Math. Phys. **162**, 85–121 (1994)
35. G.F. Lawler, J. Perlman, Loop measures and the gaussian free field. http://www.arxiv.org/abs/1403.4285 (2014)
36. Y. Le Jan, Temps local et superchamp, in *Seminaire de Probabilités XXI*, ed. by J. Azéma, M. Yor, P.A. Meyer. Lecture Notes in Mathematics, vol. 1247 (Springer, New York/Berlin, 1987)
37. J.M. Luttinger, The asymptotic evaluation of a class of path integrals. II. J. Math. Phys. **24**(8), 2070–2073 (1983)
38. N. Madras, G. Slade, *The Self-avoiding Walk* (Birkhäuser, Boston, 1993)
39. A.J. McKane, Reformulation of $n \to 0$ models using anticommuting scalar fields. Phys. Lett. A **76**, 22 (1980)
40. P.K. Mitter, Callan-Symanzik equations and ϵ expansions. Phys. Rev. D **7**, 2927–2942 (1973)
41. P.K. Mitter, B. Scoppola, The global renormalization group trajectory in a critical supersymmetric field theory on the lattice $\mathbb{Z}^3$. J. Stat. Phys. **133**(5), 921–1011 (2008)
42. E. Nelson, Construction of quantum fields from Markoff fields. J. Funct. Anal. **12**, 97–112 (1973)
43. E. Nelson, The free Markoff field. J. Funct. Anal. **12**, 211–227 (1973)

44. G. Parisi, N. Sourlas, Self-avoiding walk and supersymmetry. J. Phys. Lett. **41**, L403–L406 (1980)
45. E.C.G. Stueckelberg, A. Petermann, La normalisation des constantes dans la theorie des quanta. Helv. Phys. Acta **26**, 499–520 (1953)
46. K. Symanzik, Euclidean quantum field theory. I. Equations for a scalar model. J. Math. Phys. **7**, 510–525 (1966)
47. K. Symanzik, Euclidean quantum theory, in *Local Quantum Theory*, ed. by R. Jost (Academic, New York/London, 1969)
48. K. Symanzik, Small distance behaviour in field theory and power counting. Commun. Math. Phys. **18**(3), 227–246 (1970)
49. K. Symanzik, Small-distance-behaviour analysis and Wilson expansions. Commun. Math. Phys. **23**(1), 49–86 (1971)
50. A.-S. Sznitman, *Topics in Occupation Times and Gaussian Free Fields* (European Mathematical Society (EMS), Zürich, 2012)
51. K.G. Wilson, Feynman-graph expansion for critical exponents. Phys. Rev. Lett. **28**, 548–551 (1972)
52. K.G. Wilson, The renormalization group and critical phenomena. Rev. Mod. Phys. **55**(3), 583–600 (1983)
53. K.G. Wilson, M.E. Fisher, Critical exponents in 3.99 dimensions. Phys. Rev. Lett. **28**, 240–243 (1972)
54. K.G. Wilson, J. Kogut, The renormalization group and the ϵ expansion. Phys. Rep. (Sect. C Phys. Lett.) **12**, 75–200 (1974)
55. E. Witten, Supersymmetry and Morse theory. J. Differ. Geom. **17**(4), 661–692 (1982/1983)
56. E. Witten, Two-dimensional gauge theories revisited. J. Geom. Phys. **9**(4), 303–368 (1992)
57. J. Zinn-Justin, Critical phenomena: field theoretical approach. Scholarpedia **5**(5), 8346 (2010)

Phase Transitions in Discrete Structures

Amin Coja-Oghlan

1 Introduction

Random discrete structures such as random graphs, formulas or matrices play an important role in several disciplines, including combinatorics, coding theory, computational complexity theory, and the theory of algorithms [3, 5, 18]. Over the past 20 years, a systematic but non-rigorous statistical mechanics approach to these problems has been developed [21] that provides plausible "predictions" with respect to many important open problems. From a mathematical perspective, the obvious problem is to provide a rigorous foundation for the physics approach, the so-called *cavity method*. In this lecture, I am going to give a very brief introduction into this line of work.

In the first part of the lecture, we will see various examples of (random) discrete structures and learn about the physics approach. This part follows the monograph [21], where the various references to the original research articles can be found. The second part of the lecture deals with "classical" rigorous material from the theory of random discrete structures. Finally, in the third part we venture into the more recent, physics-inspired rigorous work. For the sake of concreteness, the second and the third part deal with the random hypergraph 2-coloring problem, although the ideas presented there have are fairly general and carry over to a wide variety of other examples as well.

A. Coja-Oghlan (✉)
Goethe University, Mathematics Institute, Postfach 111932, Fach 187, 60054 Frankfurt, Germany
e-mail: acoghlan@math.uni-frankfurt.de

M. Biskup et al. (eds.), *Random Walks, Random Fields, and Disordered Systems*,
Lecture Notes in Mathematics 2144, DOI 10.1007/978-3-319-19339-7_3

2 The Cavity Method

In this section we learn about the (non-rigorous) physics approach to a variety of problems concerning (random) discrete structures. In the physics literature, a generic method has been developed for calculating the "partition function" (a key quantity) and for locating phase transitions. This is the so-called *cavity method*. We begin by introducing a few basic concepts and examples. Then, we proceed with a discussion of the two variants of the cavity method that are most important for our purposes, the *replica symmetric ansatz* and the *1-step replica symmetry breaking ansatz*. This section closely follows the monograph [21].

2.1 Basic Concepts

Generally, the "systems" that we are going to be interested in can be captured as follows. There is a (small) finite set $\mathscr{X}$ of possible "spins" and a (large) number N of "variables" or "sites". A *configuration* assigns each variable one of the spins from $\mathscr{X}$. Hence, a configuration is just an element of $\mathscr{X}^N$. Furthermore, there are a (large) number of functions $\psi_a : \mathscr{X}^N \to \mathbf{R}_{\geq 0}$, $a = 1, \ldots, M$. These functions give rise to a probability distribution on $\mathscr{X}^N$, called the *Boltzmann distribution*, defined by

$$\mu(x) = \frac{1}{Z}\prod_{i=1}^{M}\psi_a(x), \quad \text{provided that} \quad Z = \sum_{x\in\mathscr{X}^N}\prod_{i=1}^{M}\psi_a(x) > 0.$$

The number Z is called the *partition function*.

Typically, each of the functions ψ_a only really depends upon a (small) subset $\partial a \subset \{1, \ldots, N\}$ of the variables; that is, for any two configurations $x, x' \in \mathscr{X}^n$ such that $x_i = x'_i$ for all $i \in \partial a$ we have $\psi_a(x) = \psi_a(x')$. This leads to the notion of the *factor graph*. This is a bipartite graph with two kinds of vertices: there are N *variable nodes* $x_1, \ldots, x_N$ and M *function nodes* $\psi_1, \ldots, \psi_M$, and each function node ψ_a is connected with the variable nodes x_i, $i \in \partial a$. Let us look at a few examples.

Example 2.1 (The One-Dimensional Ising Model at Inverse Temperature $\beta > 0$) We have $\mathscr{X} = \{-1, 1\}$. Moreover, there are $M = N - 1$ factors $\psi_1, \ldots, \psi_M$ defined by

$$\psi_a : \mathscr{X}^N \to \mathbf{R}_{\geq 0}, \quad x \mapsto \exp(\beta x_a x_{a+1}).$$

Thus, $\partial a = \{a, a + 1\}$. Moreover, the factor graph is just a path with alternating variable and function nodes, see Fig. 1.

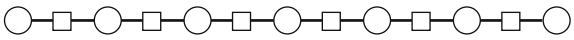

Fig. 1 The factor graph of the one-dimensional Ising model with $N = 7$. The *round vertices* represent the variable nodes and the *square vertices* the function nodes

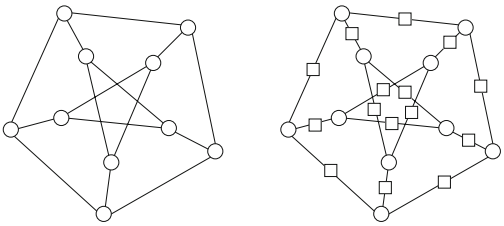

Fig. 2 A graph G (*left*) and the factor graph of the corresponding Potts antiferromagnet (*right*)

Example 2.2 (The Potts Antiferromagnet) In this case, we let $\mathscr{X} = \{1, \ldots, K\}$, where K is a given number of spins. In addition, we are given a graph $G = (V, E)$ on $N = |V|$ vertices. For each of the $M = |E|$ edges we define a factor ψ_e, $e = \{v, w\} \in E$, by letting

$$\psi_e : \mathscr{X}^V \to \mathbf{R}_{\geq 0}, \quad x \mapsto \exp(-\beta \mathbf{1}_{x_v = x_w}).$$

Thus, $\psi_e(x)$ is equal to $\exp(-\beta)$ if under the configuration x the two vertices v, w that the edge e connects have the same spin; otherwise, $\psi_e(x) = 1$. As before, $\beta > 0$ is a parameter, called the *inverse temperature*. The factor graph in this model arises simply from the graph G by placing a (square) factor node in the middle of each edge of G, see Fig. 2.

For a given model, the key quantity that we are going to be interested in is the partition function Z. The reason for this is that the partition function typically captures the key quantities of the model. More specifically, the partition function can generally be viewed as a function of the parameters that go into the definition of the functions ψ_a (such as the inverse temperature). By taking, e.g., the derivative of $\ln Z$ with respect to the inverse temperature, we obtain physically meaningful quantities such as the "average energy" of a state chosen from the Boltzmann distribution. Since the partition function typically scales exponentially in the "system size" N, we typically aim to calculate $\frac{1}{N} \ln Z$. Following [21], we refer to this quantity as the *free entropy*.

In fact, we will typically be interested in examples where it is meaningful to let N tend to infinity ("thermodynamic limit"). That is, strictly speaking we consider a *sequence* of Boltzmann distributions $(\mu[N])_N$ and their associated partition functions $(Z[N])_N$. To unclutter the notation, we typically drop the reference to N.

Thus, we would like to get a handle on

$$\lim_{N\to\infty} \frac{1}{N} \ln Z, \tag{1}$$

the so-called *free entropy density*.

This notion immediately leads to several questions. First, can we prove for a given model that the limit (1) exists? Second, if it exists, what properties does the limit have as a function of the various system parameters such as the inverse temperature? More specifically, we refer to parameter values where the free entropy density is non-analytic as *phase transitions*.

For example, it is well known that in the one-dimensional Ising model described above (parametrized by the inverse temperature β) there is no phase transition. By contrast, Onsager's well-known result shows that the Ising model on a two-dimensional lattice does exhibit a phase transition at a certain critical temperature β_c.

2.2 Belief Propagation

Generally, computing the free entropy is a very difficult problem. But in the case that the factor graph is a tree there is a systematic recipe called *Belief Propagation*. Thus, in this section we assume that the factor graph associated with the Boltzmann distribution μ is a tree. In addition, for the sake of simplicity we assume that $\mu(x) > 0$ for all configurations x. (This is the case in both the Ising and the Potts model as described in the previous section.)

Strictly speaking, Belief Propagation is a mechanism for calculating the marginal distribution of the spin of a given variable x_i under a configuration chosen from the Boltzmann distribution. To accomplish this, Belief Propagation performs a fixed point computation for a collection of so-called "messages" that travel along the edges of the factor graph.

More precisely, let $\mathscr{M}$ be the set of all families $\boldsymbol{\nu} = (\nu_{j\to a}, \hat{\nu}_{a\to j})_{a\in[M],j\in[N],j\in\partial a}$ such that $\nu_{j\to a}, \hat{\nu}_{a\to j}$ are probability distributions on $\mathscr{X}$. In other words, for every edge $\{x_j, \psi_a\}$ of the factor graph $\boldsymbol{\nu}$ features two distributions $\nu_{j\to a}, \hat{\nu}_{a\to j}$ on $\mathscr{X}$. We think of the first distribution $\nu_{j\to a}$ as a "message" from j to a, and of the second distribution $\hat{\nu}_{a\to j}$ as a "message" from a to j. Let us call $\mathscr{M}$ the *message space*.

We now define the *Belief Propagation operator* $\mathrm{BP} : \mathscr{M} \to \mathscr{M}$ as follows. For $j \in \partial a$ and $x_j \in \mathscr{X}$ we define $\boldsymbol{\eta} = \mathrm{BP}(\boldsymbol{\nu})$ by letting

$$\hat{\eta}_{a\to j}(x_j) \propto \sum_{x_{\partial a\setminus j}} \psi_a(x_{\partial a}) \prod_{i\in\partial a\setminus j} \nu_{i\to a}(x_i). \tag{2}$$

To be explicit, the above summation is over all vectors $x_{\partial a\setminus j} \in \mathscr{X}^{\partial a\setminus j}$, i.e., over all possible ways of assigning spins to the variables x_i, $i \neq j$, that affect factor ψ_a. For

each choice of $x_{\partial a\setminus j}$, we obtain a vector $x_{\partial a} = (x_i)_{i\in\partial a}$ by inserting the given value x_j. We evaluate ψ_a for this particular $x_{\partial a}$ and multiply the result by the "messages" from the other variables x_i, $i \in \partial a \setminus j$, to ψ_a. Finally, the $\propto$ sign in (2) indicates that we normalise such that $\hat{\eta}_{a\to j}$ becomes a probability distribution on $\mathscr{X}$. Spelled out in full, this means that we set

$$\hat{\eta}_{a\to j}(x_j) = \frac{\sum_{x_{\partial a\setminus j}} \psi_a(x_{\partial a}) \prod_{i\in\partial a\setminus j} \nu_{i\to a}(x_i)}{\sum_{y_j\in\mathscr{X}} \sum_{y_{\partial a\setminus j}} \psi_a(y_{\partial a}) \prod_{i\in\partial a\setminus j} \nu_{i\to a}(y_i)}. \tag{3}$$

Additionally, we define

$$\eta_{j\to a}(x_j) \propto \prod_{b\in\partial x_j\setminus a} \hat{\eta}_{b\to j}(x_j). \tag{4}$$

Here ∂x_j denotes the set of $b \in [M]$ such that x_j and ψ_b are neighbors in the factor graph. Of course, in (4) the $\propto$ symbol denotes the appropriate normalisation that turns $\eta_{j\to a}$ into a probability distribution on $\mathscr{X}$. Spelled out, (4) reads

$$\eta_{j\to a}(x_j) = \frac{\prod_{b\in\partial x_j\setminus a} \hat{\eta}_{b\to j}(x_j)}{\sum_{y_j\in\mathscr{X}} \prod_{b\in\partial x_j\setminus a} \hat{\eta}_{b\to j}(y_j)}. \tag{5}$$

The two equations (2) and (4) define the operator $\mathrm{BP} : \mathscr{M} \to \mathscr{M}$. The key property of this operator is summarised by the following theorem.

Theorem 2.3 *Assume that the factor graph is a tree.*

1. *The operator* BP *has a unique fixed point* $\boldsymbol{\nu}^*$ *(i.e.,* $\mathrm{BP}(\boldsymbol{\nu}^*) = \boldsymbol{\nu}^*$*).*
2. *If* T *is the diameter of the factor graph (i.e., the maximum distance between any two vertices), then for any* $\boldsymbol{\nu} \in \mathscr{M}$ *and any* $t > T$ *we have*

$$\mathrm{BP}^t(\boldsymbol{\nu}) = \boldsymbol{\nu}^*.$$

   That is, iterating the operator from any starting point $t > T$ *times yields the unique fixed point.*
3. *For* $j \in [N]$ *let* μ_j *be the marginal distribution of the* j*th variable under the Boltzmann distribution. Then for any* $x_j \in \mathscr{X}$ *we have*

$$\mu_j(x_j) \propto \prod_{b\in\partial j} \hat{\nu}^*_{b\to j}(x_j).$$

The proof of Theorem 2.3 is based on a relatively simple induction (over the size of the factor graph). The theorem states that Belief Propagation does indeed enable us to calculate the marginals of the individual variables under the Boltzmann distribution (if the factor graph is a tree). Indeed, part 2 of the theorem asserts that this calculation is quite efficient in terms of the number of fixed point iterations.

While we skip the proof of Theorem 2.3, let us at least provide some intuition why it is plausible that the BP fixed point $\boldsymbol{\nu}^*$ captures the marginals. The key observation (which can easily be fleshed out to prove Theorem 2.3) is that the "message" $\nu^*_{j\to a}$ is the marginal distribution of variable j in the modified Boltzmann distribution where we remove the factor ψ_a. Similarly, $\hat\nu^*_{a\to j}$ turns out to be the marginal distribution of j if we remove all factors ψ_b with $b \in \partial j \setminus a$. With this semantics in mind, it is easy to see that the fixed point equations resulting from (2) and (4) "make sense":

$$\hat\nu^*_{a\to j}(x_j) \propto \sum_{x_{\partial a\setminus j}} \psi_a(x_{\partial a}) \prod_{i\in\partial a\setminus j} \nu^*_{i\to a}(x_i), \quad \nu^*_{j\to a}(x_j) \propto \prod_{b\in\partial x_j\setminus a} \hat\nu^*_{b\to j}(x_j). \tag{6}$$

Indeed, the first equation reflects that in the absence of the other factors $b \in \partial j \setminus a$, the marginal of variable j is governed by ψ_a and the messages coming from the sub-trees rooted at the other variables $i \in \partial a \setminus j$. Similarly, the second equation reflects that in the absence of the factor ψ_a, variable j aligns itself to the sub-trees rooted at the other factors $b \neq a$ that it is adjacent to.

While Theorem 2.3 shows that the Belief Propagation fixed point captures the marginals, the following result states that it does, in fact, also yield the partition function.

Theorem 2.4 *Assume that the factor graph is a tree. Define a functional $F : \mathscr{M} \to \mathbf{R}$ by letting*

$$F(\boldsymbol{\nu}) = \sum_{a=1}^{M} F_a(\boldsymbol{\nu}) + \sum_{i=1}^{N} F_i(\boldsymbol{\nu}) - \sum_{(i,a):i\in\partial a} F_{ia}(\boldsymbol{\nu}), \text{ where}$$

$$F_a(\boldsymbol{\nu}) = \ln \sum_{x_{\partial a}} \psi_a(x_{\partial a}) \prod_{i\in\partial a} \nu_{i\to a}(x_i),$$

$$F_i(\boldsymbol{\nu}) = \ln \sum_{x_i} \prod_{b\in\partial i} \hat\nu_{b\to i}(x_i),$$

$$F_{ia}(\boldsymbol{\nu}) = \ln \sum_{x_i} \nu_{i\to a}(x_i)\hat\nu_{a\to i}(x_i).$$

If $\boldsymbol{\nu}^* \in \mathscr{M}$ is the Belief Propagation fixed point, then $\ln Z = F(\boldsymbol{\nu}^*)$.

Without going into any details, we note that the functional F, the so-called *Bethe free entropy*, fully solves the problem of calculating the free entropy in the case that the factor graph is a tree. An example of this is the one-dimensional Ising model, where Belief Propagation translates directly into the well-known transfer matrix method.

2.3 Disordered Systems

To this point we have been considering either one given factor graph or a natural family of factor graphs (e.g., in the case of the one-dimensional Ising model). But the main topic of this lecture is the case that the Boltzmann distribution (and the factor graph) under consideration is itself a random object. In physics, this case pertains to models of so-called "disordered systems" (such as glasses). If the Boltzmann distribution is itself a random object, our quantity of interest is

$$\lim_{N\to\infty} \frac{1}{N}\mathrm{E}[\ln Z],$$

the average of the free entropy density. Let us look at a few examples of "disordered systems".

Example 2.5 (The Potts Antiferromagnet on a Random Graph) Suppose that $\alpha > 0, \beta > 0, K \geq 2$ are given parameters and that $N > 0$ is an integer. Set $M = \lceil \alpha N \rceil$ and let $G(N, M)$ be a graph on the vertex set $[N]$ with precisely M edges chosen uniformly at random. Let μ be the Boltzmann distribution of the K-spin Potts antiferromagnet on the random graph. Then μ is itself a random object, and the partition function Z is a random variable. Thus, there are two levels of randomness. First, the random choice of the Boltzmann distribution. Second, once the Boltzmann distribution is fixed (or "quenched" in physics jargon), there is the random choice of a configuration.

In models of the above type, the distribution of the partition function is governed by two natural parameters: the inverse temperature β and the *density* α of the random factor graph. Hence, there is a two-dimensional phase diagram.

Example 2.6 (The Random Graph Coloring Problem) If we let $\beta \to \infty$, the Potts model leads naturally to a classical problem on random graphs that goes back to the work of Erdős and Rényi [15]. To be precise, let us define a Boltzmann distribution μ on a random graph $G(N, M)$ by setting $\psi_e(x) = 1$ if $x(v) \neq x(w)$ and $\psi_e(x) = 0$ otherwise for each edge $e = \{v, w\}$. Then μ is nothing but the uniform distribution over the K-colorings of the random graph $G(N, M)$ and Z is the number of K-colorings. Thus, Z depends on the parameter α, and figuring out its limiting behavior is a well-known open problem in the theory of random graphs. Formally, since Z may take the value 0, the quantity of interest is

$$\lim_{N\to\infty} \frac{1}{N}\mathrm{E}[\ln(Z \vee 1)].$$

Alternatively, one might consider

$$\lim_{N\to\infty} \mathrm{E}[Z^{1/N}].$$

However, it is not currently known that these limits even exists for all α.

The following well-known example is of particular relevance to theoretical computer science.

Example 2.7 (The Random K-SAT Problem) Let $K \geq 2$ be a fixed integer, let $\alpha > 0$ and let $M = \lceil \alpha N \rceil$ and $\mathscr{X} = \{0, 1\}$. We create a random bipartite factor graph as follows. There are N variable nodes $x_1, \ldots, x_N$ and M function nodes $\psi_1, \ldots, \psi_M$. We connect each function node ψ_a independently with a set of K variable nodes chosen uniformly at random among all $\binom{N}{K}$ such sets. Furthermore, for each function node ψ_a we choose a K-tuple $y_a = (y_{a,j})_{j=1,\ldots,K} \in \mathscr{X}^K$ uniformly at random. Now, if the K variables that ψ_a is connected to are $x_{a_1}, \ldots, x_{a_K}$ with $1 \leq a_1 < \cdots < a_K \leq N$, then we let

$$\psi_a(x) = 1 - \prod_{j=1}^{K} \mathbf{1}_{x_{a_j} = y_{a,j}}.$$

Thus, $\psi_a(x) = 1$ unless the K values that the K variables involved in ψ_i take precisely the "forbidden" value y_a. Much like in the graph coloring problem, the Boltzmann distribution in the random K-SAT problem is just the uniform distribution over all $x \in \mathscr{X}^N$ such that $\psi_a(x) = 1$ for all a.

Combinatorially, the random Boltzmann distribution described in the previous paragraph can be described as a Boolean formula. The variables of this formula are $x_1, \ldots, x_N$. Moreover, the function node ψ_a can be interpreted as a disjunction

$$(x_{a_1} \neq y_{a,1}) \vee \cdots \vee (x_{a_k} \neq y_{a,k}).$$

Thus, $\mu(x) > 0$ iff

$$\bigwedge_{a=1}^{M} [(x_{a_1} \neq y_{a,1}) \vee \cdots \vee (x_{a_k} \neq y_{a,k})].$$

This means that we can interpret μ as the uniform distribution over Boolean assignments that satisfy the above conjunctive normal form formula. The problem of satisfying such formulas is known in computer science as the (random) K-SAT problem ("satisfiability problem"). This problem plays a prominent role in computational complexity theory.

The K-SAT problem has a fairly intricate combinatorial nature. The following problem is a bit simpler but conceptually similar.

Example 2.8 (The Random Hypergraph 2-Coloring Problem) Let $K \geq 2$ be a fixed integer, let $\alpha > 0$ and let $M = \lceil \alpha N \rceil$ and $\mathscr{X} = \{0, 1\}$. As in the K-SAT problem, we create a random factor graph with N variable nodes $x_1, \ldots, x_N$ and M function nodes $\psi_1, \ldots, \psi_M$ by connecting each function node ψ_a independently with a set of K variable nodes chosen uniformly at random among all $\binom{N}{K}$ such sets. The factors

ψ_a are defined by

$$\psi_a(x) = 1 - \prod_{j=1}^{K} \mathbf{1}_{x_{a_j}=0} - \prod_{j=1}^{K} \mathbf{1}_{x_{a_j}=1}.$$

Thus, $\psi_a(x) = 1$ unless the K values that the K variables involved in ψ_i all take the same value (either 0 or 1).

Combinatorially, we can think of the above as a random hypergraph on N vertices in which each edge joins precisely K vertices (a so-called K-uniform hypergraph). Then $\mu(x) > 0$ iff x is a proper 2-coloring of this hypergraph, i.e., none of the edges comprises of vertices that all take the same color.

2.4 The Replica Symmetric Ansatz

In the examples from the previous section the underlying factor graphs are random. For most values of the density parameter α these factor graphs are likely to have a very complicated combinatorial structure. For instance, they feature many long cycles and they have excellent expansion properties. However, for any fixed value $\alpha > 0$ they do not contain large numbers of *short* cycles. More specifically, with probability tending to 1 as $N \to \infty$, only $o(N)$ vertices are going to be part of a cycle of length less than, say, $\ln N / \ln \ln N$. Thus, if we look at the part of the factor graph consisting of all vertices at distance less than, say, $\ln N / \ln \ln N$ from a given vertex, then very likely this sub-graph is a tree. In other words, the factor graphs from the previous section are *locally tree-like*.

This leads to the question whether Belief Propagation can be used to compute the marginals and/or the free entropy. Roughly speaking, this amounts to asking whether there are any "long-range correlations" in the above models or, more precisely, for what values of α long-range correlations prevail. Indeed, if the marginal of a variable node x_i is governed by the vertices at distance less than $\ln N / \ln \ln N$ from x_i, then we should expect that Belief Propagation yields at least a very good approximation to the marginal of x_i. The assumption that long-range correlations do not exist leads to the first variant of the cavity method, the *replica symmetric ansatz*, which is based on Belief Propagation.

Of course, we first need to adapt the Belief Propagation formalism to the specific types of trees that arise as the local neighborhoods in random factor graphs. These have two distinguished properties. First, they are random. Second, for most values of α they are infinite with probability close to one.

Instead of attempting to give a general formal account, let us work through two examples. First, let us consider the random hypergraph 2-coloring problem, arguably the simplest non-trivial case. Suppose we start with some variable node x_i, and also assume that N is large. Let us refer to the number of function nodes that

x_i is involved in as the *degree* of x_i. By construction, the average is $KM/N \sim \alpha K$ (because function nodes choose K vertices randomly). In fact, standard arguments show that as $N \to \infty$, the degree of x_i converges in distribution to a Poisson variable $\mathrm{Po}(\alpha K)$. Furthermore, each factor ψ_a that involves x_i involves another $K-1$ variables. For each such variable x_j, the number of function nodes other than ψ_a that x_j is involved in is again asymptotically a Poisson variable $\mathrm{Po}(\alpha K)$. (Due to the memorylessness property of the Poisson distribution, the number of other function nodes that x_j occurs in really has distribution $\mathrm{Po}(\alpha K)$, regardless of the fact that x_j occurs in ψ_a.)

Proceeding inductively, we see that the local neighborhood of x_i can be described by a multi-type branching process. In this process, there are two types of vertices: variable nodes and function nodes. The offspring of a variable node is a $\mathrm{Po}(\alpha K)$ number of function nodes, and the offspring of a function node consists of precisely $K-1$ variable nodes. Thus, the resulting tree is infinite with a non-vanishing probability if $\alpha K(K-1) > 1$.

While we learned about Belief Propagation applied to a finite tree, it seems difficult to extend this procedure to infinite trees. Nonetheless, the recursive structure of the random tree described above enables us to generalise the notion of a Belief Propagation fixed point. Indeed, in the scenario described above we are not specifically interested in the marginal of a specific node x_i, but rather in the *distribution* of these marginals, viewed as probability distributions on $\mathscr{X}$. That is, suppose we first choose a random Boltzmann distribution and then choose a variable node x_i randomly, what is the distribution of the marginal μ_i? Hence, the object that we are interested in is a distribution over distributions on $\mathscr{X}$.

Returning to the Belief Propagation formalism on random trees, this means that we are interested in the distribution of the fixed point messages $\nu^*_{j\to a}, \hat\nu^*_{a\to j}$ that travel along an edge of the tree. Now, if we fix a variable node x_j and the function node ψ_a that is the parent of x_j in the tree, the crucial observation is as follows. Consider the variables $x_i \neq x_j$ that occur in the function nodes $b \neq a$ that are the children of x_j. Then these variables x_i are themselves the root of a random tree with exactly the same distribution as the random tree emerging from x_j. Therefore, the *distribution* of the messages from the x_i to b should be identical, and it should also be the same as the distribution of the message from x_j to a. A similar observation applies to the messages $\hat\nu^*_{a\to j}$.

This observation inspires the idea of viewing the fixed point condition (6) as a *distributional* fixed point equation. That is, we aim to find $\boldsymbol{\nu}^* = (\nu^*, \hat\nu^*)$ such that $\nu^*, \hat\nu^*$ are probability distributions over probability distributions over $\mathscr{X}$ (sic!) such that the following system of distributional equations holds true:

$$\hat\nu^*(x_1) \stackrel{\mathrm{d}}{=} \frac{\sum_{x_2,\ldots,x_k\in\mathscr{X}} \psi(x_1,\ldots,x_k)\prod_{i=2}^k \nu_i^*(x_i)}{\sum_{y_1,\ldots,y_k} \psi(y_1,\ldots,y_k)\prod_{i=1}^k \nu_i^*(y_i)}, \tag{7}$$

$$\nu^*(x) \stackrel{\mathrm{d}}{=} \frac{\prod_{b=1}^d \hat\nu_b^*(x)}{\sum_{y\in\mathscr{X}}\prod_{b=1}^d \hat\nu_b^*(y)} \tag{8}$$

Here $\psi(x_1,\dots,x_k) = 0$ if $x_1 = \cdots = x_k$ and $\psi(x_1,\dots,x_k) = 1$ otherwise (mirroring the hypergraph 2-coloring problem). Furthermore, $\nu_1^*,\dots,\nu_{k-1}^*$ are probability distributions on $\mathscr{X}$ that are chosen independently from the distribution ν^*. In addition, $d = \mathrm{Po}(\alpha K)$ is a Poisson random variable, and $\hat\nu_1^*,\dots,\hat\nu_d^*$ are probability distributions on $\mathscr{X}$ that are chosen independently of each other and of d from $\hat\nu^*$.

In the case of hypergraph 2-coloring, the two equations (7) and (8) have the canonical solution

$$\nu^* = \hat\nu^* = \mathbf{1}_{(1/2,1/2)}. \tag{9}$$

That is, both ν^*, $\hat\nu^*$ are atoms that put the full mass on the uniform distribution on $\mathscr{X}$.

A more complicated situation arises in the case of the random K-SAT problem. Here the fixed point equations take the same form as (7) and (8). The only difference is that now

$$\psi(x_1,\dots,x_K) = 1 - \prod_{i=1}^{K} \mathbf{1}_{x_i=z_i},$$

where $(z_1,\dots,z_K) \in \mathscr{X}^k$ is a random K-tuple (chosen independently of everything else). Due to the randomness involved in ψ, there is no trivial fixed point anymore. In fact, it is expected that the (relevant) fixed point is a mixture of a distribution with a countable support (corresponding to the case that the tree is finite) and a continuous distribution. Additionally, in both the case of random hypergraph 2-coloring and random K-SAT there is no guarantee that Eqs. (7) and (8) have a *unique* solution. We will return to this question below.

Anyhow, assuming that we have got the (correct) distributional fixed point $\boldsymbol{\nu}$, it is now straightforward to derive the prediction as to the free entropy. All we need to do is to calculate the expectation of the functional F from Theorem 2.4. To be explicit, the *replica symmetric solution* reads

$$\begin{aligned}
\psi &= \alpha \cdot F_a + F_i - \alpha K \cdot F_{ia}, \text{ where}\\
F_a &= \mathrm{E}\left[\ln \sum_{x\in\mathscr{X}^k} \psi(x) \prod_{i=1}^{K} \nu_i(x_i)\right],\\
F_i &= \mathrm{E}\left[\ln \sum_{x\in\mathscr{X}} \prod_{b=1}^{d} \hat\nu_b(x)\right],\\
F_{ia} &= \mathrm{E}\left[\ln \sum_{x\in\mathscr{X}} \nu_1(x)\hat\nu_1(x)\right].
\end{aligned}$$

Here $\nu_1, \nu_2, \ldots$ are independently chosen from the distribution ν^*, $\hat{\nu}_1, \hat{\nu}_2, \ldots$ are independent samples from the distribution $\hat{\nu}^*$ (and also independent of $\nu_1, \nu_2, \ldots$), and d has a Poisson distribution with mean αK that is independent of everything else. According to the replica symmetric ansatz,

$$\psi = \lim_{N\to\infty} \frac{1}{n} \mathrm{E}[\ln(Z \vee 1)].$$

In the case of the hypergraph 2-coloring problem, it is easy to work out ψ due to the particularly simple form of the fixed point (9). Indeed, we find that

$$\psi = \ln 2 + \alpha \ln(1 - 2^{1-K}). \tag{10}$$

It is easily verified that the above is zero if $\alpha = 2^{K-1} \ln 2 - \ln 2/2 + \varepsilon_K$, where $\varepsilon_K \to 0$ in the limit of large K. (Generally, in problems such as random K-SAT or random K-uniform hypergraph coloring, one obtains fairly simple asymptotics in the limit of large K.) Thus, according to the replica symmetric ansatz, we expect that in the random hypergraph 2-coloring problem Z is positive up to

$$\alpha = 2^{K-1} \ln 2 - \ln 2/2 + \varepsilon_K, \tag{11}$$

and that $Z = 0$ for larger α with probability tending to one as $N \to \infty$.

We will see in due course that this prediction is inaccurate. The reason for this is that the key assumption behind the replica symmetric ansatz, the absence of long-range effects, does not hold for α near $2^{K-1} \ln 2 - \ln 2/2$.

2.5 Replica Symmetry Breaking

How can we formalise the presence or absence of "long-range effects"? For the sake of presentation, let us stick to the example of random hypergraph 2-coloring. For a variable node x_i let us denote by $N^{\omega}(x_i)$ the sub-graph of the factor graph consisting of all vertices at distance at most 2ω from x_i. If $\omega \leq \ln n / \ln\ln n$, then with probability tending to one this sub-graph will be a tree whose leaves are variable nodes. Analogously, let $N^{>\omega}(x_i)$ denote the sub-graph induced on the vertices at distance greater than 2ω from x_i.

The following property, which is called *Gibbs uniqueness*, is very strong way of formalising the absence of long-range effects:

$$\lim_{\omega\to\infty} \lim_{N\to\infty} \mathrm{E}\left[\sup_{x\in\mathcal{X}^N:\mu(x)>0} |\mu_i(1 \mid x_{N^{>\omega}(x_i)}) - \mu_i(1)|\right] = 0. \tag{12}$$

In words, if we condition on *any* possible configuration of the vertices that are at distance greater than 2ω from x_i, then in the limit as the system size tends to infinity the impact of that conditioning on x_i tends to 0 for large ω. Of course, in (12) the expectation refers to the choice of the Boltzmann distribution μ.

The Gibbs uniqueness condition is very strong and there have been various examples where it has been shown to be sufficient to verify the validity of the replica symmetric solution (e.g., [12]). However, the condition is so restrictive that we normally expect the replica symmetric solution to be correct for much wider ranges of the parameters. (Yet Gibbs uniqueness does generally hold whenever the parameters are chosen such that the factor graph is sub-critical, i.e., there is no "giant component". In random hypergraph 2-coloring this is true if $\alpha K(K-1) < 1$.)

A weaker condition of a similar flavour is *non-reconstruction*. In this case, rather than taking the worst possible configuration of the far away nodes, we average with respect to the Boltzmann distribution. Formally, the requirement reads

$$\lim_{\omega\to\infty}\lim_{N\to\infty}\mathrm{E}\left[\sum_{x\in\mathscr{X}^N}\mu(x)\cdot|\mu_i(1\,|\,x_{N^{>\omega}(x_i)})-\mu_i(1)|\right]=0. \tag{13}$$

In hypergraph 2-coloring, this condition is expected to be valid for $\alpha < (1-\varepsilon_K)2^{K-1}\ln K/K$, where $\varepsilon_K \to 0$ in the limit of large K. Provably, the condition is violated for $\alpha > (1+\varepsilon_K)2^{K-1}\ln K/K$. Thus, while we might expect that (13) provides a valid condition for the success of the replica symmetric ansatz, the condition is still too restrictive to get us anywhere near the replica symmetric prediction $\alpha = (1-\varepsilon_K)2^{K-1}\ln 2$ as to where 2-colorings cease to exist in a random hypergraph.

Yet another, weaker condition is that the joint distribution of a number of "far away variables" is asymptotically independent. This can be formalised by the requirement that for any fixed integer $\omega > 0$,

$$\lim_{N\to\infty}\mathrm{E}\,\|\mu_1\otimes\cdots\otimes\mu_\omega-\mu(x_1,\ldots,x_\omega)\|_{\mathrm{TV}}=0. \tag{14}$$

Here $\mu_1\otimes\cdots\otimes\mu_\omega$ denotes the product of the marginal distributions of the first ω variables $x_1,\ldots,x_\omega$. Moreover, $\mu(x_1,\ldots,x_\omega)$ signifies the joint distribution of these variables. As before, the expectation is over the choice of the random Boltzmann distribution. One might expect that the condition (14) is sufficient to warrant validity of the replica symmetric solution, but it seems that we are very far from proving this rigorously in any generality.

In the physics literature a supposedly equivalent, more combinatorial perspective on the properties (13) and (14) has been put forward. We continue to stick to the example of the random hypergraph 2-coloring problem. According to this combinatorial view, if condition (13) fails to hold, then the support of the Boltzmann distribution (in our example, the set of 2-colorings of the random hypergraph) decomposes into a large number of well-separated "clusters". Furthermore, if α is

such that (14) holds, then each of these clusters only contains an exponentially small fraction of all the 2-colorings. In physics jargon, this scenario is known as *dynamical replica symmetry breaking*. In the example of random hypergraph 2-coloring, the physics prediction is that dynamical replica symmetry breaking occurs for

$$(1 + o_K(1))2^{K-1} \ln K/K < \alpha < 2^{K-1} \ln 2 - \ln 2 + o_K(1). \tag{15}$$

Here and throughout, we denote by $o_K(1)$ a term that tends to 0 in the limit of large K.

By contrast, if α is such that (14) fails to hold, then the largest clusters are expected to have the same order of magnitude as Z, the total number of 2-colorings. Moreover, a bounded number of clusters are expected to dominate the entire set of 2-colorings. This scenario is referred to as *static replica symmetry breaking* or *condensation*. In hypergraph 2-coloring, condensation is expected to occur for

$$\alpha > 2^{K-1} \ln 2 - \ln 2 + o_K(1). \tag{16}$$

Furthermore, the replica symmetric solution is expected to be inaccurate in the condensation phase.

To be explicit, in the case of random hypergraph 2-coloring we could define the clusters as follows. We can view the set of all 2-colorings of as a subset $\Lambda \subset \{0, 1\}^N$. Now, turn Λ into a graph by connecting any two $x, x' \in \Lambda$ if their Hamming distance is bounded by, say, $N/\ln N$ (more generally, any unbounded function that is $o(N)$ would do). Then the clusters are the connected components of this auxiliary graph. Furthermore, each clusters is expected to give rise to a (near-)fixed point of Belief Propagation on the factor graph. To be precise, by iterating the Belief Propagation operator starting from any solution in the cluster, one might expect to obtain a collection of messages that is asymptotically invariant under a further application of the Belief Propagation operator for large enough N. This is not a contradiction to Theorem 2.3 because the factor graph is not a tree with probability tending to 1; thus, Belief Propagation might have multiple fixed points.

Additionally, the internal structure of the clusters is expected to be characterised by "frozen variables". More specifically, we call a variable x_i *frozen* in a cluster $\mathscr{C}$ if x_i takes the same value in all the configurations in $\mathscr{C}$. Hence, we can represent the cluster $\mathscr{C}$ by a "code word" in $\{0, 1, *\}^N$, where we represent a variable that is frozen to 0 by a 0, a variable that is frozen to 1 by a 1, and an unfrozen variable by $*$. This symbolic representation does not, of course, imply that the variables marked $*$ are perfectly independent.

To cope with (static) replica symmetry breaking, physicists have developed a stronger version of the cavity method, the so-called *1-step replica symmetry breaking ansatz*. The basic idea behind this approach is to perform a similar approach as Belief Propagation for a modified Boltzmann distribution. More precisely, imagine a decomposition of the Boltzmann distribution as a convex combination of probability

measures that correspond to the "clusters", i.e.,

$$\mu = \sum_{1 \le i \le \Sigma} w_i \cdot \mu_i.$$

For instance, in hypergraph 2-coloring the μ_i are just the uniform distributions on the various clusters, and the pre-factors w_i are proportional to the sizes of the clusters. Roughly speaking, the key idea behind the replica symmetric ansatz is to apply Belief Propagation not to the actual 2-colorings, but to the "clusters" μ_i. In the process of this, we introduce a parameter that allows us to reweigh the pre-factors w_i, the so-called *Parisi parameter*.

In the extreme case, the re-weighting is such that all clusters have exactly the same weight, i.e., we look at the uniform distribution over clusters (Parisi parameter 0). The merit of this is that it does away with the issues that condensation causes. More precisely, the condition (14) cannot be expected to hold in the presence of condensation. This is because two states x, x' chosen from the Boltzmann distribution have a constant probability of belonging to the same cluster, in which case they are strongly correlated, and a non-vanishing probability of belonging to distinct clusters. Hence, it is implausible that the joint distribution factorises as required by (14). However, if we consider the uniform distribution over *clusters*, i.e., if we disregard the volumes of the clusters, then there is a chance that a condition similar to (14) might still hold. This is because there is a huge total number of clusters (exponential in N), most of which only contain a very small number of 2-colorings. Thus, considering the uniform distribution over clusters effectively does away with the big clusters that dominate the Boltzmann distribution.

Technically, Belief Propagation can be extended to this case of considering the uniform distribution over clusters systematically. Roughly speaking, this amounts to generalising Belief Propagation to the case that we add the "joker spin" $*$ to represent unfrozen variables, and interpret the spins 0, 1 to indicate that a variable is frozen to this value. This leads to a modified message passing scheme called *Survey Propagation*. In addition, there is a generalisation of the Bethe free entropy. In the case of Survey Propagation, this formula yields a prediction as to the number Σ of *clusters*, rather than the free entropy. From this formula it is possible to make a prediction as to the threshold value of α up to which the random hypergraph admits a 2-coloring: this is expected to be the largest density up to which $\lim_{N\to\infty} \frac{1}{N}\mathrm{E}[\ln(1 \vee \Sigma)] > 0$. In the case of hypergraph 2-coloring, this threshold is expected to be

$$\alpha = 2^{K-1} \ln 2 - \left(\frac{\ln 2}{2} + \frac{1}{4}\right) + o_K(1). \tag{17}$$

In the following lectures, we will see to what extent the physics picture can be turned into rigorous mathematics. Throughout, we are going to stick to the example of the random hypergraph 2-coloring problem.

3 Classical Rigorous Results

In this section we discuss the "classical" rigorous approach to problems such as random hypergraph 2-coloring. This material is largely independent of the physics predictions from the previous section.

While much of this material generalises to other problems, we are going to focus on is the 2-coloring problem in random K-uniform hypergraphs. Thus, fix an integer $K \geq 3$, let N, M be a (large) integers, and set $\alpha = M/N$. Recall that a *K-uniform hypergraph* Φ with N vertices and M edges consists of a *vertex set* V of size $|V| = N$ and an *edge set* E of size $|E| = M$ such that each $e \in E$ is a K-element subset of V. We say that the hypergraph Φ is 2-*colorable* if there is a map $\sigma : V \to \{0, 1\}$ such that $\sigma(e) = \{0, 1\}$ for all $e \in E$. In other words, if we think of $0, 1$ as colors, then none of the edges is monochromatic. As we saw in the previous section, the hypergraph can be represented canonically by a bipartite factor graph $G(\Phi)$ whose vertices correspond to the vertices and edges of Φ, and where there is an edge between v and e iff $v \in e$.

Given K, N, M, let $\boldsymbol{\Phi} = \boldsymbol{\Phi}_K(N, M)$ denote the uniformly random K-uniform hypergraph on $V = [N]$ with M edges. Throughout, we are going to be interested in the situation that $N \to \infty$, while $\alpha = M/N$ remains fixed. We say that an event $\mathscr{A}$ occurs *with high probability* ('w.h.p.') if $\lim_{N\to\infty} \mathrm{P}[\mathscr{A}] = 1$. The questions that we are going to be dealing with are the following.

- For what α is $\boldsymbol{\Phi}$ 2-colorable w.h.p.?
- If it is, how many 2-colorings are there?

Throughout, we write $f(N) \sim g(N)$ if $\lim_{N\to\infty} f(N)/g(N) = 1$.

3.1 Basics

The following result shows that there is a "non-uniform threshold" for 2-colorability.

Theorem 3.1 ([17]) *For any $K \geq 3$ there is a sequence $\alpha_{2-col}(K, N)$ such that for any $\varepsilon > 0$ the following is true.*

- *If $M/N < (1 - \varepsilon)\alpha_{2-col}(K, N)$, then $\boldsymbol{\Phi}$ is 2-colorable w.h.p.*
- *If $M/N > (1 + \varepsilon)\alpha_{2-col}(K, N)$, then $\boldsymbol{\Phi}$ fails to be 2-colorable w.h.p.*

Unfortunately, the sequence $(\alpha_{2-col}(K, N))_{N\geq 1}$ is not currently known to converge (although it is, of course, known that

$$0 < \liminf \alpha_{2-col}(K, N) \leq \limsup \alpha_{2-col}(K, N) < \infty$$

for any $K \geq 3$). Thus, we do not currently know that there is a 2-colorability threshold $\alpha_{2-col}(K)$ that is *independent* of N. Moreover, the proof of Theorem 3.1 does not reveal the actual value of $\alpha_{2-col}(K, N)$. Theorem 3.1 is merely a corollary

of a much more general result from [17]. The proof of this more general result is based on discrete Fourier analysis, and is beyond the scope of this lecture.

Let Z be the number of 2-colorings of $\boldsymbol{\Phi}$. This corresponds to the partition function of the uniform distribution over the set of 2-colorings of $\boldsymbol{\Phi}$ (except that this set might be empty).

What can we say about the location of the 2-colorability threshold α_{2-col}? Let us start with some simple observations. First, if $\alpha < K^{-1}(K-1)^{-1}$, then the random hypergraph $\boldsymbol{\Phi}$ is well-known to have a very simple structure. Indeed, for $\alpha < K^{-1}(K-1)^{-1}$, all components of $G(\boldsymbol{\Phi})$ contain only $O(\ln N)$ vertices w.h.p. Furthermore, all but $O(1)$ components of $G(\boldsymbol{\Phi})$ are trees, while the remaining components contain only one cycle ("unicyclic components") w.h.p. [24]. In particular, each of these components is 2-colorable, and so is $\boldsymbol{\Phi}$. Additionally, Theorems 2.3 and 2.4 show that Belief Propagation and the Bethe formula yield the correct value of $\mathrm{E}\ln[Z \vee 1]$. By contrast, for $\alpha > K^{-1}(K-1)^{-1}$ the random hypergraph $\boldsymbol{\Phi}$ is going to have a connected component on $\Omega(N)$ vertices with a plethora of intersecting cycles w.h.p. As α increases, the relative size of this component grows rapidly [24].

Yet it is easy to compute the *expected* number of 2-colorings for any $\alpha > 0$. Indeed, we have

$$\frac{1}{N}\ln \mathrm{E}[Z] \sim \ln 2 + \alpha \ln(1-2^{1-K}). \tag{18}$$

To see this, observe that there are 2^N maps $\sigma : V \to \{0,1\}$. Furthermore, given *any* σ, the probability that a random edge e is bichromatic is at most $1-2^{1-K}$. This is because e consists of K random vertices, and the probability that all of them have the same color is

$$\left[\binom{|\sigma^{-1}(1)|}{K} + \binom{|\sigma^{-1}(0)|}{K}\right] \Big/ \binom{N}{K}. \tag{19}$$

In addition, (19) shows that the probability that a random edge is monochromatic is $1-2^{1-K}+o(1)$ if σ is such that $|\sigma^{-1}(1)| \sim N/2$. Since $(1-o(1))2^N$ maps $\sigma : V \to \{0,1\}$ have this property, we obtain (18).

We observe that (18) coincides with the Belief Propagation prediction (10) as to the value of $\frac{1}{N}\mathrm{E}[\ln Z \vee 1]$. This is a consequence of the particularly strong symmetry properties of the hypergraph 2-coloring problem. In other problems such as random K-SAT, the "annealed average" $\frac{1}{N}\ln \mathrm{E}[Z]$ differs from the "quenched average" $\frac{1}{N}\mathrm{E}[\ln Z \vee 1]$.

As an immediate consequence, we obtain that $\boldsymbol{\Phi}$ fails to be 2-colorable w.h.p. for any α such that

$$\ln 2 + \alpha \ln(1-2^{1-K}) < 0.$$

Since the expression is linear in α, it is easily verified that this occurs for $\alpha \geq 2^{k-1}\ln 2 - \frac{\ln 2}{2} + o_K(1)$. Thus, we have

Proposition 3.2 *If $\alpha > 2^{K-1}\ln 2 - \frac{\ln 2}{2} + o_K(1)$, then $\boldsymbol{\Phi}$ fails to be 2-colorable w.h.p.*

Once more, we observe that Proposition 3.2 matches the "replica symmetric prediction" (11). Furthermore, we emphasise that despite the use of the asymptotic notation $o_K(1)$, we do *not* take $K = K(N)$ to be a sequence that tends to infinity as $N \to \infty$. Instead, K always remains a *fixed* number, and the $o_K(1)$ simply hides an error term ε_K that we do not bother to specify but that can be made arbitrarily small by choosing K large enough.

3.2 The "Vanilla" Second Moment Method

As a next step, we aim to obtain an improved lower bound on α_{2-col}. This section follows [2].

Theorem 3.3 ([2]) *Assume that*

$$\alpha < 2^{K-1}\ln 2 - \frac{1+\ln 2}{2} - \varepsilon_K. \tag{20}$$

Then $\boldsymbol{\Phi}$ is 2-colorable w.h.p.

The proof of Theorem 3.3 is based on the *second moment method*. For the sake of technical simplicity, we are going to work with a modified random variable Z_e rather than Z. More precisely, let us call a map $\sigma : V \to \{0,1\}$ *equitable* if

$$||\sigma^{-1}(0)| - |\sigma^{-1}(1)|| \leq \sqrt{N},$$

and let Z_e be the number of equitable 2-colorings of $\boldsymbol{\Phi}$. Using Stirling's formula and (18), we check that

$$\frac{1}{N}\ln \mathrm{E}[Z_e] \sim \frac{1}{N}\ln \mathrm{E}[Z] \sim \ln 2 + \alpha \ln(1 - 2^{1-K}). \tag{21}$$

Thus, at least as far as the expectation goes, we are not giving away much by confining ourselves to equitable 2-colorings.

We are going to compare $\mathrm{E}[Z_e^2]$ to $\mathrm{E}[Z_e]^2$. More precisely, we are going to show that for α satisfying (20) we have

$$\mathrm{E}[Z_e^2] \leq C \cdot \mathrm{E}[Z_e]^2, \tag{22}$$

where $C = C(K) > 0$ is independent of N. Then the *Paley-Zygmund inequality* (valid for any non-negative random variable)

$$\mathrm{P}\left[Z_e > 0\right] \geq \frac{\mathrm{E}[Z_e]^2}{\mathrm{E}[Z_e^2]} \tag{23}$$

implies that

$$\liminf_{N\to\infty} \mathrm{P}\left[Z_e > 0\right] \geq C^{-1} > 0.$$

Hence, Theorem 3.1 implies that indeed $\boldsymbol{\Phi}$ is 2-colorable w.h.p. for any $\alpha' < \alpha$.

Thus, to prove Theorem 3.3 we "just" need to verify (22). That is, we need to work out the second moment: summing over equitable σ, τ, we have

$$\begin{aligned}
\mathrm{E}[Z_e^2] &= \sum_{\sigma,\tau} \mathrm{P}\left[\text{both } \sigma, \tau \text{ are 2-colorings}\right] \\
&= \sum_{\sigma} \sum_{\tau} \mathrm{P}\left[\tau \text{ is a 2-coloring} | \sigma \text{ is a 2-coloring}\right] \cdot \mathrm{P}\left[\sigma \text{ is a 2-coloring}\right] \\
&\leq \mathrm{E}[Z_e] \cdot \max_{\sigma} \mathrm{E}[Z_e | \sigma \text{ is a 2-coloring}].
\end{aligned}$$

In the last line, which simply follows from symmetry, σ stands for an equitable map.

Let $\mathscr{C}_\sigma$ denote the event that σ is a 2-coloring. Then (22) is equivalent to

$$\mathrm{E}\left[Z_e | \mathscr{C}_\sigma\right] \leq C \cdot \mathrm{E}[Z_e] \tag{24}$$

for all equitable σ. To compute the conditional expectation, fix an equitable σ and let

$$Z_\omega = \left|\{\tau : \mathrm{dist}(\sigma, \tau) = \omega N,\ \tau \text{ is an equitable 2-coloring}\}\right|.$$

Here $\mathrm{dist}(\sigma, \tau) = |\{v \in V : \sigma(v) \neq \tau(v)\}|$ denotes the Hamming distance. We can write $\mathrm{E}\left[Z_e | \mathscr{C}_\sigma\right]$ as

$$\mathrm{E}\left[Z_e | \mathscr{C}_\sigma\right] = \sum_{w=0}^{N} \mathrm{E}[Z_{w/N} | \mathscr{C}_\sigma].$$

It is easy to compute $\mathrm{E}[Z_{w/N} | \mathscr{C}_\sigma]$. Indeed, consider one random edge e in our hypergraph. What is the probability that e is bichromatic under both σ and another equitable τ with $\mathrm{dist}(\sigma, \tau) = \omega N$? Clearly, for symmetry reasons we have

$$\mathrm{P}\left[e \text{ is monochromatic under } \sigma\right] \sim \mathrm{P}\left[e \text{ is monochromatic under } \tau\right] \sim 2^{1-K}.$$

In addition,

$$\mathrm{P}\left[e \text{ is monochromatic under } \tau | e \text{ is monochromatic under } \sigma\right] \sim \omega^K + (1-\omega)^K.$$

In fact, for this event to occur either σ, τ must agree on all K vertices in e, or they must differ on all of them. The formula follows because these vertices are (essentially) uniformly distributed and independent. Thus, by inclusion/exclusion

$$\mathrm{P}\left[e \text{ is bichromatic under both } \sigma, \tau\right] \sim 1 - 2^{2-K} + 2^{1-K}(\omega^K + (1-\omega)^K).$$

Since the edges of $\boldsymbol{\Phi}$ are (essentially) independent, we get

$$\frac{1}{N} \ln \mathrm{P}\left[\mathscr{C}_\sigma, \mathscr{C}_\tau\right] \sim \alpha \ln(1 - 2^{2-K} + 2^{1-K}(\omega^K + (1-\omega)^K)),$$

and thus

$$\frac{1}{N} \ln \mathrm{P}\left[\mathscr{C}_\tau | \mathscr{C}_\sigma\right] \sim \alpha \ln\left(1 - \frac{1 - \omega^K - (1-\omega)^K}{2^{K-1} - 1}\right).$$

Furthermore, by Stirling's formula the logarithm of the total number of equitable τ with $\mathrm{dist}(\sigma, \tau) = \omega n$ satisfies

$$\frac{1}{N} \ln \binom{N}{\omega N} \sim H(\omega) = -\omega \ln \omega - (1-\omega) \ln(1-\omega).$$

Accounting for the various error terms carefully, we obtain

$$\begin{aligned} &\mathrm{E}[Z_\omega] \leq C \cdot \exp(f(\omega)N), \qquad \text{where} \\ &f(\omega) = H(\omega) + \alpha \ln\left(1 - \frac{1 - \omega^K - (1-\omega)^K}{2^{K-1} - 1}\right). \end{aligned} \tag{25}$$

Because $f(\omega)$ is in the exponent, the sum is dominated by the maximum function value. Indeed, using the "Laplace method", one obtains the following.

Lemma 3.4 *We have* $\mathrm{E}[Z_e | \mathscr{C}_\sigma] \leq C \cdot \exp(N \max_{\omega \in (0,1)} f(\omega))$.

Thus, we need to study the function $f(\omega)$. Since $f(\omega) = f(1-\omega)$, we have $f'(1/2) = 0$. Indeed, $\omega = 1/2$ turns out to be a local maximum, and a bit of algebra shows that

$$\exp(f(1/2)N) = \Theta(\mathrm{E}(Z_e)). \tag{26}$$

Hence, we find that

$$\mathrm{E}[Z_e | \mathscr{C}_\sigma] \leq C \cdot \mathrm{E}[Z_e] \quad \text{iff} \quad \max_{\omega \in [0,1]} f(\omega) = f(1/2).$$

Basic calculus shows that this is true for α satisfying (20), cf. Fig. 3. Thus, we have got Theorem 3.3.

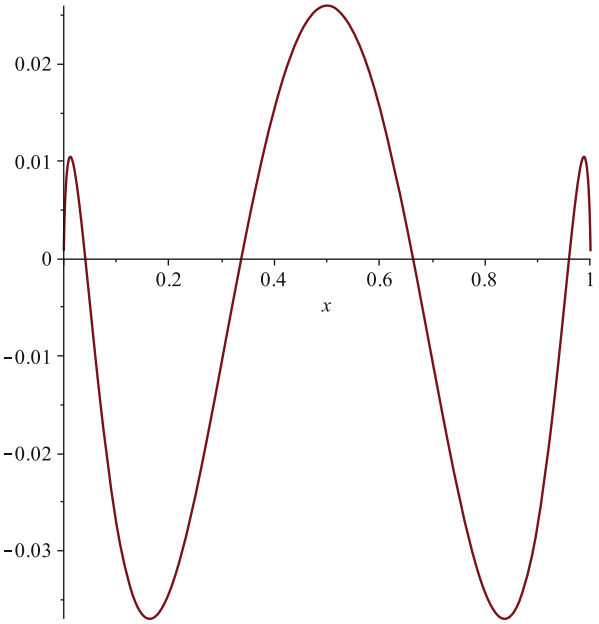

Fig. 3 The function f for $K = 7$ and $\alpha = 2^{K-1} \ln 2 - 2$

Corollary 3.5 ([1]) *Assume that (20) holds. Then* $\ln Z \sim \ln \mathrm{E}[Z]$ *w.h.p.*

Proof Due to (21) it suffices to prove that $\ln Z \sim \ln \mathrm{E}[Z_e]$ w.h.p. This follows from the strong form of the Palay-Zygmund inequality

$$\mathrm{P}\left[Z_e \geq \frac{1}{2}\mathrm{E}[Z_e]\right] \geq \frac{\mathrm{E}(Z_e)^2}{4\mathrm{E}(Z_e^2)}$$

together with a stronger version of Theorem 3.1 from [1] that provides a concentration result for Z. □

4 A Physics-Enhanced Rigorous Approach

We continue to consider the example of the random hypergraph 2-coloring problem. How do the results from the previous section compare with the physics predictions? According to the cavity method,

- dynamic 1RSB occurs at $\alpha_{d1RSB} \sim 2^{K-1} \ln K/K$. In particular, that the set of solutions "clusters" was an important aspect of the physics picture. By comparison, in Sect. 3 the idea that there are "clusters" with "frozen variables" went completely unnoticed.

- the static 1RSB phase transition occurs at

$$\alpha_{s1RSB} = 2^{K-1} \ln 2 - \ln 2 + o_K(1).$$

  This is a bit greater than the bound provided by Theorem 3.3. For $\alpha > \alpha_{s1RSB}$, we expect that $\ln Z < \ln \mathrm{E}[Z] - \Omega(N)$ w.h.p.
- we have

$$\alpha_{2-col} = 2^{K-1} \ln 2 - \left(\frac{1}{4} + \frac{\ln 2}{2}\right) + o_K(1).$$

  This is half-way between the bounds provided by Proposition 3.2 and Theorem 3.3.

In this section we will first develop a rigorous approach towards proving that, indeed, the set of 2-colorings of the random hypergraph decomposes into clusters for $\alpha > (1 + o_K(1))2^{K-1} \ln K/K$. This is by means of a trick from [1]; following [20], we refer to this technique as "quiet planting". Then we will see how this trick can be used to establish the condensation phase transition and the asymptotic threshold for the existence of 2-colorings rigorously.

4.1 Quiet Planting

In this section we are going to verify the d1RSB picture that the set of 2-colorings of the random hypergraph decomposes into clusters. Recall that the *Boltzmann distribution* that we are interested in can be described as follows.

B1. Create a random hypergraph $\boldsymbol{\Phi}$.
B2. Sample a 2-coloring σ of $\boldsymbol{\Phi}$ uniformly at random.
B3. The result is $(\boldsymbol{\Phi}, \sigma)$.

Thus, the above experiment induces a probability distribution on hypergraph/2-coloring *pairs* $(\boldsymbol{\Phi}, \sigma)$.

The Boltzmann distribution is difficult to analyse directly. By contrast, the following so-called *planted distribution* is quite accessible.

P1. Choose a map $\sigma : V \to \{0, 1\}$ uniformly at random.
P2. Choose a hypergraph $\boldsymbol{\Phi}$ that is 2-colored under σ uniformly at random.
P3. The result is $(\boldsymbol{\Phi}, \sigma)$.

We emphasise that step **P2** is indeed easy to implement: we simply choose a set of M random edges randomly out of the

$$\binom{N}{K} - \binom{|\sigma^{-1}(0)|}{K} - \binom{|\sigma^{-1}(1)|}{K}$$

edges that are bichromatic under σ. A double-counting argument implies the following connection between the two distributions.

Lemma 4.1 ("Quiet Planting Lemma", [1]) *Assume that α is such that $\ln Z \sim \ln \mathrm{E}[Z]$ w.h.p. Then for any event $\mathscr{A}$ we have*

$$\mathrm{P}_{Boltzmann}[\mathscr{A}] \leq \exp(o(N)) \cdot \mathrm{P}_{planted}[\mathscr{A}] + o(1).$$

Observe that Corollary 3.5 implies that the assumption of the quiet planting lemma is valid for α satisfying (20). Generally speaking, the quiet planting lemma allows us to upper-bound the Boltzmann probability of some "bad" event $\mathscr{A}$ by way of the planted model.

As an application of Lemma 4.1, we can now verify the prediction that the set of 2-colorings decomposes into tiny clusters for

$$(1 + \varepsilon_K)2^{K-1} \ln K/K < \alpha < (1 - \varepsilon_K)2^{K-1} \ln 2.$$

Indeed, let us say that *shattering* occurs if the set $\Lambda(\boldsymbol{\Phi})$ of 2-colorings of $\boldsymbol{\Phi}$ has a decomposition $\Lambda(\boldsymbol{\Phi}) = \bigcup_{i=1}^{\Sigma} \Lambda_i$ such that the following two conditions hold for some $\varepsilon_K > 0$.

SH1. For any $1 \leq i \leq \Sigma$ we have $\frac{1}{N} \ln |\Lambda_i| < \frac{1}{N} \ln Z(\boldsymbol{\Phi}) - \varepsilon_K$.
SH2. If $1 \leq i < j \leq \Sigma$, then $\mathrm{dist}(\Lambda_i, \Lambda_j) \geq \varepsilon_K N$.

Condition **SH1** enforces that each cluster only contains an exponentially small fraction of all 2-colorings. Moreover, **SH2** ensures that distinct clusters are separated by a linear Hamming distance.

Theorem 4.2 ([1]) *There exist a number $K_0 > 0$ and a sequence $(\varepsilon_K) \to 0$ such that for $K \geq K_0$ and for*

$$(1 + \varepsilon_K)2^{K-1} \ln K/K \leq \alpha < 2^{K-1} \ln 2 - (1 + \ln 2)/2 - \varepsilon_K$$

shattering occurs.

To prove Theorem 4.2, we combine Lemma 4.1 with a closer analysis of the function $f(\omega)$ from (25). The following is an exercise in calculus.

Lemma 4.3 *Keep the assumptions of Theorem 4.2. There is $\omega_0 \in (0, 1/2)$ such that*

$$f(\omega_0) < 0 \quad \text{and} \quad \sup_{0 \leq \omega \leq \omega_0} f(\omega) < f(1/2).$$

In words, the function f is strictly negative at some point $0 < \omega_0 < 1/2$, and the maximum value of f over the interval $(0, \omega_0)$ is strictly smaller than $f(1/2)$; Fig. 3 provides an illustration of this picture in the case $K = 7$.

Proof of Theorem 4.2 Define the *cluster* of a 2-coloring σ as

$$\mathscr{C}(\sigma) = \{\tau \in \Lambda(\boldsymbol{\Phi}) : \mathrm{dist}(\sigma, \tau) \leq \omega_0 N\}.$$

In terms of the planted model, Eq. (25) shows that $f(\omega)$ is the logarithm of the expected number of 2-colorings at Hamming distance ωn from the planted 2-coloring σ. Hence, Lemma 4.3 implies together with Markov's inequality that there is $\varepsilon_K > 0$ such that in the planted model, the following two conditions hold with probability $\geq 1 - \exp(-\varepsilon_K N)$.

- We have $|\mathscr{C}(\sigma)| \leq \mathrm{E}[Z]\exp(-\varepsilon_K N)$; this follows because $f(1/2) \sim \frac{1}{N}\ln \mathrm{E}[Z]$, cf. (26).
- For all $\tau \in \Lambda(\boldsymbol{\Phi})$ we have $\mathrm{dist}(\sigma, \tau)/N \notin [\omega_0 - \varepsilon_K, \omega_0 + \varepsilon_K]$; this follows because f is continuous.

Let us call σ *good* if the two conditions above hold.

We now construct the cluster decomposition of $\Lambda(\boldsymbol{\Phi})$ as follows (for $\boldsymbol{\Phi}$ a uniformly random hypergraph). Let Λ_0 be the set of all $\sigma \in \Lambda(\boldsymbol{\Phi})$ that are not good. By Lemma 4.1, w.h.p. we have $|\Lambda_0| \leq \exp(-\varepsilon_K N)$ for some $\varepsilon_K > 0$. Now, construct Λ_i for $i \geq 1$ as follows.

- Pick a good $\sigma \in \Lambda(\boldsymbol{\Phi}) \setminus \bigcup_{j<i} \Lambda_i$.
- Let $\Lambda_i = \mathscr{C}(\sigma) \setminus \bigcup_{j<i} \Lambda_i$.

It is easily verified that this decomposition satisfies **SH1**–**SH2**. □

4.2 Condensation

Remember that there is a gap between Corollary 3.5 and the physics prediction as to the s1RSB phase transition. In this section we are going to show how "quiet planting" can be used to remedy this discrepancy.

Theorem 4.4 ([10]) *There is a sequence* $(\varepsilon_K) \to 0$ *such that the following is true.*

1. *If* $\alpha < 2^{K-1}\ln 2 - \ln 2 - \varepsilon_K$, *then* $\ln Z \sim \ln \mathrm{E}[Z]$ *w.h.p.*
2. *If* $2^{K-1}\ln 2 - \ln 2 + \varepsilon_K < \alpha < \alpha_{2-col}$, *then* $\ln Z < \ln \mathrm{E}[Z] - \Omega(N)$ *w.h.p.*

This theorem implies that there occurs a *phase transition* at $\alpha = 2^{K-1}\ln 2 - \ln 2 + \varepsilon_K$. Indeed, (18) shows that for smaller α, the free entropy density is a *linear* function of α. Thus, by the uniqueness of analytic continuations the free entropy density is non-analytic at $\alpha = 2^{K-1}\ln 2 - \ln 2 + \varepsilon_K$.

To prove Theorem 4.4, we need to investigate the existence of *frozen variables*. Based on this concept, we are going to prove the following. Throughout, we are going to assume that $K \geq K_0$ for a certain constant $K_0 > 3$.

Lemma 4.5 *Assume that* $2^{K-1}\ln 2 - 10 \le \alpha \le 2^K \ln 2$. *Let* $(\boldsymbol{\Phi}, \sigma)$ *be chosen from the planted model. Then w.h.p. we have*

$$\frac{1}{N}\ln|\mathscr{C}(\sigma)| = \frac{\ln 2 + o_K(1)}{2^K}. \tag{27}$$

The statement of Lemma 4.5 may seem surprising since the r.h.s. of (27) does not appear to depend on α. Indeed there is a (slight) dependence on α, which is, however, hidden in the $o_K(1)$ term; the reason for this will become apparent in the following paragraphs.

To prove Lemma 4.5, let us call an edge e of the hypergraph $\boldsymbol{\Phi}$ *critical* with respect to σ if it contains $K-1$ vertices that take the same color under σ. If $v \in e$ is the unique vertex with the minority color, we say that *e blocks* v. An elementary calculation shows that in the planted model we expect to have $N \cdot (K + \varepsilon_K)\ln 2$ critical edges. Indeed, the number of edges blocking one particular vertex v is asymptotically Poisson with mean $(K+\varepsilon_K)\ln 2$, and these random variables turn out to be nearly independent. Hence, we expect that all but about $\exp(-K\ln 2) = 2^{-K}$ vertices are blocked.

In fact, for $K \ge K_0$ the number of edges blocking a vertex v is going to be, say, at least 10 for most vertices v. More precisely, let $\hat{\boldsymbol{\Phi}}$ be the largest sub-hypergraph of $\boldsymbol{\Phi}$ in which every vertex v is blocked by at least 10 edges that consist of vertices in $\hat{\boldsymbol{\Phi}}$ only. We call $\hat{\boldsymbol{\Phi}}$ the *core* of $\boldsymbol{\Phi}$. (Algorithmically, the core can be obtained by iteratively removing vertices that violated the aforementioned condition.) Using standard arguments from probabilistic combinatorics, one can show the following.

Fact 4.6 *W.h.p. the core* $\hat{\boldsymbol{\Phi}}$ *contains all but at most* $K^{11}2^{-K}N$ *vertices.*

If we try to recolor one vertex v in the core, we are going to have to recolor at least one other vertex in each of the $\ge$10 edges that block v. Since these vertices are in the core, too, it seems plausible that this is going to trigger an "avalanche" of recolorings. Indeed, the expansion properties of the random hypergraph imply the following. Recall that $\mathscr{C}(\sigma)$ denote the cluster of σ.

Fact 4.7 *Let* $\tau \in \mathscr{C}(\sigma)$. *Then for all* $v \in \hat{\boldsymbol{\Phi}}$ *we have* $\sigma(v) = \tau(v)$.

Although the core dominates the random hypergraph $\boldsymbol{\Phi}$, the number of vertices outside is still a bit bigger than $2^{-K}N$. To enhance the core, we construct the *backbone* of $\boldsymbol{\Phi}$ as follows.

BB1. Initially, let B be the vertex set of the core.
BB2. While there is a vertex $v \notin B$ that has a blocking edge e such that $e \subset B \cup \{v\}$, add v to B.

Once more based on standard arguments, we obtain the following.

Fact 4.8 *W.h.p. the backbone contains all but* $(1 + o_K(1))2^{-K}N$ *vertices.*

By construction, Fact 4.7 implies that for all v in the backbone we have $\sigma(v) = \tau(v)$ for all $\tau \in \mathscr{C}(\sigma)$. Thus, Fact 4.8 yields

$$\frac{1}{N} \ln |\mathscr{C}(\sigma)| \leq \frac{\ln 2 + o_K(1)}{2^K} \quad \text{w.h.p.}$$

Conversely, it turns out that most of the vertices outside the backbone can be colored independently. More specifically, if we reduce the hypergraph by replacing all vertices in the core by constants 0 or 1, then the factor graph of the resulting system is sub-critical. In fact, it mostly consists of isolated vertices. Working this out in detail implies (27).

Let us now call an equitable 2-coloring of $\boldsymbol{\Phi}$ *tame* if (27) holds. Moreover, let $\mathscr{Z}$ be the number of tame 2-colorings of $\boldsymbol{\Phi}$. Together with a double counting argument, Lemma 4.5 implies

$$\mathrm{E}[\mathscr{Z}] \sim \mathrm{E}[Z]. \tag{28}$$

Furthermore, the definition of "tame" implies that if $\alpha < 2^{K-1} \ln 2 - \ln 2 - \varepsilon_K$, then $|\mathscr{C}(\sigma)| \leq \mathrm{E}[Z]$ for all tame σ. As a consequence, we find that

$$\mathrm{E}[\mathscr{Z}^2] \leq C \cdot \mathrm{E}[\mathscr{Z}]^2 \quad \text{for } \alpha < 2^{K-1} \ln 2 - \ln 2 - \varepsilon_K.$$

Thus, the second moment method "works" for $\mathscr{Z}$, which implies the first part of Theorem 4.4.

The second assertion follows from a "converse quiet planting lemma". If $\alpha > 2^{K-1} \ln 2 - \ln 2 + \varepsilon_K$, then by similar reasoning as above we find that in the planted model w.h.p.

$$\frac{1}{N} \ln |\mathscr{C}(\sigma)| > \frac{1}{N} \ln \mathrm{E}[Z] + \varepsilon_K$$

for some $\varepsilon_K > 0$. This estimate can be shown to be irreconcilable with the notion that $\ln Z \sim \ln \mathrm{E}Z$ w.h.p.

Remark 4.9 The above construction of the core and the backbone also implies the existence of multiple Belief Propagation (near-)fixed points. Namely, if we start Belief Propagation from a 2-coloring σ as above, the messages of the vertices in the backbone are going to remain one-point distributions (concentrated on the color that the vertices are frozen to).

4.3 The Asymptotic 2-Colorability Threshold

Up to the ε_K error term, the following result matches the prediction of the 1-step replica symmetry breaking ansatz as to the 2-colorability transition.

Theorem 4.10 ([7]) *We have* $\alpha_{2-col} = 2^{K-1}\ln 2 - \left(\frac{1}{4} + \frac{\ln 2}{2}\right) + \varepsilon_K$.

The proof of Theorem 4.10 follows the prescription of the 1RSB cavity method. That is, we are going to mimic the effect of working with a Parisi parameter $y < 1$. In the terminology of the previous section, this means that we should work with solutions that have a larger number of frozen variables than occur in the planted model.

Capturing the notion of "frozen" accurately is technically difficult, e.g., because we are working with hypergraphs with an irregular degree sequence. As a proxy, we are going to work with a prescribed number of *blocked* variables. Thus, let Z_γ be the number of equitable 2-colorings in which $(1-\gamma)N$ variables are blocked. The idea is to carry out a second moment argument for Z_γ.

However, it turns out that $\mathrm{E}[Z_\gamma^2] \geq \exp(\delta N)\mathrm{E}\left[Z_\gamma\right]^2$ for some $\delta = \delta(K) > 0$. The reason for this is fluctuations of the degree sequence: roughly speaking, a graph with a "more regular" degree sequence is more likely to have a large number of blocked variables. Hence the moments of Z_γ are driven up by a small number of "pathological" degree sequences that are far more regular than the (Poisson) degree sequence of the random hypergraph $\boldsymbol{\Phi}$.

To get around this problem, let $\boldsymbol{d} = (d_v)_{v\in V}$ denote the degree sequence of the random hypergraph $\boldsymbol{\Phi}$. That is, d_v is the number of edges vertex v appears in. We are going to fix one particular degree sequence $\boldsymbol{d}$ and work with the random hypergraph $\boldsymbol{\Phi}_{\boldsymbol{d}}$ with that degree sequence.

In this setup, even the computation of the first moment is non-trivial. The key element is the probability that a given number of critical edges block the correct number of vertices. Hence, if we think of each vertex as a "bin" of capacity d_v and of each critical edge as a ball that is tossed randomly into a bin, we need to calculate the probability of obtaining precisely γN empty bins. That is, we need to solve a large deviations problem for an occupancy problem. Furthermore, we also need to optimise the number of critical clauses. Of course, each of these comes with a "price", i.e., we need to take into account the large deviations principle of this random variable. In the case of a regular degree sequence (i.e., $d_v = d_{v'}$ for all v), it is not very difficult to solve this problem precisely, but in the case of general degree distributions $\boldsymbol{d}$ an accurate analysis is far from trivial.

In any case, we can identify a function $g(\gamma)$ such that for a random $\boldsymbol{d}$ w.h.p. we have

$$\frac{1}{N}\ln \mathrm{E}[Z_\gamma] = g(\gamma) + \varepsilon_K 3.99^{-K}.$$

Now, by a similar token as in the previous section, if the number of blocked (or frozen) variables is $(1-\gamma)N$, we expect that the cluster size satisfies

$$\frac{1}{N}\ln|\mathscr{C}(\sigma)| = \gamma\ln 2 + \varepsilon_K 3.99^{-K}.$$

Thus, a *necessary* condition for the success of the second moment argument is that (up to the error term)

$$\gamma \ln 2 < g(\gamma). \tag{29}$$

This necessary condition turns out to be (essentially) sufficient. That is, neglecting some technical details, it is possible to carry out a second moment argument for (essentially) $Z_\gamma(\boldsymbol{\Phi}_d)$ under the assumption that (29) is satisfied.

This fact allows us to also estimate the free entropy. Namely, for a given α let $\gamma(\alpha)$ be the largest $\gamma > 2^{-K}$ for which (29) is satisfied. Then the free entropy comes to

$$\frac{1}{N}\mathrm{E}[\ln 1 \vee Z] = (1 + \varepsilon_K)g(\gamma).$$

5 Conclusions and Outlook

We have seen that the physicists' non-rigorous cavity method leads to remarkable predictions as to the properties and phase transitions of random discrete structures such as random graphs, hypergraphs or formulas. As was illustrated with the example of the random hypergraph 2-coloring problem, these predictions can be harnessed to obtain a more complete rigorous picture. Yet we are far from having a complete rigorous version of the cavity method, and the rigorous picture still has significant gaps. These gaps arise in part from the difficulty of getting a handle on the (distributional) fixed point problems that describe the results of the cavity formalism.

The techniques that we presented (mostly) in the context of the hypergraph 2-coloring problem extend to a wide range of other cases. For instance, the replica symmetric lower bound and the 1RSB upper bound on the K-colorability threshold in random graphs have been turned into rigorous results both in the case of the Erdős-Rényi random graph and in the case of random regular graphs [9, 11]. In addition, the 1RSB prediction on the random K-SAT threshold has been verified asymptotically [6, 8].

Indeed, very recently there have been the first full rigorous verifications of 1RSB predictions in random regular K-SAT and random regular K-NAESAT as well as the independent set problem (or "hardcore model") on random regular graphs [6, 13, 14]. These results are enabled by the fact that in these regular cases, the distributional fixed point problems that come out of the 1RSB ansatz simplify to fixed point problems over the real numbers.

While the aforementioned results are all based on various instalments of the second moment method, a further powerful rigorous proof method is the *interpolation method* [4, 16, 23]. This method has been used to prove the existence of the free entropy density in certain problems or to establish upper bounds on it. The

interpolation method also played a significant role in the context of the Sherrington-Kirkpatrick model [25].

Finally, the cavity method has inspired novel "message passing algorithms" that have been received with much excitement by the computer science community [22]. Experiments suggest that these algorithms perform extraordinarily well on random K-SAT instances that are well beyond the reach of "traditional" satisfiability algorithms [19]. Achieving a rigorous understanding of the performance of these algorithms is an important challenge.

Acknowledgements I am grateful to Victor Bapst for his comments on a draft version of these notes.

References

1. D. Achlioptas, A. Coja-Oghlan, Algorithmic barriers from phase transitions, in *Proceedings of 49th FOCS* (2008), pp. 793–802
2. D. Achlioptas, C. Moore, On the 2-colorability of random hypergraphs, in *Proceedings of 6th RANDOM* (2002), pp. 78–90
3. N. Alon, J. Spencer, *The Probabilistic Method*, 2nd edn. or later (Wiley, New York, 2004)
4. M. Bayati, D. Gamarnik, P. Tetali, Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. Ann. Probab. **41**, 4080–4115 (2013)
5. B. Bollobás, *Random Graphs*, 2nd edn. (Cambridge University Press, Cambridge, 2001)
6. A. Coja-Oghlan, The asymptotic k-SAT threshold (2013). arXiv:1310.2728
7. A. Coja-Oghlan, K. Panagiotou, Catching the k-NAESAT threshold, in *Proceedings of 44th STOC* (2012), pp. 899–908
8. A. Coja-Oghlan, K. Panagiotou, Going after the k-SAT threshold, in *Proceedings of 45th STOC* (2013), pp. 705–714
9. A. Coja-Oghlan, D. Vilenchik, Chasing the k-colorability threshold, in *Proceedings of 54th FOCS* (2013), pp. 380–389
10. A. Coja-Oghlan, L. Zdeborová, The condensation transition in random hypergraph 2-coloring, in *Proceedings of 23rd SODA* (2012), pp. 241–250
11. A. Coja-Oghlan, S. Hetterich, C. Efthymiou, On the chromatic number of random regular graphs (2013). arXiv:1308.4287
12. A. Dembo, A. Montanari, N. Sun, Factor models on locally tree-like graphs. Ann. Probab. **41**, 4162–4213 (2013)
13. J. Ding, A. Sly, N. Sun, Maximum independent sets on random regular graphs (2013). arXiv:1310.4787
14. J. Ding, A. Sly, N. Sun, Satisfiability threshold for random regular NAE-SAT (2013). arXiv:1310.4784
15. P. Erdős, A. Rényi, On the evolution of random graphs. Magayar Tud. Akad. Mat. Kutato Int. Kozl. **5**, 17–61 (1960)
16. S. Franz, M. Leone, Replica bounds for optimization problems and diluted spin systems. J. Stat. Phys. **111**, 535–564 (2003)
17. E. Friedgut, Sharp thresholds of graph properties, and the k-SAT problem. J. Am. Math. Soc. **12**, 1017–1054 (1999)
18. S. Janson, T. Łuczak, A. Ruciński, *Random Graphs* (Wiley, New York, 2000)
19. L. Kroc, A. Sabharwal, B. Selman, Message-passing and local heuristics as decimation strategies for satisfiability, in *Proceedings of 24th SAC* (2009), pp. 1408–1414

20. F. Krzakala, L. Zdeborová, Hiding quiet solutions in random constraint satisfaction problems. Phys. Rev. Lett. **102**, 238701 (2009)
21. M. Mézard, A. Montanari, *Information, Physics and Computation* (Oxford University Press, Oxford, 2009)
22. M. Mézard, G. Parisi, R. Zecchina, Analytic and algorithmic solution of random satisfiability problems. Science **297**, 812–815 (2002)
23. D. Panchenko, M. Talagrand, Bounds for diluted mean-fields spin glass models. Probab. Theory Relat. Fields **130**, 319–336 (2004)
24. J. Schmidt-Pruzan, E. Shamir, Component structure in the evolution of random hypergraphs. Combinatorica **5**, 81–94 (1985)
25. M. Talagrand, The Parisi formula. Ann. Math. **163**, 221–263 (2006)

Multidimensional Random Polymers: A Renewal Approach

Dmitry Ioffe

1 Introduction

Mathematical and probabilistic developments presented here draw inspiration from statistical mechanics of stretched polymers, see for instance [4, 25].

Polymers chains to be discussed in these lecture notes are modeled by paths of finite range random walks on $\mathbb{Z}^d$. We shall always assume that the underlying random walk distribution has zero mean. The word *stretched* alludes to the situation when the end-point of a polymer is pulled by an external force, or in the random walk terminology, by a drift. In the case of random walks this leads to a ballistic behaviour with limiting spatial extension described in terms of the usual law of large numbers (LLN) for independent sums. Central limit theorem (CLT) and large deviations (LD) also hold.

Polymer measures below are non-Markovian objects (see Remark 1.1), which gives rise to a rich morphology. We shall distinguish between ballistic and sub-ballistic phases and between quenched and annealed polymers. Quenched polymers correspond to pulled random walks in random potentials. Their annealed counterparts correspond to pulled random walks in a deterministic attractive self-interaction potentials.

Two main themes are the impact of the drift, that is when the model in question, annealed or quenched, becomes ballistic, and the impact of disorder, that is whether or not quenched and annealed models behave similarly.

It is instructive to compare models of stretched polymers with those of directed polymers [3]. In the latter case sub-ballistic to ballistic transition is not an issue. Furthermore, in the stretched case polymers can bend and return to the same vertices, which makes even the annealed model to be highly non-trivial (in the

D. Ioffe (✉)
Faculty of IE& M, Technion, Haifa 32000, Israel
e-mail: ieioffe@ie.technion.ac.il

M. Biskup et al. (eds.), *Random Walks, Random Fields, and Disordered Systems*,
Lecture Notes in Mathematics 2144, DOI 10.1007/978-3-319-19339-7_4

directed case the annealed model is a usual random walk). On the other hand it is unlikely that a study of stretched polymers will shed light on questions which are open in the directed context.

1.1 Class of Models

Underlying Random Walk Consider random walk on $\mathbb{Z}^d$ with an irreducible finite range step distribution. We use the notation P_d both for the random walk path measure and for the distribution of individual steps. For convenience we shall assume that nearest neighbour steps $\pm\mathsf{e}_k$ are permitted,

$$\mathsf{P}_d(\pm\mathsf{e}_k) > 0 \tag{1}$$

The size of the range is denoted R: $\mathsf{P}_d(\mathsf{X} = \mathsf{x}) > 0 \Rightarrow |\mathsf{x}| \leq R$. Without loss of generality we shall assume that $\mathsf{E}_d\mathsf{X} = 0$.

Random Environment The random environment is modeled by an i.i.d. collection $\{\mathsf{V}_\mathsf{x}^\omega\}_{\mathsf{x}\in\mathbb{Z}^d}$ of *non-negative* random variables. The notation $\mathscr{Q}$ and $\mathscr{E}$ are reserved for the corresponding product probability measure and the corresponding expectation. We shall assume:

(A1) V^ω is non-trivial and $0 \in \mathrm{supp}\,(\mathsf{V}^\omega)$.

(A2) $\mathscr{Q}(\mathsf{V}^\omega < \infty) > p_c(\mathbb{Z}^d)$, where p_c is the critical Bernoulli site percolation probability.

Polymers and Polymer Weights Polymers $\gamma = (\gamma_0, \ldots, \gamma_n)$ are paths of the underlying random walk. For each polymer γ we define $|\gamma| = n$ as the number of steps, and $\mathsf{X}(\gamma) = \gamma_n - \gamma_0$ as the displacement along the polymer.

The are two type of weights we associate with polymers: quenched random weights

$$\mathsf{W}_d^\omega(\gamma) = \exp\left\{-\beta \sum_{i=1}^{|\gamma|} \mathsf{V}_{\gamma_i}^\omega\right\} \mathsf{P}_d(\gamma), \tag{2}$$

and annealed weights

$$\mathsf{W}_d(\gamma) = \mathscr{E}\left(\mathsf{W}_d^\omega(\gamma)\right) = \mathrm{e}^{-\Phi_\beta(\gamma)}\,\mathsf{P}_d(\gamma), \tag{3}$$

where the self-interacting potential

$$\Phi_\beta(\gamma) = \sum_{\mathsf{x}} \phi_\beta(\ell_\gamma(\mathsf{x})). \tag{4}$$

Above $\ell_\gamma(\mathsf{x})$ is the local time of γ at x;

$$\ell_\gamma(x) = \sum_{i=1}^{n} \mathbb{1}_{\{\gamma_i = \mathsf{x}\}}, \tag{5}$$

and ϕ_β is given by:

$$\phi_\beta(\ell) = -\log \mathscr{E}\left(\mathrm{e}^{-\beta \ell\, \mathsf{V}^\omega}\right). \tag{6}$$

The inverse temperature $\beta > 0$ modulates the strength of disorder.

Pulling Force, Partition Functions and Probability Distributions For $h \in \mathbb{R}^d$ we shall consider quenched and annealed partition functions

$$Z_n^\omega(h) = \sum_{|\gamma|=n} \mathrm{e}^{h\cdot \mathsf{X}(\gamma)} \mathsf{W}_d^\omega(\gamma) \quad \text{and} \quad Z_n(h) = \mathscr{E}(Z_n^\omega(h)) = \sum_{|\gamma|=n} \mathrm{e}^{h\cdot \mathsf{X}(\gamma)} \mathsf{W}_d(\gamma), \tag{7}$$

and the corresponding probability distributions,

$$\mathbb{P}_n^{h,\omega}(\gamma) = \frac{1}{Z_n^\omega(h)} \mathrm{e}^{h\cdot \mathsf{X}(\gamma)} \mathsf{W}_d^\omega(\gamma) \quad \text{and} \quad \mathbb{P}_n^h(\gamma) = \frac{1}{Z_n(h)} \mathrm{e}^{h\cdot \mathsf{X}(\gamma)} \mathsf{W}_d(\gamma). \tag{8}$$

Remark 1.1 Annealed measures $\mathbb{P}_n^h$ are non-Markovian. Quenched measures $\mathbb{P}_n^{h,\omega}$ are also non-Markovian in the sense that in general $\mathbb{P}_n^{h,\omega}$ is not a marginal of $\mathbb{P}_m^{h,\omega}$ for $m > n$.

1.2 Morphology

We shall distinguish between ballistic and sub-ballistic behaviour of quenched and annealed polymers (8) and between strong an weak impact of disorder on the properties of quenched polymers (as compared to the annealed ones).

Ballistic Phase For the purpose of these lecture notes, let us say that a self-interacting random walk (or polymer) is ballistic if there exists $\delta > 0$ and a vector $\mathsf{v} \neq 0$ such that

$$\lim_{n\to\infty} \mathbb{P}_n^h(|\mathsf{X}| \le \delta n) = 0 \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{n} \mathbb{E}_n^h \mathsf{X}(\gamma) = \mathsf{v}. \tag{9}$$

The model is said to be sub-ballistic, if

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}_n^h |\mathsf{X}(\gamma)| = 0. \tag{10}$$

At this stage it is unclear whether there are models which comply neither with (9) nor with (10). It is the content of Theorem 2.1 below that for the annealed models the above dichotomy always holds.

Similarly, the quenched model is said to be in the ballistic, respectively sub-ballistic, phase if $\mathcal{Q}$-a.s

$$\lim_{n\to\infty} \mathbb{P}_n^{h,\omega}\left(|\mathsf{X}| \le \delta n\right) = 0 \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{n}\mathbb{E}_n^{h,\omega}\,\mathsf{X}(\gamma) = \mathsf{v}, \tag{11}$$

and, respectively,

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}_n^{h,\omega}\,|\mathsf{X}(\gamma)| = 0. \tag{12}$$

For quenched models it is in general open question whether the limiting spatial extension [second limit in (11)] always exists. See Theorem 2.7 below for a precise statement.

Ballistic and sub-ballistic phases correspond to very different patterns of behaviour. We focus here on the ballistic phase. In a sense sub-ballistic behaviour is more intricate than the ballistic one. In the continuous context (Brownian motion) results about sub-ballistic phase are summarized in [29]. The theory (so called *enlargement of obstacles*) was adjusted to random walks on $\mathbb{Z}^d$ in [1].

Strength of Disorder For each value of the pulling force h and the interaction β strength of disorder may be quantified on several levels:

L1. $\mathcal{Q}$-a.s $\limsup_{n\to\infty} \frac{1}{n}\log\frac{Z_n^\omega(h)}{Z_n(h)} < 0$.

Since by Assumption **(A1)** the annealed potential ϕ_β in (6) satisfies $\lim_{\beta\to\infty}\frac{\phi_\beta}{\beta} = 0$, it is not difficult to see that, at least in the case when $\{\mathsf{x} : \mathsf{V}_\mathsf{x}^\omega = 0\}$ does not percolate, **L1.** holds in any dimension whenever the strength of interaction β and the pulling force h are large enough.

Furthermore, as we shall see in Sect. 5 (and as it was originally proved in [33]) the disorder is strong in the sense of **L1.** in lower dimensions $d = 2, 3$ for any $\beta > 0$ provided that the annealed polymer is in (the interior of) the ballistic phase.

L2. $\mathcal{Q}$-a.s $\lim_{n\to\infty}\frac{Z_n^\omega(h)}{Z_n(h)} = 0$.

This is presumably always the case when the quenched model is in sub-ballistic phase. The case $h = 0$ is worked out in great detail [1, 29].

Remark 1.2 A characterization of annealed and quenched sets of sub-critical drifts; $\mathbf{K}_0$ and $\mathbf{K}_0^{\mathsf{q}}$ is given in (20) and (58) below. It always holds that $\mathbf{K}_0 \subseteq \mathbf{K}_0^{\mathsf{q}}$. The inclusion is strict in any dimension for β large enough, and it is presumably strict in dimensions $d = 2, 3$ for any $\beta > 0$. On the other hand, it seems to be an open question whether in higher dimensions $d > 3$ the two sets of sub-critical drifts coincide at sufficiently small β.

L3. Typical polymers under $\mathbb{P}_n^h$ and $\mathbb{P}_n^{h,\omega}$ have very different properties.

Ballistic phase of annealed polymers in dimensions $d \geq 2$ is by now completely understood, and we expose the core of the corresponding (Ornstein-Zernike) theory developed in [14, 17] in Sect. 3. In dimensions $d \geq 4$, the disorder happens to be weak in the sense of any of **L1–L3** in the following regime (which we shall call *very weak disorder*): Fix $h \neq 0$ and then take $\mathscr{Q}(\mathsf{V}^\omega = \infty)$ and β in (2) to be sufficiently small. These results [11, 15, 18, 32] are explained in Sect. 4.

Zero Drift Case At $h = 0$ properties of both annealed and quenched measures were described in depth in [29] and references therein, following an earlier analysis of Wiener sausage in [8, 9]. This is *not* the case we consider here. However, it is instructive to keep in mind what happens if there is no pulling force, and, accordingly, we give a brief heuristic sketch. To fix ideas consider the case of pure traps $p = \mathscr{Q}(\mathsf{V}^\omega = 0) = 1 - \mathscr{Q}(\mathsf{V}^\omega = \infty)$. If $1 - p$ is small, then $\{\mathsf{x} : \mathsf{V}_\mathsf{x}^\omega = 0\}$ percolates, and the model is non-trivial. Let us start with a quenched case. Let B_r be a lattice box $B_r = \{\mathsf{x} : |\mathsf{x}|_1 \leq r\}$ and $B_r(\mathsf{x}) = \mathsf{x} + B_r$. We say that there is an (R, r)-clearing if

$$\exists\, \mathsf{x} \in B_R \text{ such that } \mathsf{V}_\mathsf{y}^\omega = 0 \text{ for all } \mathsf{y} \in B_r(\mathsf{x}).$$

The probability

$$\mathscr{Q}(\text{there is a } (R,r) \text{ clearing}) \approx 1 - \left(1 - p^{c_1 r^d}\right)^{c_2 R^d / r^d} \approx 1 - \mathrm{e}^{-c_3 p^{c_1 r^d} \frac{R^d}{r^d}}.$$

Up to leading terms this is non-negligible if $p^{c_1 r^d} R^d \approx \text{const}$, or if $r \approx (\log R)^{1/d}$. On the other hand, a probability that a random walk will go ballistically to a box (clearing) $B_r(\mathsf{x})$ at distance of order R from the origin is of order $\mathrm{e}^{-c_4 R}$, and the probability that afterwards it will spend around n units of time in $B_r(\mathsf{x})$ is $\approx \mathrm{e}^{-c_5 n / r^2}$. We, therefore need to find an optimal balance between R and $n/r^2 \approx n/(\log R)^{2/d}$ terms, which gives, again up to leading terms, $R \approx n/(\log n)^{2/d}$. This suggests both a survival pattern for typical quenched polymer (see Fig. 1), and an asymptotic relation for the quenched partition function

$$\log Z_n^\omega \approx -\frac{n}{(\log n)^{2/d}}. \tag{13}$$

As far as the annealed model is considered for $\ell \geq 1$ define as before $\phi_\beta(\ell) = -\log \mathscr{E}\left(\mathrm{e}^{-\beta \ell \mathsf{V}_0}\right) = -\log p \stackrel{\Delta}{=} \nu$. Consider random walk which stays all n units of time inside B_R. The probabilistic price for the latter is $\approx \mathrm{e}^{-c_6 n / R^2}$. On the other hand, the self-interaction price is $\approx \mathrm{e}^{-c_7 \nu R^d}$. Choosing optimal balance leads to $R \approx n^{1/(d+2)}$. This suggests a behaviour pattern for typical annealed polymers (see Fig. 1), which is very different from the survival pattern for typical quenched

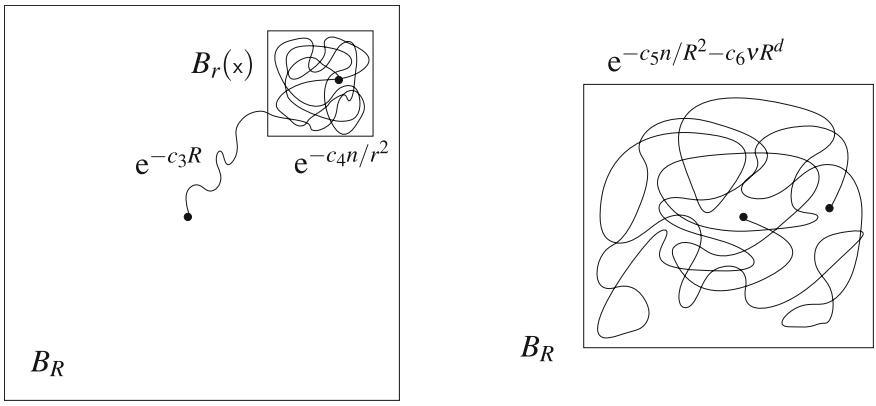

Fig. 1 *On the left*: A survival pattern for an n-step quenched polymer with $R \approx n/(\log n)^{2/d}$. *On the right*: n-step annealed polymer in B_R with $R \approx n^{1/(d+2)}$

polymer as discussed above. This also suggests the following asymptotics for the annealed partition function:

$$\log Z_n \approx -n^{d/(d+2)}. \tag{14}$$

The above discussion indicates that in the zero drift case the disorder is strong on levels **L2, L3**, but not on **L1.**

Outline of the Notes We do not attempt to give a comprehensive survey of the existing results on the subject. Neither the notes are self-contained, in many instances below we shall refer to the literature for more details on the corresponding proofs. The emphasis is on the exposition of the renewal structure behind stretched polymers in the ballistic regime, and how this might help to explore and understand various phenomena in question.

Section 2 is devoted to the thermodynamics of annealed and quenched polymers, namely to the facts which can be deduced from sub-additivity arguments and large deviation principles.

Multidimensional renewal theory (under assumption of exponential tails) is discussed in detail in Sect. 3. In Sect. 3.2 we explain renormalization procedures which lead to a reformulation of annealed models in this renewal context, which is the core of the Ornstein-Zernike theory of the latter. As a result we derive very sharp and essentially complete description of the ballistic phase of the annealed polymers on the level of invariance principles and local limit asymptotics on all deviation scales.

In Sect. 4 we explain why the annealed renewal structure persists for quenched models in the regime of very weak disorder in dimensions $d \geq 2$. More precisely, it happens that in the latter case the disorder is weak on all three levels **L1–L3**.

In Sect. 5 we explain how to check that the disorder is always strong already on level **L1** in dimensions $d = 2, 3$. A more or less complete argument is given only in two dimensions.

To facilitate references and the reading some of the back ground material on convex geometry and large deviations is collected in the Appendix.

Notation Conventions Values of finite positive constants $c, \nu, c_1, \nu_1, c_2, \nu_2, \ldots$ may change between different sections.

In the sequel we shall use the following notation for asymptotic relations: Given a set of indices $\mathscr{A}$ and two positive sequences $\{a_\alpha, b_\alpha\}_{\alpha\in\mathscr{A}}$, we say that

- $a_\alpha \overset{<}{\sim} b_\alpha$ *uniformly in* $\alpha \in \mathscr{A}$ if there exists a constant $c > 0$ such that $a_\alpha \le c b_\alpha$ for all $\alpha \in \mathscr{A}$.
- We shall use $a_\alpha \cong b_\alpha$ if both $a_\alpha \overset{<}{\sim} b_\alpha$ and $a_\alpha \overset{>}{\sim} b_\alpha$ hold.

For $1 \le p \le \infty$, the ℓ_p-norms on $\mathbb{R}^d$ are denoted

$$|x|_p = \Big(\sum_{i=1}^{d} |x_i|^p\Big)^{\frac{1}{p}}.$$

The default notation is for the Euclidean norm $|\cdot| = |\cdot|_2$.

If not explicitly stated otherwise paths $\gamma = (\gamma_0, \ldots, \gamma_n)$ are assumed to have their starting point at the origin; $\gamma_0 = 0$. A concatenation $\gamma \circ \eta$ of two paths $\gamma = (\gamma_0, \ldots, \gamma_n)$ and $\eta = (\eta_0, \ldots, \eta_m)$ is the path

$$\gamma \circ \eta = (\gamma_0, \ldots, \gamma_n, \gamma_n + \eta_1, \ldots, \gamma_n + \eta_m)\,.$$

A union of two paths $\gamma \cup \eta$, with end-points at the origin or not, is a subset of $\mathbb{Z}^d$ with multiplicities counted. In particular, local times satisfy $\ell_{\gamma\cup\eta}(\mathsf{x}) = \ell_\gamma(\mathsf{x}) + \ell_\eta(\mathsf{x})$.

2 Thermodynamics of Annealed and Quenched Models

In the sequel we shall employ the following notation for families of polymers:

$$\mathscr{P}_{\mathsf{x}} = \{\gamma \,:\, \mathsf{X}(\gamma) = \mathsf{x}\} \quad \mathscr{P}_n = \{\gamma \,:\, |\gamma| = n\} \quad \text{and} \quad \mathscr{P}_{\mathsf{x},n} = \mathscr{P}_{\mathsf{x}} \cap \mathscr{P}_n \tag{15}$$

Conjugate Ensembles Let $\lambda \ge 0$. Consider

$$G^\omega_\lambda(\mathsf{x}) = \sum_{\mathsf{X}(\gamma)=\mathsf{x}} \mathrm{e}^{-\lambda|\gamma|}\mathsf{W}^\omega_d(\gamma) \quad \text{and} \quad G_\lambda(\mathsf{x}) = \mathscr{E}\left(G^\omega_\lambda(\mathsf{x})\right) = \sum_{\mathsf{X}(\gamma)=\mathsf{x}} \mathrm{e}^{-\lambda|\gamma|}\mathsf{W}_d(\gamma). \tag{16}$$

Free Energy and Inverse Correlation Length One would like to define quenched and annealed free energies via:

$$\lambda^{\mathsf{q}}(h) = \lim_{n\to\infty} \frac{1}{n} \log Z_n^{\omega}(h) \quad \text{and} \quad \lambda(h) = \lim_{n\to\infty} \frac{1}{n} \log Z_n(h). \tag{17}$$

Similarly one would like to define and the inverse correlation lengths: For $\mathsf{x} \in \mathbb{R}^d$ set

$$\tau_\lambda^{\mathsf{q}}(\mathsf{x}) = -\lim_{r\to\infty} \frac{1}{r} \log G_\lambda^{\omega}(\lfloor r\mathsf{x} \rfloor) \quad \text{and} \quad \tau_\lambda(\mathsf{x}) = -\lim_{r\to\infty} \frac{1}{r} \log G_\lambda(\lfloor r\mathsf{x} \rfloor). \tag{18}$$

Depending on the context other names for $\lambda(h)$ are *connectivity constant* and *log-moment generating function*, and for $\tau_\lambda(\mathsf{x})$ are *Lyapunov exponent* and, for some models in two dimensions, *surface tension*.

By *Thermodynamics* we mean here statements about existence of limits in (17) and (18), and their relation to *Large Deviation* asymptotics under quenched and annealed polymer measures (8). Facts about Thermodynamics of annealed and quenched models are collected in Theorems 2.1 and 2.7 below, and, accordingly, discussed in some detail in Sects. 2.1 and 2.2.

2.1 Annealed Models in Dimensions $d \geq 2$

Theorem 2.1

A. *The free energy λ is well defined, non-negative and convex on $\mathbb{R}^d$. Furthermore,*

$$0 = \min_h \lambda(h) = \lambda(0). \tag{19}$$

The set

$$\mathbf{K}_0 \stackrel{\Delta}{=} \{h \,:\, \lambda(h) = 0\} \tag{20}$$

is a compact convex set with a non-empty interior.

B. *The inverse correlation length τ_λ is well defined for any $\lambda \geq 0$, and it can be identified as the support function of the compact convex set*

$$\mathbf{K}_\lambda \stackrel{\Delta}{=} \{h \,:\, \lambda(h) \leq \lambda\}. \tag{21}$$

Define

$$I(\mathsf{v}) = \sup_h \{h \cdot \mathsf{v} - \lambda(h)\} = \sup_\lambda \{\tau_\lambda(\mathsf{v}) - \lambda\}. \tag{22}$$

C. *For any $h \in \mathbb{R}^d$ the family of polymer measures $\mathbb{P}_n^h$ satisfies LD principle with the rate function*

$$I_h(\mathsf{v}) \stackrel{\Delta}{=} \sup_f \{f \cdot \mathsf{v} - (\lambda(f+h) - \lambda(h))\} = I(\mathsf{v}) - (h \cdot \mathsf{v} - \lambda(h))\,. \tag{23}$$

D. *For $h \in \mathrm{int}\,(\mathbf{K}_0)$ the model is sub-ballistic, whereas for any $h \notin \mathbf{K}_0$ the model is ballistic.*

E. *Furthermore, at critical drifts $h \in \partial\mathbf{K}_0$ the model is still ballistic. In other words, the ballistic to sub-ballistic transition is always of the first order in dimensions $d \geq 2$.*

Proofs of Parts A–C and of Part D for sub-critical drifts ($h \in \mathrm{int}\,(\mathbf{K}_0)$) of Theorem 2.1 are based on sub-additivity arguments. Parts D (namely existence of limiting spatial extension v in (9) for super-critical drifts $h \notin \mathbf{K}_0$) and E require a more refined multidimensional renewal analysis based on Ornstein-Zernike theory.

Sub-additivity The following result is due to Hammersley [13]:

Proposition 2.2 *Let $\{a_n\}$ and $\{b_n\}$ be two sequences such that:*

(a) For all m, n, $a_{n+m} \leq a_n + a_m + b_{n+m}$.
(b) The sequence b_n is non-decreasing and

$$\sum_n \frac{b_n}{n(n+1)} < \infty. \tag{24}$$

Then, there exists the limit

$$\xi \stackrel{\Delta}{=} \lim_{n\to\infty} \frac{a_n}{n} \quad \text{and} \quad \frac{a_n}{n} \geq \xi + \frac{b_n}{n} - 4\sum_{k=2n}^{\infty} \frac{b_k}{k(k+1)} \tag{25}$$

Note that $\xi \stackrel{\Delta}{=} \liminf_{n\to\infty} \frac{a_n}{n}$ is always defined. The usual sub-additivity statement is for $b_n \equiv 0$. In this case $\xi = \inf_n \frac{a_n}{n}$. The proof (in the case $b_n \equiv 0$) is straightforward: Indeed, iterating on the sub-additive property, we infer that for any $n \geq m$ and any k (and $N = kn + m$),

$$\frac{a_N}{N} \leq \frac{ka_n}{kn+m} + \frac{a_m}{kn+m}$$

Therefore, for any n fixed

$$\limsup_{N\to\infty} \frac{a_N}{N} \leq \frac{a_n}{n}.$$

Remark 2.3 Note that (25) implies that $\xi < \infty$, however the case $\xi = -\infty$ is not excluded by the argument. Note also that even if $\xi > -\infty$, no upper bounds (apart from $\lim \frac{a_n}{n} = \xi$) on $\frac{a_n}{n}$ are claimed. In other words the sub-additivity argument above does not give information on the speed of convergence.

Attractivity of the Interaction The interaction ϕ_β in (6) is attractive in the sense that

$$\phi_\beta(\ell + m) \leq \phi_\beta(\ell) + \phi_\beta(m). \tag{26}$$

Indeed, (26) follows from positive association of probability measures on $\mathbb{R}$. Namely, if $\mathsf{X} \in \mathbb{R}$ is a random variable, and f, g two bounded functions on $\mathbb{R}$ which are either both non-increasing or both non-decreasing, then

$$\mathbb{E} f(\mathsf{X}) g(\mathsf{X}) \geq \mathbb{E} f(\mathsf{X}) \mathbb{E} g(\mathsf{X}). \tag{27}$$

The universal validity of the latter inequality is related to total ordering of $\mathbb{R}$. If Y is an i.i.d. copy of X, then

$$(f(\mathsf{X}) - f(\mathsf{Y}))\,(g(\mathsf{X}) - g(\mathsf{Y})) \geq 0.$$

Taking expectation we deduce (27).

Exercise 2.4 Show that if ϕ is attractive, then that for any $h \in \mathbb{R}^d$ and for any λ and $\mathsf{x}, \mathsf{y} \in \mathbb{Z}^d$

$$Z_{n+m}(h) \geq Z_n(h) Z_m(h). \tag{28}$$

Note that due to a possible over-counting such line of reasoning does not imply that $G_\lambda(\mathsf{x} + \mathsf{y}) \geq G_\lambda(\mathsf{x}) G_\lambda(\mathsf{y})$.

Part A of Theorem 2.1 Since the underlying random walk has finite range R, $Z_n(h) \leq \mathrm{e}^{Rn|h|}$. On the other hand, since we assumed that $\mathsf{P}_d(\mathsf{e}_1) > 1$,

$$Z_n(h) \geq \left(\mathrm{e}^{h \cdot \mathsf{e}_1 - \phi_\beta(1)} \mathsf{P}_d(\mathsf{e}_1 > 1)\right)^n.$$

Consequently, $\lambda_n(h) \stackrel{\Delta}{=} \frac{1}{n} \log Z_n(\cdot)$ is a sequence of convex (by Hölder's inequality) locally uniformly bounded functions. By Jensen's inequality $\lambda_n(0) = \min \lambda_n$. For each h fixed, $\log Z_n(h)$ is, by (28) super-additive in n. By Proposition 2.2, $\lambda(h) = \lim_{n \to \infty} \lambda_n(h)$ exists, and, by the above, convex and finite on $\mathbb{R}^d$.

Let us check that $\lambda(0) = \min_h \lambda(h) = 0$. Assumption **(A1)** implies that

$$\phi_\beta \text{ is monotone non-decreasing}, \phi_\beta(1) = \min_\ell \phi_\beta(\ell) > 0, \tag{29}$$

and that

$$\lim_{\ell\to\infty} \frac{\phi_\beta(\ell)}{\ell} = 0. \tag{30}$$

Next we rely on the following well-known estimate for the underlying finite range random walk:

Estimate 1 There exists $c = c(\mathsf{P}_d) < \infty$ and $L_0 < \infty$, such that for any $L > L_0$,

$$\mathsf{P}_d\left(\max_{\ell\le n} |\mathsf{X}(\ell)| \le L\right) \gtrsim \mathrm{e}^{-c\frac{n}{L^2}}. \tag{31}$$

uniformly in $n \in \mathbb{N}$.

Let $\Lambda_L = \{\mathsf{x} : |\mathsf{x}|_1 \le L\}$. Then,

$$Z_n(h) \ge \mathrm{e}^{-|h|L} \sum_{\gamma\subset B_L} \mathsf{W}_d(\gamma)\mathbb{1}_{\{|\gamma|=n\}}.$$

If $|\gamma| = n$ and $\gamma \subset B_L$, then

$$\Phi_\beta(\gamma) \le (2L+1)^d \max_{\ell\le n} \phi_\beta(\ell).$$

Consequently, in view of (31),

$$\liminf_{n\to\infty} \frac{1}{n}\log Z_n(h) \ge -\frac{c}{L^2} - (2L+1)^d \liminf_{n\to\infty} \frac{\max_{\ell\le n}\phi(\ell)}{n}, \tag{32}$$

for any $L > L_0$. By (30),

$$\lim_{n\to\infty} \frac{\max_{\ell\le n}\phi_\beta(\ell)}{n} = 0.$$

It follows that $\lambda(h) \ge 0$ for any $h \in \mathbb{R}$. On the other hand, since the interaction potential ϕ_β is non-negative $Z_n = Z_n(0) \le 1$, and, consequently, $\lambda(0) \le 0$. Hence $\lambda(0) = 0 = \min_h \lambda(h)$ as claimed.

Since $\lambda \ge 0$, the set $\mathbf{K}_0$ in (20) is convex. In order to check that it contains an open neighbourhood of the origin it would be enough to show that there exists $\delta > 0$, such that

$$\sum_n Z_n(h) = \sum_{\mathsf{x}} \mathrm{e}^{h\cdot\mathsf{x}} G_0(\mathsf{x}) < \infty, \tag{33}$$

whenever, $|h| < \delta$. The convergence in (33) will follow as soon as we shall show that the critical two-point function $G_0(\mathsf{x})$ in (16) is exponentially decaying in x. We

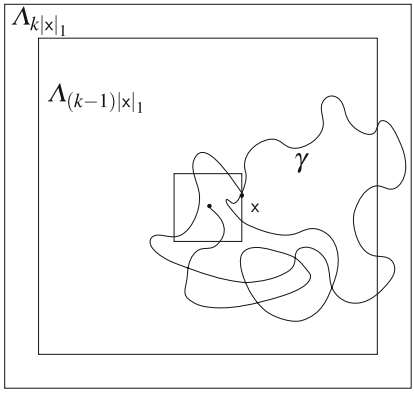

Fig. 2 Paths $\gamma \in \mathscr{P}_{\mathsf{x}}^{(k)}$

continue to employ notation $\mathscr{P}_{\mathsf{x}}$ for paths γ with $\mathsf{X}(\gamma) = \mathsf{x}$ (and, of course, with $\mathsf{P}_d(\gamma) > 0$). Consider the disjoint decomposition

$$\mathscr{P}_{\mathsf{x}} = \bigcup_{k \geq 2} \mathscr{P}_{\mathsf{x}}^{(k)},$$

where (see Fig. 2),

$$\mathscr{P}_{\mathsf{x}}^{(k)} = \left\{\gamma \in \mathscr{P}_{\mathsf{x}} \,:\, \gamma \subset \Lambda_{k|\mathsf{x}|_1}\right\} \setminus \left\{\gamma \in \mathscr{P}_{\mathsf{x}} \,:\, \gamma \subset \Lambda_{(k-1)|\mathsf{x}|_1}\right\}.$$

If $\gamma \in \mathscr{P}_{\mathsf{x}}^{(k)}$, then since the range R of the underlying random walk is finite,

$$\Phi_\beta(\gamma) \geq \frac{(k-1)\,|\mathsf{x}|_1}{R} \inf_\ell \phi_\beta(\ell) = \frac{(k-1)\,|\mathsf{x}|_1}{R} \phi_\beta(1).$$

As a result,

$$G_0^{(k)}(\mathsf{x}) \stackrel{\Delta}{=} \mathsf{W}_d\left(\mathscr{P}_{\mathsf{x}}^{(k)}\right) \leq \mathrm{e}^{-(k-1)|x|_1 \phi_\beta(1)/R} \sum_{\gamma \in \mathscr{P}_{\mathsf{x}}^{(k)}} \mathsf{P}_d(\gamma).$$

At this stage we shall rely on another well known estimate for short range zero-mean random walks:

Estimate 2 Let σ_0 be the first hitting time of 0. Then,

$$\mathsf{E}_d\left(\sum_{\ell=0}^{\sigma_0} \mathbb{1}_{\{\mathsf{X}(\ell)=\mathsf{x}\}} \,\middle|\, \mathsf{X}(0) = \mathsf{x}\right) \lesssim A_d(|\mathsf{x}|) \stackrel{\Delta}{=} \begin{cases} |\mathsf{x}|\,, & d = 1 \\ \log|\mathsf{x}|\,, & d = 2 \\ 1, & d \geq 3 \end{cases} \tag{34}$$

uniformly in $\mathsf{x} \in \mathbb{Z}^d$. By a crude application of (34),

$$\sum_{\gamma \in \mathscr{P}_{\mathsf{x}}^{(k)}} \mathsf{P}_d(\gamma) \le A_d(k\,|\mathsf{x}|_1) \;\Rightarrow\; G_0^{(k)}(\mathsf{x}) \le A_d(k\,|\mathsf{x}|_1)\mathrm{e}^{-\frac{\phi_\beta(1)}{R}(k-1)|\mathsf{x}|_1}. \tag{35}$$

Therefore,

$$G_0(\mathsf{x}) \le \sum_{k\ge 1} A_d((k+1)\,|\mathsf{x}|_1)\mathrm{e}^{-\frac{\phi_\beta(1)}{R}k|\mathsf{x}|_1},$$

and (33) follows.

Part B of Theorem 2.1 As we have already noted, due to a possible over-counting it is not obvious that $G_\lambda(\mathsf{x}+\mathsf{y}) \ge G_\lambda(\mathsf{x})G_\lambda(\mathsf{y})$. However, in view of the attractivity (26) of ϕ_β, the latter super-multiplicativity property holds for the following first-hitting time version H_λ of G_λ:

$$H_\lambda(\mathsf{x}) = \sum_{\gamma \in \mathscr{P}_{\mathsf{x}}} \mathrm{e}^{-\lambda|\gamma|}\mathsf{W}_d(\gamma)\mathbb{1}_{\ell_{\mathsf{x}}(\gamma)=1}. \tag{36}$$

In particular, the limit

$$\tau_\lambda(\mathsf{x}) = -\lim_{r\to\infty}\frac{1}{r}\log H_\lambda\left(\lfloor r\mathsf{x}\rfloor\right) \tag{37}$$

exists, and, by Proposition 2.2, is a non-negative, convex, homogeneous of order one function on $\mathbb{Z}^d$. Furthermore, $H_\lambda(\mathsf{x}) \lesssim \mathrm{e}^{-\tau_\lambda(\mathsf{x})}$.

Exercise 2.5 Prove the above statements.

We claim that the second of (18) holds with the very same τ_λ. Clearly, $H_\lambda(\mathsf{x}) \le G_\lambda(\mathsf{x})$. The proof, therefore, boils down to a derivation of a complementary upper bound, which would render negligible correction on the logarithmic scale. We shall consider two cases: Fix any $\lambda_0 > 0$.

Case 1. $\lambda > \lambda_0$. Then for any x,

$$G_\lambda(\mathsf{x}) \le H_\lambda(\mathsf{x})G_\lambda(0) \le H_\lambda(\mathsf{x})\sum_{\mathsf{y}} G_\lambda(\mathsf{y}) \le H_\lambda(\mathsf{x})\frac{1}{1-\mathrm{e}^{-\lambda_0}}. \tag{38}$$

Case 2. $\lambda \le \lambda_0$. Evidently τ_λ is non-decreasing in λ. Define:

$$k_0 = 3\frac{R}{\phi_\beta(1)}\max_{\mathsf{y}}\frac{\tau_{\lambda_0}(\mathsf{y})}{|\mathsf{y}|_1}. \tag{39}$$

By (35) ,

$$\sum_{k\geq k_0} G_\lambda^{(k)}(\mathsf{x}) \leq \sum_{k\geq k_0} G_0^{(k)} \leq \sum_{k\geq k_0} A_d(k\,|\mathsf{x}|_1)\mathrm{e}^{-\frac{\phi_\beta(1)}{R}(k-1)|\mathsf{x}|_1} \leq \mathrm{e}^{-2\tau_\lambda(\mathsf{x})}$$

is exponentially negligible with respect to $\mathrm{e}^{-\tau_\lambda(\mathsf{x})}$. Consequently,

$$G_\lambda(\mathsf{x}) \overset{<}{\sim} \sum_{k<k_0} G_\lambda^{(k)}(\mathsf{x}) \leq H_\lambda(\mathsf{x})A_d(k_0\,|\mathsf{x}|_1) \leq \mathrm{e}^{-\tau_\lambda(\mathsf{x})}A_d(k_0\,|\mathsf{x}|_1), \tag{40}$$

and the second of (18) indeed follows.

τ_λ is the Support Function of $\mathbf{K}_\lambda$ In order to see this notice that the set $\mathbf{K}_\lambda = \{h\,:\,\lambda(h)\leq\lambda\}$ is the closure of the domain of convergence

$$h \mapsto \sum_n \mathrm{e}^{-\lambda n} Z_n(h) = \sum_{\mathsf{x}} \mathrm{e}^{h\cdot\mathsf{x}} G_\lambda(\mathsf{x}). \tag{41}$$

Consider

$$\alpha_\lambda(h) = \max\{h\cdot\mathsf{x}\,:\,\tau_\lambda(\mathsf{x})\leq 1\}. \tag{42}$$

The series in (41) diverges if $\alpha_\lambda(h)>1$, whereas, $h\in\mathrm{int}(\mathbf{K}_\lambda)$ if $\alpha_\lambda(h)<1$. Hence,

$$\partial\mathbf{K}_\lambda = \{h\,:\,\alpha_\lambda(h)=1\} \quad \text{or} \quad \tau_\lambda(\mathsf{x}) = \max_{h\in\partial\mathbf{K}_\lambda} h\cdot\mathsf{x}. \tag{43}$$

Part C of Theorem 2.1 To be precise large deviations are claimed for the distribution of end-points, which we, with a slight abuse of notation, proceed to call $\mathbb{P}_n^h(\mathsf{x})$:

$$\mathbb{P}_n^h(\mathsf{x}) = \sum_{\mathsf{x}(\gamma)=\mathsf{x}} \mathbb{P}_n^h(\gamma) \overset{\Delta}{=} \frac{\mathrm{e}^{h\cdot\mathsf{x}} Z_n(\mathsf{x})}{Z_n(h)}. \tag{44}$$

Let $h\in\mathbb{R}^d$. The limiting log-moment generation function under the sequence of measures $\{\mathbb{P}_n^h\}$ is

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}_n^h\left(\mathrm{e}^{f\cdot\mathsf{X}}\right) = \lambda(h+f)-\lambda(h).$$

The function I_h in (23) is just the Legendre-Fenchel transform of the above. Since the underlying random walk has bounded range, exponential tightness is automatically ensured. By Theorem 6.26 and Exercise 6.24 of the Appendix, upper large deviation bounds hold with I_h.

We still need a matching lower bound. Let us sketch the proof which relies on sub-additivity and Lemma 6.27 of Appendix. Due to a possible over-counting the function $n \to \log Z_n(\lfloor n\mathsf{x} \rfloor)$ is not necessarily super-additive. However, its first hitting time version (see (15) for the definition of $\mathscr{P}_{\mathsf{x},n}$);

$$\hat{Z}_n(\lfloor n\mathsf{x} \rfloor) = \sum_{\gamma \in \mathscr{P}_{\mathsf{x},n}} \mathsf{W}_d(\gamma) \mathbb{1}_{\{\ell_\gamma(\mathsf{x})=1\}}, \tag{45}$$

is super-additive. By Proposition 2.2 the limit

$$J(\mathsf{x}) = -\lim_{n\to\infty} \frac{1}{n} \log \hat{Z}_n(\lfloor n\mathsf{x} \rfloor) \text{ and } \hat{Z}_n(\lfloor n\mathsf{x} \rfloor) \le c\mathrm{e}^{-nJ(\mathsf{x})} \tag{46}$$

exists and is a convex non-negative function on $\mathbb{R}^d$. Some care is needed to make this statement rigorous. Indeed, $\hat{Z}_n(\mathsf{x}) = 0$ whenever x does not belong to the (bounded) range of the underlying n-step walk, and what happens at boundary points should be explored separately. We shall ignore this issue here.

A slight modification of arguments leading to (19) imply that $J(0) = 0$. By convexity this means that for any $\alpha \in (0,1)$ and any x,

$$J(\mathsf{x}) \le \alpha J\left(\frac{\mathsf{x}}{\alpha}\right). \tag{47}$$

Since, ϕ_β is non-negative,

$$\hat{Z}_n(\lfloor n\mathsf{x} \rfloor) \le Z_n(\lfloor n\mathsf{x} \rfloor) \le \sum_{m=1}^{n} \hat{Z}_m(\lfloor n\mathsf{x} \rfloor) \le \sum_{m=1}^{n} c\mathrm{e}^{-mJ(\frac{n}{m}\mathsf{x})} \le cn\mathrm{e}^{-nJ(\mathsf{x})}, \tag{48}$$

where we used (46) and (47) on the last two steps. We conclude:

$$J(\mathsf{x}) = -\lim_{n\to\infty} \frac{1}{n} \log Z_n(\lfloor n\mathsf{x} \rfloor) \text{ and } Z_n(\lfloor n\mathsf{x} \rfloor) \le c\mathrm{e}^{-nJ(\mathsf{x})+\frac{\log n}{n}} \tag{49}$$

Now, (49) means that (213) and (217) are satisfied, the latter uniformly in x. Since the range of the underlying random walk is bounded, (220), and in particular exponential tightness, is trivially satisfied as well. Hence, by Lemma 6.27, $I = J$. Since

$$-\frac{1}{n} \log \mathbb{P}_n^h(\lfloor n\mathsf{x} \rfloor) = -\frac{1}{n} \log Z_n(\lfloor n\mathsf{x} \rfloor) - \left(h \cdot \frac{\lfloor n\mathsf{x} \rfloor}{n} - \frac{\log Z_n(h)}{n}\right),$$

claim **C** of Theorem 2.1 follows as well.

Part D of Theorem 2.1 and Limiting Spatial Extension For drifts $h \in \text{int}\,(\mathbf{K}_0)$,

$$I_h(\mathsf{v}) = \sup_g \{g \cdot \mathsf{v} - (\lambda(h+g) - \lambda(h))\} \geq |\mathsf{v}|\,\text{dist}\,(h, \partial\mathbf{K}_0) > 0,$$

for any $\mathsf{v} \neq 0$. Hence, if $h \in \text{int}\,(\mathbf{K}_0)$, then the model is sub-ballistic in the sense of (10).

On the other hand, if $h \notin \mathbf{K}_0$ or, equivalently, if $\lambda(h) > 0$, then

$$I_h(0) = \sup_g \{-\lambda(g+h) + \lambda(h)\} = \lambda(h) > 0.$$

This is a rough expression of ballisticity. It implies that the polymer is pulled away from the origin on the linear scale, but it does not imply that the limit in (9) exists.

More precisely, if λ is differentiable at $h \notin \mathbf{K}_0$, then (9) holds with $\mathsf{v} = \nabla\lambda(h)$. Indeed in the latter case I_h is strictly convex at v and, consequently $I_h(\mathsf{v}) = 0$ is the unique minimum. Furthermore, in such a case, the following law of large numbers holds: For any $\epsilon > 0$,

$$\sum_n \mathbb{P}_n^h \left(\left| \frac{\mathsf{X}}{n} - \mathsf{v} \right| \geq \epsilon \right) < \infty, \tag{50}$$

and the series converge exponentially fast.

However, the above sub-additivity based thermodynamics of annealed polymers does not imply that the sub-differential $\partial\lambda(h) \stackrel{\Delta}{=} \mathscr{M}_h = \{\mathsf{v} : I_h(\mathsf{v}) = 0\}$ is always a singleton. The general form of (50) is

$$\sum_n \mathbb{P}_n^h \left(\min_{\mathsf{v} \in \mathscr{M}_h} \left| \frac{\mathsf{X}}{n} - \mathsf{v} \right| \geq \epsilon \right) < \infty. \tag{51}$$

Therefore, in general, large deviations (Part **C** of Theorem 2.1) imply neither existence of the limit in (9), nor a LLN. The set $\mathscr{M}_h$ could be characterized as follows [10, 16]:

Lemma 2.6 *For any $h \notin \mathbf{K}_0$ the set $\mathscr{M}_h$ satisfies: Set $\mu = \lambda(h) > 0$. Then,*

$$\mathsf{v} \in \mathscr{M}_h \Longleftrightarrow \begin{cases} \tau_\mu(\mathsf{v}) = h \cdot \mathsf{v} \\ \frac{\mathrm{d}^-}{\mathrm{d}\lambda}\Big|_{\lambda=\mu} \tau_\lambda(\mathsf{v}) \leq 1 \leq \frac{\mathrm{d}^+}{\mathrm{d}\lambda}\Big|_{\lambda=\mu} \tau_\lambda(\mathsf{v}) \end{cases} \tag{52}$$

Proof By (23),

$$\mathsf{v} \in \mathscr{M}_h \Longleftrightarrow \sup_\lambda (\tau_\lambda(\mathsf{v}) - \lambda)) + (\mu - h \cdot \mathsf{v}) = 0.$$

The choice $\lambda = \mu$ implies that $\tau_\lambda(\mathsf{v}) \le h\cdot\mathsf{v}$. Since $h \in \partial\mathbf{K}_\mu$, the first of (52) follows by (43). As a result,

$$\tau_\lambda(\mathsf{v}) - \tau_\mu(\mathsf{v}) \le \lambda - \mu, \tag{53}$$

for any λ. Since the function $\lambda \to \tau_\lambda(\mathsf{v})$ is concave, left and right derivatives are well defined, and the second of (52) follows from (53).

The differentiability (and even analyticity) of λ at super-critical drifts $h \notin \mathbf{K}_0$ and, in particular, the existence of the limit in (9) and the LLN (50), is established in Sect. 3.2 as a consequence of much sharper asymptotic results based on analysis of renewal structure of ballistic polymers.

Part E of Theorem 2.1 Finally, $I_h(0) = 0$ whenever $h \in \partial\mathbf{K}_0$, which sheds little light on ballistic properties of the model at critical drifts. The critical case was worked out in [17] via refinement of the renormalization construction of the Ornstein-Zernike theory (see Sect. 3.2), and it is beyond the scope of these notes to reproduce the corresponding arguments here.

2.2 *Thermodynamics of Quenched Polymers*

The underlying random walk imposes a directed graph structure on $\mathbb{Z}^d$. Let us say that y is a neighbour of x; $\mathsf{x} \rightsquigarrow \mathsf{y}$ if $\mathsf{P}_d(\mathsf{y}-\mathsf{x}) > 0$. Because of (1) and Assumption **(A.2)** there is a unique infinite component Cl_∞ of $\{\mathsf{x} : \mathsf{V}^\omega_\mathsf{x} < \infty\}$. Clearly, non-trivial thermodynamic limits may exist only if $0 \in \mathrm{Cl}_\infty$. Furthermore, if $\mathscr{E}(\mathsf{V}^\omega) = \infty$, then $\sum_r \mathscr{Q}\left(V^\omega_{\lfloor r\mathsf{x}\rfloor} > cr\right) = \infty$ for any $c > 0$, and consequently,

$$\liminf_{r\to\infty} \frac{1}{r} \log G^\omega_\lambda(\lfloor r\mathsf{x}\rfloor) = -\infty,$$

$\mathscr{Q}$-a.s. for any $\mathsf{x} \neq 0$. Hence, in order to define inverse correlation length $\tau^{\mathsf{q}}_\lambda$ one needs either to impose more stringent requirements on disorder and use (18), or to find a more robust definition of $\tau^{\mathsf{q}}_\lambda$. A more robust definition is in terms of the so called point to hyperplane exponents:

Given $h \neq 0$ define the set $\mathscr{H}^+_{h,t} = \{\mathsf{x} : h\cdot\mathsf{x} \ge t\}$. Let $\mathscr{P}_{h,t}$ be the set of paths $\gamma = (\gamma(0), \dots, \gamma(n))$ with $\gamma(n) \in \mathscr{H}^+_{h,t}$. For $\lambda \ge 0$ consider,

$$D^\omega_{h,\lambda}(t) = \sum_{\gamma\in\mathscr{P}_{h,t}} \mathrm{e}^{-\lambda|\gamma|}\mathsf{W}^\omega_d(\gamma).$$

Assume that the limit

$$-\lim_{t\to\infty} \frac{1}{t} \log D^\omega_{h,\lambda}(t) \stackrel{\Delta}{=} \frac{1}{\alpha^{\mathsf{q}}_\lambda(h)} \tag{54}$$

exists. Should the inverse correlation length $\tau_\lambda^{\mathsf{q}}$ be also defined (and positive), the following relation should hold:

$$\frac{1}{\alpha_\lambda^{\mathsf{q}}(h)} = \min_{\mathsf{x}\in\mathscr{H}_{h,1}^+} \tau_\lambda^{\mathsf{q}}(\mathsf{x}). \tag{55}$$

If $\tau_\lambda^{\mathsf{q}}$ is the support function of a convex set $\mathbf{K}_\lambda^{\mathsf{q}}$, then, by (198) of the Appendix, $\alpha_\lambda^{\mathsf{q}}$ should be the support function of the polar set (199) or, equivalently, the Minkowski function of $\mathbf{K}_\lambda^{\mathsf{q}}$.

Conversely, if the limit $\alpha_\lambda^{\mathsf{q}}$ in (54) exists, then we may define $\tau_\lambda^{\mathsf{q}}$ via

$$\tau_\lambda^{\mathsf{q}}(\mathsf{x}) = \max\left\{h\cdot x \,:\, \alpha_\lambda^{\mathsf{q}}(h) \le 1\right\}, \tag{56}$$

even if a direct application of (18) does not make sense.

There is an extensive literature on thermodynamics of quenched models, [10, 23, 29, 31] to mention a few. The paper [23] contains state of the art information on the matter, and several conditions on the random environment were worked out there in an essentially optimal form. The treatment of $\mathscr{Q}(\mathsf{V}^\omega = \infty) > 0$ case and, more generally, of $\mathscr{E}(\mathsf{V}^\omega) = \infty$ case is based on renormalization techniques for high density site percolation and, eventually, on sub-additive ergodic theorems and large deviation arguments. It is beyond the scope of these lectures to reproduce the corresponding results here. Below we formulate some of the statements from [23] and refer to the latter paper for proofs and detailed discussions.

We assume **(A1)** and **(A2)**.

Theorem 2.7 *The following happens $\mathscr{Q}$-a.s on the event $0 \in \mathrm{Cl}_\infty$:*

A. *The free energy λ^{q} is well defined, deterministic, non-negative and convex on $\mathbb{R}^d$. Furthermore,*

$$0 = \min_h \lambda^{\mathsf{q}}(h) = \lambda^{\mathsf{q}}(0). \tag{57}$$

The set

$$\mathbf{K}_0^{\mathsf{q}} \stackrel{\Delta}{=} \left\{h \,:\, \lambda^{\mathsf{q}}(h) = 0\right\} \tag{58}$$

is a compact convex set with a non-empty interior.

B. *The point to hyperplane exponent $\alpha_\lambda^{\mathsf{q}}$ in (54) is well defined for any $\lambda \ge 0$. Consequently, the inverse correlation length $\tau_\lambda^{\mathsf{q}}$ is well defined via (56) also for any $\lambda \ge 0$, and, furthermore, it can be identified as the support function of the compact convex set*

$$\mathbf{K}_\lambda^{\mathsf{q}} \stackrel{\Delta}{=} \left\{h \,:\, \lambda^{\mathsf{q}}(h) \le \lambda\right\}. \tag{59}$$

Define

$$I^{\mathsf{q}}(\mathsf{v}) = \sup_h \left\{h \cdot \mathsf{v} - \lambda^{\mathsf{q}}(h)\right\} = \sup_\lambda \left\{\tau_\lambda^{\mathsf{q}}(\mathsf{v}) - \lambda\right\}. \tag{60}$$

C. *For any $h \in \mathbb{R}^d$ the family of polymer measures $\mathbb{P}_n^{h,\omega}$ satisfies LD principle with the rate function*

$$I_h^{\mathsf{q}}(\mathsf{v}) \stackrel{\Delta}{=} \sup_f \left\{f \cdot \mathsf{v} - \left(\lambda^{\mathsf{q}}(f+h) - \lambda^{\omega}(h)\right)\right\} = I^{\mathsf{q}}(\mathsf{v}) - \left(h \cdot \mathsf{v} - \lambda^{\mathsf{q}}(h)\right). \tag{61}$$

D. *For $h \in \operatorname{int}\left(\mathbf{K}_0^{\mathsf{q}}\right)$ the model is sub-ballistic, whereas for any $h \notin \mathbf{K}_0^{\mathsf{q}}$ the model is ballistic in the sense that $I_h^{\mathsf{q}}(0) > 0$.*

The above theorem does not imply strong limiting spatial extension form of the ballisticity condition (11) for all $h \notin \mathbf{K}_0^{\mathsf{q}}$, exactly for the same reasons as Theorem 2.1 does not imply the corresponding statement for annealed models. Existence of limiting spatial extension for quenched models in the very weak disorder regime is discussed, together with other limit theorems, in Sect. 4.

In the case of critical drifts $h \in \partial\mathbf{K}_0^{\mathsf{q}}$, a form of ballistic behaviour was established in the continuous context in [28].

3 Multidimensional Renewal Theory and Annealed Polymers

3.1 Multi-Dimensional Renewal Theory

One-Dimensional Renewals Let $\{\mathsf{f}(\mathsf{n})\}$ be a probability distribution on $\mathbb{N}$ (with strictly positive variance). We can think of f as of a probability distribution for a step T of the effective one-dimensional random walk

$$\mathsf{S}_N = \sum_1^N \mathsf{T}_i.$$

The distribution of $\{\mathsf{S}_n\}$ is governed by the product measure $\mathbb{P}$. The renewal array $\{\mathsf{t}(n)\}$ is given by

$$\mathsf{t}(0) = 1 \quad \text{and} \quad \mathsf{t}(n) = \sum_{m=1}^n \mathsf{f}(m)\mathsf{t}(n-m). \tag{62}$$

In probabilistic terms (62) reads as:

$$\mathsf{t}(n) = \mathbb{P}\left(\exists\, N \,:\, \mathsf{S}_N = n\right) = \sum_N \mathbb{P}\left(\mathsf{S}_N = n\right). \tag{63}$$

Renewal theory implies that

$$\lim_{n\to\infty} \mathsf{t}(n) = \frac{1}{\mathbb{E}\mathsf{T}} \stackrel{\Delta}{=} \frac{1}{\mu}. \tag{64}$$

A proof of (64) is based on an analysis of complex power series

$$\hat{\mathsf{t}}(z) \stackrel{\Delta}{=} \sum_n \mathsf{t}(n) z^n \quad \text{and} \quad \hat{\mathsf{f}}(z) \stackrel{\Delta}{=} \sum_n \mathsf{f}(n) z^n \tag{65}$$

Exercise 3.1 Show that $\hat{\mathsf{t}}$ is absolutely convergent and hence analytic on the interior of the unit disc $\mathbb{D}_1 = \{z : |z| < 1\}$. Check that $\hat{\mathsf{t}}(1) = \infty$.

It follows that on $\mathbb{D}_1$,

$$\hat{\mathsf{t}}(z) = \frac{1}{1 - \hat{\mathsf{f}}(z)}. \tag{66}$$

Exercise 3.2 Check that $|\mathsf{f}(z)| \le \mathsf{f}(|z|) < 1$ for any $z \in \mathbb{D}_1$. Prove (66).

Consequently, by Cauchy formula,

$$\mathsf{t}(n) = \frac{1}{2\pi i} \oint_{|z|=r} \frac{\hat{\mathsf{t}}(z)}{z^{n+1}} \mathrm{d}z = \frac{1}{2\pi i} \oint_{|z|=r} \frac{\mathrm{d}z}{z^{n+1}\left(1 - \hat{\mathsf{f}}(z)\right)}, \tag{67}$$

for any $r < 1$.

Exponential Tails Assume that there exists $\nu > 0$, such that

$$\mathsf{f}(\mathsf{n}) \stackrel{<}{\sim} \mathrm{e}^{-\nu n}, \tag{68}$$

uniformly in $n \in \mathbb{N}$.

Lemma 3.3 *Under Assumption* (68) *the convergence in* (64) *is exponentially fast in n.*

We start proving Lemma 3.3 by noting that under (68) the function $\hat{\mathsf{f}}$ is defined and analytic on $\mathbb{D}_{1+\nu}$.

Exercise 3.4 Check that there exists $\epsilon \in (0, \nu)$ such that $z = 1$ is the only zero of $1 - \mathsf{f}(z)$ on $\bar{\mathbb{D}}_{1+\epsilon}$. Furthermore, $\frac{1-z}{\hat{\mathsf{f}}(z)-1}$ is analytic on $\mathbb{D}_{1+\epsilon}$.

Recall that we defined $\mu = \sum_n n\mathsf{f}(n) = \hat{\mathsf{f}}'(1)$. Consider the representation,

$$\frac{1}{\mu} = \frac{1}{2\pi i} \int_{|z|=r} \frac{\mathrm{d}z}{\mu(1-z)z^{n+1}},$$

which holds for any $r < 1$. By (67),

$$\mathsf{t}(n) - \frac{1}{\mu} = \frac{1}{2\pi i}\int_{|z|=r} \frac{\hat{\mathsf{f}}(z) - 1 - \mu(z-1)}{(1-\hat{\mathsf{f}}(z))(1-z)z^{n+1}}\mathrm{d}z \stackrel{\Delta}{=} \frac{1}{2\pi i}\int_{|z|=r} \frac{\Delta(z)}{z^{n+1}}\mathrm{d}z. \tag{69}$$

On $\bar{\mathbb{D}}_{1+\epsilon}$ the denominator in the definition of Δ vanishes only at $z = 1$. However, since $\mu = \hat{\mathsf{f}}'(1)$, expansion of the numerator in a neighbourhood of $z = 1$ gives:

$$\hat{\mathsf{f}}(z) - 1 - \mu(z-1) = (z-1)^2 U(z),$$

with some analytic U. It follows that

$$\Delta(z) = \frac{U(z)}{(1-\hat{\mathsf{f}}(z))/(z-1)}.$$

In view of Exercise 3.4 Δ is analytic on $\bar{\mathbb{D}}_{1+\epsilon}$. As a result,

$$\frac{1}{2\pi i}\int_{|z|=r} \frac{\Delta(z)}{z^{n+1}}\mathrm{d}z = \frac{1}{2\pi i}\int_{|z|=1+\epsilon} \frac{\Delta(z)}{z^{n+1}}\mathrm{d}z.$$

By (69),

$$\left|\mathsf{t}(n) - \frac{1}{\mu}\right| \lesssim (1+\epsilon)^{-n}, \tag{70}$$

uniformly in n. This is precisely the claim of Lemma 3.3. □

Complex Renewals Suppose that (62) holds with complex $\{\mathsf{f}(n)\}$ and, accordingly, with complex $\{\mathsf{t}(n)\}$. As before, define $\hat{\mathsf{t}}(z)$ and $\hat{\mathsf{f}}(z)$ as in (65). In the sequel we shall work with complex renewals which satisfy one of the following two assumptions, Assumption 3.5 or Assumption 3.6, below.

Assumption 3.5 There exists $\epsilon > 0$, such that the function $\hat{\mathsf{f}}$ satisfies the following three properties:

(a) $\hat{\mathsf{f}}(0) = 0$ and the $\hat{\mathsf{f}}(z)$ in (65) is absolutely convergent in a neighbourhood of $\bar{\mathbb{D}}_{1+\epsilon}$.

(b) $z = 1$ is the only zero of $\left(\hat{\mathsf{f}}(z) - 1\right)$ in $\bar{\mathbb{D}}_{1+\epsilon}$.

(c) $\hat{\mathsf{f}}'(1) \neq 0$.

Under Assumption 3.5 the exponential convergence bound (70) still holds. Indeed, the only thing we have to justify is that $\hat{\mathsf{t}}(z) = \sum_n \mathsf{t}(n)z^n$ is defined and analytic on

some neighbourhood of the origin, and that

$$\hat{\mathsf{t}}(z) = \frac{1}{1 - \hat{\mathsf{f}}(z)}, \tag{71}$$

for all $|z|$ sufficiently small. Indeed, if this is the case, then (67) holds for some $r > 0$, and we may just proceed as before. However, by Assumption 3.5(a), $\sum_n |\mathsf{f}(n)|\,|z|^n < 1$ for all $|z|$ small enough. Hence (71).

Assumption 3.6 There exists $\epsilon > 0$, such that the function $\hat{\mathsf{f}}$ satisfies the following two properties:

(a) The series $\hat{\mathsf{f}}(z)$ in (65) is absolutely convergent in a neighbourhood of $\bar{\mathbb{D}}_{1+\epsilon}$.
(b) There exists $\kappa > 0$, such that $\min_{|z|\le 1+\epsilon} \left|1 - \hat{\mathsf{f}}(z)\right| \ge \kappa$.

Under Assumption 3.6, the function $\left(1 - \hat{\mathsf{f}}(z)\right)^{-1}$ is analytic in a neighbourhood of $\bar{\mathbb{D}}_{1+\epsilon}$, and the Cauchy formula (67), which again by absolute convergence of $\hat{\mathsf{f}}$ still holds for r sufficiently small, implies:

$$|\mathsf{t}(n)| \le \frac{1}{\kappa(1+\epsilon)^n}. \tag{72}$$

Multi-dimensional Renewals Let $\{\mathsf{f}(\mathsf{x}, \mathsf{n})\}$ be a probability distribution on $\mathbb{Z}^d \times \mathbb{N}$. As in the one-dimensional case we can think of f as of a probability distribution for a step $\mathsf{U} = (\mathsf{X}, \mathsf{T})$ of the effective $(d+1)$-dimensional random walk

$$\mathsf{S}_N = \sum_1^N \mathsf{U}_i.$$

The distribution of $\{\mathsf{S}_n\}$ is governed by the product measure $\mathbb{P}$. We assume:

Assumption 3.7 Random vector $\mathsf{U} = (\mathsf{X}, \mathsf{T})$ has a non-degenerate $(d+1)$-dimensional distribution. The random walk S_N is aperiodic (that is its support is not concentrated on a regular sub-lattice).

The renewal array $\{\mathsf{t}(\mathsf{x}, n)\}$ is given by

$$\mathsf{t}(\mathsf{x}, 0) = \mathbb{1}_{\{\mathsf{x}=0\}} \quad \text{and} \quad \mathsf{t}(\mathsf{x}, n) = \sum_{m=1}^{n} \sum_{\mathsf{y}} \mathsf{f}(\mathsf{y}, m)\mathsf{t}(\mathsf{x} - \mathsf{y}, n - m). \tag{73}$$

Again, as in the one dimensional case (63), in probabilistic terms (73) reads as:

$$\mathsf{t}(\mathsf{x}, n) = \mathbb{P}\left(\exists N : \mathsf{S}_N = (\mathsf{x}, n)\right). \tag{74}$$

The renewal relation is inherited by one-dimensional marginals: Set

$$\mathsf{f}(n) = \sum_{\mathsf{x}} \mathsf{f}(\mathsf{x}, n) \quad \text{and} \quad \mathsf{t}(n) = \sum_{\mathsf{x}} \mathsf{t}(\mathsf{x}, n).$$

Then, (62) holds.

We are going to explore the implications of the renewal relation (73) for a local limit analysis of conditional measures

$$\mathbb{Q}_n(\mathsf{x}) = \frac{\mathsf{t}(\mathsf{x}, n)}{\mathsf{t}(n)}. \tag{75}$$

Exponential Tails Assume that there exists $\nu > 0$, such that

$$\mathsf{f}(\mathsf{x}, \mathsf{n}) \overset{<}{\sim} \mathrm{e}^{-\nu(|\mathsf{x}|+n)}, \tag{76}$$

uniformly in $(\mathsf{x}, n) \in \mathbb{Z}^d \times \mathbb{N}$. In particular, (70) holds, and as a result we already have a sharp control over denominators in (75).

Consider the following equation

$$F(\xi, \lambda) \overset{\Delta}{=} \log \sum_{\mathsf{x},n} \mathrm{e}^{\mathsf{x}\cdot\xi - \lambda n} \mathsf{f}(\mathsf{x}, n) = 0. \tag{77}$$

Above $F : \mathbb{C}^d \times \mathbb{C} \mapsto \mathbb{C}$.

Exercise 3.8 Check that under (76) there exists $\delta > 0$ such that F is well defined and analytic on the disc $\mathbb{D}_\delta^{d+1} \subset \mathbb{C}^{d+1}$.

Shape Theorem We shall assume that δ is sufficiently small. Then by the analytic implicit function theorem [19], whose application is secured by Assumption 3.7, there is an analytic function $\lambda : \mathbb{D}_\delta^d \mapsto \mathbb{C}$ with such that for $(\xi, \lambda) \in \mathbb{D}_\delta^{d+1}$,

$$F(\xi, \lambda) = 0 \;\Leftrightarrow\; \lambda = \lambda(\xi). \tag{78}$$

For $\xi \in \mathbb{D}_\delta^d$ define:

$$\mathsf{f}_\xi(\mathsf{x}, n) = \mathsf{f}(\mathsf{x}, n)\mathrm{e}^{\xi\cdot\mathsf{x} - \lambda(\xi)n} \quad \text{and} \quad \mathsf{t}_\xi(\mathsf{x}, n) = \mathsf{t}(\mathsf{x}, n)\mathrm{e}^{\xi\cdot\mathsf{x} - \lambda(\xi)n}. \tag{79}$$

Evidently, the arrays $\{\mathsf{f}_\xi(\mathsf{x}, n)\}$ and $\{\mathsf{t}_\xi(\mathsf{x}, n)\}$ satisfy (73). Also, under (76), $\mathsf{f}_\xi(n) \overset{\Delta}{=} \sum_{\mathsf{x}} \mathsf{f}_\xi(\mathsf{x}, n)$ is well defined for all $|\xi| < \nu$.

Lemma 3.9 *There exists $\delta > 0$ and $\epsilon > 0$ such that*

$$\hat{\mathsf{f}}_\xi(z) \overset{\Delta}{=} \sum_n \mathsf{f}_\xi(n) z^n,$$

satisfies Assumption 3.5 for all $|\xi| < \delta$.

Proof Conditions **(a)** and **(c)** are straightforward. In order to check **(b)** note that it is trivially satisfied at $\xi = 0$. Which, by continuity means that we can fix $\epsilon > 0$ such that for any $\nu > 0$ fixed, the equation

$$\hat{\mathsf{f}}_\xi(z) = 1 \tag{80}$$

has no solutions in $\mathbb{D}_{1+\epsilon} \setminus \mathbb{D}_\nu(1)$ for all $|\xi| < \delta$. However, the family of analytic functions $\left\{\hat{\mathsf{f}}_\xi\right\}_{|\xi|<\delta}$ is uniformly bounded on $\bar{\mathbb{D}}_\nu(1)$. Furthermore, for $\delta > 0$ small the collection of derivatives.

$$\left\{\hat{\mathsf{f}}'_\xi(1) \stackrel{\Delta}{=} \mu(\xi) \stackrel{\Delta}{=} \sum_n n\mathsf{f}_\xi(n)\right\}_{|\xi|<\delta} \tag{81}$$

is uniformly bounded away from zero. Therefore, there exist $\nu > 0$ and $\delta = \delta(\nu) > 0$, such that $z = 1$ is the only solution of $\hat{\mathsf{f}}_\xi(z) = 1$ on $\bar{\mathbb{D}}_\nu(1)$ for all $|\xi| < \delta$.

Remark 3.10 Note that the restriction of F to $\mathbb{R}^{d+1} \cap \mathbb{D}_\delta^{d+1}$ is convex, and it is monotone non-increasing in λ. Hence, the restriction of λ to $\mathbb{R}^d \cap \mathbb{D}_\delta^d$ is convex as well. Indeed, let $\lambda_i = \lambda(\xi_i)$; $i = 1, 2$, for two vectors $\xi_1, \xi_2 \in \mathbb{R}^d \cap \mathbb{D}_\delta^d$. From convexity of level set $\{(\xi, \lambda) : F(\xi, \lambda) \le 0\}$, we infer that for any convex combination $\xi = \alpha\xi_1 + (1-\alpha)\xi_2$

$$F(\xi, \alpha\lambda_1 + (1-\alpha)\lambda_2) \le 0 \;\Rightarrow\; \lambda(\xi) \le \alpha\lambda_1 + (1-\alpha)\lambda_2.$$

The term *shape theorem* comes from the fact that in applications function λ frequently describes local parametrization of the boundary of the appropriate limiting shape.

Limit Theorems Consider the canonical measure $\mathbb{Q}_n$ defined in (75). The following proposition describes ballistic behaviour under $\mathbb{Q}_n$.

Proposition 3.11 *Under assumption on exponential tails* (76)*,*

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{Q}_n(\mathsf{X}) = \lim_{n\to\infty} \frac{1}{n}\frac{\sum_{\mathsf{x}} \mathsf{t}(\mathsf{x}, n)\mathsf{x}}{\mathsf{t}(n)} = \frac{\mathbb{E}\mathsf{X}}{\mathbb{E}\mathsf{T}} \stackrel{\Delta}{=} \mathsf{v}. \tag{82}$$

Proof Note that

$$\mathbb{Q}_n(\mathsf{X}) = \frac{\sum_{\mathsf{x}} \mathsf{t}(\mathsf{x}, n)\mathsf{x}}{\mathsf{t}(n)} = \nabla_\xi \log\left(\sum_{\mathsf{x}} \mathrm{e}^{\xi\cdot\mathsf{x}}\mathsf{t}(\mathsf{x}, n)\right)(0).$$

For $|\xi|$ small we can rely on Lemmas 3.3 and 3.9 to conclude that

$$\mathrm{e}^{-\lambda(\xi)n}\sum_{\mathsf{x}} \mathrm{e}^{\xi\cdot\mathsf{x}}\mathsf{t}(\mathsf{x}, n) = \frac{1}{\mu(\xi)}\left(1 + \mathrm{o}\left((1+\epsilon)^{-n}\right)\right), \tag{83}$$

where $\mu(\xi)$ was defined in (81). The convergence in (83) is in a sense of analytic functions on $\mathbb{D}_\delta^d$ for δ small enough. As a result, the convergence,

$$\mathrm{e}^{-\lambda(\xi)n}\mathbb{Q}_n\left(\mathrm{e}^{\xi\cdot\mathsf{X}}\right) = \frac{\mu(0)}{\mu(\xi)}\left(1+\mathrm{o}\left((1+\epsilon)^{-n}\right)\right), \tag{84}$$

is also in a sense of analytic functions on $\mathbb{D}_\delta^d$. Since $\lambda(0)=0$,

$$\mathsf{v}_n \stackrel{\Delta}{=} \frac{1}{n}\mathbb{Q}_n(\mathsf{X}) = \nabla\lambda(0) - \frac{1}{n}\nabla\log\mu(0) + \mathrm{o}\left((1+\epsilon)^{-n}\right). \tag{85}$$

Equation (82) follows, since by (77)

$$\nabla\lambda(0) = \frac{\mathbb{E}\mathsf{X}}{\mathbb{E}\mathsf{T}} = \mathsf{v}. \tag{86}$$

Integral Central Limit Theorem For $\xi\in\mathbb{C}^d$ consider

$$\phi_n(\xi) = \mathbb{Q}_n\left(\exp\left\{i(\mathsf{X}-n\mathsf{v}_n)\cdot\xi\right\}\right). \tag{87}$$

For any R fixed, $\phi_n\left(\frac{\cdot}{\sqrt{n}}\right)$ is well defined on $\mathbb{D}_R^d$. By (84) and (85),

$$\phi_n\left(\frac{\xi}{\sqrt{n}}\right) = \exp\left\{n\left(\lambda(\frac{i\xi}{\sqrt{n}}) - \nabla\lambda(0)\cdot\frac{i\xi}{\sqrt{n}}\right) + \mathrm{O}\left(\frac{R}{\sqrt{n}}\right)\right\} = \mathrm{e}^{-\frac{1}{2}\Xi\xi\cdot\xi+\mathrm{O}\left(\frac{R}{\sqrt{n}}\right)}, \tag{88}$$

in the sense of analytic functions on $\mathbb{D}_R^d$. Above $\Xi \stackrel{\Delta}{=} \mathrm{Hess}(\lambda)$. We claim:

Lemma 3.12 *Ξ is a positive definite $d\times d$ matrix.*

Proof Fix $\xi\in\mathbb{R}^d\setminus 0$ and consider (78):

$$\sum_{\mathsf{x},n} f(\mathsf{x},n)\mathrm{e}^{\epsilon\xi\cdot\mathsf{x}-n\lambda(\epsilon\xi)} \equiv 1,$$

which holds for all $|\epsilon| < \delta/|\xi|$. The second order expansion gives:

$$\mathrm{Hess}\lambda(0)\,\xi\cdot\xi = \frac{1}{\mathbb{E}\mathsf{T}}\mathbb{E}\left(\left(\mathsf{X}-\frac{\mathbb{E}\mathsf{X}}{\mathbb{E}\mathsf{T}}\mathsf{T}\right)\cdot\xi\right)^2.$$

The claim of the lemma follows from the non-degeneracy Assumption 3.7.

In view of Lemma 3.12, asymptotic formula (88) already implies the integral form of the CLT: The family of random vectors $\frac{1}{\sqrt{n}}(\mathsf{X}-n\mathsf{v}_n)$ weakly converges (under $\{\mathbb{Q}_n\}$) to $\mathcal{N}(0,\Xi)$.

Local CLT ϕ_n is related to the characteristic function of X in the following way: For any $\theta \in \mathbb{R}^d$,

$$\phi_n(\theta) = \mathrm{e}^{-in\mathsf{v}_n\cdot\theta}\mathbb{Q}_n\left(\mathrm{e}^{i\theta\cdot\mathsf{X}}\right) = \frac{\mathrm{e}^{-in\mathsf{v}_n\cdot\theta}}{\mathsf{t}(n)}\sum_{\mathsf{x}}\mathsf{t}(\mathsf{x},n)\mathrm{e}^{i\mathsf{x}\cdot\theta}.$$

The complex array $\left\{\mathsf{t}(\mathsf{x},n)\mathrm{e}^{i\mathsf{x}\cdot\theta}\right\}$ is generated via multi-dimensional renewal relation (73) by $\left\{\mathsf{f}(\mathsf{x},n)\mathrm{e}^{i\mathsf{x}\cdot\theta}\right\}$. Since $\{\mathsf{f}(\mathsf{x},n)\}$ is a non-degenerate probability distribution on $\mathbb{Z}^d$ with exponentially decaying tails, for any $\delta > 0$ one can find $\kappa = \kappa(\delta) > 0$ and $\epsilon = \epsilon(\delta)$, such that the array $\left\{\mathsf{f}(\mathsf{x},n)\mathrm{e}^{i\mathsf{x}\cdot\theta}\right\}$ satisfies Assumption 3.6 uniformly in $|\theta| \geq \delta$. We conclude:

Lemma 3.13 *For any $\delta > 0$ there exists $c_\delta > 0$ such that*

$$|\phi_n(\theta)| \leq \mathrm{e}^{-c_\delta n}, \tag{89}$$

whenever $\theta \in \mathbb{R}^d$ satisfies $|\theta| \geq \delta$.

One applies Lemma 3.13 as follows: By the Fourier inversion formula,

$$\mathbb{Q}_n(\mathsf{x}) = \frac{1}{(2\pi)^d}\int_{\mathbb{T}^d}\mathrm{e}^{-i\theta\cdot(\mathsf{x}-n\mathsf{v}_n)}\phi_n(\theta)\mathrm{d}\theta. \tag{90}$$

Choose $\delta, \epsilon > 0$ small. The above integral splits into the sum of three terms:

$$\begin{aligned}\int \mathrm{e}^{-i\theta\cdot(\mathsf{x}-n\mathsf{v}_n)}\phi_n(\theta)\mathrm{d}\theta = &\int_{A_n}\mathrm{e}^{-i\theta\cdot(\mathsf{x}-n\mathsf{v}_n)}\phi_n(\theta)\mathrm{d}\theta \\ &+\int_{B_n}\mathrm{e}^{-i\theta\cdot(\mathsf{x}-n\mathsf{v}_n)}\phi_n(\theta)\mathrm{d}\theta + \int_{|\theta|\geq\delta}\mathrm{e}^{-i\theta\cdot(\mathsf{x}-n\mathsf{v}_n)}\phi_n(\theta)\mathrm{d}\theta.\end{aligned} \tag{91}$$

Above $A_n = \left\{\theta : |\theta| < n^{-\frac{1}{2}+\epsilon}\right\}$ and $B_n = \left\{\theta : n^{-\frac{1}{2}+\epsilon} \leq |\theta| < \delta\right\}$. The third integral is negligible by Lemma 3.13. In order to control the second integral (over B_n) note that for δ small enough (84) applies, and hence, in view of positive definiteness of Ξ,

$$\left|\int_{B_n}\mathrm{e}^{-i\theta\cdot(\mathsf{x}-n\mathsf{v}_n)}\phi_n(\theta)\mathrm{d}\theta\right| \leq \int_{B_n}\mathrm{e}^{-\frac{n}{4}\Xi\,\theta\cdot\theta}\mathrm{d}\theta.$$

The first integral in (91) gives local CLT asymptotics uniformly in $|\mathsf{x}-n\mathsf{v}_n| = \mathrm{o}\left(n^{\frac{1}{2}-(d+1)\epsilon}\right)$. Namely, for such x-s

$$\int_{|\theta|<n^{-\frac{1}{2}+\epsilon}}\mathrm{e}^{-i\theta\cdot(\mathsf{x}-n\mathsf{v}_n)}\phi_n(\theta)\mathrm{d}\theta = \int_{|\theta|<n^{-\frac{1}{2}+\epsilon}}\phi_n(\theta)\mathrm{d}\theta + \mathrm{o}\left(\frac{1}{\sqrt{n^d}}\right).$$

As in (88),

$$\phi_n(\theta) = \mathrm{e}^{-\frac{n}{2}\Xi\,\theta\cdot\theta+\mathrm{o}\left(n|\theta|^3\right)},$$

uniformly in $|\theta| < n^{-\frac{1}{2}+\epsilon}$. We have proved:

Proposition 3.14 *For any fixed $\epsilon > 0$ the asymptotic relation:*

$$\mathbb{Q}_n(\mathsf{x}) = \frac{1}{\sqrt{(2\pi n)^d \det \Xi}}\left(1 + \mathrm{o}\,(1)\right), \tag{92}$$

holds uniformly in $|\mathsf{x} - n\mathsf{v}_n| \lesssim n^{\frac{1}{2}-\epsilon}$.

In fact, as it will become clear from local large deviations estimates below, it would be enough to state (92) only for $|\mathsf{x} - n\mathsf{v}_n| \leq 1$.

Local Large Deviations Estimates Assumption (76) implies that the family of measures $\{\mathbb{Q}_n\}$ is exponentially tight. Furthermore, by the very definition of t in (73)

$$\mathsf{t}(\mathsf{x}+\mathsf{y}, n+m) \geq \mathsf{t}(\mathsf{x}, n)\mathsf{t}(\mathsf{y}, m).$$

Hence, by the sub-additivity argument the function

$$J(\mathsf{u}) = -\lim_{n\to\infty} \frac{1}{n} \log \mathsf{t}(n, \lfloor n\mathsf{u} \rfloor) \tag{93}$$

is well defined and convex on $\mathbb{R}^d$. By the renewal theorem (64),

$$J(\mathsf{u}) = -\lim_{n\to\infty} \frac{1}{n} \log \mathbb{Q}_n(\lfloor n\mathsf{u} \rfloor).$$

Consequently, $\{\mathbb{Q}_n\}$ satisfies the large deviation principle with J.

A large deviation result states what it states. Obviously, J in (93) is non-negative, and $\min J = J(\mathsf{v}) = 0$, where v was defined in (82). We shall show that J has a quadratic minimum on $\mathbb{B}^d_\kappa(\mathsf{v})$, and prove a local LD asymptotic relation for any $\mathsf{u} \in \mathbb{B}^d_\kappa(\mathsf{v})$.

Recall that λ is analytic (and convex) on a (real) ball $\mathbb{B}^d_\delta$. Since, as we already know by Lemma 3.12, Hess $(\lambda)\,(0)$ is non-degenerate, and since $\nabla\lambda(0) = \mathsf{v}$, there exists $\kappa > 0$ such that

$$\mathbb{B}^d_\kappa(\mathsf{v}) \subset \nabla\lambda\big|_{\mathbb{B}^d_\delta}. \tag{94}$$

Let $\mathsf{u} \in \mathbb{B}^d_\kappa(\mathsf{v})$. Set $\mathsf{u}_n = \lfloor n\mathsf{u} \rfloor / n$ and choose $\xi_n \in \mathbb{B}^d_\delta$ such that $\mathsf{u}_n = \nabla\lambda(\xi_n)$. By (94) such ξ_n exists (at least for all n sufficiently large), and, for it is unique by the

implicit function theorem. Recall how we defined tilted function t_{ξ_n} in (79). Then,

$$\mathsf{t}(n, \lfloor n\mathsf{u}\rfloor) = \mathrm{e}^{n\lambda(\xi_n)-\xi_n\cdot\lfloor n\mathsf{u}\rfloor}\mathsf{t}_{\xi_n}(n, \lfloor n\mathsf{u}\rfloor) \tag{95}$$

The term $\mathsf{t}_{\xi_n}(n, \lfloor n\mathsf{u}\rfloor)$ obeys uniform sharp CLT asymptotics (92) with $\mathit{\Xi}(\xi_n) = \mathrm{Hess}(\lambda)(\xi_n)$. The term

$$J(\mathsf{u}_n) = \xi_n \cdot \mathsf{u}_n - \lambda(\xi_n)$$

is quadratic. Indeed, for any $\eta \in \mathbb{B}^d_\kappa(\mathsf{v})$ and $\mathsf{w} = \nabla\lambda(\eta)$, one, using $\lambda(0) = 0$, can rewrite:

$$\begin{aligned}\eta\cdot\mathsf{w} - \lambda(\eta) &= \lambda(0) - \lambda(\eta) - (-\eta)\cdot\nabla\lambda(\eta)\\ &= \left\{\int_0^1\int_0^s \mathrm{Hess}(\lambda)((1-\tau)\eta)\mathrm{d}\tau\mathrm{d}s\right\}\eta\cdot\eta,\end{aligned} \tag{96}$$

and rely on non-degeneracy of $\mathrm{Hess}(\lambda)$ on $\mathbb{B}^d_\kappa$. Incidentally, we have checked that on $\mathbb{B}^d_\kappa(\mathsf{v})$ the function $J = \lambda^*$ is real analytic with $\mathrm{Hess}(J)(\mathsf{v})$ being positive definite.

The local limit estimates we have derived reads as: Recall notation $\mathsf{u}_n = \lfloor n\mathsf{u}\rfloor/n$ and ξ_n being defined via $\mathsf{u}_n = \nabla\lambda(\xi_n)$. Then,

$$\mathbb{Q}_n\left(\lfloor n\mathsf{u}\rfloor\right) = \frac{\mu(0)}{\mu(\xi_n)\sqrt{(2\pi)^d\det\mathit{\Xi}(\xi_n)}}\mathrm{e}^{-nJ(\mathsf{u}_n)}\left(1+\mathrm{o}\left(1\right)\right), \tag{97}$$

uniformly in $\mathsf{u}\in\mathbb{B}^d_\kappa(\mathsf{v})$.

3.2 Ballistic Phase of Annealed Polymers

Recall that the reference polymer weights W_d are given by (3). $\mathit{\Phi}$ is the self-interaction potential (4) which satisfies the attractivity condition (26). Let $h \notin \mathbf{K}_0$ and, accordingly, $h \in \partial\mathbf{K}_\lambda$ with $\lambda = \lambda(h) > 0$. We shall consider the normalized weights

$$\mathsf{W}_d^{h,\lambda}(\gamma) = \mathrm{e}^{h\cdot\mathsf{X}(\gamma)-\lambda|\gamma|}\mathsf{W}_d(\gamma). \tag{98}$$

These weight are normalized for the following reason: As before define

$$\mathscr{P}_{\mathsf{x}} = \{\gamma\ :\ \mathsf{X}(\gamma) = \mathsf{x}\}\quad \mathscr{P}_n = \{\gamma\ :\ |\gamma| = n\}\quad \text{and}\quad \mathscr{P}_{\mathsf{x},n} = \mathscr{P}_{\mathsf{x}}\cap\mathscr{P}_n. \tag{99}$$

Then,

$$\sum_{\gamma\in\mathscr{P}_{\mathsf{x}}} \mathsf{W}_d^{h,\lambda}(\gamma) = \mathrm{e}^{h\cdot\mathsf{x}} G_\lambda(\mathsf{x}) \asymp \mathrm{e}^{h\cdot\mathsf{x}-\tau_\lambda(\mathsf{x})} \text{ and } \sum_{\gamma\in\mathscr{P}_n} \mathsf{W}_d^{h,\lambda}(\gamma) = \mathrm{e}^{-n\lambda} Z_n(h) \asymp 1. \tag{100}$$

If $h\cdot\mathsf{x} = \tau_\lambda(\mathsf{x})$, then the first term in (100) is also of order 1. More generally, let us define the following crucial notion:

Surcharge Function For any $\mathsf{x}\in\mathbb{R}^d$ define the surcharge function

$$\mathsf{s}_h(\mathsf{x}) = \tau_\lambda(\mathsf{x}) - h\cdot\mathsf{x} \geq 0. \tag{101}$$

In view of (40) the first of the estimates in (100) could be upgraded as follows (see (34) for the definition of A_d)

$$\mathsf{W}_d^{h,\lambda}(\mathsf{x}) \stackrel{\Delta}{=} \sum_{\gamma\in\mathscr{P}_{\mathsf{x}}} \mathsf{W}_d^{h,\lambda}(\gamma) \asymp \mathrm{e}^{-\mathsf{s}_h(\mathsf{x})} \text{ and, moreover, } \mathsf{W}_d^{h,\lambda}(\mathsf{x}) \stackrel{<}{\sim} \mathsf{A}_d(k_0\,|\mathsf{x}|_1)\mathrm{e}^{-\mathsf{s}_h(\mathsf{x})}. \tag{102}$$

Surcharge Cone Let us say that $\mathscr{Y}_1$ is a δ_1-surcharge cone with respect to h if:

(a) $\mathscr{Y}_1$ is a positive cone (meaning that its opening is strictly less than π) and it contains a lattice direction $\pm\mathsf{e}_k$ in its interior.
(b) For any $\mathsf{x}\notin\mathscr{Y}_1$ the surcharge function $\mathfrak{s}$ satisfies

$$\mathfrak{s}(\mathsf{x}) = \tau_\lambda(\mathsf{x}) - h\cdot\mathsf{x} > \delta_1\tau_\lambda(\mathsf{x}). \tag{103}$$

For the rest of this section we shall fix $\delta_1\in(0,1)$ and a δ_1-surcharge cone $\mathscr{Y}_1$ with respect to h.

Factorization Bound Assume that the path γ can be represented as a concatenation,

$$\gamma \;=\; \gamma_0\circ\eta_1\circ\gamma_1\circ\cdots\circ\gamma_m\circ\eta_{m+1}, \tag{104}$$

such that paths $\gamma_\ell = (\mathsf{u}_\ell,\ldots,\mathsf{v}_{\ell+1})$ satisfy the following two properties:

(P1) γ_i is disjoint from γ_j for all $i>j$.
(P2) For any i the local time $\ell_{\gamma_i}(\mathsf{v}_{i+1}) = 1$.

By (29) and **(P1)**,

$$\Phi_\beta(\gamma) \geq \Phi_\beta(\gamma_1\cup\cdots\cup\gamma_m) = \sum_\ell \Phi_\beta(\gamma_\ell).$$

Consequently,

$$\mathsf{W}_d(\gamma)\,\mathrm{e}^{-\lambda|\gamma|} \le \prod_{l=1}^{m} \mathsf{W}_d(\gamma_\ell)\mathrm{e}^{-\lambda|\gamma_\ell|} \cdot \prod_{k=1}^{m} \mathrm{e}^{-\lambda|\eta_k|}. \tag{105}$$

Fixing end points $\mathsf{u}_1, \mathsf{v}_2, \mathsf{u}_2, \dots$ and paths η_ℓ in (104), and summing up with respect to all paths $\gamma_1, \dots, \gamma_m$ (with $\gamma_\ell = (\mathsf{u}_\ell, \dots, \mathsf{v}_{\ell+1})$) satisfying properties **(P1)** and **(P2)** above we derive the following upper bound:

$$\sum_{\gamma_1,\dots,\gamma_m} \mathsf{W}_d(\gamma)\,\mathrm{e}^{-\lambda|\gamma|} \le \prod_{1}^{m} H_\lambda(\mathsf{v}_{\ell+1} - \mathsf{u}_\ell)\mathrm{e}^{-\lambda\sum|\eta_\ell|} \le \mathrm{e}^{-\sum\tau_\lambda(\mathsf{v}_{\ell+1}-\mathsf{u}_\ell)-\lambda\sum|\eta_\ell|}. \tag{106}$$

Let us proceed with describing our algorithm to construct representation (104) with properties **(P1)** and **(B2)** for any path $\gamma \in \mathscr{P}_{\mathsf{x}}$.

Construction of Skeletons Skeletons $\hat{\gamma}_K$ are constructed as a collection $\hat{\gamma}_K = [\mathfrak{t}_K, \mathfrak{h}_K]$, where $\mathfrak{t}_K$ is the trunk and $\mathfrak{h}_K$ is the set of hairs of $\hat{\gamma}_K$. Let $\gamma \in \mathscr{P}_x$ and choose a scale K. In the sequel we use $\mathbf{U}_\lambda^K = \{\mathsf{u} : \tau_\lambda(\mathsf{u}) \le K\}$ denote the ball of radius K with respect to τ_λ (note that since, in general, $\tau_\lambda(\mathsf{y}) \ne \tau_\lambda(-\mathsf{y})$, it does not have to be a distance). Recall that R denotes the range of the underlying random walk. Choose $r = r_\lambda = \min\{s : \mathbb{B}_R^d \subset \mathbf{U}_\lambda^s\}$, where as before $\mathbb{B}_R^d$ is the Euclidean ball of radius R. Let us first explain decomposition (104) and construction of trunks (see Fig. 3).

STEP 0. Set $\mathsf{u}_0 = 0$, $\tau_0 = 0$ and $\mathfrak{t}_0 = \{u_0\}$. Go to STEP 1.
STEP (l+1) If $(\gamma(\tau_l), \dots \gamma(n)) \subseteq \mathbf{U}_\lambda^K(u_l)$ then set $\sigma_{l+1} = n$ and stop. Otherwise, define

$$\sigma_{l+1} = \min\{i > \tau_l \ : \ \gamma(i) \notin \mathbf{U}_\lambda^K(\mathsf{u}_l)\}$$

and

$$\tau_{l+1} = 1 + \max\{i > \tau_l \ : \ \gamma(i) \in \mathbf{U}_\lambda^{K+r}(\mathsf{u}_l)\}.$$

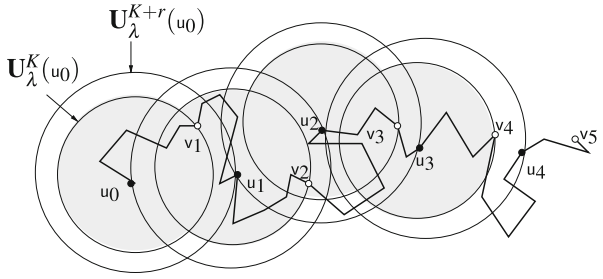

Fig. 3 Decomposition (104) and construction of the trunk $\mathfrak{t}_K = \{\mathsf{u}_1, \dots, \mathsf{u}_4\}$ of the skeleton

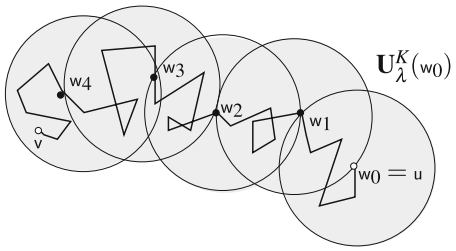

Fig. 4 Construction of a hair $\mathfrak{h}_K = \{w_1, \dots, w_4\}$

Set $v_{l+1} = \gamma(\sigma_{l+1})$ and $u_{l+1} = \gamma(\tau_{l+1})$. Update $\mathfrak{t}_K = \mathfrak{t}_K \cup \{u_{l+1}\}$ and go to STEP (l+2) □

Clearly the above algorithm leads to a decomposition of γ as in (104) with

$$\gamma_l = (\gamma(\tau_l), \dots, \gamma(\sigma_{l+1})) \quad \text{and} \quad \eta_l = (\gamma(\sigma_l), \dots, \gamma(\tau_l)),$$

and with $\gamma_1, \gamma_2, \dots$ satisfying conditions **(P1)** and **(P2)**.

The set $\mathfrak{t}_K$ is called the trunk of the skeleton $\hat{\gamma}_K$ of γ on Kth scale. The hairs $\mathfrak{h}_K$ of $\hat{\gamma}_K$ take into account those η_l-s which are long on Kth scale. Recall that $\eta_l : v_l \mapsto u_l$. It is equivalent, but, since eventually we want to keep track of vertices from the trunk $\mathfrak{t}_K$, more convenient to think about η_l as of a reversed path from u_l to v_l. Then the lth hair $\mathfrak{h}_K^l = \mathfrak{h}_K[\eta_l]$ of γ is constructed as follows:

If $\eta_l \subseteq \mathbf{U}_\lambda^K(u_l)$ then $\mathfrak{h}_K^l = \emptyset$. Otherwise, set $u = u_l$, $v = v_l$, $\eta = \eta_l$, $m = |\eta|$, and proceed with the following algorithm (see Fig. 4):

STEP 0. Set $w_0 = u$, $\tau_0 = 0$ and $\mathfrak{h}_K[\eta] = \emptyset$. Go to STEP 1.

STEP (l+1). If $(\eta(\tau_l), \dots, \eta(m)) \subseteq \mathbf{U}_\lambda^K(u_l)$ then stop. Otherwise set

$$\tau_{l+1} = \min\left\{j > \tau_l \,:\, \eta(j) \notin \mathbf{U}_\lambda^K(w_l)\right\}.$$

Define $w_{l+1} = \eta(\tau_{l+1})$, update $\mathfrak{h}_K = \mathfrak{h}_K \cup \{w_{l+1}\}$ and go to STEP (l+2). □

Control of $\mathbf{W}_d^{h,\lambda}(\hat{\gamma}_K)$ In the super-critical case $\lambda > 0$, and hairs could be controlled in a crude fashion via comparison with an underlying walk killed at rate λ.

Exercise 3.15 There exists $\epsilon = \epsilon(\mathsf{P}_d, \lambda) > 0$, such that

$$\sum_{\gamma \in \mathscr{P}_u} \mathrm{e}^{-\epsilon|\gamma|} \mathsf{P}_d(\gamma) \lesssim \mathrm{e}^{-\epsilon\tau_\lambda(u)} \tag{107}$$

uniformly in u.

Let $\hat{\gamma}_K = (\mathfrak{t}_K, \mathfrak{h}_K)$ be a skeleton. We shall carefully control the geometry of the trunk $\mathfrak{t}_K = (u_0, \dots, u_N)$. As far as the hairs are considered, we shall rely on (107), and the only thing we shall control is $\#(\mathfrak{h}_K)$—the total number of K-increments (in τ_λ-metrics).

Let us fix $\mathsf{u}_0, \mathsf{v}_1, \mathsf{u}_2, \ldots$. By (106) the contribution coming from $\gamma_\ell : \mathsf{u}_\ell \to \mathsf{v}_{\ell+1}$ paths is bounded above by the product of $\mathrm{e}^{-\tau_\lambda(\mathsf{v}_{\ell+1}-\mathsf{u}_\ell)}$. Now, by construction $\mathsf{v}_{\ell+1} \in \mathbf{U}_\lambda^{K+r_\lambda}(\mathsf{u}_\ell) \setminus \mathbf{U}_\lambda^K(\mathsf{u}_\ell)$, which means that $\tau_\lambda(\mathsf{v}_{\ell+1}-\mathsf{u}_\ell) \in [K, K+r_\lambda]$. On the other hand if $\mathsf{u}_{\ell+1}$ is defined, then $\mathsf{u}_{\ell+1} \in \mathbf{U}_\lambda^{K+2r_\lambda}(\mathsf{u}_\ell) \setminus \mathbf{U}_\lambda^{K+r_\lambda}(\mathsf{u}_\ell)$. Hence, $\tau_\lambda(\mathsf{u}_{\ell+1}-\mathsf{u}_\ell) \le \tau_\lambda(\mathsf{v}_{\ell+1}-\mathsf{u}_\ell) + 2r_\lambda$. There are $\lesssim RK^{d-1}$ possible exit points from $\mathbf{U}_\lambda^K(\mathsf{v}_\ell)$-balls which are possible candidates for $\mathsf{v}_{\ell+1}$ vertices.

Consequently, (106) and (107) imply that there exists $c = c(\lambda, \beta) > 0$ such that the following happens: Let $\hat\gamma_K = (\mathfrak{t}_K, \mathfrak{h}_K)$ be a skeleton with trunk $\mathfrak{t}_K = (\mathsf{u}_0, \ldots, \mathsf{u}_N)$, and $\mathfrak{h}_K = \{\mathfrak{h}_K^\ell\}$ collection of hairs. Notation $\gamma \sim \hat\gamma_K$ means that $\hat\gamma_K$ is the K-skeleton of γ in the sense of the two algorithms above. Then,

$$\sum_{\gamma\sim\hat\gamma_K} \mathsf{W}_d^{h,\lambda}(\gamma) \lesssim \exp\left\{h\cdot\mathsf{x} - \sum_{\ell=0}^N \tau_\lambda(\mathsf{u}_{\ell+1}-\mathsf{u}_\ell) - \epsilon K \#(\mathfrak{h}_K) + cN\log K\right\}, \tag{108}$$

uniformly in x, large enough scales K and skeletons $\hat\gamma_K$.

Kesten's Bound on the Number of Forests A forest $\mathscr{F}_N$ is a collection of N rooted trees $\mathscr{F}_N = (\mathscr{T}_1, \ldots, \mathscr{T}_N)$ of forward branching ratio at most b. The tree $\mathscr{F}_\ell$ is rooted at u_ℓ. Given $M \in \mathbb{N}$ we wish to derive an upper bound on $\#(M,N)$-number of all forests $\mathscr{F}_N$ satisfying $|\mathscr{F}_N| = M$. Above $|\mathscr{F}_N| = M$ is the number of vertices of $\mathscr{F}_N$ different from the roots $\mathsf{u}_1, \ldots, \mathsf{u}_N$. Let $\mathbb{P}_p^N$ be the product percolation measure on $\times\mathsf{T}_\ell^{\mathsf{b}}$ at the percolation value p, where T^{b} is the set of (edge) percolation configurations on the rooted tree of branching ration b. In this way $\mathscr{T}_\ell$ is viewed as a connected component of u_ℓ. Clearly,

$$\mathbb{P}_p^N(|\mathscr{F}_N| = M) \le 1. \tag{109}$$

Each realization of $\mathscr{F}_N$ with $|\mathscr{F}_N| = M$ has probability which is bounded below by $p^M(1-p)^{b(N+M)}$. Therefore, (109) implies:

$$\#(M,N) \le \left(\max_{p\in[0,1]} p^M(1-p)^{b(N+M)}\right)^{-1}. \tag{110}$$

For $x \in [0,1]$,

$$\log(1-x) = -\int_0^x \frac{\mathrm{d}t}{1-t} \ge -\frac{x}{1-x}$$

Choosing $p = \frac{1}{b}$ we, therefore, infer from (110)

$$\#(M,N) \le \mathrm{e}^{M\log b + (N+M)\frac{b}{b-1}}. \tag{111}$$

Surcharge Cost of a Skeleton A substitution of (111) with $b \cong RK^{d-1}$ into (108) implies: There exist $\epsilon', c' > 0$, such that

$$\sum_{\substack{\gamma \sim \mathfrak{t}_K \\ \#(\mathfrak{h}_K) \geq M}} \mathsf{W}_d^{h,\lambda}(\gamma) \lesssim \exp\left\{ h \cdot \mathsf{x} - \sum_{\ell=0}^{N} \tau_\lambda(\mathsf{u}_{\ell+1} - \mathsf{u}_\ell) - \epsilon' KM + c' N \log K \right\}. \tag{112}$$

uniformly in x, scales K, trunks $\mathfrak{t}_K$ and $M \in \mathbb{N}$.

With (112) in mind let us define the surcharge cost of a skeleton $\hat{\gamma}_K = [\mathfrak{t}_K, \mathfrak{h}_K]$ as follows: Recall the notation $\mathfrak{s}_h(\mathsf{u}) = \tau_\lambda(\mathsf{u}) - h \cdot \mathsf{u}$. Then,

$$\mathfrak{s}_h(\hat{\gamma}_K) \stackrel{\Delta}{=} \sum_{\ell=1}^{N} \mathfrak{s}_h(\mathsf{u}_\ell - \mathsf{u}_{\ell-1}) + \epsilon' K \#(\mathfrak{h}_K). \tag{113}$$

Above the trunk $\mathfrak{t}_K = (\mathsf{u}_0, \dots \mathsf{u}_N)$. We conclude:

Lemma 3.16 *For any $\epsilon > 0$ there exists a scale K_0, such that*

$$\sum_{\mathfrak{s}_h(\hat{\gamma}_K) > 2\epsilon|\mathsf{x}|} \mathsf{W}_d^{h,\lambda}(\hat{\gamma}_K) \lesssim \mathrm{e}^{-\epsilon|\mathsf{x}|}, \tag{114}$$

for all $K \geq K_0$ fixed and uniformly in $h \in \partial\mathbf{K}_\lambda$, $\mathsf{x} \in \mathbb{Z}^d$. By convention the summation above is with respect to skeletons $\hat{\gamma}_K$ of paths $\gamma \in \mathscr{P}_\mathsf{x}$.

Cone Points of Skeletons Recall the definition of the surcharge cone $\mathscr{Y}_1$ in (103). Fix $\delta_2 \in (\delta_1, 1)$ and $\delta_3 \in (\delta_2, 1)$, and define enlargements $\mathscr{Y}_i$ of $\mathscr{Y}_1$ as follows: For $i = 2, 3$,

$$\mathscr{Y}_i = \{\mathsf{x} \,:\, \mathfrak{s}(\mathsf{x}) \leq \delta_i \tau_\lambda(\mathsf{x})\} \tag{115}$$

Clearly, $\mathscr{Y}_i$-s are still positive cones for $i = 2, 3$. Let $A^0 \stackrel{\Delta}{=} A \setminus 0$. Then, by construction, $\mathscr{Y}_1^0 \subset \mathrm{int}\left(\mathscr{Y}_2^0\right)$ and $\mathscr{Y}_2^0 \subset \mathrm{int}\left(\mathscr{Y}_3^0\right)$.

Consider a skeleton $\hat{\gamma}_K = [\mathfrak{t}_K, \mathfrak{h}_K]$. Let us say that a vertex of the trunk $\mathsf{u}_\ell \in \mathfrak{t}_K = (\mathsf{u}_0, \dots, \mathsf{u}_\ell, \dots, \mathsf{u}_{m+1})$ is a $\mathscr{Y}_2$-cone point of the skeleton $\hat{\gamma}_K$ if

$$\hat{\gamma}_K \subset (\mathsf{u}_\ell - \mathscr{Y}_2) \cup (\mathsf{u}_\ell + \mathscr{Y}_2). \tag{116}$$

Let $\#_{\mathrm{bc}}(\hat{\gamma}_K)$ be the total number of vertices of $\hat{\gamma}_K$ which are *not* cone points.

Proposition 3.17 *There exists $\nu_2 > 0$ such that the following happens: For any $\epsilon > 0$ there exists a scale K_0, such that*

$$\sum_{\#_{\mathrm{bc}}(\hat{\gamma}_K) > \epsilon \frac{|\mathsf{x}|}{K}} \mathsf{W}_d^{h,\lambda}(\hat{\gamma}_K) \lesssim \mathrm{e}^{-\nu_2|\mathsf{x}|}, \tag{117}$$

for all $K \geq K_0$ fixed and uniformly in $\mathsf{x} \in \mathbb{Z}^d$.

A proof of Proposition 3.17 contains several steps and we refer to [14] for more details. First of all we show that, up to exponentially small corrections, most of the vertices of the trunk $\mathfrak{t}_K$ are $\mathscr{Y}_1$-cone points of the latter. We shall end up with $N' \cong \frac{|\mathsf{x}|}{K}$ $\mathscr{Y}_1$-cone points of $\mathfrak{t}_K$.

Any $\mathscr{Y}_1$-cone point of the trunk $\mathfrak{t}_K$ is evidently also a $\mathscr{Y}_2$-cone point of the latter. On the other hand, in view of (114), we can restrict attention to $\#(\mathfrak{h}_K) \le \frac{2\epsilon}{\epsilon'}\frac{|\mathsf{x}|}{K}$. For $\epsilon \ll \epsilon'$ the total number of leaves $\#(\mathfrak{h}_K)$ is only a small fraction of N'. It is clear that an addition of a leave is capable of blocking at most $c = c(\delta_1, \delta_2)$ $\mathscr{Y}_1$-cone points of $\mathfrak{t}_K$ from being a $\mathscr{Y}_2$-cone point of the whole skeleton $\hat{\gamma}_K$. It is important that the above geometric constant $c = c(\delta_1, \delta_2)$ does not depend on the running scale K. Consequently, under the reduction we are working with on large enough scales K, there are just not enough leaves to block all (and actually a small fraction of) $\mathscr{Y}_1$-cone points of $\mathfrak{t}_K$ from being a $\mathscr{Y}_2$-cone point of the whole skeleton $\hat{\gamma}_K$.

Cone Points of Paths $\gamma \in \mathscr{P}_{\mathsf{x}}$ Let us say that $\mathsf{u}_\ell \in \gamma = (\mathsf{u}_0, \dots, \mathsf{u}_\ell, \dots, \mathsf{u}_n) \in \mathscr{P}_{\mathsf{x}}$ is a cone point of γ if $0 < \ell < n$ and

$$\gamma \subset (\mathsf{u}_\ell - \mathscr{Y}_3) \cup (\mathsf{u}_\ell + \mathscr{Y}_3)\,. \tag{118}$$

Let $\#_{\text{cone}}(\gamma)$ be the total number of the cone points of γ.

Proposition 3.18 *There exist $\epsilon > 0$ and $\nu > 0$ such that:*

$$\sum_{\#_{\text{cone}}(\gamma) < \epsilon|\mathsf{x}|} \mathsf{W}_d^{h,\lambda}(\gamma) \lesssim \mathrm{e}^{-\nu|\mathsf{x}|}, \tag{119}$$

uniformly in $\mathsf{x} \in \mathbb{Z}^d$.

As before, we refer to [14] for details of the proof. Construction of cone points is depicted on Fig. 5

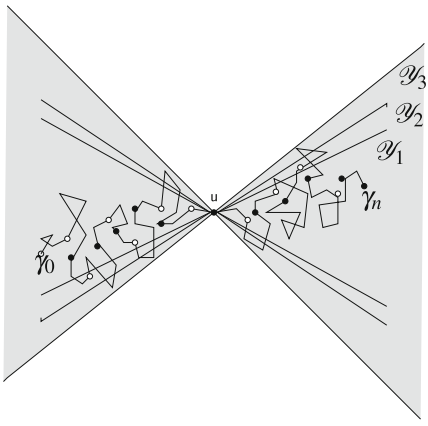

Fig. 5 u is a cone point of the path $\gamma = (\gamma_0, \dots, \gamma_n)$. *Black vertices* belong to the trunk. Paths leading from *white vertices* to *black vertices* give rise to hairs

Irreducible Decomposition of Paths in $\mathscr{P}_{\mathsf{x}}$, $\mathscr{P}_n$ and $\mathscr{P}_{\mathsf{x},n}$ In the sequel we set $\mathscr{Y} \stackrel{\Delta}{=} \mathscr{Y}_3$, where $\mathscr{Y}_3$ is the positive cone in Proposition 3.18. A path $\gamma = (\mathsf{u}_0, \ldots, \mathsf{u}_n)$ is said to be irreducible if it does not contain $\mathscr{Y}$-cone points. We shall work with three sub-families $\mathscr{F}^{[l]}$, $\mathscr{F}^{[r]}$ and $\mathscr{F} = \mathscr{F}^{[l]} \cap \mathscr{F}^{[r]}$ of irreducible paths. Those are defined as follows:

$$\mathscr{F}^{[l]} = \{\gamma \text{ irreducible} : \gamma \subset \mathsf{u}_n - \mathscr{Y}\}\,, \ \mathscr{F}^{[r]} = \{\gamma \text{ irreducible} : \gamma \subset \mathsf{u}_0 + \mathscr{Y}\}\,.$$

Note that any $\gamma = (\mathsf{u}_0, \ldots, \mathsf{u}_n) \in \mathscr{F}$ is automatically confined to the diamond shape

$$\gamma \subset D(\mathsf{u}_0, \mathsf{u}_n) \stackrel{\Delta}{=} (\mathsf{u}_0 + \mathscr{Y}) \cap (\mathsf{u}_n - \mathscr{Y})\,. \tag{120}$$

Proposition 119 implies that up to corrections of order $\mathrm{e}^{-\nu|\mathsf{x}|}$ one can restrict attention to paths $\gamma \in \mathscr{P}_{\mathsf{x}}$ which have the following decomposition into irreducible pieces (see Fig. 6):

$$\gamma = \gamma^{[l]} \circ \gamma^1 \circ \cdots \circ \gamma^N \circ \gamma^{[r]}. \tag{121}$$

Define

$$\mathsf{f}^{[l]}(\mathsf{x}, n) = \sum_{\substack{\mathsf{X}(\gamma)=\mathsf{x}, |\gamma|=n \\ \gamma \in \mathscr{F}^{[l]}}} \mathsf{W}_d^{h,\lambda}(\gamma) \text{ and } \mathsf{f}^{[r]}(\mathsf{x}, n) = \sum_{\substack{\mathsf{X}(\gamma)=\mathsf{x}, |\gamma|=n \\ \gamma \in \mathscr{F}^{[r]}}} \mathsf{W}_d^{h,\lambda}(\gamma). \tag{122}$$

Theorem 3.19 *The weights $\mathsf{f}^{[l]}$ and $\mathsf{f}^{[r]}$ have exponentially decaying tails: There exists $\nu > 0$ and, for every $\lambda > 0$, $\chi_\lambda > 0$ such that the following mass gap estimate holds uniformly in x and n:*

$$\mathsf{f}^{[l]}(\mathsf{x}, n), \mathsf{f}^{[r]}(\mathsf{x}, n) \stackrel{<}{\sim} \mathrm{e}^{-\nu|\mathsf{x}|-\chi_\lambda n}. \tag{123}$$

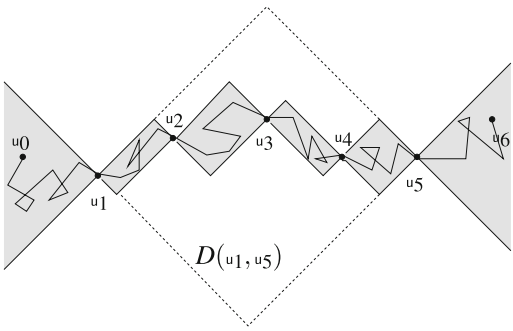

Fig. 6 Irreducible decomposition $\gamma = \gamma^{[l]} \circ \gamma^1 \circ \cdots \circ \gamma^4 \circ \gamma^{[r]}$. The path $\gamma^{[l]} : \mathsf{u}_0 \mapsto \mathsf{u}_1$ belongs to $\mathscr{F}^{[l]}$, and the path $\gamma^{[r]} : \mathsf{u}_5 \mapsto \mathsf{u}_6$ belongs to $\mathscr{F}^{[r]}$. Paths $\gamma_\ell : \mathsf{u}_\ell \mapsto \mathsf{u}_{\ell+1}$ belong to $\mathscr{F}$. Each irreducible γ_ℓ stays inside the diamond shape $D(\mathsf{u}_\ell, \mathsf{u}_{\ell+1})$, and the concatenation $\gamma_1 \circ \cdots \circ \gamma_4 \subset D(\mathsf{u}_1, \mathsf{u}_5)$

Furthermore, for each $\lambda = \lambda(h) > 0$, $\mathsf{W}_d^{h,\lambda}$ is a probability distribution on $\mathscr{F}$. In particular, the family of weights $\{\mathsf{f}(\mathsf{x}, \mathsf{n})\}$,

$$\mathsf{f}(\mathsf{x}, n) = \sum_{\substack{\mathsf{X}(\gamma)=\mathsf{x}, |\gamma|=n \\ \gamma \in \mathscr{F}}} \mathsf{W}_d^{h,\lambda}(\gamma) \stackrel{\Delta}{=} \sum_{\gamma \in \mathscr{F}_{\mathsf{x},n}} \mathsf{W}_d^{h,\lambda}(\gamma) \tag{124}$$

is a probability distribution on $\mathbb{Z}^d \times \mathbb{N}$ with exponentially decaying tails.

Remark 3.20 Note that exponential decay in n is claimed only if $\lambda > 0$. $\lambda = 0$ corresponds to the case of critical drifts $h \in \partial \mathbf{K}_0$. For critical drifts, the decay in n is sub-exponential and, furthermore, the whole coarse-graining (skeleton construction) procedure should be modified. It happens, nevertheless, that the decay in x is still exponential. We do not discuss critical case in these lecture notes, and refer to [17].

Proof For fixed $\lambda > 0$ bounds $\mathsf{f}^{[l]}(\mathsf{x}, n), \mathsf{f}^{[r]}(\mathsf{x}, n) \stackrel{<}{\sim} \mathrm{e}^{-\nu|\mathsf{x}|}$ directly follow from Proposition 3.18. The case to work out is when $|\mathsf{x}|$ is much smaller than n, say $|\mathsf{x}| < \frac{\epsilon}{\lambda} n$. But then $\mathrm{e}^{h\cdot\mathsf{x}-\lambda n} \leq \mathrm{e}^{-(\lambda-\epsilon)n}$. Since by (40) the two-point function $G_0(\cdot)$ is bounded, the decay in n indeed comes for free as long as $\lambda > 0$.

In order to see that $\{\mathsf{f}(\mathsf{x}, n)\}$ is a probability distribution recall that $\mathbf{K}_\lambda$ was characterized as the closure of the domain of convergence of $h \mapsto \sum_\gamma \mathrm{e}^{h\mathsf{x}-\lambda n} \mathsf{W}_d(\gamma)$. Thinking in terms of (121), and in view of (123) this necessarily implies that $\sum \mathsf{f}(\mathsf{x}, n) = 1$.

Local Geometry of $\partial \mathbf{K}_\lambda$ and Analyticity of $\boldsymbol{\lambda}$ Inspecting construction of the cone $\mathscr{Y}$ for $h \in \partial \mathbf{K}_\lambda$ we readily infer that the very same $\mathscr{Y}$ would do for all drifts $g \in \left(h + \mathbb{B}_\epsilon^d\right) \cap \partial \mathbf{K}_\lambda$, for some $\epsilon > 0$ sufficiently small. Similarly, it would do for all $|\mu - \lambda|$ sufficiently small. We are in the general renewal framework of (77). The following theorem is a consequence of (78), Lemma 3.12 of Sect. 3.1 and (202) of the Appendix:

Theorem 3.21 *There exists $\epsilon = \epsilon_\lambda > 0$, such that the following happens: Let h be a super-critical drift; $\lambda(h) = \lambda > 0$. Construct $\mathscr{Y}$ and, accordingly, $\{f(\mathsf{x}, n)\}$ as in Theorem 3.19. Then for $(g, \mu) \in \mathbb{B}_\epsilon^{d+1}(h, \mu)$,*

$$\mu = \lambda(g) \;\Leftrightarrow\; \sum_{\gamma \in \mathscr{F}} \mathsf{W}_d^{g,\mu}(\gamma) = \sum_{\mathsf{x},n} \mathrm{e}^{(g-h)\cdot\mathsf{x}+(\lambda-\mu)n} \mathsf{f}(\mathsf{x}, n) = 1. \tag{125}$$

As a result, $\lambda(\cdot)$ is real analytic on $\mathbb{B}_\epsilon^d(h)$ and $\Xi(h) = \mathrm{Hess}(\lambda)(h)$ is non-degenerate.

In particular,

$$g \in \left(h + \mathbb{B}_\epsilon^d\right) \cap \partial \mathbf{K}_\lambda \;\Leftrightarrow\; \sum_{\mathsf{x},n} \mathrm{e}^{(g-h)\mathsf{x}} \mathsf{f}(\mathsf{x}, n) = 1. \tag{126}$$

As a result, $\partial \mathbf{K}_\lambda$ is locally analytic and has a uniformly positive Gaussian curvature.

Ornstein-Zernike Theory Theorem 3.19 paves the way for an application of the multidimensional renewal theory, as described in Sect. 3.1 to a study of various limit properties of annealed measures $\mathbb{P}_n^h$, whenever $h \notin \mathbf{K}_0$ is a super-critical drift. For the rest of the section let us fix such h and $\lambda = \lambda(h) > 0$. By the above, this generates a cone $\mathscr{Y}$ and a probability distribution $\{\mathsf{f}(\mathsf{x}, n)\}$ with exponentially decaying tails. We declare that it is a probability distribution of a random vector $\mathsf{U} = (\mathsf{X}, \mathsf{T}) \in \mathbb{Z}^d \times \mathbb{N}$. In view of our assumptions on the underlying random walk, U satisfies the non-degeneracy Assumption 3.7, which means that it is in the framework of the theory developed therein.

Let us construct the array $\{\mathsf{t}(\mathsf{x}, n)\}$ via the renewal relation (73). The number $\mathsf{t}(\mathsf{x}, n)$ has the following meaning: Recall our definition $D(\mathsf{x}, \mathsf{y}) = (\mathsf{x} + \mathscr{Y}) \cap (\mathsf{y} - \mathscr{Y})$ of diamond shapes, and define the following three families of diamond-confined paths:

$$\mathscr{T}_{\mathsf{x}} = \{\gamma \in \mathscr{P}_{\mathsf{x}} : \gamma \subset D(0, \mathsf{x})\}\,, \ \mathscr{T}_n = \{\gamma \in \mathscr{P}_n : \gamma \subset D(0, \gamma_n)\} \text{ and } \mathscr{T}_{\mathsf{x},n} = \mathscr{T}_{\mathsf{x}} \cap \mathscr{T}_n. \tag{127}$$

As before, $\mathsf{t}(\mathsf{x}) = \sum_n \mathsf{t}(\mathsf{x}, n)$ and $\mathsf{t}(n) = \sum_{\mathsf{x}} \mathsf{t}(\mathsf{x}, n)$. Then,

$$\mathsf{t}(\mathsf{x}, n) = \sum_{\gamma \in \mathscr{T}_{\mathsf{x},n}} \mathsf{W}_d^{h,\lambda}(\gamma),\ \mathsf{t}(\mathsf{x}) = \sum_{\gamma \in \mathscr{T}_{\mathsf{x}}} \mathsf{W}_d^{h,\lambda}(\gamma) \text{ and } \mathsf{t}(n) = \sum_{\gamma \in \mathscr{T}_n} \mathsf{W}_d^{h,\lambda}(\gamma). \tag{128}$$

Asymptotics of Partition Functions By Theorem 3.19, the partition function $Z_n(h)$ in (7) satisfies:

$$\mathrm{e}^{-n\lambda} Z_n(h) = \mathsf{W}_d^{h,\lambda}(\mathscr{P}_n) = \mathrm{O}(\mathrm{e}^{-\chi_\lambda n}) + \sum_{k+m+j=n} \mathsf{f}^{[l]}(k)\mathsf{t}(m)\mathsf{f}^{[r]}(j). \tag{129}$$

Define $\kappa(h) = \left(\sum_k \mathsf{f}^{[l]}(k)\right)\left(\sum_j \mathsf{f}^{[r]}(j)\right)$ and $\mu(h) = \sum n\mathsf{f}(n) = \mathbb{E}\mathsf{T}$. By Lemma 3.3

$$\lim_{n\to\infty} \mathrm{e}^{-n\lambda(h)} Z_n(h) = \frac{\kappa(h)}{\mu(h)}, \tag{130}$$

exponentially fast.

Limiting Spatial Extension and Other Limit Theorems Since λ is differentiable at any $h \notin \mathbf{K}_0$, (9) and LLN (50) follow with $\mathsf{v} = \nabla\lambda(h)$. However, since for any $h \notin \mathbf{K}_0$ the probability distribution $\{\mathsf{f}(\mathsf{x}, \mathsf{n})\}$ in (124) has exponential tails much sharper local limit results follow along the lines of Sect. 3.1: Let $g \notin \mathbf{K}_0$ and $\mathsf{u} = \nabla\lambda(g)$. Fix δ sufficiently small and consider $\mathbb{B}_\delta^d(\mathsf{u})$. For $\mathsf{w} \in \mathbb{B}_\delta^d(\mathsf{u})$ define $\mathsf{w}_n = \lfloor n\mathsf{w} \rfloor / n$ and let g_n being defined via $\mathsf{w}_n = \nabla\lambda(g_n)$.

Theorem 3.22 *There exists a positive real analytic function ψ on $\mathbb{B}^d_\delta(\mathsf{u})$ such that*

$$\mathbb{P}^h_n(\lfloor n\mathsf{w}\rfloor) = \frac{\psi(\mathsf{w})}{\sqrt{(2\pi)^d \det \Xi(g_n)}} \mathrm{e}^{-nI_h(\mathsf{w}_n)}(1+\mathrm{o}(1)), \tag{131}$$

uniformly in $\mathsf{w} \in \mathbb{B}^d_\delta(\mathsf{u})$.

In particular (considering $\mathsf{u} = \mathsf{v}$), under $\mathbb{P}^h_n$ the distribution of the rescaled end-point $\frac{\mathsf{X}-n\mathsf{v}}{\sqrt{n}}$ converges to the d-dimensional mean-zero normal distribution with covariance matrix $\Xi(h) = \mathrm{Hess}(\lambda)(h)$.

Let us turn to the (Ornstein-Zernike) asymptotics of the two point function G_λ. Let $\mathsf{x} \neq 0$ and $h = \nabla\tau_\lambda(\mathsf{x})$, that is $h \in \partial\mathbf{K}_\lambda$ and $\tau_\lambda(\mathsf{x}) = h \cdot \mathsf{x}$. Since, as we already know, $\partial\mathbf{K}_\lambda$ is strictly convex, such h is unambiguously defined. Under the weights $\mathsf{W}^{h,\lambda}_d$ the irreducible decomposition (121) folds in the sense that the $\mathsf{W}^{h,\lambda}_d$-weight of all paths $\gamma \in \mathscr{P}_\mathsf{x}$ which do not comply with it, is exponentially negligible as compared to $G_\lambda(\mathsf{x})$. Hence:

$$\mathrm{e}^{\tau_\lambda(\mathsf{x})} G_\lambda(\mathsf{x})(1+\mathrm{o}(1)) = \sum_n \mathsf{t}(\mathsf{x}, n) = \frac{c(h)}{\sqrt{|\mathsf{x}|^{d-1}}}(1+\mathrm{o}(1)), \tag{132}$$

asymptotically in x large. This follows from (131) and Gaussian summation formula.

Invariance Principles There are two possible setups for formulating invariance principles for annealed polymers. The first is when we consider $\mathbb{P}^h_n$, and accordingly polymers γ with fixed number n of steps. In this case one defines

$$x_n(t) = \frac{1}{\sqrt{n}}\left(\gamma_{\lfloor nt\rfloor} - nt\mathsf{v}\right),$$

and concludes from Theorem 3.22 that $x_n(\cdot)$ converges to a d-dimensional Brownian motion with covariance matrix $\Xi(h)$.

A somewhat different $(d-1)$-dimensional invariance principle holds in the conjugate ensemble of crossing polymers. To define the latter fix $\mathsf{x} \neq 0$, $\lambda > 0$ and consider the following probability distribution $\mathbb{P}^\mathsf{x}_\lambda$ on the family $\mathscr{P}_\mathsf{x}$ of all polymers γ which have displacement $\mathsf{X}(\gamma) = \mathsf{x}$:

$$\mathbb{P}^\mathsf{x}_\lambda(\gamma) = \frac{1}{G_\lambda(\mathsf{x})} \mathrm{e}^{-\lambda|\gamma|} \mathsf{W}_d(\gamma). \tag{133}$$

Consider again the irreducible decomposition (121) of paths $\gamma \in \mathscr{P}_\mathsf{x}$. Let vertices $0, \mathsf{u}_1, \ldots, \mathsf{u}_{N+1}, \mathsf{x}$ be the end-points of the corresponding irreducible paths. We can approximate γ by a linear interpolation through these vertices. We employ the language of section "Curves and Surfaces" of the Appendix. Let $\mathfrak{n}(h) = \frac{\mathsf{x}}{|\mathsf{x}|}$ and $\mathfrak{v}_1, \ldots, \mathfrak{v}_{d-1}$ are unit vectors in the direction of principal curvatures of $\partial\mathbf{K}_\lambda$ at h.

In the orthogonal frame $(\mathfrak{v}_1, \ldots, \mathfrak{v}_{d-1}, \mathfrak{n}(h))$, the linear interpolation through the vertices of the irreducible decomposition of γ can be represented as a function $Y : [0, |\mathsf{x}|] \mapsto T_h \partial \mathbf{K}_\lambda$. Consider the rescaling

$$y_{\mathsf{x}}(t) = \frac{1}{\sqrt{\mathsf{x}}} Y(t\,|\mathsf{x}|).$$

Then, (132) and the quadratic expansion formula (207) of the Appendix leads to the following conclusion: Let x_m be a sequence of points with $|\mathsf{x}_m| \to \infty$ and $\lim_{m\to\infty} \frac{\mathsf{x}_m}{|\mathsf{x}_m|} = \mathfrak{n}(h)$. Then the distribution of $y_{\mathsf{x}_m}(\cdot)$ under $\mathbb{P}_\lambda^{\mathsf{x}_m}$ converges to the distribution of the $(d-1)$-dimensional Brownian bridge with the diagonal covariance matrix diag $(\chi_1(h), \ldots, \chi_{d-1}(h))$, where $\chi_i(h)$ are the principal curvatures of $\partial \mathbf{K}_\lambda$ at h.

4 Very Weak Disorder in $d \geq 4$

The notion of very weak disorder depends on the dimension $d \geq 4$ and on the pulling force $h \neq 0$. It is quantified in terms of continuous positive non-decreasing functions ζ_d on $(0, \infty)$; $\lim_{h\downarrow 0} \zeta_d(h) = 0$. Function ζ_d does not have an independent physical meaning: It is needed to ensure a certain percolation property (143), and to ensure validity of a certain $\mathbb{L}_2$-type estimate formulated in Lemma 4.6 below.

Definition 4.1 Let us say that the polymer model (8) is in the regime of very weak disorder if $h \neq 0$ and

$$\phi_\beta(1) \leq \zeta_d(|h|)\,. \tag{134}$$

Remark 4.2 Equation (134) is a technical condition, and it has three main implications as it is explained below after formulation of Theorem 4.5.

Quenched Polymers at Very Weak Disorder In the regime of very weak disorder quenched polymers behave like their annealed counter-parts. Precisely: For $h \neq 0$, continue to use $\mathsf{v}_h = \nabla\lambda(h)$ and $\Sigma_h = \mathrm{Hess}(\lambda)(h)$ for the limiting spatial extension and the diffusivity of the annealed model. Let Cl_∞ be the unique infinite connected cluster of $\{\mathsf{x} : \mathsf{V}_{\mathsf{x}}^\omega < \infty\}$.

Theorem 4.3 *Fix $h \neq 0$. Then, in the regime of very weak disorder, the following holds $\mathbb{Q}$-a.s. on the event $\{0 \in \mathrm{Cl}_\infty\}$:*

- *The limit*

$$\lim_{n\to\infty} \frac{Z_n^\omega(h)}{Z_n(h)} \tag{135}$$

exists and is a strictly positive, square-integrable random variable.

- *For every* $\epsilon > 0$,

$$\sum_n \mathbb{P}_n^{h,\omega}\left(\left|\frac{\mathsf{X}(\gamma)}{n} - \mathsf{v}_h\right| > \epsilon\right) < \infty. \tag{136}$$

- *For every* $\alpha \in \mathbb{R}^d$,

$$\lim_{n\to\infty} \mathbb{P}_n^{h,\omega}\left(\exp\left\{\frac{i\alpha}{\sqrt{n}}\left(X(\gamma) - n\mathsf{v}_h\right)\right\}\right) = \exp\left\{-\frac{1}{2}\Xi_h\alpha\cdot\alpha\right\}. \tag{137}$$

Remark 4.3 Since in the regime of very weak disorder $h \notin \mathbf{K}_0$, the series in (50) converge exponentially fast. Using $Z_n^\omega(A|h)$ for the restriction of Z_n^ω to paths from A we conclude (from exponential Markov inequality) that there exists $c = c(\epsilon) > 0$ such that

$$\sum_n \mathscr{Q}\left(Z_n^\omega\left(\left|\frac{\mathsf{X}(\gamma)}{n} - \mathsf{v}_h\right| > \epsilon\middle|h\right) > \mathrm{e}^{-cn}Z_n(h)\right) < \infty.$$

In other words, (136) routinely follows from (135) and exponential bounds on annealed polymers.

Reformulation in Terms of Basic Partition Functions For each $h \neq 0$ and $\beta > 0$ basic annealed partition functions were defined in (128). Here is the corresponding definition of basic quenched partition functions:

$$\mathsf{t}^\omega(\mathsf{x}, n) = \sum_{\gamma\in\mathscr{T}_{\mathsf{x},n}} \mathrm{e}^{h\cdot\mathsf{X}(\gamma)-\lambda|\gamma|}\mathsf{W}_d^\omega(\gamma), \tag{138}$$

and, accordingly $\mathsf{t}^\omega(\mathsf{x}) = \sum_n \mathsf{t}^\omega(\mathsf{x}, n)$ and $\mathsf{t}^\omega(n) = \sum_{\mathsf{x}} \mathsf{t}^\omega(\mathsf{x}, n)$.

Let $\mathscr{Y} = \mathscr{Y}_h$ be the cone used to define irreducible paths. Then Cl_∞^h is the infinite connected component (unique if exists) of $\{\mathsf{x} \in \mathscr{Y}_h : \mathsf{V}_{\mathsf{x}}^\omega < \infty\}$.

Theorem 4.5 *Fix* $h \neq 0$. *Then, in the regime of very weak disorder, infinite connected cluster* Cl_∞^h *exists* $\mathscr{Q}$*-a.s.. Furthermore, the following holds* $\mathscr{Q}$*-a.s. on the event* $\{0 \in \mathrm{Cl}_\infty^h\}$*:*

- *The limit*

$$\mathsf{s}^\omega \stackrel{\Delta}{=} \lim_{n\to\infty} \frac{\mathsf{t}^\omega(n)}{\mathsf{t}(n)} \tag{139}$$

exists and is a strictly positive, square-integrable random variable.

- *For every* $\alpha \in \mathbb{R}^{d+1}$,

$$\lim_{n\to\infty} \frac{1}{\mathsf{t}^\omega(n)} \sum_x \exp\left\{\frac{i\alpha}{\sqrt{n}}\cdot(x - nv)\right\} \mathsf{t}^\omega(x, n) = \exp\left\{-\frac{1}{2}\Sigma\alpha\cdot\alpha\right\}. \tag{140}$$

Below we shall explain the proof of (139). The $\mathscr{Q}$-a.s. CLT follows in a rather similar fashion, albeit with some additional technicalities, and we refer to [18] for the complete proof.

Three Properties of Very Weak Disorder The role of (technical) condition $\phi_\beta(1) \le \zeta_d(|h|)$ is threefold:

First of all, setting $p_d = \mathscr{Q}(\mathsf{V}^\omega = \infty)$ and noting that

$$\zeta_d(|h|) \ge \phi_\beta(1) = -\log \mathscr{E}\left(\mathrm{e}^{-\beta \mathsf{V}^\omega}\right) \ge -\log(1-p_d) \ge p_d, \tag{141}$$

we conclude that (134) implies that:

$$\mathscr{Q}(\mathsf{V}^\omega = \infty) \le \zeta_d(|h|). \tag{142}$$

Thus, in view of (141), condition (134) implies that the infinite cluster Cl_∞^h exists: Namely if we choose ζ_d such that $\sup_h \zeta_d(|h|)$ is sufficiently small, then

$$\mathscr{Q}\left(\text{there is an infinite cluster } \mathrm{Cl}_\infty^h\right) = 1, \tag{143}$$

for all situations in question.

The second implication is that for any $h \neq 0$ fixed, $h \notin \mathbf{K}_0(\beta)$ for all β sufficiently small. In other words, in the regime of very weak disorder the annealed model is always in the ballistic phase. Indeed, since $\phi_\beta(\ell) \le \ell \phi_\beta(1)$,

$$Z_n(h) \ge \mathrm{e}^{-n\phi_\beta(1)} \left(\mathsf{E}_d \mathrm{e}^{h\cdot \mathsf{X}}\right)^n.$$

Consequently $\lambda(h) > 0$ whenever $\log\left(\mathsf{E}_d \mathrm{e}^{h\cdot \mathsf{X}}\right) > \phi_\beta(1)$.

The third implication is an $\mathbb{L}_2$-estimate (144) below. Recall that for h fixed, annealed measures $\mathbb{P}_n^h$ have limiting spatial extension $\mathsf{v}_h = \nabla\lambda(h)$, and they satisfy sharp classical local limit asymptotics around this value.

For a subset $A \subseteq \mathbb{Z}^d$, let $\mathscr{A}$ be the σ-algebra generated by $\{\mathsf{V}_\mathsf{x}^\omega\}_{\mathsf{x}\in A}$. We shall call such σ-algebras cylindrical.

Lemma 4.6 *For any dimension $d \ge 4$ there exists a positive non-decreasing function ζ_d on $(0,\infty)$ and a number $\rho < 1/12$ such that the following holds: If $\phi_\beta(1) < \zeta_d(|h|)$, then there exist constants $c_1, c_2 < \infty$ such that*

$$\begin{aligned}&\left|\mathscr{E}\left[\mathsf{t}^\omega(\mathsf{x},\ell)\mathsf{t}^\omega(\mathsf{y},\ell)\mathscr{E}\left(\mathsf{f}^{\theta_\mathsf{x}\omega}(\mathsf{z},m) - \mathsf{f}(\mathsf{z},m)\middle|\mathscr{A}\right)\mathscr{E}\left(\mathsf{f}^{\theta_\mathsf{y}\omega}(\mathsf{w},k) - \mathsf{f}(\mathsf{w},k)\middle|\mathscr{A}\right)\right]\right| \\ &\quad \le \frac{c_1 \mathrm{e}^{-c_2(m+k)}}{\ell^{d-\rho}} \exp\left\{-c_2\left(|\mathsf{x}-\mathsf{y}| + \frac{|\mathsf{x}-\ell\mathsf{v}_h|^2}{\ell} + \frac{|\mathsf{y}-\ell\mathsf{v}_h|^2}{\ell}\right)\right\},\end{aligned} \tag{144}$$

for all $\mathsf{x},\mathsf{y},\mathsf{z},\mathsf{w},m,k$ and all cylindrical σ-algebras $\mathscr{A}$ such that both $\mathsf{t}^\omega(\mathsf{x},\ell)$ and $\mathsf{t}^\omega(\mathsf{y},\ell)$ are $\mathscr{A}$-measurable.

Remark 4.7 Since the underlying random walk is of bounded range,

$$\#\left\{\mathsf{z} : \mathsf{f}^{\omega}(\mathsf{z}, m) \neq 0\right\} \lesssim m^{d}.$$

Hence, inequality (144) also holds with $\mathsf{f}^{\theta_{\mathsf{x}}\omega}(m)$ and $\mathsf{f}^{\theta_{\mathsf{y}}\omega}(k)$ instead of $\mathsf{f}^{\theta_{\mathsf{x}}\omega}(\mathsf{z}, m)$ and $\mathsf{f}^{\theta_{\mathsf{y}}\omega}(\mathsf{w}, k)$.

There is nothing sacred about the condition $\rho < 1/12$. We just need ρ to be sufficiently small. In fact, (144) holds with $\rho = 0$, although a proof of such statement would be a bit more involved.

We refer to [18] (Lemma 2.1 there) for a proof of Lemma 4.6. The claim (144) has a transparent meaning: For $\rho = 0$, the expression

$$\frac{c_1}{\ell^d}\exp\left\{-c_2\left(\frac{|\mathsf{x}-\ell\mathsf{v}_h|^2}{\ell}+\frac{|\mathsf{y}-\ell\mathsf{v}_h|^2}{\ell}\right)\right\}$$

is just the local limit bound on the annealed quantity $\mathsf{t}(\mathsf{x}, \ell)\mathsf{t}(\mathsf{y}, \ell)$. The term $\mathrm{e}^{-c_2(k+m)}$ reflects exponential decay of irreducible terms $\left(\mathsf{f}^{\theta_{\mathsf{x}}\omega}(\mathsf{z}, m) - \mathsf{f}(\mathsf{z}, m)\right)$ and, accordingly, $\left(\mathsf{f}^{\theta_{\mathsf{y}}\omega}(\mathsf{w}, k) - \mathsf{f}(\mathsf{w}, k)\right)$. The term $\mathrm{e}^{-c_2|\mathsf{x}-\mathsf{y}|}$ appears for the following reason (see Fig. 7): By irreducibly, $\mathsf{f}^{\theta_{\mathsf{x}}\omega}(\mathsf{z}, m) - \mathsf{f}(\mathsf{z}, m)$ depends only on $\mathsf{V}^{\omega}_{\mathsf{u}}$ with u belonging to the diamond shape $D(\mathsf{x}, \mathsf{x}+\mathsf{z})$. Similarly, $\mathsf{f}^{\theta_{\mathsf{y}}\omega}(\mathsf{w}, m)-\mathsf{f}(\mathsf{w}, m)$ depends only on variables inside $D(\mathsf{y}, \mathsf{y}+\mathsf{w})$. All these terms have zero mean. Consequently,

$$\mathscr{E}\left[\mathsf{t}^{\omega}(\mathsf{x}, \ell)\mathsf{t}^{\omega}(\mathsf{y}, \ell)\mathscr{E}\left(\mathsf{f}^{\theta_{\mathsf{x}}\omega}(\mathsf{z}, m) - \mathsf{f}(\mathsf{z}, m)\big|\mathscr{A}\right)\mathscr{E}\left(\mathsf{f}^{\theta_{\mathsf{y}}\omega}(\mathsf{w}, k) - \mathsf{f}(\mathsf{w}, k)\big|\mathscr{A}\right)\right] = 0,$$

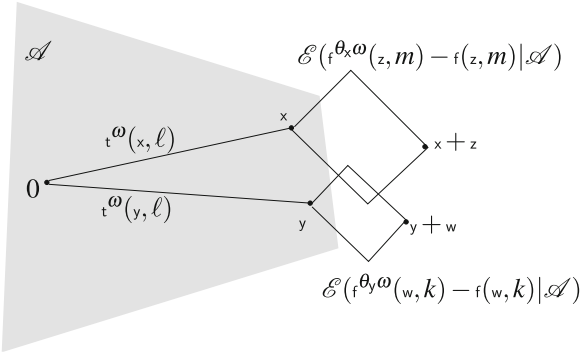

Fig. 7 Kites $\mathsf{t}^{\omega}(\mathsf{x}, \ell)\mathscr{E}\left(\mathsf{f}^{\theta_{\mathsf{x}}\omega}(\mathsf{z}, m) - \mathsf{f}(\mathsf{z}, m)\big|\mathscr{A}\right)$ and $\mathsf{t}^{\omega}(\mathsf{y}, \ell)\mathscr{E}\left(\mathsf{f}^{\theta_{\mathsf{y}}\omega}(\mathsf{w}, k) - \mathsf{f}(\mathsf{z}, k)\big|\mathscr{A}\right)$ may have a non-zero covariance only if their diamond shapes $D(\mathsf{x}, \mathsf{x}+\mathsf{z})$ and $D(\mathsf{y}, \mathsf{y}+\mathsf{w})$ intersect

whenever $D(\mathsf{x},\mathsf{x}+\mathsf{z})\cap D(\mathsf{y},\mathsf{y}+\mathsf{w})=\emptyset$. The remaining terms satisfy $\max\{\mathsf{z},\mathsf{w}\}\overset{>}{\sim}|\mathsf{x}-\mathsf{y}|$. In other words, the term $\mathrm{e}^{-c_2|\mathsf{x}-\mathsf{y}|}$ also reflects exponential decay of irreducible connections.

The disorder imposes an attractive interaction between two replicas. The impact of small $\xi_d(|h|)$ condition in the very weak disorder regime, as formulated in (144), is that this interaction is not strong enough to destroy individual annealed asymptotics.

Sinai's Decomposition of $\mathsf{t}^\omega(\mathsf{x},n)$ We rely on an expansion similar to the one employed by Sinai [27] in the context of directed polymers. By construction quantities $\mathsf{t}^\omega(\mathsf{x},n)$ satisfy the following (random) renewal relation:

$$\mathsf{t}^\omega(\mathsf{x},0)=\mathbb{1}_{\{\mathsf{x}=0\}}\quad\text{and}\quad\mathsf{t}^\omega(\mathsf{x},n)=\sum_{m=1}^{n}\sum_{\mathsf{y}}\mathsf{f}^\omega(\mathsf{y},m)\mathsf{t}^{\theta_{\mathsf{y}}\omega}(\mathsf{x}-\mathsf{y},n-m).\tag{145}$$

Iterating in (145) we obtain (for $n>0$):

$$\mathsf{t}^\omega(\mathsf{x},n)=\mathsf{f}^\omega(\mathsf{x},n)+\sum_{r=1}^{\infty}\sum_{\mathsf{x}_1,\dots\mathsf{x}_r}\sum_{n_1+\dots+n_r=n}\prod\mathsf{f}^{\theta_{\mathsf{x}_{i-1}}\omega}(\mathsf{x}_i-\mathsf{x}_{i-1},n_i),$$

where $\mathsf{x}_0\overset{\Delta}{=}0$. Writing,

$$\mathsf{f}^\omega(\mathsf{y},m)=\mathsf{f}(\mathsf{y},m)+\left(\mathsf{f}^\omega(\mathsf{y},m)-\mathsf{f}(\mathsf{y},m)\right),$$

we, after expansion and re-summation, arrive to the following decomposition:

$$\begin{aligned}\mathsf{t}^\omega(\mathsf{x},n)&=\mathsf{t}(\mathsf{x},n)\\&\quad+\sum_{\ell=0}^{n-1}\sum_{m=1}^{n-\ell}\sum_{r=0}^{n-\ell-m}\sum_{\mathsf{y},\mathsf{z}}\mathsf{t}^\omega(\mathsf{y},\ell)\left(\mathsf{f}^{\theta_{\mathsf{y}}\omega}(\mathsf{z}-\mathsf{y},m)-\mathsf{f}(\mathsf{z}-\mathsf{y},m)\right)\mathsf{t}(\mathsf{x}-\mathsf{z},r).\end{aligned}\tag{146}$$

In particular, the decomposition of $\mathsf{t}^\omega(n)$ is given by

$$\mathsf{t}^\omega(n)=\mathsf{t}(n)+\sum_{\ell=0}^{n-1}\sum_{m=1}^{n-\ell}\sum_{r=0}^{n-\ell-m}\sum_{\mathsf{y}}\mathsf{t}^\omega(\mathsf{y},\ell)\left(\mathsf{f}^{\theta_{\mathsf{y}}\omega}(m)-\mathsf{f}(m)\right)\mathsf{t}(r).\tag{147}$$

In order to prove (140) one needs to consider the full decomposition (146). As it was already mentioned, we shall not do it here and, instead, refer to [18]. From now on, we shall concentrate on proving (139) and, accordingly shall consider the reduced decomposition (147). Nevertheless, modulo additional technicalities, the proof of (139) captures all essential features of the argument.

Recall that for the annealed quantities, $\lim_{n\to\infty} \mathsf{t}(n) = \frac{1}{\mu(h)} = \frac{1}{\mu}$ exponentially fast. Writing $\mathsf{t}(m) = \frac{1}{\mu} + (\mathsf{t}(m) - \frac{1}{\mu})$ in all the corresponding terms in (147), we infer that $\mathsf{t}^\omega(n)$ can be represented as

$$\mathsf{t}^\omega(n) = \frac{1}{\mu}\mathsf{s}^\omega(n) + \epsilon_n^\omega + \left(\mathsf{t}(n) - \frac{1}{\mu}\right) \tag{148}$$

where

$$\mathsf{s}^\omega(n) = 1 + \sum_{\ell\le n}\sum_{\mathsf{x}} \mathsf{t}^\omega(\mathsf{x},\ell)\left(\mathsf{f}^{\theta_\mathsf{x}\omega} - 1\right), \tag{149}$$

and the correction term ϵ_n^ω is given by

$$\begin{aligned}\epsilon_n^\omega = \epsilon_{n,1}^\omega - \epsilon_{n,2}^\omega \stackrel{\Delta}{=} &\sum_{\ell+m+r=n}\sum_{\mathsf{x}} \mathsf{t}^\omega(\mathsf{x},\ell)\left(\mathsf{f}^{\theta_\mathsf{x}\omega}(m) - \mathsf{f}(m)\right)\left(\mathsf{t}(r) - \frac{1}{\mu}\right)\\ &+ -\frac{1}{\mu}\sum_{\substack{\ell\le n\\ m>n-\ell}}\sum_{\mathsf{x}} \mathsf{t}^\omega(x,\ell)\left(\mathsf{f}^{\theta_\mathsf{x}\omega}(m) - \mathsf{f}(m)\right).\end{aligned} \tag{150}$$

The term $\mathsf{t}(n) - \frac{1}{\mu}$ is negligible. Our target claim (139) is a direct consequence of the following proposition:

Proposition 4.8 *In the very weak disorder regime the following happens $\mathscr{P}$-a.s.:*

$$\lim_{n\to\infty} \mathsf{s}^\omega(n) = \mathsf{s}^\omega = 1 + \sum_{\mathsf{x}} \mathsf{t}^\omega(\mathsf{x})\left(\mathsf{f}^{\theta_\mathsf{x}\omega} - 1\right) \quad \text{and} \quad \sum_n \mathscr{E}(\epsilon_n^\omega)^2 < \infty. \tag{151}$$

Furthermore, $\mathsf{s}^\omega > 0$ on the set $\{0 \in \mathrm{Cl}_\infty^h\}$.

Remark 4.9 Note that the formula for s^ω is compatible with the common sense if the random walk is trapped (case $\mathscr{Q}(\mathsf{V}^\omega = \infty) > 0$.) Indeed, in such situation $\lim_{n\to\infty} \mathsf{t}^\omega(n)$ should be clearly zero. On the other hand, if the random walk is trapped, then the sum $1 + \sum_\mathsf{x} \mathsf{t}^\omega(\mathsf{x})\left(\mathsf{f}^{\theta_\mathsf{x}\omega} - 1\right)$ contains only finitely many non-zero terms. Using $\mathsf{t}^\omega(0) = 1$, let us rewrite it as

$$1 + \sum_{\mathsf{x}} \mathsf{t}^\omega(\mathsf{x})\mathsf{f}^{\theta_\mathsf{x}\omega} - \sum_{\mathsf{x}\neq 0} \mathsf{t}^\omega(\mathsf{x}) - 1 = \sum_{\mathsf{x}} \mathsf{t}^\omega(\mathsf{x})\mathsf{f}^{\theta_\mathsf{x}\omega} - \sum_{\mathsf{x}\neq 0} \mathsf{t}^\omega(\mathsf{x}).$$

However, for $\mathsf{x} \neq 0$,

$$\mathsf{t}^\omega(\mathsf{x}) = \sum_{\mathsf{y}} \mathsf{t}^\omega(\mathsf{y})\mathsf{f}^{\theta_\mathsf{y}\omega}(\mathsf{x}-\mathsf{y}) \;\Rightarrow\; \sum_{\mathsf{x}\neq 0} \mathsf{t}^\omega(\mathsf{x}) = \sum_{\mathsf{y}} \mathsf{t}^\omega(\mathsf{y})\mathsf{f}^{\theta_\mathsf{y}\omega}.$$

Mixingale Form of $\mathsf{s}^\omega(n)$ and ϵ_n^ω Let us rewrite $\mathsf{s}^\omega(n)$ as

$$\mathsf{s}^\omega(n) = 1 + \sum_{\ell \le n} \mathsf{Y}_\ell \quad \text{where} \quad \mathsf{Y}_\ell = \sum_{\mathsf{x}} \mathsf{t}^\omega(\mathsf{x}, \ell) \left(\mathsf{f}^{\theta_{\mathsf{x}}\omega} - 1 \right). \tag{152}$$

The variables Y_ℓ are mean zero, and it is easy to deduce from the basic $\mathbb{L}_2$-estimate (144) that in the regime of weak disorder, $\sum \mathscr{E}\left(\mathsf{Y}_\ell^2\right) < \infty$. Should $\{\mathsf{Y}_\ell\}$ be a martingale difference sequence (as in the case of directed polymers), we would be done. However, since, in principle, same vertices x may appear in different Y_ℓ-s, there seems to be no natural martingale structure at our disposal. Instead one should make a proper use of mixing properties of $\{\mathsf{Y}_\ell\}$. Hence the name mixingale, which was introduced in [22]. In order to prove convergence of $\mathsf{s}^\omega(n)$ we shall rely on the mixingale approach developed in [22].

Turning to the correction terms in (150), note that both $\epsilon_{n,1}^\omega$ and $\epsilon_{n,2}^\omega$ could be written in the form

$$\sum_{\ell \le n} \sum_{\mathsf{x}} \mathsf{t}^\omega(\mathsf{x}, \ell) \sum_m a^{(n)}(\ell, m) \left(\mathsf{f}^{\theta_{\mathsf{x}}\omega}(m) - \mathsf{f}(m) \right) \stackrel{\Delta}{=} \sum_{\ell \le n} \mathsf{Z}_\ell^{(n)}, \tag{153}$$

where

$$a^{(n)}(\ell, m) = \left(\mathsf{t}(n - \ell - m) - \frac{1}{\mu} \right) \mathbb{1}_{\ell + m \le n} \text{ and } a^{(n)}(\ell, m) = -\mathbb{1}_{\ell + m > n} \tag{154}$$

respectively in the cases of $\epsilon_{n,1}^\omega$ and $\epsilon_{n,2}^\omega$. Again, $\left\{\mathsf{Z}_\ell^{(n)}\right\}$ is not a martingale difference sequence, and we shall rely on the mixingale approach of [22] for deducing their second convergence statement in (151).

Below we shall formulate a particular case of the maximal inequality for mixingales [22]. To keep relation with quenched polymers, and specifically with (152) and (153), in mind, let us introduce the following filtration $\{\mathscr{A}_m\}$. Recall that the end-point of the m-step annealed polymer stays close to $m\mathsf{v}_h = m\nabla\lambda(h)$. Define half-spaces $\mathscr{H}_m^-$ and the corresponding σ-algebras $\mathscr{A}_m$ as

$$\mathscr{H}_m^- = \left\{ \mathsf{x} \in \mathbb{Z}^d \,:\, \mathsf{x} \cdot \mathsf{v}_h \le m|\mathsf{v}_h|^2 \right\} \quad \text{and} \quad \mathscr{A}_m = \sigma\left\{ \mathsf{V}_{\mathsf{x}}^\omega \,:\, x \in \mathscr{H}_m^- \right\}. \tag{155}$$

Mixingale Maximal Inequality and Convergence Theorem of McLeish Let $\mathsf{Y}_1, \mathsf{Y}_2, \ldots$ be a sequence of zero-mean, square-integrable random variables. Let also $\{\mathscr{A}_k\}_{-\infty}^{\infty}$ be a filtration of σ-algebras. Suppose that there exist $\epsilon > 0$ and numbers $d_1, d_2, \ldots$ in such a way that

$$\mathscr{E}\left(\mathbb{E}\left(\mathsf{Y}_\ell \,\middle|\, \mathscr{A}_{\ell-k} \right)^2 \right) \le \frac{d_\ell^2}{(1+k)^{1+\epsilon}} \quad \text{and} \quad \mathscr{E}\left(\mathsf{Y}_\ell - \mathbb{E}\left(\mathsf{Y}_\ell \,\middle|\, \mathscr{A}_{\ell+k} \right) \right)^2 \le \frac{d_\ell^2}{(1+k)^{1+\epsilon}} \tag{156}$$

for all $\ell = 1, 2, \ldots$ and $k \geq 0$. Then [22] there exists $K = K(\epsilon) < \infty$ such that, for all $n_1 \leq n_2$,

$$\mathscr{E}\left\{\max_{n_1 \leq r \leq n_2}\left(\sum_{n_1}^{r} \mathsf{Y}_\ell\right)^2\right\} \leq K \sum_{n_1}^{n_2} d_\ell^2. \tag{157}$$

In particular, if $\sum_\ell d_\ell^2 < \infty$, then $\sum_\ell \mathsf{Y}_\ell$ converges $\mathscr{Q}$-a.s. and in $\mathbb{L}_2$.

Convergence of $\mathsf{s}^\omega(\boldsymbol{n})$ Consider decomposition (152). Clearly,

$$\mathscr{E}\left(\mathsf{Y}_\ell \,\middle|\, \mathscr{A}_{\ell-k}\right) = \sum_{\mathsf{x} \in \mathscr{H}_{\ell-k}^-} \mathsf{t}^\omega(\mathsf{x}, \ell)\left(\mathsf{f}^{\theta_\mathsf{x}\omega} - 1\right) \tag{158}$$

Applying (144) we conclude that for any $\mathsf{x}, \mathsf{y} \in \mathscr{H}_{\ell-k}^-$,

$$\begin{aligned}&\left|\mathscr{E}\left[\mathsf{t}^\omega(\mathsf{x}, \ell)\mathsf{t}^\omega(\mathsf{y}, \ell)\mathscr{E}\left(\mathsf{f}^{\theta_\mathsf{x}\omega} - 1\middle|\mathscr{A}_{\ell-k}\right)\mathscr{E}\left(\mathsf{f}^{\theta_\mathsf{y}\omega} - 1\middle|\mathscr{A}_{\ell-k}\right)\right]\right| \\ &\quad \leq \frac{c_3}{\ell^{d-\rho}} \exp\left\{-c_2\left(|\mathsf{x} - \mathsf{y}| + \frac{|\mathsf{x} - \ell\mathsf{v}_h|^2}{\ell} + \frac{|\mathsf{y} - \ell\mathsf{v}_h|^2}{\ell}\right)\right\},\end{aligned} \tag{159}$$

Consequently, summing up with respect to $\mathsf{x}, \mathsf{y} \in \mathscr{H}_{\ell-k}^-$ we infer that for any $\epsilon \geq 2\rho$:

$$\mathscr{E}\left(\mathscr{E}\left(\mathsf{Y}_\ell\middle|\mathscr{A}_{\ell-k}\right)^2\right) \leq \frac{c_5 \mathrm{e}^{-c_4\frac{k^2}{\ell}}}{\ell^{d/2-\rho}} \leq \frac{c_6}{\ell^{(d-1)/2-\epsilon}(1+k)^{1+\epsilon}} \triangleq \frac{d_{\ell,-}^2}{(1+k)^{1+\epsilon}}. \tag{160}$$

On the last step we have relied on a trivial asymptotic inequality

$$\frac{\mathrm{e}^{-c_4\frac{k^2}{\ell}}}{\ell^{(1+\epsilon)/2}} \lesssim \frac{1}{(1+k)^{1+\epsilon}}.$$

Note that if $(d-1)/2 - \epsilon > 1$, which is compatible with $d \geq 4$ and $\rho < 1/12$, then $\sum d_{\ell,-}^2 < \infty$.

Turning to the second condition in (156) note first of all that

$$\mathscr{E}\left(\mathsf{t}^\omega(\mathsf{x}, \ell)\left(\mathsf{f}^{\theta_\mathsf{x}\omega} - 1\right)\middle|\mathscr{A}_{\ell+k}\right) = 0$$

whenever $\mathsf{x} \in \mathscr{H}_{\ell+k}^+$, and

$$\mathscr{E}\left(\mathsf{t}^\omega(\mathsf{x}, \ell)\left(\mathsf{f}^{\theta_\mathsf{x}\omega}(\mathsf{y}) - \mathsf{f}(\mathsf{y})\right)\middle|\mathscr{A}_{\ell+k}\right) = \mathsf{t}^\omega(\mathsf{x}, \ell)\left(\mathsf{f}^{\theta_\mathsf{x}\omega}(\mathsf{y}) - \mathsf{f}(\mathsf{y})\right)$$

whenever $\mathsf{x}+\mathsf{y}\in\mathscr{H}^-_{\ell+k}$. Therefore,

$$\mathsf{Y}_\ell-\mathscr{E}\left(\mathsf{Y}_\ell\middle|\mathscr{A}_{\ell+k}\right)=\sum_{\mathsf{x}\in\mathscr{H}^+_{\ell+k}}\mathsf{t}^\omega(\mathsf{x},\ell)\left(\mathsf{f}^{\theta_\mathsf{x}\omega}-1\right)$$
$$+\sum_{\substack{\mathsf{x}\in\mathscr{H}^-_{\ell+k}\\ \mathsf{y}\in\mathscr{H}^+_{\ell+k}}}\mathsf{t}^\omega(\mathsf{x},\ell)\left(\mathsf{f}^{\theta_\mathsf{x}\omega}(\mathsf{y}-\mathsf{x})-\mathscr{E}\left(\mathsf{f}^{\theta_\mathsf{x}\omega}(\mathsf{y}-\mathsf{x})\middle|\mathscr{A}_{\ell+k}\right)\right). \tag{161}$$

The first term in (161) has exactly the same structure as (158). The second term in (161) happens to be even more localized (see discussion of (2.14) in [18]). The conclusion is:

$$\mathscr{E}\left(\mathsf{Y}_\ell-\mathscr{E}\left(\mathsf{Y}_\ell\middle|\mathscr{A}_{\ell+k}\right)^2\right)\le\frac{d^2_{\ell,+}}{(1+k)^{1+\epsilon}}, \tag{162}$$

where $d^2_{\ell,+}\lesssim\ell^{-(d-1)/2+\epsilon}$.

Set $d^2_\ell=\max\left\{d^2_{\ell,-},d^2_{\ell,+}\right\}\lesssim\ell^{-(d-1)/2+\epsilon}$, we, in view of the feasible choice $(d-1)/2-\epsilon>1$, conclude from (157) that $\mathsf{s}^\omega(n)$ is indeed a $\mathscr{Q}$-a.s. converging sequence.

Correction Terms Treatment of correction terms in their mixingale representation (153) follows a similar pattern. We refer to Section 2.2 in [18] for the proof of the second claim in (151).

Positivity of s^ω As we have already checked the sum $\mathsf{s}^\omega=1+\sum_\mathsf{x}\mathsf{t}^\omega(\mathsf{x})\left(\mathsf{f}^{\theta_\mathsf{x}\omega}-1\right)$ converges $\mathscr{Q}$-a.s. and in $\mathbb{L}_2$. In particular, $\mathscr{E}(\mathsf{s}^\omega)=1$. We claim that $\mathsf{s}^\omega>0$, $\mathscr{Q}$-a.s. on the event $\left\{0\in\mathrm{Cl}^h_\infty\right\}$. In order to prove this it would be enough to check that

$$\mathscr{Q}\left(\exists\mathsf{x}\in\mathscr{Y}:\mathsf{s}^{\theta_\mathsf{x}\omega}>0\right)=1. \tag{163}$$

Let us sketch the argument: If $\mathsf{s}^{\theta_\mathsf{x}\omega}>0$, then $\mathsf{x}\in\mathrm{Cl}^h_\infty$. But there is exactly one infinite cluster in $\mathscr{Y}$. Hence 0 is connected to x by a finite path $\gamma\subset\mathrm{Cl}^h_\infty\subset\mathscr{Y}$. Now, by assumption on x, $\lim_{n\to\infty}\mathsf{t}^{\theta_\mathsf{x}\omega}(n)=\lim_{n\to\infty}\sum_\mathsf{z}\mathsf{t}^{\theta_\mathsf{x}\omega}(\mathsf{z},n)>0$. By comparison with annealed quantities [large deviations, for instance (97)] we, at least for large n, may ignore terms $\mathsf{t}^{\theta_\mathsf{x}\omega}(\mathsf{z},n)$ with $\gamma\not\subset(\mathsf{x}+\mathsf{z})-\mathscr{Y}$. Which means that $\liminf_{n\to\infty}\mathsf{t}^\omega(n)\ge\mathsf{W}^{h,\lambda,\omega}_d(\gamma)\frac{\mathsf{s}^{\theta_\mathsf{x}\omega}}{\mu}$, where $\mathsf{W}^{h,\lambda,\omega}_d(\gamma)=\mathsf{e}^{h\cdot\mathsf{x}(\gamma)-\lambda|\gamma|}\mathsf{W}^\omega_d(\gamma)>0$.

It remains to check (163). Consider sets

$$B_n=\partial\mathscr{H}^+_n\cap\mathscr{Y},\quad|B_n|\cong n^{d-1}.$$

We refer to the last subsection of [15] for the proof of the following statement:

$$\left|\mathbb{C}\mathrm{ov}(\mathsf{s}^{\theta_{\mathsf{x}}\omega}, \mathsf{s}^{\theta_{\mathsf{y}}\omega})\right| \overset{<}{\sim} \frac{1}{|\mathsf{y}-\mathsf{x}|^{d/2-1}}, \tag{164}$$

uniformly in n and in $\mathsf{x}, \mathsf{y} \in B_n$. The quantity $\frac{1}{|\mathsf{y}-\mathsf{x}|^{d/2-1}}$ in (164) represents an intersection probability for trajectories of two ballistic d-dimensional random walks (such as the effective random walks with step distribution $\mathsf{f}(\mathsf{w}, m)$) which start at x and y. The statement (164) is very similar in spirit to that of Lemma 4.6: it says that possible weak attraction due to disorder does not destroy such asymptotics.

With (164) at our disposal it is very easy to finish the proof of (163). Indeed, it implies that

$$\mathbb{V}\mathrm{ar}\left(\frac{1}{n^{d-1}}\sum_{\mathsf{x}\in B_n}\mathsf{s}^{\theta_{\mathsf{x}}\omega}\right) \overset{<}{\sim} \frac{1}{n^{d/2-1}},$$

and since $\mathscr{E}\left(\sum_{\mathsf{x}\in B_n}\mathsf{s}^{\theta_{\mathsf{x}}\omega}\right) = |B_n| \overset{\sim}{=} n^{d-1}$, the conclusion follows by Chebychev inequality and Borel-Cantelli argument.

5 Strong Disorder

In this section, we work only under Assumption **A1** of the Introduction, and we do not impose any further assumptions on the environment $\{\mathsf{V}_{\mathsf{x}}^{\omega}\}$. The case of traps; $\mathscr{Q}(\mathsf{V}^{\omega} = \infty) \in (0, 1)$, is not excluded and we even do not need **A2** or any other restriction on the size of the latter probability.

The environment is always strong in two dimensions in the following sense (level **L1** in the language of the Introduction):

Theorem 5.1 *Let $d = 2$ and $\beta, \lambda > 0$. There exists $c = c(\beta, \lambda) > 0$ such that the following holds: Let $\lambda(h) = \lambda$ (in particular $h \notin \mathbf{K}_0$). Then, $\mathscr{Q}$-a.s.*

$$\limsup_{n\to\infty}\frac{1}{n}\log\frac{Z_n^{\omega}(h)}{Z_n(h)} < -c. \tag{165}$$

In particular, $\lambda^{\omega}(h) < \lambda(h) = \lambda$ whenever λ^{ω} is well defined.

Remark 5.2 As in [21] and, subsequently, [33] proving strong disorder in dimension $d = 3$ is a substantially more delicate task.

Let us explain Theorem 5.1: By the exponential Markov inequality (and Borel-Cantelli) it is sufficient to prove that there exist $c' > 0$ and $\alpha > 0$ such that

$$\mathscr{E}\left\{\left(\frac{Z_n^{\omega}(h)}{Z_n(h)}\right)^{\alpha}\right\} \le e^{-c'n}. \tag{166}$$

We shall try to establish (166) with $\alpha \in (0,1)$. This is the fractional moment method of [21]. It has a transparent logic: Since $\mathscr{E}\left(Z_n^\omega(h)\right) = Z_n(h)$, expecting (165) means that $Z_n^\omega(h)$ takes excessive exponentially high values with exponentially small probabilities. Taking fractional moments in (166) amounts to truncating these high values.

Reduction to Basic Partition Functions Recall the definition of diamond-confined (basic) partition functions $\mathsf{f}(\mathsf{x},n) = \mathscr{E}\left(\mathsf{f}^\omega(\mathsf{x},n)\right)$, $\mathsf{t}(\mathsf{x},n) = \mathscr{E}\left(\mathsf{t}^\omega(\mathsf{x},n)\right)$ and, accordingly, $\mathsf{f}(\mathsf{x})$, $\mathsf{t}(\mathsf{x}),\dots$ in (128). Since $\lim_{n\to\infty}\mathsf{t}(n) = \mu(h)^{-1}$, theorem target statement (165) would follow from

$$\limsup_{n\to\infty}\frac{1}{n}\log\frac{\mathsf{t}^\omega(n)}{\mathsf{t}(n)} = \limsup_{n\to\infty}\frac{1}{n}\log\mathsf{t}^\omega(n) < 0. \tag{167}$$

In its turn, in view of Theorem 3.19, (167) is routinely implied by the following statement ((169) below): Let r_N^ω be the partition function of N irreducible steps:

$$\mathsf{r}_N^\omega \stackrel{\Delta}{=} \sum_{\mathsf{u}_1,\cdots,\mathsf{u}_N}\mathsf{f}^\omega(\mathsf{u}_1)\mathsf{f}^{\theta_{\mathsf{u}_1}\omega}(\mathsf{u}_2-\mathsf{u}_1)\cdots\mathsf{f}^{\theta_{\mathsf{u}_{N-1}}\omega}(\mathsf{u}_N-\mathsf{u}_{N-1}) = \sum_{\mathsf{x}}\mathsf{r}_{\mathsf{x},N}^\omega. \tag{168}$$

Then, $\mathscr{Q}$-a.s.

$$\limsup_{N\to\infty}\frac{1}{N}\log\mathsf{r}_N^\omega < 0. \tag{169}$$

Again by Borel-Cantelli and the exponential Markov inequality, (169) would follow as soon as we check that for some $\alpha > 0$,

$$\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left(\mathsf{r}_N^\omega\right)^\alpha < 0. \tag{170}$$

Fractional Moments The proof of the fractional moment bound (170) comprises several steps.

STEP 1 Following [21]: (170) is verified once we show that there exist $N \in \mathbb{N}$ and $\alpha \in (0,1)$ such that

$$\mathscr{E}\left\{\sum_x\left(\mathsf{r}_{x,N}^\omega\right)^\alpha\right\} < 1. \tag{171}$$

Indeed, first of all if $a_i \geq 0$ and $\alpha \in (0,1)$, then

$$\left(\sum a_i\right)^\alpha \leq \sum a_i^\alpha. \tag{172}$$

Equivalently (setting $p = 1/\alpha > 1$ and $b_i = a_i^\alpha$), $\sum b_i^p \le (\sum b_i)^p$. Since

$$\frac{\mathrm{d}}{\mathrm{d}b_n}\Big(\sum b_i\Big)^p \ge p b_n^{p-1} = \frac{\mathrm{d}}{\mathrm{d}b_n}\sum b_i^p,$$

the latter form of (172) follows by induction.

We proceed with proving that (171) implies (170). Evidently, by (172),

$$\mathsf{r}^\omega_{N+M} = \sum_{\mathsf{x}} \mathsf{r}^\omega_{N,\mathsf{x}} \mathsf{r}^{\theta_{\mathsf{x}}\omega}_M \;\Rightarrow\; \big(\mathsf{r}^\omega_{N+M}\big)^\alpha \le \sum_{\mathsf{x}} \big(\mathsf{r}^\omega_{N,\mathsf{x}}\big)^\alpha \Big(\mathsf{r}^{\theta_{\mathsf{x}}\omega}_M\Big)^\alpha,$$

for any $\alpha \in (0,1)$. Since $\mathsf{r}^\omega_{N,\mathsf{x}}$ and $\mathsf{r}^{\theta_{\mathsf{x}}\omega}_M$ are independent, and r^ω_M is translation invariant, it follows that

$$\mathscr{E}\left\{(\mathsf{r}^\omega_{N+M})^\alpha\right\} \le \mathscr{E}\left\{\sum_{\mathsf{x}} \big(\mathsf{r}^\omega_{N,\mathsf{x}}\big)^\alpha\right\} \mathscr{E}\left\{(\mathsf{r}^\omega_M)^\alpha\right\}.$$

Hence (171), implies exponential decay of $M \mapsto \mathscr{E}\left\{(\mathsf{r}^\omega_M)^\alpha\right\}$.

STEP 2 Let $\mathsf{v}_h = \sum_{\mathsf{x}} \mathsf{x}\mathsf{f}(\mathsf{x})$; mean displacement under probability measure $\{\mathsf{f}(\mathsf{x})\}$. By Theorem 3.19, the latter distribution has exponential tails, and classical moderate deviation results apply. For $\mathsf{y} \in \mathbb{Z}^d$ define the distance from y to the line in the direction of v_h; $\mathrm{d}_h(\mathsf{y}) = \min_a |\mathsf{y} - a\mathsf{v}_h|$. Pick K sufficiently large and ϵ small, and consider

$$A_N = \left\{\mathsf{y} \in \mathbb{Z}^d \;:\; 0 \le \mathsf{y}\cdot\mathsf{v}_h \le KN \text{ and } \mathrm{d}_h(\mathsf{y}) \le N^{\frac{1}{2}+\epsilon}\right\}.$$

Recall that $\mathsf{r}_{\mathsf{x},N}$ is the distribution of the end point of the N-step random walk with $\{\mathsf{f}(\mathsf{x})\}$ being the one step distribution. With a slight abuse of notation, we can consider r_N as a distribution on the set of all N-step trajectories of this random walk:

$$\mathsf{r}_N(\mathsf{x}_1,\dots,\mathsf{x}_N) = \mathsf{f}(\mathsf{x}_1)\mathsf{f}(\mathsf{x}_2-\mathsf{x}_1)\dots\mathsf{f}(\mathsf{x}_N-\mathsf{x}_{N-1}). \tag{173}$$

By classical (Gaussian) moderate deviation estimates, there exists $c > 0$ such that

$$\sum_{\mathsf{x}\notin A_N} \mathsf{r}^\alpha_{\mathsf{x},N} \le \mathrm{e}^{-c\alpha N^{2\epsilon}} \text{ and } \mathsf{r}_N\left(\{\mathsf{x}_1,\dots,\mathsf{x}_N\} \not\subset A_N\right) \le \mathrm{e}^{-cN^{2\epsilon}}. \tag{174}$$

Furthermore, with another slight abuse of notation we can consider $\mathsf{r}_N(\cdot)$ as the distribution on the family of all N-concatenations $\gamma = \gamma_1\circ\gamma_2\circ\dots\circ\gamma_N$ of irreducible paths $\gamma_i \in \mathscr{F}$. In this way,

$$\mathsf{r}_N(\gamma) = \prod_1^N \mathsf{W}^{h,\lambda}_d(\gamma_i).$$

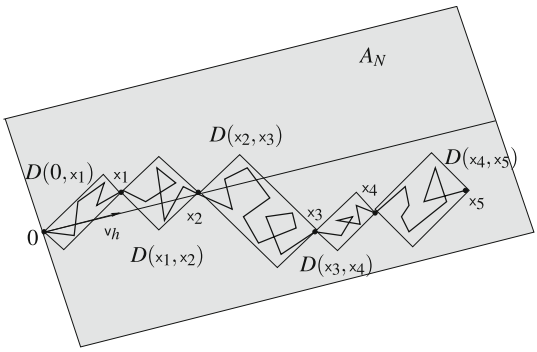

Fig. 8 Example: $N = 5$. The path $(0, \mathsf{x}_1, \ldots, \mathsf{x}_5)$ of the effective random walk, and the union of diamond shapes $\cup_i D(\mathsf{x}_{i-1}, \mathsf{x}_i) \subset A_N$

Recall from (120) that irreducible paths γ_i satisfy the following diamond confinement condition: If $\mathsf{x}_{i-1}, \mathsf{x}_i$ are the end points of γ_i, then $\gamma_i \subset D(\mathsf{x}_{i-1}, \mathsf{x}_i)$.

Exercise 5.3 Prove the following generalization of the second of (174) (see Fig. 8): There exists $c > 0$ such that

$$\mathsf{r}_N \left(\cup_i D(\mathsf{x}_{i-1}, \mathsf{x}_i) \not\subset A_N \right) \leq \mathrm{e}^{-cN^{2\epsilon}}. \tag{175}$$

Since any concatenation $\gamma = \gamma_1 \circ \gamma_2 \circ \cdots \circ \gamma_N$ of irreducible paths $\gamma_i \in \mathscr{F}$ satisfies $\gamma \subseteq \cup_i D(\mathsf{x}_{i-1}, \mathsf{x}_i)$, we readily infer that under r_N typical annealed paths stay inside A_N,

$$\mathsf{r}_N \left(\gamma \not\subset A_N \right) \leq \mathrm{e}^{-cN^{2\epsilon}} \;\Rightarrow\; \mathscr{E} \left\{ \left(\mathsf{r}_N^{\omega} \left(\gamma \not\subset A_N \right) \right)^{\alpha} \right\} \leq \mathrm{e}^{-\alpha c N^{2\epsilon}}, \tag{176}$$

for any $\alpha \in (0,1)$ (by Jensen's inequality).

Since $\mathscr{E} \left\{ \left(\mathsf{r}_{x,N}^{\omega} \right)^{\alpha} \right\} \leq r_{N,x}^{\alpha}$, we, in view of the first of (174), may restrict summation in (171) to $\mathsf{x} \in A_N$. In view of (176), it would be enough to check that

$$\sum_{x \in A_N} \mathscr{E} \left\{ \left(\mathsf{r}_{x,N}^{\omega} (\gamma \subset A_N) \right)^{\alpha} \right\} \leq |A_N| \, \mathscr{E} \left\{ \left(\mathsf{r}_N^{\omega} \left(\gamma \subset A_N \right) \right)^{\alpha} \right\} < 1. \tag{177}$$

The first inequality above is a crude over-counting, but for $d = 2$ it will do.

STEP 3 We, therefore, concentrate on proving the second inequality in (177). At this stage, we shall modify the distribution of the environment inside A_N in the following way: The modified law of the environment, which we shall denote $\mathscr{Q}_\delta$ is still product and, for every $\mathsf{x} \in A_N$,

$$\frac{\mathrm{d}\mathscr{Q}_\delta}{\mathrm{d}\mathscr{Q}} \left(\mathsf{V}_{\mathsf{x}}^{\omega} \right) \stackrel{\Delta}{=} e^{\delta \psi(\mathsf{V}_{\mathsf{x}}^{\omega}) - g(\delta)}, \quad \text{where} \quad e^{g(\delta)} = \log \mathscr{E} \left(e^{\delta \psi(\mathsf{V}^{\omega})} \right),$$

and ψ is a bounded non-decreasing function on $\mathbb{R}_+$, for instance $\psi(v) = v \wedge 1$.

The annealed potential in the modified environment is

$$\phi_\beta(\ell,\delta) = -\log \mathscr{E}_\delta\left(\mathrm{e}^{-\beta\ell \mathsf{V}^\omega}\right) = -\log\left\{\frac{\mathscr{E}\left(\mathrm{e}^{-\beta\ell\mathsf{V}^\omega+\delta\psi(\mathsf{V}^\omega)}\right)}{\mathscr{E}\left(\mathrm{e}^{\delta\psi(\mathsf{V}^\omega)}\right)}\right\}.$$

Note that for any $\ell \geq 1$,

$$\frac{\mathrm{d}\phi_\beta(\ell,\delta)}{\mathrm{d}\delta}\Big|_{\delta=0} = \mathscr{E}\left(\psi(\mathsf{V}^\omega)\right)\mathscr{E}(\mathrm{e}^{-\ell\beta\mathsf{V}^\omega}) - \mathscr{E}\left(\psi(\mathsf{V}^\omega)\mathrm{e}^{-\ell\beta\mathsf{V}^\omega}\right) > 0. \tag{178}$$

Indeed, since ψ is non-decreasing and $\mathrm{e}^{-\ell v}$ is decreasing, the last inequality follows from positive association of one-dimensional probability measures, as described in the beginning of Sect. 2.1.

By (176) we can ignore paths which do not stay inside A_N. Thus, (178) implies: There exists $c_\psi > 0$ such that for all δ sufficiently small,

$$\mathscr{E}_\delta\left(\mathsf{r}^\omega\right) = \sum_{\mathsf{x},n} \mathscr{E}_\delta\left(\mathsf{f}^\omega(\mathsf{x},n)\right) \leq 1 - c_\psi\delta \;\Rightarrow\; \mathscr{E}_\delta(\mathsf{r}_N^\omega) \leq \mathrm{e}^{-Nc_\psi\delta}. \tag{179}$$

From Hölder's inequality,

$$\begin{aligned}\mathscr{E}\left\{\left(\mathsf{r}_N^\omega(\gamma\subset A_N)\right)^\alpha\right\} &\leq \left(\mathscr{E}_\delta\left\{\left(\frac{\mathrm{d}\mathscr{Q}}{\mathrm{d}\mathscr{Q}_\delta}\right)^{1/(1-\alpha)}\right\}\right)^{1-\alpha}\left(\mathscr{E}_\delta\left\{\mathsf{r}_N^\omega(\gamma\subset A_N)\right\}\right)^\alpha \\ &\overset{(179)}{\leq} \left(\mathscr{E}_\delta\left\{\left(\frac{\mathrm{d}\mathscr{Q}}{\mathrm{d}\mathscr{Q}_\delta}\right)^{1/(1-\alpha)}\right\}\right)^{1-\alpha}\mathrm{e}^{-Nc_\psi\alpha\delta}.\end{aligned} \tag{180}$$

Now, the first term on the right hand side of (180) is

$$\begin{aligned}\mathscr{E}_\delta\left\{\left(\frac{\mathrm{d}\mathscr{Q}}{\mathrm{d}\mathscr{Q}_\delta}\right)^{1/(1-\alpha)}\right\} &= \mathscr{E}\left\{\frac{\mathrm{d}\mathscr{Q}_\delta}{\mathrm{d}\mathscr{Q}}\left(\frac{\mathrm{d}\mathscr{Q}}{\mathrm{d}\mathscr{Q}_\delta}\right)^{1/(1-\alpha)}\right\} \\ &= \left(\mathscr{E}\left\{\mathrm{e}^{\frac{\alpha}{1-\alpha}(g(\delta)-\delta\psi(\mathsf{V}^\omega))}\right\}\right)^{|A_N|}.\end{aligned} \tag{181}$$

However, the first order terms in δ cancel:

$$\frac{\mathrm{d}}{\mathrm{d}\delta}\Big|_{\delta=0}\log\mathscr{E}\left\{\mathrm{e}^{\frac{\alpha}{1-\alpha}(g(\delta)-\delta\psi(\mathsf{V}^\omega))}\right\} = \frac{\alpha}{1-\alpha}\left(g'(0) - \mathscr{E}\left\{\psi(\mathsf{V}^\omega)\right\}\right) = 0. \tag{182}$$

Consequently, by the second order expansion, there exists $v_\psi < \infty$, such that

$$\left(\mathscr{E}_\delta\left\{\left(\frac{\mathrm{d}\mathscr{Q}}{\mathrm{d}\mathscr{Q}_\delta}\right)^{1/(1-\alpha)}\right\}\right)^{1-\alpha} \leq \mathrm{e}^{\frac{v_\psi}{1-\alpha}\delta^2|A_N|}. \tag{183}$$

A substitution to (180) yields:

$$\mathscr{E}\left\{\left(\mathsf{r}_N^{\omega}\left(\gamma \subset A_N\right)\right)^{\alpha}\right\} \leq \mathrm{e}^{-Nc_{\psi}\alpha\delta+\frac{\nu_{\psi}}{1-\alpha}\delta^2|A_N|}. \tag{184}$$

We are now ready to specify the choice of $\delta = \delta_N$: In two dimensions; $d = 2$, the cardinality $|A_N| \overset{<}{\sim} N^{\frac{3}{2}+\epsilon}$. Hence, (177) follows whenever we choose

$$\frac{C\log N}{N} \ll \delta_N \ll N^{-\frac{1}{2}-\epsilon}$$

with $C = C(\alpha)$ being sufficiently large.

Appendix: Geometry of Convex Bodies and Large Deviations

In these notes we shall restrict attention to finite dimensional spaces $\mathbb{R}^d$. The principal references are [2, 24, 26] for convex geometry and [5–7, 30] for large deviations.

Convexity and Duality

Convex Functions A function $\phi : \mathbb{R}^d \to \mathbb{R} \cup \infty$ is said to be convex if

$$\phi\left(t\mathsf{x} + (1-t)\mathsf{y}\right) \leq t\phi(\mathsf{x}) + (1-t)\phi(\mathsf{y}),$$

for all $\mathsf{x}, \mathsf{y} \in \mathbb{R}^d$ and all $t \in [0,1]$.

Remark 6.1 Note that by definition we permit ∞ values, but not $-\infty$ values.

Alternatively, $\phi : \mathbb{R}^d \to \mathbb{R} \cup \infty$ is convex if the set

$$\mathrm{epi}(\phi) = \{(\mathsf{x}, \alpha) \ : \ \phi(\mathsf{x}) \leq \alpha\} \subset \mathbb{R}^{d+1} \tag{185}$$

is convex. We shall work with convex lower-semicontinuous functions: ϕ is lower-semicontinuous if for any x and any sequence x_n converging to x,

$$\phi(\mathsf{x}) \leq \lim_{n\to\infty} \phi(\mathsf{x}_n).$$

Alternatively, ϕ is lower-semicontinuous if the set $\mathrm{epi}(\phi)$ is closed.

A basic example of a convex and lower-semicontinuous (actually continuous) function is an affine function

$$\mathfrak{l}_{a,h}(\mathsf{x}) = a + h \cdot \mathsf{x}.$$

Theorem 6.2 *The following are equivalent:*

(a) $\phi : \mathbb{R}^d \to \mathbb{R} \cup \infty$ *is convex and lower-semicontinuous.*
(b) $\mathrm{epi}(\phi)$ *is convex and closed.*
(c) ϕ *can be recovered from its affine minorants:*

$$\phi(x) = \sup_{a,h} \{\mathfrak{l}_{a,h}(\mathsf{x}) \ : \ \mathfrak{l}_{a,h} \leq \phi\}. \tag{186}$$

(d) $\mathrm{epi}(\phi)$ *is the intersection of closed half-spaces*

$$\mathrm{epi}(\phi) = \bigcap_{\mathfrak{l}_{a,h} \leq \phi} \mathrm{epi}\,(\mathfrak{l}_{a,h}). \tag{187}$$

Definition 6.3 ϕ is sub-differentiable at x if there exists $\mathfrak{l}_{a,h} \leq \phi$ such that $\phi(\mathsf{x}) = \mathfrak{l}_{a,h}(\mathsf{x})$. In the latter case we write $h \in \partial\phi(\mathsf{x})$.

Example 6.4 Check that $h \in \partial\phi(\mathsf{x})$ iff $\phi(\mathsf{x}) < \infty$ and

$$\phi(\mathsf{y}) - \phi(\mathsf{x}) \geq h \cdot (\mathsf{y} - \mathsf{x}) \tag{188}$$

for any $\mathsf{y} \in \mathbb{R}^d$.

Convex functions on $\mathbb{R}^d$ are always sub-differentiable at interior points of their effective domains. In general, sets $\partial\phi(\mathsf{x})$ may be empty, may be singletons or they may contain continuum of different slopes h.

Example 6.5 Find an example with $\partial\phi(\mathsf{x}) = \emptyset$. Prove that in general $\partial\phi(\mathsf{x})$ is closed and convex. Check that a convex ϕ is differentiable at x with $\nabla\phi(\mathsf{x}) = h$ iff $\partial\phi(\mathsf{x}) = \{h\}$.

Definition 6.6 Let $\phi : \mathbb{R}^d \mapsto \mathbb{R} \cup \infty$. The Legendre-Fenchel transform, or the convex conjugate, of ϕ is

$$\phi^*(\mathsf{x}) = \sup_h \{h \cdot \mathsf{x} - \phi(h)\}. \tag{189}$$

By construction, ϕ^* is always convex and lower-semicontinuous: Indeed,

$$\mathrm{epi}(\phi^*) = \bigcap_h \mathrm{epi}\left(\mathfrak{l}_{-\phi(h),h}\right),$$

which is obviously closed and convex.

Duality Let ϕ be a convex and lower-semicontinuous function. Let us say that h and x are a pair of conjugate points if $\mathsf{x} \in \partial\phi(h)$.

Theorem 6.7 *If* ϕ *is convex and lower-semicontinuous, then* $\left((\phi)^*\right)^* = \phi$. *In the latter situation, the notion of conjugate points is symmetric, namely the following*

are equivalent:

$$\mathsf{x} \in \partial\phi(h) \Leftrightarrow h \in \partial\phi^*(\mathsf{x}) \Leftrightarrow \phi(h) + \phi^*(\mathsf{x}) = h \cdot \mathsf{x}. \tag{190}$$

Let h, x be a pair of conjugate points. Then strict convexity of ϕ at h is equivalent to differentiability of ϕ^* at x. Namely,

$$\forall\, g \neq h,\ \phi(g) - \phi(h) > \mathsf{x} \cdot (g - h) \ \Leftrightarrow\ \nabla\phi^*(\mathsf{x}) = h. \tag{191}$$

Support and Minkowski Functions Let $\mathbf{K} \subset \mathbb{R}^d$ be a compact convex set with non-empty interior around the origin $0 \in \mathrm{int}\,(\mathbf{K})$.

Definition 6.8 The function

$$\chi_{\mathbf{K}}(h) = \begin{cases} 0, & \text{if } h \in \mathbf{K} \\ \infty, & \text{otherwise} \end{cases} \tag{192}$$

is called the characteristic function of $\mathbf{K}$.

The function

$$\tau_{\mathbf{K}}(\mathsf{x}) = \sup_{h \in \mathbf{K}} h \cdot \mathsf{x} = \max_{h \in \partial\mathbf{K}} h \cdot \mathsf{x} \tag{193}$$

is called the support function of $\mathbf{K}$.

The function

$$\alpha_{\mathbf{K}}(h) = \inf\{r > 0 \ :\ h \in r\mathbf{K}\} \tag{194}$$

is called the Minkowski function of $\mathbf{K}$.

As it will become apparent below functions $\chi_{\mathbf{K}}$, $\tau_{\mathbf{K}}$ and $\phi_{\mathbf{K}}$ are convex and lower-semicontinuous.

Duality Relation Between $\chi_{\mathbf{K}}$ and $\tau_{\mathbf{K}}$ The characteristic function $\chi_{\mathbf{K}}$ is convex and lower-semicontinuous since $\mathrm{epi}\,(\chi_{\mathbf{K}}) = \mathbf{K} \times [0, \infty)$. The support function $\tau_{\mathbf{K}}$ is the supremum of linear functions. As such it is homogeneous of order one. Also, (193) could be recorded in the form which makes $\tau_{\mathbf{K}}$ to be the convex conjugate:

$$\tau_{\mathbf{K}} = \chi_{\mathbf{K}}^* \text{ and, by Theorem 6.7, } \chi_{\mathbf{K}} = \tau_{\mathbf{K}}^*. \tag{195}$$

Since $\tau_{\mathbf{K}}$ is homogeneous, the latter reads as

$$\mathbf{K} = \bigcap_{\mathsf{n} \in \mathbb{S}^{d-1}} \{h \ :\ h \cdot \mathsf{n} \leq \tau_{\mathbf{K}}(\mathsf{n})\}. \tag{196}$$

By (192) if $h \in \partial\mathbf{K}$ then $\mathsf{x} \in \partial\chi_{\mathbf{K}}(h)$ if and only if x is in the direction of the outward normal to a hyperplane which touches $\partial\mathbf{K}$ at h. Thus, $\partial\chi_{\mathbf{K}}(h)$ is always a closed convex cone (which can be just a semi-line).

Other way around, $\partial\tau_{\mathbf{K}}(\mathsf{x})$ contains all boundary points $h \in \partial\mathbf{K}$, such that x is the direction of the outward normal to a supporting hyperplane at h. In particular,

$$\partial\tau_{\mathbf{K}}(r\mathsf{x}) = \partial\tau_{\mathbf{K}}(\mathsf{x}), \tag{197}$$

and $\partial\tau_{\mathbf{K}}(\mathsf{x}) \subset \partial\mathbf{K}$ is a closed convex facet (which can be just one point).

As a consequence: $\tau_{\mathbf{K}}$ is differentiable at $\mathsf{x} \neq 0$ iff the supporting hyperplane with the outward normal direction of x touches $\partial\mathbf{K}$ at exactly one point h. In particular $\tau_{\mathbf{K}}$ is differentiable at any $\mathsf{y} \neq 0$ iff $\partial\mathbf{K}$ is strictly convex.

Polarity Relation Between $\tau_{\mathbf{K}}$ and $\alpha_{\mathbf{K}}$ As for $\alpha_{\mathbf{K}}(h)$ the assumptions $\mathbf{K}$ is bounded and $0 \in \mathrm{int}\,(\mathbf{K})$ imply that for any $h \neq 0$, the value $\alpha_{\mathbf{K}}(h)$ is positive and finite. Consequently,

$$\frac{h}{\alpha_{\mathbf{K}}(h)} \in \partial\mathbf{K} \text{ and } \exists\, \mathsf{x} \neq 0 \text{ such that } \tau_{\mathbf{K}}(\mathsf{x}) = \frac{\mathsf{x}\cdot h}{\alpha_{\mathbf{K}}(h)}.$$

On the other hand, again since $\frac{h}{\alpha_{\mathbf{K}}(h)} \in \partial\mathbf{K}$,

$$\tau_{\mathbf{K}}\,(\mathsf{y}) \geq \frac{\mathsf{y}\cdot h}{\alpha_{\mathbf{K}}(h)},$$

for any $\mathsf{y} \in \mathbb{R}^d$. Since $\tau_{\mathbf{K}}$ is homogeneous of order one, we, therefore, conclude:

$$\alpha_{\mathbf{K}}(h) = \max\{h\cdot\mathsf{y} \,:\, \tau_{\mathbf{K}}(\mathsf{y}) \leq 1\}. \tag{198}$$

In other words, $\alpha_{\mathbf{K}}$ is the support function of the closed convex set:

$$\mathbf{K}^* = \{\mathsf{y} \,:\, \tau_{\mathbf{K}}(\mathsf{y}) \leq 1\}. \tag{199}$$

In particular, $\alpha_{\mathbf{K}}$ is convex and lower-semicontinuous.

Example 6.9 For any $\mathsf{x}, h \neq 0$, $\mathsf{x}\cdot h \leq \tau_{\mathbf{K}}(\mathsf{x})\alpha_{\mathbf{K}}(h)$. Furthermore,

$$\frac{\mathsf{x}\cdot h}{\tau_{\mathbf{K}}(\mathsf{x})\alpha_{\mathbf{K}}(h)} = 1 \;\Leftrightarrow\; \frac{\mathsf{x}}{\tau_{\mathbf{K}}(\mathsf{x})} \in \partial\alpha_{\mathbf{K}}(h) \;\Leftrightarrow\; \frac{h}{\alpha_{\mathbf{K}}(h)} \in \partial\tau_{\mathbf{K}}(\mathsf{x}). \tag{200}$$

Actually, by homogeneity it would be enough to establish (200) for $\mathsf{x} \in \partial\mathbf{K}^*$ (equivalently $\tau_{\mathbf{K}}(\mathsf{x}) = 1$) and $h \in \partial\mathbf{K}$ (equivalently $\alpha_K(h) = 1$). In the latter case, let us say that $\mathsf{x} \in \partial\mathbf{K}^*$ and $h \in \partial\mathbf{K}$ are in polar relation if $\mathsf{x}\cdot h = 1$.

Example 6.10 Let x, h be in polar relation. Then

$$\partial\mathbf{K} \text{ is strictly convex (smooth) at } h \text{ iff } \partial\mathbf{K}^* \text{ is smooth (strictly convex) at } \mathsf{x}. \tag{201}$$

Remark 6.11 Most of the above notions can be defined and effectively studied in much more generality than we do. In particular, one can go beyond assumptions of finite dimensions and non-empty interior.

Curves and Surfaces

Let M be a smooth $(d-1)$-dimensional surface (without boundary) embedded in $\mathbb{R}^d$. For $u \in M$ let $\mathfrak{n}(u) \in \mathbb{S}^{d-1}$ be the normal direction at u. T_uM is the tangent space to M at u. Thus $\mathfrak{n}$ is a map $\mathfrak{n} : M \mapsto \mathbb{S}^{d-1}$. It is called the Gauss map, and its differential $\mathrm{d}\mathfrak{n}_u$ is called the Weingarten map. Since $T_uM = T_{\mathfrak{n}(u)}\mathbb{S}^{d-1}$, we may consider $\mathrm{d}\mathfrak{n}_u$ as a linear map on the tangent space T_uM.

Exercise 6.12 Check that $\mathrm{d}\mathfrak{n}_u$ is self-adjoint (with respect to the usual Euclidean scalar product on $\mathbb{R}^d$). Hence, the eigenvalues $\chi_1, \ldots, \chi_{d-1}$ of $\mathrm{d}\mathfrak{n}_u$ are real, and the corresponding normalized eigenvectors $\mathfrak{v}_1, \ldots, \mathfrak{v}_{d-1}$ form an orthonormal basis of T_uM.

Definition 6.13 Eigenvalues $\chi_1, \ldots, \chi_{d-1} \geq 0$ of $\mathrm{d}\mathfrak{n}$ are called principal curvatures of M at u. The normalized eigenvectors $\mathfrak{v}_1, \ldots, \mathfrak{v}_{d-1}$ are called directions of principal curvature. The product $\prod_\ell \chi_\ell$ is called the Gaussian curvature.

Assume that M is locally given by a level set of a smooth function $\lambda(\cdot)$, such that $\nabla\lambda(u) \neq 0$. That is, in a neighbourhood of u; $v \in M \leftrightarrow \lambda(v) = 0$. Then,

$$\mathfrak{n}_v = \frac{\nabla\lambda(v)}{|\nabla\lambda(v)|}.$$

Define $\Sigma_u = \mathrm{Hess}[\lambda](u)$.

Example 6.14 Check that for $g \in T_uM$,

$$\mathrm{d}\mathfrak{n}_u g \cdot g = \frac{1}{|\nabla\lambda(v)|}\Sigma_u g \cdot g. \tag{202}$$

Convex Surfaces Let now $M = \partial\mathbf{K}$, and $\mathbf{K}$ is a bounded convex body with non-empty interior. In the sequel we shall assume that the boundary $\partial\mathbf{K}$ is smooth (at least C_2). Let $\tau = \tau_{\mathbf{K}}$ be the support function of $\mathbf{K}$. Whenever defined the Hessian $\Xi_{\mathsf{x}} = \mathrm{Hess}[\tau](\mathsf{x})$ has a natural interpretation in terms of the curvatures of $\mathbf{K}$ at $h = \nabla\tau(\mathsf{x})$. We are following Chap. 2.5 in [26].

General Case We assume that M is smooth and that the Gaussian curvature of M is uniformly non-zero. In particular, $M = \partial\mathbf{K}$ is strictly convex, and, by duality relations, its support function τ is differentiable. In the sequel $\mathfrak{n}(h)$ is understood as the exterior normal to M at h.

Recall that for every $\mathsf{x} \neq 0$, the gradient $\nabla\tau(\mathsf{x}) = h_{\mathsf{x}} \in \partial\mathbf{K}$, and could be characterized by $h_{\mathsf{x}} \cdot \mathsf{x} = \tau(\mathsf{x})$. Consequently, $\nabla\tau(r\mathsf{x}) = \nabla\tau(\mathsf{x})$. This is a homogeneity relation. It readily implies the following: Let $\mathit{\Xi}_{\mathsf{x}}$ be the Hessian of τ at x. Then,

$$\mathit{\Xi}_{\mathsf{x}}\mathsf{x} = 0. \tag{203}$$

Let $h \in \partial\mathbf{K}$ and let $\mathfrak{n} = \mathfrak{n}(h)$ be the normal direction to $\partial\mathbf{K}$ at h. Then

$$\nabla\tau(\mathfrak{n}(h)) = h.$$

In other words, the restriction of $\nabla\tau$ to $\mathbb{S}^{d-1}$ is precisely the inverse of the Gauss map $\mathfrak{n}$. Hence the restriction $\hat{\mathit{\Xi}}_{\mathfrak{n}(h)}$ of $\mathit{\Xi}_{\mathfrak{n}}$ to $T_h\partial\mathbf{K}$ is the inverse of the Weingarten map $\mathrm{d}\mathfrak{n}_h$.

Definition 6.15 Let $\mathfrak{n} \in \mathbb{S}^{d-1}$ and $h = \nabla\tau(\mathfrak{n})$. Eigenvalues $r_\ell = 1/\chi_\ell$ of $\hat{\mathit{\Xi}}_{\mathfrak{n}}$ are called principal radii of curvature of $M = \partial\mathbf{K}$ at h.

Example: Smooth Convex Curves Let $\mathfrak{n}_\theta = (\cos\theta, \sin\theta)$. Radius of curvature $r(\theta) = 1/\chi(\theta)$ of the boundary $\partial\mathbf{K}$ at a point $h_\theta = \nabla\tau(\mathfrak{n}_\theta)$ is given by

$$r(\theta) = \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\tau(\theta) + \tau(\theta),$$

where we put $\tau(\theta) = \tau(\mathfrak{n}_\theta)$. Indeed, $\mathsf{v}_\theta = \mathfrak{n}'_\theta = (-\sin\theta, \cos\theta)$ is the unit spanning vector of $T_{h_\theta}\partial\mathbf{K}$. Note that $\frac{\mathrm{d}}{\mathrm{d}\theta}\mathsf{v}_\theta = -\mathfrak{n}_\theta$. Hence,

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\tau(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\nabla\tau(\mathfrak{n}_\theta)\cdot\mathsf{v}_\theta\right) = \mathit{\Xi}_{\mathfrak{n}_\theta}\mathsf{v}_\theta\cdot\mathsf{v}_\theta - \nabla\tau(\mathfrak{n}_\theta)\cdot\mathfrak{n}_\theta = \mathit{\Xi}_{\mathfrak{n}_\theta}\mathsf{v}_\theta\cdot\mathsf{v}_\theta - \tau(\mathfrak{n}_\theta)$$

Second Order Expansion Let $(\mathfrak{v}_1, \ldots, \mathfrak{v}_{d-1}, \mathfrak{n}(h))$ be orthonormal coordinate frame, where $(\mathfrak{v}_1, \ldots, \mathfrak{v}_{d-1})$ is a basis of T_hM. Consider matrix elements $\mathit{\Xi}_{\mathfrak{n}}(i,j)$ in this coordinates. Then the homogeneity relation (203) applied at $\mathsf{x} = \mathfrak{n}(h)$ yields:

$$\mathit{\Xi}_{\mathfrak{n}}(\ell, d) = 0 \text{ for all } \ell = 1, \ldots, d. \tag{204}$$

Which means that as a quadratic form $\mathit{\Xi}_{\mathfrak{n}}$ satisfies:

$$\mathit{\Xi}_{\mathfrak{n}}\mathsf{u}\cdot\mathsf{w} = \mathit{\Xi}_{\mathfrak{n}}\pi_h\mathsf{u}\cdot\pi_h\mathsf{v}, \tag{205}$$

where π_h is the orthogonal projection on $T_h\partial\mathbf{K}$. Furthermore,

Example 6.16 Check that the Hessian $\Xi_{\mathsf{x}} \stackrel{\Delta}{=} \mathrm{Hess}_{\mathsf{x}}\tau = \frac{1}{|\mathsf{x}|}\Xi_{\mathfrak{n}}$, where $\mathfrak{n} = \mathfrak{n}_{\mathsf{x}} \in \mathbb{S}^{d-1}$ is the unit vector in the direction of x.

Consequently, second order expansion takes the form: For any $\mathsf{x} \neq 0$ and $t \in (0,1)$

$$\tau_\lambda(t\mathsf{x}+\mathsf{v}) + \tau_\lambda((1-t)\mathsf{x}-\mathsf{v}) - \tau_\lambda(\mathsf{x}) = \frac{\Xi_{\mathfrak{n}}\mathsf{v}\cdot\mathsf{v}}{2t(1-t)\,|\mathsf{x}|} + \mathrm{o}\left(\frac{|\mathsf{v}|^2}{|\mathsf{x}|}\right). \tag{206}$$

Recording this in the (orthonormal) basis of principal curvatures, we deduce the following Corollary:

Corollary 6.17 *Let $\mathsf{x} \in \mathbb{R}^d$; $\mathfrak{n} = \mathfrak{n}_{\mathsf{x}} = \frac{\mathsf{x}}{|\mathsf{x}|} \in \mathbb{S}^{d-1}$, and let $h = \nabla\tau(\mathsf{x})$. Consider the orthogonal frame $(\mathfrak{v}_1, \ldots, \mathfrak{v}_{d-1}, \mathfrak{n})$, where $\mathfrak{v}_\ell$-s are the directions of principal curvature of $\partial\mathbf{K}$ at h. Then, for any $t \in (0,1)$ and for any $y_1, \ldots y_{d-1}$,*

$$\begin{aligned}\tau\left(t\mathsf{x} + \sum_{\ell=1}^{d-1} y_\ell \mathfrak{v}_\ell\right) + \tau\left((1-t)\mathsf{x} - \sum_{\ell=1}^{d-1} y_\ell\mathfrak{v}_\ell\right) - \tau(\mathsf{x}) \\ = \sum_{\ell=1}^{d-1} \frac{y_\ell^2}{2t(1-t)\,|\mathsf{x}|\,\chi_\ell} + \mathrm{o}\left(\frac{\sum y_\ell^2}{|\mathsf{x}|}\right).\end{aligned} \tag{207}$$

Strict Triangle Inequality If principal curvatures of $\partial\mathbf{K}$ are uniformly bounded or, equivalently, if quadratic forms $\Xi_{\mathfrak{n}}$ are uniformly (in $\mathfrak{n} \in \mathbb{S}^{d-1}$) positive definite, then there exists a constant $c > 0$ such that

$$\tau(\mathsf{x}) + \tau(\mathsf{y}) - \tau(\mathsf{x}+\mathsf{y}) \geq c\,(|\mathsf{x}| + |\mathsf{y}| - |\mathsf{x}+\mathsf{y}|)\,. \tag{208}$$

In order to prove (208) note, first of all, that since for any $\mathsf{z} \neq 0$, $\nabla\tau(\mathsf{z}) = \nabla\tau(\mathfrak{n}_{\mathsf{z}})$, one can rewrite the left hand side of (208) as

$$\tau(\mathsf{x}) + \tau(\mathsf{y}) - \tau(\mathsf{x}+\mathsf{y}) = \mathsf{x}\cdot\left(\nabla\tau(\mathfrak{n}_{\mathsf{x}}) - \nabla\tau(\mathfrak{n}_{\mathsf{x}+\mathsf{y}})\right) + \mathsf{y}\cdot\left(\nabla\tau(\mathfrak{n}_{\mathsf{y}}) - \nabla\tau(\mathfrak{n}_{\mathsf{x}+\mathsf{y}})\right).$$

Similarly,

$$|\mathsf{x}| + |\mathsf{y}| - |\mathsf{x}+\mathsf{y}| = \mathsf{x}\cdot\left(\mathfrak{n}_{\mathsf{x}} - \mathfrak{n}_{\mathsf{x}+\mathsf{y}}\right) + \mathsf{y}\cdot\left(\mathfrak{n}_{\mathsf{y}} - \mathfrak{n}_{\mathsf{x}+\mathsf{y}}\right).$$

Therefore, (208) will follow if we show that for any two unit vectors $\mathfrak{n}, \mathfrak{m} \in \mathbb{S}^{d-1}$,

$$\mathfrak{n}\cdot(\nabla\tau(\mathfrak{n}) - \nabla\tau(\mathfrak{m})) \geq c\,\mathfrak{n}\cdot(\mathfrak{n}-\mathfrak{m})\,. \tag{209}$$

Set $\Delta = \mathfrak{n} - \mathfrak{m}$. Since $\mathfrak{n} \cdot (\mathfrak{n} - \mathfrak{m}) \overset{<}{\sim} |\Delta|^2$, and since we are not pushing for the optimal value of c in (208), it would be enough to consider second order expansion in $|\Delta|$.

To this end define $\gamma_t = \mathfrak{m} + t\Delta$ and $h_t = \nabla\tau(\gamma_t)$. Then,

$$\begin{aligned} \mathfrak{n} \cdot (\nabla\tau(\mathfrak{n}) - \nabla\tau(\mathfrak{m})) &= \mathfrak{n} \cdot \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} \nabla\tau(\gamma_t)\mathrm{d}t = \int_0^1 \Xi_{\gamma_t} \mathfrak{n} \cdot \Delta \mathrm{d}t \\ &= \int_0^1 \Xi_{\gamma_t} \pi_{h_t} \mathfrak{n} \cdot \pi_{h_t} \Delta \mathrm{d}t. \end{aligned} \tag{210}$$

The last equality above is (205). By construction γ_t is orthogonal to $T_{h_t}\partial\mathbf{K}$. Hence the projection

$$\pi_{h_t}\mathfrak{n} = \mathfrak{n} - \frac{\mathfrak{n} \cdot \gamma_t}{|\gamma_t|^2}\gamma_t = (1-t)\Delta + \frac{(1-t)\Delta \cdot \gamma_t}{|\gamma_t|^2}\gamma_t = (1-t)\Delta + \mathrm{o}(|\Delta|).$$

On the other hand $\pi_{h_t}\Delta = \Delta + \mathrm{o}(|\Delta|)$. Hence, up to higher order terms in $|\Delta|$,

$$\int_0^1 \Xi_{\gamma_t} \pi_{h_t} \mathfrak{n} \cdot \pi_{h_t} \Delta \mathrm{d}t \geq \frac{1}{2} \min_{h \in \partial\mathbf{K}} \min_{\ell} r_\ell(h) \, |\mathfrak{n} - \mathfrak{m}|^2, \tag{211}$$

and (208) follows.

Large Deviations

The Setup Although the framework of the theory is much more general we shall restrict attention to probabilities on finite-dimensional spaces. Let $\{\mathbb{P}_n\}$ be a family of probability measures on $\mathbb{R}^d$.

Definition 6.18 A function $J : \mathbb{R}^d \mapsto [0, \infty]$ is said to be a rate function if it is proper ($\mathrm{Dom}(J) \overset{\Delta}{=} \{\mathsf{x} : J(\mathsf{x}) < \infty\} \neq \emptyset$) and if it has compact level sets. In particular rate functions are always lower-semicontinuous.

Definition 6.19 A family $\{\mathbb{P}_n\}$ satisfies large deviation principle with rate function J (and speed n) if:

Upper Bound For every closed $F \subseteq \mathbb{R}^d$

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}_n(F) \leq - \inf_{\mathsf{x} \in F} J(\mathsf{x}). \tag{212}$$

Lower Bound For every open $O \subseteq \mathbb{R}^d$

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_n(O) \geq -\inf_{\mathsf{x}\in O} J(\mathsf{x}). \tag{213}$$

There is an alternative formulation of the lower bound:

Exercise 6.20 Check that (213) is equivalent to: For every $\mathsf{x} \in \mathbb{R}^d$ the family $\{\mathbb{P}_n\}$ satisfies the LD lower bound at x, that is for any open neighbourhood O of x,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_n(O) \geq -J(\mathsf{x}). \tag{214}$$

All the measures we shall work with are exponentially tight:

Definition 6.21 A family $\{\mathbb{P}_n\}$ is exponentially tight if for any R one can find a compact subset K_R of $\mathbb{R}^d$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_n\left(K_R^{\mathsf{c}}\right) \leq -R. \tag{215}$$

If exponential tightness is checked then one needs derive upper bounds only for all compact sets:

Exercise 6.22 Check that if $\{\mathbb{P}_n\}$ is exponentially tight and it satisfies (213) for all open sets and (212) for all *compact* sets, then it satisfies LDP.

In particular, $\{\mathbb{P}_n\}$ satisfies an upper large deviation bound with J if

(a) It is exponentially tight.
(b) For every $\mathsf{x} \in \mathbb{R}^d$, the family $\{\mathbb{P}_n\}$ satisfies the following upper large deviation bound at x:

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_n\left(\left|\frac{\mathsf{X}}{n} - \mathsf{x}\right| \leq \delta\right) \leq -J(\mathsf{x}). \tag{216}$$

We shall mostly work with measures on $\frac{1}{n}\mathbb{Z}^d$ which are generated by scaled random variables $\frac{1}{n}\mathsf{X}$, for instance when $\mathsf{X} = \mathsf{X}(\gamma)$ is the spatial extension of a polymer or the end point of a self-interacting random walk. In the latter case we shall modify the notion (216) of point-wise LD upper bound as follows:

Definition 6.23 A family $\{\mathbb{P}_n\}$ of probability measures on $\mathbb{Z}^d$ satisfies an upper LD bound at $\mathsf{x} \in \mathbb{R}^d$ if for any $R > 0$

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_n(\mathsf{X} = \lfloor n\mathsf{x} \rfloor) \leq -J(\mathsf{x}) \wedge R. \tag{217}$$

At a first glance constant R in (217) does not seem to contribute to the statement. However, checking and formulating things this way may be convenient.

Exercise 6.24 Let J be a rate function. Check that if $\{\mathbb{P}_n\}$ is exponentially tight, if the lower bound (214) is satisfied, and if (217) is satisfied, for any $R \in [0, \infty)$, *uniformly* on compact subsets of $\mathbb{R}^d$, then $\{\mathbb{P}_n\}$ satisfies the LD principle in the sense of Definition 6.19.

Log-Moment Generating Functions and Convex Conjugates Frequently quests after LD rate functions stick to the following pattern: Assume that the (limiting) log-moment generating function

$$\lambda(h) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}_n e^{h\cdot \mathsf{x}} \tag{218}$$

is well defined (and not identically ∞) for all $h \in \mathbb{R}^d$.

Exercise 6.25 Check that if $\lambda(\cdot)$ in (218) is indeed defined, then it is convex and lower-semicontinuous.

Consider the Legendre-Fenchel transform I of λ

$$I(\mathsf{x}) = \sup_h \{h \cdot \mathsf{x} - \lambda(h)\}. \tag{219}$$

Here is one of the basic general LD results:

Theorem 6.26 *Assume that λ in (218) is well defined and proper.*

Upper Bound. *For any $\mathsf{x} \in \mathbb{R}^d$ the family $\{\mathbb{P}_n\}$ satisfies upper LD bound (217) with I at x.*

Lower Bound. *If, in addition, I is sub-differential and strictly convex at x, then $\{\mathbb{P}_n\}$ satisfies a lower LD bound at x with I in (214).*

Sub-differentiability and strict convexity over finite-dimensional spaces are studied in great generality (e.g. low-dimensional effective domains, behaviour at the boundary of relative interiors etc.) and detail [24].

Lower LD bounds with I generically *do not* hold. In particular true LD rate functions J are generically *non-convex*. However, $I = J$ in many important examples such as sums of i.i.d.-s and Markov chains. Moreover, $I = J$ for most of polymer models with purely attractive or repulsive interactions. A notable exception is provided by one-dimensional polymers with repulsion [12, 20] (which we do not discuss here). Under minor additional integrability conditions the relation between I and J could be described as follows:

Let ϕ be a function on $\mathbb{R}_+$ with a super-linear growth at ∞:

$$\lim_{t\to\infty} \frac{\phi(t)}{t} = \infty.$$

Lemma 6.27 *Assume that $\{\mathbb{P}_n\}$ satisfies LDP with rate function J, and assume that*

$$\limsup \frac{1}{n}\mathbb{E}_n e^{n\phi\left(\frac{|\mathsf{X}|}{n}\right)} < \infty. \tag{220}$$

Then [7] $\lambda(\cdot)$ in (218) is defined and equals to

$$\lambda(h) = \sup_{\mathsf{x}} \{\mathsf{x} \cdot h - J(\mathsf{x})\}.$$

Consequently, I is the convex lower-semicontinuous envelop of J, that is

$$I(\mathsf{x}) = \sup \{\mathfrak{l}_{a,h}(\mathsf{x}) \ : \ \mathfrak{l}_{a,h} \leq J\}.$$

In particular, if J convex, then $I = J$.

References

1. P. Antal, Enlargement of obstacles for the simple random walk. Ann. Probab. **23**(3), 1061–1101 (1995)
2. T. Bonnesen, W. Fenchel, *Theory of Convex Bodies*, ed. by L. Boron, C. Christenson, B. Smith (BCS Associates, Moscow, 1987). Translated from the German
3. F. Comets, T. Shiga, N. Yoshida, Probabilistic analysis of directed polymers in a random environment: a review, in *Stochastic Analysis on Large Scale Interacting Systems*. Advanced Studies in Pure Mathematics, vol. 39 (Mathematical Society of Japan, Tokyo, 2004), pp. 115–142
4. P.-G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, London 1979)
5. F. den Hollander, *Large Deviations*. Fields Institute Monographs, vol. 14 (American Mathematical Society, Providence, 2000)
6. A. Dembo, O. Zeitouni, Large deviations techniques and applications, in *Stochastic Modelling and Applied Probability*, vol. 38 (Springer, Berlin, 2010). Corrected reprint of the second (1998) edition
7. J.-D. Deuschel, D.W. Stroock, *Large Deviations*. Pure and Applied Mathematics, vol. 137 (Academic, Boston, 1989)
8. M.D. Donsker, S.R.S. Varadhan, Asymptotics for the Wiener sausage. Commun. Pure Appl. Math. **28**(4), 525–565 (1975)
9. M.D. Donsker, S.R.S. Varadhan, On the number of distinct sites visited by a random walk. Commun. Pure Appl. Math. **32**(6), 721–747 (1979)
10. M. Flury, Large deviations and phase transition for random walks in random nonnegative potentials. Stoch. Process. Appl. **117**(5), 596–612 (2007)
11. M. Flury, Coincidence of Lyapunov exponents for random walks in weak random potentials. Ann. Probab. **36**(4), 1528–1583 (2008)
12. A. Greven, F. den Hollander, A variational characterization of the speed of a one-dimensional self-repellent random walk. Ann. Appl. Probab. **3**(4), 1067–1099 (1993)
13. J.M. Hammersley, Generalization of the fundamental theorem on subadditive functions. Proc. Camb. Philos. Soc. **58**(2), 235–238 (1962)

14. D. Ioffe, Y. Velenik, Ballistic phase of self-interacting random walks, in *Analysis and Stochastics of Growth Processes and Interface Models* (Oxford University Press, Oxford, 2008), pp. 55–79
15. D. Ioffe, Y. Velenik, Crossing random walks and stretched polymers at weak disorder. Ann. Probab. **40**(2), 714–742 (2012)
16. D. Ioffe, Y. Velenik, Stretched polymers in random environment, in *Probability in Complex Physical Systems*. Springer Proceedings in Mathematics, vol. 11 (Springer, Berlin/Heidelberg, 2012), pp. 339–369
17. D. Ioffe, Y. Velenik, Self-attracting random walks: the case of critical drifts. Commun. Math. Phys. **313**, 209–235 (2012)
18. D. Ioffe, Y. Velenik, An almost sure CLT for stretched polymers. Electron. J. Probab. **18**(97), 1–20 (2013)
19. L. Kaup, B. Kaup, *Holomorphic Functions of Several Variables*. de Gruyter Studies in Mathematics, vol. 3 (Walter de Gruyter, Berlin, 1983)
20. W. König, The drift of a one-dimensional self-repellent random walk with bounded increments. Probab. Theory Relat. Fields **100**(4), 513–544 (1994)
21. H. Lacoin, New bounds for the free energy of directed polymers in dimension $1+1$ and $1+2$. Commun. Math. Phys. **294**(2), 471–503 (2010)
22. D.L. McLeish, A maximal inequality and dependent strong laws. Ann. Probab. **3**, 829–839 (1975)
23. J.-C. Mourrat, Lyapunov exponents, shape theorems and large deviations for the random walk in random potential. ALEA Lat. Am. J. Probab. Math. Stat. **9**, 165–211 (2012)
24. R.T. Rockafellar, *Convex Analysis*. Princeton Mathematical Series, vol. 28 (Princeton University Press, Princeton, 1970)
25. M. Rubinstein, R.H. Colby, *Polymer Physics* (Oxford University Press, Oxford, 2003)
26. R. Schneider, *Convex Bodies: The Brunn-Minkowski Theory*. Encyclopedia of Mathematics and Its Applications, vol. 44 (Cambridge University Press, Cambridge, 1993)
27. Y.G. Sinai, A remark concerning random walks with random potentials. Fund. Math. **147**(2), 173–180 (1995)
28. A.-S. Sznitman, Crossing velocities and random lattice animals. Ann. Probab. **23**(3), 1006–1023 (1995)
29. A.-S. Sznitman, *Brownian Motion, Obstacles and Random Media* (Springer, Berlin, 1998)
30. S.R.S. Varadhan, *Large Deviations and Applications*. CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 46 (Society for Industrial and Applied Mathematics, Philadelphia, 1984)
31. M.P.W. Zerner, Directional decay of the Green's function for a random nonnegative potential on $\mathbf{Z}^d$. Ann. Appl. Probab. **8**(1), 246–280 (1998)
32. N. Zygouras, Lyapounov norms for random walks in low disorder and dimension greater than three. Probab. Theory Relat. Fields **143**(3–4), 615–642 (2009)
33. N. Zygouras, Strong disorder in semidirected random polymers. Ann. Inst. Henri Poincaré Probab. Stat. **49**(3), 753–780 (2013)

Loop Measures and the Gaussian Free Field

Gregory F. Lawler and Jacob Perlman

1 Introduction

Loop measures have become important in the analysis of random walks and fields arising from random walks. Such measures appear in work of Symanzik [9] but the recent revival came from the Brownian loop soup [7] which arose in the study of the Schramm-Loewner evolution. The random walk loop soup is a discrete analogue for which one can show convergence to the Brownian loop soup. The study of such measures and soups has continued: in continuous time by Le Jan [8] and in discrete time in [5, 6]. The purpose of this note is to give an introduction to the discrete time measures and to discuss two of the applications: the relation with loop-erased walk and spanning trees, and a distributional identity between a function of the loop soup and the square of the Gaussian free field. This paper is not intended to be a survey but only a sample of the uses of the loop measure.

While the term "loop measure" may seem vague, we are talking about a specific measure from which a probabilistic construction, the "loop soup" is derived. We are emphasizing the loop measure rather than the loop soup which is a Poissonian realization of the measure because we want to allow the loop measure to take negative or complex values. However, we do consider the loop soup as a complex measure. Measures with negative and complex weights can arise even when studying probabilistic objects; for example, sharp asymptotics for the planar loop-erased random walk were derived in [4] using a loop measure with signed weights.

We will start with some basic definitions. In many ways, the loop measure can be considered a way to understand matrices, especially determinants, and some of the results have very classical counterparts. Most of the theorems about the basic

G.F. Lawler (✉) • J. Perlman
Department of Mathematics, University of Chicago, 5734 S. University Avenue, Chicago, IL 60637, USA
e-mail: lawler@math.uchicago.edu

M. Biskup et al. (eds.), *Random Walks, Random Fields, and Disordered Systems*,
Lecture Notes in Mathematics 2144, DOI 10.1007/978-3-319-19339-7_5

properties can be found in [5, Chap. 9] although that book restricted itself to positive measures. We redo some proofs just to show that positivity of the entries is not important. A key fact is that the total mass of the loop measure is the negative of the logarithm of the determinant of the Laplacian.

We next introduce the loop-erased random walk and show how one can use loop measures to give a short proof of Kirchhoff matrix-tree theorem by using an algorithm due to David Wilson for generating uniform spanning trees.

Our next section describes an isomorphism theorem found by Le Jan that is related to earlier isomorphism theorems of Brydges et al. [1, 2] and Dynkin [3]. In this case, one shows that the local time of a continuous time version of the loop soup has the same distribution as the square of a Gaussian field. Le Jan established this by constructing a continuous-time loop soup. We choose a slightly different, but essentially equivalent, method of using the discrete loop soup and then adding exponential waiting times. This is similar to the construction of continuous time Markov chains by starting with a discrete time chain and then adding the waiting times. In order to get the formulas to work, one needs to consider a correction term that is given by "trivial loops".

We finally give some discussion of complex Gaussian fields with positive definite Hermitian weights. We first consider real (signed) weights and relate this to the real Gaussian free field. Finally we consider a complex Gaussian field and show that it can be considered as a pair of real Gaussian fields.

2 Definitions

We will consider edge weights, perhaps complex valued, on a finite state space A. A set of weights is the same thing as a matrix Q indexed by A.

- We call Q *acceptable* if the matrix with entries $|Q(x, y)|$ has all eigenvalues in the interior of the unit disc. (This is not a standard term, but we will use it for convenience.)
- We say Q is *positive* if the entries are nonnegative and Q is *real* if the entries are real.
- As usual, we say that Q is *symmetric* if $Q(x, y) = Q(y, x)$ for all x, y and Q is *Hermitian* if $Q(x, y) = \overline{Q(y, x)}$ for all x, y.
- If Q is Hermitian we say that Q is *positive definite* if all the eigenvalues are strictly greater than zero, or equivalently if $\overline{\mathbf{x}} \cdot Q\mathbf{x} > 0$ for all non-zero $\mathbf{x}$.

If $A \subsetneq A'$ and Q is the transition matrix for an irreducible Markov chain on A', then Q restricted to A is positive and acceptable. This is one of the main examples of interest. If Q is any matrix, then λQ is acceptable for $\lambda > 0$ sufficiently small.

If $V \subset A$ with k elements, we will write Q_V for the $k \times k$ matrix obtained by restricting Q to V. A *path* in A of length n is a finite sequence of points

$$\omega = [\omega_0, \ldots, \omega_n], \quad \omega_j \in A.$$

We write $|\omega| = n$ for the number of steps in the path and ω^R for the reversed path

$$\omega^R = [\omega_n, \ldots, \omega_0].$$

We allow the trivial paths with $|\omega| = 0$. We write $\mathscr{K}_{x,y}(A)$ for the set of all paths in A with $\omega_0 = x, \omega_n = y$; if $x = y$, we include the trivial path.

The matrix Q gives the *path measure* defined by

$$Q(\omega) = \prod_{j=1}^{n} Q(\omega_{j-1}, \omega_j), \quad \omega = [\omega_0, \ldots, \omega_n] \in \bigcup_{x,y \in A} \mathscr{K}_{x,y}(A),$$

where $Q(\omega) = 1$ if $|\omega| = 0$. Note that if Q is Hermitian, then $Q(\omega^R) = \overline{Q(\omega)}$. A path ω is a *(rooted) loop (rooted at ω_0)* if $\omega_0 = \omega_n$. Note that we write Q both for the edge weights (matrix entries) and for the induced measure on paths.

We let $\Delta = I - Q$ denote the Laplacian. We write $G(x, y) = G^Q(x, y)$ for the Green's function that can be defined either as

$$G = \Delta^{-1} = \sum_{j=0}^{\infty} Q^j$$

or by

$$G(x, y) = Q[\mathscr{K}_{x,y}(A)] = \sum_{\omega \in \mathscr{K}_{x,y}(A)} Q(\omega).$$

Provided Q is acceptable, these sums converge absolutely. We write

$$G(x, y) = G_R(x, y) + i\, G_I(x, y),$$

where G_R, G_I are real matrices.

Let

$$f_x = \sum Q(\omega) \tag{1}$$

where the sum is over all paths ω from x to x of length at least one that have no other visits to x. A standard renewal argument shows that

$$G(x, x) = \sum_{k=0}^{\infty} f_x^k, \tag{2}$$

and since the sum is convergent,

$$|f_x| < 1.$$

If $V \subset A$, we will write

$$G_V(x,y) = G^{Q_V}(x,y) = \sum_{\omega \in \mathcal{K}_{x,y}(V)} Q(\omega),$$

for the corresponding Green's function associated to paths in V. The next proposition is a well known relation between the determinant of the Laplacian and the Green's function.

Proposition 2.1 *If $A = \{x_1, \ldots, x_n\}$ and $A_j = A \setminus \{x_1, \ldots, x_{j-1}\}$,*

$$\frac{1}{\det \Delta} = \prod_{j=1}^{n} G_{A_j}(x_j, x_j).$$

Proof By induction on n. If $n = 1$ and $q = Q(x_1, x_1)$, there is exactly one path of length k in A_1 and it has measure q^k. Therefore

$$G_{A_1}(x_1, x_1) = \sum_{k=0}^{\infty} q^k = \frac{1}{1-q}.$$

Assume the result is true for each $A_j \subsetneq A$, and note that if $g(x) = G_{A_j}(x, x_j)$, then

$$[I - Q_{A_j}]\, g = \delta_{x_j}$$

Using Cramer's rule to solve this linear system. we see that

$$G_{A_j}(x_j, x_j) = \frac{\det[I - Q_{A_{j+1}}]}{\det[I - Q_{A_j}]}.$$

Proposition 2.2 *If Q is a Hermitian acceptable matrix, then for each x, $G(x,x) > 0$. In particular, Δ and $G = \Delta^{-1}$ are positive definite Hermitian matrices.*

Proof It is immediate that Δ and G are Hermitian. If ω is a path in (1), then so is ω^R. Since $Q(\omega^R) = \overline{Q(\omega)}$, we can see that $\Im[f_x] = 0$, and hence $-1 < f_x < 1$. As in (2), we can write

$$G(x,x) = \sum_{k=0}^{\infty} f_x^k = \frac{1}{1 - f_x} > 0.$$

Combining this with Proposition 2.1, we see that each principal minor of Δ is positive and hence Δ is positive definite.

3 Loop Measures

3.1 Definition

Let $\mathscr{O} = \mathscr{O}(A)$ denote the set of rooted loops of strictly positive length. If Q is an acceptable weight, then the *(rooted) loop measure (associated to Q)* is the complex measure $m = m^Q$ on $\mathscr{O}$, given by

$$m(\omega) = \frac{Q(\omega)}{|\omega|}.$$

Note that the loop measure is *not* the same thing as the path measure restricted to loops. An *unrooted loop* is an equivalence class of rooted loops in $\mathscr{O}$ under the equivalence relation generated by

$$[\omega_0, \ldots, \omega_n] \sim [\omega_1, \ldots, \omega_n, \omega_1].$$

In other words, an unrooted loop is a loop for which one forgets the "starting point". We will write $\tilde{\omega}$ for unrooted loops and we let $\tilde{\mathscr{O}}$ denote the set of unrooted loops. We write $\omega \sim \tilde{\omega}$ if ω is in the equivalence class $\tilde{\omega}$. The measure m induces a measure that we call $\tilde{m}$ by

$$\tilde{m}(\tilde{\omega}) = \sum_{\omega \sim \tilde{\omega}} m(\omega).$$

We make several remarks.

- Unrooted loops have forgotten their roots but have not lost their orientation. In particular, $\tilde{\omega}$ and $\tilde{\omega}^R$ may be different unrooted loops.
- Since $Q(\omega)$ and $|\omega|$ are functions of the unrooted loop, we can write $Q(\tilde{\omega}), |\tilde{\omega}|$. If Q is Hermitian, then $Q(\tilde{\omega}^R) = \overline{Q(\tilde{\omega})}$.
- Let $d(\tilde{\omega})$ denote the number of rooted loops ω with $\omega \sim \tilde{\omega}$. Note that $d(\tilde{\omega})$ is an integer that divides $|\tilde{\omega}|$, but it is possible that $d(\tilde{\omega}) < |\tilde{\omega}|$. For example, if a, b, c are distinct elements and $\tilde{\omega}$ is the unrooted loop with representative

  $$\omega = [a, b, c, a, b, a, b, c, a, b, a],$$

  then $|\tilde{\omega}| = 10$ and $d(\tilde{\omega}) = 5$. Note that

  $$\tilde{m}(\tilde{\omega}) = \frac{d(\tilde{\omega})}{|\tilde{\omega}|} Q(\tilde{\omega}).$$

- Suppose that an unrooted loop $\tilde{\omega}$ with $|\tilde{\omega}| = n$ has $d = d(\tilde{\omega})$ rooted representative. In other words, the loop "repeats" itself after d steps and does n/d such repetitions. Suppose $k > 0$ of these rooted representatives are rooted

at x. In the example above, $k = 2$ for $x = a$ and $x = b$ and $k = 1$ for $x = c$. Then the total number of times that the loop visits x is $k(n/d)$. Suppose that we give each of the k loops that are rooted at x measure $Q(\tilde{\omega})/[kn/d]$ and give all the other rooted representatives of $\tilde{\omega}$ measure zero. Then the induced measure on unrooted loops is the same as the usual unrooted loop measure, giving measure $(d/n)\, Q(\tilde{\omega})$ to $\tilde{\omega}$.

- In other words, if we give each *rooted* loop rooted at x measure $Q(\omega)/k$ where k is the number of visits to x, then the induced measure on unrooted loops restricted to loops that intersect x is the same as $\tilde{m}$.
- One reason that the unrooted loop measure is useful is that one can move the root around to do calculations. The next lemma is an example of this.

Let

$$F(A) = F^{Q}(A) = \exp\left(\sum_{\tilde{\omega}\in\tilde{\mathscr{O}}} \tilde{m}(\tilde{\omega})\right) = \exp\left(\sum_{\omega\in\mathscr{O}} m(\omega)\right).$$

If $V \subset A$, we let

$$F_V(A) = \exp\left(\sum_{\tilde{\omega}\in\tilde{\mathscr{O}},\tilde{\omega}\cap V\neq\emptyset} \tilde{m}(\tilde{\omega})\right).$$

Note that $F_A(A) = F(A)$. If $V = \{x\}$, we write just $F_x(A)$. The next lemma relates the Green's function to the exponential of the loop measure; considering the case where Q is positive shows that the sum converges absolutely. As a corollary, we will have a relationship between the determinant of the Laplacian and the loop measure.

Lemma 3.1

$$F_x(A) = G(x,x).$$

More generally, if $V = \{x_1,\dots,x_l\} \subset A$ *and* $A_j = A \setminus \{x_1,\dots,x_{j-1}\}$, *then*

$$F_V(A) = \prod_{j=1}^{l} G_{A_j}(x_j,x_j).$$

Proof Let $\mathscr{A}_k$ denote the set of $\tilde{\omega} \in \tilde{\mathscr{O}}$ that have k different representatives that are rooted at x. By spreading the mass evenly over these k representatives, as described in the second and third to last bullets above, we can see that

$$\tilde{m}\,[\mathscr{A}_k] = \frac{1}{k} f_x^k.$$

Hence,

$$\tilde{m}\left[\bigcup_{k=1}^{\infty}\mathscr{A}_k\right] = \sum_{k=1}^{\infty}\frac{1}{k}f_x^k = -\log[1-f_x] = \log G(x,x).$$

This gives the first equality and by iterating this fact, we get the second equality.

Corollary 3.2

$$F(A) = \frac{1}{\det \Delta}.$$

Proof Let $A = \{x_1, \ldots, x_n\}, A_j = \{x_j, \ldots, x_n\}$. By Proposition 2.1 and Lemma 3.1,

$$\frac{1}{\det \Delta} = \prod_{j=1}^{n} G_{A_j}(x_j, x_j) = F(A).$$

Suppose f is a complex valued function defined on A to which we associate the diagonal matrix

$$D_f(x,y) = \delta_{x,y} f(x).$$

Let $Q_f = D_{1/(1+f)}\, Q$, that is,

$$Q_f(x,y) = \frac{Q(x,y)}{1+f(x)}.$$

If Q is acceptable, then for f sufficiently small, Q_f will be an acceptable matrix for which we can define the loop measure m_f. More specifically, if $\omega = [\omega_0, \ldots, \omega_n] \in \mathscr{O}$, then

$$m_f(\omega) = \frac{Q_f(\omega)}{|\omega|} = m(\omega)\prod_{j=1}^{n}\frac{1}{1+f(\omega_j)},$$

$$\tilde{m}_f(\tilde{\omega}) = \tilde{m}(\tilde{\omega})\prod_{j=1}^{n}\frac{1}{1+f(\omega_j)}.$$

Hence, if $G_f = G^{Q_f}$,

$$\det G_f = \exp\left(\sum_{\omega\in\mathscr{O}} m_f(\omega)\right) = \exp\left(\sum_{\tilde{\omega}\in\tilde{\mathscr{O}}} \tilde{m}_f(\tilde{\omega})\right).$$

Example Consider a one-point space $A = \{x\}$ with $Q(x,x) = q \in (0,1)$. For each $n > 0$, there is exactly one loop ω^n of length n with $Q(\omega^n) = q^n, m(\omega^n) = q^n/n$. Then, Δ is the 1×1 matrix with entry $1 - q$,

$$G_A(x,x) = \sum_{n=0}^{\infty} q^n = \frac{1}{1-q},$$

and

$$\sum_{\omega \in \mathscr{O}} m(\omega) = \sum_{n=1}^{\infty} \frac{q^n}{n} = -\log[1-q].$$

3.2 Relation to Loop-Erased Walk

Suppose $\overline{A}$ is a finite set, $A \subsetneq \overline{A}, \partial A = \overline{A} \setminus A$, and Q is a an acceptable matrix on $\overline{A}$. Let $\mathscr{K}(A)$ denote the set of paths $\omega = [\omega_0, \ldots, \omega_n]$ with $\omega_n \in \partial A$ and $\{\omega_0, \ldots, \omega_{n-1}\} \subset A$. For each path ω, there exists a unique loop-erased path $LE(\omega)$ obtained from ω by chronological loop-erasure as follows.

- Let $j_0 = \max\{j : \omega_j = \omega_0\}$.
- Recursively, if $j_k < n$, then $j_{k+1} = \max\{j : \omega_j = \omega_{j_k+1}\}$.
- If $j_k = n$, then $LE(\omega) = [\omega_{j_0}, \ldots, \omega_{j_k}]$.

If $\eta = [\eta_0, \ldots, \eta_k]$ is a self-avoiding path in $\mathscr{K}(A)$, we define its loop-erased measure by

$$\hat{Q}(\eta; A) = \sum_{\omega \in \mathscr{K}(A),\, LE(\omega) = \eta} Q(\omega).$$

The loop measure gives a convenient way to describe $\hat{Q}(\eta; A)$.

Proposition 3.3

$$\hat{Q}(\eta; A) = Q(\eta)\, F_\eta(A).$$

Proof We can decompose any path ω with $LE(\omega) = \eta$ uniquely as

$$l^0 \oplus [\eta_0, \eta_1] \oplus l^1 \oplus [\eta_1, \eta_2] \oplus \cdots \oplus l^{k-1} \oplus [\eta_{k-1}, \eta_k]$$

where l^j is a rooted loop rooted at η_j that is contained in $A_j := A \setminus \{\eta_0, \eta_1, \ldots, \eta_{j-1}\}$. By considering all the possibilities, we see that the measure of all walks with

$LE(\omega) = \eta$ is

$$G_A(\eta_0, \eta_0)\, Q(\eta_0, \eta_1)\, G_{A_1}(\eta_1, \eta_1) \cdots G_{A_{k-1}}(\eta_{k-1}, \eta_{k-1})\, Q(\eta_{k-1}, \eta_k),$$

which can be written as

$$Q(\eta) \prod_{j=0}^{k-1} G_{A_j}(\eta_j, \eta_j) = Q(\eta)\, F_\eta(A).$$

There is a nice application of this to spanning trees. Let $A = \{x_0, x_1, \ldots, x_n\}$ be the vertex set of a finite connected graph, and let Q be the transition probability for simple random walk on the graph, that is, $Q(x, y) = 1/d(x)$ if x and y are adjacent, where $d(x)$ is the degree of x. Consider the following algorithm due to David Wilson [10] to choose a spanning tree from A:

- Start with the trivial tree consisting of a single vertex, x_0, and no edges.
- Start a random walk at x_1 and run it until it reaches x_0. Erase the loops (chronologically) and add the edges of the loop-erased walk to the tree.
- Let x_j be the vertex of smallest index that has not been added to the tree yet. Start a random walk at x_j, let it run until it hits a vertex that has been added to the tree. Erase loops and add the remaining edges to the tree.
- Continue until we have a spanning tree.

It is a straightforward exercise using the last proposition to see that for any tree, the probability that it is chosen is exactly

$$\left[\prod_{j=1}^{n} d(x_j) \right]^{-1} F(A')$$

which by Corollary 3.2 can be written as

$$\left[\det[I - Q_{A'}] \prod_{j=1}^{n} d(x_j) \right]^{-1} = \frac{1}{\det[D - K]}.$$

Here $D(x, y) = \delta_{x,y}\, d(x)$ is the diagonal matrix of degrees and K is the adjacency matrix, both restricted to A'. (The matrix $D - K$ is what graph theorists call the Laplacian.) We can therefore conclude the following. The second assertion is a classical result due to Kirchhoff called the *matrix-tree theorem*.

Theorem 3.4 *Every spanning tree is equally likely to be chosen in Wilson's algorithm. Moreover, the total number of spanning trees is* $\det[D-K]$. *In particular,* $\det[D - K]$ *does not depend on the ordering* $\{x_0, \ldots, x_n\}$ *of the vertices of A.*

4 Loop Soup and Gaussian Free Field

4.1 Soups

If $\lambda > 0$, then the Poisson distribution on $\mathbb{N} = \{0, 1, 2, \dots\}$ is given by

$$q^\lambda(k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

We can use this formula to define the Poisson "distribution" for $\lambda \in \mathbb{C}$. In this case q^λ is a complex measure supported on $\mathbb{N}$ with variation measure $|q^\lambda|$ given by

$$|q^\lambda|(k) = |e^{-\lambda}| \frac{|\lambda|^k}{k!} = e^{-\Re(\lambda)} \frac{|\lambda|^k}{k!},$$

and total variation

$$\|q^\lambda\| = \sum_{k=0}^{\infty} |q^\lambda|(k) = \exp\{|\lambda| - \Re(\lambda)\} \le e^{2|\lambda|}.$$

Note that

$$\sum_{k=1}^{\infty} |q^\lambda|(k) = \exp\{|\lambda| - \Re(\lambda)\} [1 - e^{-|\lambda|}] \le |\lambda| e^{2|\lambda|}.$$

The usual convolution formula $q^{\lambda_1} * q^{\lambda_2} = q^{\lambda_1 + \lambda_2}$ holds, and if

$$\sum_{j=1}^{\infty} |\lambda_j| < \infty,$$

we can define the infinite convolution

$$\prod_j^* q^{\lambda_j} = \lim_{n \to \infty} (q^{\lambda_1} * \cdots * q^{\lambda_n}) = q^{\sum \lambda_j}.$$

If $\lambda > 0$ and M_t is a Poisson process with parameter λ, then the distribution of M_t is

$$q_t(\{k\}) = q^{t\lambda}(k) = e^{-t\lambda} \frac{(t\lambda)^k}{k!}, \quad k = 0, 1, 2, \dots \tag{3}$$

The family of measures $\{q_t\}$ satisfy the semigroup law $q_{s+t} = q_s * q_t$. If we are only interested in the measure q_t, then we may choose λ in (3) to be complex. In this case

the measures $\{q_t\}$ are not probability measures but they still satisfy the semigroup law. We call this the Poisson semigroup of measures with parameter λ and note that the Laplace transform is given by

$$\sum_{k=0}^{\infty} e^{k\alpha} q_t(\{k\}) = \exp(t\lambda(e^{\alpha} - 1)).$$

Suppose m is a complex measure on a countable set X, that is, a complex function with

$$\sum_{x \in X} |m(x)| < \infty.$$

Then we say that the soup generated by m is the semigroup of measures $\{q_t : t \geq 0\}$ on $\mathbb{N}^X$ where q_t is the product measure of $\{q_t^x : x \in X\}$ where $\{q_t^x : t \geq 0\}$ is a Poisson semigroup of measures with parameter $m(x)$. Pushing forward q_t along the map $\phi \mapsto \sum_{x \in X} \phi(x)$ to a measure on $\mathbb{N} \cup \{\infty\}$, we see that it agrees with $\prod^* q_t^x$ on $\mathbb{N}$ and thus q_t is supported on the pre-image of $\mathbb{N}$, the set of $\phi \in \mathbb{N}^X$ with finite support which we will call $\mathbb{N}^X_{\text{fin}}$. The complex measure q_t satisfies

$$\|q_t\| \leq \prod_{x \in X} \|q_t^x\| \leq \exp\left\{ 2t \sum_{x \in X} |m(x)| \right\}.$$

Soups were originally defined when m is a positive measure on X, in which case it is defined as an independent collection of Poisson processes $\{M_t^x : x \in X\}$ where M_t^x has rate $m(x)$. A realization $\mathscr{C}_t$ of the soup at time t is a multiset of X in which the element x appears M_t^x times. In this case q_t gives the distribution of the vector $(M_t^x : x \in X)$.

4.2 Loop Soup

Suppose Q is an acceptable weight with associated loop measure m. Let $0 < \epsilon < 1$ be such that the matrix with entries $P_\epsilon(x, y) := e^{\epsilon} |Q(x, y)|$ is still acceptable. Let m be the rooted loop measure associated to Q and note that

$$\sum_{\omega \in \mathscr{O}} |\omega| \, e^{\epsilon |\omega|} \, |m(\omega)| = \sum_{\omega \in \mathscr{O}} P_\epsilon(\omega) < \infty. \tag{4}$$

The *(rooted) loop soup* is a "Poissonian realization" of the measure m. To be more precise, recall that $\mathscr{O}$ is the set of rooted loops in A with positive length. A multiset $\mathscr{C}$ of loops is a generalized subset of $\mathscr{O}$ in which loops can appear more than once. In other words it is an element $\{\mathscr{C}(\omega) : \omega \in \mathscr{O}\}$ of $\mathbb{N}^{\mathscr{O}}$ where $\mathscr{C}(\omega)$

denotes the number of times that ω appears in $\mathscr{C}$. Then the rooted loop soup is the semigroup of measures $\mathscr{M}_t = \mathscr{M}_{t,m}$ on $\mathbb{N}^{\mathscr{O}}$ given by the product measure of the Poisson semigroups $\{\mathscr{M}_t^{\omega} : \omega \in \mathscr{O}\}$ where $\mathscr{M}_t^{\omega}$ has parameter $m(\omega)$. The measure $\mathscr{M}_t^{\omega}$ is Poisson with parameter $tm(\omega)$ and hence $\|\mathscr{M}_t^{\omega}\| \le \exp\{2t|m(\omega)|\}$ and

$$\sum_{k=1}^{\infty} |\mathscr{M}_t^{\omega}(k)| \le t|m(\omega)|\, e^{2t|m(\omega)|}. \tag{5}$$

For any $x \in A$ and rooted loop ω, we define the *(discrete) local time* $N^{\omega}(x)$ to be the number of visits of ω to x :

$$N^{\omega}(x) = \sum_{j=0}^{|\omega|-1} 1\{\omega_j = x\} = \sum_{j=1}^{|\omega|} 1\{\omega_j = x\}.$$

Note that this is a function of an unrooted loop, so we can also write $N^{\tilde{\omega}}(x)$. Also $N^{\omega^R}(x) = N^{\omega}(x)$. We define the additive function $L : \mathbb{N}^{\mathscr{O}}_{\text{fin}} \to \mathbb{N}^A$ by

$$L_{\mathscr{C}}(x) = \sum_{\omega \in \mathscr{O}} \mathscr{C}(\omega)\, N^{\omega}(x).$$

By pushing forward by L, the loop soup $\mathscr{M}_t$ induces a measure on $\mathbb{N}^A$ which we denote by $\mu_t = \mu_{t,m}$ and refer to as the *discrete occupation field.* Indeed, since $\mathscr{M}_t$ is a product measure, we can write μ_t as

$$\mu_t = \prod_{\omega \in \mathscr{O}}^{*} \mu_t^{\omega}$$

where the notation $\prod^*$ means convolution and μ_t^{ω} denotes the measure supported on $\{kN^{\omega} : k = 0, 1, 2, \ldots\}$ with

$$\mu_t^{\omega}(kN^{\omega}) = e^{-tm(\omega)}\, \frac{[tm(\omega)]^k}{k!}.$$

For future reference we note that since $N^{\omega} = N^{\omega^R}$,

$$[\mu_t^{\omega} * \mu_t^{\omega^R}](kN^{\omega}) = e^{-t[m(\omega)+m(\omega^R)]}\, \frac{t^k\, [m(\omega) + m(\omega^R)]^k}{k!},$$

and hence,

$$\mu_{2t} = \prod_{\omega \in \mathscr{O}}^{*} \mu_{2t}^{\omega} = \prod_{\omega \in \mathscr{O}}^{*} \mu_t^{\omega} * \mu_t^{\omega} = \prod_{\omega \in \mathscr{O}}^{*} \mu_t^{\omega} * \mu_t^{\omega^R} = \mu_{t,m^R}, \tag{6}$$

where

$$m^R(\omega) = m(\omega) + m(\omega^R).$$

4.3 A Continuous Occupation Field

In order to get a representation of the Gaussian free field, we need to change the discrete occupation field to a continuous time occupation field. We will do so in a simple way by replacing $N^\omega(x)$ with a sum of $N^\omega(x)$ independent rate one exponential random variables. This is similar to the method of constructing continuous time Markov chains from discrete time chains by adding exponential waiting times.

We say that a process $Y(t)$ is a *gamma process* if it has independent increments, $Y(0) = 0$, and for any $t, s \geq 0$, $Y(t+s) - Y(t)$ has a Gamma$(s, 1)$ distribution. In particular, $Y(n)$ is distributed as the sum of n independent rate one exponential random variables. Let $\{Y^x : x \in A\}$ be a collection of independent gamma processes. If $\bar{s} = \{s_x : x \in A\} \in [0,\infty)^A$, we write $Y(\bar{s})$ for the random vector $(Y^x(s_x))$. The Laplace transform is well known,

$$\mathbb{E}\left[\exp\{-Y(\bar{s}) \cdot f\}\right] = \prod_{x \in A} \frac{1}{[1+f(x)]^{s_x}},$$

provided that $\|f\|_\infty < 1$. In particular, if $\mathscr{C} \in \mathbb{N}_{\text{fin}}^{\mathscr{O}}$, then

$$\begin{aligned}
\mathbb{E}\left[\exp\{-Y(L_{\mathscr{C}}) \cdot f\}\right] &= \prod_{x \in A} \frac{1}{[1+f(x)]^{L_{\mathscr{C}}(x)}} \\
&= \prod_{\omega \in \mathscr{O}} \prod_{x \in A} \frac{1}{[1+f(x)]^{\mathscr{C}(\omega) N^\omega(x)}} \\
&= \prod_{\omega \in \mathscr{O}} \exp\left[-\mathscr{C}(\omega)(\ln(1+f) \cdot N^\omega)\right]. \qquad (7)
\end{aligned}$$

For positive Q, we could then define a continuous occupation field in terms of random variables, and we let $\mathscr{L}_t = Y(L_{\mathscr{C}_t})$ by taking $\mathscr{C}_t$ as an independent loop soup corresponding to $|Q|$. In order to handle the general case, we define the "distribution" of the continuous occupation field at time t to be the complex measure $\nu_t = \nu_{t,m}$ on $[0,\infty)^A$ given by

$$\nu_t(V) = \sum_{\mathscr{C} \in \mathbb{N}_{\text{fin}}^{\mathscr{O}}} \mathscr{M}_t(\mathscr{C})\, \mathbb{P}\{Y(L_{\mathscr{C}}) \in V\} = \sum_{\bar{k} \in \mathbb{N}^A} \mu_t(\bar{k})\, \mathbb{P}\{Y(\bar{k}) \in V\}, \qquad (8)$$

where $V \subset [0,\infty)^A$. We will write

$$\nu_t[h(\mathscr{L})] = \int_{[0,\infty)^A} h(\mathscr{L}) d\nu_t(\mathscr{L})$$

provided that $\int_{[0,\infty)^A} |h(\mathscr{L})|\, d|\nu_t|(\mathscr{L}) < \infty$.

Lemma 4.1 *If* $\mathbb{E}[|h(\mathscr{L}_t)|] < \infty$, *then* $|\nu_t|[|h(\mathscr{L})|] < \infty$.

Proof First, note that

$$\begin{aligned} |\mathscr{M}_t(\mathscr{C})| &= \left| \prod_{\omega \in \mathscr{C}} \mathscr{M}_t^{\omega}(\mathscr{C}(\omega)) \right| \\ &= \left| e^{-t \sum_{\omega \in \mathscr{C}} m(\omega)} \right| \prod_{\omega \in \mathscr{C}} \frac{(t|m(\omega)|)^{\mathscr{C}(\omega)}}{\mathscr{C}(\omega)!} = \alpha \mathbb{P}\{\mathscr{C}_t = \mathscr{C}\} \end{aligned}$$

with $\alpha = e^{t \sum_{\omega \in \mathscr{C}} |m(\omega)| - \Re(m(\omega))}$. Thus, taking sup over all finite partitions $\{V_i\}_{i=1}^n$ of V into measurable sets,

$$\begin{aligned} |\nu_t|(V) &= \sup \sum_{i=1}^n |\nu_t(V_i)| = \sup \sum_{i=1}^n \left| \sum_{\mathscr{C} \in \mathbb{N}_{\text{fin}}^{\mathscr{C}}} \mathscr{M}_t(\mathscr{C})\, \mathbb{P}\{Y(L_{\mathscr{C}}) \in V_i\} \right| \\ &\le \sup \sum_{i=1}^n \sum_{\mathscr{C} \in \mathbb{N}_{\text{fin}}^{\mathscr{C}}} |\mathscr{M}_t(\mathscr{C})|\, \mathbb{P}\{Y(L_{\mathscr{C}}) \in V_i\} \\ &= \sum_{\mathscr{C} \in \mathbb{N}_{\text{fin}}^{\mathscr{C}}} |\mathscr{M}_t(\mathscr{C})|\, \mathbb{P}\{Y(L_{\mathscr{C}}) \in V\} = \alpha \mathbb{P}\{\mathscr{L}_t \in V\}. \end{aligned}$$

We will compute the Laplace transform of the measure ν_t, but first we need use the following lemma.

Lemma 4.2 *Suppose* S *is a countable set and* $F : S \times \mathbb{N} \to \mathbb{C}$ *is a function with* $F(s,0) = 1$ *for all* $s \in S$,

$$\sum_{s \in S} \left| \sum_{n=1}^{\infty} F(s,n) \right| < \infty$$

and

$$\sum_{\psi \in \mathbb{N}_{\text{fin}}^S} \left| \prod_{s \in S} F(s, \psi(s)) \right| < \infty.$$

Then,

$$\prod_{s\in S}\sum_{n=0}^{\infty} F(s,n) = \sum_{\psi\in\mathbb{N}^S_{\text{fin}}}\prod_{s\in S} F(s,\psi(s)).$$

Proof Since $\sum_{s\in S}|\sum_{n=1}^{\infty} F(s,n)| < \infty$, the product on the left-hand side does not depend on the order. For this reason we may assume that S is the positive integers and write

$$\begin{aligned}\prod_{s=1}^{\infty}\sum_{n=0}^{\infty} F(s,n) &= \lim_{J\to\infty}\prod_{s=1}^{J}\sum_{n=0}^{\infty} F(s,n)\\ &= \lim_{J\to\infty}\sum_{\psi\in\mathbb{N}^J}\prod_{s=1}^{J} F(s,\psi(s)) = \sum_{\psi\in\mathbb{N}^\infty_{\text{fin}}}\prod_{s=1}^{\infty} F(s,\psi(s)).\end{aligned}$$

The last equality uses the absolute convergence of the final sum.

Proposition 4.3 *For f sufficiently small,*

$$\nu_t[\exp(-\mathscr{L}\cdot f)] = \left(\frac{\det G_f}{\det G}\right)^t. \tag{9}$$

Proof We first claim that there exists $\delta > 0$ such that if $\|f\|_\infty < \delta$,

$$\mathbb{E}[|\exp\{-\mathscr{L}_t\cdot f\}|] < \infty,$$

so that the left hand side of (9) is well defined. Indeed, if $\|f\|_\infty < \delta$, and $(1-\delta) = e^{-\epsilon}$, then for any $\mathscr{C}\in\mathbb{N}^\omega_{\text{fin}}$

$$\mathbb{E}\,[|\exp\{-Y(L_{\mathscr{C}})\cdot f\}|] \le \prod_{\omega\in\mathscr{O}}|1-\delta|^{-\mathscr{C}(\omega)\,|\omega|} = \prod_{\omega\in\mathscr{O}} e^{\epsilon\,|\omega|\,\mathscr{C}(\omega)},$$

and hence

$$\begin{aligned}\mathbb{E}[|\exp\{-Y(L_{\mathscr{C}_t})\cdot f\}|] &= \mathbb{E}\left[\mathbb{E}[|\exp\{-Y(L_{\mathscr{C}_t})\cdot f\}|\big|\mathscr{C}_t]\right]\\ &\le \mathbb{E}\left[\prod_{\omega\in\mathscr{O}} e^{\epsilon|\omega|\mathscr{C}_t(\omega)}\right]\\ &= \prod_{\omega\in\mathscr{O}}\exp\left(t|m(\omega)|(e^{\epsilon|\omega|}-1)\right)\end{aligned}$$

which is finite for ϵ sufficiently small by (4).

We assume that $\|f\|_\infty < \delta$. Using (7) we get

$$\begin{aligned}
\nu_t[\exp(-\mathscr{L}\cdot f)] &= \sum_{\mathscr{C}\in\mathbb{N}_{\rm fin}^{\mathscr{O}}} \mathscr{M}_t(\mathscr{C})\,\mathbb{E}\left[\exp(-Y(L_{\mathscr{C}})\cdot f)\right] \\
&= \sum_{\mathscr{C}\in\mathbb{N}_{\rm fin}^{\mathscr{O}}} \prod_{\omega\in\mathscr{O}} \mathscr{M}_t^{\omega}(\mathscr{C}(\omega)) \exp\left[\mathscr{C}(\omega)(-\ln(1+f)\cdot N^{\omega})\right] \\
&= \prod_{\omega\in\mathscr{O}} \sum_{n=0}^{\infty} \mathscr{M}_t^{\omega}(n) \exp\left[n(-\ln(1+f)\cdot N^{\omega})\right] \\
&= \prod_{\omega\in\mathscr{O}} \exp\left[tm(\omega)(\exp(-\ln(1+f)\cdot N^{\omega})-1)\right].
\end{aligned}$$

The third equality uses Lemma 4.2, which is valid as

$$\sum_{\omega\in\mathscr{O}} \left| \sum_{n=1}^{\infty} \frac{(tm(\omega))^n}{n!} \exp\left[n(-\ln(1+f)\cdot N^{\omega})\right] \right| = \sum_{\omega\in\mathscr{O}} \left|\exp\left[tm_f(\omega)\right]-1\right| < \infty$$

and

$$\sum_{\mathscr{C}\in\mathbb{N}_{\rm fin}^{\mathscr{O}}} \left| \prod_{\omega\in\mathscr{O}} \mathscr{M}_t^{\omega}(\mathscr{C}(\omega)) \exp\left[\mathscr{C}(\omega)(-\ln(1+f)\cdot N^{\omega})\right] \right| < \infty$$

for sufficiently small f since $|\mathscr{M}_t|$ is a finite measure on $\mathbb{N}_{\rm fin}^{\mathscr{O}}$. Above we used that

$$\begin{aligned}
\exp(-\ln(1+f)\cdot N^{\omega}) &= \prod_{x\in A} \exp\ln\left[\left(\frac{1}{1+f(x)}\right)^{N^{\omega}(x)}\right] \\
&= \prod_{j=0}^{|\omega|-1} \frac{1}{1+f(\omega_j)},
\end{aligned}$$

which also gives us

$$\begin{aligned}
\nu_t[\exp(-\mathscr{L}\cdot f)] &= \exp\left[t\sum_{\omega\in\mathscr{O}} m(\omega) \prod_{j=0}^{|\omega|-1} \frac{1}{1+f(\omega_j)}\right] \exp\left[-t\sum_{\omega\in\mathscr{O}} m(\omega)\right] \\
&= \exp\left[t\sum_{\omega\in\mathscr{O}} m_f(\omega)\right] \exp\left[-t\sum_{\omega\in\mathscr{O}} m(\omega)\right] \\
&= \left(\frac{\det G_f}{\det G}\right)^t.
\end{aligned}$$

The last equality is by Corollary 3.2.

Consider the one-point example at the end of Sect. 3.1. If we let s denote the function f taking the value s, then

$$m_s(\omega^n) = \frac{1}{n}\left[\frac{q}{1+s}\right]^n,$$

and

$$\sum_{\omega^n \in \mathscr{O}} m_s(\omega^n) = \sum_{n=1}^{\infty} \frac{1}{n}\left[\frac{q}{1+s}\right]^n = -\log\left[1 - \frac{q}{1+s}\right].$$

Therefore, if $\mathscr{L}_t$ denotes the continuous time occupation field at time t,

$$\mathbb{E}\left[e^{-s\mathscr{L}_t}\right] = \left(\frac{\det G_s}{\det G}\right)^t = \left[\frac{1+s-q(1+s)}{1+s-q}\right]^t. \tag{10}$$

We recall that we have defined the (discrete time) loop measure $m(\omega) = \frac{Q(\omega)}{|\omega|}$ and then we have added continuous holding times. Another approach, which is the original one taken by Le Jan [8], is to construct a loop measure on continuous time paths. Here we start with Q, add the waiting times to give a measure on continuous time loops, and then divide the measure by the (continuous) length. Considered as a measure on unrooted continuous time loops, the two procedures are essentially equivalent (although using discrete time loops makes it easier to have "jumps" from a site to itself).

4.4 Trivial Loops

We will see soon that the loop soup and the square of the Gaussian free field are closely related, but because our construction of the loop soup used discrete loops and only added continuous time afterwards, we restricted our attention to loops of positive length. We will need to add a correction factor to the occupation time to account for these trivial loops which are formed by viewing the continuous time process before its first jump.

Consider the one-point example at the end of Sect. 3.1. The Gaussian free field with covariance matrix $[I-Q]^{-1}$ is just a centered normal random variable Z with variance $1/(1-q)$ which we can write as $N/\sqrt{1-q}$ where N is a standard normal. Since N^2 has a χ^2 distribution with one degree of freedom, we see that

$$\mathbb{E}\left[e^{-sZ^2/2}\right] = \sqrt{\frac{1-q}{1-q+s}}$$

If we compare this to (10), we can see that

$$\mathbb{E}\left[e^{-sZ^2/2}\right] = \mathbb{E}\left[e^{-s\mathscr{L}_{\frac{1}{2}}}\right] [1+s]^{-1/2}.$$

The second term on the right-hand side is the moment generating function for a Gamma$(\frac{1}{2}, 1)$ random variable. Hence we can see that $Z^2/2$ has the same distribution at $\mathscr{L}_{\frac{1}{2}} + Y$ where Y is an independent Gamma$(\frac{1}{2}, 1)$ random variable.

The trivial loops we will add are not treated in the same way as the other loops. To be specific, we add another collection of independent gamma processes $\{Y^x_{\text{trivial}}\}_{x\in A}$ and define the *occupation field of the trivial loops* as

$$\mathscr{T}_t(x) = Y^x_{\text{trivial}}(t).$$

When viewed in terms of the discrete time loop measure, this seems unmotivated. It is useful to consider the continuous time loop measure in terms of continuous time Markov chains. For any t prior to the first jump of the Markov chain, the path will form a trivial loop of time duration t. As the Markov chain has exponential holding times, the path measure (analogue of Q) to assign to such a trivial loop is $e^{-t}\,dt$, and so the loop measure (analogue of m) should be $t^{-1}\,e^{-t}\,dt$. Hence in the continuous time measure, we give trivial loops of time duration t weight e^{-t}/t. Since e^{-t}/t is the intensity measure for the jumps of a gamma process, we see that the added occupation time at x corresponds to $\mathscr{T}_t(x)$.

We write $\nu_t^{\mathscr{T}}$ for the probability distribution of $\mathscr{T}_t$. In other words, it is the distribution of independent gamma processes $\{Y^x(t) : x \in A\}$. Note that if $\mathscr{L} \in [0,\infty)^A$,

$$\nu_t^{\mathscr{T}}[\exp\{-f\cdot\mathscr{L}\}] = \prod_{x\in A}\frac{1}{[1+f(x)]^t} = [\det D_{1+f}]^{-t}. \tag{11}$$

We will also write

$$\rho_t = \nu_t * \nu_t^{\mathscr{T}},$$

which using (8) can also be written as

$$\rho_t(V) = \sum_{\mathscr{C}\in\mathbb{N}^{\mathscr{C}}_{\text{fin}}} \mathscr{M}_t(\mathscr{C})\,\mathbb{P}\{Y(L_{\mathscr{C}} + \bar{t}) \subset V\},$$

where $\bar{t}$ denotes the vector each of whose components equals t.

4.5 Relation to the Real Gaussian Free Field

If A is a finite set with $|A| = n$, and G is a symmetric, positive definite real matrix, then the *(centered, discrete) Gaussian free field* on A with covariance matrix G is the random function $\phi : A \to \mathbb{R}$, defined by having density

$$\frac{1}{(2\pi)^{n/2}\sqrt{\det G}} \exp\left(-\frac{1}{2}\phi \cdot G^{-1}\phi\right) = \frac{1}{(2\pi)^{n/2}\sqrt{\det G}} \exp\left(-\frac{1}{2}|J\phi|^2\right)$$

with respect to Lebesgue measure on $\mathbb{R}^n$. Here J is a positive definite, symmetric square root of G^{-1}. In other words, ϕ is a $|A|$-dimensional mean zero normal random variable with covariance matrix $\mathbb{E}[\phi(x)\phi(y)] = G(x, y)$.

Lemma 4.4 *Suppose G is a symmetric positive definite matrix, $\Delta = G^{-1}$, and let ϕ denote a Gaussian free field with covariance matrix G. Then for all f sufficiently small,*

$$\mathbb{E}\left[\exp\left(-\frac{1}{2}\phi^2 \cdot f\right)\right] = \frac{1}{\sqrt{\det(\Delta + D_f)}} \frac{1}{\sqrt{\det G}}. \tag{12}$$

Proof This is a standard calculation,

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(-\frac{1}{2}\phi^2 \cdot f\right)\right] \\
&\quad = \frac{1}{(2\pi)^{n/2}\sqrt{\det G}} \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\phi^2 \cdot f\right) \exp\left(-\frac{1}{2}\phi \cdot G^{-1}\phi\right) d\phi \\
&\quad = \frac{1}{(2\pi)^{n/2}\sqrt{\det G}} \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\phi \cdot (\Delta + D_f)\phi\right) d\phi.
\end{aligned}$$

If f is sufficiently small, then $\Delta + D_f$ is a positive definite symmetric matrix and so has a positive definite square root, call it R_f. Then

$$\begin{aligned}
\int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\phi \cdot (\Delta + D_f)\phi\right) d\phi &= \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}R_f\phi \cdot R_f\phi\right) d\phi \\
&= \frac{1}{\det R_f} \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\phi \cdot \phi\right) d\phi \\
&= \frac{(2\pi)^{n/2}}{\sqrt{\det(\Delta + D_f)}}.
\end{aligned}$$

Theorem 4.5 *If Q is a symmetric, acceptable real matrix, and ϕ is the discrete Gaussian free field on A with covariance matrix G, then the distribution of $\frac{1}{2}\phi^2$ is $\rho_{\frac{1}{2}}$.*

Proof It suffices to show that the Laplace transforms for $\frac{1}{2}\phi^2$ and $\rho_{\frac{1}{2}}$ exist and agree on a neighborhood of zero. We have calculated the transforms for $\mathscr{L}_{\frac{1}{2}}$ and ϕ^2 in (9) and (11) giving

$$\begin{aligned}
\rho_{\frac{1}{2}}\left[\exp\{\mathscr{L}\cdot f\}\right] &= \nu_{\frac{1}{2}}\left[\exp\{\mathscr{L}\cdot f\}\right]\, \nu_{\frac{1}{2}}^{\mathscr{T}}\left[\exp\{\mathscr{L}\cdot f\}\right] \\
&= \det(D_{1+f}^{-1})^{\frac{1}{2}} \left(\frac{\det G_f}{\det G}\right)^{\frac{1}{2}} \\
&= \det(D_{1+f}^{-1})^{\frac{1}{2}} \left(\frac{\det(I - D_{1+f}^{-1}Q)^{-1})}{\det G}\right)^{\frac{1}{2}} \\
&= \left(\frac{\det(D_{1+f}-Q)^{-1}}{\det G}\right)^{\frac{1}{2}} \\
&= \frac{1}{\sqrt{\det(\Delta + D_f)}}\,\frac{1}{\sqrt{\det G}}.
\end{aligned}$$

Comparing this to (12) completes the proof.

Conversely, suppose that a symmetric, positive definite real matrix G is given, indexed by the elements of A and let $\{\phi(x) : x \in A\}$ denote the Gaussian free field. If the matrix $Q := I - G^{-1}$ is positive definite and acceptable, then we can use loops to give a representation of $\{\phi(x)^2 : x \in A\}$. If G has negative entries then so must Q (since the Green's function for positive weights is always positive).

4.6 Complex Weights

There is also a relation between complex, Hermitian weights and a complex Gaussian field. Let A be a finite set with n elements. Suppose G' is a positive definite Hermitian matrix and let K be a positive definite Hermitian square root of $(G')^{-1}$. The *(centered) complex Gaussian free field on A with covariance matrix* G' is defined to be the measure on complex functions $h : \mathbb{R}^A \to \mathbb{C}$ with density

$$\frac{1}{\pi^n \det G'} \exp\left(-\overline{h}\cdot (G')^{-1}h\right) = \frac{1}{\pi^n \det G'} \exp\left(-|Kh|^2\right)$$

with respect to Lebesgue measure on $\mathbb{C}^n$ (or $\mathbb{R}^{2n}$). Equivalently, the function $\psi = \sqrt{2}\,h$ has density

$$\frac{1}{(2\pi)^n \det G'} \exp\left(-\frac{1}{2}\overline{\psi}\cdot (G')^{-1}\psi\right) = \frac{1}{(2\pi)^n \det G'} \exp\left(-\frac{1}{2}\,|K\psi|^2\right), \qquad (13)$$

It satisfies the covariance relations

$$\mathbb{E}\left[\overline{h}(x)\, h(y)\right] = \frac{1}{2}\,\mathbb{E}\left[\overline{\psi(x)}\,\psi(y)\right] = G'(x,y), \tag{14}$$
$$\mathbb{E}\left[h(x)\, h(y)\right] = 0.$$

The complex Gaussian free field on a set of n elements can be considered as a real field on $2n$ elements by viewing the real and imaginary parts as separate components. The next proposition makes this precise. Let $A^* = \{x^* : x \in A\}$ be another copy of A and $\overline{A} = A \cup A^*$. We can view $\overline{A}$ as a "covering space" of A and let $\Phi : \overline{A} \to A$ be the covering map, that is, $\Phi(x) = \Phi(x^*) = x$. We call A and A^* the two "sheets" in $\overline{A}$. Let $G' = G_R + iG_I$ and define G on $\overline{A}$ by

$$\begin{aligned} G(x,y) =& G(x^*,y^*) = G_R(x,y), \\ G(x,y^*) =& -G(x^*,y) = -G_I(x,y). \end{aligned}$$

Note that G is a real, symmetric, positive definite matrix.

Proposition 4.6 *Suppose $G' = G_R + iG_I$ is a positive definite Hermitian matrix indexed by A and suppose G is the positive definite, symmetric matrix indexed by $\overline{A}$,*

$$G = \begin{matrix} \\ A \\ A^* \end{matrix} \begin{matrix} \begin{matrix} A & A^* \end{matrix} \\ \begin{pmatrix} G_R & -G_I \\ G_I & G_R \end{pmatrix} \end{matrix}.$$

Let $\{\phi_z : z \in \overline{A}\}$ be a centered Gaussian free field on $\overline{A}$ with covariance matrix G. If

$$\psi_x = \phi_x + i\,\phi_{x^*}, \tag{15}$$

then $\{\psi_x : x \in A\}$ is a complex centered Gaussian free field with covariance matrix $2G'$.

Proof Let $K = K_R + iK_I$ be the Hermitian positive definite square root of $(G')^{-1}$ and write $(G')^{-1} = \Delta_R + i\,\Delta_I$. The relation $K^2 = (G')^{-1}$ implies

$$K_R^2 - K_I^2 = \Delta_R, \quad K_R K_I + K_I K_R = \Delta_I.$$

and $G'\,(G')^{-1} = I$ implies

$$G_R\,\Delta_R - G_I\,\Delta_I = I, \quad G_R\,\Delta_I + G_I\,\Delta_R = 0.$$

Therefore,

$$G^{-1} = \begin{bmatrix} \Delta_R & -\Delta_I \\ \Delta_I & \Delta_R \end{bmatrix},$$

and $J^2 = G^{-1}$ where

$$J = \begin{bmatrix} K_R & -K_I \\ K_I & K_R \end{bmatrix}.$$

In particular, $|J\phi|^2 = |K\psi|^2$. Moreover if $\lambda > 0$ is an eigenvalue of G' with eigenvector $\mathbf{x} + i\mathbf{y}$, then

$$G_R\,\mathbf{x} - G_I\mathbf{y} = \lambda\,\mathbf{x}, \quad G_R\,\mathbf{y} + G_I\,\mathbf{x} = \lambda\,\mathbf{y},$$

from which we see that

$$G\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} = \lambda\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}, \quad G\begin{bmatrix}-\mathbf{y}\\ \mathbf{x}\end{bmatrix} = \lambda\begin{bmatrix}-\mathbf{y}\\ \mathbf{x}\end{bmatrix}.$$

Since the eigenvalues of G are the eigenvalues of G' with double the multiplicity,

$$\det G = [\det G']^2.$$

Therefore, (13) can be written as

$$\frac{1}{(\sqrt{2\pi})^{2n}\sqrt{\det G}}\,\exp\left(-\frac{1}{2}\,|J\phi|^2\right),$$

which is the density for the centered real field on $\overline{A}$ with covariance matrix G.

We will discuss the analogue to Theorem 4.5 for complex Hermitian weights. We can either use the complex weights $Q' = I - (G')^{-1} = Q_R + i\,Q_I$ on A to give a representation of $\{|\psi(x)|^2 : x \in A\}$ or we can use the weights on $\overline{A}$ given by

$$Q = \begin{bmatrix} Q_R & -Q_I \\ Q_I & Q_R \end{bmatrix} = I - G^{-1}, \tag{16}$$

to give a representation of $\{|\phi(z)|^2 : z \in \overline{A}\}$. The latter contains more information so we will do this. Note that Q is a positive definite symmetric matrix, but may not be acceptable even if Q' is.

Provided that Q' and Q are acceptable, let $\hat{m}, m$ denote the loop measures derived from them respectively. As before, let $\mathscr{O}$ denote the set of (rooted) loops of positive length in A. Let $\overline{\mathscr{O}}$ be the set of such loops in $\overline{A}$. Note that $\hat{m}$ is a complex measure on $\mathscr{O}$ and m is a real measure on $\overline{\mathscr{O}}$. We write

$$\hat{m}(\omega) = \hat{m}_R(\omega) + i\,\hat{m}_I(\omega).$$

Recall that $\Phi : \overline{A} \to A$ is the covering map. We also write $\Phi : \overline{\mathscr{O}} \to \mathscr{O}$ for the projection, that is, if $\omega' = [\omega'_0, \ldots, \omega'_k] \in \overline{\mathscr{O}}$ then $\Phi(\omega')$ is the loop of length k

whose jth component is $\Phi(\omega'_j)$. We define the pushforward measure $\Phi_* m$ on $\mathscr{O}$ by

$$\Phi_* m(\omega) = m\left[\Phi^{-1}(\omega)\right] = \sum_{\Phi(\omega')=\omega} m(\omega').$$

Proposition 4.7

$$\Phi_* m(\omega) = 2\,\hat{m}_R(\omega) = \hat{m}(\omega) + \hat{m}(\omega^R).$$

Proof Let $\mathscr{S}_k = \{R, I\}^k$ and if $\pi = (\pi^1, \ldots, \pi^k) \in \mathscr{S}_k$ we write $d(\pi)$ for the number of components that equal I. Let $\mathscr{S}_k^e$ denote the set of sequences $\pi \in \mathscr{S}_k$ with $d(\pi)$ even.

Suppose $\omega = [\omega_0, \ldots, \omega_k] \in \mathscr{O}$. There are 2^k loops $\omega' = [\omega'_0, \omega'_1, \ldots, \omega'_k] \in \overline{\mathscr{O}}$ such that $\Phi(\omega') = \omega$. We can write each such loop as an ordered triple (ω, θ, π). Here $\theta \in \{0, *\}$ and $\pi \in \mathscr{S}_k^e$. We obtain ω' from (ω, θ, π) as follows. If $\theta = 0$ then $\omega'_0 = \omega_0$, and otherwise $\omega'_0 = \omega_0^*$. For $j \geq 1$, $\omega'_j \in \{\omega_j, \omega_j^*\}$. If $\pi^j = R$, then ω'_j is chosen to be in the same sheet as ω'_{j-1}. If $\pi^j = I$, then ω'_j is chosen in the opposite sheet to ω'_{j-1}. Since $d(\pi)$ is even, we see that $\omega'_n = \omega'_0$ so this gives a loop in $\overline{\mathscr{O}}$ with $\Phi(\omega') = \omega$.

By expanding the product we see that

$$\begin{aligned} Q'(\omega) &= \prod_{j=1}^{k} \left[Q_R(\omega_{j-1}, \omega_j) + i\, Q_I(\omega_{j-1}, \omega_j)\right] \\ &= \sum_{\pi \in \mathscr{S}_k} i^{d(\pi)} \prod_{j=1}^{k} Q_{\pi^j}(\omega_{j-1}, \omega_j), \\ \operatorname{Re}\left[Q'(\omega)\right] &= \sum_{\pi \in \mathscr{S}_k^e} i^{d(\pi)} \prod_{j=1}^{k} Q_{\pi^j}(\omega_{j-1}, \omega_j), \end{aligned}$$

Note that

$$\begin{aligned} Q(\omega'_{j-1}, \omega'_j) &= -\,Q_I(\omega_{j-1}, \omega_j), \quad \omega'_{j-1} \in A, \quad \omega'_j \in A', \\ Q(\omega'_{j-1}, \omega'_j) &= Q_I(\omega_{j-1}, \omega_j), \quad \omega'_{j-1} \in A', \quad \omega'_j \in A, \\ Q(\omega'_{j-1}, \omega'_j) &= Q_R(\omega_{j-1}, \omega_j), \quad \text{otherwise}. \end{aligned}$$

If $d[\pi]$ is even, then $d[\pi]/2$ denotes the number of times that the path ω' goes from A^* to A. Using this we can write

$$\operatorname{Re}\left[Q'(\omega)\right] = \frac{1}{2}\, Q\left[\Phi^{-1}(\omega)\right] = \frac{1}{2} \sum_{\Phi(\omega')=\omega} Q(\omega').$$

The factor $1/2$ compensates for the initial choice of ω_0'. Since $Q'(\omega^R) = \overline{Q'(\omega)}$, we see that

$$Q'(\omega) + Q'(\omega^R) = Q\left[\Phi^{-1}(\omega)\right].$$

Since

$$\hat{m}(\omega) = \frac{Q'(\omega)}{|\omega|}, \quad \Phi_* m(\omega) = \frac{Q\left[\Phi^{-1}(\omega)\right]}{|\omega|},$$

we get the result.

Since

$$\det G' = \exp\left\{\sum_{\omega \subset A} \hat{m}(\omega)\right\}, \quad \det G = \exp\left\{\sum_{\omega' \subset \overline{A}} m(\omega)\right\},$$

we get another derivation of the relation $\det G = [\det G']^2$.

Given the loop measure $\hat{m}$ on A (or the loop measure m on $\overline{A}$), we can consider the discrete occupation field at time t as a measure $\mu_{t,\hat{m}}$ on $\mathbb{N}^A$ (or $\mu_{t,m}$ on $\mathbb{N}^{\overline{A}}$, respectively). The measure $\mu_{t,m}$ pushes forward to a measure $\Phi_* \mu_{t,m}$ on $\mathbb{N}^A$ by adding the components of x and x^*. It follows from (6) and Proposition 4.7 that

$$\Phi_* \mu_{t,m} = \mu_{2t,\hat{m}}.$$

Also the "trivial loop occupation field" on $\overline{A}$ at time t induces an occupation field on A by addition. This has the same distribution as the trivial loop occupation field on A at time $2t$ since there are two points in $\overline{A}$ corresponding to each point in A. Hence $\Phi_* \rho_{t,m}$ has the same distribution as $\rho_{2t,\hat{m}}$.

Using Theorem 4.5 and Proposition 4.6 we get the following.

- Suppose Q' is a positive definite acceptable Hermitian matrix indexed by A. Let $G' = (I - Q')^{-1}$.
- Let Q be the positive definite real matrix on $\overline{A}$ as in (16). Let $G = (I - Q)^{-1}$.
- Let $\{\phi(z) : z \in \overline{A}\}$ be a centered Gaussian free field on $\overline{A}$ with covariance matrix G provided Q is acceptable.
- If $h(x) = [\phi(x) + i\,\phi(x^*)]/\sqrt{2}$, then h is a complex Gaussian free field on A with covariance matrix G'.
- If ρ_t denotes the continuous occupation field on $\overline{A}$ (including trivial loops) given by Q at time t then $\{\frac{1}{2}\phi(z)^2 : z \in \overline{A}\}$ has distribution $\rho_{\frac{1}{2}}$.
- If ρ_t' denotes the continuous occupation field on A (including trivial loops) given by Q' at time t then $\{|h(z)|^2 : z \in A\}$ has distribution ρ_1'.

Acknowledgements This research was supported by National Science Foundation grant DMS-0907143.

The authors would like to thank Dapeng Zhan for bringing an error in an earlier version of this paper to our attention.

References

1. D. Brydges, J. Fröhlich, T. Spencer, The random walk representation of classical spin systems and correlations inequalities. Commun. Math. Phys. **83**, 123–190 (1982)
2. D. Brydges, J. Imbrie, G. Slade, Functional integral representations for self-avoiding walk. Probab. Surv. **6**, 34–61 (2009)
3. E.B. Dynkin, Markov processes as a tool in field theory. J. Funct. Anal. **50**, 167–187 (1983)
4. G. Lawler, The probability that loop-erased random walk uses a given edge. Electron. Commun. Probab. **19**, Article no. 51 (2014)
5. G. Lawler, V. Limic, *Random Walk: A Modern Introduction*. Cambridge Studies in Advanced Mathematics, vol. 123 (Cambridge University Press, Cambridge, 2010)
6. G. Lawler, J. Trujillo Ferreras, Random walk loop soup. Trans. Am. Math. Soc. **359**(2), 767–787 (2007)
7. G. Lawler, W. Werner, The Brownian loop soup. Probab. Theory Relat. Fields **128**(4), 565–588 (2004)
8. Y. Le Jan, *Markov Paths, Loops, and Fields. Ecole e'Eté de Probabilités de St. Flour*. Lecture Notes in Mathematics, vol. 2026 (Springer, Berlin, 2012)
9. K. Symanzik, Euclidean quantum field theory, in *Scuola Internazionale di Fisica "Enrico Fermi", XLV Coso* (Academic, New York, 1969), pp. 152–223
10. D. Wilson, Generating random spanning trees more quickly than the cover time, in *Proceedings of STOC96* (1996), pp. 296–303

Index

M. Biskup et al. (eds.), *Random Walks, Random Fields, and Disordered Systems*,
Lecture Notes in Mathematics 2144, DOI 10.1007/978-3-319-19339-7

LECTURE NOTES IN MATHEMATICS

Springer

Edited by J.-M. Morel, B. Teissier; P.K. Maini

Editorial Policy (for Multi-Author Publications: Summer Schools / Intensive Courses)

1. Lecture Notes aim to report new developments in all areas of mathematics and their applications - quickly, informally and at a high level. Mathematical texts analysing new developments in modelling and numerical simulation are welcome. Manuscripts should be reasonably selfcontained and rounded off. Thus they may, and often will, present not only results of the author but also related work by other people. They should provide sufficient motivation, examples and applications. There should also be an introduction making the text comprehensible to a wider audience. This clearly distinguishes Lecture Notes from journal articles or technical reports which normally are very concise. Articles intended for a journal but too long to be accepted by most journals, usually do not have this "lecture notes" character.

2. In general SUMMER SCHOOLS and other similar INTENSIVE COURSES are held to present mathematical topics that are close to the frontiers of recent research to an audience at the beginning or intermediate graduate level, who may want to continue with this area of work, for a thesis or later. This makes demands on the didactic aspects of the presentation. Because the subjects of such schools are advanced, there often exists no textbook, and so ideally, the publication resulting from such a school could be a first approximation to such a textbook. Usually several authors are involved in the writing, so it is not always simple to obtain a unified approach to the presentation.
   For prospective publication in LNM, the resulting manuscript should not be just a collection of course notes, each of which has been developed by an individual author with little or no coordination with the others, and with little or no common concept. The subject matter should dictate the structure of the book, and the authorship of each part or chapter should take secondary importance. Of course the choice of authors is crucial to the quality of the material at the school and in the book, and the intention here is not to belittle their impact, but simply to say that the book should be planned to be written by these authors jointly, and not just assembled as a result of what these authors happen to submit.
   This represents considerable preparatory work (as it is imperative to ensure that the authors know these criteria before they invest work on a manuscript), and also considerable editing work afterwards, to get the book into final shape. Still it is the form that holds the most promise of a successful book that will be used by its intended audience, rather than yet another volume of proceedings for the library shelf.

3. Manuscripts should be submitted either online at www.editorialmanager.com/lnm/ to Springer's mathematics editorial, or to one of the series editors. Volume editors are expected to arrange for the refereeing, to the usual scientific standards, of the individual contributions. If the resulting reports can be forwarded to us (series editors or Springer) this is very helpful. If no reports are forwarded or if other questions remain unclear in respect of homogeneity etc, the series editors may wish to consult external referees for an overall evaluation of the volume. A final decision to publish can be made only on the basis of the complete manuscript; however a preliminary decision can be based on a pre-final or incomplete manuscript. The strict minimum amount of material that will be considered should include a detailed outline describing the planned contents of each chapter.
   Volume editors and authors should be aware that incomplete or insufficiently close to final manuscripts almost always result in longer evaluation times. They should also be aware that parallel submission of their manuscript to another publisher while under consideration for LNM will in general lead to immediate rejection.

4. Manuscripts should in general be submitted in English. Final manuscripts should contain at least 100 pages of mathematical text and should always include

   – a general table of contents;
   – an informative introduction, with adequate motivation and perhaps some historical remarks: it should be accessible to a reader not intimately familiar with the topic treated;
   – a global subject index: as a rule this is genuinely helpful for the reader.

   Lecture Notes volumes are, as a rule, printed digitally from the authors' files. We strongly recommend that all contributions in a volume be written in the same LaTeX version, preferably LaTeX2e. To ensure best results, authors are asked to use the LaTeX2e style files available from Springer's web-server at
   ftp://ftp.springer.de/pub/tex/latex/svmonot1/ (for monographs) and
   ftp://ftp.springer.de/pub/tex/latex/svmultt1/ (for summer schools/tutorials).
   Additional technical instructions, if necessary, are available on request from:
   lnm@springer.com.

5. Careful preparation of the manuscripts will help keep production time short besides ensuring satisfactory appearance of the finished book in print and online. After acceptance of the manuscript authors will be asked to prepare the final LaTeX source files and also the corresponding dvi-, pdf- or zipped ps-file. The LaTeX source files are essential for producing the full-text online version of the book. For the existing online volumes of LNM see:
   http://www.springerlink.com/openurl.asp?genre=journal&issn=0075-8434.
   The actual production of a Lecture Notes volume takes approximately 12 weeks.

6. Volume editors receive a total of 50 free copies of their volume to be shared with the authors, but no royalties. They and the authors are entitled to a discount of 33.3 % on the price of Springer books purchased for their personal use, if ordering directly from Springer.

7. Commitment to publish is made by letter of intent rather than by signing a formal contract. Springer-Verlag secures the copyright for each volume. Authors are free to reuse material contained in their LNM volumes in later publications: a brief written (or e-mail) request for formal permission is sufficient.

Addresses:

Professor J.-M. Morel, CMLA,
École Normale Supérieure de Cachan,
61 Avenue du Président Wilson, 94235 Cachan Cedex, France
E-mail: morel@cmla.ens-cachan.fr

Professor B. Teissier, Institut Mathématique de Jussieu,
UMR 7586 du CNRS, Équipe "Géométrie et Dynamique",
175 rue du Chevaleret,
75013 Paris, France
E-mail: teissier@math.jussieu.fr

For the "Mathematical Biosciences Subseries" of LNM:

Professor P. K. Maini, Center for Mathematical Biology,
Mathematical Institute, 24-29 St Giles,
Oxford OX1 3LP, UK
E-mail: maini@maths.ox.ac.uk

Springer, Mathematics Editorial I,
Tiergartenstr. 17,
69121 Heidelberg, Germany,
Tel.: +49 (6221) 4876-8259
Fax: +49 (6221) 4876-8259
E-mail: lnm@springer.com