Extremes of hierarchical fields

DISSERTATION
ZUR ERLANGUNG DES DOKTORGRADES
DER NATURWISSENSCHAFTEN

VORGELEGT BEIM FACHBEREICH INFORMATIK UND MATHEMATIK
DER JOHANN WOLFGANG GOETHE -UNIVERSITÄT
IN FRANKFURT AM MAIN

von
Marius Alexander Schmidt
aus Rüsselsheim

Frankfurt 2018
(D30)

vom Fachbereich Informatik und Mathematik der

Johann Wolfgang Goethe - Universität als Dissertation angenommen.

Dekan:

Prof. Dr. Andreas Bernig

Gutachter :

Prof. Dr. Nicola Kistler

Prof. Dr. David Belius

Prof. Dr. Louis-Pierre Arguin

Datum der Disputation :

Contents

1 Introduction

Preface

The goal of this thesis is to give insight into the study of hierarchical fields and their application today. It is also a nice opportunity to introduce the reader to the credo "If there is a tree, there is a way.", meaning once the hierarchical structure of a model is nailed down we can unravel its inner workings, which was (maybe half jokingly) introduced to me by my Ph.D. adviser Nicola Kistler, but held true ever since. To this end we begin our journey in Section 1.1 with a short introduction to hierarchical fields and some well known results for Derrida's random energy model and generalized random energy model, which are some of the simplest yet paradigmatic hierarchical fields. We proceed to explain in Section 1.2 the connection of so called scales in hierarchical models with the behavior of their maximum culminating in a summary of the results and ideas of Kistler and Schmidt [42], in which a class of models is discussed for which the second order of the maximum is directly related to the number of scales, clarifying the meaning of the constant in aforementioned second order correction. In Section 1.3 thereafter we outline the state of the art for branching Brownian motion type models that are studied extensively to this day not only for their theoretical appeal, but also for their far reaching prototypical role and connections to other fields of research. We also explain the contributions made to the study of these models by Glenz, Kistler and Schmidt [36] and give some intuition for the result. We conclude the introduction by Section 1.4 explaining the strong connection between cover times in two dimensions and hierarchical fields; again giving the reader some intuition for the results and stating the contributions of Schmidt [52] to the research efforts. Overall the aim of this introduction is to give the reader who may not be an expert for hierarchical fields a good view of "the big picture" without forcing the fine details onto him or her. For this reason, in order to keep the introduction and summary brief, concise and readable as well as to familiarize the reader with some intuition that directs the arguments of attached papers we deliberately forgo being absolutely rigorous in these sections. For more details on the discussed topics as well as full proofs of all mentioned results we refer the reader to given references. After the introduction aforementioned papers [36, 42, 52] are attached. We then proceed to give a quick summary of the insights obtained in English and German language. As the introduction explains also the interlinking of discussed topics it is encouraged to read it start to finish, whereas the attached papers are written to be read independently.

Acknowledgments. *Before going into more detail, the author wishes to express his gratitude to the many people who in one way or another helped him on his way, some of whom are Julius Achenbach, Magdalena Anthes, Alexander Becker, Daniel Fröhlich, Rita Gerlach, Nicola Kistler, Luis Kuhn, Anna Meiser, Ralph Neininger, Sebastian Oliva, Evelyn Praefke, Claudia Schmidt, Frank Schmidt, Karlheinz Schmidt, Ursula Schmidt, Doris Stürmer, Anton Wakolbinger, Tobias Weth and Stephan Wezorke.*

1.1 From independence to hierarchical models

Since the inception of probability theory the concept of independence has remained most central to this day. There are innumerable results on independent random variables: Law of large numbers, central limit theorem, Cramér's theorem, Fisher-Tippett-Gnedenko theorem, Poisson limit theorem or local limit theorems are only a few prominent representatives of this huge class of results. Slightly more general there are just as many results on approximately independent behavior. De Finetti's theorem, mixing, Martingale difference sequences, Galton Watson processes, subcritical Erdös Rényi Graphs or Chen-Stein method are some exponents of this class. Asking the most natural question "What about dependent behavior?" one is quickly convinced that this is too vague a question to answer. This leaves no choice, but to look for large classes of models that are still confined enough to say something meaningful. One such class is the class of hierarchically dependent fields which, as in the independent case, allows one to consider approximately hierarchical behavior as well. We consider any field constructed as follows to be hierarchical:

1. Pick a possibly random rooted tree.
2. Given the tree associate independent random variables to the edges.
3. Consider the field indexed by the leafs that is obtained by associating each leaf to the sum of the random variables from the root to the leaf.

Typically one considers an in some sense consistent sequence of hierarchical fields with growing number of leafs and is interested in e.g. the maximum, minimum, extremal process or how many leafs are associated to random variables near a given value. Note that elements of the index set of a field are called leafs for now, but for some fields may be called spins or particles depending on the context in order to stay in line with the literature. The concept of so called scales will play an important role. Given a tree scale refers simply to the distance to the root, hence the behavior on small scales is the behavior near the root and the number of scales is the number of levels on which the tree is branching. Even for not exactly hierarchical models we introduce scales to indicate where the suggested branching structure is to be found. The simplest hierarchical field

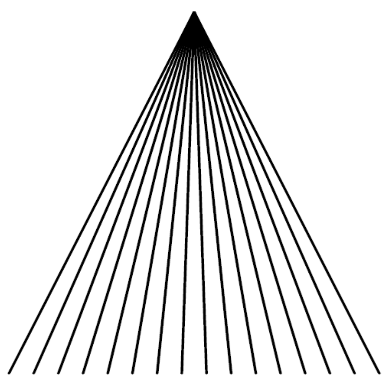

Figure 1: Random energy model

and first model we discuss is the random energy model (short REM) introduced by Derrida [33]. It corresponds to the tree that is only the root with 2^N leafs and centered Gaussian random variables of variance N on all edges (see Fig. 1). Clearly the field at hand simply consists of 2^N independent Gaussians. Hence considering the maximum or extremal process is a classical problem:

Theorem 1. *Subtracting*

$$a_N^{\mathrm{REM}} \equiv \sqrt{2\ln 2}N - \frac{1}{2\sqrt{2\ln 2}}\ln N \tag{1}$$

from the REM the maximum of the field converges to a Gumbel distribution and the extremal process converges to a Poisson point process with intensity

$$\kappa e^{-\sqrt{2\ln 2}x}dx, \tag{2}$$

where $\kappa > 0$ is a numerical constant, which is explicitly known.

Behavior similar to this is not special to the Gaussian distribution, but known for any distribution such that there exists a renormalization that emits a sensible limit. Details are given by the Fisher-Tippett-Gnedenko theorem and related results. The next model we consider is the generalized random energy model (short GREM) introduced by Derrida [32]. For $K \in \mathbb{N}$ and $a_1, a_2, ..., a_K \geq 0$ it is described by the tree of depth K where each non leaf vertex has $2^{N/K}$ children. The edges connecting a depth i and a depth $i+1$ vertex are equipped with centered Gaussians of variance $a_{i+1}N$. To compare only different correlation structures one typically considers the case $\sum_{i=1}^K a_i = 1$ fixing the variance to N. One could consider different amounts of children on each level, which is usually not done as it has about the same effect as a variance change, which we allow to be chosen freely.

We will focus on $K = 2$ (see Fig. 2), since the most relevant phenomena are already present in this case and the exposition can be handled without heavy notation. For an in depth treatment of the GREM see Bovier and Kurkova [24] or see Gayard and Kistler [41] for an intuitive introduction to the model. The $K = 2$ GREM has three regimes: $a_1 < a_2$, $a_1 = a_2$ and $a_1 > a_2$. If $a_1 < a_2$ we are in the so called REM phase this name becomes clear in view of

Theorem 2. *For $a_1 < a_2$ subtracting a_N^{REM} from the GREM gives convergence of the extremal process to a Poisson point process of intensity $\kappa e^{-\sqrt{2\ln 2}x}dx$ for some known constant $\kappa > 0$.*

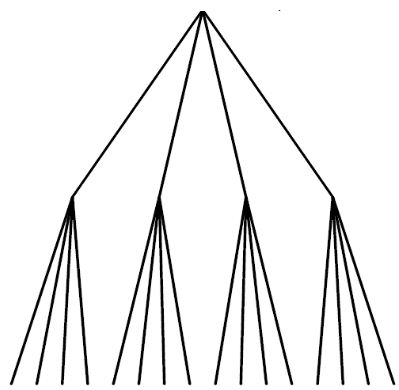

Figure 2: Generalized random energy model: $K = 2$, $N = 4$

For the critical regime $a_1 = a_2$ the same centering and convergence are correct, but the constant κ is different. Although this result is very similar the critical case is the hardest to handle and arguably has the most interesting underlying behavior. For $a_1 > a_2$ we have the centering

$$a_N^{\mathrm{GREM}} \equiv (\sqrt{a_1} + \sqrt{a_2})\left(\sqrt{\ln 2}N - \frac{1}{2\sqrt{\ln 2}}\ln N\right). \tag{3}$$

Let $(\zeta_i^{(1)})_{i\in\mathbb{N}}$ be a Poisson point process of intensity $\kappa^{(1)}e^{-x\sqrt{ln2/a_1}}dx$ and consider independent Poisson point processes $(\zeta_{i,j}^{(2)})_{j\in\mathbb{N}}$ with intensity $\kappa^{(2)}e^{-x\sqrt{ln2/a_2}}dx$ for $i \in \mathbb{N}$ all independent of $\zeta^{(1)}$.

Theorem 3. *The extremal process of the GREM with $a_1 > a_2$ after subtracting a_N^{GREM} from the field converges to the process defined by*

$$\sum_{i,j=1}^{\infty} \delta_{\zeta_i^{(1)}+\zeta_{i,j}^{(2)}} \tag{4}$$

for some known constants $\kappa^{(1)}, \kappa^{(2)} > 0$.

Generalizing this limiting process to arbitrary amount of scales (from 2 scales here) gives the class of Derrida-Ruelle Cascades, which play an important role even beyond GREM like models. For details on Derrida-Ruelle Cascades see e.g. Ruelle [51]. Prominent examples in the field of spin glasses incorporating these processes are the Parisi theory [47] or Guerra's interpolation technique [38]. Before we continue to the discussion of result and intuition of attatched papers we mention the main techniques used. At the center of all three attached papers stands the hierarchical structure of the model, which is to my knowledge best exploited employing a multiscale refinement of the second moment method. This flexible and powerful method is employed in all three papers. For a comprehensive introduction to the method see Kistler [41, pages 71-120]. To handle convergence of extremal processes Kallenberg [39] and [40] have proven very useful. Either the Laplace transform of the extremal process can be controlled directly or alternatively one can employ Chen-Stein methods (see e.g. Barbour, Holst and Janson [11]) to control the avoidance function as is done in Kistler and Schmidt [42].

1.2 From generalized random energy model to branching random walk

While models with a fixed number of scales are very well understood, models with growing number of scales have some unanswered questions still. Focusing on what is known first, we introduce the critical branching random walk which is the straight forward generalization of the critical GREM to $K = N$ scales. This is the model constructed by using the complete binary tree of depth N and attaching a standard Gaussian to each edge. For this model we consider the following result, which by linear rescaling is a direct consequence of Aïdékon [1, Theorem 1.1]:

Theorem 4. *Subtracting the normalization*

$$a_N^{\mathrm{BRW}} \equiv \sqrt{2\ln 2}N - \frac{3}{2\sqrt{2\ln 2}}\ln N \tag{5}$$

from the critical branching random walk yields convergence to a randomly shifted Gumbel distribution.

Note that results of this type often are universal: e.g. Aïdékon [1, Theorem 1.1] implies that as long as branching and edge random variables are constructed in the same way for the entire tree (resulting in a self similar model), then we are still in the critical regime and up different normalizing constants we obtain the same result (given reasonable tail behavior). One immediately notices the difference to the critical GREM which has only 1/3 the log-correction. This is in fact not only true for $K = 2$ but for any fixed number of scales. Explaining the emergence of the extra factor 3 in the log-correction and constructing models with any log-correction in between is the topic of Kistler and Schmidt [42] the first paper of this thesis, which closed the before unexplained gap between the 1 of the REM and the 3, which was first seen in Bramson [27] in the case of branching Brownian motion (which we discuss in the next section).

We now discuss the main result of Kistler and Schmidt [42]: Consider the balanced tree with 2^N leafs and N^α scales for some $0 < \alpha < 1$. This entails that any non leaf vertex has $2^{(N^{1-\alpha})}$ child vertices (see Fig.3). Associating each edge to a Gaussian random variable of variance $N^{1-\alpha}$ finishes the description of the model. Note that this model as well as the REM, GREM and BRW we consider are all normalized to have 2^N leafs that are all associated to Gaussians of variance N, allowing for a comparison of dependencies only. This interpolating model gives a first idea of the root cause of the change in log-correction in view of the main result of Kistler and Schmidt [42]: We identify the space of vertices by strings of length up to N^α with values in $\{1,...,2^{(N^{1-\alpha})}\}$, hence the set of leafs is $\Sigma_N \equiv \{1,...,2^{(N^{1-\alpha})}\}^{(N^\alpha)}$. We refer to the random variable associated to

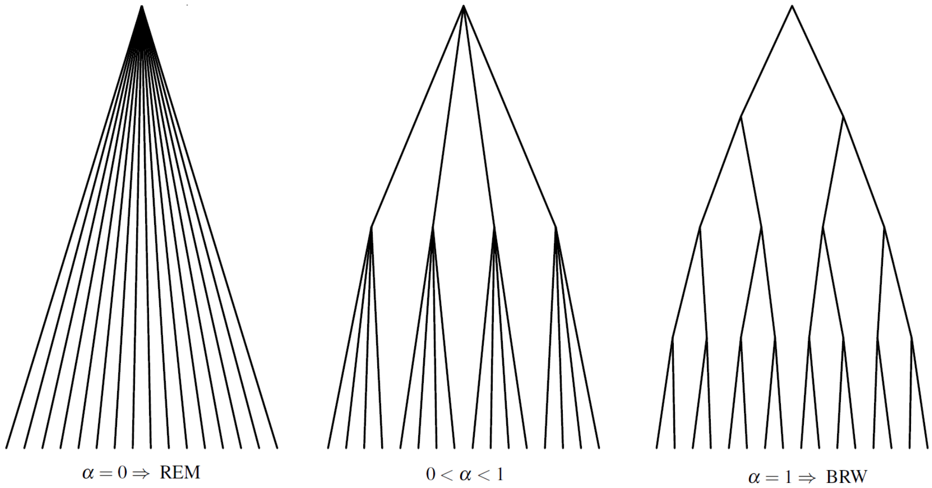

Figure 3: Trees interpolating between REM and BRW

the edge from $(\sigma_1, ..., \sigma_{i-1})$ to $(\sigma_1, ..., \sigma_i)$ by $X_{\sigma_1,...,\sigma_i}$. Finally we define the field associated to the leafs $\sigma \in \Sigma_N$ by

$$X_\sigma = \sum_{i=1}^{N^\alpha} X_{\sigma_1,..,\sigma_i}. \tag{6}$$

Theorem 5. *Setting*

$$a_N^{(\alpha)} \equiv \sqrt{2\ln 2}N - \frac{1+2\alpha}{2\sqrt{2\ln 2}} \ln N \tag{7}$$

we have for Ξ *a Poisson point process with density* $\kappa e^{-\sqrt{2\ln 2}x} dx$

$$\sum_{\sigma \in \Sigma_N} \delta_{X_\sigma - a_N^{(\alpha)}} \to \Xi \tag{8}$$

for a known $\kappa > 0$ *weakly in the large N limit.*

While this gives a precise statement how the number of scales or equivalently strength of correlation influences the log-correction, we need to take a closer look to truly find the root cause of this behavior. Note that we have for any compact set A of positive Lebesgue measure

$$\mathbb{E}\left[\sum_{\sigma \in \Sigma_N} \delta_{X_\sigma - a_N^{(\alpha)}}(A)\right] = 2^N \int_A \exp\left[-\frac{\left(a_N^{(\alpha)} + x\right)^2}{2N}\right] \frac{dx}{\sqrt{2\pi N}} \sim \kappa_A N^\alpha, \tag{9}$$

where $\kappa_A > 0$ is a constant dependent on A only. This would on first inspection suggest that the recentering $a_N^{(\alpha)}$ is too small, this is however misleading as the main contribution of this expectation is carried by paths that are so rare that they have no contribution to the limiting process. We understand a path as the sequence of partial sums from the root to the leaf picking up the random variable from the edge we traverse in each step, i.e

$$S^\sigma = (S_k^\sigma, k \le N^\alpha), \quad S_k^\sigma \equiv \sum_{j \le k} X_{\sigma_1,..\sigma_j}. \tag{10}$$

Defining

$$U_N(k) \equiv \sqrt{2\ln 2} k N^{1-\alpha} + \ln(N), \; k = 1,..,N^\alpha. \tag{11}$$

we check by Markov inequality that there are most likely no paths above U as

$$\mathbb{P}\left(\exists k \leq N^{\alpha}, \sigma \in \Sigma_N : S_k^{\sigma} > U_N(k)\right) \leq \sum_{k \leq N^{\alpha}} 2^{kN^{1-\alpha}} \mathbb{P}\left(S_k^{\sigma} > U_N(k)\right), \tag{12}$$

is vanishing in view of the following Gaussian tail estimate and as $\sqrt{2\ln 2} > 1 > \alpha$

$$\mathbb{P}\left(S_k^{\sigma} > U_N(k)\right) \leq \frac{\sqrt{kN^{1-\alpha}}}{U_N(k)} \exp\left(-\frac{U_N(k)^2}{2kN^{1-\alpha}}\right) \leq \frac{1}{N^{\sqrt{2\ln 2}}} 2^{-kN^{1-\alpha}}. \tag{13}$$

As there are most likely no path above U_N restricting the consideration to leafs with path below U_N does not change the limiting process. Considering the restricted extremal process yields

$$\mathbb{E}\left[\sum_{\sigma \in \Sigma_N} \mathbb{1}_{\{S^{\sigma} \leq U_N\}} \delta_{X_{\sigma} - a_N^{(\alpha)}}(A)\right] = \\ 2^N \int_A \mathbb{P}\left(S^{\sigma} \leq U_N | X_{\sigma} = a_N^{(\alpha)} + x\right) \exp\left[-\frac{\left(a_N^{(\alpha)} + x\right)^2}{2N}\right] \frac{dx}{\sqrt{2\pi N}}. \tag{14}$$

Up to small error $\mathbb{P}\left(S^{\sigma} \leq U_N | X_{\sigma} = a_N^{(\alpha)} + x\right)$ is the probability that a discrete Brownian bridge from 0 to $\sqrt{2\ln 2}N$ exceeds its expectation no more than logarithmically. This is equally likely to a discrete Brownian bridge from 0 to 0 staying below a logarithmic barrier. Renormalizing to standard Brownian bridges this comes essentially down to a discrete Brownian bridge of length $K = N^{\alpha}$ being non positive. The probability of this happening is well known by the ballot theorem and is exactly $K^{-1} = N^{-\alpha}$. This gives precisely the contribution necessary to push (9) down to (13) which we now see is of order one, as we expect from the correct centering. With this it is clear that in the critical GREM this term only contributes a factor of order one as K is fixed in that case.

Expanding on this intuition we aim to explain the reasons for the Poissonian nature of the limiting process for $\alpha < 1$ next. Discrete Brownian bridge path have fluctuations of order of the standard deviation. Hence asking such a bridge to stay non positive, forces it to be negative and roughly of size of the standard deviation. This phenomenon is known as entropic repulsion and an integral observation of Bramson [26] needed for the treatment of extreme values of hierarchical fields. With this observation we are now in a position to make the following simple statement which has far reaching consequences: The best leafs are not the children of vertices with the highest paths. Of course this entails that away from the starting point many paths are potential parents for future maxima, which in turn makes it unlikely to branch at some time into two particles that much later have both some near maximal child. This results in the fact that two near maximal leafs have paths that are either disjoint up to a common part of order one from the early evolution where there are only few vertices in existence or nearly identical paths up to branching of order one away from the leaf. This is unavoidable as relative only different by a random variable of order one from the maximum is near the maximum. The common part near the root typically gives the limiting process a random shift encoding the success of the early evolution and the branching near time N gives a clustering phenomenon. For $\alpha < 1$ however the first and last step are larger than order one, hence neither random shift nor clustering occurs making the limiting process Poissonian.

1.3 On branching Brownian motion

Standard branching Brownian motion is a model very similar to the branching random walk we discussed in the last section. It can be constructed by running a Yule process with rate one up to some time t. Then equipping each edge with a centered Gaussian of variance equal to the length of the edge in the Yule tree. This is equivalent to starting a particle at zero performing a Brownian motion for an exponentially distributed time, then splitting into two particles that progress like independent copies of the first particle from the splitting point onwards (see Fig. 4). Branching

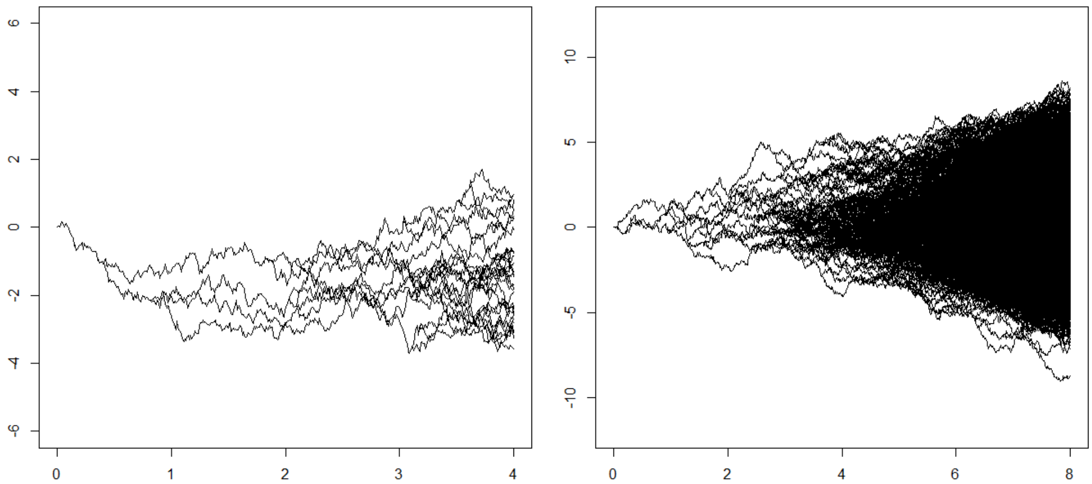

Figure 4: Two realizations of branching brownian motion

Brownian motion (short BBM) today is a very well analyzed model. The interest in the model, especially in its maximum, was amplified by the connection to the Kolmogorov-Petrovskii-Piskunov (or Fisher-Kolmogorov-Petrovskii-Piskunov) equation first observed by McKean [46]. The model is also relevant in the theory of disordered systems see e.g. Bovier and Kurkova [25] or Derrida and Spohn [34]. Bramson [27] showed two years after McKeans's observation that the maximum is up to error of order one given by $\sqrt{2}t - \frac{3}{2\sqrt{2}} \ln t$. The missing $\ln 2$ compared to the BRW simply stems from the fact that at time t BBM has about e^t particles whereas a BRW at time N has 2^N particles. This turns the BRW $\ln 2$ term to a $\ln e = 1$. Hence we already notice that changing from deterministic discrete binary branching to continuous random branching with rate one only changes the model slightly seen here by the fact that the maxima differ only by order one. The question of the distribution of the maximum was resolved by Lalley and Sellke [43] and the limiting extremal process was found independently by Aïdékon, Berestycki, Brunet and Shi [2] as well as Arguin, Bovier and Kistler [8]. Recently even finer results were established see Bovier and Hartung [21] or Cortines, Hartung and Louidor [30] for details. Also variants of the standard branching Brownian motion have been studied and are still far from fully solved one of which being variable speed branching Brownian motion allowing the variance of the Brownian motions used in the construction to depend on time, see e.g. Bovier and Hartung [23]. One model in the class of variable speed branching Brownian motions is two-speed branching Brownian motion introduced by Derrida and Spohn [34] investigated in some detail by Fang and Zeitouni [35] and the extremal process was established in Bovier and Hartung [22]. A simulation of the model is given in Fig. 5, fixing the variance at time 8 to that of standard branching Brownian motion and comparing a branching Brownian motion which fluctuates faster up to time 4 and slower thereafter (left) with one that fluctuates slower up to time 4 and faster thereafter (right). The weak correlation regime and its extremal processes of two-speed branching Brownian motion and also variable speed branching Brownian motion are intimately intertwined with the number of so called high points of standard branching Brownian motion, which is analyzed in the second paper of this theses: Glenz, Kistler and Schmidt [36]. For $\{x_k(t), k \leq n(t)\}$ the points of a branching Brownian motion we consider point $k \leq n(t)$ to be a high point of parameter $\alpha \in (0, \sqrt{2})$ if $x_k(t) \geq \left(\sqrt{2} - \alpha\right) t$. Hence the

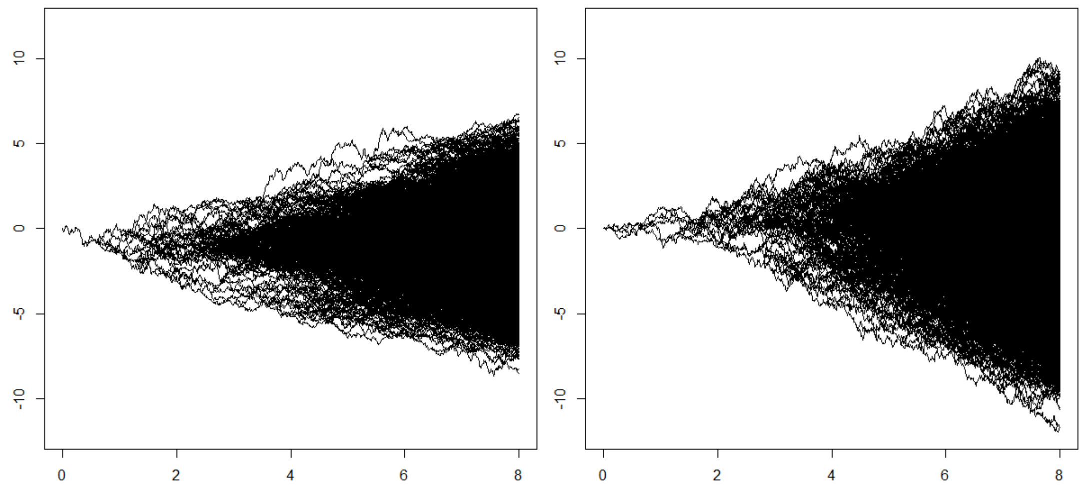

Figure 5: Two-speed branching Brownian motion, strong correlation (left), weak correlation (right)

number of α-high points is given by

$$Z_\alpha(t) \equiv \#\left\{k \le n(t) : x_k(t) \ge \left(\sqrt{2}-\alpha\right)t\right\}. \tag{15}$$

As there is a growing amount of particles in each region that is traversed by a typical α-high point except for the beginning, it is a natural guess that given the early evolution $Z_\alpha(t)$ should be practically known. Hence consider the conditional expectation of $Z_\alpha(t)$ conditioned on everything that happens up to some time $r \in (0,t)$ as a good approximation of $Z_\alpha(t)$ for r large enough. To compute said expectation some notation is needed. To this end let $\Delta_\alpha \equiv \sqrt{2}-\alpha$, $n(r)$ the number of particles at time r and let $n_i(t-r)$ the number of children particle $i \le n(r)$ at time r has at time t. By grouping particles at time t in groups of common ancestor at time r we identify

$$\{x_k(t), k \le n(t)\} = \{x_i(r) + x_{i,j}(t-r), i \le n(r), j \le n_i(t-r)\}. \tag{16}$$

Aforementioned conditional expectation is now computed to leading order by

$$\begin{aligned}
\mathbb{E}\left[Z_\alpha(t) \mid \mathscr{F}_r\right] &= \mathbb{E}\left[\sum_{k \le n(t)} 1\left\{x_k(t) \ge \Delta_\alpha t\right\} \middle| \mathscr{F}_r\right] = \\
&= \mathbb{E}\left[\sum_{i \le n(r)} \sum_{j \le n_i(t-r)} 1\left\{x_{i,j}(t-r) \ge \Delta_\alpha(t-r) - (x_i(r) - \Delta_\alpha r)\right\} \middle| \mathscr{F}_r\right] \\
&= \sum_{i \le n(r)} e^{t-r} \mathbb{P}\left[x_1(t-r) \ge \Delta_\alpha(t-r) - (x_i(r) - \Delta_\alpha r) \middle| \mathscr{F}_r\right] \\
&\sim \left(\Delta_\alpha \sqrt{2\pi}\right)^{-1} \exp\left[\left(1 - \Delta_\alpha^2/2\right)t - \frac{1}{2}\log(t)\right] Y_\alpha(r), \quad \text{a.s.},
\end{aligned} \tag{17}$$

where

$$Y_\alpha(r) \equiv \sum_{k \le n(r)} \exp\left[-r\left(1 + \frac{1}{2}\Delta_\alpha^2\right) + \Delta_\alpha x_k(r)\right]. \tag{18}$$

The last step by a standard tail estimate for Gaussian random variables and using that r is much smaller than t. As is to be expected by the derivation as a conditional expectation of non negative random variables $Y_\alpha(r)$ is a non negative martingale, which turns out to be square integrable for $\alpha \in (0,\sqrt{2})$ and therefore has a nontrivial limit. $Y_\alpha(r)$ is known as McKean's martingale who first discovered it in the context of branching Brownian motion. More details are available in Bovier and Hartung [22]. Realizing that

$$\mathbb{E}\left[Z_\alpha(t)\right] \sim \left(\Delta_\alpha\sqrt{2\pi}\right)^{-1} \exp\left[\left(1-\Delta_\alpha^2/2\right)t - \frac{1}{2}\log(t)\right] \tag{19}$$

as $Y_\alpha(0) = 1$ the main result of Glenz, Kistler and Schmidt [36] should come as no surprise to the reader:

Theorem 6. *(Strong law of large numbers for high points of BBM) For any* $0 < \alpha < \sqrt{2}$*,*

$$\lim_{t\to\infty} \frac{Z_\alpha(t)}{\mathbb{E}Z_\alpha(t)} = \lim_{r\to\infty} Y_\alpha(r), \quad \textit{almost surely.} \tag{20}$$

1.4 On cover times

To any finite Graph we can associate a random walk by considering the Markov chain that goes in one step to one neighbor of the momentary position, each being equally likely. Now the cover time of the Graph is given by the first time each vertex has been hit at least once. For some results on these discrete time cover times see e.g. Aldous [3]. The continuous analog is considering Brownian motion on a compact and smooth Riemannian manifold without boundary (or reflecting the Brownian motion on the boundary) and defining the ε-cover time as the first time all ε balls with centers on the manifold are hit. The cover time is also given by the first time the ε Wiener sausage of the Brownian motion covers the entire manifold. For some results concerning continuous cover times see e.g. Matthews [45], who establishes the $\varepsilon \to 0$ asymptotic of cover times on spheres of dimension at least 3. The two-dimensional case, discrete or continuous, regardless of the choice of manifold remained open for quite some time. Aldous [4] conjectured the upper

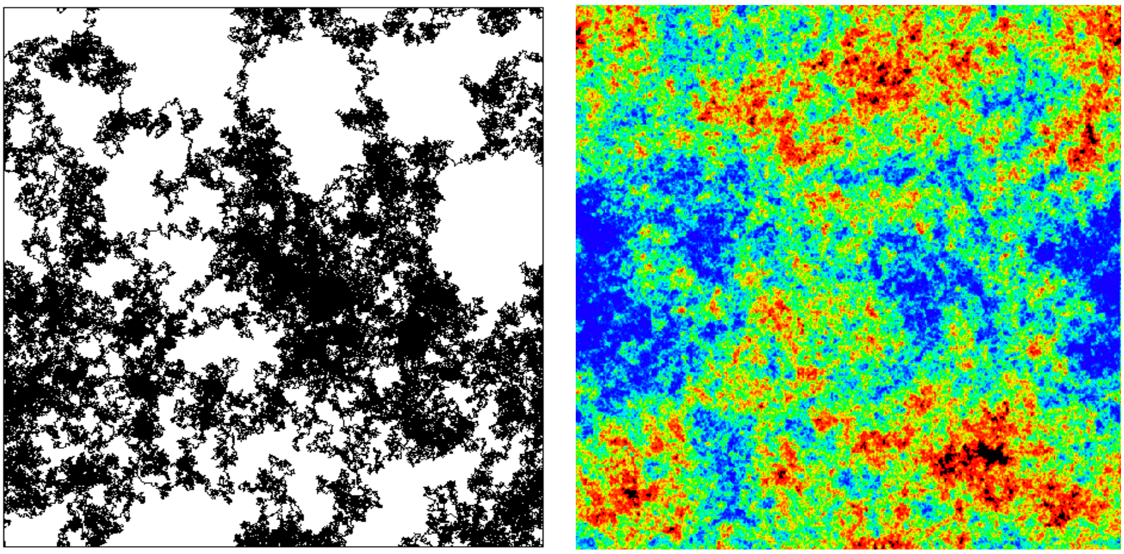

Figure 6: Brownian motion on the torus and its occupation times

bound $\frac{4}{\pi}(n\ln n)^2$ to be sharp for the n by n discrete torus. Zuckerman [53] provided a first lower bound of the correct order, which was sharpened by Lawler [44] and the conjecture was finally proven by Dembo, Peres, Rosen and Zeitouni (short DPRZ) [31]. They solved the discrete problem by proving the asymptotics for the cover time of the continuous two-dimensional unit torus first and then deducing the result for the n by n torus by a coupling argument. They also argue that the method of proof extends to arbitrary smooth, compact manifolds without boundary. This makes the following result of DPRZ [31] the center piece of leading order considerations for cover times in two dimensions:

Theorem 7. *For T_ε the cover time of the two dimensional unit torus we have*

$$\frac{T_\varepsilon}{(\ln \varepsilon)^2} \to \frac{2}{\pi} \qquad \textit{almost surely as } \varepsilon \to 0. \tag{21}$$

Today even more details are known as Belius and Kistler [13] established the next correction term and very recently Belius, Rosen and Zeitouni [14] showed tightness for the recentered cover time of the unit Sphere and from there generalized to arbitrary smooth, compact, connected, two-dimensional Riemannian manifolds without boundary. As the field of cover times and the techniques of analyzing approximately hierarchical fields have evolved over the years, we are able today to give a simple proof of Theorem 7 laying bare the underlying phenomena driving the model in Schmidt [52], which is the third paper attached to this thesis. To give the reader a first impression why the result is at least plausible we make some rough computations. Consider a point x on the unit torus and two circles around it of radii r and R satisfying $0 < r < R < \frac{1}{2}$. Using explicit asymptotics on Green's function on the torus it is not too hard to establish that one excursion from r to R and back to r takes on average about $\frac{1}{\pi} \ln \frac{R}{r}$ time. Not very surprisingly there is a law of large numbers for the time needed to perform many excursions and also exponential tail bounds hold. This concentration is sufficiently strong to justify replacing large times t with the time needed to perform the first $\frac{t}{\frac{1}{\pi}\ln\frac{R}{r}}$ many excursions from r to R and back. The probability that in one such excursion the ε-ball is hit is $\frac{\ln R - \ln r}{\ln R - \ln \varepsilon}$, which is easy to compute as the scenario can be identified with the same scenario on $\mathbb{R}^2$ and therefore the probability in question is rotationally invariant harmonic function of the starting point. Hence the probability of one small ε-ball not being hit up to some time t is roughly

$$\left(1 - \frac{\ln R - \ln r}{\ln R - \ln \varepsilon}\right)^{\frac{t}{\frac{1}{\pi}\ln\frac{R}{r}}} \approx \exp\left(\pi t (\ln \varepsilon)^{-1}\right). \tag{22}$$

As we can find of order ε^{-2} many disjoint ε-balls in a torus we can hope that the dependencies between them are not too strong and match the expected number of avoided balls among these ε^{-2} many to 1. This gives a guess for the critical time around which covering should happen by

$$\varepsilon^{-2} \exp\left(\pi t (\ln \varepsilon)^{-1}\right) \overset{!}{=} O(1), \tag{23}$$

which gives precisely $t = \frac{2}{\pi}(\ln \varepsilon)^2$. While this simple line of reasoning can be refined to establish an upper bound rigorously it shines not the slightest bit of light on why the dependencies are sufficiently weak, hence giving no idea how to find a matching lower bound. It turns out that the dependencies at hand are almost the same as the dependencies of branching Brownian motion making them barely weak enough for this first moment calculation to hit the leading order term precisely. To go into more detail about aforementioned analogy we need to make some observations on the behavior of the model first. We consider for some $R \in (0, 1/2)$ and $K \in \mathbb{N}$ the radii

$$r_i \equiv R \left(\frac{\varepsilon}{R}\right)^{i/K} \tag{24}$$

for $0 \leq i \leq K$ and associate to each point x on the torus the circles $(\partial B_{r_i}(x))_{i \leq K}$ which we call scales. Controlling the model is done by counting the number of excursions the Brownian motion performs up to some large time t. These excursion counts can be viewed as proxy for the occupation times displayed in Fig. 6. By identifying the circles with circles in $\mathbb{R}^2$ we see that starting at some circle r_i, $i \neq 0, K$ it is equally likely to hit the next smaller or the next larger circle first due to relative sizes of neighboring circles being constant. Hence tracking visits to circles (excluding consecutive visits to the same circle) and stopping upon hitting scale 0 gives a simple random walk stopped in 0 due to the strong Markov property of the Brownian motion being inherited. Starting the counting with the first visit to scale 1, stopping when hitting scale 0 and starting the next excursion when scale 1 is hit again we can read independent excursions of a SRW from 1 to 0 off the path of Brownian motion on the torus. Note that the independence of different excursions is due to the strong Markov property of Brownian motion and rotational invariance making the distribution of future hits of scales independent of the choice of starting point on scale 1. Taking the path displayed in Fig. 7 as an example we start at the black dot and track the path to the first hit of scale 1, which is marked by the blue dot. From there we follow the path writing down each hit to a non most recently visited scale, i.e. following the path and noting down the circle numbers along the red dots. As hitting scale K is the same as hitting an ε-ball, tracking these SRW excursions

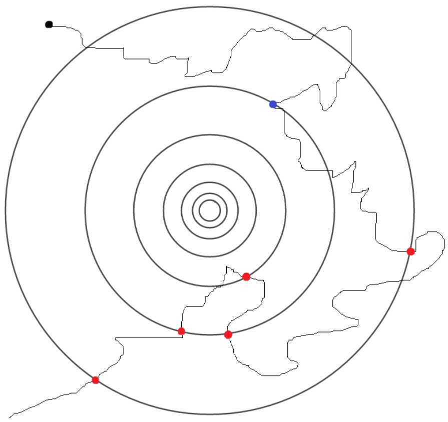

Figure 7: Reading off the excursions $1 \to 0$ and $1 \to 2 \to 1 \to 0$

is sufficient to decide weather an ε-ball is hit or not, given the information how many excursions W performs from scale 1 to scale 0. The number of excursions from scale 1 to scale 0 up to some large time t is however concentrated enough to replace these excursion numbers with constants in the proof. Establishing some notation for excursion counts we set

$$\begin{aligned} \mathcal{N}_l^x(n) \equiv \text{ number of excursions of } W \text{ from } \partial B_{r_l}(x) \text{ to } \partial B_{r_{l+1}}(x) \text{ within the} \\ \text{first } n \text{ excursions from } \partial B_{r_l}(x) \text{ to } \partial B_{r_{l-1}}(x) \text{ after time } \tau_{r_1}, \end{aligned} \tag{25}$$

for W Brownian motion on the torus and τ_{r_1} its first hitting time of scale 1. Note that for fixed x the $\mathcal{N}_l^x(n)$, $l \in \{1, ..., K-1\}$, $n \in \mathbb{N}$ are independent and distributed like sums of n independent geometrically distributed random variables of parameter $1/2$. Both independencies are due to the strong Markov property of the simple random walk. The geometrical distribution simply appears as the answer to the question"How often does a simple random walk started in j go from j to $j+1$ before hitting $j-1$?". While this gives very strong control over the probability of single ε-ball being hit or not we need to also keep the correlations between $\mathcal{N}^x$ and $\mathcal{N}^y$ in mind. As the circle sizes (associated to the scales) decay exponentially and the relative size difference becomes larger with ε getting smaller, the circle around x and the circle around y associated to some scale i are either practically identical or disjoint. No matter the distance of x and y this effect holds true up to at most 1 scale, which produces for a big number of scales (K large) only a small error. The

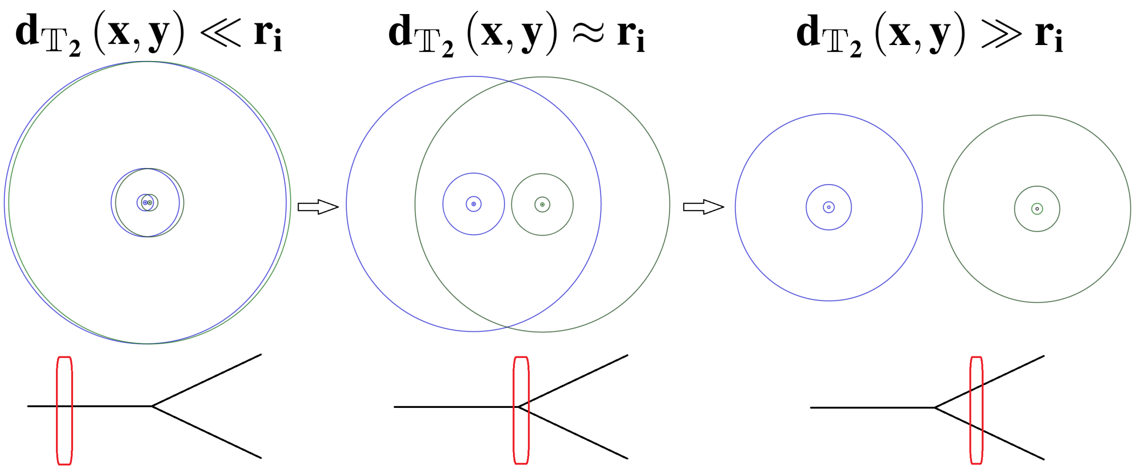

Figure 8: Scales seen by zooming towards two points.

phenomenon becomes almost obvious considering Figure 8: left the circles are almost identical, zooming in the circles are neither similar nor disjoint but only one scale later the circles are disjoint (right). This has the following two crucial consequences. On one hand if the circles associated to scale l around x and y are practically identical then so are the excursion counts $\mathcal{N}_l^x$ and $\mathcal{N}_l^y$. On the other hand if two circles are disjoint then conditionally on the exterior of both circles what happens inside one circle is independent of the events in the other by the strong Markov property of W. Hence $\mathcal{N}_l^x$ and $\mathcal{N}_m^y$ are perfectly independent if $B_{r_l}(x)$ and $B_{r_m}(y)$ are disjoint. This reveals that the model is approximately hierarchical and fixes up to small error the dependence structure. This effect is indicated schematically in the bottom of Figure 8. These ideas are the guiding principles of Schmidt [52]. We however establish some additional intuition for the model to see the strong connection to branching Brownian motion, which has played a major role in the control of the subleading order established by Belius and Kistler [13]. Considering

$$N_l^x(t) \equiv \text{Number of excursions } W \text{ completes from } \partial B_{r_l}(x) \text{ to } \partial B_{r_{l+1}}(x) \text{ before time } t, \tag{26}$$

it should be not too outlandish of a statement to the reader, that $\sqrt{N_l^x(t)}$ has very strong analogies to a branching Brownian motion. Let us draw a comparison considering two particles of branching Brownian motion. Given when the paths of these two particles split the increments are identical up to that point and given the past independent thereafter. Partitioning into K increments gives some identical pairs of increments in the beginning, some independent pairs of increments at the end and one mixed pair that for large K has only little influence. This is up to only approximately identical increments exactly what we observed for excursion counts of cover times. Also from the representation as sum of independent geometrical random variables we get that the increments $\sqrt{N_l^x(t)} - \sqrt{N_{l-1}^x(t)}$ given the past (i.e. given $\sqrt{N_{l-1}^x(t)}$) have the tail of a centered Gaussian of variance not depending l. This is exactly the case for branching Brownian motion. Finally we have to check the branching. In branching Brownian motion particles branch at constant rate, which is analogous to the ratio of neighboring circles being constant, as this keeps ratio between the number of circles with radius r_l we can place disjointly into the torus to the number of circles of radius r_{l+1} we can place disjointly into the torus asymptotically constant as well. Hence morally the "rate of branching" is essentially constant. As the correlation structure and tail behavior of increments match we expect to see the same behavior of extrema. This turns out to be true for all known results, which nail down the cover time up to an error of order one. This of course includes our first moment inspired guess (23) being sharp as is the case for branching Brownian motion. The reader may be interested to know that cover times in two dimensions is one of a plethora of models for which such an analogy holds and approximately hierarchical correlations are present. Some prominent examples of such models are the two-dimensional Gaussian free field [15, 16, 17, 18, 19, 28], characteristic polynomials of random unitary matrices [5, 29, 49] and extreme values of the Riemann zeta function on the critical line [6, 7, 10, 48].

Electron. Commun. Probab. **20** (2015), no. 47, 1–12.
DOI: 10.1214/ECP.v20-4189
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

From Derrida's random energy model to branching random walks: from 1 to 3

Nicola Kistler* Marius A. Schmidt†

Abstract

We study the extremes of a class of Gaussian fields with in-built hierarchical structure. The number of scales in the underlying trees depends on a parameter $\alpha \in [0,1]$: choosing $\alpha = 0$ yields the random energy model by Derrida (REM), whereas $\alpha = 1$ corresponds to the branching random walk (BRW). When the parameter α increases, the level of the maximum of the field decreases smoothly from the REM- to the BRW-value. However, as long as $\alpha < 1$ strictly, the limiting extremal process is always Poissonian.

Keywords: extreme value theory; extremal process; Gaussian hierarchical fields.
AMS MSC 2010: 60J80, 60G70, 82B44.
Submitted to ECP on March 17, 2015, final version accepted on June 9, 2015.
Supersedes arXiv:1503.04064.

1 Introduction and main result

The Gaussian fields we consider are constructed as follows. Let $\alpha \in [0,1]$ and $N \in \mathbb{N}$. We refer to the parameter N as the *size of the system*. For $j = 1 \dots N^\alpha$ and $\sigma_j = 1 \dots 2^{\left(N^{1-\alpha}\right)}$, consider the vectors $\sigma = (\sigma_1, \dots, \sigma_{N^\alpha})$. (We assume, without loss of generality, that N and α are such that N^α and $N^{1-\alpha}$ are both integers). We refer to the indices $j = 1 \dots N^\alpha$ as *scales*, and to the labels σ as *configurations*. The space of configurations is denoted by $\Sigma_N^{(\alpha)}$. Remark that, by construction, $\sharp\Sigma_N^{(\alpha)} = 2^N$. For scales $j \leq N^\alpha$ and $(\sigma_1, \dots, \sigma_j)$, consider independent centered Gaussian random variables $X_{\sigma_1,\dots,\sigma_j}^{(\alpha,j)}$ with variance $N^{1-\alpha}$ defined on some common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. To given configuration $\sigma \in \Sigma_N^{(\alpha)}$ we associate the *energies*

$$X_\sigma^{(\alpha,N)} \equiv \sum_{j=1}^{N^\alpha} X_{\sigma_1,\dots,\sigma_j}^{(\alpha,j)} \tag{1.1}$$

The collection $X^{(\alpha,N)} \equiv \left\{X_\sigma^{(\alpha,N)}, \sigma \in \Sigma_N^{(\alpha)}\right\}$ defines a centered Gaussian field with

$$\text{var}\left[X_\sigma^{(\alpha,N)}\right] = N, \quad \text{and} \quad \text{cov}\left[X_\sigma^{(\alpha,N)}, X_\tau^{(\alpha,N)}\right] = (\sigma \wedge \tau)\, N^{1-\alpha},$$

where $\sigma \wedge \tau \equiv \inf\{j \leq N^\alpha : (\sigma_1, \dots, \sigma_j) = (\tau_1, \dots, \tau_j) \text{ and } \sigma_{j+1} \neq \tau_{j+1}\}$. In spin glass terminology, $\sigma \wedge \tau$ is the *overlap* of the configurations σ and τ. In other words, the

*J.W. Goethe-Universität Frankfurt, Germany. E-mail: kistler@math.uni-frankfurt.de
†J.W. Goethe-Universität Frankfurt, Germany. E-mail: mschmidt@math.uni-frankfurt.de

Gaussian field $X^{(\alpha,N)}$ is *hierarchically* correlated. The parameter α governs the number of scales in the underlying "trees". The choice $\alpha = 0$ yields the celebrated REM of Derrida [12]; in this case the tree consists of a single scale (only for this boundary case is the field uncorrelated). The choice $\alpha = 1$ yields the (classical) BRW, also known as the directed polymer on Cayley trees [15]: in this model, the number of scales grows linearly with the size of the system. In this sense, the fields $X^{(\alpha,N)}$ interpolate between REM and BRW (remark that these boundary cases are, within our class, the least resp. the most correlated fields). See Figure 1 below for a graphical representation.

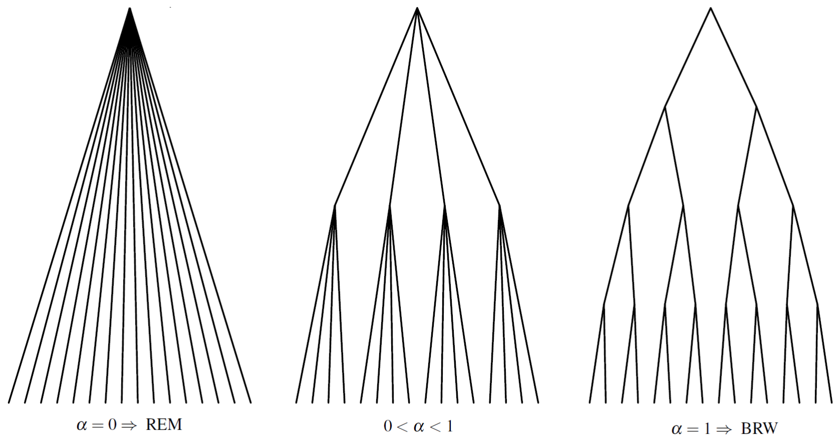

Figure 1: Trees interpolating between REM and BRW

A fundamental question in the study of random fields concerns the behavior of the extreme values in the limit of large system-size. The case of independent random variables is simple, and completely understood, see e.g. the classic [21]. On the other hand, the study of the extremes of *correlated* random fields is a much harder question. There is good reason to develop an extreme value theory for Gaussian fields defined on trees: besides being typically amenable to a detailed analysis (see e.g. [3, 5, 7, 8, 9, 10, 16, 22]), Gaussian hierarchical fields should be some sort of "universal attractors" in the limit of large system-size; this claim is a major pillar of the Parisi theory [24] which has remained to these days rather elusive (see however [19] and references therein for some recent advances). Our main result provides a characterization of the weak limit of the extremes of the hierarchical field (1.1).

Theorem 1.1. *Assume* $\alpha \in [0, 1)$. *Let*

$$a_N^{(\alpha)} \equiv \beta_c N - \frac{1+2\alpha}{2\beta_c} \log N, \quad \textit{where } \beta_c \equiv \sqrt{2\log 2},$$

and consider the random Radon measure on the real line

$$\Xi_N^{(\alpha)} \equiv \sum_{\sigma \in \Sigma_N^{(\alpha)}} \delta_{X_\sigma^{(\alpha,N)} - a_N^{(\alpha)}} .$$

Then $\Xi_N^{(\alpha)}$ *converges weakly to a Poisson process* Ξ *of intensity* $\mu(A) \equiv \int_A e^{-\beta_c x} dx/\sqrt{2\pi}$.

The weak limits of the extremes of Gaussian hierarchical fields with a *fixed* number of scales, the generalized random energy models by Derrida [13], have been rigorously derived in [10]. On the other hand, apart from the case $\alpha = 0$, the picture depicted in Theorem 1.1 seems to be new. There is good reason to leave out the case $\alpha = 1$: to clarify this, and to shed further light on our main result, let us spend a few words. First, the theorem implies that $a_N^{(\alpha)}$ is the *level of the maximum* of the random field $X^{(\alpha,N)}$, and $\Xi_N^{(\alpha)}$ is then the *extremal process*. It steadily follows from the convergence of the extremal process that the maximum of the field, recentered by its level, weakly converges to a Gumbel distribution. As expected under the light of (say) Slepian's Lemma, the level of the maximum *decreases* when α (hence the amount of correlations) *increases*. However, this feature is only detectable at the level of the second order, logarithmic corrections; curiously, the pre-factor $1 + 2\alpha$ interpolates smoothly between the REM- and the BRW-values ("from 1 to 3"). Notwithstanding, as long as $\alpha < 1$ strictly, and in spite of what might look at first sight as severe correlations, all our models fall into the universality class of the REM, which is indeed characterized by convergence towards Poissonian extremal processes. In the boundary case of the BRW, the picture is only partially correct: the logarithmic correction is still given by $a_N^{(\alpha)}$ with $\alpha = 1$, see [1, 2, 11], yet the weak limit of the maximum is no longer a Gumbel distribution [20], nor is the limiting extremal process a simple Poisson process [3, 5, 14, 22].

We conclude this section with a sketch of the proof of our main result. A natural approach would be to choose $a_N^{(\alpha)}$ such that the expected number of extremal configurations in any given compact $A \subset \mathbb{R}$ is of order one in the large N-limit. However, with the level of the maximum as given by Theorem 1.1, classical Gaussian estimates steadily yield

$$\mathbb{E}\left[\Xi_N^{(\alpha)}(A)\right] = 2^N \int_A \exp\left[-\left(x - a_N^{(\alpha)}\right)^2/(2N)\right] \frac{dx}{\sqrt{2\pi N}} = N^\alpha(1 + o(1)) \qquad (N \to \infty),$$

which is exploding as soon as $\alpha > 0$ strictly. The reason for this is easily identified: by linearity of the expectation, we are completely omitting correlations, but these turn out to be strong enough to affect the level of the maximum. To overcome this problem, we rely on the multi-scale analysis which has emerged in the study of the extremes of branching Brownian motion (see e.g. [19]). To formalize, we need some notation. First, for a given $\sigma \in \Sigma_N^{(\alpha)}$, we refer to the process

$$S^\sigma = (S_k^\sigma, k \le N^\alpha), \quad S_k^\sigma \equiv \sum_{j \le k} X_{\sigma_1,..\sigma_j}^{(\alpha,j)},$$

as the *path* of a configuration. (The process S^σ is a random walk with Gaussian increments, i.e. a discrete Brownian motion). We refer to any function $F_N : \{0 \dots N^\alpha\} \to \mathbb{R}$, $k \mapsto F_N(k)$, as *barrier*. Given a barrier F_N, we denote by

$$\Xi_{N,F_N}^{(\alpha)} \equiv \sum_{\sigma \in \Sigma_N} \delta_{X_\sigma^{(\alpha,N)} - a_N^{(\alpha)}} \mathbf{1}_{\left\{S_k^\sigma \le F_N(k) \text{ for all } k \in \{1,..,N^\alpha\}\right\}}$$

the *modified (extremal) process*. A key step in the proof is to identify a barrier E_N, see (2.8) below for its explicit form, such that for any compact $A \subset \mathbb{R}$,

$$\lim_{N \to \infty} \mathbb{P}\left[\Xi_N^{(\alpha)}(A) = \Xi_{N,E_N}^{(\alpha)}(A)\right] = 1. \tag{1.2}$$

This naturally entails that the weak limit of the extremal process and that of the modified process must coincide (provided one of the two exists). We will thus focus our attention on the modified process $\Xi_{N,E_N}^{(\alpha)}$, thereby proving that *mean of the process* as well as its *avoidance functions* converge to the Poissonian limit as given by Theorem 1.1, to wit:

$$\lim_{N\to\infty} \mathbb{E}\left[\Xi_{N,E_N}^{(\alpha)}(A)\right] = \mu(A) \qquad \text{(Convergence of mean)} \tag{1.3}$$

and

$$\lim_{N\to\infty} \mathbb{P}\left(\Xi_{N,E_N}^{(\alpha)}(A) = 0\right) = \mathbb{P}\left(\Xi(A) = 0\right) \qquad \text{(Avoidance functions)} \tag{1.4}$$

By (1.3) and (1.4), it follows from Kallenberg's theorem on Poissonian convergence [18], that the modified process weakly converges to the Poisson point process Ξ; but by (1.2), the same must be true for the extremal process, settling the proof of Theorem 1.1.

The rest of the paper is devoted to the proof of (1.2), (1.3) and (1.4). Since $\alpha \in [0,1)$ is fixed throughout, we lighten notations by dropping the α-dependence whenever no confusion can possibly arise, writing e.g. Σ_N for $\Sigma_N^{(\alpha)}$, X_σ for $X_\sigma^{(\alpha,N)}$, a_N for $a_N^{(\alpha)}$, etc.

2 Barriers, and the modified processes

The goal of this section is to construct the barrier E_N to which we alluded in the introduction, and to give a proof of (1.2) and (1.3). In a first step, we construct a barrier which is not "optimal", but which provides important *a priori* information:

Lemma 2.1. *Consider the barrier*

$$U_N(k) \equiv \beta_c k N^{1-\alpha} + \ln(N)\,,\ k = 0,..,N^\alpha.$$

It then holds:

$$\lim_{N\to\infty} \mathbb{P}\left(S_k^\sigma \le U_N(k) \quad \forall k \in \{1,..,N^\alpha\}, \sigma \in \Sigma_N\right) = 1\,.$$

Proof. By Markov inequality, and simple counting, it holds:

$$\begin{aligned} &\mathbb{P}\left(\exists \sigma \in \Sigma_N : \sum_{i\le j} X^{(i)}_{\sigma_1,..,\sigma_i} > U_N(j), \text{ for some } j \le N^\alpha\right) \\ &\qquad\qquad \le \sum_{j\le N^\alpha} \exp\left(jN^{1-\alpha}\ln 2\right) \mathbb{P}\left(\sum_{i\le j} X^{(j)}_{1,\dots,1} > \beta_c j N^{1-\alpha} + \ln N\right). \end{aligned} \tag{2.1}$$

By classical Gaussian estimates, the probability on the r.h.s. above is at most

$$\frac{\sqrt{jN^{1-\alpha}}}{\sqrt{2\pi}\left(\beta_c j N^{1-\alpha} + \ln N\right)} \exp\left[-\frac{\left(\beta_c j N^{1-\alpha} + \ln N\right)^2}{2jN^{1-\alpha}}\right].$$

Using this, and straightforward estimates, we get

$$(2.1) \le \exp\left[\left(\frac{3\alpha - 1}{2} - \beta_c\right)\ln N\right],$$

which is evidently vanishing in the large N-limit, since $\frac{3\alpha-1}{2} < \beta_c$. ☐

The above Lemma immediately implies that the weak limit of the extremal process Ξ_N and the weak limit of the modified process Ξ_{N,U_N} must necessarily coincide (provided one of the two exists). We now identify conditions under which this remains true for barriers which lie even lower than U_N.

Lemma 2.2. *Consider a barrier F_N with the following properties:*

i) $F_N \leq U_N$, *i.e.* $F_N(k) \leq U_N(k)$ *for all* k;

ii) for $A \subset \mathbb{R}$ *compact, it holds:*

$$\lim_{N\to\infty} \mathbb{E}\left[\Xi_{N,F_N}(A)\right] = \lim_{N\to\infty} \mathbb{E}\left[\Xi_{N,U_N}(A)\right]$$

Then the weak limits of Ξ_{N,F_N} *and* Ξ_{N,U_N} *coincide (provided one of the two exists).*

Proof. The Lemma steadily follows from the claim

$$\mathbb{P}\left(\Xi_{N,U_N}(A) = \Xi_{N,F_N}(A)\right) \geq 1 - \mathbb{E}\left[\Xi_{N,U_N}(A) - \Xi_{N,F_N}(A)\right]. \tag{2.2}$$

The proof of (2.2) is straightforward. Simple rearrangements and subadditivity imply that the for probability of the complementary event, it holds:

$$\begin{aligned}
&\mathbb{P}\left(\Xi_{N,U_N}(A) \neq \Xi_{N,F_N}(A)\right)\\
&= \mathbb{P}\left(\exists \sigma \in \Sigma_N : X_\sigma - a_N \in A,\ \forall_{j=1\dots N^\alpha}:\ S_j^\sigma \leq U_N(j) \text{ but } \exists_{j=1\dots N^\alpha}\ : S_j^\sigma > F_{j,N}\right)\\
&\leq \sum_{\sigma\in\Sigma_N} \mathbb{P}\left(X_\sigma - a_N \in A, \forall_{j=1\dots N^\alpha}:\ S_j^\sigma \leq U_N(j) \text{ but } \exists_{j=1\dots N^\alpha}\ : S_j^\sigma > F_{j,N}\right)\\
&= 2^N \mathbb{P}\left(X_\sigma - a_N \in A, \forall_{j=1\dots N^\alpha}:\ S_j^\sigma \leq U_N(j) \text{ but } \exists_{j=1\dots N^\alpha}\ : S_j^\sigma > F_{j,N}\right)\\
&= 2^N \mathbb{P}\left(X_\sigma - a_N \in A,\ S^\sigma \leq U_N\right) - 2^N \mathbb{P}\left(X_\sigma - a_N \in A,\ S^\sigma \leq F_N\right)\\
&= \mathbb{E}\left[\Xi_{N,U_N}(A) - \Xi_{N,F_N}(A)\right].
\end{aligned}$$

Building the complement, (2.2) immediately follows. ☐

By the previous Lemma, and in view of a proof of the main theorem, it is crucial to identify conditions for which the mean(s) of the modified process(es) converge to a finite limit. This is done by

Proposition 2.3. *Consider a barrier of the form* $F_N = U_N + f_N$, *where* f_N *is such that*

i) $f_N(0) = f_N(N^\alpha) = 0$

ii) $\sup_{k\in\{1,..,N^\alpha\}} |f_N(k)| = o\left(N^{\frac{1-\alpha}{2}}\right)$ *for* $N \uparrow \infty$.

For $A \subset \mathbb{R}$ *compact, and* μ *as in Theorem 1.1, it holds:*

$$\lim_{N\to\infty} \mathbb{E}\left[\Xi_{N,F_N}(A)\right] = \mu(A).$$

Proof. By linearity of the expectation, and by conditioning on the "terminal event",

$$\begin{aligned}
&\mathbb{E}\left[\Xi_{N,F_N}(A)\right] =\\
&\quad = 2^N \int_A \mathbb{P}\left(\forall_{k\in\{1,..,N^\alpha\}}:\ S_k^\sigma \leq F_N(k) \,\middle|\, X_\sigma - a_N = x\right) \mathbb{P}\left(X_\sigma - a_N \in dx\right).
\end{aligned} \tag{2.3}$$

Let us focus on the conditional probability: we first write this as

$$\begin{aligned}
&\mathbb{P}\left(\forall_{k\in\{1,..,N^\alpha\}}:\ S_k^\sigma \leq F_N(k) \,\middle|\, X_\sigma - a_N = x\right)\\
&\quad = \mathbb{P}\left(\forall_{k\in\{1,..,N^\alpha\}}:\ S_k^\sigma - \frac{k}{N^\alpha} X_\sigma \leq F_N(k) - \frac{k}{N^\alpha}(a_N + x) \,\middle|\, X_\sigma - a_N = x\right).
\end{aligned} \tag{2.4}$$

Inspection of the covariances shows that the Gaussian vector $\left(S_k^\sigma - \frac{k}{N^\alpha} X_\sigma, k = 1 \ldots N^\alpha\right)$ is, in fact, *independent* of X_σ. Using this, and rescaling by $N^{-\frac{1-\alpha}{2}}$ yields

$$(2.4) = \mathbb{P}\left(\forall_{k\in\{1,..,N^\alpha\}} :\ N^{-\frac{1-\alpha}{2}} \left[S_k^\sigma - \frac{k}{N^\alpha} S_{N^\alpha}^\sigma \right] \leq N^{-\frac{1-\alpha}{2}} \left[F_N(k) - \frac{k}{N^\alpha}\left(a_N + x\right) \right] \right). \tag{2.5}$$

Again by inspection of the covariances, one immediately realizes that the law of the Gaussian vector $\left(N^{-\frac{1-\alpha}{2}} \left[S_k^\sigma - \frac{k}{N^\alpha} S_{N^\alpha}^\sigma \right], k = 0 \ldots N^\alpha \right)$ is that of a (discrete) Brownian bridge of lifespan N^α, starting and ending in 0. To lighten notations, let $(B_{N^\alpha}(k), k \leq N^\alpha)$ be such a Brownian bridge, and shorten

$$\widetilde{F}_N(k,x) \equiv N^{-\frac{1-\alpha}{2}} \left[F_N(k) - \frac{k}{N^\alpha}\left(a_N + x\right) \right].$$

It thus holds:

$$(2.4) = \mathbb{P}\left(\forall_{k\in\{1,..,N^\alpha\}} :\ B_{N^\alpha}(k) \leq \widetilde{F}_N(k,x) \right).$$

One immediately checks that within our choice of the barrier F_N, and since $\alpha < 1$ strictly,

$$\lim_{N\uparrow\infty} \sup_{k\leq N^\alpha, x\in A} \left| \widetilde{F}_N(k,x) \right| = 0.$$

in which case it follows from the lemmata in the Appendix that

$$\begin{aligned} \mathbb{P}\left(\forall_{k\in\{1,..,N^\alpha\}} :\ B_{N^\alpha}(k) \leq \widetilde{F}_N(k,x) \right) &= \mathbb{P}\left(\forall_{k\in\{1,..,N^\alpha\}} :\ B_{N^\alpha}(k) \leq 0 \right)(1+o(1)) \\ &= N^{-\alpha}\left(1+o(1)\right), \end{aligned} \tag{2.6}$$

uniformly for x in compacts, and for $N \uparrow \infty$. Plugging this into (2.3), we have

$$\mathbb{E}\left[\Xi_{N,F_N}(A)\right] = 2^N N^{-\alpha}\left(1+o(1)\right) \int_A \mathbb{P}\left(X_\sigma - a_N \in dx\right).$$

The claim of the Proposition then immediately follows by straightforward estimates on the Gaussian density. □

Remark 2.4. The proof of Proposition 2.3 breaks down in the limiting case $\alpha = 1$: for this choice of the parameter, $F_N(k) - \frac{k}{N^\alpha}(a_N + x)$ is of logarithmic size, hence $\widetilde{F}_N(k,x)$ does not vanish in the large N limit. As technical as it may look, this is in fact a structural issue: for $\alpha = 1$ the extremal process cannot be a simple Poisson point process.

We can finally specify our choice of the barrier E_N alluded to in the introduction. The optimal choice is (by far) not unique, and depends on an additional free parameter γ. The only requirement is that

$$0 < \gamma < \frac{1-\alpha}{2}. \tag{2.7}$$

With any γ satisfying (2.7), and U_N as in Lemma 2.1, we set

$$E_N(k) \equiv U_N(k) - N^\gamma \mathbf{1}_{k\neq 0, N^\alpha} \tag{2.8}$$

This choice of a barrier clearly satisfies the assumptions of Proposition 2.3 and also Lemma 2.2. This has two fundamental consequences: first, the weak limit of the modified process $\Xi_{N,E_N}^{(\alpha)}$ and that of extremal process $\Xi_N^{(\alpha)}$ must necessarily coincide (provided one of the two exists); second, the mean of the modified process converges to the alleged limit, i.e. (1.3) holds with E_N as a barrier. Theorem 1.1 will thus follow as soon as we prove that avoidance functions (1.4) also converge with the very same choice for the barrier. This will be done in the next section by means of the Chen-Stein method. Before that, let us spend a few words "on what really stands behind" the choice of the barrier (2.8). (The discussion is intentionally informal: for details, the reader is referred e.g. to [19].)

First, we remark that by Lemma 2.1, the path of extremal configurations (the process $k \mapsto S_k^\sigma$ for σ s.t. $X_\sigma \approx a_N$) must necessarily satisfy the "U_N-barrier condition". As we have seen in Proposition 2.3, conditioning onto the terminal event turns the path into a Brownian bridge which is required to stay below 0 during its lifespan. It is well known that in order to achieve this, the bridge will behave within good approximation as the path of its modulus in the negative, $k \mapsto -\mid S_k^\sigma \mid$, which is typically much lower than the shift $-N^\gamma \mathbf{1}_{k\neq 0, N^\alpha}$ for $\gamma < (1-\alpha)/2$ (this is the so-called entropic repulsion, see e.g. [4]). In other words, requiring that the paths stay below E_N is no stricter requirement than asking them to stay below U_N. However, and crucially, the E_N-barrier is low enough to force the expected number of correlated extremal pairs to vanish in the large N-limit: it is this specific feature which stands behind the Chen-Stein method which we implement below.

3 Convergence of the avoidance functions

The goal of this section is to prove (1.4), which we recall reads

$$\lim_{N\to\infty} \mathbb{P}\left(\Xi_{N,E_N}(A) = 0\right) = \mathbb{P}\left(\Xi(A) = 0\right), \tag{3.1}$$

where E_N is given by (2.8), A is any compact set, and Ξ is a Poisson point process with density $\mu(A) = \int_A e^{-\beta_c x} dx/\sqrt{2\pi}$. To do so, we will use the so-called Chen-Stein method [6, Theorem 1A]. We begin with a warm-up computation. In what follows, we write $\mathcal{E}_N(\sigma)$ for the event that a configuration σ satifies the "E_N-barrier condition", more precisely:

$$\mathcal{E}_N(\sigma) \equiv \{\omega \in \Omega : S_k^\sigma(\omega) \le E_N(k), k = 1 \dots N^\alpha\}.$$

Recall that for two configurations $\sigma, \tau \in \Sigma_N^{(\alpha)}$, we denote by $\sigma \wedge \tau$ their overlap, namely the first scale at which the two configurations do not coincide.

Lemma 3.1 (Extremal pairs)**.** *Let $A \subset \mathbb{R}$ be compact. With the above notations, it holds:*

$$\mathbb{E}\left[\sharp\left\{\sigma,\tau:\ \sigma\wedge\tau \neq 0, N^\alpha,\ \text{and } X_\sigma - a_N \in A, \mathcal{E}_N(\sigma);\ X_\tau - a_N \in A, \mathcal{E}_N(\tau)\right\}\right] = o(1),$$

as $N \to \infty$.

It follows from Lemma 3.1 that energies of extremal configurations are, in fact, *independent* random variables. It will come hardly as a surprise that this feature stands behind the onset of the Poisson point process in the large N-limit.

Proof of Lemma 3.1. By linearity of the expectation, and re-arranging the ensuing sum according to the possible overlap-values, it holds:

$$\begin{aligned} &\mathbb{E}\left[\sharp\left\{\sigma,\tau:\ \sigma\wedge\tau \neq 0, N^\alpha,\ \text{and } X_\sigma - a_N \in A, \mathcal{E}_N(\sigma);\ X_\tau - a_N \in A, \mathcal{E}_N(\tau)\right\}\right] \\ &\quad = \sum_{K=1}^{N^\alpha - 1} \#\left\{(\sigma,\tau)\,|\sigma\wedge\tau = K\right\} \mathbb{P}\left(X_\sigma - a_N \in A, \mathcal{E}_N\left(\sigma\right),\ X_\tau - a_N \in A, \mathcal{E}_N\left(\tau\right)\right) \end{aligned} \tag{3.2}$$

Let us focus on the probability on the r.h.s. above: since σ and τ coincide up to scale K, by conditioning on the "trunk" which is shared by σ and τ, we get

$$\mathbb{P}\left(X_\sigma - a_N \in A, \mathcal{E}_N\left(\sigma\right); X_\tau - a_N \in A, \mathcal{E}_N\left(\tau\right)\right) = \int_{-\infty}^{E_{K,N}} (P) \times \mathbb{P}\left(S_K^\sigma \in dx\right), \tag{3.3}$$

where

$$(P) \equiv \mathbb{P}\left(x + (S^{\sigma}_{N^{\alpha}} - S^{\sigma}_{K}) - a_N \in A, \mathcal{E}_N(\sigma);\; x + (S^{\tau}_{N^{\alpha}} - S^{\tau}_{K}) - a_N \in A, \mathcal{E}_N(\tau) \,\middle|\, S^{\sigma}_{K} = x\right).$$

On the event appearing in (P) we drop the $\mathcal{E}$-requirements: by independence of the paths after the "branching point", this leads to

$$(3.3) \leq \int_{-\infty}^{E_{K,N}} \mathbb{P}\left(x + (S^{\sigma}_{N^{\alpha}} - S^{\sigma}_{K}) - a_N \in A\right)^2 \mathbb{P}\left(S^{\sigma}_{K} \in dx\right). \tag{3.4}$$

This steadily implies that the r.h.s. of (3.4) is at most

$$\begin{aligned} &\int_{-\infty}^{E_{K,N}} \left[\int_{A+a_N-x} \exp\left(-\frac{z^2}{2N^{1-\alpha}(N^{\alpha}-K)}\right) dz\right]^2 \exp\left(-\frac{x^2}{2N^{1-\alpha}K}\right) dx \\ &\leq 2^{-2N+KN^{1-\alpha}} \lambda(A) \times \\ &\quad \times \sup_{a\in A} \int_{-\infty}^{0} \exp\left(-\frac{\left(a_N^{(\alpha)} - E_{K,N} - x + a\right)^2}{N^{1-\alpha}(N^{\alpha}-K)} - \frac{(x+E_{K,N})^2}{2N^{1-\alpha}K} + \left(2N - KN^{1-\alpha}\right)\ln 2\right) dx, \end{aligned} \tag{3.5}$$

where λ denotes Lebesgue measure. The argument of the exponential in (3.5) is easily seen to be bounded by $\beta_c\left(3\ln N - N^{\gamma} + x - 2a\right)$, hence

$$\begin{aligned} (3.3) &\leq 2^{-2N+KN^{1-\alpha}} \lambda(A) \sup_{a\in A} \int_{-\infty}^{0} \exp\left[\beta_c\left(3\ln N - N^{\gamma} + x - 2a\right)\right] dx \\ &\leq 2^{-2N+KN^{1-\alpha}} C_A \exp\left[\beta_c\left(3\ln N - N^{\gamma}\right)\right], \end{aligned} \tag{3.6}$$

with $C_A \equiv \frac{1}{\beta_c}\lambda(A)\exp\left[-2\beta_c \inf\{A\}\right]$. Plugging (3.6) into (3.2), and using that

$$\#\left\{(\sigma,\tau)\,|\sigma\wedge\tau = K\right\} \times 2^{-2N+KN^{1-\alpha}} \leq 1,$$

we get

$$(3.2) \leq \sum_{K=1}^{N^{\alpha}-1} C_A \exp\left[\beta_c\left(3\ln N - N^{\gamma}\right)\right],$$

which is evidently vanishing in the large N-limit. □

We can now finally move to the last missing piece, namely a proof of convergence of the avoidance functions (3.1). As mentioned, the main technical device here will be the so-called Chen-Stein method, [6, Theorem 1A]. To implement this, we need to introduce some notation. For compact $A \subset \mathbb{R}$, we shorten

$$\mu_N(A) \equiv \mathbb{E}\left[\Xi_{N,E_N}(A)\right]$$

and denote by $\mathcal{L}_N(A)$ the *law* of the random variable $\Xi_{N,E_N}(A)$. For a (sigma-finite) measure ν on $\mathbb{R}$, we denote by $\mathrm{Pois}_{\nu(A)}$ the law of a Poisson random variable with mean $\nu(A)$. For $\rho, \rho' \in \mathcal{M}_1(\mathbb{R})$ two probability measures on $\mathbb{R}$ we denote by $d_{TV}(\rho,\rho')$ their distance in total variation. In order to closely stick to the notation in [6], we write

$$\Xi_{N,E_N}(A) = \sum_{\sigma\in\Sigma_N^{(\alpha)}} I_{\sigma}, \quad I_{\sigma} \equiv \mathbf{1}_{\{X_{\sigma}-a_N\in A, \mathcal{E}_N(\sigma)\}},$$

and define, for given $\sigma \in \Sigma_N$,

$$Z_\sigma \equiv \sum_{\tau \in \Sigma_N, \tau \wedge \sigma \neq 0, N^\alpha} I_\tau \,.$$

For a last piece of notation, we shorten $p_\sigma \equiv \mathbb{E}[I_\sigma]$.

Coming back to our main task of proving (3.1), with $\mu(A)$ as in Theorem 1.1, it holds:

$$\begin{aligned} \Big| \mathbb{P}\left(\Xi_{N,E_N}(A) = 0\right) - \mathbb{P}\left(\Xi(A) = 0\right) \Big| &\leq d_{TV}\left(\mathcal{L}_N(A), \text{Pois}_{\mu(A)}\right) \\ &\leq d_{TV}\left(\mathcal{L}_N(A), \text{Pois}_{\mu_N(A)}\right) + d_{TV}\left(\text{Pois}_{\mu_N(A)}, \text{Pois}_{\mu(A)}\right) \end{aligned} \tag{3.7}$$

The convergence of $\mu_N(A)$ towards $\mu(A)$ is guaranteed by Proposition 2.3; in virtue of simple properties of Poisson random variables, this convergence implies that the second term on the r.h.s. above vanishes in the limit of large N. Concerning the first term on the r.h.s. of (3.7): the Chen-Stein method [6, Theorem 1A] yields the bound

$$d_{TV}\left(\mathcal{L}_N(A), \text{Pois}_{\mu_N(A)}\right) \leq \sum_{\sigma \in \Sigma_N} \Big(p_\sigma^2 + p_\sigma \mathbb{E}\left[Z_\sigma\right] + \mathbb{E}\left[I_\sigma Z_\sigma\right] \Big) \tag{3.8}$$

Since for any $\sigma \in \Sigma_N$, $p_\sigma = 2^{-N}\mu_N(A)$, one immediately gets

$$\sum_{\sigma \in \Sigma_N} p_\sigma^2 = 2^{-N} \mu_N(A)^2, \tag{3.9}$$

and by simple counting,

$$\sum_{\sigma \in \Sigma_N} p_\sigma \mathbb{E}\left[Z_\sigma\right] \leq 2^{-N^{1-\alpha}} \mu_N(A)^2. \tag{3.10}$$

Plugging (3.9) and (3.10) in (3.8) we get

$$\begin{aligned} d_{TV}\left(\mathcal{L}_N(A), \text{Pois}_{\mu_N(A)}\right) &\leq 2^{-N}\mu_N(A)^2 + 2^{-N^{1-\alpha}}\mu_N(A)^2 + \sum_{\sigma \in \Sigma_N^{(\alpha)}} \mathbb{E}\left[I_\sigma Z_\sigma\right] \\ &= 2^{-N}\mu_N(A)^2 + 2^{-N^{1-\alpha}}\mu_N(A)^2 + \sum_{\sigma \wedge \tau \neq 0, N^\alpha} \mathbb{E}\left[I_\sigma I_\tau\right], \end{aligned} \tag{3.11}$$

the last equality by definition of Z_σ. Since $\mu_N(A)$ converges to a finite limit (by Proposition 2.3), the first two terms in the last display of (3.11) vanish in the limit of large N; the third term is exactly what was analyzed in Lemma 3.1, and therefore also vanishing. All in all, (3.7) is vanishing, hence (3.1) holds and the proof of Theorem 1.1 is concluded.

Appendix

A fundamental ingredient in the proof of Theorem 1.1 are the estimates (2.6) on Brownian bridge probabilities appearing in the proof of Proposition 2.3; these are somewhat classical [17], sometimes going under the name of "ballot theorems". For the reader's convenience, we give here a short proof of the estimates as needed in our framework.

Lemma 3.2. *Let $(\Delta_i)_{i\in\{0,..,n-1\}}$ be i.i.d random variables having a density with respect to the Lebesgue measure and $(B_n(j), j \in \{1,..,n\})$ the related bridge, i.e.*

$$B_n(j) \equiv \sum_{i=0}^{j-1} \Delta_i - \frac{j}{n}\sum_{i=0}^{n-1} \Delta_i,$$

then it holds:

$$\mathbb{P}\left(B_n(j) \leq 0 \textit{ for all } j \in \{1,..,n\}\right) = \frac{1}{n} \tag{3.12}$$

Proof. We refer to $(\Delta_i)_{i\in\{0,..,n-1\}}$ as *increments*. The event in (3.12) is equivalent to the maximum of the bridge being lower than zero. Let $m \in \{0, .., n-1\}$ be the position of the maximum; remark that this is almost surely unique by the density-assumption. One steadily checks that applying a cyclic permutation, say π, to the increments of the bridge, shifts the position of the maximum to $\pi^{-1}m$. There is one cyclic permutation only, say $\hat{\pi}$, which shifts the position of the maximum to the origin, i.e. for which $\hat{\pi}^{-1}m = 0$. On the other hand, the distribution of the bridge is not affected by any permutation, hence $\hat{\pi}$ must be uniformly distributed among the n possible cyclic permutations: since the event in (3.12) is equivalent to $\hat{\pi}$ being the identity, the Lemma follows. □

In other words, the probability that a discrete bridge stays below zero during its lifetime decays as the inverse of the length of the bridge. On the other hand, since our bridges have "square-root fluctuations", one expects that whether the bridge is required to stay below zero or below a straight line shouldn't alter (much) the asymptotic behavior of these probabilities. This is indeed the case:

Lemma 3.3. *Let $(B_n(j), j = \dots n)$ be a (discrete) Brownian bridge of length n. Then there exists $c > 0$ independent of n such that for any n and $|\varepsilon| \le c^{-1}$, it holds:*

$$\Big|\mathbb{P}\left(B_n(j) \le 0, j = 1\dots n-1\right) - \mathbb{P}\left(B_n(j) \le \varepsilon, j = 1\dots n-1\right)\Big| \le c\frac{|\varepsilon|}{n}.$$

Proof. We proceed by induction on the length of the bridge.

Base case. For $n = 2$, it clearly holds:

$$\mathbb{P}\left(B_n(j) \le \varepsilon, j \in \{1\}, \exists j \in \{1\} : B_n(j) > 0\right) = \mathbb{P}\left(B_n(1) \in [0, \varepsilon]\right) \le \frac{2}{\sqrt{\pi}}\frac{|\varepsilon|}{n} \quad \text{for } \varepsilon > 0.$$

$$\mathbb{P}\left(B_n(j) \le 0, j \in \{1\}, \exists j \in \{1\} : B_n(j) > \varepsilon\right) = \mathbb{P}\left(B_n(1) \in [\varepsilon, 0]\right) \le \frac{2}{\sqrt{\pi}}\frac{|\varepsilon|}{n} \quad \text{for } \varepsilon < 0.$$

The proof in the cases $\varepsilon > 0$ and $\varepsilon < 0$ are similar, we thus consider only the first case.

Induction step. For $n \ge 3$, assume the claim is true for all $k \le n-1$. By Markov inequality,

$$\begin{aligned} &\mathbb{P}\left(B_n(j) \le \varepsilon, j = 1\dots n-1 \text{ but } \exists i = 1\dots n-1 : B_n(i) > 0\right) \\ &\le \sum_{i=1}^{n-1} \mathbb{P}\left(B_n(j) \le \varepsilon, j = 1\dots n-1 \text{ but } B_i > 0\right) \\ &= \sum_{i=1}^{n-1} \int_0^{\varepsilon} \mathbb{P}\left(B_n(j) \le \varepsilon,\ j = 1\dots n-1 \,\Big|\, B_n(i) = x\right) \mathbb{P}\left(B_n(i) \in dx\right). \end{aligned} \tag{3.13}$$

By the Markov property of Brownian bridges, the conditional probability above reads

$$\begin{aligned} &\mathbb{P}\left(\forall_{j\in\{1,..,i\}} :\ B_n(j) \le \varepsilon \,\Big|\, B_n(i) = x\right) \mathbb{P}\left(\forall_{j\in\{i,..,n-1\}}\ B_n(j) \le \varepsilon \Big| B_n(i) = x\right) \\ &\quad \le \mathbb{P}\left(\forall_{j\in\{1,..,i\}} :\ B_n(j) \le \varepsilon \,\Big|\, B_n(i) = 0\right) \mathbb{P}\left(\forall_{j\in\{i,..,n-1\}}\ B_n(j) \le \varepsilon \Big| B_n(i) = 0\right) \end{aligned}$$

where the inequality follows by monotonicity in x. Using this, and Gaussian estimates,

$$\begin{aligned} (3.13) \le \sum_{i=1}^{n-1} &\mathbb{P}\left(\forall_{j\in\{1,..,i\}} : B_n(j) \le \varepsilon | B_n(i) = 0\right) \times \\ &\times \mathbb{P}\left(\forall_{j\in\{i,..,n-1\}} :\ B_n(j) \le \varepsilon | B_n(i) = 0\right) \frac{\varepsilon}{\sqrt{2\pi\left(i - \frac{i^2}{n}\right)}}. \end{aligned}$$

Given that $B_n(i) = 0$, the process $(B_n(j), j = 1 \dots n)$ is a Brownian bridge of length $i \leq n-1$; analogously, the second probability involves a Brownian bridge of length $n-i$. The assumption therefore applies, and the above is at most

$$\sum_{i=1}^{n-1} \left(\mathbb{P}\left(\forall_{j\in\{1,..,i\}} : B_i(j) \leq 0\right) + c\frac{\varepsilon}{i} \right) \times \\ \times \left(\mathbb{P}\left(\forall_{j\in\{1,..,n-i\}} : B_{n-i}(j) \leq 0\right) + c\frac{\varepsilon}{n-i} \right) \frac{\varepsilon}{\sqrt{2\pi\left(i - \frac{i^2}{n}\right)}}. \tag{3.14}$$

It then holds:

$$(3.14) \leq \varepsilon \frac{4\sqrt{n}}{\sqrt{2\pi}} \sum_{i=1}^{n-1} \left(\frac{1}{i\,(n-i)} \right)^{3/2} \leq \varepsilon \frac{8\sqrt{n}}{\sqrt{2\pi}} \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} \left(\frac{1}{i\,(n-i)} \right)^{3/2} \leq c\frac{\varepsilon}{n}\,,$$

the last inequality using the bound $[i(n-i)]^{-3/2} \leq [i(n/2)]^{-3/2}$, for $i \leq \lfloor n/2 \rfloor$. $\square$

References

[1] L. Addario-Berry and B. Reed, *Minima in branching random walks*, Ann. of Prob. 37, 1044-1079 (2009) MR-2537549

[2] E. Aidekon, *Convergence in law of the minimum of a branching random walk*, Ann. of Prob. 41, 362-1426 (2013) MR-3098680

[3] E. Aidekon, J. Berestycki, E. Brunet, and Z. Shi. *Branching Brownian motion seen from its tip*, Probab. Theor. Rel. Fields, 157, 405-451 (2013). MR-3101852

[4] L.-P. Arguin, A. Bovier, and N. Kistler, *Genealogy of extremal particles of branching Brownian motion*, Comm. Pure and Appl. Math. 64, 1647-1676 (2011).MR-2838339

[5] L.-P. Arguin, A. Bovier, and N. Kistler, *The extremal process of branching Brownian motion*, Probab. Theor. Rel. Fields 157, 535-574 (2013). MR-3129797

[6] A.D. Barbour, L. Holst, and S. Janson, *Poisson Approximation*, Oxford: Clarendon Press, (1992). MR-1163825

[7] A. Bovier and L. Hartung, *The extremal process of two-speed branching Brownian motion*, Elect. J. Probab. 19, No. 18, 1-28 (2014) MR-3164771

[8] A. Bovier and L. Hartung, *Variable speed branching Brownian motion 1. Extremal processes in the weak correlation regime*, arXiv e-prints (2014)

[9] A. Bovier and L. Hartung, *Extended Convergence of the Extremal Process of Branching Brownian Motion*, arXiv e-prints (2014).

[10] A. Bovier, I. Kurkova. *Derrida's generalized random energy models. I. Models with finitely many hierarchies*. Ann. Inst. H. Poincare. Prob. et Statistiques (B) Prob. Stat. 40, 439-480 (2004). MR-2070334

[11] M. Bramson, *Maximal displacement of branching Brownian motion*, Comm. Pure Appl. Math. 31, 531-581 (1978). MR-0494541

[12] B. Derrida, *Random-energy model: An exactly solvable model of disordered systems*, Phys. Rev. B (3), 24 (1981) MR-0627810

[13] B. Derrida, *A generalization of the random energy model which includes correlations between energies*, Jour. de Phys. Lett. 46, 401-407 (1985).

[14] E. Brunet and B. Derrida, *A branching random walk seen from the tip*, Jour. Stat. Phys. 143, 420-446 (2010). MR-2799946

[15] B. Derrida and H. Spohn, *Polymers on disordered trees, spin glasses, and travelling waves*, Jour. Stat. Phys. 51, 817-840 (1988). MR-0971033

[16] M. Fang, and O. Zeitouni, *Slowdown for time inhomogeneous branching Brownian motion*, Jour. Stat. Phys. 149, 1-9 (2012). MR-2981635

[17] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. I, Wiley (1968) MR-0228020

[18] O. Kallenberg, *Random measures*, Akademie-Verlag, Berlin (1983). MR-0818219

[19] N. Kistler, *Derrida's random energy models. From spin glasses to the extremes of correlated random fields*, Springer Lecture Notes, to appear (2015+)

[20] S.P. Lalley and T. Sellke, *A conditional limit theorem for the frontier of a branching Brownian motion*, Ann. of Prob, 15, 1052-1061 (1987). MR-0893913

[21] M. R. Leadbetter, G. Lindgren, and H. Rootzen. *Extremes and related properties of random sequences and processes*. Vol. 21. New York: Springer-Verlag (1983). MR-0691492

[22] T. Madaule, *Convergence in law for the branching random walk seen from its tip*, arXiv e-prints (2011).

[23] H.P. McKean, *Application of Brownian Motion to the equation of Kolmogorov-Petrovskii-Piskunov*, Comm.Pure. Appl. Math. 28: 323-331 (1976). MR-0400428

[24] M. Mezard, M.A. Virasoro, and G. Parisi, *Spin glass theory and beyond*. World scientific (1987). MR-1026102

Acknowledgments. This paper owes much to conversations with Bernard Derrida, who raised in particular the question whether models with an increasing number of scales can provide (possibly quantitative) insights into the fractal structure of the extremal process of BRW/branching Brownian motion. Unfortunately, our main result shows that this is not the case, at least as long as $\alpha < 1$. It is tempting to believe that letting α depend on the size of the system (i.e. $\alpha(N) \to 1$, as $N \uparrow \infty$) gives rise to more interesting extremal processes.

ON MCKEAN'S MARTINGALE IN THE BOVIER-HARTUNG EXTREMAL PROCESS

CONSTANTIN GLENZ, NICOLA KISTLER, AND MARIUS A. SCHMIDT

Abstract. It has been proved by Bovier & Hartung [*Elect. J. Probab.* **19** (2014)] that the maximum of a variable-speed branching Brownian motion (BBM) in the weak correlation regime converges to a randomly shifted Gumbel distribution. The random shift is given by the almost sure limit of McKean's martingale, and captures the early evolution of the system. In the Bovier-Hartung extremal process, McKean's martingale thus plays a role which parallels that of the derivative martingale in the classical BBM. In this note, we provide an alternative interpretation of McKean's martingale in terms of a law of large numbers for *high-points* of BBM, i.e. particles which lie at a macroscopic distance from the edge. At such scales, 'McKean-like martingales' are naturally expected to arise in all models belonging to the BBM-universality class.

1. Introduction

Over the last years, one has witnessed an explosion of activity in the study of the extremes of Branching Brownian Motion, BBM for short. The list of papers on the subject is way too long to be given here: below, we shall only mention those works which are indispensable for the discussion, and refer the reader to Bovier's monograph [14] for an exhaustive overview of the literature.

The classical, supercritical and *time-homogeneous* BBM is constructed as follows. A single particle performs standard Brownian motion $x(t)$, starting at 0 at time 0. After an exponential random time T of mean one and independent of x, the particle splits into two (say) particles. The positions of these particles are independent Brownian motions starting at $x(T)$. Each of these processes have the same law as the first Brownian particle. Thus, after a time $t > 0$, there will be $n(t)$ particles located at $x_1(t), \dots, x_{n(t)}(t)$, with $n(t)$ being the random number of offspring generated up to that time (note that $\mathbb{E}n(t) = e^t$).

A fundamental link between BBM and partial differential equations was observed by McKean [27], who showed that the law of the maximal displacement of BBM solves the celebrated KPP-equation [25]. Thanks to the cumulative works [25, 27, 16, 26] it is now known that the maximum of BBM weakly converges, upon recentering, to a random shift of the Gumbel distribution. More precisely, let

$$m(t) \equiv \sqrt{2}t - \frac{3}{2\sqrt{2}} \log t, \qquad M(t) \equiv \max_{k \leq n(t)} x_k(t) - m(t) \ , \tag{1.1}$$

Date: December 13, 2017.

2000 *Mathematics Subject Classification.* 60J80, 60G70, 82B44.

Key words and phrases. (in)homogeneous Branching Brownian motions, McKean martingale, Bovier-Hartung extremal process.

and consider the so-called *derivative martingale*

$$Z(t) \equiv \sum_{k \leq n(t)} \left(\sqrt{2}t - x_k(t)\right) \exp -\sqrt{2}\left(\sqrt{2}t - x_k(t)\right) . \tag{1.2}$$

Leaning on the work of McKean [27] and Bramson [16], Lalley and Sellke [26] proved that

$$\lim_{t\uparrow\infty} Z(t) = Z \text{ a.s.}, \tag{1.3}$$

with Z a positive random variable, and that

$$\lim_{t\to\infty} \mathbb{P}\left[M(t) \leq x\right] = \mathbb{E} \exp -CZe^{-\sqrt{2}x} . \tag{1.4}$$

where $C > 0$ is a numerical constant. Inspecting the proof of this result, one gathers that the derivative martingale captures the early evolution of the system. Perhaps a more intuitive interpretation of the derivative martingale has been given by Arguin, Bovier and Kistler [5] in the form of an ergodic theorem, to wit:

$$\lim_{t\uparrow\infty} \frac{1}{t}\int_0^t \mathbb{1}\{M(s) \leq x\}\, ds = \exp\left(-CZe^{-\sqrt{2}x}\right) \quad \text{almost surely.} \tag{1.5}$$

The derivative martingale may thus be seen as a *measure of success*, capturing the fraction of particles which reach maximal heights.

Also of interest are *time-inhomogeneous* BBMs. These have been first introduced by Derrida and Spohn [22], and are constructed as follows: one considers a BBM where, at time s, all particles move independently as Brownian motions with time-dependent variance

$$\sigma^2(s) = \begin{cases} \sigma_1^2 & 0 \leq s < t/2 \\ \sigma_2^2 & t/2 \leq s \leq t. \end{cases} \tag{1.6}$$

In the above, σ_1, σ_2 are (positive) parameters chosen in such a way that the total variance is normalised to unity, to wit: $\sigma_1^2/2 + \sigma_2^2/2 = 1$. (One may also consider $K > 2$ distinct variance-regimes, but the qualitative picture does not change much, as long as K remains finite). Denoting by $\hat{n}(s)$ the number of particles at time s, and by $\{\hat{x}_k(s), k \leq \hat{n}(s)\}$ their position, it has been proved by Fang and Zeitouni [23] that

$$\max_{k\leq\hat{n}(t)} \hat{x}_k(t) = \begin{cases} \sqrt{2}t - \frac{1}{2\sqrt{2}}\log(t) + O_{\mathbb{P}}(1) & \text{if } \sigma_1 < \sigma_2 \\ \sqrt{2}(\sigma_1/2 + \sigma_2/2) - \frac{3}{2\sqrt{2}}(\sigma_1+\sigma_2)\log(t) + O_{\mathbb{P}}(1) & \text{if } \sigma_1 > \sigma_2, \end{cases} \tag{1.7}$$

The second case above, namely $\sigma_1 > \sigma_2$, is easily understood: the maximum of the time-inhomogeneous process is given by the superposition of the relative maxima at time $t/2$ and the maxima of their offspring after a $t/2$-lifespan.

The first case, $\sigma_1 < \sigma_2$, is arguably more interesting as it shows that the level of the maximum coincides with that of Derrida's REM [21]: in spite of what may look as severe correlations between the Brownian particles, the extremes behave as if they were coming from a field of independent random variables. This, however, is only true at the above level of precision, as the weak limit of the maximum does detect the underlying correlations. To formulate precisely, let us shorten

$$m_{\text{REM}}(t) \equiv \sqrt{2}t - \frac{1}{2\sqrt{2}}\log(t), \quad \hat{M}(t) \equiv \max_{k\leq\hat{n}(t)} \hat{x}_k(t) - m_{\text{REM}}(t). \tag{1.8}$$

Bovier and Hartung [15] prove that in such a regime of "REM-collapse" ($\sigma_1 < \sigma_2$),

$$\lim_{t\to\infty} \mathbb{P}\left[\hat{M}(t) \le x\right] = \mathbb{E}\exp -\hat{C}\hat{Z}e^{-\sqrt{2}x}, \tag{1.9}$$

for a numerical constant $\hat{C} > 0$ and $\hat{Z}$ a positive random variable. Much akin to the homogeneous BBM, the weak limit of the time-inhomogeneous process is thus given by a random shift of the Gumbel distribution. For our purposes, it is however crucial to emphasize that the random shift in case of REM-collapse is *not* given by the derivative martingale, but by the so-called *McKean's martingale.* To see how the latter comes about, recall the *time-homogeneous* BBM $\{x_k(t), k \le n(t)\}$, and define *McKean's martingale* by

$$\hat{Z}(t) \equiv \sum_{k\le n(t)} \exp\left[-t(1+\sigma_1^2) + \sqrt{2}\sigma_1 x_k(t)\right]. \tag{1.10}$$

Bovier and Hartung [15] show that this is, in fact, a square integrable martingale, *provided* that $\sigma_1 < 1$ strictly. It therefore converges almost surely to a well defined random variable whose law coincides with that of the $\hat{Z}$-random variable shifting the maximum (1.9) of the *time-inhomogeneous* process[1].

The analogy with the homogeneous BBM goes even further: an inspection of the proof of the weak convergence (1.9) shows that McKean's martingale captures, in fact, the early evolution of the system (remark, in particular, that McKean's martingale depends solely on σ_1).

The purpose of these notes is to present yet another interpretation of McKean's martingale, somewhat close in spirit to the ergodic theorem (1.5). To formulate precisely, let

$$\alpha \in (0, \sqrt{2}), \quad \Delta_\alpha \equiv \sqrt{2} - \alpha, \quad \text{and} \quad Z_\alpha(t) \equiv \sharp\{k \le n(t) : x_k(t) \ge \Delta_\alpha t\}. \tag{1.11}$$

The random variable $Z_\alpha(t)$ thus counts the "α-high-points", those particles which lag behind the leader at time t by a macroscopic distance αt. Finally, consider the McKean's martingale

$$Y_\alpha(t) \equiv \sum_{k\le n(t)} \exp\left[-t\left(1+\frac{1}{2}\Delta_\alpha^2\right) + \Delta_\alpha x_k(t)\right]. \tag{1.12}$$

Through the matching $\alpha \equiv \sqrt{2}(1-\sigma_1)$, and by the aforementioned result of Bovier and Hartung we see that this is, for $\alpha > 0$, a square integrable martingale whose limit

$$\lim_{t\to\infty} Y_\alpha(t) =: Y_\alpha \tag{1.13}$$

exists almost surely. Here is our main result.

Theorem 1.1. *(Strong law of large numbers) For any* $0 < \alpha < \sqrt{2}$*, and* Y_α *as in* (1.13),

$$\lim_{t\to\infty} \frac{Z_\alpha(t)}{\mathbb{E}Z_\alpha(t)} = Y_\alpha, \quad \textit{almost surely.} \tag{1.14}$$

[1]We emphasize tha the REM-collapse holds only for $\sigma_1 < \sigma_2$; given the normalization $\sigma_1^2/2+\sigma_2^2/2 = 1$, this is equivalent to $\sigma_1 < \sqrt{2}/2$. The square integrability of McKean's martingale holds however for any $\sigma_1 < 1$. The choice $\sigma_1 = 1$ corresponds to a boundary case where square integrability no longer holds. It has furthermore been proved by Lalley and Sellke [26] that for $\sigma_1 = 1$ the limit of McKean's martingale vanishes, in which case it is the derivative martingale that enters the picture for the weak limit of the maximum of (the time-homogeneous) BBM.

According to the SLLN, the random shift entering the weak limit (1.9) in the Bovier-Hartung extremal process thus captures the average number of *successful* particles. It is however noteworthy that the definition of *success* comes here with a twist: it pertains to those particles reaching[2] heights which lie *macroscopically lower* than the level of the maximum; this should be contrasted with (1.5), where successful particles reach *extremal* heights. Half-jokingly, we may thus say that Lalley and Sellke's derivative martingale is way more elitist than McKean's martingale!

The proof of the SLLN follows the strategy of [5]. Precisely, for some $r = o(t)$ to be specified later, and $\mathcal{F}_r \equiv \sigma\left(x_k(r), k \leq n(r)\right)$ the standard filtration of BBM, we first decompose telescopically

$$\frac{Z_\alpha(t)}{\mathbb{E}Z_\alpha(t)} = \frac{\mathbb{E}\left[Z_\alpha(t) \mid \mathcal{F}_r\right]}{\mathbb{E}Z_\alpha(t)} + \frac{Z_\alpha(t) - \mathbb{E}\left[Z_\alpha(t) \mid \mathcal{F}_r\right]}{\mathbb{E}Z_\alpha(t)}. \tag{1.15}$$

The next Proposition will seamlessly follow from the strong Markov property of BBM and classical Gaussian estimates.

Proposition 1.2. *(Onset of McKean's martingale) It holds:*

$$\frac{\mathbb{E}\left[Z_\alpha(t) \mid \mathcal{F}_r\right]}{\mathbb{E}Z_\alpha(t)} = (1+o(1))Y_\alpha(r), \qquad \textit{almost surely.} \tag{1.16}$$

By the above, and with our main theorem in mind, an important ingredient is therefore to prove that the second term on the r.h.s. of the telescopic decomposition (1.15) yields an irrelevant contribution in the limit of large times. This is guaranteed by the following

Proposition 1.3. *There exists $\kappa_\alpha > 0$ such that for $r = o(t)$ as $t \to \infty$,*

$$\mathbb{P}\left(\left|\frac{Z_\alpha(t) - \mathbb{E}\left[Z_\alpha(t) \mid \mathcal{F}_r\right]}{\mathbb{E}Z_\alpha(t)}\right| > c\right) \leq (1+o(1))\frac{c+1}{c^2}e^{-\kappa_\alpha r}\,, \tag{1.17}$$

A small remark concerning the conceptual picture behind Proposition 1.3 is perhaps at place. As mentioned, the proof strategy and, in particular, the telescopic decomposition (1.15), are borrowed from [5]. In the latter paper, the counterpart of Proposition 1.3, namely [5, Theorem 3] holds thanks to a delicate decorrelation at specific timescales of the extremal particles of BBM which, in turns, is a consequence of the picture derived in [4]. It is however unreasonable to expect here a similar decorrelation: α-high-particles, namely those with $x_k(t) \geq \Delta_\alpha t$, are unlikely to come from (genealogically) distant ancestors. Indeed, quite the contrary is true: a wealth of random variables contributing to $Z_\alpha(t)$ turn out to be strongly correlated, but these are *washed out*, in the limit of large times, by the exponentially large normalization $\mathbb{E}Z_\alpha(t)$.

Assuming Proposition 1.2 and 1.3, our main theorem steadily follows.

Proof of Theorem 1.1. We use the above telescopic decomposition

$$\frac{Z_\alpha(t)}{\mathbb{E}Z_\alpha(t)} = \frac{\mathbb{E}\left[Z_\alpha(t) \mid \mathcal{F}_r\right]}{\mathbb{E}Z_\alpha(t)} + \frac{Z_\alpha(t) - \mathbb{E}\left[Z_\alpha(t) \mid \mathcal{F}_r\right]}{\mathbb{E}Z_\alpha(t)} =: (A) + (B)\,. \tag{1.18}$$

[2]in a "distant past", as will become clear in the course of the proof.

By Proposition 1.3, the (B)-term on the r.h.s. above vanishes in probability as $t \uparrow \infty$ first, and $r \uparrow \infty$ next. In fact, we may lift this to an almost sure statement in the single limit $t \uparrow \infty$: simply choose $r = r(t) \to \infty$ such that the r.h.s of (1.17) becomes integrable (any choice of the form $r = (\log t)^\epsilon$ with $\epsilon > 1$ will do) and appeal to the Borel-Cantelli Lemma together with standard approximation arguments (for the latter, see e.g. [4]). But for such choice $r = r(t)$, and the almost sure convergence of the McKean's martingale established by Bovier-Hartung, the (A)-term will converge, almost surely, to Y_α. This settles the proof of the SLLN. □

The rest of the paper is devoted to the proofs of Proposition 1.2 and 1.3. Before giving the details, we conclude this section with the following

Conjecture. *A SLLN as in Theorem 1.1 holds true,* mutatis mutandis*, in all models belonging to the BBM-universality class, such as the 2-dim Gaussian free field* [13, 17, 9, 10, 11, 12]*, the 2-dim cover times* [20, 7, 8]*, the characteristic polynomials of random unitary matrices* [1, 19, 29]*, and the extreme values of the Riemann zeta function on the critical line* [2, 28, 3, 6]*. In particular, we expect that an* approximate McKean's martingale *will capture in all such models the almost sure limit of the normalized number of high-points. (What stands behind this wording becomes, of course, model-dependent).*

Acknowledgments. It is a pleasure to thank David Belius for raising the question of high-points in the 2dim GFF, which lead to the writing of this paper.

2. Proofs

2.1. **Some preliminaries, and Onset of McKean's martingale.**

We will make constant use of some classical Gaussian tail-estimates:

Lemma 2.1. *Let $X \sim \mathcal{N}(0, \sigma^2)$ be centered Gaussian random variables. Then*

$$\mathbb{P}[X > a] = (2\pi)^{-1/2} (\sigma/a) \exp\left[-\frac{a^2}{2\sigma^2}\right] \left(1 + O\left(\sigma^2/a^2\right)\right) \qquad (a/\sigma \to \infty), \tag{2.1}$$

with the r.h.s. above without *error term being an upper bound valid for any $a > 0$.*

The following is also elementary.

Lemma 2.2.

$$\mathbb{E}Z_\alpha(t) \sim \left(\Delta_\alpha \sqrt{2\pi}\right)^{-1} \exp\left[\left(1 - \Delta_\alpha^2/2\right) t - \frac{1}{2}\log(t)\right]. \tag{2.2}$$

(Here and throughout, we use $f(t) \sim g(t)$ if the ratio converges, as $t \to \infty$, to one). In order to exploit the strong Markov property of BBM, for t and r as in Proposition 1.2 , we re-label particles at time t according to their ancestor at time r:

$$(x_k(t))_{k \le n(t)} = (x_i(r) + x_{i,j}(t-r))_{i \le n(r), j \le n_i(t-r)}. \tag{2.3}$$

Precisely: $x_i(r)$ is the position of the i-th particle at time r, $n(r)$ is the number of such particles, $n_i(t-r)$ stands for the number of offspring such particle has produced in the timespan $t-r$, and finally $x_{i,j}(t-r)$ denotes the displacement of the j-th offspring of particle i from its starting position $x_i(r)$.

Proof of Proposition 1.2 (Onset of McKean's martingale).

$$\begin{aligned}
\mathbb{E}\left[Z_\alpha(t) \mid \mathcal{F}_r\right] &= \mathbb{E}\left[\sum_{k\le n(t)} \mathbf{1}\left\{x_k(t) \ge \Delta_\alpha t\right\} \Big| \mathcal{F}_r\right] = \\
&= \mathbb{E}\left[\sum_{i\le n(r)} \sum_{j\le n_i(t-r)} \mathbf{1}\left\{x_{i,j}(t-r) \ge \Delta_\alpha(t-r) - (x_i(r)-\Delta_\alpha r)\right\} \Big| \mathcal{F}_r\right] \\
&= \sum_{i\le n(r)} e^{t-r}\mathbb{P}\left[x_1(t-r) \ge \Delta_\alpha(t-r) - (x_i(r)-\Delta_\alpha r) \Big| \mathcal{F}_r\right] \\
&\sim \left(\Delta_\alpha\sqrt{2\pi}\right)^{-1} \exp\left[\left(1-\Delta_\alpha^2/2\right)t - \frac{1}{2}\log(t)\right] Y_\alpha(r), \quad \text{a.s..}
\end{aligned} \tag{2.4}$$

The last step by combining Lemma 2.1 with the fact that $(x_k(r) - \Delta_\alpha r)\,/\,(t-r) \to 0$ almost surely for $r = o(t)$, see e.g. [24] (where, in fact, control at even finer levels is provided). The claim then follows by Lemma 2.2. $\square$

2.2. **Vanishing correlations, via 2nd moment.** We will prove Proposition 1.3 by a 2nd moment estimate. For these computations to go through, we need however to localize paths of contributing particles. (This approach is by now classical in the BBM-field, see for instance [5] for a closely related setting). As for the localization, let $\varepsilon > 0$ and consider

$$Z_\alpha^{>}(t) \equiv \sharp\left\{k \le n(t) : x_k(t) \ge \Delta_\alpha t, \exists s \in [r,t] : x_k(s) > (\Delta_\alpha+\varepsilon)s\right\}, \tag{2.5}$$

This random variable thus counts paths which overshoot at some point (in time) the straight line connecting 0 to $(\Delta_\alpha+\varepsilon)t$. As it turns out, such particles do not contribute, upon $\mathbb{E}[Z_\alpha(t)]$-normalization, to the α-high-points. Here is the precise statement.

Lemma 2.3. *(Paths-localization) For $r = o(t)$, r, t both sufficiently large, it holds:*

$$\mathbb{P}\Big(Z_\alpha^{>}(t) \ge c\mathbb{E}Z_\alpha(t)\Big) \le \frac{1}{c}\exp\left(-r\frac{\varepsilon^2}{4}\right), \tag{2.6}$$

and

$$\mathbb{P}\Big(\mathbb{E}[Z_\alpha^{>}(t)|\mathcal{F}_r] \ge c\mathbb{E}Z_\alpha(t)\Big) \le \frac{1}{c}\exp\left(-r\frac{\varepsilon^2}{4}\right). \tag{2.7}$$

Proof.

$$\mathbb{E}Z_\alpha^{>}(t) = e^t \int_{\Delta_\alpha t}^{\infty} \mathbb{P}(x_1(t)\in dy)\mathbb{P}(\exists s\in[r,t] : x_1(s) > (\Delta_\alpha+\varepsilon)s | x_1(t) = y) \tag{2.8}$$

$$= e^t \int_{\Delta_\alpha t}^{\infty} \mathbb{P}(x_1(t)\in dy)\mathbb{P}(\exists s\in[r,t] : b(s) > (\Delta_\alpha+\varepsilon-\frac{y}{t})s), \tag{2.9}$$

where $b(s) \equiv x_1(s) - \frac{s}{t}x_1(t)$ is a Brownian bridge of length t. Consider the line l from $(0, \varepsilon r/2)$ to $(t, (\Delta_\alpha+\varepsilon/2)t - y)$. One easily checks that $l(s) \le (\Delta_\alpha+\varepsilon-\frac{y}{t})s$ for all

$s \in [r,t]$. Hence the probability involving the Brownian brige is at most

$$\mathbb{P}(\exists s \in [0,t] : b(s) > l(s)) = \exp\left(-2\frac{l(0)l(t)}{t}\right), \tag{2.10}$$

by a well-known formula (see e.g. [31]). Using this, (2.9) is therefore at most

$$\begin{aligned} e^t \int_{\Delta_\alpha t}^{\infty} & \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{y^2}{2t} - \varepsilon r\left(\Delta_\alpha + \frac{\varepsilon}{2} - \frac{y}{t}\right)\right) dy \\ &= \exp\left(t - r\left(\Delta_\alpha \varepsilon + \frac{\varepsilon^2}{2}\right) + \frac{\varepsilon^2 r^2}{2t}\right) \int_{\Delta_\alpha t}^{\infty} \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(y-\varepsilon r)^2}{2t}\right) dy, \end{aligned} \tag{2.11}$$

which is, by Lemma 2.1,

$$\begin{aligned} &\sim \frac{\sqrt{t}}{(\Delta_\alpha t - \varepsilon r)\sqrt{2\pi}} \exp\left(t - r\left(\Delta_\alpha \varepsilon + \frac{\varepsilon^2}{2}\right) + \frac{\varepsilon^2 r^2}{2t} - \frac{(\Delta_\alpha t - \varepsilon r)^2}{2t}\right) \\ &\sim \mathbb{E}[Z_\alpha(t)] \exp\left(-r\frac{\varepsilon^2}{2} + o(r)\right), \end{aligned} \tag{2.12}$$

the second asymptotical equivalence by Lemma 2.2. The claim of Lemma 2.3 thus follows from Markov inequality. □

As mentioned, we will prove Proposition 1.3 by means of a (truncated) second moment computation. The following is the key estimate.

Lemma 2.4. *(Pair-counting) Let $I_{i,j}$ be the the Indicator of the event that the j-th offspring at time t of particle i at time r contributes to $Z_\alpha^{\leq}(t)$, i.e the event*

$$\{x_i(r) + x_{i,j}(t-r) \geq \Delta_\alpha t, \forall s \in [r,t] : x_i(r) + x_{i,j}(s-r) \leq (\Delta_\alpha + \varepsilon)s\} \tag{2.13}$$

for $i \leq n(r)$ and $j \leq n_i(r)$. Then, for any $\alpha \in (0, \sqrt{2})$ there exists ε_α and $\kappa_\alpha, r(\alpha) > 0$ such that

$$\mathbb{E}\left[\sum_{i=1}^{n(r)} \sum_{j \neq j'=1}^{n_i(t-r)} I_{i,j} I_{i,j'}\right] \leq (1+o(1))\mathbb{E}[Z_\alpha(t)]^2 e^{-\kappa_\alpha r}, \qquad \text{as } t \to \infty, \tag{2.14}$$

for $r > r_\alpha$.

Proof. Denoting by φ_γ the density of a centered Gaussian of variance γ, and with x a standard Brownian motion, it holds that

$$\begin{aligned} \frac{1}{2}\mathbb{E}&\left[\sum_{i=1}^{n(r)} \sum_{j \neq j'=1}^{n_i(t-r)} I_{i,j} I_{i,j'}\right] = \\ &= \int_r^t e^{2t-\gamma} \int_{-\infty}^{(\Delta_\alpha+\varepsilon)\gamma} \varphi_\gamma(y) \mathbb{P}(\forall s \in [r,\gamma] : x(s) \leq (\Delta_\alpha + \varepsilon)s | x(\gamma) = y) \times \\ &\quad \times \mathbb{P}\left(y + x(t-\gamma) \geq \Delta_\alpha t, \forall s \in [\gamma, t] : y + x(s-\gamma) \leq (\Delta_\alpha + \varepsilon)s\right)^2 dy d\gamma\,. \end{aligned} \tag{2.15}$$

(See e.g. [30] for a rigorous derivation of a similar "two-point formula"). Dropping the path-constraint appearing in the integrand (yet *keeping* those in the domain of integration), (2.15) is at most

$$
\begin{aligned}
&\int_r^t e^{2t-\gamma} \int_{-\infty}^{(\Delta_\alpha+\varepsilon)\gamma} \varphi_\gamma(y)\mathbb{P}\left(x(t-\gamma) \geq \Delta_\alpha t - y\right)^2 dyd\gamma \\
&= \int_r^t e^{2t-\gamma} \int_{-\infty}^{(\Delta_\alpha+\varepsilon)\gamma} \mathbb{1}_{\{y \geq (\Delta_\alpha-\varepsilon)t\}}(\cdot) dyd\gamma + \int_r^t e^{2t-\gamma} \int_{-\infty}^{(\Delta_\alpha+\varepsilon)\gamma} \mathbb{1}_{\{y<(\Delta_\alpha-\varepsilon)t\}}(\cdot)dyd\gamma\,,
\end{aligned}
\tag{2.16}
$$

by distinguishing whether at time of splitting particles are above (respectively below) a threshold which is slightly below the target.

As for the first scenario, we clearly have

$$
\int_r^t e^{2t-\gamma} \int_{-\infty}^{(\Delta_\alpha+\varepsilon)\gamma} \mathbb{1}_{\{y \geq (\Delta_\alpha-\varepsilon)t\}}(\cdot)dyd\gamma \leq \int_{(1-2\varepsilon/\Delta_\alpha)t}^{t} e^{2t-\gamma} \int_{\Delta_\alpha t-\varepsilon t}^{(\Delta_\alpha+\varepsilon)\gamma} \varphi_\gamma(y)dyd\gamma\,, \tag{2.17}
$$

with the r.h.s. of (2.17) being at most

$$
\frac{4\varepsilon^2 t^2}{\Delta_\alpha\sqrt{2\pi t}} \exp\left((1+2\varepsilon/\Delta_\alpha)t - \frac{(\Delta_\alpha t-\varepsilon t)^2}{2t}\right) \leq \exp\left((1-\Delta_\alpha^2/2+\varepsilon(2/\Delta_\alpha+\Delta_\alpha))t\right), \tag{2.18}
$$

for t large enough. Since $1-\Delta_\alpha^2/2+\varepsilon(2/\Delta_\alpha+\Delta_\alpha) < 2-\Delta_\alpha^2$ for some $\varepsilon=\varepsilon_\alpha$, it follows that (2.18) grows at most exponentially (in t) with rate smaller than $2-\Delta_\alpha^2$, and therefore yields a negligible contribution.

It thus remains to analyse the second scenario in (2.16). By Lemma 2.1 and shortening $\zeta \equiv \min\{(\Delta_\alpha+\varepsilon)\gamma, \Delta_\alpha t-\varepsilon t\}$, we have that

$$
\begin{aligned}
&\int_r^t e^{2t-\gamma} \int_{-\infty}^{(\Delta_\alpha+\varepsilon)\gamma} \mathbb{1}_{\{y<(\Delta_\alpha-\varepsilon)t\}}(\cdot)dyd\gamma \\
&\leq \int_r^t e^{2t-\gamma} \int_{-\infty}^{\zeta} \frac{t-\gamma}{2\pi\sqrt{\gamma}(\Delta_\alpha t-y)^2} \exp\left(-\frac{y^2}{2\gamma} - \frac{(\Delta_\alpha t-y)^2}{(t-\gamma)}\right) dyd\gamma \\
&= \int_r^t e^{2t-\gamma} \int_{-\infty}^{\zeta} \frac{t-\gamma}{2\pi\sqrt{\gamma}(\Delta_\alpha t-y)^2} \exp\left(-\frac{\Delta_\alpha^2 t^2}{t+\gamma} - \frac{(y-\frac{2\gamma\Delta_\alpha t}{t+\gamma})^2}{2\gamma(t-\gamma)/(t+\gamma)}\right) dyd\gamma.
\end{aligned}
\tag{2.19}
$$

Since $\Delta_\alpha t - y \geq \varepsilon t$ on the entire domain of integration, and rearranging, (2.19) is at most

$$
\int_r^t \exp\left(2t-\gamma-\frac{\Delta_\alpha^2 t^2}{t+\gamma}\right) \frac{\sqrt{t^2-\gamma^2}}{\varepsilon^2 t^2 \gamma\sqrt{2\pi}} \mathbb{P}\left(x\left(\frac{\gamma(t-\gamma)}{t+\gamma}\right) < \zeta - \frac{2\gamma\Delta_\alpha t}{t+\gamma}\right) d\gamma. \tag{2.20}
$$

We split (2.20) again into two regions: the first concerns $\gamma > (1-\delta)t$, with $\delta \equiv \frac{3\varepsilon}{\Delta_\alpha+\varepsilon}$. In this case, estimating the probability by one and the remaining integrand by a rough bound on its maximum yields a contribution of at most

$$
\begin{aligned}
&\int_{(1-\delta)t}^{t} \exp\left(2t-\gamma-\frac{\Delta_\alpha^2 t^2}{t+\gamma}\right)\frac{\sqrt{t^2-\gamma^2}}{\varepsilon^2 t^2\gamma\sqrt{2\pi}}\mathbb{P}\left(x\left(\frac{\gamma(t-\gamma)}{t+\gamma}\right)<\zeta-\frac{2\gamma\Delta_\alpha t}{t+\gamma}\right)d\gamma \\
&\qquad \leq \delta t\exp\left((1-\frac{\Delta_\alpha^2}{2}+\delta)t\right)\frac{\sqrt{t^2-\gamma^2}}{\varepsilon^2 t^2\gamma\sqrt{2\pi}} \leq \exp\left((1-\frac{\Delta_\alpha^2}{2}+2\delta)t\right),
\end{aligned}
\tag{2.21}
$$

for t large enough. Therefore, for δ or equivalently ε small enough (depending on α only), this term is also negligeable.

The second case in (2.20) pertains to $\gamma < (1-\delta)t$: in this region, and due to the choice of δ, we have

$$
\zeta = (\Delta_\alpha+\varepsilon)\gamma \Longrightarrow \zeta - \frac{2\gamma\Delta_\alpha t}{t+\gamma} = \gamma\left(\varepsilon - \Delta_\alpha\frac{t-\gamma}{t+\gamma}\right) < 0,
\tag{2.22}
$$

(the last estimate again due to the choice of δ, and for sufficiently small ε depending on α only). This, together with Lemma 2.1, implies that

$$
\begin{aligned}
&\int_{r}^{(1-\delta)t} \exp\left(2t-\gamma-\frac{\Delta_\alpha^2 t^2}{t+\gamma}\right)\frac{\sqrt{t^2-\gamma^2}}{\varepsilon^2 t^2\gamma\sqrt{2\pi}}\mathbb{P}\left(x\left(\frac{\gamma(t-\gamma)}{t+\gamma}\right)<\zeta-\frac{2\gamma\Delta_\alpha t}{t+\gamma}\right)d\gamma \\
&\qquad \leq \int_{r}^{(1-\delta)t} \exp\left(2t-\gamma-\frac{\Delta_\alpha^2 t^2}{t+\gamma}-\frac{\gamma\left(\Delta_\alpha(t-\gamma)-\varepsilon(t+\gamma)\right)^2}{2(t^2-\gamma^2)}\right)\times \\
&\qquad\qquad \times \frac{t^2-\gamma^2}{\left(\Delta_\alpha(t-\gamma)-\varepsilon(t+\gamma)\right)\varepsilon^2 t^2\gamma\sqrt{2\pi\gamma}}d\gamma\,.
\end{aligned}
\tag{2.23}
$$

Using that $-\frac{\gamma\varepsilon^2(t+\gamma)^2}{2(t^2-\gamma^2)} < 0$, and by simple algebra, (2.23) is at most

$$
\begin{aligned}
&\int_{r}^{(1-\delta)t} \exp\left((2-\Delta_\alpha^2)t+(\frac{\Delta_\alpha^2}{2}-1+\Delta_\alpha\varepsilon)\gamma\right)\frac{2}{\Delta_\alpha(1-\frac{3\varepsilon}{\Delta_\alpha\delta})\varepsilon^2\gamma\sqrt{\gamma}2\pi t}d\gamma \\
&\sim \mathbb{E}[Z_\alpha(t)]^2\frac{2\Delta_\alpha}{(1-\frac{3\varepsilon}{\Delta_\alpha\delta})\varepsilon^2}\int_{r}^{(1-\delta)t}\gamma^{-3/2}\exp\left((\frac{\Delta_\alpha^2}{2}-1+\Delta_\alpha\varepsilon)\gamma\right)d\gamma\,,
\end{aligned}
\tag{2.24}
$$

by Lemma 2.2. But for ε sufficiently small $\Delta_\alpha^2/2-1+\Delta_\alpha\varepsilon < 0$, hence the integral in (2.24) vanishes exponentially fast in r. In other words, (2.24) can be bounded by $\mathbb{E}[Z_\alpha(t)]^2\exp(-\kappa_\alpha r)$ for any $\kappa_\alpha < 1-\Delta_\alpha^2/2$ and r large enough. Combining this with the fact that the two error-terms (2.18) and (2.21) are of negligeable size compared to $\mathbb{E}[Z_\alpha(t)]^2$ finishes the proof. □

We can now move to the

Proof of Proposition 1.3. By the paths-localization from Lemma 2.3, and for t large enough,

$$\begin{aligned}\mathbb{P}\left(\left|\frac{Z_\alpha(t)-\mathbb{E}\left[Z_\alpha(t)|\mathcal{F}_r\right]}{\mathbb{E}\left[Z_\alpha(t)\right]}\right|\geq c\right) &\leq \mathbb{P}\left(\left|\frac{Z_\alpha^{\leq}(t)-\mathbb{E}\left[Z_\alpha^{\leq}(t)|\mathcal{F}_r\right]}{\mathbb{E}\left[Z_\alpha(t)\right]}\right|\geq c/2\right)+\frac{8}{c}\exp\left(-r\frac{\varepsilon^2}{4}\right)\\ &\leq \frac{4\mathbb{E}\left[\left(Z_\alpha^{\leq}(t)-\mathbb{E}\left[Z_\alpha^{\leq}(t)|\mathcal{F}_r\right]\right)^2\right]}{c^2\mathbb{E}\left[Z_\alpha(t)\right]^2}+\frac{8}{c}\exp\left(-r\frac{\varepsilon^2}{4}\right),\end{aligned} \tag{2.25}$$

the last step by Markov inequality. We thus need sufficiently good (upper) bounds for

$$\mathbb{E}\left[\left(Z_\alpha^{\leq}(t)-\mathbb{E}\left[Z_\alpha^{\leq}(t)|\mathcal{F}_r\right]\right)^2\right]=\mathbb{E}\left[\mathbb{E}\left[(Z_\alpha^{\leq}(t))^2|\mathcal{F}_r\right]-\mathbb{E}\left[Z_\alpha^{\leq}(t)|\mathcal{F}_r\right]^2\right]. \tag{2.26}$$

Adopting the notation of Lemma 2.4,

$$Z_\alpha^{\leq}(t)=\sum_{i\leq n(r)}\sum_{j\leq n_i(t-r)}I_{i,j}, \tag{2.27}$$

in which case

$$\begin{aligned}&\mathbb{E}\left[(Z_\alpha^{\leq}(t))^2|\mathcal{F}_r\right]-\mathbb{E}\left[Z_\alpha^{\leq}(t)|\mathcal{F}_r\right]^2=\\ &=\sum_{i,i'=1}^{n(r)}\mathbb{E}\left[\left(\sum_{j=1}^{n_i(t-r)}I_{i,j}\right)\left(\sum_{j'=1}^{n_{i'}(t-r)}I_{i',j'}\right)\middle|\mathcal{F}_r\right]\\ &\qquad-\mathbb{E}\left[\sum_{j=1}^{n_i(t-r)}I_{i,j}\middle|\mathcal{F}_r\right]\mathbb{E}\left[\sum_{j'=1}^{n_{i'}(t-r)}I_{i',j'}\middle|\mathcal{F}_r\right].\end{aligned} \tag{2.28}$$

In the above, and for $i\neq i'$, particles (i,j) and (i',j') have branched off before time r: they are thus independent, conditionally upon $\mathcal{F}_r$. This leads to a perfect cancellation of all terms $i\neq i'$, and reduces the above formula to

$$\mathbb{E}\left[(Z_\alpha^{\leq}(t))^2|\mathcal{F}_r\right]-\mathbb{E}\left[Z_\alpha^{\leq}(t)|\mathcal{F}_r\right]^2=\sum_{i=1}^{n(r)}\mathbb{E}\left[\left(\sum_{j=1}^{n_i(t-r)}I_{i,j}\right)^2\middle|\mathcal{F}_r\right]-\mathbb{E}\left[\sum_{j=1}^{n_i(t-r)}I_{i,j}\middle|\mathcal{F}_r\right]^2. \tag{2.29}$$

Dropping the second term, and taking expectations, we thus obtain

$$\mathbb{E}\left[\left(Z_\alpha^{\leq}(t)-\mathbb{E}\left[Z_\alpha^{\leq}(t)|\mathcal{F}_r\right]\right)^2\right]\leq\mathbb{E}\left[\sum_{i=1}^{n(r)}\sum_{j,j'=1}^{n_i(t-r)}I_{i,j}I_{i,j'}\right]. \tag{2.30}$$

Collecting the terms $j=j'$ yields $Z_\alpha^{\leq}(t)$, while all other terms sum over all unordered pairs of particles that have split after time r. Choosing $\varepsilon=\varepsilon_\alpha$ small enough and $r=o(t)$ large enough, by Lemma 2.4 there exists $\kappa_\alpha>0$ such that

$$(2.30)\leq\mathbb{E}[Z_\alpha^{\leq}(t)]+(1+o(1))\mathbb{E}[Z_\alpha(t)]^2e^{-\kappa_\alpha r}=(1+o(1))\mathbb{E}[Z_\alpha(t)]^2e^{-\kappa_\alpha r}. \tag{2.31}$$

The claim thus follows by plugging (2.31) into (2.25). □

References

[1] Arguin, Louis-Pierre, David Belius, and Paul Bourgade. *Maximum of the characteristic polynomial of random unitary matrices.* Communications in Mathematical Physics 349.2 (2017): 703-751.

[2] Arguin, Louis-Pierre, David Belius, and Adam J. Harper. *Maxima of a randomized Riemann zeta function, and branching random walks.* The Annals of Applied Probability 27.1 (2017): 178-215.

[3] Arguin, Louis-Pierre, David Belius, Paul Bourgade, Maksym Radziwiłł and Kannan Soundararajan. *Maximum of the Riemann zeta function on a short interval of the critical line.* arXiv preprint: 1612.08575.

[4] Arguin, Louis-Pierre, Anton Bovier, and Nicola Kistler. *Genealogy of extremal particles of branching Brownian motion*, Communications on Pure and Applied Mathematics, Vol. LXIV (2011):, 16471676.

[5] Arguin, Louis-Pierre, Anton Bovier, and Nicola Kistler. *An ergodic theorem for the frontier of branching Brownian motion.* Electronic Journal of Probability 18, No 53 (2013): 1-25.

[6] Arguin, Louis-Pierre, and Warren Tai. *Is the Riemann zeta function in a short interval a 1-RSB spin glass?.* arXiv preprint:1706.08462 (2017).

[7] Belius, David, and Nicola Kistler. *The subleading order of two dimensional cover times.* Probability Theory and Related Fields 167.1-2 (2017): 461-552.

[8] Belius, David, Jay Rosen, and Ofer Zeitouni. *Tightness for the Cover Time of* $\mathbf{S}^2$. arXiv preprint:1711.02845 (2017).

[9] Biskup, Marek, Jian Ding, and Subhajit Goswami. *Return probability and recurrence for the random walk driven by two-dimensional Gaussian free field.* arXiv preprint:1611.03901 (2016).

[10] Biskup, Marek, and Oren Louidor. *Extreme local extrema of two-dimensional discrete Gaussian free field.* Communications in Mathematical Physics 345.1 (2016): 271-304.

[11] Biskup, Marek, and Oren Louidor. *Full extremal process, cluster law and freezing for two-dimensional discrete Gaussian Free Field.* arXiv preprint: 1606.00510 (2016).

[12] Biskup, Marek, and Oren Louidor. *On intermediate level sets of two-dimensional discrete gaussian free field.* arXiv preprint:1612.01424 (2016).

[13] Bolthausen, Erwin, Jean-Dominique Deuschel, and Giambattista Giacomin. *Entropic repulsion and the maximum of the two-dimensional harmonic.* The Annals of Probability 29.4 (2001): 1670-1692.

[14] Bovier, Anton. *Gaussian processes on trees: From spin glasses to branching Brownian motion.* Cambridge Studies in Advanced Mathematics Vol. 163, Cambridge University Press (2016).

[15] Bovier, Anton, and Lisa Hartung, *The extremal process of two-speed branching Brownian motion*, Electronic Journal of Probability 19, No. 18: 1-28 (2014).

[16] Bramson, Maury. *Maximal displacement of branching Brownian motion*, Comm. Pure Appl. Math. 31 (1978): 531-581.

[17] Bramson, Maury, Jian Ding, and Ofer Zeitouni. *Convergence in Law of the Maximum of the Two-Dimensional Discrete Gaussian Free Field.* Communications on Pure and Applied Mathematics 69.1 (2016): 62-123.

[18] Bramson, Maury, and Ofer Zeitouni. *Tightness of the recentered maximum of the two-dimensional discrete Gaussian free field.* Communications on Pure and Applied Mathematics 65.1 (2012): 1-20.

[19] Chhaibi, Reda, Joseph Najnudel, and Thomas Madaule. *On the maximum of the CβE field.* arXiv preprint:1607.00243 (2016).

[20] Dembo, Amir, Yuval Peres, Jay Rosen and Ofer Zeitouni. *Cover times for Brownian motion and random walks in two dimensions.* Annals of Mathematics (2004): 433-464.

[21] Derrida, Bernard. *Random-energy model: An exactly solvable model of disordered systems.* Physical Review B 24.5 (1981): 2613.

[22] Derrida, Bernard, and Herbert Spohn. *Polymers on disordered trees, spin glasses, and traveling waves.* Journal of Statistical Physics 51.5 (1988): 817-840.

[23] Fang, Ming, and Ofer Zeitouni. *Slowdown for time inhomogeneous branching Brownian motion.* Journal of Statistical Physics 149 (2012).1: 1-9.

[24] Hu, Yueyun, and Zhan Shi. *Minimal position and critical martingale convergence in branching random walks, and directed polymers on disordered trees.* The Annals of Probability 37.2 (2009): 742-789.

[25] A. Kolmogorov, I. Petrovsky, and 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 (1937): 1-25.

[26] Lalley Steven and Thomas Sellke, *A conditional limit theorem for the frontier of a branching Brownian motion*, Ann. Probab. 15 (1987): 1052-1061.

[27] McKean, Henry P. *Application of Brownian Motion to the equation of Kolmogorov-Petrovskii-Piskunov*, Comm. Pure. Appl. Math. 28 (1976): 323-331.

[28] Najnudel, Joseph. *On the extreme values of the Riemann zeta function on random intervals of the critical line.* arXiv preprint: 1611.05562 (2016).

[29] Paquette, Elliot, and Ofer Zeitouni. *The maximum of the CUE field.* International Mathematics Research Notices (2017): rnx033.

[30] Sawyer, Stanley. *Branching diffusion processes in population genetics.* Advances in Applied Probability 8.4 (1976): 659-689.

[31] Scheike, Thomas H. *A boundary-crossing result for Brownian motion.* Journal of Applied Probability 29.2 (1992): 448-453.

Constantin Glenz, J.W. Goethe-Universität Frankfurt, Germany.
Email address: constantin.glenz@gmail.com

Nicola Kistler, J.W. Goethe-Universität Frankfurt, Germany.
Email address: kistler@math.uni-frankfurt.de

Marius A. Schmidt, J.W. Goethe-Universität Frankfurt, Germany.
Email address: mschmidt@math.uni-frankfurt.de

A SIMPLE PROOF OF THE DPRZ-THEOREM FOR 2d COVER TIMES.

Marius A. Schmidt
J.W. Goethe-Universität Frankfurt, Germany
mschmidt@math.uni-frankfurt.de

May 24, 2018

The ε*-cover time* of the two dimensional unit torus $\mathbb{T}_2$ by Brownian motion (BM) is the time for the process to come within distance $\varepsilon > 0$ from any point. Denoting by $T_\varepsilon(x)$ the first time BM hits the ε-ball centered in $x \in \mathbb{T}_2$, the ε-cover time is thus given by

$$T_\varepsilon \equiv \sup_{x \in \mathbb{T}_2} T_\varepsilon(x). \tag{1}$$

The purpose of these short notes is to provide a concise proof of a celebrated theorem by Dembo, Peres, Rosen and Zeitouni, DPRZ for short, which settles the leading order in the small-ε regime:

Theorem 1. *(The DPRZ-Theorem, [3]) Almost surely,*

$$\lim_{\varepsilon \downarrow 0} \frac{T_\varepsilon}{(\ln \varepsilon)^2} = \frac{2}{\pi}. \tag{2}$$

A key idea in the DPRZ-approach is to relate hitting times of ε-balls on $\mathbb{T}_2$ to excursion counts between circles of mesoscopic sizes around these balls [6]; the DPRZ-proof of the theorem goes then through an involved multiscale analysis in the form of a second moment computation with truncation. We take here a similar point of view but with a number of twists which altogether lead to a considerable streamlining of the arguments. In particular, we implement the multiscale refinement of the second moment method emerged in the recent studies of Derrida's GREM and branching Brownian motion [5]. This tool brings to the fore the *true* process of covering [1] with the help of minimal infrastructure only; it also efficiently replaces the delicate tracking of points which DPRZ refer to as 'n-successful', and requires the use of finitely many scales only. All these features simplify substantially the proof of the DPRZ-theorem.

We believe the route taken here also considerably streamlines the deep DPRZ-results on late and thin/thick points of BM [2], and, what is perhaps more, it will be useful in the study of the *finer* properties. In fact, our approach carries over, *mutatis mutandis*, to these issues as well: when backed with [1], the present notes suggest that in order to address lower order corrections, one "simply" needs to increase the number of scales.

These notes are self-contained. Although, as mentioned, some key insights are taken from [3], no knowledge of the latter is assumed and detailed proofs to all statements are given.

1 The (new) road to the DPRZ-Theorem

We identify the unit torus $\mathbb{T}_2$ with $[0,1) \times [0,1) \subset \mathbb{R}^2$, endowed with the metric

$$d_{\mathbb{T}_2}(x,y) = \min\{||x - y + (e_1, e_2)|| : e_1, e_2 \in \{-1, 0, 1\}\}.$$

We construct BM on $\mathbb{T}_2$ by $W_t \equiv \left(\hat{W}_1(t) \mod 1, \hat{W}_2(t) \mod 1\right)$, where $\hat{W}$ is standard BM on $\mathbb{R}^2$.

By monotonicity of T_ε and Borel-Cantelli Lemma, the DPRZ-Theorem steadily follows from

Theorem 2. *For $\delta > 0$ small enough there exist constants $c(\delta), c'(\delta) > 0$ such that the following bounds hold for any $0 < \varepsilon < c'(\delta)$:*

1) (upper bound)

$$\mathbb{P}\left(T_\varepsilon > (1+\delta)\frac{2}{\pi}(\ln\varepsilon)^2\right) \leq \varepsilon^{c(\delta)}, \tag{3}$$

2) (lower bound)

$$\mathbb{P}\left(T_\varepsilon < (1-\delta)\frac{2}{\pi}(\ln\varepsilon)^2\right) \leq \varepsilon^{c(\delta)}. \tag{4}$$

Theorem 2 will be proved by relating the natural timescale of the covering process to the excursion-counts of an embedded random walk, and a multiscale analysis of the latter which exploits some underlying, approximate hierarchical structure in the spirit of [1].

1.1 Scales, embedded random walks and excursion-counts

For $R \in \left(0, \frac{1}{2}\right)$ and $K \geq 1$ we consider *scales* $i = 0, 1, .., K$ and associate to each such scale a radius

$$r_i \equiv R\left(\frac{\varepsilon}{R}\right)^{i/K}. \tag{5}$$

BM started on ∂B_{r_i} hits $\partial B_{r_{i+1}}$ before $\partial B_{r_{i-1}}$ with probability $1/2$: by the strong Markovianity and rotational invariance, it follows that the process obtained by tracking the order in which BM visits the scales (with respect to one fixed center point and not counting multiple consecutive hits to the same scale) during one excursion from scale 1 to scale 0 is a simple random walk (SRW) started at 1 and stopped in 0. Keeping track of all BM-excursions up to some time thus yields a collection of independent SRW-excursions from 1 to 0. (The evolution of the SRW-excursions can

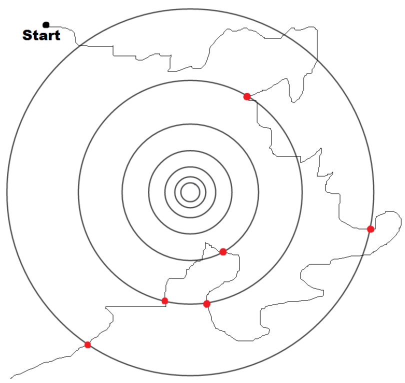

Figure 1: Reading off the SRW excursions $1 \to 0$ and $1 \to 2 \to 1 \to 0$

be unambiguously read off the BM-path, see Figure 1). For $x \in \mathbb{T}_2$, we set

$$D_n(x) \equiv \text{time at which } W \text{ completes the } n\text{-th excursion from } \partial B_{r_1}(x) \text{ to } B^c_{r_0}(x). \tag{6}$$

Proposition 1. *(Concentration of excursion-counts) For $\delta, R \in \left(0, \frac{1}{2}\right)$ and $x \in \mathbb{T}_2$, it holds*

$$\mathbb{P}\left(D_n(x) \geq (1+\delta)n\frac{1}{\pi}\ln\frac{r_0}{r_1}\right) \leq \exp\left(-n\left(\frac{\delta^2}{8} + o_{r_1}(1)\right)\right) \tag{7}$$

$$\mathbb{P}\left(D_n(x) \leq (1-\delta)n\frac{1}{\pi}\ln\frac{r_0}{r_1}\right) \leq \exp\left(-n\left(\frac{\delta^2}{4} + o_{r_1}(1)\right)\right) \tag{8}$$

for all $n \in \mathbb{N}$ as $r_1 \to 0$.

Proposition 1 will bear fruits when combined with the following

Proposition 2. *(First moment of hitting times) There exists an universal constant* $C>0$*, such that*

$$\left|\mathbb{E}_y[\tau_{B_r(x)}]-\frac{1}{\pi}\ln\frac{d_{\mathbb{T}_2}(x,y)}{r}\right|\leq C \tag{9}$$

for all $x\in\mathbb{T}_2$*,* $r>0$ *and* $y\in\mathbb{T}_2\setminus B_r(x)$*. Also*

$$\mathbb{E}_y\left[\tau_{B_r^c(x)}\right]=\frac{r^2-d_{\mathbb{T}_2}(x,y)^2}{2} \tag{10}$$

for all $x\in\mathbb{T}_2$*,* $r\in\left(0,\frac{1}{2}\right)$ *and* $y\in B_r(x)$*.*

Propositions 1 and 2 make precise the intuition that $D_n(x)\approx n\mathbb{E}_{B_{r_0}}[\tau_{B_{r_1}}]$, allowing in particular to switch from the natural timescale to the excursion-counts. Armed with the above results, which will be proved in Section 2.1, we discuss the main steps behind Theorem 2. The upper bound is easy: we address that first.

Here and below, L_ε will denote the square lattice of mesh size $\lceil\varepsilon^{-1}\rceil^{-1}$. Remark that $|L_\varepsilon|\approx\varepsilon^{-2}$.

1.2 The upper bound

We will show that, with overwhelming probability, at time

$$t_\varepsilon(\delta)\equiv(1+\delta)\frac{2}{\pi}(\ln\varepsilon)^2, \tag{11}$$

each ε-ball with center on L_ε has been hit by BM and extend this to the entire torus thereafter.

Lemma 1. *For* $\delta>0$ *small enough there exist constants* $c,c'>0$ *depending on* δ *only such that*

$$\mathbb{P}(\exists x\in L_\varepsilon \text{ such that } T_\varepsilon(x)>t_\varepsilon(\delta))\leq\varepsilon^c \tag{12}$$

holds for all $0<\varepsilon<c'$*.*

Proof. We set

$$n_\varepsilon(\delta)=-(1+\delta/2)2K\ln(\varepsilon), \tag{13}$$

which is slightly larger then the typical amount of excursions at time $t_\varepsilon(\delta)$. For an ε-ball to be avoided up to some time: either *i)* BM needs to complete less than $n_\varepsilon(\delta)$ excursions from scale 1 to scale 0 in that time or *ii)* scale K, corresponding to the ε-ball, has to be avoided for at least $n_\varepsilon(\delta)$ many excursions. Therefore setting

$$\mathscr{T}(x)\equiv\text{number of the first excursion from } \partial B_{r_1}(x) \text{ to } B_{r_0}^c(x) \text{ that hits } B_{r_K}(x).$$

we have

$$\mathbb{P}(\exists x\in L_\varepsilon \text{ s.t. } T_\varepsilon(x)>t_\varepsilon(\delta))\leq\mathbb{P}\left(\exists x\in L_\varepsilon \text{ s.t. } \mathscr{T}(x)>n_\varepsilon(\delta) \text{ or } D_{n_\varepsilon(\delta)}(x)\geq t_\varepsilon(\delta)\right). \tag{14}$$

By Markov inequality and union bound

$$(14)\leq\sum_{x\in L_\varepsilon}\mathbb{P}(\mathscr{T}(x)>n_\varepsilon(\delta))+\mathbb{P}\left(D_{n_\varepsilon(\delta)}(x)\geq t_\varepsilon(\delta)\right). \tag{15}$$

The probability that $n_\varepsilon(\delta)$ independent excursions of a SRW starting in 1 all hit 0 before K is given by $(1-1/K)^{n_\varepsilon(\delta)}$, while the second probability on the r.h.s of (15) is estimated by Proposition 1. This shows that the above is at most

$$|L_\varepsilon|\left[\left(1-\frac{1}{K}\right)^{n_\varepsilon(\delta)}+\exp\left(-\frac{\delta^2}{72}n_\varepsilon(\delta)\right)\right]\leq\varepsilon^\delta(1+o_\varepsilon(1)), \tag{16}$$

for K large enough, the last inequality since $1-\frac{1}{K}\leq e^{-1/K}$, and $|L_\varepsilon|\approx\varepsilon^{-2}$. ∎

Coming back to the upper bound in Theorem 2,

$$\mathbb{P}(T_\varepsilon > t_\varepsilon(\delta)) = \mathbb{P}(\exists x \in \mathbb{T}_2 : T_\varepsilon(x) > t_\varepsilon(\delta)) \le \mathbb{P}\left(\exists x \in L_{\varepsilon/10} : T_{\varepsilon/10}(x) > t_\varepsilon(\delta)\right), \tag{17}$$

the last step using that any ε-ball contains a ball of radius $\varepsilon/10$ with center in $L_{\varepsilon/10}$. For $\varepsilon > 0$ small enough depending on δ we have $t_\varepsilon(\delta) \ge t_{\varepsilon/10}(\delta/2)$, therefore it holds that

$$(17) \le \mathbb{P}\left(\exists x \in L_{\varepsilon/10} : T_{\varepsilon/10}(x) > t_{\varepsilon/10}(\delta/2)\right). \tag{18}$$

Applying Lemma 1 with $\varepsilon/10$ for ε and $\delta/2$ yields the upper bound in Theorem 2.

1.3 The lower bound

We show that with overwhelming probability there exists $x \in \mathbb{T}_2$ with avoided ε-ball at time

$$\mathfrak{t} = \mathfrak{t}(\varepsilon, \delta) \equiv (1-\delta)^4 \frac{2}{\pi} (\ln \varepsilon)^2 . \tag{19}$$

Theorem 2 will then follow immediately by considering[1] $\hat{\delta} \equiv 1 - (1-\delta)^4$. Set

$$\mathfrak{n}(j) = \mathfrak{n}(j; \varepsilon, \delta, K) \equiv -2K(1-\delta)^j \ln \varepsilon, \qquad (j \in \mathbb{N}). \tag{20}$$

With $\tau_r \equiv \tau_r(x)$ denoting the first time BM hits the r-ball around $x \in \mathbb{T}_2$, we define the events

$$\mathscr{R} \equiv \bigcap_{x \in L_\varepsilon} \left\{ D_{\mathfrak{n}(3)}(x) > \mathfrak{t} \right\} \quad \text{and} \tag{21}$$

$$\mathscr{R}^x \equiv \{\tau_{r_1} < \tau_{r_K}\} \cap \{\text{At most } \mathfrak{n}(2) \text{ excursions } \lceil \delta k \rceil \to \lceil \delta k \rceil - 1 \text{ during first } \mathfrak{n}(3) \text{ excursions } 1 \to 0\}. \tag{22}$$

For $n \in \mathbb{N}$ and $l \in \{1, .., K-1\}$, let

$$\begin{aligned} \mathscr{N}_l^x(n) \equiv \;& \text{number of excursions of } W \text{ from } \partial B_{r_l}(x) \text{ to } \partial B_{r_{l+1}}(x) \text{ within the} \\ & \text{first } n \text{ excursions from } \partial B_{r_l}(x) \text{ to } \partial B_{r_{l-1}}(x) \text{ after time } \tau_{r_1}. \end{aligned} \tag{23}$$

For $x \in \mathbb{T}_2$ define the events

$$A^x \equiv \bigcap_{l = \lceil \delta K \rceil}^{K-1} A_l^x, \quad \text{where} \quad A_l^x \equiv \left\{ \mathscr{N}_l^x \left(n \left(1 - \frac{l}{K}\right)^2 \right) \le n \left(1 - \frac{l+1}{K}\right)^2 \right\}. \tag{24}$$

The events $A, \mathscr{R}$ are motivated by the following observations. First, it can be checked via Doob's h-transform that the expected number of excursions from l to $l+1$ performed by a SRW started at 1 and stopped at 0 and conditioned not to hit K, is approximately $[1-(l+1)/K]^2$. The events A^x thus describe the natural avoidance strategy of scale K by n independent such SRW, which is in turn equivalent to specifying the avoidance strategy of an ε-ball. Second, we claim that

$$\mathscr{R} \cap \mathscr{R}^x \cap A^x \subset \{B_\varepsilon(x) \text{ is not hit up to time } t\}. \tag{25}$$

Remark in fact that on $\mathscr{R}^x$, the ball $B_\varepsilon(x)$ is not hit before $\partial B_{r_1}(x)$, hence the ε-ball can only be hit in an excursion from B_{r_1} to B_{r_0}. $\mathscr{R}$ ensures that there are at most $\mathfrak{n}(3)$-excursions before time $\mathfrak{t}$. Therefore, on $\mathscr{R}^x \cap \mathscr{R}$, there are at most $\mathfrak{n}(2)$ excursions from scale $\lceil \delta K \rceil \to \lceil \delta K \rceil - 1$ at time $\mathfrak{t}$. But on A^x, none of these excursions reaches scale K, hence the ε-ball is not hit, and (25) holds.

In light of (25), and in view of the lower bound in Theorem 2, estimates on the probabilities of the $\mathscr{R}, A$-events are needed. This information is provided by Lemma 2 and 3 below, whose proofs are deferred to Section 2.2. Concerning the $\mathscr{R}$-event we state

[1] This is notationally convenient, but holds no deeper meaning.

Lemma 2. *For all $\delta > 0$ and large enough $K = K(\delta) \in \mathbb{N}$ there exist constants $\kappa, \kappa' > 0$ depending on δ, K only such that*

$$\inf_{x \in L_\varepsilon \setminus B_{r_1}(W_0), \varepsilon \in (0,\kappa')} \mathbb{P}(\mathscr{R}^x), \mathbb{P}(\mathscr{R}) \geq 1 - \varepsilon^\kappa. \tag{26}$$

Concerning the A-events,

Lemma 3. *(One-point estimates) For K large, $\varepsilon > 0$ small enough (depending on δ)*

$$\varepsilon^{2-1.99\delta} \leq \mathbb{P}(A^x) \leq \varepsilon^{2-2.01\delta}, \tag{27}$$

Coming back to the lower bound, restricting to the set $L_\varepsilon^* \equiv L_\varepsilon \setminus B_{r_1}(W_0)$ yields that

$$\begin{aligned} \mathbb{P}\left(\sup_{x \in \mathbb{T}_2} T_\varepsilon(x) > \mathfrak{t}\right) &\geq \mathbb{P}(\exists x \in L_\varepsilon^* \text{ such that } B_\varepsilon(x) \text{ is not hit up to time } \mathfrak{t}) \\ &\overset{(25)}{\geq} \mathbb{P}(\mathscr{R} \text{ and } \exists x \in L_\varepsilon^* \text{ such that } \mathscr{R}^x \cap A^x) \\ &\geq \frac{\mathbb{E}[\#\{x \in L_\varepsilon^* : \mathscr{R}^x \cap A^x\}]^2}{\mathbb{E}\left[\#\{x \in L_\varepsilon^* : \mathscr{R}^x \cap A^x\}^2\right]} - \mathbb{P}(\mathscr{R}^c), \end{aligned} \tag{28}$$

by Paley-Zygmund inequality. Rotational invariance and strong Markovianity imply that $\mathscr{R}^x$ and A^x are independent, hence the above is *at least*

$$\left[\sum_{x \in L_\varepsilon^*} \mathbb{P}(\mathscr{R}^x)\mathbb{P}(A^x)\right]^2 \Big/ \left[\sum_{x,y \in L_\varepsilon^*} \mathbb{P}(A^x \cap A^y)\right] - \mathbb{P}(\mathscr{R}^c). \tag{29}$$

We now analyse the denominator. First, remark that for $d_{\mathbb{T}_2}(x,y) > 2r_{\lceil \delta K \rceil - 1}$, the A-events decouple: in fact, they are rotationally invariant and depend on disjoint excursions, hence the strong Markov property yields $\mathbb{P}(A^x \cap A^y) = \mathbb{P}(A^x)\mathbb{P}(A^y)$. Shortening

$$\mathscr{A} \equiv \sum_{x \in L_\varepsilon^*} \mathbb{P}(A^x), \quad \mathscr{B} \equiv \sum_{x,y \in L_\varepsilon} \mathbb{1}_{\{d_{\mathbb{T}_2}(x,y) \leq 2r_{\lceil \delta K \rceil - 1}\}} \mathbb{P}(A^x \cap A^y),$$

by Lemma 2 and the exact decoupling we thus have that

$$\begin{aligned} (29) &\geq (1-\varepsilon^\kappa)^2 \frac{\mathscr{A}^2}{\mathscr{A}^2 + \mathscr{B}} - \varepsilon^\kappa \geq (1-\varepsilon^\kappa)^2 \left(1 - \frac{\mathscr{B}}{\mathscr{A}^2}\right) - \varepsilon^\kappa \\ &\geq (1-\varepsilon^\kappa)^2 \left(1 - \frac{\mathscr{B}}{\varepsilon^{-3.96\delta}}\right) - \varepsilon^\kappa, \end{aligned} \tag{30}$$

the last step by Lemma 3 and using that $|L_\varepsilon| \geq \varepsilon^{-2+0.01\delta}$. It thus remains to analyze the $\mathscr{B}$-term: by regrouping terms according to the distance,

$$\mathscr{B} \leq \sum_{i = \lceil \delta K \rceil - 2}^{K} \sum_{x,y \in L_\varepsilon} \mathbb{1}_{\{d_{\mathbb{T}_2}(x,y) \in [r_{i+1}, r_i]\}} \mathbb{P}(A^x \cap A^y). \tag{31}$$

To get a handle on the two-points probabilities appearing in (31), we follow the recipe from [5, Sec. 3.1.1 p. 97-98], exploiting the approximate hierarchical structure which underlies the excursion-counts, and which is best explained with the help of a picture, see Figure 2 below. First, the circles associated to x, y on small scales i (left) are almost identical and so are the excursion counts; this suggests that $A_i^x \cap A_i^y$ is well represented by A_i^x alone. Dropping one of the events is an estimate by worst case scenario known in this context as "REM approximation". For larger i (middle) this approximation is not sharp, but only a single scale can fall into this case as we can choose ε arbitrarily small for given K. Choosing K large makes the influence of a single

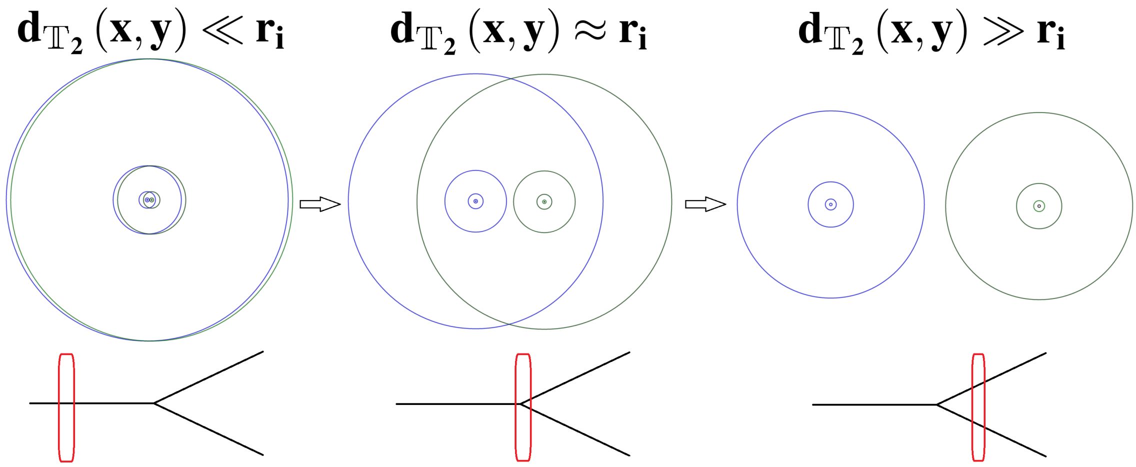

Figure 2: Common branch on small scales (left) and decoupling on large scales (right).

scale comparatively small. For i large (right), balls are disjoint, which by rotational invariance and strong Markovianity yields independent excursion counts. Such approximate tree-structure of excursion counts is summarized in the lower picture, with the red box corresponding to the scale at hand. By these considerations, for $i \geq \lceil \delta K \rceil - 2$ and $d_{\mathbb{T}_2}(x,y) \in [r_{i+1}, r_i]$, we write

$$\begin{aligned}
\mathbb{P}(A^x \cap A^y) &= \mathbb{P}\left(\bigcap_{l=\lceil \delta K \rceil}^{K-1} A_l^x \cap \bigcap_{l=\lceil \delta K \rceil}^{K-1} A_l^y \right) \\
&\leq \mathbb{P}\left(\bigcap_{l=\lceil \delta K \rceil, l \neq i, i+1}^{K-1} A_l^x \cap \bigcap_{l=i+1}^{K-1} A_l^y \right) \qquad (\text{"REM approximation"}) \\
&= \prod_{l=\lceil \delta K \rceil, l \neq i, i+1}^{K-1} \mathbb{P}(A_l^x) \prod_{l=i+1}^{K-1} \mathbb{P}(A_l^y) \qquad (\text{exact decoupling}) \\
&\leq \varepsilon^{4-2.01\delta - 2\frac{i+1}{K}} \qquad (\text{Lemma 3 / one-point estimates}).
\end{aligned} \tag{32}$$

There are at most $2\varepsilon^{-4}\pi r_i^2$ pairs of points on L_ε with distance at most r_i: using that $r_i \leq \varepsilon^{i/K}$, and (32) in (31) we get

$$\mathscr{B} \leq \sum_{i=\lceil \delta K \rceil - 2}^{K} 2\pi \varepsilon^{-2.01\delta - \frac{4}{K}} \leq \varepsilon^{-2.02\delta}. \tag{33}$$

Applying this estimate to (30) and putting $\hat{\delta} \equiv 1 - (1-\delta)^4$ we therefore see that

$$\mathbb{P}\left(\sup_{x \in \mathbb{T}_2} T_\varepsilon(x) > \left(1 - \hat{\delta}\right) \frac{2}{\pi} (\ln \varepsilon)^2 \right) \geq 1 - \varepsilon^{\hat{c}}, \tag{34}$$

for $\hat{c} \equiv \frac{1}{2} \min\{\kappa, 1.94\delta\}$, settling the lower bound of Theorem 2.

2 Proofs

2.1 Hitting times and excursion-counts

The study of hitting times for BM is closely related to Green's functions. Estimates on the torus have however proofs which are either opaque or hard to find: we include here an elementary treatment based on Fourier analysis for the reader's convenience.

Lemma 4. *The function*

$$F(x,y) \equiv G_x(y) - \frac{1}{2\pi}\ln d_{\mathbb{T}_2}(x,y), \quad \textit{where} \quad G_x(y) \equiv -\sum_{p\in 2\pi\mathbb{Z}^2\setminus\{0\}} \frac{1}{|p|^2} e^{ip(x-y)} \tag{35}$$

is bounded on $\mathbb{T}_2^2 \setminus \{(x,x) : x \in \mathbb{T}_2\}$.

Proof. It suffices to consider y in a small neighborhood of x, as otherwise the result is trivial. So let $z \equiv x - y$ and assume that $2|z_1| \geq |z|$ (swapping coordinates otherwise). We have

$$\left| \sum_{\substack{p\in 2\pi\mathbb{Z}^2\setminus\{0\} \\ |p|>|z|^{-1}}} \frac{1}{|p|^2} e^{ipz} \right| = \left| \sum_{\substack{p\in 2\pi\mathbb{Z}^2\setminus\{0\} \\ |p|>|z|^{-1}}} \frac{1}{1-e^{i2\pi z_1}} \frac{1}{|p|^2} \left(e^{ipz} - e^{i(p+(2\pi,0))z} \right) \right|. \tag{36}$$

Shifting the difference from the exponential to $|p|^{-2}$ by collecting terms with the same exponent, and by the triangle inequality, one obtains boundedness uniformly over $z \neq 0$ in a small enough neighborhood of 0. The extra terms due to the boundary of the summation domain are easily shown to be bounded. By combining the summand p and $-p$ we see that sums of this form are real valued. Therefore

$$\sum_{\substack{p\in 2\pi\mathbb{Z}^2\setminus\{0\} \\ |p|\leq|z|^{-1}}} \frac{1}{|p|^2} e^{ipz} = \sum_{\substack{p\in 2\pi\mathbb{Z}^2\setminus\{0\} \\ |p|\leq|z|^{-1}}} \frac{1}{|p|^2} \cos(pz). \tag{37}$$

Since $|pz| \leq 1$ for all summands contained in this sum we can estimate $\cos(x) \leq 1 - x^2/4$. Hence

$$\left| G_x(y) - \sum_{\substack{p\in 2\pi\mathbb{Z}^2\setminus\{0\} \\ |p|\leq|z|^{-1}}} \frac{1}{|p|^2} \right| \tag{38}$$

is uniformly bounded for y in a small neighborhood of x. The claim of Lemma 4 then follows by rearranging summands into groups $C_j \equiv \{p \in 2\pi\mathbb{Z}^2 \setminus \{0\} : |p|^2 \in \left((j-1)^3, j^3\right]\}$, estimating $|p|^{-2}$ by best/worst case scenario within each group, and using that $|C_j| = \frac{3}{4\pi} j^2 + O(j^{3/2})$. □

Proof of Proposition 2: first moment of hitting times. Let $\mu(y) \equiv \mathbb{E}_y\left[\tau_{B_r(x)}\right]$. For Δ the Laplacian with periodic boundary condition on $\mathbb{T}_2$ we have Poisson's equation $\Delta\mu = -2$ on $\mathbb{T}_2 \setminus B_r(x)$ with $\mu = 0$ on $B_r(x)$. Plainly,

$$G_x(y) \equiv -\sum_{p\in 2\pi\mathbb{Z}^2\setminus\{0\}} \frac{1}{|p|^2} e^{ip(x-y)} \tag{39}$$

is a Green function, i.e. solution of $\Delta G_x = 1 - \delta_x$ on the torus. In particular, $\mu + 2G_x$ is harmonic on $\mathbb{T}_2 \setminus B_r(x)$. By the maximum principle, and since $\mu \equiv 0$ on $\partial B_r(x)$,

$$2 \inf_{z\in\partial B_r(x)} G_x(z) \leq \mu(y) + 2G_x(y) \leq 2 \sup_{z\in\partial B_r(x)} G_x(z) \tag{40}$$

holds. It follows from Lemma 4 that $\mu(y) - \frac{1}{\pi}\ln[d_{\mathbb{T}_2}(x,y)/r]$ is bounded, and the first claim (9) is proved. The second claim (10) is elementary as we can identify the ball on $\mathbb{T}_2$ with the ball in $\mathbb{R}^2$ and exploit rotational invariance to solve Poisson's equation explicitly. □

Proof of Proposition 1: concentration of excursion-counts. By Kac's moment formula [4],

$$\mathbb{E}_x\left[\tau_A^i\right] \leq i! \sup_{x\in\mathbb{T}} \mathbb{E}_x\left[\tau_A\right]^i, \qquad A \subset \mathbb{T} \text{ closed}. \tag{41}$$

By monotone convergence, Taylor-expanding the exponential function, and by the above estimate,

$$\mathbb{E}_x\left[e^{\theta\tau_A}\right] \le 1+\theta\mathbb{E}_x[\tau_A]+\sum_{i=2}^{\infty}\left(\theta\sup_{x\in\mathbb{T}}\mathbb{E}_x[\tau_A]\right)^i \le \exp\left(\theta\mathbb{E}_x[\tau_A]+2\theta^2\sup_{x\in\mathbb{T}}\mathbb{E}_x[\tau_A]^2\right) \tag{42}$$

for $0<\theta<\frac{1}{2}\left(\sup_{x\in\mathbb{T}}\mathbb{E}_x[\tau_A]\right)^{-1}$. Using $e^{-x}\le 1-x+x^2$ for positive x gives

$$\mathbb{E}_x\left[e^{-\theta\tau_A}\right] \le 1-\theta\mathbb{E}_x[\tau_A]+\theta^2\sup_{x\in\mathbb{T}}\mathbb{E}_x[\tau_A]^2 \le \exp\left(-\theta\mathbb{E}_x[\tau_A]+\theta^2\sup_{x\in\mathbb{T}}\mathbb{E}_x[\tau_A]^2\right). \tag{43}$$

Consider $\tau^{(i\leftarrow)}$ the time it takes W to get from $\partial B_{r_1}(x)$ to $B^c_{r_0}(x)$ the $i-th$ time; $\tau^{i\rightarrow}$ the time W needs to get from $\partial B_{r_0}(x)$ to $B_{r_1}(x)$ the i-th time after $B_{r_1}(x)$ has been hit the first time and τ_{r_1} the time it takes W to get from the starting point to $\partial B_{r_1}(x)$. Now by definition we have

$$D_n(x)=\tau_{r_1}+\sum_{i=1}^{n-1}\tau^{(i\rightarrow)}+\sum_{i=1}^{n}\tau^{(i\leftarrow)}. \tag{44}$$

Exponential Markov inequality gives for any $t,\theta>0$

$$\mathbb{P}(D_n(x)\ge t)\le e^{-\theta t}\mathbb{E}\left[e^{\theta D_n(x)}\right] \tag{45}$$

Using (44), by strong Markovianity and estimating by worst starting points this is

$$\le e^{-\theta t}\left(\sup_{z\in\mathbb{T}_2}\mathbb{E}_z\left[e^{\theta\tau_{r_1}}\right]\right)\left(\sup_{z\in B_{r_0}(x)}\mathbb{E}_z\left[e^{\theta\tau^{(1\rightarrow)}}\right]\right)^{n-1}\left(\sup_{z\in B_{r_1}(x)}\mathbb{E}_z\left[e^{\theta\tau^{(1\leftarrow)}}\right]\right)^{n} \tag{46}$$

Using (42) with $\theta=-\frac{\pi\delta}{4\ln r_1}$, and applying Proposition 2, we obtain

$$\begin{aligned}
\sup_{z\in\mathbb{T}_2}\mathbb{E}_z\left[e^{\theta\tau_{r_1}}\right] &\le e^{\frac{\delta}{4}+\frac{\delta^2}{8}+o_{r_1}(1)}\\
\sup_{z\in B_{r_0}(x)}\mathbb{E}_z\left[e^{\theta\tau^{(1\rightarrow)}}\right]^{n-1} &\le e^{(n-1)\left(\frac{\delta}{4}+\frac{\delta^2}{8}+o_{r_1}(1)\right)}\\
\text{and }\sup_{z\in B_{r_1}(x)}\mathbb{E}_z\left[e^{\theta\tau^{(1\leftarrow)}}\right]^{n} &\le e^{no_{r_1}(1)}.
\end{aligned} \tag{47}$$

With $t=(1+\delta)n\frac{1}{\pi}\ln\frac{r_0}{r_1}$, and by the above estimates, (46) reads

$$\mathbb{P}\left(D_n(x)\ge(1+\delta)n\frac{1}{\pi}\ln\frac{r_0}{r_1}\right)\le e^{-n\left(\frac{\delta}{4}+\frac{\delta^2}{4}+o_{r_1}(1)\right)}e^{n\left(\frac{\delta}{4}+\frac{\delta^2}{8}+o_{r_1}(1)\right)}, \tag{48}$$

settling (7). As for (8): for any $n\in\mathbb{N}$ and $\theta>0$ we have

$$\mathbb{P}(D_n(x)\le t)\le e^{\theta t}\mathbb{E}e^{-\theta D_n(x)}\le e^{\theta t}\mathbb{E}\left[e^{-\theta\tau^{(1\rightarrow)}}\right]^{n-1}. \tag{49}$$

Choosing $\theta=\frac{\pi\delta}{2\ln r_1}$ and $t=(1-\delta)n\frac{1}{\pi}\ln\frac{r_0}{r_1}$, applying (43) together with Proposition 2 yields the second claim and concludes the proof of Proposition 1. □

2.2 Estimates for $\mathscr{R}$ and A

Proof of Lemma 2. For $x \in L_\varepsilon^*$, $\{\tau_{r_1} < \tau_{r_K}\}$ almost surely. By rotational invariance and strong Markovianity, the number of excursions from scale $\lceil \delta K \rceil$ to scale $\lceil \delta K \rceil - 1$ in different excursions from scale 1 to scale 0 are independent of each other. The number of excursions from scale $\lceil \delta K \rceil$ to scale $\lceil \delta K \rceil - 1$ in one excursion from scale 1 to scale 0 is distributed like the product of a Bernoulli distributed and an independent geometrically distributed random variable, both with parameter $\lceil \delta K \rceil^{-1}$. (This product has expectation 1). By Cramér's theorem,

$$\begin{aligned} &\mathbb{P}\left(\text{more than } \mathfrak{n}(2) \text{ times } \lceil \delta K \rceil \to \lceil \delta K \rceil - 1 \text{ in the first } \mathfrak{n}(3) \text{ excursions } 1 \to 0\right) \\ &\qquad \leq \exp\left(-\mathfrak{n}(3) I\left(\frac{1}{1-\delta}\right)\right) = \varepsilon^{2K(1-\delta)^3 I\left(\frac{1}{1-\delta}\right)}, \end{aligned} \tag{50}$$

with I the rate function of a Bernoulli$(1/\lceil \delta K \rceil)\times$ geometric$(1/\lceil \delta K \rceil)$. It follows that $\mathbb{P}\left(\left(\mathscr{R}^x\right)^c\right)$ vanishes polynomially in ε for fixed δ and K. Taking the complement yields the first claim.

By Proposition 1 we have

$$\mathbb{P}\left(D_{\mathfrak{n}(3)}(x) \leq \mathfrak{t}\right) \leq \varepsilon^{2K(1-\delta)^3\left(\delta^2/4 + o_{r_1}(1)\right)}, \tag{51}$$

which vanishes faster then, say, ε^3 for K sufficiently large. The second claim thus follows by union bound over all $x \in L_\varepsilon$ on the complements. $\square$

Proof of Lemma 3. The number of times a SRW goes from l to $l+1$ before going from l to $l-1$ is geo$(1/2)$-distributed. Therefore $\mathscr{N}_l^x(n)$ is, by strong Markovianity and rotational invariance, the sum of n independent geo$(1/2)$-distributed r.v.'s. Hence by Cramér's theorem

$$\begin{aligned} \mathbb{P}(A^x) = \prod_{l=\lceil \delta K \rceil}^{K-1} \mathbb{P}(A_l^x) &= \prod_{l=\lceil \delta K \rceil}^{K-1} \exp\left(-n\left(1-\frac{l}{K}\right)^2 I\left(\frac{\left(1-\frac{l+1}{K}\right)^2}{\left(1-\frac{l}{K}\right)^2}\right) + o_\varepsilon(n)\right) \\ &= \exp\left(-\frac{n}{K^2}\sum_{l=\lceil \delta K \rceil}^{K-1} (K-l)^2 I\left(\left(1-\frac{1}{K-l}\right)^2\right) + o_\varepsilon(n)\right), \end{aligned} \tag{52}$$

where $I(x) = x\ln(x) - (1+x)\ln\left(\frac{1+x}{2}\right)$ is the geo(1/2)-rate function. Using $I(1) = I'(1) = 0$ and $I''(1) = \frac{1}{2}$ one quickly obtains $j^2 I\left((1-1/j)^2\right) = 1 + o_j(1)$ as $j \to \infty$, and therefore

$$\mathbb{P}(A^x) = \exp\left(-\frac{n}{K}(1-\delta)(1+o_K(1)) + o_\varepsilon(n)\right) = \varepsilon^{2(1-\delta)(1+o_K(1))+o_\varepsilon(1)}, \tag{53}$$

concluding the proof of the Lemma. $\square$

Acknowledgments. *I am grateful to David Belius, Giuseppe Genovese, Nicola Kistler, Benjamin Schlein and Tobias Weth for enlightening conversations.*

References

[1] *D. Belius, and N. Kistler,* The subleading order of two dimensional cover times, *Probability Theory and Related Fields (2016)*

[2] *A. Dembo,* Simple random covering, disconnection, late and favorite points. *Proceedings of the International Congress of Mathematicians, Madrid. Vol. 3. (2006).*

[3] *A. Dembo, Y. Peres, J. Rosen, and O. Zeitouni,* Cover times for Brownian motion and random walks in two dimensions. *Annals of Mathematics (2) Vol. 160 (2004): 433-464.*

[4] *P.J. Fitzsimmons, and J. Pitman,* Kac's moment formula and the Feynman-Kac formula for additive functionals of a Markov process*, Stochastic Processes and their Applications, Vol. 79 (1999): 117-134.*

[5] *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 and N. Kistler (eds), Springer Lecture Notes in Mathematics, Vol. 2143 (2015)*

[6] *D. Ray,* Sojourn times and the exact Hausdorff measure of the sample path for planar Brownian motion. *Transactions of the American Mathematical Society 106 (1963): 436-444.*

5 Summary

The goal of this thesis is to give insight into the study of hierarchical fields and its application today. We call any field hierarchical that is constructed by generating a rooted tree, given the tree equipping the edges with independent random variables and then considering the field of random variables indexed by leafs, that associates a leaf to the sum of random variables along the path from the root to the leaf. Typically one is interested how the extremes or related functionals like extremal process or high points of such fields develop when considering an in some sense consistent sequence of hierarchical fields growing the number of leafs to infinity. In the discussion of hierarchical and approximately hierarchical models the notion of scales is central. If the model at hand is constructed from a tree, scale refers to the distance to the root, i.e. behavior on small scales is behavior close to the root and number of scales is the (maximal) distance from the root to the leafs. In not exactly hierarchical models however we still use the notion in order to indicate where the suggested tree like structure is to be found. Prominent hierarchical models are e.g. Derrida's random energy model (REM), Derrida's generalized random energy model (GREM), branching random walk (BRW) and branching Brownian motion (BBM). The REM was introduced by Derrida [33] and simply considers a collection of independent random variables, studying the extremes of which is classical and discussed in the Fisher-Tippett-Gnedenko theorem and related results. The GREM model corresponds to a regular tree with finite number of scales and centered Gaussian edge weights, which was introduced by Derrida [32] and extensively analyzed by Bovier and Kurkova [24]. The GREM is very well analyzed, but for this discussion we will only note the following result describing the critical case:

Theorem 8. *Let $K \in \mathbb{N}$, N be a multiple of K and consider the complete tree of height K where each non leaf vertex has $2^{N/K}$ children. Equipping each edge with independent Gaussians of variance N/K and considering the field associated to the leafs $(X_\sigma^{(N)})_{\sigma\in\Sigma_N}$, we have that*

$$\max_{\sigma\in\Sigma_N} X_\sigma^{(N)} - \left(\sqrt{2\ln 2}N - \frac{1}{2\sqrt{2\ln 2}}\ln N\right) \tag{27}$$

converges to a Gumbel distribution as $N \to \infty$ for any fixed K.

For details see [41] or [24]. We compare this result to the BRW analogue. BRWs in general are constructed by adding an independent copy of a point process as children to each vertex, where the number of points of the process is the number of children and the displacement gives the edge weights, then considering the field of vertices of depth N. The analogue case to GREM setting we considered is choosing the point process consisting of two independent standard Gaussian points and considering the maximum at scale N. This construction is the same as considering the binary tree of depth N with standard Gaussian edge weights. Which in turn is identical to the critical GREM with $K = N$ levels. In this case we have

Theorem 9. *Let $(X_\sigma^{(N)})_{\sigma\in\Sigma_N}$ be defined as before, then we have that*

$$\max_{\sigma\in\Sigma_N} X_\sigma^{(N)} - \left(\sqrt{2\ln 2}N - \frac{3}{2\sqrt{2\ln 2}}\ln N\right) \tag{28}$$

converges to a randomly shifted Gumbel distribution for $K = N$.

This is a consequence of [1, Theorem 1.1], but not a straight forward extension of Theorem 8 as one notices the extra factor 3 in the log-correction. Of course such an effect needs to have an explanation how and why the stronger correlations in this latter case translate to the different recentering. This is precisely the topic of the first paper of this thesis Kistler and Schmidt [42], wherein we discuss a class of models interpolating between the cases $K = 1$ and $K = N$, see Figure 9.

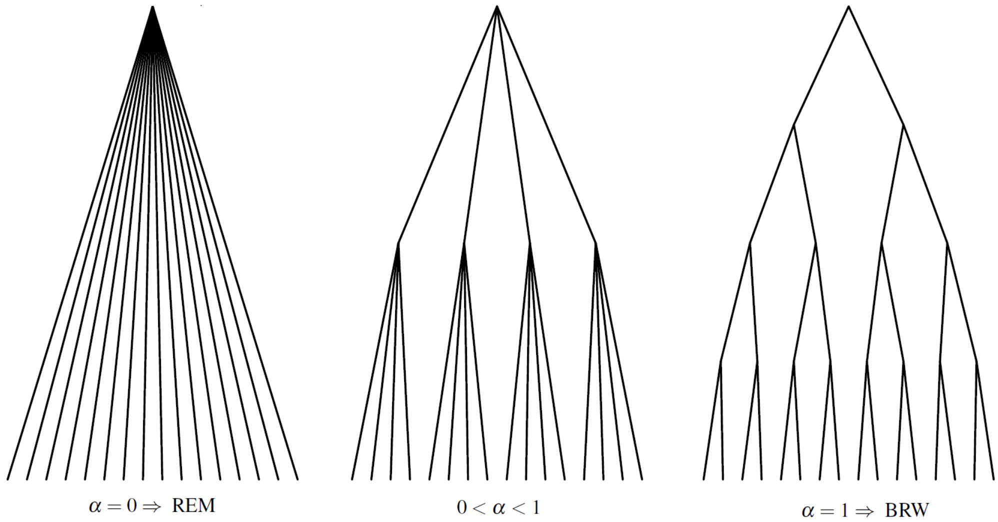

Figure 9: Trees interpolating between REM and BRW

The main result being

Theorem 10. *Let $(X_\sigma^{(N)})_{\sigma\in\Sigma_N}$ be defined as before with $K = N^\alpha$ for some $\alpha \in (0,1)$, then we have that*

$$\sum_{\sigma\in\Sigma_N} \delta_{X_\sigma^{(N)} - \left(\sqrt{2\ln 2}N - \frac{1+2\alpha}{2\sqrt{2\ln 2}}\ln N\right)} \to \Xi, \tag{29}$$

weakly as $N \to \infty$, where Ξ is a Poisson point process of intensity $\frac{1}{\sqrt{2\pi}}e^{-\sqrt{2\ln 2}x}dx$.

For comparison to the results before note that this implies weak convergence of the recentered maximum to the Gumbel distribution. We see that the log-correction interpolates linearly between REM and BRW in α. The result is correct for $\alpha = 0$ also which is the REM case, but cannot be correct for $\alpha = 1$ in view of Theorem 9. This is very intuitive as fluctuations in the beginning and close relatives of maximal particles in the end produce non Poissonian effects that have to carry over to the limit. Some intuition for Theorem 10 is given in Section 1.2 of this thesis.

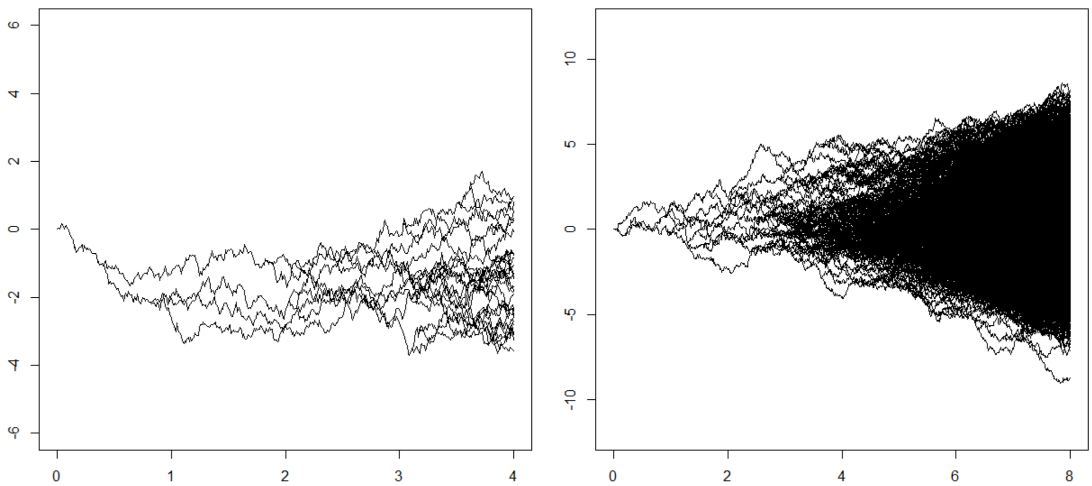

Figure 10: Two realizations of branching Brownian motion

We continue with the study of a model with many similarities to the BRW, namely branching Brownian motion (BBM), see Fig. 10 for two realizations. BBM can be constructed by considering one particle starting at 0 following a standard Brownian motion for an exponentially distributed time of parameter one. After said time the particle splits into two independent copies of itself progressing and splitting independently of each other and of the past from the splitting point onwards. The extremes of BBM have been studied extensively (see e.g. [2, 8, 21, 27, 43]), not only for the theoretical appeal but also for the connection to the FKPP equation [46] and the relevance to disordered systems [25, 34]. Also the generalizations to time dependent speed of the Brownian motions is studied [22, 23, 34, 35]. One such model is two-speed BBM, which simply considers one speed up to some time $t/2$ and another from time $t/2$ to t, see Fig. 11. If

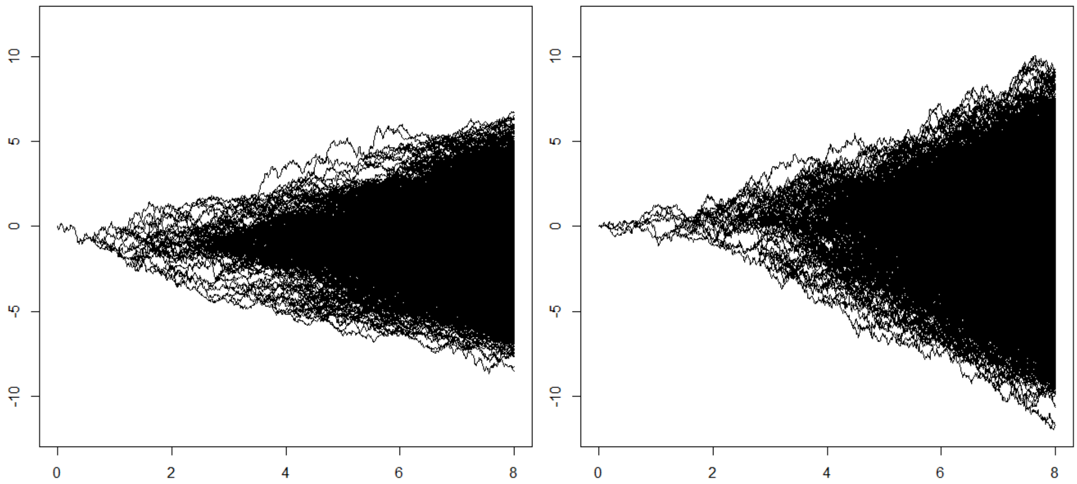

Figure 11: Two-speed BBM, strong correlation (left), weak correlation (right)

the fluctuations in the beginning are weaker we are in the weak correlation regime and if they are stronger we are in the strong correlation regime. Roughly speaking stronger correlations stem from less chances to improve later, hence increasing the value of a good candidate, which in turn results in there being less candidates at time $t/2$ that might produce offspring that are optimal at time t. The bigger the fluctuations at the end the more the balance between entropy and energy in the beginning moves towards entropy. Explained in more simple words considering the time $t/2$ as now: If large changes are bound to happen chances are the best in the future are far from the best now, simply due to the fact that many particles now are suboptimal but only few are very good. On the other hand if things change in the future by not as much the advantage of good particles today becomes more pronounced. Standard BBM is the critical case in between these two regimes. In the second paper of this thesis Glenz, Kistler and Schmidt [36] we discuss the number of highpoints of BBM, which has strong connections to the weak correlation regime of two-speed BBM. For $(X_i(t), i \leq n(t))$ the position of particles of a standard BBM at time t we define the number of α-high points by

$$Z_\alpha(t) \equiv \#\left\{k \leq n(t) : x_k(t) \geq \left(\sqrt{2} - \alpha\right) t\right\}. \tag{30}$$

This is sensible as there are no 0-high points for large t and roughly every second particle would be a $\sqrt{2}$-high point. The main result of Glenz, Kistler and Schmidt [36] is the following strong law of large numbers for high points:

Theorem 11. *For any* $0 < \alpha < \sqrt{2}$

$$\lim_{t\to\infty} \frac{Z_\alpha(t)}{\mathbb{E}Z_\alpha(t)} = Y_\alpha, \quad \textit{almost surely}, \tag{31}$$

where Y_α is given as the almost sure limit of McKean's martingale:

$$Y_\alpha(t) \equiv \sum_{k\leq n(t)} \exp\left[-t\left(1+\frac{1}{2}\Delta_\alpha^2\right)+\Delta_\alpha x_k(t)\right]. \tag{32}$$

To explain the intuition behind this result let $\Delta_\alpha \equiv \sqrt{2}-\alpha$, $n(r)$ the number of particles at time r and let $n_i(t-r)$ the number of children particle $i \leq n(r)$ at time r has at time t. By grouping particles at time t in groups of common ancestor at time r we identify

$$\{x_k(t), k\leq n(t)\} = \{x_i(r)+x_{i,j}(t-r), i\leq n(r), j\leq n_i(t-r)\}. \tag{33}$$

Computing the conditional expectation of $Z_\alpha(t)$ given the events up to time r we have

$$\begin{aligned}
\mathbb{E}\left[Z_\alpha(t) \mid \mathscr{F}_r\right] &= \mathbb{E}\left[\sum_{k\leq n(t)} 1\left\{x_k(t)\geq \Delta_\alpha t\right\}\Big|\mathscr{F}_r\right] = \\
&= \mathbb{E}\left[\sum_{i\leq n(r)}\sum_{j\leq n_i(t-r)} 1\left\{x_{i,j}(t-r)\geq \Delta_\alpha(t-r)-(x_i(r)-\Delta_\alpha r)\right\}\Big|\mathscr{F}_r\right] \\
&= \sum_{i\leq n(r)} e^{t-r}\mathbb{P}\left[x_1(t-r)\geq \Delta_\alpha(t-r)-(x_i(r)-\Delta_\alpha r)\Big|\mathscr{F}_r\right] \\
&\sim \left(\Delta_\alpha\sqrt{2\pi}\right)^{-1}\exp\left[\left(1-\Delta_\alpha^2/2\right)t-\frac{1}{2}\log(t)\right]Y_\alpha(r), \quad \text{a.s.},
\end{aligned} \tag{34}$$

The last step by a standard tail estimate for Gaussian random variables and using that r is much smaller than t. This gives us

$$\frac{\mathbb{E}\left[Z_\alpha(t) \mid \mathscr{F}_r\right]}{\mathbb{E}\left[Z_\alpha(t)\right]} = \frac{\mathbb{E}\left[Z_\alpha(t) \mid \mathscr{F}_r\right]}{\mathbb{E}\left[Z_\alpha(t)\mathscr{F}_0\right]} \sim \frac{Y_\alpha(r)}{Y_\alpha(0)} = Y_\alpha(r) \tag{35}$$

making the claim of Theorem 11 what one expects, as it is easy to predict that the randomness in the beginning carries over to the limit as each fluctuation in the early evolution influences a non vanishing fraction of the entire population. The proof of Theorem 11 is controlling the difference of $Z_\alpha(t)$ and its conditional expectation by a multiscale refinement of the second moment method truncating off particles moving too far above the optimal strategy.

The third paper of this thesis is Schmidt [52]. It discusses how the methods developed for hierarchical fields can be applied to only approximately hierarchical models in the case of two-dimensional cover times giving a simple proof to the famous result by Dembo, Peres, Rosen and Zeitouni (short DPRZ) [31]:

Theorem 12. *For T_ε the ε-cover time of the two-dimensional unit torus we have*

$$\frac{T_\varepsilon}{(\ln\varepsilon)^2} \to \frac{2}{\pi} \qquad \textit{almost surely as } \varepsilon\to 0. \tag{36}$$

To define this cover time precisely we identify the unit torus $\mathbb{T}_2$ with $[0,1)\times[0,1)\subset\mathbb{R}^2$, endowed with the metric

$$d_{\mathbb{T}_2}(x,y) = \min\left\{||x-y+(e_1,e_2)|| : e_1,e_2\in\{-1,0,1\}\right\}. \tag{37}$$

Brownian motion on $\mathbb{T}_2$ is given by $W_t \equiv \left(\hat{W}_1(t) \mod 1, \hat{W}_2(t) \mod 1\right)$, where $\hat{W}$ is standard Brownian motion on $\mathbb{R}^2$. The ε-cover time T_ε is then given as the first time the path of W is within

ε distance of any point on the torus. Considering the $T_\varepsilon(x)$ the hitting time of the ε-ball around $x \in \mathbb{T}_2$ we have the identity

$$T_\varepsilon = \sup_{x \in \mathbb{T}_2} T_\varepsilon(x). \tag{38}$$

To make Theorem 12 plausible we do some rough calculations. Consider $0 < r < R < 1/2$ and some reference point $x \in \mathbb{T}_2$, then one can show that it takes Brownian motion on the torus on average $\frac{1}{\pi} \ln \frac{R}{r}$ long to perform one excursion from $\partial B_R(x)$ to $\partial B_r(x)$ and back up to bounded error. Hence for r small and t large enough W completes about $t\pi / \ln \frac{R}{r}$ such excursions up to time t. The chance of $B_\varepsilon(x)$ being hit by W in one such excursion is easily seen to be exactly $\frac{\ln R - \ln r}{\ln R - \ln \varepsilon}$. Hence in total the chance to avoid on ε-ball is roughly

$$\left(1 - \frac{\ln R - \ln r}{\ln R - \ln \varepsilon}\right)^{\frac{t}{\frac{1}{\pi} \ln \frac{R}{r}}} \approx \exp\left(\pi t (\ln \varepsilon)^{-1}\right). \tag{39}$$

As one can place of order ε^{-2} disjoint ε-balls on the unit torus it is plausible to think that the torus should be ε-covered around the time all these balls are hit. Hoping that the correlations are not to strong we can conjecture that this happens around the time the expected number of unhit ε-balls is of order one, which suggests:

$$\varepsilon^{-2} \exp\left(\pi t (\ln \varepsilon)^{-1}\right) \stackrel{!}{=} O(1). \tag{40}$$

This hits precisely the result. One can even make this line of reasoning rigorous to establish an upper bound. For a matching lower bound it is however necessary to handle the correlations, which is much more delicate. This is done by controlling excursion counts between circles of mesoscopic sizes exploiting the strong Markovianity of W to its fullest. This point of view reveals a decoupling effect resulting in a hierarchical structure, which is best explained with Figure 12. Consider two

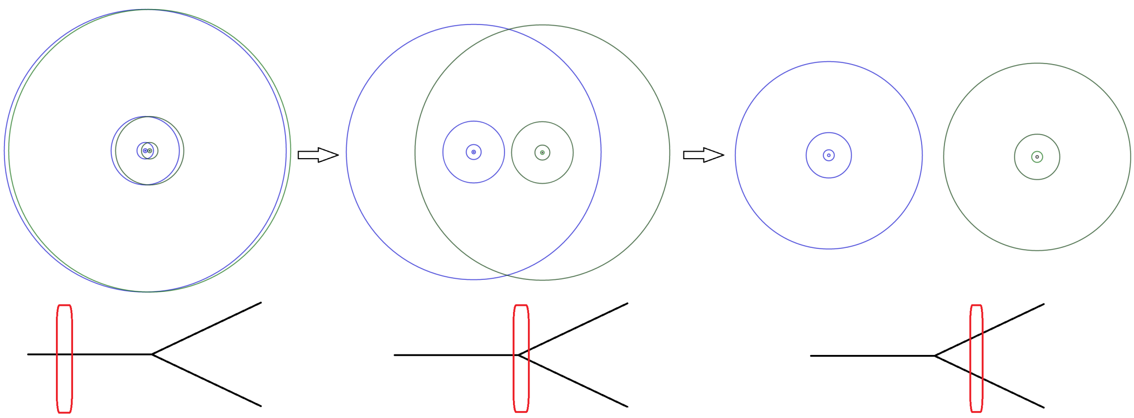

Figure 12: Hierarchical structure of cover times.

points on the torus and circles of different sizes around them. If the radii of considered circles are large in comparison to the distance of the two points the circles are similar resulting in similar excursion counts. This is depicted on the left and gives morally the common trunk of these two points. If we consider circles of size that is of the same order as the distance, then we see the middle picture which is a complicated situation, however as this can happen only for very specific circle sizes the error caused by this effect can be controlled. If the distance is of larger order than the circle sizes, then those circles are disjoint (right), which means conditionally on the exterior what happens inside them is independent of each other. As excursion counts in between circles have a distribution that does not depend on the starting point by rotational invariance we obtain perfect independence of excursion counts in this case. This observation allows a multiscale analysis treating the model as though it was hierarchical to be successful.

6 Zusammenfassung (German summary)

Das Ziel dieser Ausarbeitung ist einen Einblick in die momentane Erforschung hierarchischer Felder zu geben. Wir nennen jedes Feld hierarchisch, dass konstruiert ist, indem ein verwurzelter Baum generiert wird, gegeben dem Baum seine Kanten mit unabhängigen Zufallsvariablen versehen werden und das Feld indiziert durch die Blätter des Baumes betrachtet wird, dass jedem Blatt die Summe der Kantengewichte entlang des Weges von Wurzel zu Blatt zuordnet. Typischerweise gilt das Interesse dem Verhalten der Extreme solcher Felder oder verwandter Funktionale, wie dem Extremalprozess oder der Anzahl sogenannter highpoints, entlang einer auf eine Art konsistenten Folge von hierarchischen Feldern mit wachsender Blattzahl. Ein zentraler Begriff im Studium hierarchischer Felder oder approximativ hierarchischer Felder ist der Begriff der Skalen. Ist ein Modell wie beschrieben durch einen Baum konstruiert, so bezieht sich der Begriff Skala auf den Abstand zur Wurzel. Das bedeutet, "Verhalten auf kleinen Skalen" meint das Verhalten nahe der Wurzel und "Anzahl der Skalen" ist nichts weiter als die (maximale) Tiefe des betrachteten Baumes. Für nicht exakt hierarchische Modelle verwenden wir den Begriff der Skalen um aufzuzeigen wo die suggerierte Baumstruktur zu finden ist. Prominente Beispiele für hierarchische Modelle sind Derridas random energy model (REM) und generalized random energy model (GREM), branching random walk (BRW) und branching Brownian motion (BBM). Im von Derrida bekannt gemachten REM betrachtet man unabhängige identisch verteilte Zufallsvariablen. Die Betrachtung der Extreme des REM ist somit ein klassisches Problem der Extremwerttheorie und kann durch den Satz von Fisher-Tippett-Gnedenko und verwandte Resultate behandelt werden. Das GREM wurde ebenfalls von Derrida [32] eingeführt. Es entsteht durch Betrachtung eines gleichmäßigen Baumes mit fester Stufenzahl und zentrierten normalverteilten Kantengewichten. Für eine ausführliche Behandlung dieses Modells siehe Bovier und Kurkova [24]. Für unsere Zwecke beschränken wir uns auf die Angabe von folgendem Resultat, das den kritischen Fall beschreibt:

Theorem 13. *Sei $K \in \mathbb{N}$, N ein Vielfaches von K und betrachte den vollständigen Baum der Tiefe K bei dem jeder Knoten, der kein Blatt ist, $2^{N/K}$ Kinder besitzt. Das Versehen der Kanten mit zentriert normalverteilten Zufallsvariablen mit Varianz N/K schließt die Konstruktion ab. Für $(X_\sigma^{(N)})_{\sigma \in \Sigma_N}$, das zu den Blättern assoziierte Feld, gilt*

$$\max_{\sigma \in \Sigma_N} X_\sigma^{(N)} - \left(\sqrt{2 \ln 2} N - \frac{1}{2\sqrt{2 \ln 2}} \ln N \right) \tag{41}$$

konvergiert gegen die Gumbelverteilung für $N \to \infty$ und K fest.

Wir vergleichen dieses Resultat mit seinem BRW-Analogon. Im Allgemeinen sind BRWs konstruiert, indem an die Wurzel ein Punktprozess angehängt wird, wobei jeder Punkt einem Kind entspricht und die Auslenkung des Punktes dem Kantengewicht. Dieser Prozess wird für jedes Kind der Wurzel unabhängig wiederholt, um eine zweite Stufe anzuhängen, usw. bis N Stufen entstanden sind. Das Analogon zum kritischen GREM in diesem setting ist die Wahl des Punktprozesses, der aus zwei unabhängigen standardnormalverteilten Zufallsvariablen entsteht. Diese Konstruktion ist äquivalent zur Betrachtung des binären Baumes mit Tiefe N und standard normalverteilten Kantengewichten oder zur Betrachtung eines GREMs mit $K = N$ Skalen. Für dieses Modell gilt:

Theorem 14. *Sei $(X_\sigma^{(N)})_{\sigma\in\Sigma_N}$ gegeben wie zuvor, dann konvergiert*

$$\max_{\sigma\in\Sigma_N} X_\sigma^{(N)} - \left(\sqrt{2\ln 2}N - \frac{3}{2\sqrt{2\ln 2}}\ln N\right) \tag{42}$$

gegen eine zufällig verschobene Gumbelverteilung für $K = N$.

Dieses Resultat folgt aus [1, Theorem 1.1], ist jedoch keine simple Verallgemeinerung von Theorem 13 wegen des extra Faktors 3 in der log-Korrektur. Natürlich muss erklärlich sein, wie und warum die stärkeren Korrelationen eine veränderte Zentrierung zur Folge haben. Dies ist präzise das Thema des ersten Papers dieser Arbeit Kistler und Schmidt [42] in der eine Klasse von Modellen beschrieben wird die zwischen $K = 1$ (REM) und $K = N$ (BRW) interpolieren. Siehe Abbildung 13.

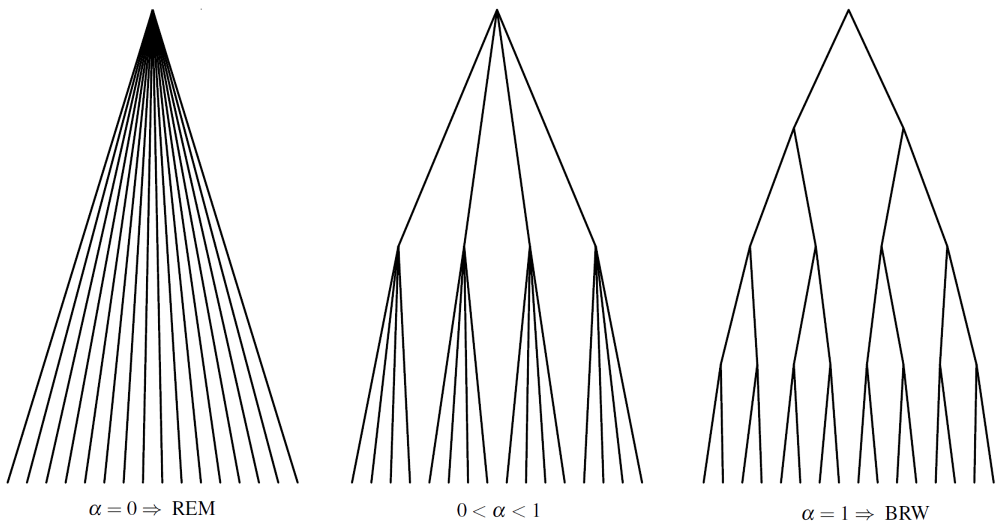

Figure 13: Interpolierende Bäume von REM bis BRW

Das Hauptresultat von Kistler und Schmidt [42] ist

Theorem 15. *Sei $(X_\sigma^{(N)})_{\sigma\in\Sigma_N}$ definiert wie zuvor für $K = N^\alpha$ und ein $\alpha \in (0,1)$, dann gilt*

$$\sum_{\sigma\in\Sigma_N} \delta_{X_\sigma^{(N)} - \left(\sqrt{2\ln 2}N - \frac{1+2\alpha}{2\sqrt{2\ln 2}}\ln N\right)} \to \Xi, \tag{43}$$

schwach für $N \to \infty$ mit Ξ einem Poisson Punkt Prozess mit Intensität $\frac{1}{\sqrt{2\pi}}e^{-\sqrt{2\ln 2}x}dx$.

Zu Vergleichszwecken sei angemerkt, dass dies direkt schwache Konvergenz des rezentrierten Maximums gegen die Gumbelverteilung zur Folge hat. Wir sehen, dass die log-Korrektur linear in α zwischen REM und BRW interpoliert. Gegebenes Resultat ist für $\alpha = 0$ (REM Fall) immer noch korrekt, kann jedoch nicht korrekt sein im BRW Fall $\alpha = 1$ in Anbetracht von Theorem 14. Dies ist zu erwarten, da in diesem Fall sowohl Fluktuationen in der frühen Entwicklung als auch enge Verwandte der maximalen Partikel, die sich nur um Ordnung 1 unterscheiden, existieren. Diese beiden Effekte haben einen nicht poissonschen Beitrag zum Extremalprozess, für den es keinen Grund gibt nach Grenzübergang nicht mehr sichtbar zu sein. Die intuitiven Gründe für die Gültigkeit von Theorem 15 werden detaillierter in Kapitel 1.2 dieser Arbeit behandelt.

Wir gehen zur Betrachtung eines mit dem BRW verwandten Modells über: Branching Brownian motion (BBM), siehe Abbildung 14. Zur Konstruktion einer BBM betrachten wir ein in 0

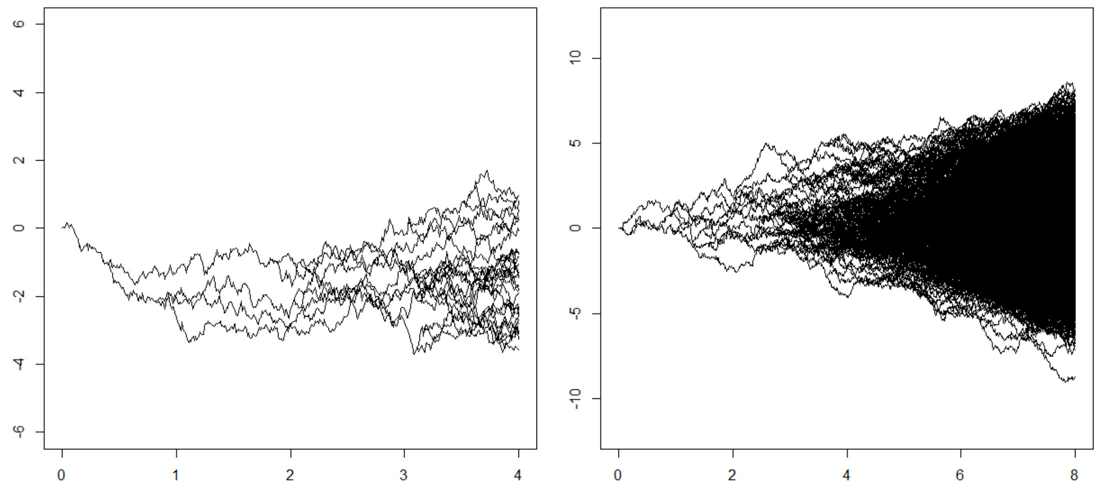

Figure 14: Zwei Realisationen einer branching Brownian motion

startendes Partikel, dass für eine exponentialverteilte Zeit zum Parameter 1 einer standard brownschen Bewegung folgt und sich dann in zwei Partikel spaltet, die sich vom Spaltungspunkt aus wie unabhängige Kopien des ersten Partikels verhalten. Die Extreme der BBM wurden ausführlich studiert (siehe z.B. [2, 8, 21, 27, 43]). Dies ist nicht nur auf Interesse rein theoretischer Natur zurückzuführen, sondern auch bedingt durch den Zusammenhang zur FKPP Gleichung [46] und die Relevanz für ungeordnete Systeme [25, 34]. Auch die Verallgemeinerungen auf zeitabhänge Geschwindigkeit der brownschen Bewegung wird untersucht [22, 23, 34, 35]. Ein Modell in dieser Klasse ist two-speed BBM, die bis zur Zeit $t/2$ eine Geschwindigkeit und von $t/2$ bis t eine andere Geschwindigkeit betrachtet (siehe Abbildung 15). Fluktuiert die brownsche Bewegung im ersten

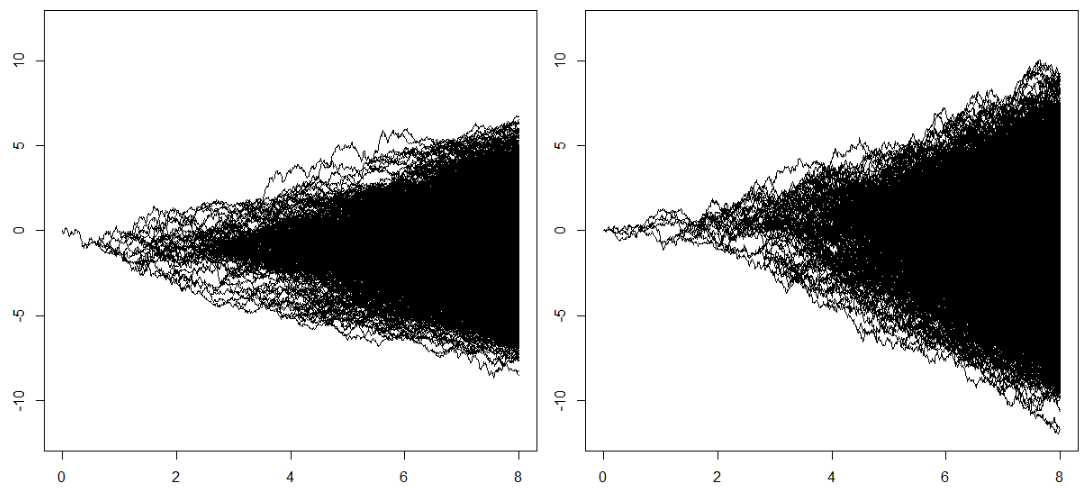

Figure 15: Two-speed BBM, starke Abhängigkeiten (links), schwache Abhängigkeiten (rechts)

Teil weniger als im Zweiten, so sind wir im Regime schwacher Abhängigkeiten. Umgekehrt sind die Fluktuationen am Anfang stärker, so befinden wir uns im Regime starker Abhängigkeiten. Grob kann man dies wie folgt begründen: Weniger Fluktuationen am Ende reduzieren die Chancen auf eine große Verbesserung in der zweiten Hälfte; dies erhöht den Wert eines guten Kandidaten zur Zeit $t/2$, was dazu führt, dass es weniger Kandidaten zur Zeit $t/2$ gibt, die eine Chance darauf haben, einen optimalen Nachfahren zur Zeit t zu erzeugen. Je größer die Fluktuationen am

Ende umso mehr verschiebt sich das Energie-Entropie-Gleichgewicht in Richtung Entropie. In einfachen Worten: Wenn große Veränderungen anstehen, sind die optimalen Partikel von morgen, Nachfahren suboptimaler Partikel von heute, allein aus dem Grund, dass diese enorm zahlreich sind. Auf der anderen Seite, wenn stabile Zeiten anstehen, ist der Vorteil von heute von größerer Bedeutung. Standard BBM ist der kritische Fall zwischen diesen Regimen. Im zweiten Paper dieser Arbeit Glenz, Kistler und Schmidt [36] besprechen wir das Verhalten der Anzahl von high-points einer BBM. Diese Größe ist stark mit dem schwachen Abhängigkeitsregime der two-speed BMM verbunden. Mit $(X_i(t), i \leq n(t))$ den Positionen der Partikel einer standard BBM zur Zeit t definieren wir die Anzahl der α-highpoints durch

$$Z_\alpha(t) \equiv \#\left\{k \leq n(t) : x_k(t) \geq \left(\sqrt{2}-\alpha\right)t\right\} \tag{44}$$

für $\alpha \in (0, \sqrt{2})$. Dies ist sinnvoll, da für $\alpha = 0$ keine Partikel mehr existieren und für $\alpha = \sqrt{2}$ praktisch jeder zweite Partikel ein highpoint ist. Das Hauptresultat von Glenz, Kistler und Schmidt [36] ist folgendes starke Gesetz der großen Zahlen für highpoints:

Theorem 16. *Für $0 < \alpha < \sqrt{2}$ gilt*

$$\lim_{t\to\infty} \frac{Z_\alpha(t)}{\mathbb{E}Z_\alpha(t)} = Y_\alpha, \quad \textit{fast sicher}, \tag{45}$$

wobei Y_α der fast sichere Grenzwert von McKean's Martingal $Y_\alpha(t)$ ist:

$$Y_\alpha(t) \equiv \sum_{k\leq n(t)} \exp\left[-t\left(1+\frac{1}{2}\Delta_\alpha^2\right) + \Delta_\alpha x_k(t)\right]. \tag{46}$$

Zur Erklärung der Intuition hinter diesem Resultat sei $\Delta_\alpha \equiv \sqrt{2}-\alpha$, $n(r)$ die Anzahl der Partikel zur Zeit r und sei $n_i(t-r)$ die Anzahl der Kinder die ein Partikel $i \leq n(r)$ zur Zeit r bis Zeit t erzeugt. Durch sortieren der Partikel zur Zeit t in Gruppen mit gemeinsamem Vorfahren zur Zeit r identifizieren wir

$$\{x_k(t), k \leq n(t)\} = \{x_i(r) + x_{i,j}(t-r), i \leq n(r), j \leq n_i(t-r)\}. \tag{47}$$

Berechnen wir den bedingten Erwartungswert von $Z_\alpha(t)$, gegeben die Ereignisse bis zur Zeit r, so ergibt sich

$$\begin{aligned}
\mathbb{E}\left[Z_\alpha(t) \mid \mathscr{F}_r\right] &= \mathbb{E}\left[\sum_{k\leq n(t)} 1\left\{x_k(t) \geq \Delta_\alpha t\right\} \middle| \mathscr{F}_r\right] = \\
&= \mathbb{E}\left[\sum_{i\leq n(r)} \sum_{j\leq n_i(t-r)} 1\left\{x_{i,j}(t-r) \geq \Delta_\alpha(t-r) - (x_i(r) - \Delta_\alpha r)\right\} \middle| \mathscr{F}_r\right] \\
&= \sum_{i\leq n(r)} e^{t-r}\mathbb{P}\left[x_1(t-r) \geq \Delta_\alpha(t-r) - (x_i(r)-\Delta_\alpha r) \middle| \mathscr{F}_r\right] \\
&\sim \left(\Delta_\alpha\sqrt{2\pi}\right)^{-1} \exp\left[\left(1-\Delta_\alpha^2/2\right)t - \frac{1}{2}\log(t)\right] Y_\alpha(r), \quad \text{f.s..}
\end{aligned} \tag{48}$$

Der letzte Schritt durch gaussche Tailabschätzungen unter Verwendung, dass r wesentlich kleiner als t ist. Somit sehen wir, dass

$$\frac{\mathbb{E}\left[Z_\alpha(t) \mid \mathscr{F}_r\right]}{\mathbb{E}\left[Z_\alpha(t)\right]} = \frac{\mathbb{E}\left[Z_\alpha(t) \mid \mathscr{F}_r\right]}{\mathbb{E}\left[Z_\alpha(t)\mathscr{F}_0\right]} \sim \frac{Y_\alpha(r)}{Y_\alpha(0)} = Y_\alpha(r) \tag{49}$$

gilt, was die Aussage von Theorem 16 zum natürlichen Resultat macht. Natürlich auch daher, dass es leicht ist vorherzusagen, dass die Fluktuationen am Anfang der Evolution im Grenzwert sichtbar sind, da diese einen Einfluss auf einen nicht verschwindenden Anteil der Gesamtpopulation haben. Der Beweis von Theorem 16 besteht somit im Wesentlich daraus, zu zeigen, dass die Differenz zwischen $Z_\alpha(t)$ und seinem bedingten Erwartungswert verschwindet. Dies geschieht mittels einer Multiskalenverbesserung der zwei Momenten Methode, die Partikel zu weit oberhalb der optimalen Strategie abschneidet.

Das dritte Paper dieser Arbeit ist Schmidt [52]. Es beschäftigt sich damit, wie die entwickelten Methoden für hierarchische Felder Anwendung auf approximativ hierarchische Felder finden, im Falle von Abdeckzeiten (cover times) in zwei Dimensionen. Dieses Studium manifestiert sich in einem wesentlich vereinfachten Beweis des berühmten Resultats von Dembo, Peres, Rosen und Zeitouni (kurz DPRZ) [31]:

Theorem 17. *Für T_ε die ε-Abdeckzeit des zweidimensionalen Einheitstorus gilt*

$$\frac{T_\varepsilon}{(\ln \varepsilon)^2} \to \frac{2}{\pi} \qquad \textit{fast sicher für } \varepsilon \to 0. \tag{50}$$

Um Abdeckzeiten präzise zu definieren, identifizieren wir den Einheitstorus $\mathbb{T}_2$ mit $[0,1) \times [0,1) \subset \mathbb{R}^2$, ausgestattet mit der Metrik

$$d_{\mathbb{T}_2}(x,y) = \min\{||x-y+(e_1,e_2)|| : e_1, e_2 \in \{-1,0,1\}\}. \tag{51}$$

Eine brownsche Bewegung auf $\mathbb{T}_2$ ist durch $W_t \equiv \left(\hat{W}_1(t) \mod 1, \hat{W}_2(t) \mod 1\right)$ definiert, wobei $\hat{W}$ eine standard brownsche Bewegung auf $\mathbb{R}^2$ ist. Die ε-Abdeckzeit T_ε ist nun gegeben durch die erste Zeit zu der W bis auf Abstand ε an jedem Punkt war. Betrachten wir $T_\varepsilon(x)$ die erste Treffzeit des ε-Balls um $x \in \mathbb{T}_2$, dann gilt folgende Identität

$$T_\varepsilon = \sup_{x \in \mathbb{T}_2} T_\varepsilon(x). \tag{52}$$

Um Theorem 17 zu plausibilisieren führen wir einige Überschlagsrechnungen durch. Betrachten wir $0 < r < R < 1/2$ und einen Referenzpunkt $x \in \mathbb{T}_2$, dann kann gezeigt werden, dass eine brownsche Bewegung in Erwartung $\frac{1}{\pi} \ln \frac{R}{r}$ Zeiteinheiten für eine Exkursion von $\partial B_R(x)$ nach $\partial B_r(x)$ und zurück benötigt (bis auf beschränkten Fehler). Daher gilt für r klein und t groß genug, dass W circa $t\pi / \ln \frac{R}{r}$ solche Exkursionen bis zur Zeit t absolviert. Die Chance, dass $B_\varepsilon(x)$ von W in einer solchen Exkursion getroffen wird beträgt exakt $\frac{\ln R - \ln r}{\ln R - \ln \varepsilon}$. Folglich ist die Wahrscheinlichkeit einen ε-Ball zu vermeiden grob

$$\left(1 - \frac{\ln R - \ln r}{\ln R - \ln \varepsilon}\right)^{\frac{t}{\frac{1}{\pi}\ln \frac{R}{r}}} \approx \exp\left(\pi t (\ln \varepsilon)^{-1}\right). \tag{53}$$

Da grob ε^{-2} disjunkte ε-Bälle auf dem Einheitstorus platziert werden können, ist es plausibel zu denken, dass ε-Abdeckung ungefähr zu der Zeit stattfindet, zu der all diese Bälle getroffen worden sind. Wenn wir nun für einen kurzen Moment blind darauf hoffen, dass die Korrelationen nicht zu stark sind, sollte beschriebenes Ereignis grob zu der Zeit eintreten, zu der die erwartete Anzahl ungetroffener Bälle von Ordnung 1 ist. Somit ergibt sich folgende Vermutung:

$$\varepsilon^{-2} \exp\left(\pi t (\ln \varepsilon)^{-1}\right) \stackrel{!}{=} O(1). \tag{54}$$

Diese trifft exakt Theorem 17. Man kann diese grobe Betrachtung als obere Schranke rigoros beweisen. Für eine passende untere Schranke ist es jedoch notwendig die Abhängigkeiten genauer zu

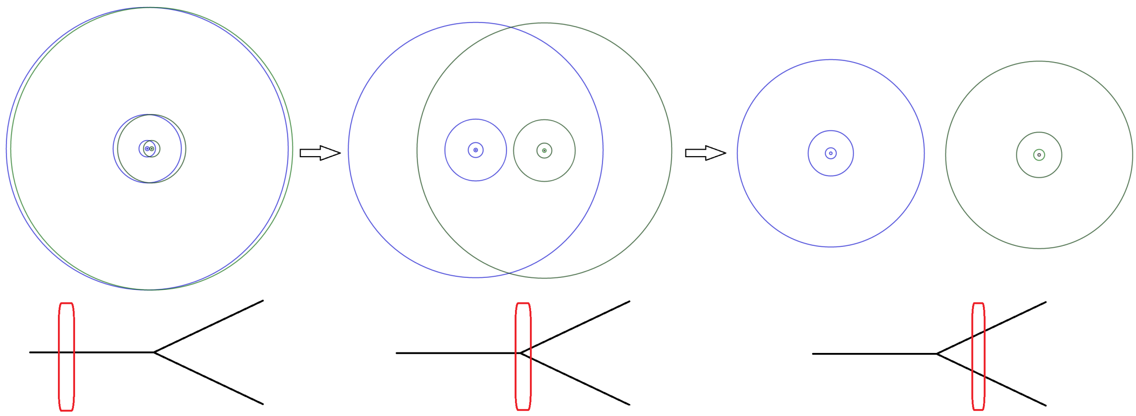

Figure 16: Hierarchische Struktur von Abdeckzeiten.

studieren. Dies erfolgt durch Kontrolle von Exkursionsanzahlen zwischen Kreisen mesoskopischer Größe mittels Nutzung der starken Markoveigenschaft von W. Dieser Blickwinkel offenbart einen Entkopplungseffekt, der eine (approximativ) hierarchische Struktur erzeugt. Um das Phänomen zu beschreiben betrachten wir Abbildung 16. Betrachten wir zwei Punkte auf dem Torus und Kreise verschiedener Größe um sie herum. Wenn Kreisradien von größerer Ordnung als der Abstand der Punkte sind, dann sind jeweils die zwei Kreise gleicher Größe mit verschiedenen Mittelpunkten praktisch identisch und selbiges gilt demnach auch für betrachtete Exkursionsanzahlen. Dies ist abgebildet links und gibt uns den gemeinsamen "Stamm" der zwei Punkte. Betrachtet man Kreise mit Radius von gleicher Größenordnung wie der Abstand, so ist die Situation komplizierter, wie wir im mittleren Bild sehen. Da dies jedoch nur für vergleichsweise wenige Kreisgrössen eintritt, kann dieser Fall grob abgeschätzt werden, ohne problematische Fehler zu verursachen. Ist der Abstand zwischen den Punkten größer als die betrachteten Radien, sind die Kreise disjunkt (rechts). Disjunkte Kreise bedeuten, dass gegeben dem Geschehen außerhalb, ist das Geschehen innerhalb der Kreise unabhängig voneinander. Da Exkursionsanzahlen zwischen Kreisen mit dem selben Mittelpunkt in Verteilung nicht von dem Startpunkt auf dem Rand abhängen, erhalten wir perfekte Unabhängigkeit von Exkursionsanzahlen disjunkter Kreise. Diese Beobachtung erlaubt eine Multiskalenanalyse des Modells, als ob es ein hierarchisches Modell wäre, die letztendlich zum Erfolg führt.

References

[1] E. Aïdékon, *Convergence in law of the minimum of a branching random walk.* Ann. Probab. 41 (2013), no. 3A, 1362–1426. doi: 10.1214/12-AOP750.

[2] E. Aïdékon, J. Berestycki, É. Brunet and Z. Shi, *Branching Brownian motion seen from its tip*, Probab. Theory Relat. Fields (2013) 157: 405. doi: 10.1007/s00440-012-0461-0

[3] D. Aldous, *An introduction to covering problems for random walks on graphs*, Journal of Theoretical Probability January 1989, Volume 2, Issue 1, pp 87-89

[4] D. Aldous, *Probability Approximations via the Poisson Clumping Heuristic*, Applied Mathematical Sciences 77, Springer-Verlag, New York, 1989.

[5] L.-P. Arguin, D. Belius, and P. Bourgade. *Maximum of the characteristic polynomial of random unitary matrices.* Communications in Mathematical Physics 349.2 (2017): 703-751.

[6] L.-P. Arguin, D. Belius, P. Bourgade, M. Radziwill and K. Soundararajan, *Maximum of the Riemann zeta function on a short interval of the critical line.* arXiv preprint: 1612.08575 (2016).

[7] L.-P. Arguin, D. Belius and A.J. Harper, *Maxima of a randomized Riemann zeta function, and branching random walks.* The Annals of Applied Probability 27.1 (2017): 178-215.

[8] L.-P. Arguin, A. Bovier and N. Kistler, *The extremal process of branching Brownian motion*, Probab. Theory Relat. Fields (2013) 157: 535. doi: 10.1007/s00440-012-0464-x

[9] L.-P. Arguin, A. Bovier and N. Kistler, *The genealogy of extremal particles of Branching Brownian Motion.* CPAM, 64, 1647-1676 (2011)

[10] L.-P. Arguin and W. Tai, *Is the Riemann zeta function in a short interval a 1-RSB spin glass?.* arXiv preprint:1706.08462 (2017).

[11] A. D. Barbour, L. Holst and S. Janson, *Poisson Approximation (Oxford Studies in Probability)*, Clarendon Press. 1992.

[12] D. Belius, *Gumbel fluctuations for cover times in the discrete torus.* Probab. Theory Related Fields, 157(3-4):635-689, 2013.

[13] D. Belius and N. Kistler, *THE SUBLEADING ORDER OF TWO DIMENSIONAL COVER TIMES*, Probab. Theory Relat. Fields (2016)

[14] D. Belius, J. Rosen and O. Zeitouni, *Tightness for the Cover Time of compact two dimensional manifolds*, arXiv preprint, arXiv:1711.02845 [math.PR]

[15] M. Biskup, J. Ding and S. Goswami, *Return probability and recurrence for the random walk driven by two-dimensional Gaussian free field.* arXiv preprint:1611.03901 (2016).

[16] M. Biskup and O. Louidor, *Extreme local extrema of two-dimensional discrete Gaussian free field.* Communications in Mathematical Physics 345.1 (2016): 271-304.

[17] M. Biskup and O. Louidor, *Full extremal process, cluster law and freezing for two-dimensional discrete Gaussian Free Field.* arXiv preprint: 1606.00510 (2016).

[18] M. Biskup and O. Louidor, *On intermediate level sets of two-dimensional discrete Gaussian free field.* arXiv preprint:1612.01424 (2016).

[19] E. Bolthausen, J.-D. Deuschel and G. Giacomin. *Entropic repulsion and the maximum of the two-dimensional harmonic.* The Annals of Probability 29.4 (2001): 1670-1692.

[20] E. Bolthausen and N. Kistler, *On a nonhierarchical version of the Generalized Random Energy Model.* Ann. Appl. Prob. Vol. 16, No. 1, 1-14 (2006)

[21] A. Bovier and L. Hartung, *Extended convergence of the extremal process of branching Brownian motion.* Ann. Appl. Probab. 27 (2017), no. 3, 1756–1777. doi: 10.1214/16-AAP1244.

[22] A. Bovier and L. Hartung, *The extremal process of two-speed branching Brownian motion.* Electron. J. Probab. 19 (2014), paper no. 18, 28 pp. doi: 10.1214/EJP.v19-2982.

[23] A. Bovier and L. Hartung, *Variable Speed Branching Brownian Motion 1. Extremal Processes in the Weak Correlation Regime*, ALEA, Lat. Am. J. Probab. Math. Stat. 12 (1), 261-291 (2015)

[24] A. Bovier and I. Kurkova, *Derrida's Generalised Random Energy models 1: models with finitely many hierarchies.* Annals de l' Institut Henri Poincaré 40 (2004)

[25] A. Bovier and I. Kurkova *Derrida's generalized random energy models. 2. Models with continuous hierarchies.* Ann. Inst. H. Poincare. Prob. et Statistiques (B) Prob. Stat. 40, pp. 481-495 (2004)

[26] M. Bramson, *Convergence of solutions of the Kolmogorov equation to traveling waves*, Mem. Amer. Math. Soc. 44, no. 285, iv+190 pp. (1983)

[27] M. Bramson, *Maximal displacement of branching brownian motion.* Comm. Pure Appl. Math (1978). 31: 531-581. doi: 10.1002/cpa.3160310502

[28] M. Bramson, J. Ding and O. Zeitouni, *Convergence in Law of the Maximum of the Two-Dimensional Discrete Gaussian Free Field.* Communications on Pure and Applied Mathematics 69.1 (2016): 62-123.

[29] R. Chhaibi, J. Najnudel and T. Madaule, *On the maximum of the $C\beta E$ field.* arXiv preprint:1607.00243 (2016).

[30] A. Cortines, L. Hartung and O. Louidor, *The Structure of Extreme Level Sets in Branching Brownian Motion*, arXiv preprint, arXiv:1703.06529 [math.PR]

[31] A. Dembo, Y. Peres, J. Rosen, and O. Zeitouni, *Cover times for Brownian motion and random walks in two dimensions.* Ann. of Math. (2) 160 (2004), no. 2, 433-464. 60J65 (60G50)

[32] B. Derrida, *A generalization of the random energy model that includes correlations between the energies*, J.Phys.Lett. 46 (1985)

[33] B. Derrida, *Random energy model: An exactly solvable model of disordered systems.* Phys. Rev. B 24, 2613-2626 (1981)

[34] B. Derrida and H. Spohn, *Polymers on disordered trees, spin glasses, and travelling waves*, J. Statist. Phys. 51, no. 5-6, pp. 817-840 (1988)

[35] M. Fang and O. Zeitouni, *Slowdown for time inhomogeneous branching Brownian motion.* J. Stat. Phys., 149(1):1-9, 2012.

[36] C. Glenz, N. Kistler and M. A. Schmidt, *On McKean's martingale in the Bovier-Hartung extremal process*, arXiv preprint, arXiv:1712.04325 [math.PR].

[37] J. Goodman and F. Hollander, *Extremal geometry of a brownian porous medium.* Probab. Theory and Related Fields, 2013.

[38] F. Guerra, *Broken replica symmetry bounds in the mean field spin glass model.* Comm. Math. Phys. 233, no. 1, 1-12. 2003.

[39] O. Kallenberg, *Random Measures*, Akademie-Verlag. 1976.

[40] O. Kallenberg, *Random Measures, Theory and Applications.*, Springer International Publishing. 2017. DOI: 10.1007/978-3-319-41598-7.

[41] 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 and N. Kistler (eds), Springer Lecture Notes in Mathematics, Vol. 2143 (2015)

[42] N. Kistler and M. A. Schmidt, *From Derrida's random energy model to branching random walks: from 1 to 3.* Electron. Commun. Probab. 20 (2015), paper no. 47, 12 pp. doi: 10.1214/ECP.v20-4189.

[43] S.P. Lalley and T. Sellke, *A conditional limit theorem for the frontier of a branching brownian motion*, Ann. Probab. 15, no.3, pp. 1052-1061 (1987)

[44] G. F. Lawler, *On the covering time of a disc by simple random walk in two dimensions.* In Seminar on Stochastic Processes, 1992 (Seattle, WA, 1992), volume 33 of Progr. Probab., pages 189-207. Birkhäuser Boston, Boston, MA, 1993.

[45] P. Matthews. *Covering problems for Brownian motion on spheres.* Ann. Probab., 16(1):189-199, 1988.

[46] H.P. McKean, *Application of Brownian Motion to the equation of Kolmogorov-Petrovskii-Piskunov*, Comm. Pure. Appl. Math. 28: 323-331 (1976).

[47] M. Mezard, G. Parisi and M. Virasoro, *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications*, Volume 9: World Scientific Lecture Notes in Physics, World Scientific Publishing Company, 1987

[48] J. Najnudel, *On the extreme values of the Riemann zeta function on random intervals of the critical line.* arXiv preprint: 1611.05562 (2016).

[49] E. Paquette and O. Zeitouni, *The maximum of the CUE field.* International Mathematics Research Notices (2017): rnx033.

[50] D. Ray, *Sojourn times and the exact Hausdorff measure of the sample path for planar Brownian motion.* Transactions of the American Mathematical Society 106 (1963): 436-444.

[51] D. Ruelle, *A mathematical reformulation of Derrida's REM and GREM. Comm.* Math. Phys. 108, no. 2, 225 - 239. 1987.

[52] M. A. Schmidt, *A SIMPLE PROOF OF THE DPRZ-THEOREM FOR 2d COVER TIMES.* arXiv preprint: 1805.09744

[53] D. Zuckerman, *A technique for lower bounding the cover time*, SIAM J. Discrete Math. 5 (1992), 81-87.