The Annals of Applied Probability
2017, Vol. 27, No. 1, 178–215
DOI: 10.1214/16-AAP1201

MAXIMA OF A RANDOMIZED RIEMANN ZETA FUNCTION, AND BRANCHING RANDOM WALKS

BY LOUIS-PIERRE ARGUIN[*,†,1], DAVID BELIUS[‡,2] AND ADAM J. HARPER[§,3]

Université de Montréal[*], *City University of New York*[†], *New York University*[‡] *and Jesus College*[§]

A recent conjecture of Fyodorov–Hiary–Keating states that the maximum of the absolute value of the Riemann zeta function on a typical bounded interval of the critical line is $\exp\{\log\log T - \frac{3}{4}\log\log\log T + O(1)\}$, for an interval at (large) height T. In this paper, we verify the first two terms in the exponential for a model of the zeta function, which is essentially a randomized Euler product. The critical element of the proof is the identification of an approximate tree structure, present also in the actual zeta function, which allows us to relate the maximum to that of a branching random walk.

1. Introduction. The Riemann zeta function is defined for $\mathrm{Re}(s) > 1$ by a sum over integers, or equivalently by an *Euler product* over primes, as

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ primes}} (1 - p^{-s})^{-1}, \tag{1}$$

and by analytic continuation for other complex s. The behavior of the function on the critical line $\mathrm{Re}(s) = 1/2$ is a major theme in number theory, the most important questions of course concerning the zeroes (e.g., the Riemann hypothesis).

This paper is motivated by the study of the large values of $|\zeta(s)|$ on the critical line $s = 1/2 + it$. Little is known about the behavior on long intervals, say $0 \le t \le T$ for T large. The Lindelöf hypothesis, which is implied by the Riemann hypothesis, states that $\max_{0\le t\le T} |\zeta(1/2 + it)|$ grows slower than any small power of T. See the paper by Farmer, Gonek and Hughes [13] for more precise conjectures about this maximum size, and the paper of Soundararajan [26] and the recent work of Bondarenko and Seip [8] for rigorous lower bounds. More recently, Fyodorov, Hiary and Keating considered the maximum on bounded intervals of the critical line. They made the following conjecture:

Received June 2015; revised April 2016.

[1]Supported by NSF Grant DMS-15-13441 and PSC-CUNY Research Award 68784-00 46.

[2]Supported by the Swiss National Science Foundation, the Centre de Recherches Mathématiques and the Institut des Sciences Mathématiques, for part of the research leading up to this work.

[3]Supported by a research fellowship at Jesus College, Cambridge.

MSC2010 subject classifications. 60G70, 11M06.

Key words and phrases. Extreme value theory, Riemann zeta function, branching random walk.

CONJECTURE 1 (Fyodorov–Hiary–Keating [15, 16]). *For τ sampled uniformly from* $[0, T]$,

$$\max_{h\in[0,1]} \log\left|\zeta\left(1/2+i(\tau+h)\right)\right| = \log\log T - \frac{3}{4}\log\log\log T + O_P(1), \tag{2}$$

where $O_P(1)$ is a term that is stochastically bounded as $T \to \infty$.

The main result of this paper is a proof of the validity of the first two terms in (2) for a *random model* of ζ defined in (5) below, which is essentially a randomized Euler product. Until now, such precise estimates were not known rigorously even for models of zeta.

The conjecture is intriguing for many reasons. From a number theory point of view, the precision of the prediction is striking. From a probability point of view, the leading and subleading order of the maximum correspond exactly to those of the maximum of a branching random walk (which is a collection of correlated random walks indexed by the leaves of a tree), as will be explained below. In fact, the key element of the proof for the random model will be the identification of an approximate tree structure for the zeta function.

1.1. *Modeling the zeta function.* If we take logarithms and Taylor expand the Euler product formula for the zeta function, we find for $\mathrm{Re}(s) > 1$,

$$\log\zeta(s) = -\sum_p \log\left(1-p^{-s}\right) = \sum_{k=1}^{\infty}\frac{1}{k}\sum_p \frac{1}{p^{ks}} = \sum_p \frac{1}{p^s} + O(1), \tag{3}$$

since the total contribution from all proper prime powers (p^{ks} with $k \geq 2$) is uniformly bounded. One of the great challenges of analytic number theory is to understand how the influence of the Euler product may persist for general $s \in \mathbb{C}$. The definition of our random model is based on a rigorous result in that direction, assuming the truth of the Riemann hypothesis, proved by Harper [20] by adapting a method of Soundararajan [27] (which itself builds heavily on classical work of Selberg [25]).

PROPOSITION 1.1 (See Proposition 1 of Harper [20]). *Assume the Riemann hypothesis. For T large enough, there exists a set $\mathcal{H} \subseteq [T, T+1]$, of measure at least* 0.99, *such that*

$$\log\left|\zeta(1/2+it)\right| = \mathrm{Re}\left(\sum_{p\leq T}\frac{1}{p^{1/2+it}}\frac{\log(T/p)}{\log T}\right) + O(1) \qquad \forall t \in \mathcal{H}. \tag{4}$$

The set $\mathcal{H}$ produced in Proposition 1.1 consists of values t that are not abnormally close, in a certain averaged sense, to many zeros of the zeta function. It seems reasonable to think that one should not typically find maxima very close to zeros. Moreover, if one only wants an upper bound then the restriction to the set $\mathcal{H}$ can in fact be removed, at the cost of a slightly more complicated right-hand side. Therefore, to understand the typical size of $\max_{0\le h\le 1}\log|\zeta(1/2+i(\tau+h))|$ as τ varies we should try to understand the typical size of $\max_{0\le h\le 1}\sum_{p\le T}\mathrm{Re}(\frac{1}{p^{1/2+i(\tau+h)}}\frac{\log(T/p)}{\log T})$. The factor $\log(T/p)/\log T$ is a smoothing introduced for technical reasons. For simplicity, we shall ignore it in our model.

Since the values of $\log p$ are linearly independent for distinct primes, it is easy to check by computing moments that the finite-dimensional distributions of the process $(p^{-i\tau},\ p \text{ primes})$, where τ is sampled uniformly from $[0,T]$, converge as $T\to\infty$ to those of a sequence of independent random variables distributed uniformly on the unit circle. Following [20], this observation suggests to build a model from a probability space $(\Omega,\mathcal{F},\mathbb{P})$ with random variables $(U_p, p \text{ primes})$ which are uniform on the unit circle, and independent. For $T>0$ and $h\in\mathbb{R}$, we consider the random variables $\sum_{p\le T}p^{-1/2}\mathrm{Re}(U_p p^{-ih})$. In view of Proposition 1.1, the process

$$\left(\sum_{p\le T}\frac{\mathrm{Re}(U_p p^{-ih})}{p^{1/2}}, h\in[0,1]\right) \tag{5}$$

seems like a reasonable model for the large values of $(\log|\zeta(1/2+i(\tau+h))|, h\in[0,1])$.

1.2. *Main result.* In this paper, we provide evidence in favor of Conjecture 1 by proving a similar statement for the random model (5). At the same time, we hope to outline a possible approach to tackle the conjecture for the Riemann zeta function itself.

THEOREM 1.2. *Let $(U_p, p$ primes$)$ be independent random variables on $(\Omega,\mathcal{F},\mathbb{P})$, distributed uniformly on the unit circle. Then*

$$\max_{h\in[0,1]}\sum_{p\le T}\frac{\mathrm{Re}(U_p p^{-ih})}{p^{1/2}} = \log\log T - \frac{3}{4}\log\log\log T + o_P(\log\log\log T), \tag{6}$$

where the sum is over the primes less than or equal to T and the error term converges to 0 *in probability when divided by* $\log\log\log T$.

An outline of the proof of the theorem is given in Section 1.4 below. The technical tools needed are developed in Section 2, and finally the proof is given in Section 3.

1.3. *Relations to previous results.* The leading order term $\log\log T$ in (6) was proved in [20], where it was also shown that the second-order term must lie between $-2\log\log\log T$ and $-(1/4)\log\log\log T$. As well as giving a stronger result, our analysis here is ultimately based on a control of the joint distribution of only two points h_1 and h_2 of the random process at a time, which could feasibly be achieved for the zeta function itself. In contrast, the lower bound analysis in [20] depends on a Gaussian comparison inequality that requires control of $\log T$ points.

Fyodorov, Hiary and Keating motivated Conjecture 1 in [15, 16] using a connection to random matrices. There is convincing evidence (see, e.g., [21]) that the values of the zeta function in an interval of the critical line are well modeled by the characteristic polynomial $P_N(x)$ of an $N \times N$ matrix sampled uniformly from the unitary group, for $x = e^{i\theta}$ on the unit circle. In this spirit, they compute in [16] the moments of the partition function $Z_N(\beta) = \int_0^{2\pi} |P_N(e^{i\theta})|^\beta \, d\theta$. They argue that these coincide with those previously obtained for a logarithmically correlated Gaussian field [14]. For large β, this leads to the conjecture that the maximum of the characteristic polynomial behaves like the maximum of the Gaussian model. Unfortunately, the analogue of Conjecture 1 for this random matrix model is not known rigorously even to leading order (see [29] for recent developments at low β and its relation to Gaussian multiplicative chaos). The conjecture is also expected to hold for other random matrix models such as the Gaussian Unitary Ensembles; see [17]. One advantage of the model (5) is that it can be analysed rigorously to a high level of precision with current probabilistic techniques.

As explained in Section 1.4, the proof of Theorem 1.2 uses in a crucial way an *approximate tree structure* present in our model and also in the actual zeta function. This structure explains the observed agreement between the high values of the zeta function and those of log-correlated random fields. The approach to control subleading orders of log-correlated Gaussian fields and branching random walks was first developed by Bramson [10] in his seminal work on the maximum of branching Brownian motion. It has since been extended to more general branching random walks by several authors, for example, [1, 2, 11], and to log-correlated Gaussian fields see, for example, [12, 23]. This type of argument can also be applied to obtain the joint distribution of the near-maxima; see, for example, [3, 4, 7]. Recently, Kistler introduced a multiscale refinement of the second moment method to control the maximum of processes with neither a priori Gaussianity nor exact tree structure, to leading and subleading order. It was successfully implemented in [5] to obtain the subleading order of cover times on the two-dimensional torus. The proof of Theorem 1.2 follows the same approach.

It is instructive to consider the conjecture in the light of the statistics of typical values of the zeta function. One beautiful result is the *Selberg central limit theorem* [25], which asserts that if τ is sampled uniformly from the interval $[0, T]$ then $(\frac{1}{2}\log\log T)^{-1/2} \log|\zeta(1/2 + i\tau)|$ converges in law to a standard Gaussian variable. Thus, to obtain a rough prediction for the order of the maximum on $[0, 1]$, one may compare it to the maximum of independent Gaussian variables of mean 0 and variance $\frac{1}{2}\log\log T$. For $\log T$ such variables, it is not hard to show that the order of the maximum is $\log\log T - \frac{1}{4}\log\log\log T + O(1)$. The leading order agrees with Conjecture 1, but the constant in the subleading correction is different.

Our proof shows how to modify this "independent" heuristic to account for the "extra" $-\frac{1}{2}\log\log\log T$ present in Conjecture 1. Bourgade showed a multivariate version of Selberg's theorem where the correlations are logarithmic in the limit [9]. However, the convergence is too weak to describe the maximum on an interval.

1.4. *Outline of the proof.* The proof of Theorem 1.2 is based on an analogy between the process (5) and a branching random walk (also known as hierarchical random field). We make this connection precise here, and indicate for an unfamiliar reader how to analyse the maximum of a branching random walk.

We will work in the case where $T = e^{2^n}$ for some large natural number n. In this setup, the process of interest in Theorem 1.2 is

$$(X_n(h), h \in [0,1]) \qquad \text{where } X_n(h) = \sum_{p \le e^{2^n}} \frac{\mathrm{Re}(U_p p^{-ih})}{p^{1/2}} \tag{7}$$

is a continuous function of h. Since $\log\log T = n\log 2$ and $\log\log\log T = \log n + O(1)$, Theorem 1.2 can be restated as

$$\lim_{n\to\infty} \mathbb{P}\Big[m_n(-\varepsilon) \le \max_{h\in[0,1]} X_n(h) \le m_n(\varepsilon)\Big] = 1 \qquad \text{for all } \varepsilon > 0, \tag{8}$$

$$\text{where } m_n(\varepsilon) = n\log 2 - \frac{3}{4}\log n + \varepsilon \log n. \tag{9}$$

In other words, with large probability, the maximum of the process lies in an arbitrarily small window (of order $\log n$) around $n\log 2 - \frac{3}{4}\log n$.

By symmetry of U_p we have $\mathbb{E}[X_n(h)] = 0$ for any $h \in [0,1]$. Also a simple computation shows that $\mathbb{E}[\mathrm{Re}(U_p p^{-ih})\,\mathrm{Re}(U_p p^{-ih'})] = (1/2)\cos(|h-h'|\log p)$, so the covariance $\mathbb{E}[X_n(h)X_n(h')]$ equals $\frac{1}{2}\sum_{\log p \le 2^n} p^{-1}\cos(|h-h'|\log p)$. Using well-known results on primes (cf. Lemma 2.1), it is possible to estimate this as

$$\mathbb{E}\big[X_n(h)X_n(h')\big] \approx \frac{1}{2}\log|h-h'|^{-1}, \tag{10}$$

for any $h, h' \in [0,1]$ provided $|h-h'| \ge 2^{-n}$. If instead $|h-h'| < 2^{-n}$, then the covariance is almost $n(\log 2)/2$, that is, $X_n(h)$ and $X_n(h')$ are almost perfectly correlated. Therefore, one can think of the maximum over $h \in [0,1]$ as a maximum over 2^n equally spaced points.

The key point of the proof is that the logarithmic nature of the correlations can be understood in a more structural way using a *multiscale decomposition*. Precisely, we rewrite the process as

$$X_n(h) = \sum_{k=0}^{n} Y_k(h) \qquad \text{where } Y_k(h) = \sum_{2^{k-1} < \log p \le 2^k} \frac{\mathrm{Re}(U_p p^{-ih})}{p^{1/2}} \tag{11}$$

is the *increment at "scale" k* of $X_n(h)$. It is not hard to show (see Section 2.1) that for k large,

$$\mathbb{E}[Y_k(h)^2] \approx \frac{\log 2}{2} \quad \text{and}$$

$$\mathbb{E}[Y_k(h)Y_k(h')] \approx \begin{cases} \dfrac{\log 2}{2} & \text{if } |h-h'| \le 2^{-k}, \\ 0 & \text{if } |h-h'| > 2^{-k}. \end{cases} \tag{12}$$

In view of (12), for given h, h', one can think of the partial sums $X_k(h) = \sum_{j=1}^k Y_j(h)$ and $X_k(h') = \sum_{j=1}^k Y_j(h')$ as random walks, where the increments $Y_j(h), Y_j(h')$ are almost perfectly correlated (so roughly the same) for those j such that $2^j \le |h-h'|^{-1}$, and where they are almost perfectly decorrelated (so essentially independent) when $2^j > |h-h'|^{-1}$. A similar, but exact, behavior would be obtained as follows: Consider 2^n equally spaced points in [0, 1], thought of as leaves of a binary tree of depth n. Place on each edge of the binary tree an independent Gaussian with mean zero and variance $(\log 2)/2$, and associate to a leaf the random walk given by the partial sums of the Gaussians on the path from root to leaf; see Figure 1. With this construction, the first k increments of the random walks of two leaves will be exactly the same, where k is the level of the most recent common ancestor, and the rest of the increments will be perfectly independent. This tree construction is an example of branching random walk. For the model (7) of zeta, the *branching point k* where the paths $X_k(h)$ and $X_k(h')$ roughly decorrelate is

$$h \wedge h' := \lfloor \log_2 |h-h'|^{-1} \rfloor. \tag{13}$$

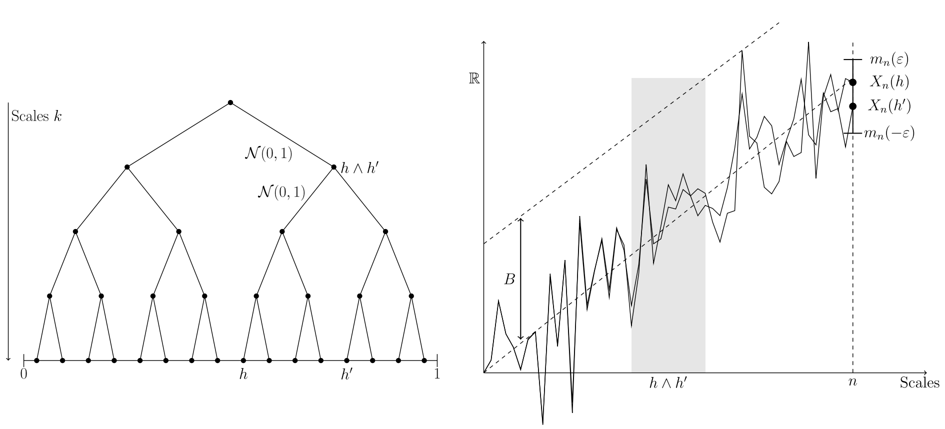

FIG. 1. (*Left*) *An illustration of the correlation structure of a branching random walk.* (*Right*) *A realization of two paths of an approximate branching random walk with increments almost equal before the branching point $h \wedge h'$ and almost decoupled after. The barrier below which the paths must stay is also shown.*

So h and h' correspond to leaves whose most recent common ancestor is in level $k = h \wedge h'$. We note that the different nature of the correlations for different ranges of p was already exploited in early work of Halász [19], although without drawing any connection to branching. Conversely, in Bourgade's work on the multivariate central limit theorem for zeta, the branching nature of the correlations was observed in the limit $T \to \infty$ but without the connection to different ranges of p.

A compelling method to analyse the maximum of a branching random walk, and of log-correlated processes in general, is a *multiscale refinement of the second moment method* as proposed in [22], which we shall implement in the approximate branching setting described above. First, proceeding naïvely, one could consider the number of variables whose value exceeds a given value m, that is, the *number of exceedances*,

$$Z(m) = \#\{j \le 2^n : X_n(j/2^n) \ge m\}. \tag{14}$$

Clearly, $\max_{j \le 2^n} X_n(j/2^n) \ge m$ if and only if $Z(m) \ge 1$. Thus, an upper bound for the maximum can be obtained by the union bound

$$\mathbb{P}(Z(m) \ge 1) \le \mathbb{E}[Z(m)] = 2^n \mathbb{P}(X_n(0) \ge m). \tag{15}$$

On the other hand, a lower bound can be obtained by the Paley–Zygmund inequality,

$$\mathbb{P}(Z(m) \ge 1) \ge \frac{\mathbb{E}[Z(m)]^2}{\mathbb{E}[Z(m)^2]}. \tag{16}$$

More precisely, one would choose $m = m(n)$ large enough in (15) so that $\mathbb{E}[Z(m)] = o(1)$, and m small enough in (16) so that $\mathbb{E}[Z]^2 = (1 + o(1))\mathbb{E}[Z^2]$, and thus $\mathbb{P}(Z(m) \ge 1) = 1 + o(1)$. For this, one needs large deviation estimates: if we think of $X_n(h)$ as Gaussian with variance $n \log 2/2$, then a standard Gaussian estimate yields that $P(X_n(h) \ge m)$ is approximately $\frac{\sqrt{n}}{m} e^{-m^2/((\log 2)n)}$. Thus, $2^n P(X_n(0) \ge m) = o(1)$ when $m = (\log 2)n - \frac{1}{4} \log n + \varepsilon \log n$. This would in fact be the correct answer (the union bound would be sharp) if the random variables $X_n(j/2^n)$ were independent. However, if $m = (\log 2)n - \frac{3}{4} \log n + \varepsilon \log n$, then $2^n \mathbb{P}(X_n(0) \ge m) \ge cn^{1-\varepsilon} \to \infty$, so (15) cannot prove the upper bound we seek in Theorem 1.2. Similarly, the right-hand side of (16) will tend to zero unless $m \le \frac{\log 2}{2} n$, since strong correlation between exceedance events for nearby h, h' inflates the second moment. Thus, the lower bound obtained is not close to what we seek even to leading order.

To get good bounds, one needs to modify the definition of the number of exceedances using an insight from the underlying approximate tree structure. For branching random walk there are exactly 2^k distinct partial sums up to the k-level, one for each vertex at that level. By analogy one expects that the "variation" in $X_k(h)$ (i.e., in the partial sums up to the kth level) for different $h \in [0, 1]$ should be captured by just 2^k equally spaced points in $[0, 1]$. Even if they were independent, it would be very unlikely that one of these 2^k values exceeded $k \log 2 + B$, for $B > 0$ growing slowly with n, and it turns out that positive correlations only make it less likely. In other words, with high probability, all random walks $X_k(h)$ must lie below the *barrier* $k \mapsto k \log 2 + B$. This suggests to look at the modified number of exceedances

$$\widetilde{Z}(m) = \#\{j \leq 2^n : X_n(j/2^n) \geq m, X_k(j/2^n) < k \log 2 + B, \forall k \leq n\}. \tag{17}$$

It turns out that replacing Z by $\widetilde{Z}$ in the first moment bound (15) and (with slight modifications) in the second moment bound (16) will yield the correct answer. To see this in the former case, we write the first moment by conditioning on the end point:

$$\begin{aligned}\mathbb{E}[\widetilde{Z}(m)] = {} & 2^n \mathbb{P}(X_n(0) > m) \\ & \times \mathbb{P}(X_k(0) < k \log 2 + B, \forall k \leq n \mid X_n(0) > m).\end{aligned} \tag{18}$$

By the earlier naïve discussion, the first two terms amount to $O(n^{1-\varepsilon})$ when we set $m = n \log 2 - \frac{3}{4} \log n + \varepsilon \log n$. The third term is the probability that a *random walk bridge* starting at 0 and ending at $m = n \log 2 - \frac{3}{4} \log n + \varepsilon \log n$ avoids the barrier $k \log 2 + B$. This probability turns out to be n^{-1}, as shown by the *ballot theorem*; cf. Lemma 2.12. Therefore, $\mathbb{E}[\widetilde{Z}(m)] = O(n^{-\varepsilon}) \to 0$, for all $\varepsilon > 0$. A similar analysis can be done for the lower bound, where we have the obvious inequality $\mathbb{P}(Z(m) \geq 1) \geq \mathbb{P}(\widetilde{Z}(m) \geq 1)$. The extra barrier condition turns out to reduce correlations between exceedance events sufficiently so that the second moment is now essentially the first moment squared when $m \leq n \log 2 - \frac{3}{4} \log n - \varepsilon \log n$. [This analysis also indicates why the second moment of $Z(m)$ is too large: in the exponentially unlikely event that a path manages to go far above the barrier, it has exponentially many "offspring" that end up far above the typical level of the maximum.]

The form of the subleading correction is thus explained by the extra "cost" n^{-1} of satisfying the barrier condition. And the barrier condition arises because of "tree-like" correlations present in the values of (the model of) the zeta function. This suggests the possibility that the partial sums of the Euler product (3) of the actual zeta function behave similarly where the zeta function is large.

To prove Theorem 1.2, we must address the imprecisions in the above discussion. The necessary large deviation estimates are derived in Section 2.1. The claim that $X_k(h)$ does not vary much below scale 2^{-k} is proved in Section 2.2 using a *chaining argument*. Another issue is that our process is not exactly a branching random walk because increments are never perfectly independent (for different h, h') nor exactly identical. To deal with this, we use a Berry–Esseen approximation in Section 2.3 to show that the random walks are very close to being Gaussian. This then allows for an explicit comparison with Gaussian random walks with i.i.d. increments. Moreover, to get a sharp lower bound with the second moment method, it is necessary to "cut off the first r scales" and consider

$$X_{r,k}(h) = X_k(h) - X_r(h) \qquad \text{for } h \in \mathbb{R}, \tag{19}$$

for an appropriately chosen r growing slowly with n. We note that in order to improve the error term in (6) to order one, it would also be necessary to take r of order one relative to n, which would complicate the analysis.

Finally, it should be stressed that the approach relies only on controlling first and second moments, which means that the estimates we need only involve at most two random walks simultaneously.

2. Preliminaries. Throughout the paper, we will write c for absolute constants whose value may change at different occurrences. A sum over the variable p always denotes a sum over primes.

2.1. *Large deviation estimates.* In this section, we derive the large deviation properties of the increments $(Y_k(h), h \in [0, 1])$ and their sum. We first derive basic facts on their distribution and in particular on their correlations.

Recall that the random variables $(U_p, p \text{ primes})$ are i.i.d. and uniform on the unit circle. For simplicity, we denote the pth term of the sum over primes in (6) by

$$W_p(h) = \frac{\mathrm{Re}(U_p p^{-ih})}{p^{1/2}}, \qquad h \in \mathbb{R}. \tag{20}$$

Note that the law of the process $(W_p(h), h \in \mathbb{R})$ is translation-invariant on the real line and also invariant under the reflection $h \mapsto -h$. A straightforward computation using the law of the U_p's and translation invariance gives

$$\mathbb{E}[W_p(h)W_p(h')] = \frac{1}{2p}\cos(|h-h'|\log p) \qquad \text{for all } h, h'. \tag{21}$$

In this notation, the increments defined in (11) are

$$Y_k(h) = \sum_{2^{k-1} < \log p \le 2^k} W_p(h), \qquad h \in \mathbb{R}. \tag{22}$$

Using (21) and the independence of the U_p's, the variance of $Y_k(h)$ becomes

$$\sigma_k^2 = \mathrm{Var}(Y_k(h)) = \sum_{2^{k-1} < \log p \le 2^k} \frac{1}{2p}, \tag{23}$$

and the covariance of $Y_k(h)$ and $Y_k(h')$ is

$$\rho_k(h, h') = \mathbb{E}[Y_k(h)Y_k(h')] = \sum_{2^{k-1} < \log p \le 2^k} \frac{1}{2p}\cos(|h-h'|\log p). \tag{24}$$

The next lemma formalizes (12), giving bounds on how close the variance of the increments is to

$$\sigma^2 = (\log 2)/2, \tag{25}$$

and (for $h \neq h'$) giving bounds on how close the covariance is to the variance before the "branching point" $h \wedge h'$ [defined in (13)], and on how fast the covariance decays after.

LEMMA 2.1. *For $h, h' \in \mathbb{R}$ and $k \geq 1$,*

$$\sigma_k^2 = \mathbb{E}\big[Y_k(h)^2\big] = \sigma^2 + O\big(e^{-c\sqrt{2^k}}\big), \tag{26}$$

$$\begin{aligned} \rho_k(h,h') &= \mathbb{E}\big[Y_k(h)Y_k(h')\big] \\ &= \begin{cases} \sigma^2 + O\big(2^{-2(h\wedge h'-k)}\big) + O\big(e^{-c\sqrt{2^k}}\big) & \text{if } k \leq h\wedge h', \\ O\big(2^{-(k-h\wedge h')}\big) & \text{if } k > h\wedge h'. \end{cases} \end{aligned} \tag{27}$$

Note that in both cases the error term decays exponentially in k.

PROOF. We use a strong form of the prime number theorem (see Theorem 6.9 of [24]) which states that

$$\#\{p \leq x : p \text{ prime}\} = \int_2^x \frac{1}{\log u}\,du + O\big(xe^{-c\sqrt{\log x}}\big). \tag{28}$$

By replacing the sum $\sum_{P<p\leq Q} \frac{1}{p}$ with the integral $\int_P^Q \frac{1}{u\log u}\,du$ using (28) and integration by parts, one obtains

$$\sum_{P<p\leq Q} \frac{1}{p} = \log\log Q - \log\log P + O\big(e^{-c\sqrt{\log P}}\big) \qquad \text{for all } 2 \leq P \leq Q.$$

This together with (23) yields (26). Similarly, (28) implies that

$$\rho_k(h,h') = \frac{1}{2}\int_{e^{2^{k-1}}}^{e^{2^k}} \frac{\cos(|h-h'|\log u)}{u\log u}\,du + O\big((1+|h-h'|)e^{-c\sqrt{2^{k-1}}}\big).$$

When $2^k|h-h'| = 2^{k-h\wedge h'} \leq 1$, the claim (27) follows by using that $\cos(|h - h'|\log u) = 1 + O(|h-h'|^2(\log u)^2)$. When $2^{-k}|h-h'|^{-1} = 2^{-k+h\wedge h'} < 1$, we use integration by parts. After the change of variable $v = \log u$, the integral becomes

$$\frac{\sin(|h-h'|v)}{|h-h'|v}\bigg|_{2^{k-1}}^{2^k} + \int_{2^{k-1}}^{2^k} \frac{\sin(|h-h'|v)}{|h-h'|v^2}\,du.$$

Both terms are $O(2^{-k}|h-h'|^{-1})$. $\square$

REMARK 1. A similar but easier argument using (28) shows that

$$\sum_{P<p\leq Q} \frac{(\log p)^m}{p} = O\big((\log Q)^m\big) \qquad \text{for all } 1 \leq P \leq Q. \tag{29}$$

The main results of this section are explicit expressions for the cumulant generating functions of the increments, from which we will deduce large deviation estimates. For fixed $h, h' \in \mathbb{R}$, we will often drop the dependence on h and h' when it is clear from context and define

$$\boldsymbol{Y}_k = \big(Y_k(h), Y_k(h')\big).$$

The covariance matrix of $\boldsymbol{Y}_k$ is then denoted by

$$\boldsymbol{\Sigma}_k = \mathrm{Cov}(\boldsymbol{Y}_k) = \begin{pmatrix} \sigma_k^2 & \rho_k \\ \rho_k & \sigma_k^2 \end{pmatrix}.$$

The eigenvalues of Σ_k are $\sigma_k^2 \pm \rho_k$.

The cumulant generating functions are

$$(30) \qquad \psi_k^{(1)}(\lambda) = \log \mathbb{E}[\exp(\lambda Y_k)], \qquad \psi_k^{(2)}(\boldsymbol{\lambda}) = \log \mathbb{E}[\exp(\boldsymbol{\lambda} \cdot \boldsymbol{Y}_k)],$$

where $\lambda \in \mathbb{R}$, $\boldsymbol{\lambda} \in \mathbb{R}^2$ and "$\cdot$" is the inner product in $\mathbb{R}^2$. The following change of measure will also be needed in the proof of Theorem 1.2:

$$(31) \qquad \begin{aligned} \frac{d\mathbb{Q}_\lambda}{d\mathbb{P}} &= \prod_{k=1}^{n} \frac{e^{\lambda Y_k}}{e^{\psi_k^{(1)}(\lambda)}} \qquad \text{for } \lambda \in \mathbb{R}, \\ \frac{d\mathbb{Q}_{\boldsymbol{\lambda}}}{d\mathbb{P}} &= \prod_{k=1}^{n} \frac{e^{\boldsymbol{\lambda} \cdot \boldsymbol{Y}_k}}{e^{\psi_k^{(2)}(\boldsymbol{\lambda})}} \qquad \text{for } \boldsymbol{\lambda} \in \mathbb{R}^2. \end{aligned}$$

Recall that in the univariate case,

$$(32) \qquad \mathbb{Q}_\lambda[Y_k] = \frac{d}{d\lambda}\psi_k^{(1)}(\lambda), \qquad \mathrm{Var}_{\mathbb{Q}_\lambda}(Y_k) = \frac{d^2}{d\lambda^2}\psi_k^{(1)}(\lambda),$$

and in the multivariate case,

$$(33) \qquad \mathbb{Q}_{\boldsymbol{\lambda}}[\boldsymbol{Y}_k] = \nabla \psi_k^{(2)}(\boldsymbol{\lambda}), \qquad \mathrm{Cov}_{\mathbb{Q}_{\boldsymbol{\lambda}}}(\boldsymbol{Y}_k) = \mathrm{Hess}\, \psi_k^{(2)}(\boldsymbol{\lambda}).$$

The results also provide bounds on these quantities. We first state the result for the univariate case. The proof is omitted since it is a special case of the multivariate bound in Proposition 2.4.

PROPOSITION 2.2. *Let $C > 0$. For all $0 < \lambda < C$ and k large enough (depending on C), the cumulant generating function $\psi_k^{(1)}(\lambda)$ satisfies*

$$(34) \qquad \psi_k^{(1)}(\lambda) = \frac{\lambda^2 \sigma_k^2}{2} + O(e^{-2^{k-1}}).$$

Moreover, for such k, the measure $\mathbb{Q}_\lambda$ in (31) satisfies

$$(35) \qquad \mathbb{Q}_\lambda[Y_k] = \lambda \sigma_k^2 + O(e^{-2^{k-1}}), \qquad \mathrm{Var}_{\mathbb{Q}_\lambda}[Y_k] = \sigma_k^2 + O(e^{-2^{k-1}}).$$

One useful consequence of the proposition is the following one-point large deviation estimate, which (after being strengthened to a bound for the maximum over a small interval, in Corollary 2.6) will be crucial to impose the barrier in Lemma 3.4. Recall from (19) that $X_{r,k}(h) = X_k(h) - X_r(h) = \sum_{l=r+1}^{k} Y_l(h)$.

COROLLARY 2.3. *Let $C > 0$. For any $0 \le r \le k-1$, $0 < x < C(k-r)$ and $h \in \mathbb{R}$,*

$$\mathbb{P}\big[X_{r,k}(h) > x\big] \le c\exp\biggl(-\frac{x^2}{2(k-r)\sigma^2}\biggr), \tag{36}$$

where the constant c depends on C.

PROOF. Using the exponential Chebyshev's inequality, the probability in (36) is bounded above by $\exp(\sum_{l=r+1}^{k}\psi_l^{(1)}(\lambda)-\lambda x)$, for all $\lambda > 0$. By Proposition 2.2 (with, say, $10C$ in place of C), we get that if $\lambda \le 10C$,

$$\begin{aligned}\mathbb{P}\big[X_{r,k}(h) > x\big] &\le \exp\Biggl(c+\frac{\lambda^2}{2}\sum_{l=r+1}^{k}\sigma_l^2-\lambda x+O\bigl(e^{-c2^r}\bigr)\Biggr)\\ &\le c\exp\Biggl(\frac{\lambda^2}{2}\sum_{l=r+1}^{k}\sigma_l^2-\lambda x\Biggr)\le c\exp\biggl(\frac{\lambda^2}{2}(k-r)\sigma^2-\lambda x\biggr),\end{aligned}$$

where (26) is used in the last inequality. If l is too small for (34) to be applied, we simply use that $\psi_l(\lambda)$ is bounded. Setting $\lambda = x((k-r)\sigma^2)^{-1} \le 10C$ gives the result. $\square$

We now prove the bounds in the multivariate case.

PROPOSITION 2.4. *Let $C > 0$. For all $\boldsymbol{\lambda} = (\lambda, \lambda')$, where $0 < \lambda, \lambda' < C$, and k large enough (depending on C), the cumulant generating function $\psi_k^{(2)}(\boldsymbol{\lambda})$ satisfies*

$$\psi_k^{(2)}(\boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{\lambda}\cdot\boldsymbol{\Sigma}_k\boldsymbol{\lambda}+O\bigl(e^{-2^{k-1}}\bigr). \tag{37}$$

Moreover, for such k, the measure $\mathbb{Q}_{\boldsymbol{\lambda}}$ in (31) satisfies

$$\mathbb{Q}_{\boldsymbol{\lambda}}[\boldsymbol{Y}_k] = \boldsymbol{\Sigma}_k\boldsymbol{\lambda}+O\bigl(e^{-2^{k-1}}\bigr) \quad \textit{and} \quad \mathrm{Cov}_{\mathbb{Q}_{\lambda}}[\boldsymbol{Y}_k] = \boldsymbol{\Sigma}_k+O\bigl(e^{-2^{k-1}}\bigr). \tag{38}$$

PROOF. We first compute

$$\begin{aligned}\psi_p^W(\boldsymbol{\lambda}) &= \log\mathbb{E}\bigl[\exp\bigl(\lambda W_p(0)+\lambda' W_p\bigl(|h-h'|\bigr)\bigr)\bigr]\\ &= \log\frac{1}{2\pi}\int_0^{2\pi}\exp\biggl(\frac{\lambda}{p^{1/2}}\cos(\theta)\\ &\qquad+\frac{\lambda'}{p^{1/2}}\cos\bigl(\theta+|h-h'|\log p\bigr)\biggr)\,d\theta.\end{aligned} \tag{39}$$

Recall that for any $a, b \in \mathbb{R}$,

$$\frac{1}{2\pi}\int_0^{2\pi}\exp\bigl(a\cos(\theta)+b\sin(\theta)\bigr)\,d\theta = I_0\bigl(\sqrt{a^2+b^2}\bigr), \tag{40}$$

where I_0 denotes the Bessel function of the first kind. The identity $\cos(\theta+\eta)=\cos(\theta)\cos(\eta)-\sin(\theta)\sin(\eta)$ can be used together with (40) to write the integral in the bottom line of (39) as

$$\begin{aligned}
&I_0\Bigg(\sqrt{\frac{1}{p}\big(\lambda+\cos(|h-h'|\log p)\lambda'\big)^2+\frac{1}{p}\big(\sin(|h-h'|\log p)\lambda'\big)^2}\Bigg)\\
&\quad=I_0\Bigg(\sqrt{\frac{1}{p}\big(\lambda^2+2\lambda\lambda'\cos(|h-h'|\log p)+\lambda'^2\big)}\Bigg)\\
&\quad=I_0(\sqrt{2\boldsymbol{\lambda}\cdot\boldsymbol{M}_p\boldsymbol{\lambda}}),
\end{aligned}\tag{41}$$

where

$$\boldsymbol{M}_p=\frac{1}{2p}\begin{pmatrix}1 & \cos(|h-h'|\log p)\\ \cos(|h-h'|\log p) & 1\end{pmatrix},$$

is the covariance matrix of $(W_p(h),W_p(h'))$, see (21). Thus, writing

$$f(x)=\log I_0(\sqrt{2x}),\tag{42}$$

we have $\psi_p^W(\boldsymbol{\lambda})=f(\boldsymbol{\lambda}\cdot\boldsymbol{M}_p\boldsymbol{\lambda})$. Recall that $I_0(x)$ has Taylor expansion $I_0(x)=1+\frac{x^2}{4}+\frac{x^4}{64}+O(x^6)$ [which can be verified by expanding in (40)], so that f has Taylor expansion

$$f(x)=\frac{x}{2}-\frac{x^2}{16}+O(x^3).\tag{43}$$

Now since the random variables U_p are independent,

$$\psi_k^{(2)}(\boldsymbol{\lambda})=\sum_{2^{k-1}<\log p\le 2^k}\psi_p^W(\boldsymbol{\lambda})=\sum_{2^{k-1}<\log p\le 2^k}f(\boldsymbol{\lambda}\cdot\boldsymbol{M}_p\boldsymbol{\lambda}).$$

The bound (43) implies that for k large enough (depending on C),

$$\psi_k^{(2)}(\boldsymbol{\lambda})=\sum_{2^{k-1}<\log p\le 2^k}\Big(\frac{1}{2}\boldsymbol{\lambda}\cdot\boldsymbol{M}_p\boldsymbol{\lambda}+O(p^{-2})\Big)=\frac{1}{2}\boldsymbol{\lambda}\cdot\boldsymbol{\Sigma}_k\boldsymbol{\lambda}+O(e^{-2^{k-1}}).$$

This proves (37).

The first claim of (38) follows similarly after noting that the gradient of the map $\boldsymbol{\lambda}\to f(\boldsymbol{\lambda}\cdot\boldsymbol{M}_p\boldsymbol{\lambda})$ is $\boldsymbol{M}_p\boldsymbol{\lambda}f'(\boldsymbol{\lambda}\cdot\boldsymbol{M}_p\boldsymbol{\lambda})$, and using the bound $f'(x)=\frac{1}{2}+O(x)$, valid for $x\in[0,1]$. Finally, the second claim of (38) follows by noting that the Hessian of the aforementioned map is

$$\boldsymbol{M}_p f'(\boldsymbol{\lambda}\cdot\boldsymbol{M}_p)+(\boldsymbol{M}_p\boldsymbol{\lambda})(\boldsymbol{M}_p\boldsymbol{\lambda})^T f''(\boldsymbol{\lambda}\cdot\boldsymbol{M}_p\boldsymbol{\lambda}),$$

and using the previous bound for $f'(x)$, and that $f''(x)$ is bounded in $[0,1]$. □

2.2. *Continuity estimates*. The main result of this section is a maximal inequality which shows that the maximum over an interval of length 2^{-k} of the field $X_{r,k}(h)$ is close to the value of the field at the mid-point of the interval, where $X_{r,k}(h)$ is defined in (19). One of the upshots is to reduce the proof of the upper bound of the maximum of the process on [0, 1] to an upper bound on the maximum over a discrete set of points in Section 3.1.

PROPOSITION 2.5. *Let $C > 0$. For any $0 \le r \le k-1$, $0 \le x \le C(k-r)$, $2 \le a \le 2^{2k} - x$ and $h \in \mathbb{R}$,*

$$\mathbb{P}\Big[\max_{h':|h'-h|\le 2^{-k-1}} X_{r,k}(h') > x + a, X_{r,k}(h) \le x\Big] \tag{44}$$
$$\le c \exp\Big(-\frac{x^2}{2(k-r)\sigma^2} - ca^{3/2}\Big),$$

where the constants c depend on C.

The proof of the proposition is postponed until the end of the section. It is based on a *chaining argument* and an estimate on joint large deviations of $X_{r,k}(h)$ and of the difference $X_{r,k}(h') - X_{r,k}(h)$ for $|h'-h| \le 2^{-k-1}$, see Lemma 2.7 below. The exponent of the a term is probably not optimal. A direct consequence of the proposition is the following large deviation bound of the maximum of $X_k(h)$ over an interval of length 2^{-k}.

COROLLARY 2.6. *Let $C > 0$. For any $0 \le r \le k-1$, $h \in \mathbb{R}$ and $0 \le x \le C(k-r)$,*

$$\mathbb{P}\Big[\max_{h':|h'-h|\le 2^{-k-1}} X_{r,k}(h') > x\Big] \le c \exp\Big(-\frac{x^2}{2(k-r)\sigma^2}\Big), \tag{45}$$

where the constant c depends on C.

PROOF. The left-hand side of (45) is at most

$$\mathbb{P}\Big[\max_{h':|h'-h|\le 2^{-k-1}} X_{r,k}(h') > (x-2) + 2, X_{r,k}(h) \le x - 2\Big]$$
$$+ \mathbb{P}\big[X_{r,k}(h) > x - 2\big].$$

The bound follows by (44) with $x-2$ in place of x and $a = 2$, and the bound (36). □

REMARK 2. A union bound over 2^n intervals of length 2^{-n} yields

$$\mathbb{P}\Big[\max_{h\in[0,1]} X_n(h) \ge (1+\delta)n \log 2\Big] \le c2^{-n\delta} \qquad \text{for all } \delta > 0, \tag{46}$$

where (45) is used with $r = 0$ and $k = n$ [note that $X_n(h) = Y_0(h) + X_{0,n}(h)$ and Y_0 is bounded]. This proves that $\max_{h\in[0,1]} X_n(h)$ is at most $(1 + o(1))n \log 2$, which is tight to leading order, but does not include the subleading correction present in (8) and (9).

To prove Proposition 2.5 we will use the following large deviation estimate for $X_{r,k}(0)$ and the difference $X_{r,k}(h_2) - X_{r,k}(h_1)$ (jointly), where $|h_2 - h_1| \le 2^{-k}$. It shows that on a large deviation scale the two quantities are essentially independent, and that the difference decays rapidly with $|h_2 - h_1|$. The latter is a consequence of the covariance of the field $X_{r,k}(h)$ losing its log-correlation structure below scale 2^{-k}, and instead decaying linearly with distance.

LEMMA 2.7. *Let $C > 0$. For any $0 \le r \le k-1$, $0 \le x \le C(k-r)$, $0 \le y \le 2^{2k}$ and any distinct $-2^{-k-1} \le h_1, h_2 \le 2^{-k-1}$,*

$$\begin{aligned}\mathbb{P}\big[X_{r,k}(0) \ge x, X_{r,k}(h_2) - X_{r,k}(h_1) \ge y\big]\\ \le c \exp\biggl(-\frac{x^2}{2(k-r)\sigma^2} - \frac{cy^{3/2}}{2^k|h_2 - h_1|}\biggr),\end{aligned} \tag{47}$$

where the constants c depend on C.

PROOF. Observe first that we may assume y is bigger than a large constant depending on C times $2^k|h_2 - h_1|$ (and therefore also bigger than a large constant times $2^{2k}|h_2 - h_1|^2$), because otherwise the required bound follows from (36).

For any $\lambda_1, \lambda_2 > 0$, the left-hand side of (47) is bounded above by

$$\mathbb{E}\big[\exp\big(\lambda_1 X_{r,k}(0) + \lambda_2\big(X_{r,k}(h_2) - X_{r,k}(h_1)\big)\big)\big] \exp(-\lambda_1 x - \lambda_2 y). \tag{48}$$

We will show that if $\lambda_1 \le 10C$ and $1 \le \lambda_2 \le |h_2 - h_1|^{-1}$,

$$\begin{aligned}&\mathbb{E}\big[\exp\big(\lambda_1 X_{r,k}(0) + \lambda_2\big(X_{r,k}(h_2) - X_{r,k}(h_1)\big)\big)\big]\\ &\quad\le c \exp\biggl(\frac{\lambda_1^2\sigma^2}{2}(k-r) + c\lambda_2 2^k|h_2 - h_1| + c\big(\lambda_2 2^k|h_2 - h_1|\big)^2\biggr).\end{aligned} \tag{49}$$

The result then follows by choosing $\lambda_1 = x((k-r)\sigma^2)^{-1}$ and $\lambda_2 = cy^{1/2}2^{-k}|h_2 - h_1|^{-1}$ in (48) and (49), for a suitable small c, and using our assumption that y is bigger than a large constant times $2^k|h_2 - h_1|$. Note that the assumptions on x, y, h_1 and h_2 ensure that $\lambda_1 \le 10C$ and $1 \le \lambda_2 \le |h_2 - h_1|^{-1}$.

We now prove (49). First, we note that similarly to the argument from (39) to (41),

$$\mathbb{E}\big[\exp\big(\lambda_1 W_p(0) + \lambda_2\big(W_p(h_2) - W_p(h_1)\big)\big)\big], \tag{50}$$

can be written explicitly as

$$I_0\Bigg(\bigg(\frac{1}{p}\big(\lambda_1+\big(\cos(h_2\log p)-\cos(h_1\log p)\big)\lambda_2\big)^2 + \frac{1}{p}\big(\big(\sin(h_2\log p)-\sin(h_1\log p)\big)\lambda_2\big)^2\bigg)^{1/2}\Bigg). \tag{51}$$

Recall from (43) that $\log I_0(\sqrt{x}) = \frac{1}{4}x + O(x^2)$, and that $\cos(h_2\log p) - \cos(h_1\log p) = O(|h_2-h_1|\log p)$ and $\sin(h_2\log p)-\sin(h_1\log p) = O(|h_2 - h_1|\log p)$. Thus, provided $\lambda_1 \le 10C$, $1 \le \lambda_2 \le |h_2-h_1|^{-1}$ and p is large enough, the logarithm of the quantity in (50) is at most

$$\begin{aligned}&\frac{1}{4p}\big(\lambda_1 + c\lambda_2|h_2-h_1|\log p\big)^2 + \frac{c}{p}\big(\lambda_2|h_2-h_1|\log p\big)^2 + cp^{-2}\\ &\quad\le \frac{\lambda_1^2}{4p} + \frac{c}{p}\lambda_2|h_2-h_1|\log p + \frac{c}{p}\big(\lambda_2|h_2-h_1|\log p\big)^2 + cp^{-2}.\end{aligned} \tag{52}$$

Here, we used the fact that $\lambda_1 \le 10C$. After summing over $2^r < \log p \le 2^k$, we get that

$$\begin{aligned}&\log\mathbb{E}\big[\exp\big(\lambda_1 X_{r,k}(0) + \lambda_2\big(X_{r,k}(h_2) - X_{r,k}(h_1)\big)\big)\big]\\ &\quad\le c + \sum_{2^r<\log p\le 2^k}\frac{\lambda_1^2}{4p} + c\sum_{2^r<\log p\le 2^k}\frac{\log p}{p}\lambda_2|h_2-h_1|\\ &\qquad + c\sum_{2^r<\log p\le 2^k}\frac{(\log p)^2}{p}\big(\lambda_2|h_2-h_1|\big)^2.\end{aligned}$$

In the above, if p is too small for (52) to be an upper bound, we simply use that (50) is bounded. The claim (49) now follows from the bounds (26) and (29). $\square$

We are now ready to prove Proposition 2.5. We will use the following notation: for $k \in \mathbb{N}$, let

(53) $\mathcal{H}_k$ be the set $\frac{1}{2^k}\mathbb{Z}$ of dyadic rationals, so that $\mathcal{H}_0 \subset \mathcal{H}_1 \subset \cdots \subset \mathcal{H}_k \subset \cdots \subset \mathbb{R}$ is a nested sequence of sets of equally spaced points and $|\mathcal{H}_k \cap [0, 1)| = 2^k$.

PROOF OF PROPOSITION 2.5. Without loss of generality, we may assume $h = 0$. We can also round x up and decrease a so that we may assume that x is an integer and $a \ge 1$. Define the events

$$B_q = \big\{X_{r,k}(0) \in [x-q-1, x-q]\big\}, \qquad q = 0, 1, \ldots, x-1$$

and

$$B_x = \big\{X_{r,k}(0) \le 0\big\}.$$

Note that the left-hand side of (44) is at most

$$\sum_{q=0}^{x} \mathbb{P}\Big[B_q \cap \Big\{\max_{h' \in A}\{X_{r,k}(h') - X_{r,k}(0)\} \geq a+q\Big\}\Big], \tag{54}$$

where $A = [-2^{-k-1}, 2^{-k-1}]$. Let $(h_i, i \geq 0)$ be a dyadic sequence such that $h_0 = 0$, $h_i \in \mathcal{H}_{k+i} \cap A$ and $\lim_{i\to\infty} h_i = h'$, so that $|h_{i+1} - h_i| \in \{0, 2^{-k-i-1}\}$ for all i. Because the map $h \mapsto X_{r,k}(h)$ is almost surely continuous,

$$X_{r,k}(h') - X_{r,k}(0) = \sum_{i=0}^{\infty}\big(X_{r,k}(h_{i+1}) - X_{r,k}(h_i)\big).$$

The right-hand side converges almost surely, since $\sum_{i=0}^{l}(X_{r,k}(h_{i+1}) - X_{r,k}(h_i)) = X_{r,k}(h_{l+1}) - X_{r,k}(0) \to X_{r,k}(h') - X_{r,k}(0)$, because $X_{r,k}(h)$ is continuous almost surely. Since $\sum_{i=0}^{\infty} \frac{1}{2(i+1)^2} \leq 1$, we have the inclusion of events,

$$\big\{X_{r,k}(h') - X_{r,k}(0) \geq a+q\big\} \subset \bigcup_{i=0}^{\infty}\Big\{X_{r,k}(h_{i+1}) - X_{r,k}(h_i) \geq \frac{a+q}{2(i+1)^2}\Big\}.$$

This implies that $\{\max_{h'\in A}(X_{r,k}(h') - X_{r,k}(0)) \geq a+q\}$ is included in

$$\bigcup_{i=0}^{\infty} \bigcup_{\substack{h_1 \in \mathcal{H}_{k+i}\cap A,\\ h_2 = h_1 \pm 2^{-k-i-1}}} \Big\{X_{r,k}(h_2) - X_{r,k}(h_1) \geq \frac{a+q}{2(i+1)^2}\Big\},$$

where we have ignored the case $h_1 = h_2$ since then event $\{X_{r,k}(h_2) - X_{r,k}(h_1) \geq \frac{a+q}{2(i+1)^2}\}$ is the empty set. Because $|\mathcal{H}_{k+i} \cap A| \leq c2^i$, the qth summand in (54) is at most,

$$\sum_{i=0}^{\infty} c2^i \sup_{\substack{h_1 \in \mathcal{H}_{k+i}\cap A,\\ h_2 = h_1 \pm 2^{-k-i-1}}} \mathbb{P}\Big[B_q \cap \Big\{X_{r,k}(h_2) - X_{r,k}(h_1) \geq \frac{a+q}{2(i+1)^2}\Big\}\Big].$$

Note that $a+q \leq a+x \leq 2^{2k}$ by assumption. Inequality (47) can thus be applied to get that the above is at most

$$c\sum_{i=0}^{\infty} 2^i \exp\Big(-\frac{(x-q-1)^2}{2(k-r)\sigma^2} - c2^i \frac{(a+q)^{3/2}}{(i+1)^3}\Big) \leq ce^{-\frac{(x-q-1)^2}{2(k-r)\sigma^2} - c(a+q)^{3/2}}.$$

Since $e^{-c(a+q)^{3/2}} \leq e^{-ca^{3/2} - cq^{3/2}}$, (54) is thus at most

$$\begin{aligned} ce^{-ca^{3/2}} \sum_{q=0}^{x} e^{-(x-q-1)^2/(2(k-r)\sigma^2) - cq^{3/2}} &\leq ce^{-x^2/(2(k-r)\sigma^2) - ca^{3/2}} \sum_{q=0}^{x} e^{c(q+1) - cq^{3/2}} \\ &\leq ce^{-x^2/(2(k-r)\sigma^2) - ca^{3/2}}, \end{aligned}$$

where we used the assumption $x \leq C(k-r)$. This proves (44). $\square$

2.3. *Gaussian approximation*. The purpose of this section is to compare the increments $Y_k(h)$ to Gaussian random variables with mean and variance independent of k, both for a single $h \in \mathbb{R}$ and for vectors $(Y_k(h_1), Y_k(h_2))$ for $h_1 \neq h_2 \in \mathbb{R}$. This will be used in the subsequent sections to apply the ballot theorem and derive bounds on the probability that $X_{r,k}(h_1)$ and $X_{r,k}(h_2)$ satisfy a barrier condition. One reason to pass to Gaussian random variables is that the standard ballot theorem provides such bounds for random walks with i.i.d. increments. It does not immediately apply to the process $k \mapsto X_{r,k}(h)$, whose increments $Y_k(h)$ have slightly different distributions for different k. Moreover, we need to show that the increments $Y_k(h_1)$ and $Y_k(h_2)$ for two points $h_1 \neq h_2$ become roughly independent when k is beyond the branching point $h_1 \wedge h_2$; cf. (13). To quantify this, we introduce a parameter Δ and refer to the scale $h_1 \wedge h_2 + \Delta$ as the *decoupling point*. Passing to Gaussian variables facilitates the proof of the decoupling, since in the Gaussian case we can investigate independence solely by controlling the covariance and the mean.

Our main tool is the following multivariate Berry–Esseen approximation for independent random vectors. For the remainder of the paper, $\eta_{\boldsymbol{\mu},\Sigma}$ will denote the Gaussian measure with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ.

LEMMA 2.8 (Corollary 17.2 in [6], see also Theorem 1.3 in [18]). *Let $(\boldsymbol{W}_j, j \geq 1)$ be a sequence of independent random vectors on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), P)$ with mean $E[\boldsymbol{W}_j]$ and covariance matrix $\mathrm{Cov}(\boldsymbol{W}_j)$. Define*

$$\boldsymbol{\mu}_m = \sum_{j=1}^{m} E[\boldsymbol{W}_j] \quad \text{and} \quad \Sigma_m = \sum_{j=1}^{m} \mathrm{Cov}(\boldsymbol{W}_j).$$

Let λ_m be the smallest eigenvalue of Σ_m and Q_m be the law of $\boldsymbol{W}_1 + \cdots + \boldsymbol{W}_m$.

There exists an absolute constant c depending only on the dimension d such that

$$\sup_{A \in \mathcal{A}} \left| Q_m(A) - \eta_{\boldsymbol{\mu}_m, \Sigma_m}(A) \right| \leq c \lambda_m^{-3/2} \sum_{j=1}^{m} E\left[\| \boldsymbol{W}_j - E[\boldsymbol{W}_j] \|^3 \right],$$

where $\mathcal{A}$ is the collection of Borel measurable convex subsets of $\mathbb{R}^d$.

Before stating the results, we recall the notation from Section 2.1: for fixed $h_1, h_2 \in \mathbb{R}$, we write $\boldsymbol{Y}_k = (Y_k(h_1), Y_k(h_2))$, $\boldsymbol{X}_{r,n} = (X_{r,n}(h_1), X_{r,n}(h_2))$. Furthermore, $\mathbb{Q}_{\boldsymbol{\lambda}}$ is the product measure from (31), where in this section we apply the same tilt to both components by setting

$$\boldsymbol{\lambda} = (\lambda, \lambda), \qquad \lambda \in \mathbb{R}.$$

We show that beyond the decoupling point $h_1 \wedge h_2 + \Delta$, the increments under $\mathbb{Q}_{\boldsymbol{\lambda}}$ are close (in terms of Δ) to being independent Gaussians with mean $\lambda\sigma^2$ and variance $\sigma^2 = (\log 2)/2$.

PROPOSITION 2.9. *Let $\lambda \in \mathbb{R}$ and $\Delta > 0$. Let $h_1, h_2 \in \mathbb{R}$, $m \geq h_1 \wedge h_2 + \Delta$ and $\mu = \lambda\sigma^2$. For any convex subsets $A_k \subseteq \mathbb{R}^2$, $k = m+1, \dots, n$, we have*

$$\begin{aligned} &\mathbb{Q}_{\boldsymbol{\lambda}}[\boldsymbol{X}_{m,k} \in A_k \ \forall m < k \leq n] \\ &\quad = \bigl(1 + O\bigl(e^{-c\Delta}\bigr)\bigr)\eta_{\mu,\sigma^2}^{\times 2(n-m)}\Biggl\{\boldsymbol{y} \in \mathbb{R}^{2\times(n-m)} : \sum_{j=1}^{k} \boldsymbol{y}_j \in A_{k+m} \\ &\qquad \forall k = 1, \dots, n-m\Biggr\} + O\bigl(e^{-e^{c\Delta}}\bigr), \end{aligned} \tag{55}$$

where $\eta_{\mu,\sigma^2}^{\times 2(n-m)}$ denotes the product measure on $2(n-m)$ independent Gaussians each with mean μ and variance σ^2.

PROOF. Recall that $\boldsymbol{Y}_k = \sum_{2^{k-1} < \log p \leq 2^k} \boldsymbol{W}_p$ where $\boldsymbol{W}_p = (W_p(h_1), W_p(h_2))$. The proof has two steps. First, Lemma 2.8 is applied successively for each k from $k = n$ down to $k = m+1$ to pass to a Gaussian measure. The resulting measure is the product of $(n-m)$ bivariate Gaussian measures with mean $\tilde{\boldsymbol{\mu}}_k = \tilde{\mu}_k(1,1)$ for $\tilde{\mu}_k = \mathbb{Q}_{\boldsymbol{\lambda}}[Y_k(h_1)] = \mathbb{Q}_{\boldsymbol{\lambda}}[Y_k(h_2)]$ and covariance matrix $\tilde{\Sigma}_k = \mathrm{Cov}_{\mathbb{Q}_{\boldsymbol{\lambda}}}[\boldsymbol{Y}_k]$. This measure is denoted by $\bigotimes_{k=m+1}^{n} \eta_{\tilde{\boldsymbol{\mu}}_k, \tilde{\Sigma}_k}$. Second, we explicitly compare the resulting Gaussian measure to the decoupled measure $\eta_{\mu,\sigma^2}^{\times 2(n-m)}$.

Conditioning on the values of $\boldsymbol{Y}_j$ for all $m+1 \leq j \leq n-1$, then applying Lemma 2.8 to the $\boldsymbol{W}_p$ with $2^{n-1} < \log p \leq 2^n$, and finally integrating over $\boldsymbol{Y}_j$ we obtain

$$\begin{aligned} &\Biggl|\mathbb{Q}_{\boldsymbol{\lambda}}[\boldsymbol{X}_{m,k} \in A_k \ \forall m < k \leq n] \\ &\qquad - \mathbb{Q}_{\boldsymbol{\lambda}} \times \eta_{\tilde{\boldsymbol{\mu}}_n, \tilde{\Sigma}_n}\Biggl[\sum_{j=m+1}^{k} \boldsymbol{Y}_j \in A_k \ \forall m < k \leq n-1, \\ &\qquad \boldsymbol{y}_n \in \Biggl(A_n - \sum_{j=m+1}^{n-1} \boldsymbol{Y}_j\Biggr)\Biggr]\Biggr| \\ &\qquad \leq c\lambda_n^{-3/2} \sum_{2^{n-1} < \log p \leq 2^n} \mathbb{Q}_{\boldsymbol{\lambda}}\bigl[\|\boldsymbol{W}_p - \mathbb{Q}_{\boldsymbol{\lambda}}[\boldsymbol{W}_p]\|^3\bigr], \end{aligned} \tag{56}$$

where $\boldsymbol{y}_n$ is sampled from $\eta_{\tilde{\boldsymbol{\mu}}_n, \tilde{\Sigma}_n}$, λ_n is the smallest eigenvalue of $\tilde{\Sigma}_n$, and $A_n - \boldsymbol{y}$ denotes the set A_n translated by $\boldsymbol{y}$. Lemma 2.8 can be applied the same way to the $\boldsymbol{W}_p$'s contributing to $\boldsymbol{Y}_{n-1}$, $\boldsymbol{Y}_{n-2}$, and so on. In each case, the relevant target subset for the sum is convex as an intersection of convex sets is convex. For example, the subset for $\sum_{j=m+1}^{n-1} \boldsymbol{Y}_{n-1}$ from (56) is $A_{n-1} \cap (A_n - \boldsymbol{y}_n)$. The resulting estimate is

$$\begin{aligned}(57)\qquad &\Bigg|\mathbb{Q}_{\boldsymbol{\lambda}}[\boldsymbol{X}_{m,k} \in A_k \ \forall m < k \leq n] \\ &- \bigotimes_{k=m+1}^{n} \eta_{\tilde{\boldsymbol{\mu}}_k, \tilde{\Sigma}_k}\Bigg\{\boldsymbol{y} \in \mathbb{R}^{2\times(n-m)} : \sum_{j=m+1}^{k} \boldsymbol{y}_j \in A_k \ \forall k = m+1, \dots, n\Bigg\}\Bigg| \\ &\leq c \sum_{k=m+1}^{n} \sum_{2^{k-1} < \log p \leq 2^k} \lambda_k^{-3/2} \mathbb{Q}_{\boldsymbol{\lambda}}\big[\|\boldsymbol{W}_p - \mathbb{Q}_{\boldsymbol{\lambda}}[\boldsymbol{W}_p]\|^3\big].\end{aligned}$$

For $k > h_1 \wedge h_2 + \Delta$, the eigenvalues λ_k are uniformly bounded away from 0. Indeed, observe that by (38), and the discussion preceding (30), and Lemma 2.1,

$$\lambda_k = \sigma_k^2 - \rho_k + O\big(e^{-2^{k-1}}\big) = \sigma^2 + O\big(e^{-c\sqrt{2^k}} + e^{-c\Delta}\big) \geq c > 0,$$

for Δ large enough but fixed. Also by construction, the norm of the vector $\boldsymbol{W}_p$ is bounded by $cp^{-1/2}$. Hence, the error term in (57) is bounded by

$$(58)\qquad c \sum_{2^m < \log p \leq 2^n} p^{-3/2} \leq ce^{-2^{m-1}} \leq e^{-e^{c\Delta}}.$$

It remains to compare the measure $\bigotimes_{k=m+1}^{n} \eta_{\tilde{\boldsymbol{\mu}}_k, \tilde{\Sigma}_k}$ with the measure $\eta_{\mu,\sigma^2}^{\times 2(n-m)}$. The specifics of the considered event play no role at this point, so we write B for a generic measurable subset of $\mathbb{R}^2$. We show

$$(59)\qquad \begin{aligned}\eta_{\tilde{\boldsymbol{\mu}}_k, \tilde{\Sigma}_k}[B] &= \big(1 + O\big(e^{-c(k-h_1\wedge h_2)}\big)\big)\eta_{\mu,\sigma^2}[B] \\ &\quad + O\big(e^{-e^{c(k-h_1\wedge h_2)}}\big) \qquad \forall k > m.\end{aligned}$$

Together with (58) and (57), this implies the proposition since the estimate (59) can be applied successively integrating in each coordinate to get for any $A \subseteq \mathbb{R}^{2(n-m)}$

$$\begin{aligned}\bigotimes_{k=m+1}^{n} \eta_{\tilde{\boldsymbol{\mu}}_k, \tilde{\Sigma}_k}[A] &= \prod_{k=m+1}^{n} \big(1 + O\big(e^{-c(k-h_1\wedge h_2)}\big)\big)\eta_{\mu,\sigma^2}^{\times 2(n-m)}[A] \\ &\quad + \sum_{k=m+1}^{n} O\big(e^{-e^{c(k-h_1\wedge h_2)}}\big) \\ &= \big(1 + O\big(e^{-c\Delta}\big)\big)\eta_{\mu,\sigma^2}^{\times 2(n-m)}[A] + O\big(e^{-e^{c\Delta}}\big).\end{aligned}$$

To prove (59), we compare densities. Proposition 2.4 and Lemma 2.1 give

$$(60)\qquad \begin{aligned}\tilde{\mu}_k &= \mu + O\big(2^{-(k-h_1\wedge h_2)}\big), \\ \tilde{\Sigma}_k &= \sigma^2\mathbb{1} + O\big(2^{-(k-h_1\wedge h_2)}\big),\end{aligned}$$

where $\mathbb{1}$ is the 2×2 identity matrix, using that $k > m > h_1 \wedge h_2 + \Delta$. Consider the set,

$$E_k = \{\boldsymbol{y} \in \mathbb{R}^2 : \|\boldsymbol{y} - \tilde{\boldsymbol{\mu}}_k\| \leq 2^{(k-h_1\wedge h_2)/4}\}.$$

A straightforward Gaussian estimate yields

$$\eta_{\tilde{\boldsymbol{\mu}}_k, \tilde{\Sigma}_k}[E_k^c] \leq \exp\left(-c\frac{2^{(k-h_1\wedge h_2)/2}}{\sigma^2}\right) \leq e^{-e^{c(k-h_1\wedge h_2)}},$$

and similarly for $\eta_{\mu,\sigma^2}^{\times 2}[E_k^c]$. Therefore, it suffices to prove (59) for $B \subset E_k$. The density of $\eta_{\tilde{\boldsymbol{\mu}}_k, \tilde{\Sigma}_k}$ with respect to Lebesgue measure is

$$\frac{1}{2\pi(\det \tilde{\Sigma}_k)^{1/2}} e^{-(\boldsymbol{y}-\tilde{\boldsymbol{\mu}}_k)\cdot\tilde{\Sigma}_k^{-1}(\boldsymbol{y}-\tilde{\boldsymbol{\mu}}_k)/2}. \tag{61}$$

By (60),

$$(\det \tilde{\Sigma}_k)^{-1/2} = \sigma^{-2}\big(1 + O\big(2^{-(k-h_1\wedge h_2)}\big)\big).$$

Furthermore, for all $\boldsymbol{y} \in \mathbb{R}^2$,

$$\begin{aligned}(\boldsymbol{y} - \tilde{\boldsymbol{\mu}}_k) \cdot \tilde{\Sigma}_k^{-1}(\boldsymbol{y} - \tilde{\boldsymbol{\mu}}_k)& \\ = \sigma^{-2}\|\boldsymbol{y} - \tilde{\boldsymbol{\mu}}_k\|^2 &+ (\boldsymbol{y} - \tilde{\boldsymbol{\mu}}_k) \cdot \big(\tilde{\Sigma}_k^{-1} - \sigma^{-2}\mathbb{1}\big)(\boldsymbol{y} - \tilde{\boldsymbol{\mu}}_k).\end{aligned}$$

By (60) and the definition of E_k, the error term is

$$(\boldsymbol{y} - \tilde{\boldsymbol{\mu}}_k) \cdot \big(\tilde{\Sigma}_k^{-1} - \sigma^{-2}\mathbb{1}\big)(\boldsymbol{y} - \tilde{\boldsymbol{\mu}}_k) = O\big(2^{-(k-h_1\wedge h_2)/4}\big).$$

Thus, on E_k, the density (61) equals $(1 + O(e^{-c(k-h_1\wedge h_2)}))\frac{1}{2\pi\sigma^2}e^{-\|\boldsymbol{y}-\tilde{\boldsymbol{\mu}}_k\|^2/2\sigma^2}$. In particular,

$$\eta_{\tilde{\boldsymbol{\mu}}_k, \tilde{\Sigma}_k}[B] = \big(1 + O\big(e^{-c(k-h_1\wedge h_2)}\big)\big)\eta_{\tilde{\mu}_k,\sigma^2}^{\times 2}[B] \qquad \text{for any } B \subset E_k.$$

It remains to compare the densities of $\eta_{\tilde{\mu}_k,\sigma^2}$ and η_{μ,σ^2}. We have that

$$(y - \tilde{\mu}_k)^2 = (y - \mu)^2 + (\tilde{\mu}_k - \mu)^2 - 2(y - \mu)(\tilde{\mu}_k - \mu).$$

The second term is $O(2^{-(k-h_1\wedge h_2)})$ by (60). The third term can be estimated using the fact that $|y - \tilde{\mu}_k| = O(2^{(k-h_1\wedge h_2)/4})$:

$$\begin{aligned}\big|(y - \mu)(\tilde{\mu}_k - \mu)\big| &\leq \big(|y - \tilde{\mu}_k| + |\tilde{\mu}_k - \mu|\big)|\tilde{\mu}_k - \mu| \\ &= O\big(2^{-3(k-h_1\wedge h_2)/4}\big).\end{aligned}$$

This implies that on $B \subset E_k$

$$\eta_{\tilde{\mu}_k,\sigma^2}^{\times 2}[B] = \big(1 + O\big(e^{-c(k-h_1\wedge h_2)}\big)\big)\eta_{\mu,\sigma^2}^{\times 2}[B].$$

This completes the proof of the claim (59). □

The next proposition provides a Gaussian comparison before the branching point. Before this point the increments are highly correlated, so the walks behave essentially as one. When we apply this proposition in Lemma 3.9 we will therefore drop the condition on the second walk at negligible cost. This is the reason the event involves only the walk $X_{m,k}(h_1)$, even though $\mathbb{Q}_{\boldsymbol{\lambda}}$ tilts both $X_{m,k}(h_1)$ and $X_{m,k}(h_2)$. The proof of the proposition is omitted, as it follows the previous one closely, with μ replaced by $2\lambda\sigma^2$ in (60).

PROPOSITION 2.10. *Let $\lambda \in \mathbb{R}$, $\Delta > 0$ and $\mu = 2\lambda\sigma^2$. Let $h_1, h_2 \in \mathbb{R}$ with $l = h_1 \wedge h_2$ and $m \leq l - \Delta$. For any convex subsets $A_k \subseteq \mathbb{R}$, $k = m+1, \dots, l-\Delta$, we have*

$$\begin{aligned}\mathbb{Q}_{\boldsymbol{\lambda}}&\big[X_{m,k}(h_1) \in A_k \ \forall m < k \leq l - \Delta\big]\\ &= \big(1 + O\big(e^{-c\Delta}\big)\big)\eta_{\mu,\sigma^2}^{\times(l-\Delta-m)}\Bigg\{\boldsymbol{y} \in \mathbb{R}^{\times(l-\Delta-m)} : \sum_{j=1}^{k} y_j \in A_{k+m}\\ &\qquad \forall k = 1, \dots, l - \Delta - m\Bigg\} + O\big(e^{-e^{c\Delta}}\big).\end{aligned} \tag{62}$$

A one-point Gaussian approximation for the measure $\mathbb{Q}_\lambda$ from (31) will also be needed. The proof is again similar to the proof of Proposition 2.9 and is omitted. One noticeable difference is in (60) where the covariance estimate is replaced by $\sigma_k^2 = \sigma^2 + O(e^{-e^{ck}})$ because of (26). The additive error $e^{-e^{c\Delta}}$ is then replaced by $e^{-e^{cm}}$. The multiplicative error $1 + O(e^{-c\Delta})$ becomes $1 + O(e^{-e^{cm}})$, and can thus be "absorbed" in the additive error.

PROPOSITION 2.11. *Let $\lambda \in \mathbb{R}$, $h \in \mathbb{R}$, $0 \leq m < n$ and $\mu = \lambda\sigma^2$. For any convex subsets $A_k \subseteq \mathbb{R}$, $k = m+1, \dots, n$, we have*

$$\begin{aligned}\mathbb{Q}_\lambda&\big[X_{m,k}(h) \in A_k \ \forall m < k \leq n\big]\\ &= \eta_{\mu,\sigma^2}^{\times(n-m)}\Bigg\{y \in \mathbb{R}^{\times(n-m)} : \sum_{j=1}^{k} y_j \in A_{k+m} \ \forall k = 1, \dots, n-m\Bigg\}\\ &\quad + O\big(e^{-e^{cm}}\big).\end{aligned} \tag{63}$$

2.4. *Ballot theorem.* The ballot theorem provides an estimate for the probability that a random walk stays below a certain value and ends up in an interval. We state the case we need, which is that of Gaussian random walk with increments of mean 0 and variance σ^2.

LEMMA 2.12. *Let $(X_n)_{n\geq 0}$ be a Gaussian random walk with increments of mean 0 and variance $\sigma^2 > 0$, with $X_0 = 0$. Let $\delta > 0$. There is a constant $c =$*

$c(\sigma, \delta)$ *such that for all* $a > 0$, $b \le a - \delta$ *and* $n \ge 1$

$$P\big[X_n \in (b, b+\delta) \text{ and } X_k \le a \text{ for } 0 < k < n\big] \le c\frac{(1+a)(1+a-b)}{n^{3/2}}. \tag{64}$$

Also provided $\delta < 1$,

$$\frac{1}{cn^{3/2}} \le P\big[X_n \in (0, \delta) \text{ and } X_k \le 1 \text{ for } 0 < k < n\big]. \tag{65}$$

PROOF. Note that $(X_k)_{0\le k\le n}$ has the law of $(\sigma B_k)_{0\le k\le n}$, where $(B_t)_{t\ge 0}$ is standard Brownian motion. Thus, we see that the probability in (64) conditioned on $X_n = y$ can be written as the probability that a Brownian bridge avoids a barrier at integer times. The bound (6.4) of [28] shows, after shifting by a/σ and reflecting, that this conditional probability is at most $c(1+a/\sigma)(1+(a-b-\delta)/\sigma)/n$. Noting that $P[X_n \in (b, b+\delta)] \le cn^{-1/2}$ then yields (64). In a similar fashion, the display below (6.4) in [28] gives (65). □

3. Proof of Theorem 1.2. In this section, we prove (8), that is,

$$\lim_{n\to\infty} \mathbb{P}\Big[m_n(-\varepsilon) \le \max_{h\in[0,1]} X_n(h) \le m_n(\varepsilon)\Big] = 1 \qquad \text{for all } \varepsilon > 0. \tag{66}$$

This proves Theorem 1.2 for the subsequence $T = e^{2^n}$, $n \in \mathbb{N}$. The extension of the argument to general sequences T follows by trivial adjustments. We will need to consider the process $X_{r,n}(h)$ with the first r scales cutoff; see (19). Throughout this section, we use

$$r = \big\lfloor (\log\log n)^2 \big\rfloor. \tag{67}$$

First, we show that the difference between $\max_{h\in[0,1]} X_{r,n}(h)$ and $\max_{h\in[0,1]} X_n(h)$ is negligible compared to the subleading correction term.

LEMMA 3.1. *For all* $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\Big[\max_{h\in[0,1]} X_n(h) \ge m_n(2\varepsilon), \max_{h\in[0,1]} X_{r,n}(h) \le m_{n-r}(\varepsilon)\Big] = 0, \tag{68}$$

$$\lim_{n\to\infty} \mathbb{P}\Big[\max_{h\in[0,1]} X_n(h) \le m_n(-2\varepsilon), \max_{h\in[0,1]} X_{r,n}(h) \ge m_{n-r}(-\varepsilon)\Big] = 0. \tag{69}$$

PROOF. The event in the probability in (68) implies $\max_{h\in[0,1]} X_r(h) \ge (\log 2)r + \varepsilon\log(n-r) \ge 100(\log 2)r$, where the last inequality holds for n large enough. But (46), with $n = r$, gives

$$\mathbb{P}\Big[\max_{h\in[0,1]} X_r(h) \ge 100(\log 2)r\Big] \le 2^{-99r} \to 0 \qquad \text{as } r \to \infty.$$

Since the laws of $\max_{h\in[0,1]} X_r(h)$ and $-\min_{h\in[0,1]} X_r(h)$, coincide we also have that the probability $\mathbb{P}[\min_{h\in[0,1]} X_r(h) \le -100(\log 2)r]$ tends to 0 as $r \to \infty$, which similarly implies (69). □

In the proof of (66), we will use a change of measure under which the process $X_{r,n}$ has an upward drift of

$$\mu(\varepsilon) = \frac{m_{n-r}(\varepsilon)}{n-r} = \frac{(n-r)\log 2 - \frac{3}{4}\log(n-r) + \varepsilon \log(n-r)}{n-r}. \tag{70}$$

We use the following consequence of (9) and (25) several times:

$$\frac{\mu(\varepsilon)^2}{2\sigma^2} = \log 2 - \left(\frac{3}{2} - 2\varepsilon\right)\frac{\log(n-r)}{n-r} + o(n^{-1}). \tag{71}$$

3.1. *Proof of the upper bound.* In this section we prove the upper bound part of (66). By Lemma 3.1, it suffices to prove the following upper bound for $\max_{h\in[0,1]} X_{r,n}(h)$.

PROPOSITION 3.2. *For all $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\Big[\max_{h\in[0,1]} X_{r,n}(h) \geq m_{n-r}(\varepsilon)\Big] = 0. \tag{72}$$

The first step is to reduce the proof to a bound on the maximum over the discrete set $\mathcal{H}_n \cap [0, 1]$ [as defined in (53)] using the continuity estimates from Section 2.2.

LEMMA 3.3. *For all $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\Big[\max_{h\in[0,1]} X_{r,n}(h) \geq m_{n-r}(2\varepsilon), \max_{h\in\mathcal{H}_n\cap[0,1]} X_{r,n}(h) \leq m_{n-r}(\varepsilon)\Big] = 0. \tag{73}$$

PROOF. Using translation invariance and a union bound on 2^n intervals, the probability in (73) is at most

$$2^n\mathbb{P}\Big[\max_{h:|h|\leq 2^{-n-1}} X_{r,n}(h) \geq m_{n-r}(2\varepsilon), X_{r,n}(0) \leq m_{n-r}(\varepsilon)\Big].$$

Proposition 2.5 can be applied with $k = n$, $x = m_{n-r}(\varepsilon) = (n-r)\mu(\varepsilon)$ and $a = m_{n-r}(2\varepsilon) - m_{n-r}(\varepsilon) = \varepsilon \log(n-r) < 2^{2n} - x$. This gives the upper bound

$$c2^n \exp\left(-(n-r)\frac{\mu(\varepsilon)^2}{2\sigma^2} - c\varepsilon^{3/2}\big(\log(n-r)\big)^{3/2}\right). \tag{74}$$

Using (71) and (67), we get that (74) is at most

$$c2^n\big(2^{r-n}(n-r)^{\frac{3}{2}-2\varepsilon}e^{-c\varepsilon^{3/2}(\log(n-r))^{3/2}}\big) = o(1). \qquad \square$$

The second step is to show that for each $h \in [0, 1]$ the process $k \to X_{r,k}(h)$ satisfies a barrier condition with very high probability. This simply requires a union bound together with continuity estimates.

LEMMA 3.4. *For all* $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\big[\exists h \in [0,1], k \in \{\lfloor \log n\rfloor^2, \dots, n\} \textit{ s.t.} \\ X_{r,k}(h) > (k-r)\mu(\varepsilon) + (\log n)^2\big] = 0. \tag{75}$$

PROOF. By two successive union bounds, first over the scales $k = \lfloor \log n\rfloor^2, \dots, n$, and then, for each of those scales, over 2^k intervals (together with translation invariance), the probability in (75) is at most

$$\sum_{k=\lfloor \log n\rfloor^2}^{n} 2^k \mathbb{P}\Big[\max_{h:|h|\le 2^{-k-1}} X_{r,k}(h) \ge (k-r)\mu(\varepsilon) + (\log n)^2\Big].$$

The maximal inequality (45) can be applied since the right-hand side of the inequality in the probability is less than a constant times $(k-r)$. Thus, the sum is bounded above by

$$c \sum_{k=\lfloor \log n\rfloor^2}^{n} 2^k \exp\Big(-\frac{((k-r)\mu(\varepsilon) + (\log n)^2)^2}{2(k-r)\sigma^2}\Big).$$

Using (71), the argument in the exponential is at least

$$(k-r)\log 2 - \frac{3}{2}\log(n-r) + c(\log n)^2.$$

We conclude that the probability in (75) is at most

$$c \sum_{k=\lfloor \log n\rfloor^2}^{n} 2^k \big(2^{r-k} n^{3/2} e^{-c(\log n)^2}\big) = c 2^r n^{5/2} e^{-c(\log n)^2} = o(1).$$

□

Lemma 3.3 and Lemma 3.4 show that $\max_{h\in[0,1]} X_{r,n}(h)$ exceeds $m_{n-r}(2\varepsilon)$ only if, for some $h \in \mathcal{H}_n \cap [0,1]$, $X_{r,n}(h)$ exceeds $m_{n-r}(\varepsilon)$ and the process $(X_{r,k}(h), \lfloor \log n\rfloor^2 \le k \le n)$ stays below a linear barrier. The number of $h \in \mathcal{H}_n$ that manage this feat is

$$Z^+ = \sum_{h\in\mathcal{H}_n\cap[0,1]} \mathbf{1}_{J^+(h)}, \qquad \text{where} \\ J^+(h) = \{X_{r,n}(h) \ge m_{n-r}(\varepsilon), \\ X_{r,k}(h) \le (k-r)\mu(\varepsilon) + (\log n)^2 \ \forall k \ge \lfloor \log n\rfloor^2\}. \tag{76}$$

We show $\mathbb{P}[Z^+ > 0] \le c 2^r (\log n)^6 (n-r)^{-2\varepsilon}$, thereby proving Proposition 3.2 since the right-hand side is $o(1)$ by the definition (67) of r. Here, we shall use the previous Gaussian approximation results and the ballot theorem.

PROPOSITION 3.5. *For all $\varepsilon > 0$,*

$$\mathbb{P}[Z^+ > 0] \leq \mathbb{E}[Z^+] \leq c2^r (\log n)^6 (n-r)^{-2\varepsilon}. \tag{77}$$

PROOF. By translation invariance and linearity of expectation, we have $\mathbb{E}[Z^+] = 2^n \mathbb{P}[J^+(0)]$. We show that

$$\mathbb{P}[J^+(0)] \leq c2^{r-n} (\log n)^6 (n-r)^{-2\varepsilon}, \tag{78}$$

thus yielding (77). To prove (78), let $\lambda = \mu(\varepsilon)/\sigma^2$, and recall the definition of $\mathbb{Q}_\lambda$ from (31). We have that

$$\mathbb{P}[J^+(0)] \leq \mathbb{Q}_\lambda[J^+(0)] e^{\sum_{k=r+1}^n \psi_k^{(1)}(\lambda) - \lambda(n-r)\mu(\varepsilon)}, \tag{79}$$

because $X_{r,n}(0) \geq (n-r)\mu(\varepsilon)$ on the event $J^+(0)$. Using the estimates (34) and (26), we get that

$$\sum_{k=r+1}^{n} \psi_k^{(1)}(\lambda) - \lambda(n-r)\mu(\varepsilon) = -(n-r)\frac{\mu(\varepsilon)^2}{2\sigma^2} + O(e^{-c\sqrt{2^r}}). \tag{80}$$

By (71), the exponential in (79) is thus at most $c2^{r-n}(n-r)^{\frac{3}{2}-2\varepsilon}$. It remains to show

$$\mathbb{Q}_\lambda[J^+(0)] \leq c(\log n)^6 (n-r)^{-3/2}. \tag{81}$$

The event $J^+(0)$ takes the form in Proposition 2.11 with $m = r$. Thus, $\mathbb{Q}_\lambda[J^+(0)]$ is at most $\eta_{\mu(\varepsilon),\sigma^2}^{\times(n-r)}(E_1) + O(e^{-e^{cr}})$, where

$$E_1 = \left\{ \boldsymbol{y} \in \mathbb{R}^{n-r} : \sum_{l=1}^{k} (y_l - \mu(\varepsilon)) \leq (\log n)^2 \right.$$
$$\left. \forall k \geq \lfloor \log n \rfloor^2 - r, \sum_{l=1}^{n-r} (y_l - \mu(\varepsilon)) \geq 0 \right\}.$$

After recentering, the probability of E_1 is simply

$$\eta_{0,\sigma^2}^{\times(n-r)} \left\{ \boldsymbol{y} \in \mathbb{R}^{n-r} : \sum_{l=1}^{k} y_l \leq (\log n)^2 \ \forall k \geq \lfloor \log n \rfloor^2 - r, \sum_{l=1}^{n-r} y_l \geq 0 \right\}. \tag{82}$$

By conditioning on $\sum_{l=1}^{\lfloor \log n \rfloor^2 - r} y_l = q$, we may bound the above by the supremum over $q \in [-(\log n)^2, (\log n)^2]$ of $\eta_{0,\sigma^2}^{\times(n-(\log n)^2)}(E_2) + O(ce^{-(\log n)^2})$, where

$$E_2 = \left\{ \boldsymbol{y} \in \mathbb{R}^{n-(\log n)^2} : \sum_{l=1}^{k} y_l \leq (\log n)^2 - q \ \forall k \geq 0, \sum_{l=1}^{n-\lfloor \log n \rfloor^2} y_l \geq -q \right\}. \tag{83}$$

This is because of the standard Gaussian bound

$$\eta_{0,\sigma^2}^{\times((\log n)^2-r)}\left\{\boldsymbol{y}\in\mathbb{R}^{(\log n)^2-r}:\sum_{l=1}^{(\log n)^2-r}y_l\leq-(\log n)^2\right\}$$
$$\leq c\exp\left(-c\frac{(\log n)^4}{(\log n)^2-r}\right).$$

For a given q, the probability of the event in (83) may be bounded above by a union bound over a partition of $[-q,(\log n)^2-q]$ into intervals of length 1, and the ballot theorem (Lemma 2.12). This gives an upper bound for (82) of

$$\sup_{-(\log n)^2\leq q\leq(\log n)^2}(\log n)^2\times c\frac{(1+(\log n)^2-q)(2(\log n)^2)}{(n-r)^{3/2}}$$
$$\leq c(\log n)^6(n-r)^{-3/2}.$$

This proves (81), and thus also (78) and (77). □

3.2. *Proof of the lower bound*. In this section, we prove the lower bound part of (66). The proof is reduced to a lower bound on $\max_{h\in[0,1]}X_{r,n}(h)$ by Lemma 3.1. We show the following.

PROPOSITION 3.6. *For all* $\varepsilon>0$,

$$\lim_{n\to\infty}\mathbb{P}\Big[\max_{h\in[0,1]}X_{r,n}(h)\geq m_{n-r}(-\varepsilon)\Big]=1. \tag{84}$$

As for the upper bound, we consider a modified number of exceedances with a barrier. For $\delta>0$, let

$$J^-(h)=\big\{X_{r,n}(h)\in\big[m_{n-r}(-\varepsilon),m_{n-r}(-\varepsilon)+\delta\big],$$
$$X_{r,k}(h)\leq(k-r)\mu(-\varepsilon)+1\ \forall k=r+1,\dots,n\big\}.$$

We omit the dependence on the parameter δ in the notation for simplicity. Consider the random variable

$$Z^-=\sum_{h\in\mathcal{H}_n\cap[0,1)}\mathbf{1}_{J^-(h)}.$$

Clearly, $\max_{h\in[0,1]}X_{r,n}(h)\geq m_{n-r}(-\varepsilon)$ if and only if $Z^-\geq1$. The Paley–Zygmund inequality implies that

$$\mathbb{P}\big(Z^-\geq1\big)\geq\frac{\mathbb{E}[Z^-]^2}{\mathbb{E}[(Z^-)^2]}.$$

We will prove the following estimates for the first and second moments of Z^-. Let

$$A=\left\{\boldsymbol{y}\in\mathbb{R}^{n-r}:\sum_{k=1}^{n-r}y_k\in[0,\delta],\sum_{k=1}^{l-r}y_k\leq1\ \forall l=r+1,\dots,n\right\}. \tag{85}$$

LEMMA 3.7. *For* $\delta > 0$,

$$\mathbb{E}[Z^-] \geq (1+o(1))e^{-c\delta}2^r(n-r)^{\frac{3}{2}+2\varepsilon}\eta_{0,\sigma^2}^{\times(n-r)}[A]. \tag{86}$$

LEMMA 3.8. *For* $\delta > 0$,

$$\mathbb{E}[(Z^-)^2] \leq (1+o(1))\big(2^r(n-r)^{\frac{3}{2}+2\varepsilon}\eta_{0,\sigma^2}^{\times(n-r)}[A]\big)^2. \tag{87}$$

The lower bound (84) follows directly from these two lemmas, even without estimating the probability $\eta_{0,\sigma^2}^{\times(n-r)}[A]$ precisely. However, it is important to observe that (65) of the ballot theorem (Lemma 2.12) ensures that

$$\eta_{0,\sigma^2}^{\times(n-r)}[A] \geq c(n-r)^{-3/2}. \tag{88}$$

In particular, this implies that $\mathbb{E}[Z^-] \to \infty$, as n goes to infinity.

PROOF OF PROPOSITION 3.6. By the Paley–Zygmund inequality, Lemma 3.7 and Lemma 3.8, we have

$$\begin{aligned}\mathbb{P}\Big[\max_{h\in[0,1]} X_{r,n}(h) \geq m_{n-r}(-\varepsilon)\Big] &\geq \mathbb{P}(Z^- \geq 1)\\ &\geq \frac{\mathbb{E}[Z^-]^2}{\mathbb{E}[(Z^-)^2]}\\ &\geq (1+o(1))e^{-2c\delta}.\end{aligned}$$

The result follows by taking the limits $n \to \infty$, then $\delta \to 0$. □

We now prove the bound on $\mathbb{E}[Z^-]$.

PROOF OF LEMMA 3.7. Translation invariance implies $\mathbb{E}[Z^-] = 2^n\mathbb{P}[J^-(0)]$. Consider the probability $\mathbb{Q}_\lambda$ from (31), where $\lambda = \mu(-\varepsilon)/\sigma^2$. (By (35) and (26), this choice of λ implies that $\mathbb{Q}_\lambda[Y_k(0)]$ is approximately $\mu(-\varepsilon)$.) Since on the event $J^-(0)$, we have that $X_{r,n} \leq (n-r)\mu(-\varepsilon)+\delta$, the definition of $\mathbb{Q}_\lambda$ implies that

$$\mathbb{P}[J^-(0)] \geq \mathbb{Q}_\lambda[J^-(0)]e^{\sum_{k=r+1}^n \psi_k^{(1)}(\lambda)-\lambda(n-r)\mu(-\varepsilon)-c\delta}. \tag{89}$$

Proceeding as in (80) to estimate $\sum_{k=r+1}^n \psi_k^{(1)}(\lambda) - \lambda(n-r)\mu(-\varepsilon)$, and using (71), we get

$$\mathbb{P}[J^-(0)] \geq (1+o(1))e^{-c\delta}2^{-(n-r)}(n-r)^{3/2+2\varepsilon}\mathbb{Q}_\lambda[J^-(0)].$$

The event $J^-(0)$ is of the form appearing in the Berry–Esseen approximation of Proposition 2.11. The result can be applied with $m = r$, and after recentering the increments by their mean $\mu = \lambda\sigma^2 = \mu(-\varepsilon)$ we get

$$\mathbb{Q}_\lambda[J^-(0)] = \eta_{0,\sigma^2}^{\times(n-r)}[A] + O(e^{-e^{cr}}).$$

By (88), $\eta_{0,\sigma^2}^{\times(n-r)}[A]$ dominates $e^{-e^{cr}}$, since $r = \lfloor (\log\log n)^2 \rfloor$. This proves the lemma. $\square$

REMARK 3. We note for future reference that the same reasoning [using that $X_{r,n} \geq (n-r)\mu(-\varepsilon)$ on $J^-(0)$; cf. (89)] gives the upper bound,

$$\mathbb{P}\big[J^-(0)\big] \leq \big(1+o(1)\big)2^{-(n-r)}(n-r)^{3/2+2\varepsilon}\eta_{0,\sigma^2}^{\times(n-r)}[A]. \tag{90}$$

To prove the second moment bound in Lemma 3.8, we use the identity

$$\mathbb{E}\big[(Z^-)^2\big] = \sum_{h_1,h_2 \in \mathcal{H}_n \cap [0,1)} \mathbb{P}\big[J^-(h_1) \cap J^-(h_2)\big]. \tag{91}$$

We thus seek bounds on $\mathbb{P}[J^-(h_1) \cap J^-(h_2)]$ for $h_1 \neq h_2$. This is the key additional difficulty in the lower bound calculation. In essence, these bounds are obtained by conditioning on the values of the processes $k \mapsto X_{r,k}(h_i)$, close to the "branching point" $h_1 \wedge h_2$ [defined in (13)], and then applying the following two lemmas. Lemma 3.9 gives an estimate for the part of the event before the branching point (where the processes are coupled), and Lemma 3.10 for the part after (where they are decoupled). To get sufficiently strong coupling and decoupling, each estimate must be applied for scales that are respectively slightly before and slightly after the branching point. To quantify this, we use for the decoupling parameter Δ the value

$$\Delta = r/100. \tag{92}$$

For convenience, define the recentered process

$$\overline{X}_{r,k}(h) = X_{r,k}(h) - (k-r)\mu(-\varepsilon).$$

LEMMA 3.9. *Let $h_1, h_2 \in \mathbb{R}$ and $l = h_1 \wedge h_2$. For $i = 1, 2$ and any $q \geq 0$, define the event*

$$\begin{aligned} A_i(q) = \big\{&\overline{X}_{r,l-\Delta}(h_i) \in [-q, -q+1], \\ &\overline{X}_{r,k}(h_i) \leq 1 \textit{ for } k = r+1, \ldots, l-\Delta\big\}. \end{aligned} \tag{93}$$

Then for any $q_1, q_2 \geq 0$,

$$\mathbb{P}\big[A_1(q_1) \cap A_2(q_2)\big] \leq c\frac{e^{-(l-\Delta-r)\frac{\mu(-\varepsilon)^2}{2\sigma^2}}}{(l-\Delta-r)^{3/2}}(1+q_1)e^{\frac{1}{2}\frac{\mu(-\varepsilon)}{\sigma^2}(q_1+q_2)}. \tag{94}$$

PROOF. Let $\lambda = \mu(-\varepsilon)/(2\sigma^2)$ and $\boldsymbol{\lambda} = \lambda(1,1)$. We recall the definition of $\mathbb{Q}_{\boldsymbol{\lambda}}$ from (31). The choice of λ ensures that $\mathbb{Q}_{\boldsymbol{\lambda}}[\boldsymbol{Y}_k(0)]$ is approximately $\mu(-\varepsilon)(1,1)$.

By the definition of $\mathbb{Q}_{\boldsymbol{\lambda}}$,

$$\begin{aligned}
&\mathbb{P}\big[A_1(q_1) \cap A_2(q_2)\big] \\
&\quad = \mathbb{Q}_{\boldsymbol{\lambda}}\Big[\mathbf{1}_{A_1(q_1)\cap A_2(q_2)} \prod_{i=1,2} e^{-\lambda \overline{X}_{r,l-\Delta}(h_i)}\Big] \\
&\qquad \times \exp\Bigg(\sum_{k=r+1}^{l-\Delta} \big\{\psi_k^{(2)}(\boldsymbol{\lambda}) - 2\lambda\mu(-\varepsilon)\big\}\Bigg),
\end{aligned} \tag{95}$$

where $\mathbf{1}_{A_1(q_1)\cap A_2(q_2)}$ denotes the indicator function of the event. Using Proposition 2.4 as well as the covariance estimates (26) and (27), we have that

$$\begin{aligned}
\sum_{k=r+1}^{l-\Delta} \psi_k^{(2)}(\boldsymbol{\lambda}) &= \lambda^2 \sum_{k=r+1}^{l-\Delta} \big(\sigma_k^2 + \rho_k + O\big(e^{-2^{k-1}}\big)\big) \\
&\le \lambda^2 (l-\Delta-r) 2\sigma^2 + O(1) \\
&= (l-\Delta-r)\frac{\mu(-\varepsilon)^2}{2\sigma^2} + O(1).
\end{aligned}$$

This proves that the second exponential in (95) is at most $ce^{-(l-\Delta-r)\frac{\mu(-\varepsilon)^2}{2\sigma^2}}$. Also on the event $A_1(q_1) \cap A_2(q_2)$, the first exponential is at most $ce^{\lambda q_1 + \lambda q_2}$. Thus,

$$\begin{aligned}
&\mathbb{P}\big[A_1(q_1) \cap A_2(q_2)\big] \\
&\quad \le ce^{-(l-\Delta-r)\frac{\mu(-\varepsilon)^2}{2\sigma^2} + \frac{1}{2}\frac{\mu(-\varepsilon)}{\sigma^2}(q_1+q_2)} \mathbb{Q}_{\boldsymbol{\lambda}}\big[A_1(q_1) \cap A_2(q_2)\big].
\end{aligned} \tag{96}$$

It remains to bound $\mathbb{Q}_{\boldsymbol{\lambda}}[A_1(q_1) \cap A_2(q_2)]$. In fact, we drop the condition on h_2 and bound $\mathbb{Q}_{\boldsymbol{\lambda}}[A_1(q_1)]$. We expect not to lose much by this because the behavior at h_1 and h_2 should be very similar. The event $A_1(q_1)$ is of the right form to use Proposition 2.10 with $m = r$ and $n = l - \Delta$. After recentering of the increments by $\mu(-\varepsilon)$, we get that $\mathbb{Q}_{\boldsymbol{\lambda}}[A_1(q_1)]$ is

$$\begin{aligned}
&\big(1 + O\big(e^{-cr}\big)\big)\eta_{0,\sigma^2}^{\times(l-\Delta-r)}\Bigg\{\boldsymbol{y} \in \mathbb{R}^{l-\Delta-r} : \sum_{l'=1}^{k} y_{l'} \le 1 \\
&\qquad \text{for } k = 1, \dots, l-\Delta-r, \sum_{l'=1}^{l-\Delta-r} y_{l'} \in [-q_1, -q_1+1]\Bigg\} + O\big(e^{-e^{cr}}\big).
\end{aligned}$$

By (64) of the ballot theorem (Lemma 2.12) with $b = -q_1$ and $\delta = 1$, the probability on the right-hand side is at most $c\frac{1+q_1}{(l-\Delta-r)^{3/2}}$. Together with (96), this proves (94). $\square$

We now prove the bound for scales after the decoupling point. One notable difference with the proof of the previous lemma is that the change of measure is now done for a λ which is twice the one of Lemma 3.9. This reflects the fact that, before the branching point, the two processes are essentially coupled, therefore, a tilt for one process is also a tilt for the other.

LEMMA 3.10. *Let $h_1, h_2 \in \mathbb{R}$. For any $h_1 \wedge h_2 + \Delta \le j \le n$, and $\delta, \delta' > 0$, define for $i = 1, 2$ and $q \ge 0$ the events*

$$B_i(q) = \{\overline{X}_{j,n}(h_i) - q \in [-\delta', \delta], \overline{X}_{j,k}(h_i) - q \le 1 \text{ for } k = j+1, \dots, n\},$$

$$\overline{B}_i(q) = \Bigg\{ \boldsymbol{y} \in (\mathbb{R}^2)^{\times(n-j)} : \sum_{k=1}^{n-j} (\boldsymbol{y}_k)_i - q \in [-\delta', \delta], \tag{97}$$
$$\sum_{k=1}^{j'} (\boldsymbol{y}_k)_i - q \le 1, \ \forall j' = 1, \dots, n-j \Bigg\},$$

where $\boldsymbol{y} = (((\boldsymbol{y}_k)_1, (\boldsymbol{y}_k)_2), k = 1, \dots, n-j)$. Then for $q_1, q_2 \in \mathbb{R}$,

$$\begin{aligned} &\mathbb{P}[B_1(q_1) \cap B_2(q_2)] \\ &\quad \le (1 + o(1)) \\ &\quad\quad \times e^{c\delta'} \prod_{i=1,2} \{ e^{-(n-j)\frac{\mu(-\varepsilon)^2}{2\sigma^2} - \frac{\mu(-\varepsilon)}{\sigma^2} q_i} (\eta_{0,\sigma^2}^{\times(n-j)}[\overline{B}_i(q_i)] + e^{-e^{c\Delta}}) \}. \end{aligned} \tag{98}$$

PROOF. Let $\lambda = \frac{\mu(-\varepsilon)}{\sigma^2}$, $\boldsymbol{\lambda} = \lambda(1, 1)$ and recall the definition of $\mathbb{Q}_{\boldsymbol{\lambda}}$ from (31). The choice of $\boldsymbol{\lambda}$ ensures that $\mathbb{Q}_{\boldsymbol{\lambda}}[\boldsymbol{Y}_k]$ is approximately $\mu(-\varepsilon)(1, 1)$. The definition of $\mathbb{Q}_{\boldsymbol{\lambda}}$ gives

$$\begin{aligned} &\mathbb{P}[B_1(q_1) \cap B_2(q_2)] \\ &\quad = \mathbb{Q}_{\boldsymbol{\lambda}} \Big[\mathbf{1}_{B_1(q_1) \cap B_2(q_2)} \prod_{i=1,2} e^{-\lambda \overline{X}_{j,n}(h_i)} \Big] e^{\sum_{k=j+1}^{n} \psi_k^{(2)}(\boldsymbol{\lambda}) - 2\lambda(n-j)\mu(-\varepsilon)}. \end{aligned} \tag{99}$$

By Proposition 2.4, (26) and (27),

$$\begin{aligned} \psi_k^{(2)}(\boldsymbol{\lambda}) &= \lambda^2(\sigma_k^2 + \rho_k) + O(e^{-2^{k-1}}) \\ &= \lambda\mu(-\varepsilon) + O(2^{-(k - h \wedge h_2)}). \end{aligned}$$

We deduce that $\sum_{k=j+1}^{n} \psi_k^{(2)}(\boldsymbol{\lambda})$ is at most $(n-j)\lambda\mu(-\varepsilon) + c2^{-\Delta}$. Therefore, the second exponential in (99) is $(1 + o(1))e^{-2(n-j)\frac{\mu(-\varepsilon)^2}{2\sigma^2}}$. On the event $B_1(q_1) \cap B_2(q_2)$, the first exponential in (99) is at most $e^{c\delta'} e^{-\frac{\mu(-\varepsilon)}{\sigma^2} q_1 - \frac{\mu(-\varepsilon)}{\sigma^2} q_2}$. In view of this, it only remains to show

$$\mathbb{Q}_{\boldsymbol{\lambda}}\big[B_1(q_1)\cap B_2(q_2)\big] \le \big(1+o(1)\big)\prod_{i=1,2}\eta_{0,\sigma^2}^{\times(n-j)}\big[\overline{B}_i(q_i)\big] + ce^{-e^{c\Delta}}. \tag{100}$$

Note that the event $B_1(q_1)\cap B_2(q_2)$ takes the form considered in Proposition 2.9. Applying Proposition 2.9 with j in place of m and then recentering yields

$$\mathbb{Q}_{\boldsymbol{\lambda}}\big[B_1(q_1)\cap B_2(q_2)\big] \le \big(1+ce^{-c\Delta}\big)\eta_{0,\sigma^2}^{\times 2(n-j)}\big[\overline{B}_1(q_1)\cap\overline{B}_2(q_2)\big] + ce^{-e^{c\Delta}}.$$

By independence, it is plain that

$$\eta_{0,\sigma^2}^{\times 2(n-j)}\big[\overline{B}_1(q_1)\cap\overline{B}_2(q_2)\big] = \prod_{i=1,2}\eta_{0,\sigma^2}^{\times(n-j)}\big[\overline{B}_i(q_i)\big].$$

This proves (100) and, therefore, also (98). □

The previous lemmas will now be used to prove bounds on $\mathbb{P}[J^-(h_1)\cap J^-(h_2)]$ in three cases: (i) $h_1\wedge h_2\le r-\Delta$, (ii) $r-\Delta<h_1\wedge h_2\le r+\Delta$ and (iii) $r+\Delta< h_1\wedge h_2\le n-\Delta$. The case $h_1\wedge h_2>n-\Delta$ is easy and will be handled directly in the proof of Lemma 3.8.

If $h_1\wedge h_2\le r-\Delta$, then h_1 and h_2 are sufficiently far apart so that the scale r is well beyond the "branching point" of h_1 and h_2, and the events $J^-(h_1)$ and $J^-(h_2)$ decouple:

LEMMA 3.11. *Let $h_1,h_2\in\mathbb{R}$ be such that $1\le h_1\wedge h_2\le r-\Delta$. Then*

$$\mathbb{P}\big[J^-(h_1)\cap J^-(h_2)\big] \le \big(1+o(1)\big)\bigg(\frac{(n-r)^{\frac{3}{2}+2\varepsilon}}{2^{n-r}}\eta_{0,\sigma^2}^{\times(n-r)}[A]\bigg)^2, \tag{101}$$

where A is the event defined in (85).

PROOF. Let $j=r$. By assumption, we have $h_1\wedge h_2+\Delta\le j$, so Lemma 3.10 can be applied with $q_1=q_2=0$ and $\delta'=0$ to give

$$\mathbb{P}\big[J^-(h_1)\cap J^-(h_2)\big] \le \big(1+o(1)\big)\big(e^{-\frac{\mu(-\varepsilon)^2}{2\sigma^2}(n-r)}\big(\eta_{0,\sigma^2}^{\times(n-r)}[A]+e^{-e^{c\Delta}}\big)\big)^2.$$

By (88) and (92), the probability $\eta_{0,\sigma^2}^{\times(n-r)}[A]$ dominates $e^{-e^{c\Delta}}$, so the claim follows by (71). □

In the case where h_1 and h_2 are such that their "branching point" happens after the scale $r+\Delta$, there is no hope of a decoupling of $J^-(h_1)$ and $J^-(h_2)$. Instead, we need to split the probability into a coupled part and a decoupled part and use Lemmas 3.9 and 3.10 separately.

LEMMA 3.12. *Let $h_1, h_2 \in \mathbb{R}$ and $l = h_1 \wedge h_2$. If $r + \Delta < l \le n - \Delta$, then*

$$\mathbb{P}\big[J^-(h_1) \cap J^-(h_2)\big] \le c2^{-(2n-l)} 2^{19\Delta + r} \frac{(n-r)^{(\frac{3}{2}+2\varepsilon)(2-\frac{l+3\Delta-r}{n-r})}}{(n-l-\Delta)^3 (l-\Delta-r)^{3/2}}. \tag{102}$$

PROOF. Write $\overline{X}_{r,n}(h) = \overline{X}_{r,l-\Delta}(h) + \overline{X}_{l-\Delta,n}(h)$ and decompose the event $J^-(h_1) \cap J^-(h_2)$ over the values of $\overline{X}_{r,l-\Delta}(h)$ as follows:

$$\bigcup_{q_1,q_2=0}^{\infty} \left(J^-(h_1) \cap J^-(h_2) \cap \bigcap_{i=1}^{2} \{\overline{X}_{r,l-\Delta}(h_i) \in [-q_i, -q_i + 1]\} \right).$$

For fixed q_1, q_2, the event in the union is contained in $\bigcap_{i=1,2} A_i(q_i) \cap C_i(q_i)$, where the events $A_i(q)$ are defined in (93) and for $i = 1, 2$,

$$\begin{aligned} C_i(q) = \big\{ &\overline{X}_{l-\Delta,n}(h_i) \in [q-1, q+\delta], \\ &\overline{X}_{l-\Delta,k}(h_i) \le q+1 \text{ for } k = l-\Delta, \dots, n \big\}. \end{aligned}$$

Now note that $(\overline{X}_{r,k}(h_i))_{r \le k \le l-\Delta}$ are independent from $(\overline{X}_{l-\Delta,k}(h_i))_{l-\Delta \le k \le n}$, for $i = 1, 2$. Altogether we get that

$$\begin{aligned} &\mathbb{P}\big[J^-(h_1) \cap J^-(h_2)\big] \\ &\quad \le \sum_{q_1,q_2=0}^{\infty} \mathbb{P}\big[A_1(q_1) \cap A_2(q_2)\big] \mathbb{P}\big[C_1(q_1) \cap C_2(q_2)\big]. \end{aligned} \tag{103}$$

Lemma 3.9 gives

$$\mathbb{P}\big[A_1(q_1) \cap A_2(q_2)\big] \le c \frac{e^{-(l-\Delta-r)\frac{\mu(-\varepsilon)^2}{2\sigma^2}}}{(l-\Delta-r)^{3/2}} (1+q_1) e^{\frac{\mu(-\varepsilon)}{2\sigma^2}(q_1+q_2)}.$$

In order to use Lemma 3.10, we express the probability on the event C_i's by conditioning on $\overline{X}_{l-\Delta,l+\Delta}(h_i)$, which are independent of $\overline{X}_{l+\Delta,n}(h_i)$. We have

$$\begin{aligned} &\mathbb{P}\big[C_1(q_1) \cap C_2(q_2)\big] \\ &\quad = \int_{\mathbb{R}^2} \mathbb{P}\big[B_1(q_1 - y_1) \cap B_1(q_2 - y_2)\big] f(y_1, y_2)\, dy_1\, dy_2, \end{aligned} \tag{104}$$

where $f(y_1, y_2)$ is the density of $(\overline{X}_{l-\Delta,l+\Delta}(h_i), i = 1, 2)$, and the events B_i's are as in (97) with $\delta' = 1$. Lemma 3.10 then gives

$$\begin{aligned} &\mathbb{P}\big[B_1(q_1 - y_1) \cap B_2(q_2 - y_2)\big] \\ &\quad \le c \frac{e^{-2(n-l-\Delta)\frac{\mu(-\varepsilon)^2}{2\sigma^2}}}{(n-l-\Delta)^3} \prod_{i=1,2} (1 + q_i - y_i) e^{-\frac{\mu(-\varepsilon)}{\sigma^2}(q_i - y_i)}, \end{aligned} \tag{105}$$

using also that $\eta_{0,\sigma^2}^{\times(n-l)}[\overline{B}_i(q_i - y_i)] \le c(1+q_i-y_i)/(n-l-\Delta)^{3/2}$ by (64) of the ballot theorem with $\delta'+\delta$ in place of δ and $b = q_i - y_i - \delta'$. Thus,

$$\begin{aligned}
&\mathbb{P}\big[J^-(h_1) \cap J^-(h_2)\big] \\
&\quad \le \int_{\mathbb{R}^2} \prod_{i=1,2} (1+q_i-y_i) e^{-\frac{\mu(-\varepsilon)}{\sigma^2}(q_i-y_i)} f(y_1,y_2)\, dy_1\, dy_2.
\end{aligned}$$

To handle the integral, note that Proposition 2.4 implies

$$\begin{aligned}
&\mathbb{E}\big[e^{\frac{\mu(-\varepsilon)}{\sigma^2}(\sum_{i=1,2}\overline{X}_{l-\Delta,l+\Delta}(h_i))}\big] \\
&\quad \le c \exp\Bigg(\sum_{k=l-\Delta+1}^{l+\Delta} \frac{\mu(-\varepsilon)^2}{\sigma^4}(\sigma_k^2+\rho_k)\Bigg) \le c e^{\Delta 16 \log 2},
\end{aligned} \tag{106}$$

where the last inequality follows from (25), (26) and the inequalities $\rho_k \le \sigma_k^2 \le 2\sigma^2$ and $\mu(-\varepsilon)/\sigma^2 \le 2$ [see (71)]. Using (38), the same estimate holds for

$$\mathbb{E}\big[\overline{X}_{l-\Delta,l+\Delta}(h_1) e^{\frac{\mu(-\varepsilon)}{\sigma^2}(\sum_{i=1,2}\overline{X}_{l-\Delta,l+\Delta}(h_i))}\big]$$

and

$$\mathbb{E}\Bigg[\prod_{i=1,2} \overline{X}_{l-\Delta,l+\Delta}(h_i) e^{\frac{\mu(-\varepsilon)}{\sigma^2}\overline{X}_{l-\Delta,l+\Delta}(h_i)}\Bigg].$$

After expanding the product in the integral, this altogether implies

$$\begin{aligned}
&\int_{\mathbb{R}^2} \prod_{i=1,2} (1+q_i-y_i) e^{\frac{\mu(-\varepsilon)}{\sigma^2}(y_i)} f(y_1,y_2)\, dy_1\, dy_2 \\
&\qquad \le c(1+q_1)(1+q_2) e^{\Delta 16 \log 2}.
\end{aligned} \tag{107}$$

Thus, equations (103) to (107) yield

$$\mathbb{P}\big[J^-(h_1) \cap J^-(h_2)\big] \le c 2^{16\Delta} \frac{e^{-(2(n-l-\Delta)+(l-\Delta-r))\frac{\mu(-\varepsilon)^2}{2\sigma^2}}}{(l-\Delta-r)^{3/2}(n-l-\Delta)^3},$$

where we used the fact that $\sum_{q_1,q_2=0}^{\infty}(1+q_1)^2(1+q_2)e^{-cq_1-cq_2}$ is finite. The claim then follows from (71). $\square$

The case where the branching point is between $r-\Delta$ and $r+\Delta$ is handled similarly.

LEMMA 3.13. *Let $h_1, h_2 \in \mathbb{R}$ with $l = h_1 \wedge h_2$ be such that $r-\Delta \le l \le r+\Delta$. Then*

$$\mathbb{P}\big[J^-(h_1) \cap J^-(h_2)\big] \le c 2^{18\Delta} 2^{-2(n-l-\Delta)}(n-r)^{4\varepsilon}. \tag{108}$$

PROOF. Since $r-\Delta<l\le r+\Delta$, we have the decomposition $\overline{X}_{r,n}(h)=\overline{X}_{r,l+\Delta}(h)+\overline{X}_{l+\Delta,n}(h)$. We proceed as in Lemma 3.12 by conditioning on $\overline{X}_{r,l+\Delta}(h_i)$, $i=1,2$, and then drop the barrier condition on $\overline{X}_{r,l+\Delta}(h_i)$ for both $i=1$ and $i=2$. Following (104) and (105), this gives

$$\begin{aligned}&\mathbb{P}\bigl[J^-(h_1)\cap J^-(h_2)\bigr]\\&\quad\le c\frac{e^{-2(n-l-\Delta)\frac{\mu(-\varepsilon)^2}{2\sigma^2}}}{(n-l-\Delta)^3}\int_{\mathbb{R}^2}\prod_{i=1,2}(1-y_i)e^{\frac{\mu(-\varepsilon)}{\sigma^2}y_i}f(y_1,y_2)\,dy_1\,dy_2,\end{aligned}$$

where $f(y_1,y_2)$ is now the density of $(\overline{X}_{r,l+\Delta}(h_i),i=1,2)$. The integral can be estimated using Proposition 2.4 as in (106). It is smaller than $c2^{16\Delta}$. By (71), the fraction in front of the integral is

$$2^{-2(n-l-\Delta)}(n-r)^{4\varepsilon\frac{n-l-\Delta}{n-r}}(n-r)^{3\frac{n-l-\Delta}{n-r}}/(n-l-\Delta)^3.$$

Since $r-\Delta<l<r+\Delta$, this is smaller than $c2^{-2(n-l-\Delta)}(n-r)^{4\varepsilon}$ as claimed. □

We now have the necessary two-point estimates to prove the upper bound on $\mathbb{E}[(Z^-)^2]$.

PROOF OF LEMMA 3.8. We split the sum in (91) into four terms depending on the branching point $h_1\wedge h_2$ of the pair $h_1,h_2\in\mathcal{H}_n\cap[0,1)$:

$$\begin{aligned}&\underbrace{\sum_{h_1,h_2:h_1\wedge h_2\le r-\Delta}(\cdot)}_{(I)}+\underbrace{\sum_{h_1,h_2:r-\Delta<h_1\wedge h_2\le r+\Delta}(\cdot)}_{(II)}\\&\quad+\underbrace{\sum_{h_1,h_2:r+\Delta<h_1\wedge h_2<n-\Delta}(\cdot)}_{(III)}+\underbrace{\sum_{h_1,h_2:h_1\wedge h_2\ge n-\Delta}(\cdot)}_{(IV)}.\end{aligned}$$

Using that $\#\mathcal{H}_n\cap[0,1)=2^n$ and the bound (101), we get

$$(I)\le\bigl(1+o(1)\bigr)\bigl(2^r(n-r)^{\frac{3}{2}+2\varepsilon}\eta_{0,\sigma^2}^{\times(n-r)}[A]\bigr)^2.$$

By (88), the right-hand side is at least $c2^{2r}(n-r)^{4\varepsilon}$. We now show that (II), (III) and (IV) are negligible compared to this, and thus (I) is the dominant term in the sum. Note that the number of pairs $h_1,h_2\in\mathcal{H}_n\cap[0,1)$ such that $2^{-l-1}\le|h_1-h_2|\le2^{-l}$ is at most $c2^{2n-l}$. Thus, the contribution of (II), by Lemma 3.13, is at most

$$\begin{aligned}(II)&\le c\sum_{l=r-\Delta+1}^{r+\Delta}2^{2n-l}2^{16\Delta}2^{-2(n-l-\Delta)}(n-r)^{4\varepsilon}\\&\le c2^{19\Delta}2^r(n-r)^{4\varepsilon},\end{aligned}$$

which is negligible compared to $2^{2r}(n-r)^{4\varepsilon}$, because of the choice $\Delta = r/100$. Similarly, the contribution of (*III*) can be bounded as

$$(III) \le \sum_{l=r+\Delta+1}^{n-\Delta-1} 2^{2n-l} \max_{h\in[2^{-l-1},2^l]} \mathbb{P}\big[J^-(0)\cap J^-(h)\big].$$

Lemma 3.12 then yields

$$\begin{aligned}(III) &\le c2^{r+19\Delta}(n-r)^{4\varepsilon} \sum_{l=r+\Delta+1}^{n-\Delta-1} \frac{(n-r)^{\frac{3}{2}(2-\frac{l+3\Delta-r}{n-r})}}{(n-l-\Delta)^3(l-\Delta-r)^{3/2}} \\ &= c2^{r+19\Delta}(n-r)^{4\varepsilon} \sum_{a=1}^{m-2\Delta-1} \frac{m^{\frac{3}{2}(2-(a+2\Delta)/m)}}{(m-a-2\Delta)^3 a^{3/2}} \qquad \text{for } m=n-r \\ &\le c2^{r+19\Delta}(n-r)^{4\varepsilon},\end{aligned}$$

where the last inequality follows from the fact that the sum over a stays finite as $m\to\infty$. Since $\Delta = r/100$, the bound on (*III*) is negligible relative to the bound on (I). Finally, for (*IV*), the event $J^-(h_2)$ can be dropped. There are at most $2^{n+\Delta}$ pairs $h_1, h_2 \in \mathcal{H}_n \cap [0,1)$ such that $|h_1-h_2| \le 2^{-n+\Delta}$. A union bound using the one-point bound (90) gives

$$(IV) \le 2^{n+\Delta}\mathbb{P}\big[J^-(0)\big] \le \big(1+o(1)\big)2^{r+\Delta}(n-r)^{2\varepsilon}.$$

Again, this is negligible relative to the bound on (I). Therefore,

$$(I)+(II)+(III)+(IV) \le \big(1+o(1)\big)\big(2^r(n-r)^{\frac{3}{2}+2\varepsilon}\eta_{0,\sigma^2}^{\times(n-r)}[A]\big)^2,$$

which proves the lemma. □

This bound on the second moment of Z^- concludes the proof of lower bound Proposition 3.6 and, therefore, also for the main result Theorem 1.2

Acknowledgments. We would like to thank the referees of this paper for their detailed and helpful comments and corrections.

REFERENCES

[1] ADDARIO-BERRY, L. and REED, B. (2009). Minima in branching random walks. *Ann. Probab.* **37** 1044–1079. MR2537549

[2] AÏDÉKON, E. (2013). Convergence in law of the minimum of a branching random walk. *Ann. Probab.* **41** 1362–1426. MR3098680

[3] AÏDÉKON, E., BERESTYCKI, J., BRUNET, É. and SHI, Z. (2013). Branching Brownian motion seen from its tip. *Probab. Theory Related Fields* **157** 405–451. MR3101852

[4] ARGUIN, L.-P., BOVIER, A. and KISTLER, N. (2013). The extremal process of branching Brownian motion. *Probab. Theory Related Fields* **157** 535–574. MR3129797

[5] BELIUS, D. and KISTLER, N. (2016). The subleading order of two dimensional cover times. *Probab. Theory Related Fields*. To appear. DOI:10.1007/s00440-015-0689-6.

[6] BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0436272

[7] BISKUP, M. and LOUIDOR, O. (2016). Extreme local extrema of two-dimensional discrete Gaussian free field. *Comm. Math. Phys.* **345** 271–304. MR3509015

[8] BONDARENKO, A. and SEIP, K. (2015). Large GCD sums and extreme values of the Riemann zeta function. Preprint. Available at arXiv:1507.05840.

[9] BOURGADE, P. (2010). Mesoscopic fluctuations of the zeta zeros. *Probab. Theory Related Fields* **148** 479–500. MR2678896

[10] BRAMSON, M. D. (1978). Maximal displacement of branching Brownian motion. *Comm. Pure Appl. Math.* **31** 531–581. MR0494541

[11] BRAMSON, M., DING, J. and ZEITOUNI, O. (2014). Convergence in law of the maximum of nonlattice branching random walk. Preprint. Available at arXiv:1404.3423.

[12] BRAMSON, M., DING, J. and ZEITOUNI, O. (2016). Convergence in law of the maximum of the two-dimensional discrete Gaussian free field. *Comm. Pure Appl. Math.* **69** 62–123. MR3433630

[13] FARMER, D. W., GONEK, S. M. and HUGHES, C. P. (2007). The maximum size of L-functions. *J. Reine Angew. Math.* **609** 215–236. MR2350784

[14] FYODOROV, Y. V. and BOUCHAUD, J.-P. (2008). Freezing and extreme-value statistics in a random energy model with logarithmically correlated potential. *J. Phys. A* **41** 372001. MR2430565

[15] FYODOROV, Y. V., HIARY, G. A. and KEATING, J. P. (2012). Freezing transition, characteristic polynomials of random matrices, and the Riemann zeta function. *Phys. Rev. Lett.* **108** 170601.

[16] FYODOROV, Y. V. and KEATING, J. P. (2014). Freezing transitions and extreme values: Random matrix theory, and disordered landscapes. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **372** 20120503. MR3151088

[17] FYODOROV, Y. V. and SIMM, N. J. (2015). On the distribution of maximum value of the characteristic polynomial of gue random matrices. Preprint. Available at arXiv:1503.07110.

[18] GÖTZE, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724–739. MR1106283

[19] HALÁSZ, G. (1983). On random multiplicative functions. In *Hubert Delange Colloquium* (*Orsay*, 1982). *Publ. Math. Orsay* **83** 74–96. Univ. Paris XI, Orsay. MR0728404

[20] HARPER, A. J. (2013). A note on the maximum of the Riemann zeta function, and log-correlated random variables. Preprint. Available at arXiv:1304.0677.

[21] KEATING, J. P. and SNAITH, N. C. (2000). Random matrix theory and $\zeta(1/2+it)$. *Comm. Math. Phys.* **214** 57–89. MR1794265

[22] KISTLER, N. (2015). Derrida's random energy models. From spin glasses to the extremes of correlated random fields. In *Correlated Random Systems*: *Five Different Methods*. *Lecture Notes in Math.* **2143** 71–120. Springer, Cham. MR3380419

[23] MADAULE, T. (2015). Maximum of a log-correlated Gaussian field. *Ann. Inst. Henri Poincaré Probab. Stat.* **51** 1369–1431. MR3414451

[24] MONTGOMERY, H. L. and VAUGHAN, R. C. (2007). *Multiplicative Number Theory. I. Classical Theory*. *Cambridge Studies in Advanced Mathematics* **97**. Cambridge Univ. Press, Cambridge. MR2378655

[25] SELBERG, A. (1946). Contributions to the theory of the Riemann zeta-function. *Arch. Math. Naturvidensk.* **48** 89–155. MR0020594

[26] SOUNDARARAJAN, K. (2008). Extreme values of zeta and L-functions. *Math. Ann.* **342** 467–486. MR2425151

[27] SOUNDARARAJAN, K. (2009). Moments of the Riemann zeta function. *Ann. of Math.* (2) **170** 981–993. MR2552116

[28] WEBB, C. (2011). Exact asymptotics of the freezing transition of a logarithmically correlated random energy model. *J. Stat. Phys.* **145** 1595–1619. MR2863721

[29] WEBB, C. (2015). The characteristic polynomial of a random unitary matrix and Gaussian multiplicative chaos—The L^2-phase. *Electron. J. Probab.* **20** 104. MR3407221

L.-P. ARGUIN
DÉPARTEMENT DE MATHÉMATIQUES
ET STATISTIQUE
UNIVERSITÉ DE MONTRÉAL
MONTRÉAL, QUEBEC H3T 1J4
CANADA
AND
DEPARTMENT OF MATHEMATICS
BARUCH COLLEGE AND GRADUATE CENTER
CITY UNIVERSITY OF NEW YORK
NEW YORK, NEW YORK 10010
USA
E-MAIL: louis-pierre.arguin@baruch.cuny.edu

D. BELIUS
COURANT INSTITUTE
OF MATHEMATICAL SCIENCES
NEW YORK UNIVERSITY
NEW YORK, NEW YORK 10012
USA
E-MAIL: david.belius@cantab.net

A. J. HARPER
JESUS COLLEGE
CAMBRIDGE CB5 8BL
UNITED KINGDOM
E-MAIL: A.J.Harper@dpmms.cam.ac.uk