LEVEL-SET PERCOLATION OF THE GAUSSIAN FREE FIELD ON REGULAR GRAPHS II: FINITE EXPANDERS

Angelo Abächerli * Jiří Černý †

Preliminary draft

Abstract

We consider the zero-average Gaussian free field on a certain class of finite d-regular graphs for fixed $d \geq 3$. This class includes d-regular expanders of large girth and typical realisations of random d-regular graphs. We show that the level set of the zero-average Gaussian free field above level h exhibits a phase transition at level $h_\star$, which agrees with the critical value for level-set percolation of the Gaussian free field on the *infinite d-regular tree*. More precisely, we show that, with probability tending to one as the size of the *finite* graphs tends to infinity, the level set above level h does not contain any connected component of larger than logarithmic size whenever $h > h_\star$, and on the contrary, whenever $h < h_\star$, a linear fraction of the vertices is contained in connected components of the level set above level h having a size of at least a small fractional power of the total size of the graph. It remains open whether in the supercritical phase $h < h_\star$, as the size of the graphs tends to infinity, one observes the emergence of a (potentially unique) *giant* connected component of the level set above level h. The proofs in this article make use of results from the accompanying paper [AČ19].

0 Introduction

In this article we study level-set percolation of the zero-average Gaussian free field on a class of large d-regular graphs with $d \geq 3$. This class contains d-regular expanders of large girth and typical realisations of random d-regular graphs. Through suitable local approximations of the zero-average Gaussian free field by the Gaussian free field on the infinite d-regular tree we are able to establish a phase transition for level-set percolation of the zero-average Gaussian free field which occurs at the critical value for level-set percolation in the infinite model, that is, on the d-regular tree.

Level-set percolation and the local picture of the zero-average Gaussian free field have been previously studied by the first author in [Abä19] for the situation where the underlying sequence of finite graphs is given by the discrete tori of growing side length in dimension $d \geq 3$. The motivation for investigating the zero-average Gaussian free field on the different class of finite graphs considered here (see (0.1)–(0.3) below) stems from the insight that analysing probabilistic models on these types of finite graphs has led to often very explicit and strong results over the years. Examples include the emergence of a giant connected component for Bernoulli bond percolation (see e.g. [ABS04] and recently

*Departement Mathematik, ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland

†Departement Mathematik und Informatik, University of Basel, Spiegelgasse 1, 4051 Basel, Switzerland

[KLS18]), cutoff phenomena for random walks (see e.g. [LS10]) and the appearance of a giant connected component in the vacant set of simple random walk (see e.g. [ČTW11]). Actually, we will borrow the assumptions (0.1)–(0.3) on the finite graphs from [ČTW11].

From a more general perspective, level-set percolation of the Gaussian free field is a significant representative of a percolation model with long-range dependencies and it has attracted attention for a long time, dating back to [MS83], [LS86] and [BLM87]. More recent developments can be found for instance in [RS13], [PR15], [Szn15], [DPR18b] and [DPR18a]. For the particular case of the Gaussian free field on regular trees we also refer to [Szn16], [Szn19] and [AČ19]; for more general transient trees to [AS18].

We now describe our results more precisely. We let $d \geq 3$ and assume that $(\mathcal{G}_n)_{n\geq 1}$ is a sequence of graphs satisfying the following conditions.

Assumptions. There exist some $\alpha, \beta > 0$ and an increasing sequence of positive integers $(N_n)_{n\geq 1}$ with $N_n \xrightarrow{n\to\infty} \infty$ such that for all $n \geq 1$

- $\mathcal{G}_n$ is d-regular, connected and has N_n vertices (0.1)
- for all $x \in \mathcal{G}_n$ there is at most one cycle in the ball of radius $\lfloor \alpha \log_{d-1}(N_n) \rfloor$ around x (0.2)
- the spectral gap of $\mathcal{G}_n$, denoted by $\lambda_{\mathcal{G}_n}$, satisfies $\lambda_{\mathcal{G}_n} \geq \beta$. (0.3)

Here by spectral gap we mean the smallest non-zero eigenvalue of $I - P$, where I is the identity matrix and P is the transition matrix of the simple random walk on the graph (see also [SC97], Definition 2.1.3 and beneath it). For an explanation of why these assumptions are satisfied by d-regular expanders of large girth and by typical realisations of random d-regular graphs we refer to [ČTW11], Section 2.2 and Remark 1.4.

On $\mathcal{G}_n$ we consider the zero-average Gaussian free field (see Section 1.2 for more details about it) with law $\mathbb{P}^{\mathcal{G}_n}$ on $\mathbb{R}^{\mathcal{G}_n}$ and canonical coordinate process $(\Psi_{\mathcal{G}_n}(x))_{x\in\mathcal{G}_n}$ so that,

under $\mathbb{P}^{\mathcal{G}_n}$, $(\Psi_{\mathcal{G}_n}(x))_{x\in\mathcal{G}_n}$ is a centred Gaussian field on $\mathcal{G}_n$ with covariance $\mathbb{E}^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(x)\Psi_{\mathcal{G}_n}(y)] = G_{\mathcal{G}_n}(x,y)$ for all $x, y \in \mathcal{G}_n$, where $G_{\mathcal{G}_n}(\cdot,\cdot)$ is the zero-average Green function on $\mathcal{G}_n$ (see (1.16)). (0.4)

The zero-average Gaussian free field is a natural version of the Gaussian free field for finite graphs. However, due to the zero-average property (see below (1.18)), it comes with some peculiarities like the *lack* of an FKG-inequality and of the domain Markov property.

Our main interest lies in analysing the size (i.e. the number of contained vertices) of the connected components of the level sets of $\Psi_{\mathcal{G}_n}$, i.e. of

$$E^{\geq h}_{\Psi_{\mathcal{G}_n}} := \{x \in \mathcal{G}_n \,|\, \Psi_{\mathcal{G}_n}(x) \geq h\} \text{ for } h \in \mathbb{R}. \tag{0.5}$$

In order to do so, it will be helpful to locally describe $\Psi_{\mathcal{G}_n}$ via the Gaussian free field on the infinite d-regular tree $\mathbb{T}_d$ with root denoted by o, that is, the centred Gaussian field on $\mathbb{T}_d$ with law $\mathbb{P}^{\mathbb{T}_d}$ on $\mathbb{R}^{\mathbb{T}_d}$ and canonical coordinate process $(\varphi_{\mathbb{T}_d}(x))_{x\in\mathbb{T}_d}$ so that,

under $\mathbb{P}^{\mathbb{T}_d}$, $(\varphi_{\mathbb{T}_d}(x))_{x\in\mathbb{T}_d}$ is a centred Gaussian field on $\mathbb{T}_d$ with covariance $\mathbb{E}^{\mathbb{T}_d}[\varphi_{\mathbb{T}_d}(x)\varphi_{\mathbb{T}_d}(y)] = g_{\mathbb{T}_d}(x,y)$ for all $x, y \in \mathbb{T}_d$, where $g_{\mathbb{T}_d}(\cdot,\cdot)$ is the Green function of simple random walk on $\mathbb{T}_d$ (see (1.6)). (0.6)

The Gaussian free field on $\mathbb{T}_d$ has first been studied in [Szn16]. Recently, more refined results have been obtained by the authors in the accompanying paper [AČ19]. These results lay the groundwork for the present article and they will be central in our analysis of the zero-average Gaussian free field on the graphs $(\mathcal{G}_n)_{n\geq 1}$. For now, we only recall the critical value of level-set percolation of $\varphi_{\mathbb{T}_d}$, that is,

$$h_\star := \inf \Big\{ h \in \mathbb{R} \,\Big|\, \mathbb{P}^{\mathbb{T}_d}\big[\big|\mathcal{C}_{\mathrm{o}}^{\mathbb{T}_d,h}\big| = \infty\big] = 0 \Big\}, \tag{0.7}$$

where $\mathcal{C}_{\mathrm{o}}^{\mathbb{T}_d,h}$ is the connected component of the level set $E_{\varphi_{\mathbb{T}_d}}^{\geq h} := \{x \in \mathbb{T}_d \,|\, \varphi_{\mathbb{T}_d}(x) \geq h\}$ of $\varphi_{\mathbb{T}_d}$ above level h containing the root $\mathrm{o} \in \mathbb{T}_d$. There is a crucial spectral characterisation of $h_\star$ derived in [Szn16], which leads to the proof of $0 < h_\star < \infty$ on $\mathbb{T}_d$ for $d \geq 3$ (see [Szn16], Proposition 3.3 and Corollary 4.5). Actually, in the accompanying paper [AČ19] we make heavy use of this characterisation to obtain new results about $\varphi_{\mathbb{T}_d}$ on $\mathbb{T}_d$.

Our main results concerning the size of the connected components of the level sets of $\Psi_{\mathcal{G}_n}$ on the finite graphs $(\mathcal{G}_n)_{n\geq 1}$ satisfying (0.1)–(0.3) are the following: we show in essence that (see Section 3, Theorem 3.1, for the precise statement)

> in the subcritical phase $h > h_\star$, with high probability for large n, the level set $E_{\Psi_{\mathcal{G}_n}}^{\geq h}$ of $\Psi_{\mathcal{G}_n}$ only contains microscopic connected components (i.e. containing at most a logarithmic number of vertices of $\mathcal{G}_n$); (0.8)

and furthermore that (see Section 4, Theorem 4.1, for the precise statement)

> in the supercritical phase $h < h_\star$, with high probability for large n, a linear fraction of the vertices of $\mathcal{G}_n$ is contained in at least mesoscopic connected components of the level set $E_{\Psi_{\mathcal{G}_n}}^{\geq h}$ of $\Psi_{\mathcal{G}_n}$ (i.e. containing a fractional power of the number of vertices of $\mathcal{G}_n$). (0.9)

Although giving a strong hint to, the result (0.9) leaves open whether in the supercritical phase $h < h_\star$, with high probability for large n, there actually is a macroscopic (*giant*) connected component in the level set above level h, i.e. containing a number of vertices comparable to $\mathcal{G}_n$. Furthermore, in the affirmative, one could ask if this giant component is unique, that is, if the second-largest connected component of the level set above level $h < h_\star$ only contains a negligible number of vertices compared to $\mathcal{G}_n$ (see also Remark 4.7).

As a comparison, the emergence of a unique giant connected component in the supercritical phase has been shown for Bernoulli bond percolation on d-regular expanders of large girth in [ABS04] (see also [KLS18]) and for vacant-set percolation of simple random walk on exactly the same graphs $(\mathcal{G}_n)_{n\geq 1}$ like here in [ČTW11]. In the latter, this result is achieved by relating the model to vacant-set percolation of random interlacements on $\mathbb{T}_d$. Subsequently, more refined results have been obtained about the vacant set of simple random walk on random regular graphs in [CF13] and [ČT13].

In the models mentioned above, the assertion of existence and uniqueness of a giant component in the supercritical phase is achieved by a 'sprinkling argument' starting from a statement like (0.9). In our situation, it would correspond to showing that distinct mesoscopic connected components of $E_{\Psi_{\mathcal{G}_n}}^{\geq h}$ for a supercritical level $h < h_\star$ are going to be connected at a slightly smaller level $h' < h$ with high probability, thus forming large clusters. As [ČTW11] shows, it can be very involved to carry out sprinkling arguments in the non-i.i.d. setting. At present we have not been able to do it in our context, one of the main restrictions stemming from the defining zero-average property of the fields we are considering (see below (1.18)). We point out that sprinkling techniques have been already applied in the discussion of level-set percolation of the Gaussian free field in [DR15] to construct an infinite connected component with the underlying graph being $\mathbb{Z}^d$ for high dimension d.

Let us now comment on the proofs of Theorem 3.1 and Theorem 4.1 (corresponding to (0.8) and (0.9)). In both cases, the general philosophy is to locally approximate $\Psi_{\mathcal{G}_n}$ on the finite graphs by $\varphi_{\mathbb{T}_d}$ on the d-regular tree and by that reduce the analysis to the infinite model, which is easier to understand. A similar strategy has been successfully carried out in [ABS04] and [ČTW11] where the connected components in question are locally approximated by Galton-Watson trees. In our setting the situation is considerably more complicated since neither the connected components of the level sets of $\Psi_{\mathcal{G}_n}$ nor the connected components of the level sets of $\varphi_{\mathbb{T}_d}$ (used in the approximation) are locally Galton-Watson trees, even if the connected components of $E^{\geq h}_{\varphi_{\mathbb{T}_d}}$ share some global properties with them, as shown in [AČ19]. The exact way how the local approximation by $\varphi_{\mathbb{T}_d}$ is performed differs considerably between the subcritical and supercritical phase.

In the supercritical phase $h < h_\star$, we use an approximation of $\Psi_{\mathcal{G}_n}$ by $\varphi_{\mathbb{T}_d}$ via local charts around vertices of $\mathcal{G}_n$ with a tree-like neighbourhood (Theorem 2.1). Then the proof of Theorem 4.1 (corresponding to (0.9)) is, roughly said, a second moment computation based on this local approximation and involving a good control of the supercritical level sets of $\varphi_{\mathbb{T}_d}$, obtained in the accompanying paper [AČ19].

More precisely, to show (0.9) we prove that the number of vertices contained in mesoscopic connected components of the level set $E^{\geq h}_{\Psi_{\mathcal{G}_n}}$ concentrates around its expectation, which we show to grow linearly in the total number of vertices. The concentration follows by a variance computation and a second moment inequality. Actually, when estimating the expectation and variance, it is enough to consider only vertices with a tree-like neighbourhood since the assumption (0.2) (together with (0.1)) guarantees that the number of vertices having a tree-like neighbourhood is comparable to the total number of vertices in $\mathcal{G}_n$ (Remark 4.3). Thanks to the approximation of $\Psi_{\mathcal{G}_n}$ by $\varphi_{\mathbb{T}_d}$ around such vertices (Theorem 2.1 mentioned above), we are able to transfer the computations to the regular tree. The linear lower bound on the expectation ((4.9) in Lemma 4.4) now follows rather direct from this approximation and from [AČ19], Theorem 4.3, showing that connected components of the level sets of $\varphi_{\mathbb{T}_d}$ are mesoscopic with positive probability in the supercritical phase. The control of the variance follows along similar lines (Lemma 4.6). It requires the approximation of $\Psi_{\mathbb{T}_d}$ by $\varphi_{\mathbb{T}_d}$ on neighbourhoods of vertices with a tree-like *and* disjoint neighbourhood. This is provided by Theorem 2.1 as well. Once we have reduced the computations to quantities for $\varphi_{\mathbb{T}_d}$ on $\mathbb{T}_d$, we can apply a decoupling inequality ([PR15], Corollary 1.3) and deduce the bound on the variance again from results on $\varphi_{\mathbb{T}_d}$ developed in the accompanying paper [AČ19].

For the subcritical phase $h > h_\star$ (Theorem 3.1 corresponding to (0.8)) the local approximation of $\Psi_{\mathcal{G}_n}$ by $\varphi_{\mathbb{T}_d}$ around vertices with tree-like neighbourhood is not good enough. On the one hand, the connected components of $E^{\geq h}_{\Psi_{\mathcal{G}_n}}$ may have a diameter that is larger than the diameter of those neighbourhoods (at least if h is close to $h_\star$). On the other hand, one expects that the connected components are typically 'thin'. These two points of 'thinness' and of 'escaping the local charts' suggest that the approximation of $\Psi_{\mathcal{G}_n}$ by $\varphi_{\mathbb{T}_d}$ should rather be carried out *along* the connected components of $E^{\geq h}_{\Psi_{\mathcal{G}_n}}$. We achieve this by employing an exploration process uncovering the connected component of the level set containing a given vertex (Algorithm 1 in Section 3). Roughly said, by exploring $\Psi_{\mathcal{G}_n}$ vertex by vertex we are able to couple it vertex by vertex to a number of independent copies of $\varphi_{\mathbb{T}_d}$ on $\mathbb{T}_d$, hence bringing back the problem to the tree. Results from [AČ19] on $\varphi_{\mathbb{T}_d}$ in the subcritical phase then conclude the proof.

More precisely, the exploration process aggregates the vertices found in the connected component of $E^{\geq h}_{\Psi_{\mathcal{G}_n}}$ containing a fixed $x \in \mathcal{G}_n$ into a union of disjoint subtrees of $\mathcal{G}_n$. The decomposition into a union of disjoint subtrees is determined *during* the exploration and it is dictated by the *geometric* properties of the graph $\mathcal{G}_n$ and of the evolving set of explored vertices. These geometric conditions guarantee that for each of the disjoint subtrees we can approximate the zero-average Gaussian free field $\Psi_{\mathcal{G}_n}$ on the subtree by an independent copy of the Gaussian free field $\varphi_{\mathbb{T}_d}$ on $\mathbb{T}_d$ (Lemma 3.4). In order to do so, it is crucial to have a good understanding of the conditional distribution of the zero-average Gaussian free field (Lemma 2.6 and Proposition 2.7). As a consequence, the size of each disjoint subtree of $\mathcal{G}_n$ constructed by the exploration process is dominated by the size of the connected component containing the root $\mathrm{o} \in \mathbb{T}_d$ of the level set of $\varphi_{\mathbb{T}_d}$ above a slightly lower level $h - \varepsilon$ (Corollary 3.5). The last two ingredients for the proof of (0.8) are now a control on the number of disjoint subtrees (Lemma 3.3, already proven in [ČTW11]) and a control on the exponential moments of the size of the connected component of the level set of $\varphi_{\mathbb{T}_d}$ containing the root $\mathrm{o} \in \mathbb{T}_d$ in the subcritical phase (see [AČ19], Theorem 5.1).

Incidentally, let us point out that exploration processes are frequently used in the Bernoulli percolation literature and actually, a variant of such an algorithm was applied in [ČTW11] to deal with the vacant set of simple random walk in the subcritical phase. However, in our setting we cannot follow the 'standard' procedure. Usually, to show statements like (0.8), a good control on the termination time of the exploration process is necessary, i.e. on the time by when the connected component is completely uncovered. This is typically done by comparing the number of yet unexplored vertices to a random walk of negative drift. In our case this is not possible, essentially again because locally the connected components of $E^{\geq h}_{\Psi_{\mathcal{G}_n}}$ are not approximated by Galton-Watson trees (as mentioned earlier).

The structure of the article is as follows. In Section 1 we collect the notation and some results on the Gaussian free fields on both the finite graphs and the infinite tree. In particular, in Section 1.1 we recall results on $\varphi_{\mathbb{T}_d}$ from [Szn16] and [AČ19]. Then in Section 2 we investigate the local picture of the zero-average Gaussian free field on $\mathcal{G}_n$ and its connection to the Gaussian free field on $\mathbb{T}_d$. The content of these first two sections will be subsequently used to show Theorem 3.1 (corresponding to (0.8)) and Theorem 4.1 (corresponding to (0.9)). More precisely, in Section 3 we deal with the subcritical phase, ultimately proving the non-existence of connected components of $E^{\geq h}_{\Psi_{\mathcal{G}_n}}$ for $h > h_\star$ of larger than logarithmic size (Theorem 3.1). Finally, in Section 4 we conclude with the proof of Theorem 4.1 showing that for $h < h_\star$ most vertices of $\mathcal{G}_n$ live in a connected component of $E^{\geq h}_{\Psi_{\mathcal{G}_n}}$ of at least mesoscopic size.

Acknowledgements. The authors wish to express their gratitude to A.-S. Sznitman for suggesting the problem and for the valuable comments made at various stages of the project.

1 Notation and useful results

In this section we introduce our main notation and recall the essential material about the Gaussian free field on the d-regular tree $\mathbb{T}_d$ that will be needed in the study of the zero-average Gaussian free field on the finite graphs $(\mathcal{G}_n)_{n\geq 1}$ (Section 1.1). We end the section with results on the zero-average Green function and some basic properties of the zero-average Gaussian free field on $\mathcal{G}_n$ (Section 1.2).

As mentioned earlier, we consider for fixed $d \geq 3$ the d-regular graphs $(\mathcal{G}_n)_{n\geq 1}$, satisfying the assumptions (0.1)–(0.3). For the constants α and β appearing in these assumptions we assume without loss of generality that

$$\alpha \leq 1 \qquad \text{and} \qquad \beta \leq 2. \tag{1.1}$$

Indeed, for α this is trivial and for β it follows from the fact that the matrix P (see below (0.3)) is a symmetric stochastic matrix and thus all its eigenvalues are contained in the interval $[-1, 1]$. Consequently the eigenvalues of $I - P$ are contained in $[0, 2]$.

For the general graph notation introduced in the next two paragraphs, $\mathcal{G}$ stands either for $\mathcal{G}_n$ or for $\mathbb{T}_d$ with root o.

By $x \in \mathcal{G}$ resp. $U \subseteq \mathcal{G}$ we mean a vertex resp. a subset of vertices of the graph $\mathcal{G}$. We let $d_{\mathcal{G}}(\cdot,\cdot)$ denote the graph distance on $\mathcal{G}$. For any $U \subseteq \mathcal{G}$, $|U|$ stands for its cardinality, and $\partial_{\mathcal{G}} U := \{y \in \mathcal{G} \setminus U \,|\, y \text{ has some neighbour } x \in U \text{ in } \mathcal{G}\}$ denotes its (outer) boundary in $\mathcal{G}$. For any $R \geq 0$ and $x \in \mathcal{G}$ we define the balls and spheres of radius R around x to be $B_{\mathcal{G}}(x, R) := \{y \in \mathcal{G} \,|\, d_{\mathcal{G}}(x, y) \leq R\}$ and $S_{\mathcal{G}}(x, R) := \{y \in \mathcal{G} \,|\, d_{\mathcal{G}}(x, y) = R\}$. The maximum number of edges that can be deleted from the subgraph of $\mathcal{G}$ induced by some connected subset $U \subseteq \mathcal{G}$ while keeping it connected is called tree excess of U and we denote it by $\texttt{tx}(U)$. Note that $\texttt{tx}(U) = 0$ if and only if (the subgraph induced by) U is a tree. (In particular, the assumption (0.2) could be rewritten as $\texttt{tx}(B_{\mathcal{G}_n}(x, \lfloor \alpha \log_{d-1}(N_n) \rfloor)) \leq 1$ for all $n \geq 1$ and $x \in \mathcal{G}_n$.) For $x, z \in \mathcal{G}$ a path from x to z is a sequence of vertices $x = y_0, y_1, \ldots, y_m = z$ in $\mathcal{G}$ for some $m \geq 0$ such that y_i and y_{i-1} are neighbours for all $i = 1, \ldots, m$ (if $m \geq 1$). It is a *non-backtracking* path from x to z if in addition $y_i \neq y_{i-2}$ for all $i = 2, \ldots, m$ (if $m \geq 2$).

We write $P_x^{\mathcal{G}}$ for the canonical law of the simple random walk on $\mathcal{G}$ starting at $x \in \mathcal{G}$ as well as $E_x^{\mathcal{G}}$ for the corresponding expectation. The canonical process for the discrete-time walk is denoted by $(X_k)_{k\geq 0}$. For the continuous-time walk with i.i.d. mean-one exponential holding times we write $(\overline{X}_t)_{t\geq 0}$. Given $U \subseteq \mathcal{G}$ we write $T_U := \inf\{k \geq 0 \,|\, X_k \notin U\}$ for the exit time from U and $H_U := \inf\{k \geq 0 \,|\, X_k \in U\}$ for the entrance time in U of the discrete-time walk (here we set $\inf \emptyset := \infty$). For the continuous-time simple random walk T_U and H_U are defined accordingly. In the special case of $U = \{z\}$ we use H_z in place of $H_{\{z\}}$.

For $\mathcal{G} = \mathbb{T}_d$ we need some extra notation. In this case, there is a unique non-backtracking path of length $d_{\mathbb{T}_d}(x, z)$ between any two vertices $x, z \in \mathbb{T}_d$ (namely the *geodesic* path). For $x \in \mathbb{T}_d \setminus \{\text{o}\}$ let $\overline{x}$ be the unique neighbour of x on the non-backtracking path from x to o. Moreover, let $\overline{\text{o}} \in \mathbb{T}_d$ denote a fixed neighbour of the root $\text{o} \in \mathbb{T}_d$. For $x \in \mathbb{T}_d$ we define

$$U_x := \{z \in \mathbb{T}_d \,|\, \text{the non-backtracking path from } z \text{ to } x \text{ does not contain } \overline{x}\}. \tag{1.2}$$

In particular $\mathbb{T}_d = \{\text{o}\} \cup \bigcup_{i=1}^{d} U_{x_i}$ if $S_{\mathbb{T}_d}(\text{o}, 1) =: \{x_1, \ldots, x_d\}$. In the special case of $x = \text{o}$

we write $\mathbb{T}_d^+ \coloneqq U_{\rm o}$. We also set $B^+_{\mathbb{T}_d}({\rm o},R) \coloneqq \{y \in \mathbb{T}_d^+ \,|\, d_{\mathbb{T}_d}({\rm o},y) \le R\}$ and similarly $S^+_{\mathbb{T}_d}({\rm o},R) \coloneqq \{y \in \mathbb{T}_d^+ \,|\, d_{\mathbb{T}_d}({\rm o},y) = R\}$ for $R \ge 0$.

Finally, some notation for the finite graphs $(\mathcal{G}_n)_{n\ge1}$. For all $n \ge 1$ and $x \in \mathcal{G}_n$ we fix a cover tree $\pi_{n,x}$ of $\mathcal{G}_n$ at x, that is, a surjective map $\pi_{n,x} : \mathbb{T}_d \to \mathcal{G}_n$ such that $\pi_{n,x}({\rm o}) = x$ and such that for all $y \in \mathbb{T}_d$ one has $\pi_{n,x}(S_{\mathbb{T}_d}(y,1)) = S_{\mathcal{G}_n}(\pi_{n,x}(y),1)$, meaning that $\pi_{n,x}$ preserves the neighbourhood of radius 1 of any $y \in \mathbb{T}_d$. Note that:

- if $x \in \mathcal{G}_n$ with $\texttt{tx}(B_{\mathcal{G}_n}(x,R)) = 0$ for some $R \ge 0$, then the map $\pi_{n,x}$ restricted to $B_{\mathbb{T}_d}({\rm o},R)$ induces a graph isomorphism from $B_{\mathbb{T}_d}({\rm o},R)$ to $B_{\mathcal{G}_n}(x,R)$ (1.3)
- a sequence of vertices ${\rm o} = y_0, y_1, \dots, y_m \in \mathbb{T}_d$, $m \ge 0$, is a non-backtracking path in $\mathbb{T}_d$ starting at o if and only if $x = \pi_{n,x}(y_0), \pi_{n,x}(y_1), \dots, \pi_{n,x}(y_m) \in \mathcal{G}_n$ is a non-backtracking path in $\mathcal{G}_n$ starting at x. (1.4)

Furthermore, for the cover tree $\pi_{n,x}$ of $\mathcal{G}_n$ at x, the process $(\pi_{n,x}(X_k))_{k\ge0}$ under $P^{\mathbb{T}_d}_{\rm o}$ has the same law as $(X_k)_{k\ge0}$ under $P^{\mathcal{G}_n}_x$. Hence

$$P^{\mathcal{G}_n}_x[X_k \in U] = P^{\mathbb{T}_d}_{\rm o}[\pi_{n,x}(X_k) \in U] = P^{\mathbb{T}_d}_{\rm o}[X_k \in \pi^{-1}_{n,x}(U)] \quad \text{for } U \subseteq \mathcal{G}_n,\ k \ge 0. \tag{1.5}$$

A final word on the convention followed concerning constants: by $c, c', \dots$ we denote positive constants with values changing from place to place and which only depend on the dimension d and the constants α and β from the assumptions (0.1)–(0.3). Numbered constants $c_0, c_1, \dots$ are defined in the place of first occurrence and thereafter remain fixed. The dependence of constants on additional parameters appears in the notation.

1.1 Some properties of the Gaussian free field on regular trees

In this section we recall basic facts related to the Green function and the Gaussian free field on $\mathbb{T}_d$. We also restate a couple of results about $\varphi_{\mathbb{T}_d}$ that were derived by the authors in the accompanying paper [AČ19] and that will be used in several occasions throughout the rest of this article.

The Green function $g_{\mathbb{T}_d}(\cdot,\cdot)$ of simple random walk on $\mathbb{T}_d$ is (see [Woe00], Lemma 1.24, for the explicit computation)

$$g_{\mathbb{T}_d}(x,y) \coloneqq E^{\mathbb{T}_d}_x\Big[\sum_{k=0}^{\infty} \mathbf{1}_{\{X_k=y\}}\Big] = \frac{d-1}{d-2}\Big(\frac{1}{d-1}\Big)^{d_{\mathbb{T}_d}(x,y)} \quad \text{for } x,y \in \mathbb{T}_d. \tag{1.6}$$

For $U \subseteq \mathbb{T}_d$ the Green function $g^U_{\mathbb{T}_d}(\cdot,\cdot)$ of simple random walk on $\mathbb{T}_d$ killed when exiting U is $g^U_{\mathbb{T}_d}(x,y) \coloneqq E^{\mathbb{T}_d}_x\big[\sum_{0\le k<T_U} \mathbf{1}_{\{X_k=y\}}\big]$. The functions $g_{\mathbb{T}_d}(\cdot,\cdot)$ and $g^U_{\mathbb{T}_d}(\cdot,\cdot)$ are related by the identity

$$g_{\mathbb{T}_d}(x,y) = g^U_{\mathbb{T}_d}(x,y) + E^{\mathbb{T}_d}_x\big[g_{\mathbb{T}_d}(X_{T_U},y)\mathbf{1}_{\{T_U<\infty\}}\big] \quad \text{for } x,y \in \mathbb{T}_d. \tag{1.7}$$

We continue by collecting known results and properties of $\varphi_{\mathbb{T}_d}$. Recall from (0.6) that $(\varphi_{\mathbb{T}_d}(x))_{x\in\mathbb{T}_d}$ is the centred Gaussian field with covariance given by $g_{\mathbb{T}_d}(\cdot,\cdot)$. An important feature of the Gaussian free field is the domain Markov property: for $U \subseteq \mathbb{T}_d$ let $(\varphi^U_{\mathbb{T}_d}(x))_{x\in\mathbb{T}_d}$ be a new field defined by

$$\varphi^U_{\mathbb{T}_d}(x) \coloneqq \varphi_{\mathbb{T}_d}(x) - E^{\mathbb{T}_d}_x[\varphi_{\mathbb{T}_d}(X_{T_U})\mathbf{1}_{\{T_U<\infty\}}] \quad \text{for } x \in \mathbb{T}_d.$$

Then,

under $\mathbb{P}^{\mathbb{T}_d}$, $(\varphi^U_{\mathbb{T}_d}(x))_{x\in\mathbb{T}_d}$ is a centred Gaussian field on $\mathbb{T}_d$ which is independent from $(\varphi_{\mathbb{T}_d}(x))_{x\in\mathbb{T}_d\setminus U}$ and has covariance $\mathbb{E}^{\mathbb{T}_d}[\varphi^U_{\mathbb{T}_d}(x)\varphi^U_{\mathbb{T}_d}(y)] = g^U_{\mathbb{T}_d}(x,y)$ for all $x,y\in\mathbb{T}_d$. (1.8)

As a consequence of (1.8), the Gaussian free field on $\mathbb{T}_d$ can be obtained by the following recursive construction (explained in detail in [AČ19], Section 1.1). Let $(Y_x)_{x\in\mathbb{T}_d}$ be a collection of independent centred Gaussian variables defined on some auxiliary probability space $(\Omega,\mathcal{A},\mathbb{P})$ such that $Y_{\mathrm{o}}\sim\mathcal{N}(0,g_{\mathbb{T}_d}(\mathrm{o},\mathrm{o}))=\mathcal{N}(0,\frac{d-1}{d-2})$ and $Y_x\sim\mathcal{N}(0,g^{U_x}_{\mathbb{T}_d}(x,x))=\mathcal{N}(0,\frac{d}{d-1})$ for $x\neq\mathrm{o}$. Define recursively

$$\widetilde{\varphi}(\mathrm{o}) \coloneqq Y_{\mathrm{o}} \quad \text{and} \quad \widetilde{\varphi}(x) \coloneqq \frac{1}{d-1}\widetilde{\varphi}(\overline{x}) + Y_x \quad \text{for } x\in\mathbb{T}_d\setminus\{\mathrm{o}\}. \tag{1.9}$$

Then,

$$\text{under } \mathbb{P}, \text{ the law of } (\widetilde{\varphi}(x))_{x\in\mathbb{T}_d} \text{ is } \mathbb{P}^{\mathbb{T}_d}, \tag{1.10}$$

so that (1.9) can be used as an alternative description of $(\varphi_{\mathbb{T}_d}(x))_{x\in\mathbb{T}_d}$. In particular, it gives a representation of the conditional distribution of $\varphi_{\mathbb{T}_d}$ given $\varphi_{\mathbb{T}_d}(\mathrm{o})=a\in\mathbb{R}$,

$$\mathbb{P}^{\mathbb{T}_d}_a\big[(\varphi_{\mathbb{T}_d}(y))_{y\in\mathbb{T}_d}\in\cdot\,\big] \coloneqq \mathbb{P}^{\mathbb{T}_d}\big[(\varphi_{\mathbb{T}_d}(y))_{y\in\mathbb{T}_d}\in\cdot\,\big|\,\varphi_{\mathbb{T}_d}(\mathrm{o})=a\big], \tag{1.11}$$

with corresponding expectation $\mathbb{E}^{\mathbb{T}_d}_a$.

We turn to known results about level-set percolation of the Gaussian free field on $\mathbb{T}^d$ from [Szn16] and [AČ19]. First, there is a characterisation of the critical value $h_\star$ through eigenvalues $(\lambda_h)_{h\in\mathbb{R}}$ of certain self-adjoint operators $(L_h)_{h\in\mathbb{R}}$ (see [Szn16], Section 3, summarised in [AČ19], Proposition 1.1). Important for us will be that (see [Szn16], Proposition 3.3)

the map $h\mapsto\lambda_h$ is a decreasing homeomorphism from $\mathbb{R}$ to $(0,d-1)$ and $h_\star$ is the unique value in $\mathbb{R}$ such that $\lambda_{h_\star}=1$. (1.12)

To restate the other results we remind that $\mathcal{C}^{\mathbb{T}_d,h}_{\mathrm{o}}$ denotes the connected component of the level set $E^{\geq h}_{\varphi_{\mathbb{T}_d}}$ above level h containing the root $\mathrm{o}\in\mathbb{T}_d$ (see below (0.7)). The second result says that (see [AČ19], Theorem 4.1)

the 'forward percolation probability' $h\mapsto\eta^+(h)$ given by $\eta^+(h)\coloneqq\mathbb{P}^{\mathbb{T}_d}\big[|\mathcal{C}^{\mathbb{T}_d,h}_{\mathrm{o}}\cap\mathbb{T}^+_d|=\infty\big]$ is continuous and positive on $(-\infty,h_\star)$ and vanishes on $(h_\star,\infty)$. (1.13)

The third result controls the subcritical behaviour (see [AČ19], Theorem 5.1). It shows that

for $h>h_\star$ there exists $\delta_h>0$ such that $g_h(a)\coloneqq\mathbb{E}^{\mathbb{T}_d}_a\big[(1+\delta_h)^{|\mathcal{C}^{\mathbb{T}_d,h}_{\mathrm{o}}\cap\mathbb{T}^+_d|}\big]$ defines a finite function, continuous on $[h,\infty)$. Furthermore, $g_h(a)=(1+\delta_h)\mathbb{E}^Y\big[g_h(\frac{a}{d-1}+Y)\big]^{d-1}$ for all $a\geq h$, where $Y\sim\mathcal{N}(0,\frac{d}{d-1})$ and $\mathbb{E}^Y$ is taken with respect to Y. Moreover, there exist $c_h,c'_h>0$ such that $g_h(a)\leq c_{h,\gamma}\exp(c'_h a^{3/2})$ for all $a\geq h$. (1.14)

Finally, the last result about $\varphi_{\mathbb{T}_d}$ needed in the sequel in the supercritical regime is the following fact in which the λ_h, $h\in\mathbb{R}$, from (1.12) appear: by [AČ19], Theorem 4.3,

$$\text{for } h<h_\star \text{ it holds that } \lim_{k\to\infty}\mathbb{P}^{\mathbb{T}_d}\Big[\big|\mathcal{C}^{\mathbb{T}_d,h}_{\mathrm{o}}\cap S^+_{\mathbb{T}_d}(\mathrm{o},k)\big|\geq\frac{\lambda^k_h}{k^2}\Big]=\eta^+(h)>0. \tag{1.15}$$

1.2 The Green function and the zero-average Gaussian free field on $\mathcal{G}_n$

We now introduce the zero-average Green function associated to the simple random walk on $\mathcal{G}_n$ and prove an upper bound on it (Proposition 1.1). Along the way we also remind of a basic property of the zero-average Gaussian free field on $\mathcal{G}_n$ of similar type as (1.8) (see (1.19) and (1.20)).

The zero-average Green function $G_{\mathcal{G}_n}(\cdot,\cdot)$ associated with the simple random walk on $\mathcal{G}_n$ is given by

$$G_{\mathcal{G}_n}(x,y) \coloneqq \int_0^\infty \Big(P_x^{\mathcal{G}_n}[\overline{X}_t = y] - \frac{1}{N_n} \Big)\, dt \quad \text{for } x,y \in \mathcal{G}_n. \tag{1.16}$$

It is symmetric, finite and positive-semidefinite, i.e. for any $f : \mathcal{G}_n \to \mathbb{R}$ one has $\sum_{x,y\in\mathcal{G}_n} f(x) G_{\mathcal{G}_n}(x,y) f(y) \geq 0$ (see [Abä19], Remark 1.2). For $U \subseteq \mathcal{G}_n$ we define $g^U_{\mathcal{G}_n}(\cdot,\cdot)$ to be the Green function of simple random walk on $\mathcal{G}_n$ killed when exiting U, that is,

$$g^U_{\mathcal{G}_n}(x,y) \coloneqq E_x^{\mathcal{G}_n}\Big[\sum_{0\leq k<T_U} \mathbf{1}_{\{X_k=y\}} \Big] = \sum_{k=0}^\infty P_x^{\mathcal{G}_n}[X_k = y, k < T_U] \quad \text{for } x,y\in\mathcal{G}_n. \tag{1.17}$$

As $g^U_{\mathbb{T}_d}(\cdot,\cdot)$ it is symmetric, finite and vanishes for $x \notin U$ or $y \notin U$. The functions $G_{\mathcal{G}_n}(\cdot,\cdot)$ and $g^U_{\mathcal{G}_n}(\cdot,\cdot)$ are related by a similar expression as the identity (1.7) for the Green functions on $\mathbb{T}_d$. More precisely, for $U \subsetneq \mathcal{G}_n$ it holds (see [Abä19], Lemma 1.4)

$$G_{\mathcal{G}_n}(x,y) = g^U_{\mathcal{G}_n}(x,y) + E_x^{\mathcal{G}_n}\big[G_{\mathcal{G}_n}(X_{T_U},y)\big] - \frac{1}{N_n} E_x^{\mathcal{G}_n}[T_U] \quad \text{for } x,y\in\mathcal{G}_n. \tag{1.18}$$

(Lemma 1.4 in [Abä19] is stated in the case of a discrete d-dimensional torus as underlying graph. However, its proof applies as well to the graph $\mathcal{G}_n$.)

Recall from (0.4) that $(\Psi_{\mathcal{G}_n}(x))_{x\in\mathcal{G}_n}$ is the centred Gaussian field with covariance given by $G_{\mathcal{G}_n}(\cdot,\cdot)$. We point out that the Green function $G_{\mathcal{G}_n}(\cdot,\cdot)$ is called 'zero-average' since its average over $\mathcal{G}_n$ in any of the two arguments is zero. This implies that the average of $\Psi_{\mathcal{G}_n}(x)$ over $x\in\mathcal{G}_n$ vanishes $\mathbb{P}^{\mathcal{G}_n}$-almost surely and explains the name 'zero-average Gaussian free field'.

In the same way as the identity (1.7) allows for the property (1.8) of the Gaussian free field $\varphi_{\mathbb{T}_d}$ on $\mathbb{T}_d$, the identity (1.18) implies a similar (but not equal) property of the zero-average Gaussian free field $\Psi_{\mathcal{G}_n}$ on $\mathcal{G}_n$. It is given below and follows from [Abä19], Lemma 1.7. There it is stated and proved for the zero-average Gaussian free field on the discrete d-dimensional torus but the proof applies, with the obvious adjustments, also to our situation. For $U \subsetneq \mathcal{G}_n$ set

$$\varphi^U_{\mathcal{G}_n}(x) \coloneqq \Psi_{\mathcal{G}_n}(x) - E_x^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(X_{T_U})] \quad \text{for } x\in\mathcal{G}_n. \tag{1.19}$$

Then,

$$\begin{aligned}&\text{under } \mathbb{P}^{\mathcal{G}_n},\ (\varphi^U_{\mathcal{G}_n}(x))_{x\in\mathcal{G}_n} \text{ is a centred Gaussian field on } \mathcal{G}_n \text{ with covariance}\\ &\mathbb{E}^{\mathcal{G}_n}[\varphi^U_{\mathcal{G}_n}(x)\varphi^U_{\mathcal{G}_n}(y)] = g^U_{\mathcal{G}_n}(x,y) \text{ for all } x,y\in\mathcal{G}_n.\end{aligned} \tag{1.20}$$

Note that $(\varphi^U_{\mathcal{G}_n}(x))_{x\in\mathcal{G}_n}$ cannot be independent from $(\Psi_{\mathcal{G}_n}(x))_{x\in\mathcal{G}_n\setminus U}$ due to the zero-average property of $\Psi_{\mathcal{G}_n}$.

We conclude Section 1 with an upper bound on $G_{\mathcal{G}_n}(\cdot,\cdot)$ which is going to be of particular use in the proof of Proposition 2.5 needed for the supercritical phase. Note that the obtained bound (1.23) resembles the expression for the Green function $g_{\mathbb{T}_d}(\cdot,\cdot)$ on $\mathbb{T}_d$ (see (1.6)). We first define the new constant

$$c_0 := \frac{\alpha\beta}{d-1} \overset{(1.1)}{\in} (0,1). \tag{1.21}$$

Proposition 1.1. *For all $n \geq 1$ and $x, y \in \mathcal{G}_n$ it holds that*

$$G_{\mathcal{G}_n}(x,y) \leq \frac{16}{7}\frac{d-1}{d-2}\Big(\frac{1}{d-1}\Big)^{d_{\mathcal{G}_n}(x,y)} + 2\ln(N_n)N_n^{-\frac{c_0}{\beta}} + \frac{1}{\beta N_n^{c_0}}. \tag{1.22}$$

In particular, for all n large enough and $x, y \in \mathcal{G}_n$ with $d_{\mathcal{G}_n}(x,y) \leq \frac{c_0}{3}\log_{d-1}(N_n)$ it holds that

$$G_{\mathcal{G}_n}(x,y) \leq 3\frac{d-1}{d-2}\Big(\frac{1}{d-1}\Big)^{d_{\mathcal{G}_n}(x,y)}. \tag{1.23}$$

Proof. We set $t_{\mathcal{G}_n} := \frac{c_0}{\beta}\ln(N_n) = \frac{\alpha}{d-1}\ln(N_n) \overset{(1.1)}{\leq} \ln(N_n)$. By [SC97], Corollary 2.1.5, one then has (the stationary distribution of $(\overline{X}_t)_{t\geq 0}$ is the uniform distribution on $\mathcal{G}_n$ due to (0.1))

$$\int_{t_{\mathcal{G}_n}}^{\infty} \Big|P_x^{\mathcal{G}_n}[\overline{X}_t = y] - \frac{1}{N_n}\Big|\,dt \leq \int_{t_{\mathcal{G}_n}}^{\infty} e^{-\lambda_{\mathcal{G}_n} t}\,dt = \frac{e^{-\lambda_{\mathcal{G}_n} t_{\mathcal{G}_n}}}{\lambda_{\mathcal{G}_n}} \overset{(0.3)}{\leq} \frac{1}{\beta N_n^{c_0}}. \tag{1.24}$$

On the other hand, by switching to the discrete-time walk $(X_k)_{k\geq 0}$ and with $M_t \sim \text{Poi}(t)$ for $t \geq 0$ describing the number of jumps of the continuous-time simple random walk up to time t, we have

$$\begin{aligned}
&\int_0^{t_{\mathcal{G}_n}} \Big|P_x^{\mathcal{G}_n}[\overline{X}_t = y] - \frac{1}{N_n}\Big|\,dt \leq \int_0^{t_{\mathcal{G}_n}} \sum_{k=0}^{\infty} \mathbb{P}[M_t = k]P_x^{\mathcal{G}_n}[X_k = y]\,dt + \frac{t_{\mathcal{G}_n}}{N_n} \\
&\leq \sum_{k=0}^{\lfloor \alpha\log_{d-1}(N_n)\rfloor} P_x^{\mathcal{G}_n}[X_k = y]\underbrace{\int_0^{\infty}\frac{t^k}{k!}e^{-t}\,dt}_{=1} + \int_0^{t_{\mathcal{G}_n}} \mathbb{P}[M_t \geq \alpha\log_{d-1}(N_n)]\,dt + \frac{t_{\mathcal{G}_n}}{N_n}.
\end{aligned} \tag{1.25}$$

Note that for $0 \leq t \leq t_{\mathcal{G}_n}$ by Markov's inequality one has $\mathbb{P}[M_t \geq \alpha\log_{d-1}(N_n)] = \mathbb{P}[(d-1)^{M_t} \geq N_n^{\alpha}] \leq N_n^{-\alpha}\mathbb{E}[e^{\ln(d-1)M_t}] = N_n^{-\alpha}\exp(t(d-2)) \leq N_n^{-\alpha}\exp(t_{\mathcal{G}_n}(d-2)) = N_n^{-\alpha}\exp(\alpha\ln(N_n))\exp(-t_{\mathcal{G}_n}) = \exp(-t_{\mathcal{G}_n}) = N_n^{-c_0/\beta}$. Therefore (1.25) implies

$$\begin{aligned}
\int_0^{t_{\mathcal{G}_n}} \Big|P_x^{\mathcal{G}_n}[\overline{X}_t = y] - \frac{1}{N_n}\Big|\,dt &\leq \sum_{k=0}^{\lfloor \alpha\log_{d-1}(N_n)\rfloor} P_x^{\mathcal{G}_n}[X_k = y] + t_{\mathcal{G}_n}N_n^{-\frac{c_0}{\beta}} + \frac{t_{\mathcal{G}_n}}{N_n} \\
&\underset{(1.21)}{\overset{(1.5)}{\leq}} \sum_{k=0}^{\lfloor \alpha\log_{d-1}(N_n)\rfloor} P_{\mathrm{o}}^{\mathbb{T}_d}[X_k \in \pi_{n,x}^{-1}(\{y\})] + 2\ln(N_n)N_n^{-\frac{c_0}{\beta}}
\end{aligned} \tag{1.26}$$

for the cover tree $\pi_{n,x}$ of $\mathcal{G}_n$ at x. To bound the sum appearing on the right hand side of (1.26) we consider different cases for $\pi_{n,x}^{-1}(\{y\}) \cap B_{\mathbb{T}_d}(\mathrm{o}, \lfloor \alpha\log_{d-1}(N_n)\rfloor)$.

If $\big|\pi_{n,x}^{-1}(\{y\}) \cap B_{\mathbb{T}_d}(\mathrm{o}, \lfloor \alpha\log_{d-1}(N_n)\rfloor)\big| = 0$, then the sum on the last line of (1.26) vanishes and together with (1.24) this shows (1.22).

If $\left|\pi_{n,x}^{-1}(\{y\}) \cap B_{\mathbb{T}_d}(\mathrm{o}, \lfloor \alpha \log_{d-1}(N_n) \rfloor)\right| = 1$, say the intersection is $\{u\}$ (this is in particular the case if $y \in B_{\mathcal{G}_n}(x, \lfloor \alpha \log_{d-1}(N_n) \rfloor)$ and $\mathtt{tx}(B_{\mathcal{G}_n}(x, \lfloor \alpha \log_{d-1}(N_n) \rfloor)) = 0$), then the sum appearing on the right hand side of (1.26) can be rewritten as

$$\sum_{k=0}^{\lfloor \alpha \log_{d-1}(N_n) \rfloor} P_{\mathrm{o}}^{\mathbb{T}_d}[X_k \in \pi_{n,x}^{-1}(\{y\})] = \sum_{k=0}^{\lfloor \alpha \log_{d-1}(N_n) \rfloor} P_{\mathrm{o}}^{\mathbb{T}_d}[X_k = u] \overset{(1.6)}{\leq} g_{\mathbb{T}_d}(\mathrm{o}, u)$$
$$\overset{(1.6)}{=} \frac{d-1}{d-2}\Big(\frac{1}{d-1}\Big)^{d_{\mathbb{T}_d}(\mathrm{o},u)} = \frac{d-1}{d-2}\Big(\frac{1}{d-1}\Big)^{d_{\mathcal{G}_n}(x,y)}$$

and together with (1.24) this shows (1.22).

It remains to consider the last case, that is, $\left|\pi_{n,x}^{-1}(\{y\}) \cap B_{\mathbb{T}_d}(\mathrm{o}, \lfloor \alpha \log_{d-1}(N_n) \rfloor)\right| \geq 2$. Then $B_{\mathcal{G}_n}(x, \lfloor \alpha \log_{d-1}(N_n) \rfloor)$ contains a (unique by (0.2)) cycle of some length ℓ. Let us abbreviate $B \coloneqq B_{\mathbb{T}_d}(\mathrm{o}, \lfloor \alpha \log_{d-1}(N_n) \rfloor)$ and define for $m \geq 0$ the disjoint intervals $I_m \coloneqq [d_{\mathcal{G}_n}(x,y) + m\ell, d_{\mathcal{G}_n}(x,y) + (m+1)\ell)$ of length ℓ. We claim that one has the disjoint union

$$\begin{aligned} &\pi_{n,x}^{-1}(\{y\}) \cap B = \bigcup_{m=0}^{\infty} \{z \in \pi_{n,x}^{-1}(\{y\}) \cap B \,|\, d_{\mathbb{T}_d}(\mathrm{o}, z) \in I_m\} \\ &\text{with } \left|\{z \in \pi_{n,x}^{-1}(\{y\}) \cap B \,|\, d_{\mathbb{T}_d}(\mathrm{o}, z) \in I_m\}\right| \leq 2 \text{ for } m \geq 0. \end{aligned} \tag{1.27}$$

This fact is a direct consequence of Lemma 1.2 stated and proved below. We first conclude the proof of Proposition 1.1 assuming (1.27). The sum on the last line of (1.26) can be bounded, in case $\left|\pi_{n,x}^{-1}(\{y\}) \cap B_{\mathbb{T}_d}(\mathrm{o}, \lfloor \alpha \log_{d-1}(N_n) \rfloor)\right| \geq 2$, by

$$\begin{aligned} &\sum_{k=0}^{\lfloor \alpha \log_{d-1}(N_n) \rfloor} P_{\mathrm{o}}^{\mathbb{T}_d}[X_k \in \pi_{n,x}^{-1}(\{y\})] \leq \sum_{k=0}^{\infty} \sum_{z \in \pi_{n,x}^{-1}(\{y\}) \cap B} P_{\mathrm{o}}^{\mathbb{T}_d}[X_k = z] \\ &\overset{(1.6)}{=} \sum_{z \in \pi_{n,x}^{-1}(\{y\}) \cap B} \frac{d-1}{d-2}\Big(\frac{1}{d-1}\Big)^{d_{\mathbb{T}_d}(\mathrm{o},z)} \overset{(1.27)}{\leq} 2\frac{d-1}{d-2} \sum_{m=0}^{\infty} \Big(\frac{1}{d-1}\Big)^{d_{\mathcal{G}_n}(x,y)+m\ell} \\ &= 2\frac{d-1}{d-2}\Big(\frac{1}{d-1}\Big)^{d_{\mathcal{G}_n}(x,y)} \frac{1}{1 - \big(\frac{1}{d-1}\big)^{\ell}} \leq \frac{16}{7}\frac{d-1}{d-2}\Big(\frac{1}{d-1}\Big)^{d_{\mathcal{G}_n}(x,y)}, \end{aligned} \tag{1.28}$$

where in the last step we use that $d \geq 3$ and $\ell \geq 3$, too, since ℓ is the length of a cycle. The combination of (1.24), (1.26) and (1.28) concludes the proof of (1.22) also in this case, once (1.27) is asserted. To derive (1.23) from (1.22) it is enough to recall that $c_0 \leq 1$ and $\beta \leq 2$ (see (1.21) and (1.1)). Hence one has $\frac{c_0}{3} < \min\{c_0, \frac{c_0}{\beta}\}$ and therefore for n large enough also

$$2\ln(N_n) N_n^{-\frac{c_0}{\beta}} + \frac{1}{\beta N_n^{c_0}} \leq \frac{1}{N_n^{\frac{c_0}{3}}} = \Big(\frac{1}{d-1}\Big)^{\frac{c_0}{3}\log_{d-1}(N_n)} \leq \underbrace{\frac{5}{7}\frac{d-1}{d-2}}_{\geq 1} \Big(\frac{1}{d-1}\Big)^{d_{\mathcal{G}_n}(x,y)}, \tag{1.29}$$

assuming $x, y \in \mathcal{G}_n$ are such that $d_{\mathcal{G}_n}(x,y) \leq \frac{c_0}{3}\log_{d-1}(N_n)$. We can combine (1.22) with (1.29) to obtain (1.23).

To conclude the proof of Proposition 1.1 it only remains to show (1.27), which follows directly from the next lemma.

Lemma 1.2. *Let $x \in \mathcal{G}_n$, $R \geq 0$ and assume $B_{\mathcal{G}_n}(x,R)$ contains a unique cycle of length ℓ. Recall that $\pi_{n,x}$ is the fixed cover tree of $\mathcal{G}_n$ at x and assume $y \in B_{\mathcal{G}_n}(x,R)$. Then*

$$\left|\left\{z \in \pi_{n,x}^{-1}(\{y\}) \cap B_{\mathbb{T}_d}(\mathrm{o},R) \,\middle|\, d_{\mathbb{T}_d}(\mathrm{o},z) \in [0, d_{\mathcal{G}_n}(x,y))\right\}\right| = 0. \tag{1.30}$$

Moreover, for all $k \geq 0$ one has

$$\left|\left\{z \in \pi_{n,x}^{-1}(\{y\}) \cap B_{\mathbb{T}_d}(\mathrm{o},R) \,\middle|\, d_{\mathbb{T}_d}(\mathrm{o},z) \in [k, k+\ell)\right\}\right| \leq 2. \tag{1.31}$$

Proof. For any vertex $z \in \pi_{n,x}^{-1}(\{y\}) \cap B_{\mathbb{T}_d}(\mathrm{o},R)$ there is a unique non-backtracking path of length $d_{\mathbb{T}_d}(\mathrm{o},z)$ from o to z in $B_{\mathbb{T}_d}(\mathrm{o},R)$. Therefore, by the one-to-one correspondence from (1.4), every such z uniquely determines a non-backtracking path of length $d_{\mathbb{T}_d}(\mathrm{o},z)$ connecting x to y in $B_{\mathcal{G}_n}(x,R)$. Thus (1.30) is clear and for (1.31) it is enough to show that for all $k \geq 0$ one has

$$\left|\left\{\text{non-backtracking paths from } x \text{ to } y \text{ in } B_{\mathcal{G}_n}(x,R) \text{ of length in } [k,k+\ell)\right\}\right| \leq 2. \tag{1.32}$$

Let us denote by $C \coloneqq \{c_1,\dots,c_\ell\} \subseteq \mathcal{G}_n$ the unique cycle of length ℓ in $B_{\mathcal{G}_n}(x,R)$ and by $x = x_0,\dots,x_i$ for some $i \geq 0$ the unique non-backtracking path in $B_{\mathcal{G}_n}(x,R)$ from x to C such that $x_i \in C$ and $x_0,\dots,x_{i-1} \notin C$ (if $i \geq 1$). This path is unique for if $x = \widetilde{x}_0,\dots,\widetilde{x}_j$ was another such path, then one could find a cycle different from C in $\{x_0,\dots,x_i,\widetilde{x}_0,\dots,\widetilde{x}_j,c_1,\dots,c_\ell\} \subseteq B_{\mathcal{G}_n}(x,R)$. Analogously, we let $y = y_0,\dots,y_j$ for some $j \geq 0$ be the unique non-backtracking path in $B_{\mathcal{G}_n}(x,R)$ from y to C such that $y_j \in C$ and $y_0,\dots,y_{j-1} \notin C$ (if $j \geq 1$). We distinguish two cases: either $\{x_0,\dots,x_i\} \cap \{y_0,\dots,y_j\} = \emptyset$ or the intersection is not empty.

In the first case any non-backtracking path from x to y in $B_{\mathcal{G}_n}(x,R)$ starts with the segment $x_0,\dots,x_i$ from x to C and ends with the segment $y_j,\dots,y_0$ from C to y because a non-backtracking path $v_0,\dots,v_s$ from x to y in $B_{\mathcal{G}_n}(x,R)$ with $(v_0,\dots,v_i) \neq (x_0,\dots,x_i)$ or $(v_{s-j},\dots,v_s) \neq (y_j,\dots,y_0)$ would imply the existence of a cycle different from C in $\{v_0,\dots,v_s,c_1,\dots,c_\ell,x_0,\dots,x_i,y_0,\dots,y_j\} \subseteq B_{\mathcal{G}_n}(x,R)$. In between the segments $x_0,\dots,x_i$ and $y_j,\dots,y_0$ any of those non-backtracking paths can only visit vertices in C (else there would be another cycle in $B_{\mathcal{G}_n}(x,R)$) and they can only do so in clockwise or anti-clockwise direction (because they are non-backtracking). To wrap up: any non-backtracking path from x to y in $B_{\mathcal{G}_n}(x,R)$ starts with the segment $x_0,\dots,x_i$, then goes M times (for some $M \geq 0$ and some direction) around the cycle C from x_i to x_i, then continues (in the same direction) along the cycle from x_i to y_j (note that $x_i \neq y_j$ by assumption) and then ends with the segment $y_j,\dots,y_0$.

In the second case, i.e. if $\{x_0,\dots,x_i\} \cap \{y_0,\dots,y_j\} \neq \emptyset$, let $m \in \{0,\dots,i\}$ and $m' \in \{0,\dots,j\}$ be such that $x_m = y_{m'}$ and $\{x_0,\dots,x_{m-1}\} \cap \{y_0,\dots,y_{m'-1}\} = \emptyset$. In other words, $x_m = y_{m'}$ is the first common vertex of the paths $x_0,\dots,x_i$ and $y_0,\dots,y_j$. Any non-backtracking path from x to y in $B_{\mathcal{G}_n}(x,R)$ starts with the segment $x_0,\dots,x_m$ and ends with the segment $y_{m'},\dots,y_0$ because a non-backtracking path $v_0,\dots,v_s$ from x to y in $B_{\mathcal{G}_n}(x,R)$ with $(v_0,\dots,v_m) \neq (x_0,\dots,x_m)$ or $(v_{s-m'},\dots,v_s) \neq (y_{m'},\dots,y_0)$ would imply the existence of a cycle in $\{v_0,\dots,v_s,c_1,\dots,c_\ell,x_0,\dots,x_i,y_0,\dots,y_j\} \subseteq B_{\mathcal{G}_n}(x,R)$ different from C. In between the segments $x_0,\dots,x_m$ and $y_{m'},\dots,y_0$ any of those non-backtracking paths either does not do anything (possible since $x_m = y_{m'}$ by definition, i.e. the full path is $x_0,\dots,x_m,y_{m'-1},\dots,y_0$) or it has to form a non-backtracking path from x_m to itself of non-zero length. Note that in any graph a non-backtracking path (of non-zero length) from a vertex to itself necessarily contains vertices of a cycle. In our situation C is the only cycle in $B_{\mathcal{G}_n}(x,R)$ and so any non-backtracking path (of non-zero length) from x_m to x_m necessarily touches C. Therefore, it has to start with the segment $x_m,\dots,x_i$ from x_m to C and end with the segment $x_i,\dots,x_m$ from C to x_m (else there would be a cycle different from C in $B_{\mathcal{G}_n}(x,R)$). Between the segments $x_m,\dots,x_i$ and $x_i,\dots,x_m$ it can only visit vertices in C (else there would be another cycle in $B_{\mathcal{G}_n}(x,R)$) and it has to do at least one full turn around the cycle in clockwise or anti-clockwise direction (because non-backtracking). To wrap up: any non-backtracking path from x to y in $B_{\mathcal{G}_n}(x,R)$ is either of the form $x_0,\dots,x_m,y_{m'-1},\dots,y_0$ or between the initial segment $x_0,\dots,x_m$ and the final segment $y_{m'},\dots,y_0$ it continues with the segment $x_m,\dots,x_i$, then goes M times (for some $M\geq 1$ and some direction) around the cycle C from x_i to x_i and then goes back to $y_{m'}$ through $x_i,\dots,x_m$.

In any of the two cases, different non-backtracking paths from x to y in $B_{\mathcal{G}_n}(x,R)$ differ by at least ℓ in length (the length of the cycle) except if they go around the full cycle both M times but in different directions (clockwise or anti-clockwise). This shows (1.32) and concludes the proof of Lemma 1.2 and hence also of Proposition 1.1. □

2 The local picture of the zero-average Gaussian free field

In this section we investigate the local behaviour of the zero-average Gaussian free field and we derive key results and estimates that will be used in Section 3 and Section 4 for proving the main theorems of this article (Theorem 3.1 and Theorem 4.1 corresponding to (0.8) and (0.9)). The results in this section support the intuition that the *local picture* of the zero-average Gaussian free field $\Psi_{\mathcal{G}_n}$ on $\mathcal{G}_n$ is given by the Gaussian free field $\varphi_{\mathbb{T}_d}$ on $\mathbb{T}_d$. We will see two instances here: first we show in Section 2.1 that one can locally approximate $\Psi_{\mathcal{G}_n}$ around vertices of $\mathcal{G}_n$ with a tree-like neighbourhood (Theorem 2.1). This will be the type of approximation of $\Psi_{\mathcal{G}_n}$ by $\varphi_{\mathbb{T}_d}$ needed to deal with the supercritical phase in Section 4 and to prove Theorem 4.1 (corresponding to (0.9)). Then in Section 2.2 we compute conditional distributions of $\Psi_{\mathcal{G}_n}$ (Lemma 2.6) and we derive that in certain situations they resemble conditional distributions of $\varphi_{\mathbb{T}_d}$ (Proposition 2.7, see also (2.24)). This will be the crucial ingredient for approximating $\Psi_{\mathcal{G}_n}$ by $\varphi_{\mathbb{T}_d}$ along the connected components of subcritical level sets and ultimately proving Theorem 3.1 (corresponding to (0.8)) in Section 3.

2.1 A local approximation of $\Psi_{\mathcal{G}_n}$ by $\varphi_{\mathbb{T}_d}$ on tree-like neighbourhoods

The goal of this section is to prove Theorem 2.1 below, stating the approximation of the zero-average Gaussian free field $\Psi_{\mathcal{G}_n}$ on neighbourhoods of vertices with tree-like surroundings by the Gaussian free field $\varphi_{\mathbb{T}_d}$ on $\mathbb{T}_d$. This supports the intuition that the local picture of $\Psi_{\mathcal{G}_n}$ on $\mathcal{G}_n$ is given by $\varphi_{\mathbb{T}_d}$ on $\mathbb{T}_d$. The approximation derived here will be used in Section 4 to prove the main result (0.9), i.e. that a linear fraction of the vertices of $\mathcal{G}_n$ is contained in mesoscopic connected components of the level set above level h if $h<h_\star$. Theorem 2.1 will allow us to reduce the required computations on $\Psi_{\mathcal{G}_n}$ to computations on $\varphi_{\mathbb{T}_d}$.

For the remainder of Section 2.1 we introduce some notation. If $n\geq 1$, $x\in\mathcal{G}_n$ and $R\geq 1$ with $\mathtt{tx}(B_{\mathcal{G}_n}(x,R))=0$, then (see (1.3)) let $\rho_{x,R}:B_{\mathcal{G}_n}(x,R)\to B_{\mathbb{T}_d}(\mathrm{o},R)$ denote the graph isomorphism given by $(\pi_{n,x}|_{B_{\mathbb{T}_d}(\mathrm{o},R)})^{-1}$. Furthermore, for all $n\geq 1$

and pairs $x, x' \in \mathcal{G}_n$ we fix $z_{x,x'} \in \pi_{n,x}^{-1}(\{x'\}) \subseteq \mathbb{T}_d$. Finally, if $n \geq 1$, $x, x' \in \mathcal{G}_n$, $R \geq 1$ with $\mathtt{tx}(B_{\mathcal{G}_n}(x,R)) = 0$, $\mathtt{tx}(B_{\mathcal{G}_n}(x',R)) = 0$ and $B_{\mathcal{G}_n}(x,R) \cap B_{\mathcal{G}_n}(x',R) = \emptyset$, then let $\rho_{x,x',R} : B_{\mathcal{G}_n}(x,R) \cup B_{\mathcal{G}_n}(x',R) \to B_{\mathbb{T}_d}(\mathrm{o},R) \cup B_{\mathbb{T}_d}(z_{x,x'},R)$ denote the graph isomorphism given by $(\pi_{n,x}|_{B_{\mathbb{T}_d}(\mathrm{o},R)\cup B_{\mathbb{T}_d}(z_{x,x'},R)})^{-1}$. Finally, recall the constant c_0 from (1.21). The main result of this section is the following

Theorem 2.1. *For all n large enough, $x, x' \in \mathcal{G}_n$, $1 \leq r < R \leq \frac{c_0}{6}\log_{d-1}(N_n)$ such that $\mathtt{tx}(B_{\mathcal{G}_n}(x,2R)) = 0$, $\mathtt{tx}(B_{\mathcal{G}_n}(x',2R)) = 0$ and $B_{\mathcal{G}_n}(x,2R) \cap B_{\mathcal{G}_n}(x',2R) = \emptyset$, there exists a coupling $\mathbb{Q}_n$ of $\Psi_{\mathcal{G}_n}$ and $\varphi_{\mathbb{T}_d}$ such that for all $\varepsilon > 0$*

$$\begin{aligned}\mathbb{Q}_n\Big[\sup_{y\in B_{\mathcal{G}_n}(x,r)\cup B_{\mathcal{G}_n}(x',r)} \big|\Psi_{\mathcal{G}_n}(y) - \varphi_{\mathbb{T}_d}(\rho_{x,x',2R}(y))\big| > \varepsilon\Big] \\ \leq 8d(d-1)^r \exp\Big(-\frac{\varepsilon^2(d-1)(d-2)}{24d^2}(d-1)^{R-2r}\Big).\end{aligned} \tag{2.1}$$

In particular, for all n large enough, $x \in \mathcal{G}_n$, $1 \leq r < R \leq \frac{c_0}{6}\log_{d-1}(N_n)$ such that $\mathtt{tx}(B_{\mathcal{G}_n}(x,2R)) = 0$, there exists a coupling $\mathbb{Q}_n$ of $\Psi_{\mathcal{G}_n}$ and $\varphi_{\mathbb{T}_d}$ such that for all $\varepsilon > 0$ the same bound as in (2.1) applies to $\mathbb{Q}_n\big[\sup_{y\in B_{\mathcal{G}_n}(x,r)} \big|\Psi_{\mathcal{G}_n}(y) - \varphi_{\mathbb{T}_d}(\rho_{x,2R}(y))\big| > \varepsilon\big]$.

We now proceed with some preparations for the proof of Theorem 2.1. The first goal is an easy preliminary coupling of $\Psi_{\mathcal{G}_n}$ and $\varphi_{\mathbb{T}_d}$ around vertices of $\mathcal{G}_n$ with tree-like neighbourhood (Lemma 2.3). In its proof we use the following observation.

Remark 2.2. Let $x, x' \in \mathcal{G}_n$ and $R \geq 1$ satisfy $\mathtt{tx}(B_{\mathcal{G}_n}(x,R)) = 0$, $\mathtt{tx}(B_{\mathcal{G}_n}(x',R)) = 0$ and $B_{\mathcal{G}_n}(x,R) \cap B_{\mathcal{G}_n}(x',R) = \emptyset$. Assume $U \subseteq B_{\mathcal{G}_n}(x,R-1) \cup B_{\mathcal{G}_n}(x',R-1)$, so that $\partial_{\mathcal{G}_n}U \subseteq B_{\mathcal{G}_n}(x,R) \cup B_{\mathcal{G}_n}(x',R)$. Then for any $y \in B_{\mathcal{G}_n}(x,R) \cup B_{\mathcal{G}_n}(x',R) \subseteq \mathcal{G}_n$ the image under $\pi_{n,x}$ of the law of the simple random walk on $\mathbb{T}_d$ started at $\rho_{x,x',R}(y) \in B_{\mathbb{T}_d}(\mathrm{o},R) \cup B_{\mathbb{T}_d}(z_{x,x'},R) \subseteq \mathbb{T}_d$ and stopped when exiting $\rho_{x,x',R}(U)$ is the same as the law of the simple random walk on $\mathcal{G}_n$ started at y and stopped when exiting U. In particular, the hitting distribution of the boundary $\partial_{\mathcal{G}_n}U$ of the walk on $\mathcal{G}_n$ is the image under $\pi_{n,x}$ of the hitting distribution of $\partial_{\mathbb{T}_d}\rho_{x,x',R}(U)$ of the walk on $\mathbb{T}_d$, that is

$$P_y^{\mathcal{G}_n}[X_{T_U} = z] = P^{\mathbb{T}_d}_{\rho_{x,x',R}(y)}[X_{T_{\rho_{x,x',R}(U)}} = \rho_{x,x',R}(z)] \quad \text{for all } y \in U \text{ and } z \in \partial_{\mathcal{G}_n}U. \tag{2.2}$$

Similarly, for any $x \in \mathcal{G}_n$ with $\mathtt{tx}(B_{\mathcal{G}_n}(x,R)) = 0$, $U \subseteq B_{\mathcal{G}_n}(x,R-1)$, and $y \in B_{\mathcal{G}_n}(x,R) \subseteq \mathcal{G}_n$ the image under $\pi_{n,x}$ of the law of the simple random walk on $\mathbb{T}_d$ started at $\rho_{x,R}(y) \in B_{\mathbb{T}_d}(\mathrm{o},R) \subseteq \mathbb{T}_d$ and stopped when exiting $\rho_{x,R}(U)$ is the same as the law of the simple random walk on $\mathcal{G}_n$ started at y and stopped when exiting U. So (2.2) holds for $\rho_{x,x',R}$ replaced by $\rho_{x,R}$. □

As a direct implication of the above Remark 2.2 we obtain a straightforward way to couple $\Psi_{\mathcal{G}_n}$ on $B_{\mathcal{G}_n}(x,R) \cup B_{\mathcal{G}_n}(x',R)$ with $\varphi_{\mathbb{T}_d}$ on $B_{\mathbb{T}_d}(\mathrm{o},R) \cup B_{\mathbb{T}_d}(z_{x,x'},R)$.

Lemma 2.3. *Assume $x, x' \in \mathcal{G}_n$ with $\mathtt{tx}(B_{\mathcal{G}_n}(x,R)) = 0$ and $\mathtt{tx}(B_{\mathcal{G}_n}(x',R)) = 0$ satisfy $B_{\mathcal{G}_n}(x,R) \cap B_{\mathcal{G}_n}(x',R) = \emptyset$ for some $R \geq 1$. Let $U \subseteq B_{\mathcal{G}_n}(x,R-1) \cup B_{\mathcal{G}_n}(x',R-1)$. Then there exists a coupling of $\Psi_{\mathcal{G}_n}$ and $\varphi_{\mathbb{T}_d}$ such that*

$$\begin{aligned}\Psi_{\mathcal{G}_n}(y) - E_y^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(X_{T_U})] = \varphi_{\mathbb{T}_d}(\rho_{x,x',R}(y)) - E^{\mathbb{T}_d}_{\rho_{x,x',R}(y)}[\varphi_{\mathbb{T}_d}(X_{T_{\rho_{x,x',R}(U)}})] \\ \textit{for all } y \in B_{\mathcal{G}_n}(x,R) \cup B_{\mathcal{G}_n}(x',R).\end{aligned} \tag{2.3}$$

Similarly, if we only have $x \in \mathcal{G}_n$ with $\mathtt{tx}(B_{\mathcal{G}_n}(x,R)) = 0$ for some $R \geq 1$ and $U \subseteq B_{\mathcal{G}_n}(x,R-1)$, then (2.3) holds for all $y \in B_{\mathcal{G}_n}(x,R)$ with $\rho_{x,x',R}$ replaced by $\rho_{x,R}$.

Proof. The proof is analogous to the proof of Lemma 1.10 in [Abä19]. Since both sides of (2.3) describe centred Gaussian fields, it is enough to check that the covariance is the same. By (1.20) resp. by (1.8) the covariance of the field for $y, z \in B_{\mathcal{G}_n}(x, R) \cup B_{\mathcal{G}_n}(x', R)$ is $g^U_{\mathcal{G}_n}(y, z)$ on the left resp. $g^{\rho_{x,x',R}(U)}_{\mathbb{T}_d}(\rho_{x,x',R}(y), \rho_{x,x',R}(z))$ on the right hand side. These two covariances are equal by Remark 2.2 and hence the proof is complete. □

We can now lay out the strategy for proving Theorem 2.1. The idea is to combine the coupling of $\Psi_{\mathcal{G}_n}$ and $\varphi_{\mathbb{T}_d}$ from Lemma 2.3 (for some suitable choice of U) with uniform bounds on the variance of the expectations appearing in (2.3). These uniform bounds are shown in Proposition 2.5 and will ultimately lead to the proof of Theorem 2.1. Before that, we show a simple estimate of the hitting distribution of a sphere by the simple random walk on $\mathbb{T}_d$ (Lemma 2.4). This estimate is needed for the proof of the bounds in Proposition 2.5.

Lemma 2.4. *Let $R \geq 0$. Then for all $y \in B_{\mathbb{T}_d}(\text{o}, R)$ and $z \in S_{\mathbb{T}_d}(\text{o}, R)$ one has*

$$P_y^{\mathbb{T}_d}[X_{H_{S_{\mathbb{T}_d}(\text{o},R)}} = z] \leq \Big(\frac{1}{d-1}\Big)^{R - d_{\mathbb{T}_d}(y,\text{o})}. \tag{2.4}$$

Proof. Note that the statement we need to prove only depends on the distance of the vertex y to the centre of $B_{\mathbb{T}_d}(\text{o}, R)$. We denote by $\text{o} = y_0, y_1, \dots, y_R$ a fixed non-backtracking path from o to $S_{\mathbb{T}_d}(\text{o}, R) =: S$, so that $d_{\mathbb{T}_d}(y_k, \text{o}) = k$ for $k = 0, \dots, R$. First, we argue that $P^{\mathbb{T}_d}_{y_k}[X_{H_S} = z] \leq P^{\mathbb{T}_d}_{y_k}[X_{H_S} = y_R]$ for all $z \in S$ and $k = 0, \dots, R$. Indeed, fix $z \in S$ and $k \in \{0, \dots, R\}$ and let

$$i_0 \coloneqq \max\{i \in \{0, \dots, R\} \,|\, y_i \text{ is on the non-backtracking path from o to } z\},$$

so that y_{i_0} is the last common vertex of the two non-backtracking paths from o to z resp. to y_R. Note that any path from y_k to z in $\mathbb{T}_d$ has to pass through y_{i_0} and also that $P^{\mathbb{T}_d}_{y_{i_0}}[X_{H_S} = z] = P^{\mathbb{T}_d}_{y_{i_0}}[X_{H_S} = y_R]$ because $z, y_R \in S$ and $d_{\mathbb{T}_d}(y_{i_0}, z) = d_{\mathbb{T}_d}(y_{i_0}, y_R)$ by definition of y_{i_0}. As claimed, one obtains

$$\begin{aligned} P^{\mathbb{T}_d}_{y_k}[X_{H_S} = z] &= P^{\mathbb{T}_d}_{y_k}[X_{H_S} = z, H_{y_{i_0}} \leq H_S] \overset{(*)}{=} P^{\mathbb{T}_d}_{y_k}[H_{y_{i_0}} \leq H_S] P^{\mathbb{T}_d}_{y_{i_0}}[X_{H_S} = z] \\ &= P^{\mathbb{T}_d}_{y_k}[H_{y_{i_0}} \leq H_S] P^{\mathbb{T}_d}_{y_{i_0}}[X_{H_S} = y_R] \overset{(*)}{=} P^{\mathbb{T}_d}_{y_k}[X_{H_S} = y_R, H_{y_{i_0}} \leq H_S] \leq P^{\mathbb{T}_d}_{y_k}[X_{H_S} = y_R], \end{aligned}$$

where in both $(*)$ we use the strong Markov property.

It remains to show $P^{\mathbb{T}_d}_{y_k}[X_{H_S} = y_R] \leq (d-1)^{-(R-k)}$ for $k = 0, \dots, R$. To this end, let $A_k \coloneqq S_{\mathbb{T}_d}(\text{o}, R) \cap U_{y_k}$ (see (1.2)). By definition we have $y_R \in A_k$ and $|A_k| = (d-1)^{R-k}$. Moreover, by symmetry it holds $P^{\mathbb{T}_d}_{y_k}[X_{H_S} = z] = P^{\mathbb{T}_d}_{y_k}[X_{H_S} = y_R]$ for all $z \in A_k$. Hence

$$1 \geq P^{\mathbb{T}_d}_{y_k}[X_{H_S} \in A_k] = \sum_{z \in A_k} P^{\mathbb{T}_d}_{y_k}[X_{H_S} = z] = (d-1)^{R-k} P^{\mathbb{T}_d}_{y_k}[X_{H_S} = y_R],$$

from which the required claim follows directly. □

Proposition 2.5. *For all $R \geq 1$ and $y \in B_{\mathbb{T}_d}(\text{o}, R)$ one has*

$$\operatorname{Var}_{\mathbb{P}^{\mathbb{T}_d}}\Big(E^{\mathbb{T}_d}_y[\varphi_{\mathbb{T}_d}(X_{H_{S_{\mathbb{T}_d}(\text{o},R)}})]\Big) \leq \frac{d^2}{(d-1)(d-2)} \Big(\frac{1}{d-1}\Big)^{R - 2d_{\mathbb{T}_d}(y,\text{o})}. \tag{2.5}$$

Also, for all n large enough, $x \in \mathcal{G}_n$ and $1 \le R \le \frac{c_0}{6} \log_{d-1}(N_n)$ with $\mathtt{tx}(B_{\mathcal{G}_n}(x, 2R)) = 0$ and $y \in B_{\mathcal{G}_n}(x, R)$ one has

$$\operatorname{Var}_{\mathbb{P}^{\mathcal{G}_n}}\left(E_y^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(X_{H_{S_{\mathcal{G}_n}(x,R)}})]\right) \le \frac{3d^2}{(d-1)(d-2)}\Big(\frac{1}{d-1}\Big)^{R-2d_{\mathcal{G}_n}(y,x)}. \tag{2.6}$$

Proof. We start with (2.5). Let us abbreviate $S \coloneqq S_{\mathbb{T}_d}(\mathrm{o}, R)$. We first expand the variance to obtain

$$\begin{aligned}\operatorname{Var}_{\mathbb{P}^{\mathbb{T}_d}}\left(E_y^{\mathbb{T}_d}[\varphi_{\mathbb{T}_d}(X_{H_S})]\right) &= \sum_{z_1,z_2\in S} P_y^{\mathbb{T}_d}[X_{H_S} = z_1]P_y^{\mathbb{T}_d}[X_{H_S} = z_2]g_{\mathbb{T}_d}(z_1, z_2) \\ &\overset{(2.4)}{\underset{(1.6)}{\le}} \frac{d-1}{d-2}\Big(\frac{1}{d-1}\Big)^{2(R-d_{\mathbb{T}_d}(y,\mathrm{o}))} \sum_{z_1,z_2\in S}\Big(\frac{1}{d-1}\Big)^{d_{\mathbb{T}_d}(z_1,z_2)}.\end{aligned} \tag{2.7}$$

Fix $z_1 \in S$. Note that all vertices of S are at even distance from z_1 and more precisely that in S

$$\begin{cases}\text{there is one vertex at distance 0 from } z_1 \text{ (namely } z_1 \text{ itself)},\\ \text{there are } (d-2)(d-1)^{j-1} \text{ vertices at distance } 2j \text{ from } z_1 \text{ for } 1 \le j \le R-1,\\ \text{there are } (d-1)^R \text{ vertices at distance } 2R \text{ from } z_1.\end{cases} \tag{2.8}$$

This implies that for fixed $z_1 \in S$ it holds

$$\begin{aligned}\sum_{z_2\in S}\Big(\frac{1}{d-1}\Big)^{d_{\mathbb{T}_d}(z_1,z_2)} &= 1 + \sum_{j=1}^{R-1}(d-2)(d-1)^{j-1}\Big(\frac{1}{d-1}\Big)^{2j} + (d-1)^R\Big(\frac{1}{d-1}\Big)^{2R} \\ &= 1 + (d-2)\sum_{j=1}^{R-1}\Big(\frac{1}{d-1}\Big)^{j+1} + \Big(\frac{1}{d-1}\Big)^{R} \\ &= 1 + \Big(\frac{1}{d-1}\Big)\Big(1 - \Big(\frac{1}{d-1}\Big)^{R-1}\Big) + \Big(\frac{1}{d-1}\Big)^{R} = \frac{d}{d-1}.\end{aligned} \tag{2.9}$$

Since $|S| = d(d-1)^{R-1}$, we can combine (2.7) and (2.9) to obtain

$$\operatorname{Var}_{\mathbb{P}^{\mathbb{T}_d}}\left(E_y^{\mathbb{T}_d}[\varphi_{\mathbb{T}_d}(X_{H_S})]\right) \le \frac{d-1}{d-2}\Big(\frac{1}{d-1}\Big)^{2(R-d_{\mathbb{T}_d}(y,\mathrm{o}))} d(d-1)^{R-1}\frac{d}{d-1},$$

which is equal to the right hand side of (2.5) and concludes the proof of the first part.

For the proof of (2.6) we proceed similarly. Let us abbreviate $S' \coloneqq S_{\mathcal{G}_n}(x, R)$ and note that $P_y^{\mathcal{G}_n}$-almost surely $H_{S'} = T_{B_{\mathcal{G}_n}(x,R-1)}$. Since by assumption we have $\mathtt{tx}(B_{\mathcal{G}_n}(x, 2R)) = 0$, Remark 2.2 implies that for every $z \in S'$ we have

$$\begin{aligned}P_y^{\mathcal{G}_n}[X_{H_{S'}} = z] &= P^{\mathbb{T}_d}_{\rho_{x,R}(y)}[X_{H_{\rho_{x,R}(S')}} = \rho_{x,R}(z)] \\ &\overset{(2.4)}{\le} \Big(\frac{1}{d-1}\Big)^{R-d_{\mathbb{T}_d}(\rho_{x,R}(y),\mathrm{o})} = \Big(\frac{1}{d-1}\Big)^{R-d_{\mathcal{G}_n}(y,x)}.\end{aligned}$$

Furthermore, for n large enough, the inequality (1.23) in Proposition 1.1 applies to $G_{\mathcal{G}_n}(z_1, z_2)$ with $z_1, z_2 \in S'$ since $d_{\mathcal{G}_n}(z_1, z_2) \le 2R \le \frac{c_0}{3}\log_{d-1}(N_n)$ by assumption on R. Therefore, by expanding the variance we obtain similarly to (2.7) the inequality

$$\operatorname{Var}_{\mathbb{P}^{\mathcal{G}_n}}\left(E_y^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(X_{H_{S'}})]\right) \le 3\frac{d-1}{d-2}\Big(\frac{1}{d-1}\Big)^{2(R-d_{\mathcal{G}_n}(y,x))}\sum_{z_1,z_2\in S'}\Big(\frac{1}{d-1}\Big)^{d_{\mathcal{G}_n}(z_1,z_2)}, \tag{2.10}$$

assuming n is large enough. We now argue that for fixed $z_1 \in S'$ the vertices in S' can be again characterised by (2.8). Indeed, the assumption $\mathtt{tx}(B_{\mathcal{G}_n}(x,2R)) = 0$ implies that any shortest path from z_1 to some $z_2 \in S'$ necessarily remains in $B_{\mathcal{G}_n}(x,R)$ for which $\mathtt{tx}(B_{\mathcal{G}_n}(x,R)) = 0$ holds. Therefore, $d_{\mathcal{G}_n}(z_1,z_2)$ can be computed by only considering the shortest connection in $B_{\mathcal{G}_n}(x,R)$ between z_1 and z_2 and so we are in the tree-like situation of (2.8). Thus, the same computation as in (2.9) leads to $\sum_{z_2\in S'}\left(\frac{1}{d-1}\right)^{d_{\mathcal{G}_n}(z_1,z_2)} = \frac{d}{d-1}$. This combined with (2.10) concludes the proof of (2.6) since $|S'| = d(d-1)^{R-1}$ as $\mathtt{tx}(B_{\mathcal{G}_n}(x,R)) = 0$. □

We now have all the ingredients for the proof of Theorem 2.1, by which we conclude Section 2.1.

Proof of Theorem 2.1. Let us abbreviate $V \coloneqq B_{\mathcal{G}_n}(x,r)\cup B_{\mathcal{G}_n}(x',r)$. Under the assumptions of the theorem we can apply Lemma 2.3 with $U \coloneqq B_{\mathcal{G}_n}(x,R-1)\cup B_{\mathcal{G}_n}(x',R-1) \supseteq V$. Thus we obtain a coupling $\mathbb{Q}_n$ of $\Psi_{\mathcal{G}_n}$ and $\varphi_{\mathbb{T}_d}$ such that for all $\varepsilon > 0$

$$\begin{aligned}
&\mathbb{Q}_n\Big[\sup_{y\in V}\big|\Psi_{\mathcal{G}_n}(y) - \varphi_{\mathbb{T}_d}(\rho_{x,x',R}(y))\big| > \varepsilon\Big] \qquad (2.11)\\
&\le \mathbb{Q}_n\Big[\sup_{y\in V}\Big|E^{\mathbb{T}_d}_{\rho_{x,x',R}(y)}[\varphi_{\mathbb{T}_d}(X_{T_{\rho_{x,x',R}(U)}})]\Big| > \frac{\varepsilon}{2}\Big] + \mathbb{Q}_n\Big[\sup_{y\in V}\Big|E^{\mathcal{G}_n}_y[\Psi_{\mathcal{G}_n}(X_{T_U})]\Big| > \frac{\varepsilon}{2}\Big],
\end{aligned}$$

where $\rho_{x,x',R}(U) = B_{\mathbb{T}_d}(\mathrm{o},R-1)\cup B_{\mathbb{T}_d}(z_{x,x'},R-1) \subseteq B_{\mathbb{T}_d}(\mathrm{o},2R)\cup B_{\mathbb{T}_d}(z_{x,x'},2R)$. We now consider the two terms on the right hand side of (2.11) separately. For the first term a union bound leads to, abbreviating $S \coloneqq S_{\mathbb{T}_d}(\mathrm{o},R)$ and $S' \coloneqq S_{\mathbb{T}_d}(z_{x,x'},R)$,

$$\begin{aligned}
&\mathbb{Q}_n\Big[\sup_{y\in V}\Big|E^{\mathbb{T}_d}_{\rho_{x,x',R}(y)}[\varphi_{\mathbb{T}_d}(X_{T_{\rho_{x,x',R}(U)}})]\Big| > \frac{\varepsilon}{2}\Big]\\
&= \mathbb{P}^{\mathbb{T}_d}\Big[\sup_{y\in B_{\mathbb{T}_d}(\mathrm{o},r)\cup B_{\mathbb{T}_d}(z_{x,x'},r)}\Big|E^{\mathbb{T}_d}_y[\varphi_{\mathbb{T}_d}(X_{H_{S\cup S'}})]\Big| > \frac{\varepsilon}{2}\Big]\\
&\le \sum_{y\in B_{\mathbb{T}_d}(\mathrm{o},r)}\mathbb{P}^{\mathbb{T}_d}\Big[\Big|E^{\mathbb{T}_d}_y[\varphi_{\mathbb{T}_d}(X_{H_S})]\Big| > \frac{\varepsilon}{2}\Big] + \sum_{y\in B_{\mathbb{T}_d}(z_{x,x'},r)}\mathbb{P}^{\mathbb{T}_d}\Big[\Big|E^{\mathbb{T}_d}_y[\varphi_{\mathbb{T}_d}(X_{H_{S'}})]\Big| > \frac{\varepsilon}{2}\Big] \qquad (2.12)\\
&= 2\sum_{y\in B_{\mathbb{T}_d}(\mathrm{o},r)}\mathbb{P}^{\mathbb{T}_d}\Big[\Big|E^{\mathbb{T}_d}_y[\varphi_{\mathbb{T}_d}(X_{H_S})]\Big| > \frac{\varepsilon}{2}\Big],
\end{aligned}$$

where the last equality follows by symmetry. Now for each $y \in B_{\mathbb{T}_d}(\mathrm{o},r)$ the expectation appearing inside the probability on the right hand side of (2.12) is a centred Gaussian variable with respect to $\mathbb{P}^{\mathbb{T}_d}$. Thus the exponential Markov inequality implies that

$$\begin{aligned}
2\sum_{y\in B_{\mathbb{T}_d}(\mathrm{o},r)}\mathbb{P}^{\mathbb{T}_d}\Big[\Big|E^{\mathbb{T}_d}_y[\varphi_{\mathbb{T}_d}(X_{H_S})]\Big| > \frac{\varepsilon}{2}\Big] &\le 4\sum_{y\in B_{\mathbb{T}_d}(\mathrm{o},r)}\exp\bigg(-\frac{(\varepsilon/2)^2}{2\mathrm{Var}_{\mathbb{P}^{\mathbb{T}_d}}\big(E^{\mathbb{T}_d}_y[\varphi_{\mathbb{T}_d}(X_{H_S})]\big)}\bigg)\\
&\overset{(2.5)}{\le} 4\big|B_{\mathbb{T}_d}(\mathrm{o},r)\big|\exp\bigg(-\frac{\varepsilon^2(d-1)(d-2)}{8d^2}(d-1)^{R-2r}\bigg). \qquad (2.13)
\end{aligned}$$

For the second term on the right hand side of (2.11) we similarly have by a union bound that, abbreviating $\overline{S} \coloneqq S_{\mathcal{G}_n}(x,R)$ and $\overline{S}' \coloneqq S_{\mathcal{G}_n}(x',R)$,

$$\begin{aligned}\mathbb{Q}_n\Big[\sup_{y\in V}\Big|E_y^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(X_{T_U})]\Big| > \frac{\varepsilon}{2}\Big] &= \mathbb{P}^{\mathcal{G}_n}\Big[\sup_{y\in V}\Big|E_y^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(X_{H_{\overline{S}\cup\overline{S}'}})]\Big| > \frac{\varepsilon}{2}\Big] \\ &\le \sum_{y\in B_{\mathcal{G}_n}(x,r)}\mathbb{P}^{\mathcal{G}_n}\Big[\Big|E_y^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(X_{H_{\overline{S}}})]\Big| > \frac{\varepsilon}{2}\Big] + \sum_{y\in B_{\mathcal{G}_n}(x',r)}\mathbb{P}^{\mathcal{G}_n}\Big[\Big|E_y^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(X_{H_{\overline{S}'}})]\Big| > \frac{\varepsilon}{2}\Big].\end{aligned} \tag{2.14}$$

The expectations appearing inside the probabilities on the right hand side of (2.14) are centred Gaussian variables with respect to $\mathbb{P}^{\mathcal{G}_n}$. By (2.6) their variance can be bounded by $\frac{3d^2}{(d-1)(d-2)}(\frac{1}{d-1})^{R-2r}$. Hence the exponential Markov inequality implies that

$$\begin{aligned}&\sum_{y\in B_{\mathcal{G}_n}(x,r)}\mathbb{P}^{\mathcal{G}_n}\Big[\Big|E_y^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(X_{H_{\overline{S}}})]\Big| > \frac{\varepsilon}{2}\Big] + \sum_{y\in B_{\mathcal{G}_n}(x',r)}\mathbb{P}^{\mathcal{G}_n}\Big[\Big|E_y^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(X_{H_{\overline{S}'}})]\Big| > \frac{\varepsilon}{2}\Big] \\ &\quad\le 2\Big(\big|B_{\mathcal{G}_n}(x,r)\big| + \big|B_{\mathcal{G}_n}(x',r)\big|\Big)\exp\Big(-\frac{\varepsilon^2(d-1)(d-2)}{24d^2}(d-1)^{R-2r}\Big).\end{aligned} \tag{2.15}$$

The combination of (2.11)–(2.15) concludes the proof of Theorem 2.1 since $|B_{\mathcal{G}_n}(x,r)| = |B_{\mathcal{G}_n}(x',r)| = |B_{\mathbb{T}_d}(\mathrm{o},r)| = \frac{d(d-1)^r-2}{d-2} \le d(d-1)^r$ as $\mathtt{tx}(B_{\mathcal{G}_n}(x,r)) = \mathtt{tx}(B_{\mathcal{G}_n}(x',r)) = 0$ by assumption. □

2.2 Conditional distribution of the zero-average Gaussian free field

In this section we investigate the conditional distributions of the zero-average Gaussian free field. Their detailed understanding will be needed in Section 3 to control the behaviour of the exploration process used in the proof of the main subcritical result (0.8). We start with the exact computation of the conditional distribution of $\Psi_{\mathcal{G}_n}(x)$ for $x \in \mathcal{G}_n$ given $\Psi_{\mathcal{G}_n}$ on some $A \subsetneq \mathcal{G}_n$ (Lemma 2.6). We then see that, under certain geometric conditions on x and A (see (2.26)–(2.28)), the conditional distribution of $\Psi_{\mathcal{G}_n}(x)$ given $\Psi_{\mathcal{G}_n}$ on $A \subsetneq \mathcal{G}_n$ shows strong similarities with the conditional distribution of the Gaussian free field $\varphi_{\mathbb{T}_d}$ on $\mathbb{T}_d$ (Proposition 2.7, see also (2.24)). This feature reflects the general philosophy that the local picture of $\Psi_{\mathcal{G}_n}$ on $\mathcal{G}_n$ is given by $\varphi_{\mathbb{T}_d}$ on $\mathbb{T}_d$.

Lemma 2.6. *Let $A \subsetneq \mathcal{G}_n$ non-empty and $x \in \mathcal{G}_n$. Then $\mathbb{P}^{\mathcal{G}_n}$-almost surely*

$$\mathbb{E}^{\mathcal{G}_n}\big[\Psi_{\mathcal{G}_n}(x)\big|\sigma(\Psi_{\mathcal{G}_n}(y), y\in A)\big] = E_x^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(X_{H_A})] - \frac{E_x^{\mathcal{G}_n}[H_A]}{E_\pi^{\mathcal{G}_n}[H_A]}E_\pi^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(X_{H_A})] \tag{2.16}$$

and

$$\begin{aligned}&\mathrm{Var}_{\mathbb{P}^{\mathcal{G}_n}}\big(\Psi_{\mathcal{G}_n}(x)\big|\sigma(\Psi_{\mathcal{G}_n}(y), y\in A)\big) \\ &\qquad = G_{\mathcal{G}_n}(x,x) - E_x^{\mathcal{G}_n}[G_{\mathcal{G}_n}(X_{H_A},x)] + \frac{E_x^{\mathcal{G}_n}[H_A]}{E_\pi^{\mathcal{G}_n}[H_A]}E_\pi^{\mathcal{G}_n}[G_{\mathcal{G}_n}(X_{H_A},x)].\end{aligned} \tag{2.17}$$

Here $E_\pi^{\mathcal{G}_n}$ is the expectation with respect to $\frac{1}{N_n}\sum_{z\in\mathcal{G}_n}P_z^{\mathcal{G}_n}$, i.e. the canonical law of simple random walk on $\mathcal{G}_n$ starting at a uniformly chosen vertex.

Proof. We will abbreviate $U \coloneqq \mathcal{G}_n \setminus A \subsetneq \mathcal{G}_n$. In particular $T_U = H_A$. Note that by (1.19) one can write $\Psi_{\mathcal{G}_n}(x) = \varphi_{\mathcal{G}_n}^U(x) + E_x^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(X_{H_A})]$, the second term actually being $\sigma(\Psi_{\mathcal{G}_n}(y), y\in A)$-measurable. Hence $\mathbb{E}^{\mathcal{G}_n}\big[\Psi_{\mathcal{G}_n}(x)\big|\sigma(\Psi_{\mathcal{G}_n}(y), y\in A)\big] = E_x^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(X_{H_A})]+$

$\mathbb{E}^{\mathcal{G}_n}\left[\varphi^U_{\mathcal{G}_n}(x)\middle|\sigma(\Psi_{\mathcal{G}_n}(y), y\in A)\right]$ and moreover also $\mathrm{Var}_{\mathbb{P}^{\mathcal{G}_n}}\left(\Psi_{\mathcal{G}_n}(x)\middle|\sigma(\Psi_{\mathcal{G}_n}(y), y\in A)\right) = \mathrm{Var}_{\mathbb{P}^{\mathcal{G}_n}}\left(\varphi^U_{\mathcal{G}_n}(x)\middle|\sigma(\Psi_{\mathcal{G}_n}(y), y\in A)\right)$. For (2.16) it is therefore enough to show that $\mathbb{P}^{\mathcal{G}_n}$-almost surely

$$\mathbb{E}^{\mathcal{G}_n}\left[\varphi^U_{\mathcal{G}_n}(x)\,\middle|\,\sigma(\Psi_{\mathcal{G}_n}(y), y\in A)\right] = -\frac{E^{\mathcal{G}_n}_x[T_U]}{E^{\mathcal{G}_n}_\pi[T_U]}E^{\mathcal{G}_n}_\pi[\Psi_{\mathcal{G}_n}(X_{T_U})]. \tag{2.18}$$

On the other hand, for (2.17) it is enough to show (use (1.18) to manipulate the first two terms on the right hand side of (2.17))

$$\mathrm{Var}_{\mathbb{P}^{\mathcal{G}_n}}\left(\varphi^U_{\mathcal{G}_n}(x)\middle|\sigma(\Psi_{\mathcal{G}_n}(y), y\in A)\right) = g^U_{\mathcal{G}_n}(x,x) + \frac{E^{\mathcal{G}_n}_x[T_U]}{E^{\mathcal{G}_n}_\pi[T_U]}\left(E^{\mathcal{G}_n}_\pi[G_{\mathcal{G}_n}(X_{T_U},x)] - \frac{E_\pi[T_U]}{N_n}\right). \tag{2.19}$$

Let us fix $x_0\in A$. We claim that

$$\sigma(\Psi_{\mathcal{G}_n}(y), y\in A) = \sigma\Big(\textstyle\sum_{z\in U}\varphi^U_{\mathcal{G}_n}(z), \Psi_{\mathcal{G}_n}(y) - \Psi_{\mathcal{G}_n}(x_0), y\in A\Big). \tag{2.20}$$

To see (2.20) first note that $\sigma(\Psi_{\mathcal{G}_n}(y), y\in A) = \sigma(\Psi_{\mathcal{G}_n}(x_0), \Psi_{\mathcal{G}_n}(y) - \Psi_{\mathcal{G}_n}(x_0), y\in A)$. Moreover, by the zero-average property of $\Psi_{\mathcal{G}_n}$ (see below (1.18)), one $\mathbb{P}^{\mathcal{G}_n}$-almost surely has

$$\begin{aligned}
\Psi_{\mathcal{G}_n}(x_0) &= -\frac{1}{N_n}\sum_{z\in\mathcal{G}_n}(\Psi_{\mathcal{G}_n}(z) - \Psi_{\mathcal{G}_n}(x_0)) \\
&\overset{(1.19)}{=} -\frac{1}{N_n}\sum_{z\in\mathcal{G}_n}(\varphi^U_{\mathcal{G}_n}(z) + E^{\mathcal{G}_n}_z[\Psi_{\mathcal{G}_n}(X_{T_U})] - \Psi_{\mathcal{G}_n}(x_0)) \\
&= -\frac{1}{N_n}\sum_{z\in U}\varphi^U_{\mathcal{G}_n}(z) - \frac{1}{N_n}\sum_{z\in\mathcal{G}_n}E^{\mathcal{G}_n}_z[\Psi_{\mathcal{G}_n}(X_{T_U}) - \Psi_{\mathcal{G}_n}(x_0)].
\end{aligned}$$

The latter sum is $\sigma(\Psi_{\mathcal{G}_n}(y) - \Psi_{\mathcal{G}_n}(x_0), y\in A)$-measurable. Thus $\sigma(\Psi_{\mathcal{G}_n}(x_0), \Psi_{\mathcal{G}_n}(y) - \Psi_{\mathcal{G}_n}(x_0), y\in A) = \sigma\big(\sum_{z\in U}\varphi^U_{\mathcal{G}_n}(z), \Psi_{\mathcal{G}_n}(y) - \Psi_{\mathcal{G}_n}(x_0), y\in A\big)$, which shows (2.20).

Now note that

$$\text{for } z\in\mathcal{G}_n \text{ and } y\in A \text{ the Gaussian random variables } \varphi^U_{\mathcal{G}_n}(z) \text{ and } \Psi_{\mathcal{G}_n}(y) - \Psi_{\mathcal{G}_n}(x_0) \text{ are independent.} \tag{2.21}$$

Indeed, $\mathbb{E}^{\mathcal{G}_n}\left[\varphi^U_{\mathcal{G}_n}(z)(\Psi_{\mathcal{G}_n}(y) - \Psi_{\mathcal{G}_n}(x_0))\right] = G_{\mathcal{G}_n}(z,y) - E^{\mathcal{G}_n}_z[G_{\mathcal{G}_n}(X_{T_U},y)] - G_{\mathcal{G}_n}(z,x_0) + E^{\mathcal{G}_n}_z[G_{\mathcal{G}_n}(X_{T_U},x_0)]$ by (1.19) and (0.4), which is equal to $g^U_{\mathcal{G}_n}(z,y) - g^U_{\mathcal{G}_n}(z,x_0) = 0$ by (1.18) and (1.17) (since $y, x_0\notin U$).

Recall that for random variables U, Y, Z such that U is integrable and Z is independent of $\sigma(U,Y)$ one has $\mathbb{E}[U|\sigma(Y,Z)] = \mathbb{E}[U|\sigma(Y)]$ almost surely (see e.g. [Wil91], 9.7(k)). Hence we get $\mathbb{E}^{\mathcal{G}_n}\left[\varphi^U_{\mathcal{G}_n}(x)\middle|\sigma(\Psi_{\mathcal{G}_n}(y), y\in A)\right] = \mathbb{E}^{\mathcal{G}_n}\left[\varphi^U_{\mathcal{G}_n}(x)\middle|\sigma(\sum_{z\in U}\varphi^U_{\mathcal{G}_n}(z))\right]$ $\mathbb{P}^{\mathcal{G}_n}$-almost surely by (2.20) and (2.21). Due to the general formula $\mathrm{Var}(X|\sigma(Y)) = \mathbb{E}[X^2|\sigma(Y)] - \mathbb{E}[X|\sigma(Y)]^2$, the same observation also shows that $\mathrm{Var}_{\mathbb{P}^{\mathcal{G}_n}}\left(\varphi^U_{\mathcal{G}_n}(x)\middle|\sigma(\Psi_{\mathcal{G}_n}(y), y\in A)\right) = \mathrm{Var}_{\mathbb{P}^{\mathcal{G}_n}}\left(\varphi^U_{\mathcal{G}_n}(x)\middle|\sigma(\sum_{z\in U}\varphi^U_{\mathcal{G}_n}(z))\right)$. Therefore, the conditional expectation/variance to be considered in (2.18) and (2.19) are actually only with respect to the sigma-algebra generated by the single Gaussian random variable $\sum_{z\in U}\varphi^U_{\mathcal{G}_n}(z)$. So by the formula for conditional expectation/variance of the bivariate centred Gaussian distribution we have

$$\begin{aligned}\mathbb{E}^{\mathcal{G}_n}\big[\varphi^U_{\mathcal{G}_n}(x)\,\big|\,\sigma(\Psi_{\mathcal{G}_n}(y),y\in A)\big] &= \frac{\mathbb{E}^{\mathcal{G}_n}\big[\varphi^U_{\mathcal{G}_n}(x)\sum_{z\in U}\varphi^U_{\mathcal{G}_n}(z)\big]}{\mathbb{E}^{\mathcal{G}_n}\big[\big(\sum_{z\in U}\varphi^U_{\mathcal{G}_n}(z)\big)^2\big]}\sum_{z\in U}\varphi^U_{\mathcal{G}_n}(z),\\ \mathrm{Var}_{\mathbb{P}^{\mathcal{G}_n}}\big(\varphi^U_{\mathcal{G}_n}(x)\big|\sigma(\Psi_{\mathcal{G}_n}(y),y\in A)\big) &= \mathbb{E}^{\mathcal{G}_n}\big[\varphi^U_{\mathcal{G}_n}(x)^2\big]-\frac{\mathbb{E}^{\mathcal{G}_n}\big[\varphi^U_{\mathcal{G}_n}(x)\sum_{z\in U}\varphi^U_{\mathcal{G}_n}(z)\big]^2}{\mathbb{E}^{\mathcal{G}_n}\big[\big(\sum_{z\in U}\varphi^U_{\mathcal{G}_n}(z)\big)^2\big]}.\end{aligned} \tag{2.22}$$

We observe that for $u\in\mathcal{G}_n$ one has $\sum_{z\in U}\mathbb{E}^{\mathcal{G}_n}\big[\varphi^U_{\mathcal{G}_n}(u)\varphi^U_{\mathcal{G}_n}(z)\big]=\sum_{z\in U}g^U_{\mathcal{G}_n}(u,z)=E^{\mathcal{G}_n}_u[T_U]$ by (1.20) and (1.17). By applying this and (1.8) inside (2.22) we obtain

$$\begin{aligned}\mathbb{E}^{\mathcal{G}_n}\big[\varphi^U_{\mathcal{G}_n}(x)\,\big|\,\sigma(\Psi_{\mathcal{G}_n}(y),y\in A)\big] &= \frac{E^{\mathcal{G}_n}_x[T_U]}{\sum_{z\in U}E^{\mathcal{G}_n}_z[T_U]}\sum_{z\in U}\varphi^U_{\mathcal{G}_n}(z),\\ \mathrm{Var}_{\mathbb{P}^{\mathcal{G}_n}}\big(\varphi^U_{\mathcal{G}_n}(x)\big|\sigma(\Psi_{\mathcal{G}_n}(y),y\in A)\big) &= g^U_{\mathcal{G}_n}(x,x)-\frac{E^{\mathcal{G}_n}_x[T_U]^2}{\sum_{z\in U}E^{\mathcal{G}_n}_z[T_U]}.\end{aligned} \tag{2.23}$$

We are almost done. Observe that by (1.20), (1.19) and the zero-average property of $\Psi_{\mathcal{G}_n}$ it $\mathbb{P}^{\mathcal{G}_n}$-almost surely holds

$$\sum_{z\in U}\varphi^U_{\mathcal{G}_n}(z)=\sum_{z\in\mathcal{G}_n}\varphi^U_{\mathcal{G}_n}(z)=\sum_{z\in\mathcal{G}_n}\big(\Psi_{\mathcal{G}_n}(z)-E^{\mathcal{G}_n}_z[\Psi_{\mathcal{G}_n}(X_{T_U})]\big)=-\sum_{z\in\mathcal{G}_n}E^{\mathcal{G}_n}_z[\Psi_{\mathcal{G}_n}(X_{T_U})].$$

This combined with (2.23) shows (2.18). On the other hand, by the formula above (2.23), (1.18) and the zero-average property of $G_{\mathcal{G}_n}(\cdot,\cdot)$ (see below (1.18)) one has

$$\begin{aligned}-\frac{E^{\mathcal{G}_n}_x[T_U]}{N_n} &= -\frac{1}{N_n}\sum_{z\in U}g^U_{\mathcal{G}_n}(z,x)=-\frac{1}{N_n}\sum_{z\in\mathcal{G}_n}g^U_{\mathcal{G}_n}(z,x)\\ &= \frac{1}{N_n}\sum_{z\in\mathcal{G}_n}\Big(E^{\mathcal{G}_n}_z[G_{\mathcal{G}_n}(X_{T_U}),x]-\frac{1}{N_n}E^{\mathcal{G}_n}_z[T_U]-G_{\mathcal{G}_n}(z,x)\Big)\\ &= E^{\mathcal{G}_n}_\pi[G_{\mathcal{G}_n}(X_{T_U},x)]-\frac{E_\pi[T_U]}{N_n}.\end{aligned}$$

This combined with (2.23) shows (2.19) and concludes the proof of Lemma 2.6. □

Lemma 2.6 above shows that for any $x\in\mathcal{G}_n$, $\Psi_{\mathcal{G}_n}(x)$ conditionally on $(\Psi_{\mathcal{G}_n}(y))_{y\in A}$ for $A\subsetneq\mathcal{G}_n$ non-empty is a Gaussian random variable with mean and variance given by the right hand sides of (2.16) and (2.17). Comparable (but easier) statements for the Gaussian free field $\varphi_{\mathbb{T}_d}$ on $\mathbb{T}_d$ follow directly from (1.8). In particular, if $x'\in\mathbb{T}_d$ and $A':=\mathbb{T}_d\setminus U_{x'}$ (recall definition (1.2)), then by (1.9) and (1.10) one has

$$\begin{aligned}\mathbb{E}^{\mathbb{T}_d}\big[\varphi_{\mathbb{T}_d}(x')\big|\sigma(\varphi_{\mathbb{T}_d}(y),y\in A')\big] &= \tfrac{1}{d-1}\varphi_{\mathbb{T}_d}(\overline{x}'),\\ \mathrm{Var}_{\mathbb{P}^{\mathbb{T}_d}}\big(\varphi_{\mathbb{T}_d}(x')\big|\sigma(\varphi_{\mathbb{T}_d}(y),y\in A')\big) &= \tfrac{d}{d-1}.\end{aligned} \tag{2.24}$$

As we will show in Proposition 2.7 below, a similar behaviour can be observed for the zero-average Gaussian free field $\Psi_{\mathcal{G}_n}$ on $\mathcal{G}_n$, at least in specific situations. We now introduce the requirements on $x\in\mathcal{G}_n$ and $A\subseteq\mathcal{G}_n$. Define for $A\subsetneq\mathcal{G}_n$ non-empty and $r\geq 1$ the set $B_{\mathcal{G}_n}(A,r):=\{z\in\mathcal{G}_n\,|\,z\in B_{\mathcal{G}_n}(w,r)\text{ for some }w\in A\}$. Moreover, for $x\in\partial_{\mathcal{G}_n}A$ we set

$$F_A(x,r):=\{z\in B_{\mathcal{G}_n}(A,r)\setminus A\,|\,z\text{ is connected to }x\text{ in }B_{\mathcal{G}_n}(A,r)\setminus A\}.$$

In particular $x \in F_A(x, r)$. We set

$$s_n \coloneqq \max\{1\,,\ \lfloor 8 \log_{d-1}(\log_{d-1}(N_n)) \rfloor\} \quad \text{for } n \geq 1. \tag{2.25}$$

and say that $x \in \partial_{\mathcal{G}_n} A$ is a *good vertex at the boundary of* A if the following properties hold

- $|B_{\mathcal{G}_n}(x,1) \cap A| = 1$, write $\overline{x} \in A$ for the unique vertex in this intersection (note that for $x' \in \mathbb{T}_d$ the notation $\overline{x}'$ has been defined above (1.2)) (2.26)
- $\mathtt{tx}(F_A(x, s_n)) = 0$ (2.27)
- for all $y \in \partial_{\mathcal{G}_n} A \setminus \{x\}$ every path in $\mathcal{G}_n \setminus A$ from y to x leaves $B_{\mathcal{G}_n}(A, s_n)$. (2.28)

Equivalently, $F_A(x, s_n)$ is *proper* in the notation of [ČTW11] (see Figure 1 for an illustration of the conditions (2.26)–(2.28)).

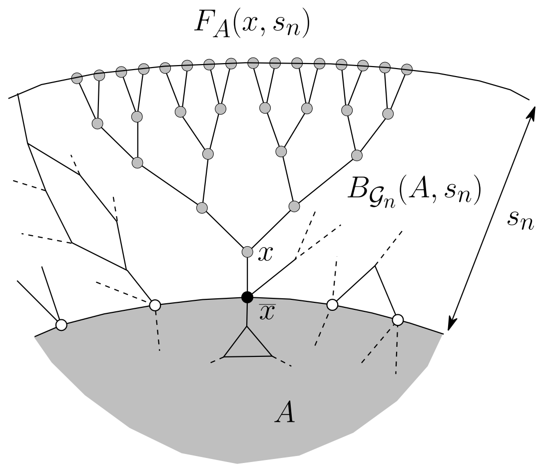

Figure 1: (adapted from [ČTW11]) The point $x \in \partial_{\mathcal{G}_n} A$ is a good vertex at the boundary of A, the points in $F_A(x, s_n)$ are marked grey.

For $A \subsetneq \mathcal{G}_n$ non-empty we set

$$G_A \coloneqq \{\text{good vertices at the boundary of } A\}. \tag{2.29}$$

We are now ready to state Proposition 2.7. Observe the analogies between its statement and (2.24).

Proposition 2.7. *For every $b, b' > 0$ there exists $c_{b,b'} > 0$ such that for $n \geq 1$, $A \subseteq \mathcal{G}_n$ non-empty with $|A| \leq b \ln(N_n)$, $x \in G_A$ and on the event $\left\{ \sup_{z \in A} |\Psi_{\mathcal{G}_n}(z)| \leq b' \sqrt{\ln(N_n)} \right\}$ it holds*

$$\left| \mathbb{E}^{\mathcal{G}_n}\left[\Psi_{\mathcal{G}_n}(x) \,\middle|\, \sigma(\Psi_{\mathcal{G}_n}(y), y \in A) \right] - \frac{1}{d-1} \Psi_{\mathcal{G}_n}(\overline{x}) \right| \leq c_{b,b'} (\ln(N_n))^{-2}, \tag{2.30}$$

$$\left| \mathrm{Var}_{\mathbb{P}^{\mathcal{G}_n}}\left(\Psi_{\mathcal{G}_n}(x) \,\middle|\, \sigma(\Psi_{\mathcal{G}_n}(y), y \in A) \right) - \frac{d}{d-1} \right| \leq c_{b,b'} (\ln(N_n))^{-3}. \tag{2.31}$$

Recall that $\overline{x} \in A$ denotes the unique neighbour of x in A (see (2.26)).

To show Proposition 2.7 and conclude this section we will manipulate the explicit expressions for the conditional expectation and variance obtained in Lemma 2.6. In these expressions one considers the hitting time of A for the simple random walk on $\mathcal{G}_n$ and in the proof of Proposition 2.7 we will look at different situations for when the hitting happens (see the beginning of the proof of Proposition 2.7 below). Since $x \in G_A$ in the statement, the simple random walk started at x has to leave $F_A(x, s_n)$ to hit A and so $P_x^{\mathcal{G}_n}$-almost surely either $H_A = T_{F_A(x,s_n)}$ or $H_A > T_{F_A(x,s_n)}$. We will further split the latter case into whether H_A happens before or after an additional time

$$t_n := \frac{1}{\lambda_{\mathcal{G}_n}}(\ln(N_n))^2, \tag{2.32}$$

by which the distribution of the simple random walk is very close to the stationary distribution (here the uniform distribution on $\mathcal{G}_n$). This follows from e.g. [SC97], Corollary 2.1.5. So in the proof of Proposition 2.7 we will consider the three situations $H_A = T_{F_A(x,s_n)}$, $T_{F_A(x,s_n)} < H_A < T_{F_A(x,s_n)} + t_n$ and $H_A \geq T_{F_A(x,s_n)} + t_n$ separately. Before that, we collect in Lemma 2.8 some preliminary observations about the simple random walk on $\mathcal{G}_n$ and subsequently start with the proof of Proposition 2.7. For the rest of this section we will abbreviate $F_A := F_A(x, s_n)$. It is also convenient to consider the *continuous-time* simple random walk $(\overline{X}_t)_{t\geq 0}$. We remind that for the exit time from $U \subseteq \mathcal{G}_n$ (resp. for the entrance time in $U \subseteq \mathcal{G}_n$) of this walk we use the same notation T_U (resp. H_U) as for the discrete-time simple random walk.

Lemma 2.8. *For $n \geq 1$, $A \subseteq \mathcal{G}_n$ non-empty and $x \in G_A$ one has*

(i) $\frac{1}{d-1} - c(\ln(N_n))^{-7} \leq P_x^{\mathcal{G}_n}[H_A = T_{F_A}] \leq \frac{1}{d-1}$ *and* $E_x^{\mathcal{G}_n}[T_{F_A}] \leq c\ln(N_n)$ (2.33)

(ii) $P_x^{\mathcal{G}_n}[H_A > T_{F_A}] = \sum_{z \in B_{\mathcal{G}_n}(A,s_n)^c} P_x^{\mathcal{G}_n}[\overline{X}_{T_{F_A}} = z, H_A > T_{F_A}].$ (2.34)

Moreover, for every $b > 0$ there exists $c_b > 0$ such that for $n \geq 1$, $A \subseteq \mathcal{G}_n$ non-empty with $|A| \leq b\ln(N_n)$ and $x \in G_A$ one has

(iii) $P_x^{\mathcal{G}_n}[T_{F_A} < H_A < T_{F_A} + t_n] \leq c_b(\ln(N_n))^{-5}$ (2.35)

(iv) $\sum_{w \in \mathcal{G}_n} \left|P_z^{\mathcal{G}_n}[\overline{X}_{t_n} = w, H_A \geq t_n] - \frac{1}{N_n}\right| \leq c_b(\ln(N_n))^{-5}$ *for $z \in \mathcal{G}_n$* (2.36)

(v) $\left|\frac{E_x^{\mathcal{G}_n}[H_A]}{E_\pi^{\mathcal{G}_n}[H_A]} - P_x^{\mathcal{G}_n}[H_A > T_{F_A}]\right| \leq c_b(\ln(N_n))^{-3}.$ (2.37)

Proof. Due to (2.27), the probability $P_x^{\mathcal{G}_n}[H_A = T_{F_A}]$ is equal to the probability that a (discrete-time) random walk on $\mathbb{Z}$ started at 1 and jumping with probability $\frac{d-1}{d}$ to the right and $\frac{1}{d}$ to the left hits 0 before hitting $s_n + 1$. Similarly, $E_x^{\mathcal{G}_n}[T_{F_A}]$ is equal to the expected time until this random walk hits 0 or $s_n + 1$. Thus (see e.g. [Fel68], (2.4) and (3.4) in Chapter 14) it holds

$$\begin{aligned} P_x^{\mathcal{G}_n}[H_A = T_{F_A}] &= 1 - \tfrac{d-2}{d-1}(1 - (\tfrac{1}{d-1})^{s_n+1})^{-1} \leq 1 - \tfrac{d-2}{d-1} = \tfrac{1}{d-1}, \\ E_x^{\mathcal{G}_n}[T_{F_A}] &= \tfrac{d}{d-2}\Big((s_n+1)\tfrac{d-2}{d-1}\frac{1}{1-(\frac{1}{d-1})^{s_n+1}} - 1\Big) \leq 2\tfrac{d}{d-1}(s_n+1) \overset{(2.25)}{\leq} c\ln(N_n). \end{aligned}$$

Since $(1 - (\frac{1}{d-1})^{s_n+1})^{-1} \leq (1 - (\log_{d-1}(N_n))^{-8})^{-1} \leq 1 + c(\ln(N_n))^{-7}$, one also has $P_x^{\mathcal{G}_n}[H_A = T_{F_A}] \geq 1 - \frac{d-2}{d-1}(1 + c(\ln(N_n))^{-7}) \geq \frac{1}{d-1} - c(\ln(N_n))^{-7}$. Thus (2.33) is shown.

To see (2.34) observe that on the event $\{H_A > T_{F_A}\}$, at the moment the simple random walk started at x leaves F_A, it is in some $z \in \partial_{\mathcal{G}_n} F_A \cap B_{\mathcal{G}_n}(A, s_n)^{\mathsf{c}}$ (note that indeed $z \notin B_{\mathcal{G}_n}(A, s_n)$ since else there would exist a path like those excluded by (2.28)). In other words,

$$P_x^{\mathcal{G}_n}\text{-almost surely } \overline{X}_{T_{F_A}} \in \partial_{\mathcal{G}_n} F_A \cap B_{\mathcal{G}_n}(A, s_n)^{\mathsf{c}} \text{ on the event } \{H_A > T_{F_A}\}. \tag{2.38}$$

This shows (2.34). To derive (2.35) we apply the strong Markov property of simple random walk for time T_{F_A} and obtain for $n \geq 1$

$$P_x^{\mathcal{G}_n}[T_{F_A} < H_A < T_{F_A} + t_n] \overset{(2.38)}{\leq} \sup_{z \in B_{\mathcal{G}_n}(A, s_n)^{\mathsf{c}}} P_z^{\mathcal{G}_n}[H_A < t_n]. \tag{2.39}$$

Roughly speaking, the right hand side of (2.39) is small since it is difficult for the simple random walk to hit A within time t_n because it starts at distance larger than s_n from A and the environment is nearly treelike (see (0.2)). More precisely, we can apply [ČTW11], Lemma 3.4 (for $T := t_n$, $r := 0$, $s := s_n$ and using (0.2)) to find $c, c' > 0$ such that for $z \in B_{\mathcal{G}_n}(A, s_n)^{\mathsf{c}}$ one has for $n \geq 1$

$$P_z^{\mathcal{G}_n}[H_A < t_n] \leq \sum_{y \in A} P_z^{\mathcal{G}_n}[H_y < t_n] \leq |A|\big(ct_n(d-1)^{-s_n} + e^{-c't_n}\big) \overset{(2.25)}{\leq} c_b(\ln(N_n))^{-5},$$

where the last inequality also uses the assumption on A, (2.32) and (0.3). This combined with (2.39) gives (2.35).

For (2.36) the idea is that on the event $\{H_A \geq t_n\}$ the simple random walk started at z has, roughly speaking, reached the stationary distribution by time t_n without having hit A. We observe that for $z, w \in \mathcal{G}_n$ one has

$$\begin{aligned}
\big|P_z^{\mathcal{G}_n}[\overline{X}_{t_n} = w, H_A \geq t_n] - \tfrac{1}{N_n}\big| &\leq P_z^{\mathcal{G}_n}[\overline{X}_{t_n} = w, H_A < t_n] + \big|P_z^{\mathcal{G}_n}[\overline{X}_{t_n} = w] - \tfrac{1}{N_n}\big| \\
&\overset{(*)}{\leq} P_z^{\mathcal{G}_n}[\overline{X}_{t_n} = w, H_A < t_n] + \exp(-\lambda_{\mathcal{G}_n} t_n) \\
&\overset{(2.32)}{=} P_z^{\mathcal{G}_n}[\overline{X}_{t_n} = w, H_A < t_n] + \exp(-(\ln(N_n))^2),
\end{aligned}$$

where in $(*)$ we apply [SC97], Corollary 2.1.5. Hence for $n \geq 1$, $z \in \mathcal{G}_n$, one has

$$\sum_{w \in \mathcal{G}_n} \big|P_z^{\mathcal{G}_n}[\overline{X}_{t_n} = w, H_A \geq t_n] - \tfrac{1}{N_n}\big| \leq P_z^{\mathcal{G}_n}[H_A < t_n] + N_n \exp(-(\ln(N_n))^2),$$

which together with the above estimate on $P_z^{\mathcal{G}_n}[H_A < t_n]$ gives (2.36). It remains to show (2.37). We start by computing (using also (3.20) of [ČTW11] in the second inequality)

$$\frac{E_x^{\mathcal{G}_n}[H_A \mathbf{1}_{H_A = T_{F_A}}]}{E_\pi^{\mathcal{G}_n}[H_A]} \leq \frac{E_x^{\mathcal{G}_n}[T_{F_A}]}{E_\pi^{\mathcal{G}_n}[H_A]} \overset{(2.33)}{\leq} c\ln(N_n)\frac{4|A|}{N_n} \leq \frac{c_b(\ln(N_n))^2}{N_n}. \tag{2.40}$$

Now by (2.38) and the strong Markov property of simple random walk for time T_{F_A} one has $E_x^{\mathcal{G}_n}[H_A \mathbf{1}_{H_A > T_{F_A}}] = \sum_{z \in B_{\mathcal{G}_n}(A, s_n)^{\mathsf{c}}} P_x^{\mathcal{G}_n}[\overline{X}_{T_{F_A}} = z, H_A > T_{F_A}] E_z^{\mathcal{G}_n}[H_A]$. This combined with (2.34) shows

$$\left|\frac{E_x^{\mathcal{G}_n}[H_A \mathbf{1}_{H_A > T_{F_A}}]}{E_\pi^{\mathcal{G}_n}[H_A]} - P_x^{\mathcal{G}_n}[H_A > T_{F_A}]\right| \leq \sup_{z \in B_{\mathcal{G}_n}(A, s_n)^{\mathsf{c}}} \left|\frac{E_z^{\mathcal{G}_n}[H_A]}{E_\pi^{\mathcal{G}_n}[H_A]} - 1\right|. \tag{2.41}$$

By [ČTW11], Proposition 3.5, we can bound the absolute value on the right hand side of (2.41) by $c|A|(d-1)^{-s_n}(\ln(N_n))^4 \leq c_b(\ln(N_n))^{-3}$. Since $P_x^{\mathcal{G}_n}$-almost surely either $H_A = T_{F_A}$ or $H_A > T_{F_A}$, the combination of (2.40) and (2.41) concludes the proof. $\square$

Proof of Proposition 2.7. We start with the basic observation that by (2.16) one has $\big|\mathbb{E}^{\mathcal{G}_n}\big[\Psi_{\mathcal{G}_n}(x)\,\big|\,\sigma(\Psi_{\mathcal{G}_n}(y), y\in A)\big] - \frac{1}{d-1}\Psi_{\mathcal{G}_n}(\overline{x})\big| \le U^{\mathcal{G}_n}_{A,x} + V^{\mathcal{G}_n}_{A,x} + W^{\mathcal{G}_n}_{A,x}$, where

$$\begin{aligned}
U^{\mathcal{G}_n}_{A,x} &:= \Big|E^{\mathcal{G}_n}_x\big[\Psi_{\mathcal{G}_n}(\overline{X}_{H_A})\mathbf{1}_{\{H_A=T_{F_A}\}}\big] - \tfrac{1}{d-1}\Psi_{\mathcal{G}_n}(\overline{x})\Big|,\\
V^{\mathcal{G}_n}_{A,x} &:= \big|E^{\mathcal{G}_n}_x\big[\Psi_{\mathcal{G}_n}(\overline{X}_{H_A})\mathbf{1}_{\{T_{F_A}<H_A<T_{F_A}+t_n\}}\big]\big|,\\
W^{\mathcal{G}_n}_{A,x} &:= \Big|E^{\mathcal{G}_n}_x\big[\Psi_{\mathcal{G}_n}(\overline{X}_{H_A})\mathbf{1}_{\{H_A\ge T_{F_A}+t_n\}}\big] - \tfrac{E^{\mathcal{G}_n}_x[H_A]}{E^{\mathcal{G}_n}_\pi[H_A]}E^{\mathcal{G}_n}_\pi\big[\Psi_{\mathcal{G}_n}(\overline{X}_{H_A})\big]\Big|.
\end{aligned}$$

Hence the proof of (2.30) follows once we show that

$$\text{there exists } c_{b,b'}>0 \text{ such that for } n\ge 1,\ A\subseteq\mathcal{G}_n \text{ non-empty with } |A|\le b\ln(N_n),\ x\in G_A \text{ and on the event } \Big\{\sup_{z\in A}|\Psi_{\mathcal{G}_n}(z)|\le b'\sqrt{\ln(N_n)}\Big\} \text{ one has } U^{\mathcal{G}_n}_{A,x}+V^{\mathcal{G}_n}_{A,x}+W^{\mathcal{G}_n}_{A,x}\le c_{b,b'}(\ln(N_n))^{-2}. \tag{2.42}$$

Similarly we have $\big|\mathrm{Var}_{\mathbb{P}^{\mathcal{G}_n}}\big(\Psi_{\mathcal{G}_n}(x)\,\big|\,\sigma(\Psi_{\mathcal{G}_n}(y), y\in A)\big) - \frac{d}{d-1}\big| \le \overline{U}^{\mathcal{G}_n}_{A,x} + \overline{V}^{\mathcal{G}_n}_{A,x} + \overline{W}^{\mathcal{G}_n}_{A,x}$ by (2.17), where

$$\begin{aligned}
\overline{U}^{\mathcal{G}_n}_{A,x} &:= \Big|G_{\mathcal{G}_n}(x,x) - E^{\mathcal{G}_n}_x\big[G_{\mathcal{G}_n}(X_{H_A},x)\mathbf{1}_{\{H_A=T_{F_A}\}}\big] - \tfrac{d}{d-1}\Big|,\\
\overline{V}^{\mathcal{G}_n}_{A,x} &:= E^{\mathcal{G}_n}_x\big[G_{\mathcal{G}_n}(X_{H_A},x)\mathbf{1}_{\{T_{F_A}<H_A<T_{F_A}+t_n\}}\big],\\
\overline{W}^{\mathcal{G}_n}_{A,x} &:= \Big|E^{\mathcal{G}_n}_x\big[G_{\mathcal{G}_n}(X_{H_A},x)\mathbf{1}_{\{H_A\ge T_{F_A}+t_n\}}\big] - \tfrac{E^{\mathcal{G}_n}_x[H_A]}{E^{\mathcal{G}_n}_\pi[H_A]}E^{\mathcal{G}_n}_\pi[G_{\mathcal{G}_n}(X_{H_A},x)]\Big|.
\end{aligned}$$

Thus the proof of (2.31) follows once we show that

$$\text{there exists } c_b>0 \text{ such that for } n\ge 1,\ A\subseteq\mathcal{G}_n \text{ non-empty with } |A|\le b\ln(N_n) \text{ and } x\in G_A \text{ one has } \overline{U}^{\mathcal{G}_n}_{A,x}+\overline{V}^{\mathcal{G}_n}_{A,x}+\overline{W}^{\mathcal{G}_n}_{A,x}\le c_b(\ln(N_n))^{-3}. \tag{2.43}$$

It remains to show (2.42) and (2.43). For (2.42) we bound the three terms $U^{\mathcal{G}_n}_{A,x}$, $V^{\mathcal{G}_n}_{A,x}$ and $W^{\mathcal{G}_n}_{A,x}$ separately. On $\{H_A = T_{F_A}\}$ one has $P^{\mathcal{G}_n}_x$-almost surely $\Psi_{\mathcal{G}_n}(\overline{X}_{H_A}) = \Psi_{\mathcal{G}_n}(\overline{x})$ due to $x\in G_A$. Therefore we deduce $U^{\mathcal{G}_n}_{A,x} = |\Psi_{\mathcal{G}_n}(\overline{x})|\cdot\big|P^{\mathcal{G}_n}_x[H_A = T_{F_A}] - \frac{1}{d-1}\big| \le b'\sqrt{\ln(N_n)}c(\ln(N_n))^{-7}$ by (2.33), where in the last inequality we also use that $\overline{x}\in A$. This shows $U^{\mathcal{G}_n}_{A,x}\le c_{b'}(\ln(N_n))^{-6}$.

We turn to $V^{\mathcal{G}_n}_{A,x}$. By (2.35) we have $V^{\mathcal{G}_n}_{A,x} \le \sup_{y\in A}|\Psi_{\mathcal{G}_n}(y)|\cdot P^{\mathcal{G}_n}_x[T_{F_A} < H_A < T_{F_A}+t_n] \le b'\sqrt{\ln(N_n)}c_b(\ln(N_n))^{-5}$. This shows $V^{\mathcal{G}_n}_{A,x}\le c_{b,b'}(\ln(N_n))^{-4}$.

Finally, we consider $W^{\mathcal{G}_n}_{A,x}$. Let us define

$$\begin{aligned}
Y^{\mathcal{G}_n}_{A,x} &:= \Big|\tfrac{E^{\mathcal{G}_n}_x[H_A]}{E^{\mathcal{G}_n}_\pi[H_A]} - P^{\mathcal{G}_n}_x[H_A > T_{F_A}]\Big|,\\
Z^{\mathcal{G}_n}_{A,x} &:= \Big|E^{\mathcal{G}_n}_x\big[\Psi_{\mathcal{G}_n}(\overline{X}_{H_A})\mathbf{1}_{\{H_A\ge T_{F_A}+t_n\}}\big] - P^{\mathcal{G}_n}_x[H_A > T_{F_A}]E^{\mathcal{G}_n}_\pi\big[\Psi_{\mathcal{G}_n}(\overline{X}_{H_A})\big]\Big|.
\end{aligned} \tag{2.44}$$

By adding and subtracting $\frac{1}{P^{\mathcal{G}_n}_x[H_A>T_{F_A}]}\frac{E^{\mathcal{G}_n}_x[H_A]}{E^{\mathcal{G}_n}_\pi[H_A]}E^{\mathcal{G}_n}_x\big[\Psi_{\mathcal{G}_n}(\overline{X}_{H_A})\mathbf{1}_{\{H_A\ge T_{F_A}+t_n\}}\big]$ inside the expression for $W^{\mathcal{G}_n}_{A,x}$ we obtain

$$W^{\mathcal{G}_n}_{A,x} \le \frac{|E^{\mathcal{G}_n}_x[\Psi_{\mathcal{G}_n}(\overline{X}_{H_A})\mathbf{1}_{\{H_A\ge T_{F_A}+t_n\}}]|}{P^{\mathcal{G}_n}_x[H_A>T_{F_A}]}Y^{\mathcal{G}_n}_{A,x} + \frac{1}{P^{\mathcal{G}_n}_x[H_A>T_{F_A}]}\frac{E^{\mathcal{G}_n}_x[H_A]}{E^{\mathcal{G}_n}_\pi[H_A]}Z^{\mathcal{G}_n}_{A,x}. \tag{2.45}$$

To the first term on the right hand side of (2.45) we apply $P^{\mathcal{G}_n}_x[H_A > T_{F_A}] \ge \frac{d-2}{d-1}$ (by (2.33)) as well as (2.37) and the assumption on the supremum of $\Psi_{\mathcal{G}_n}$ on A. For the second term we first observe (2.37) and then again use $P_x^{\mathcal{G}_n}[H_A > T_{F_A}] \geq \frac{d-2}{d-1}$. In this way we obtain

$$W_{A,x}^{\mathcal{G}_n} \leq c_{b,b'}(\ln(N_n))^{-2} + (1 + c_b(\ln(N_n))^{-3})Z_{A,x}^{\mathcal{G}_n}. \tag{2.46}$$

We proceed to bound $Z_{A,x}^{\mathcal{G}_n}$. By (2.38) and the strong Markov property for time T_{F_A} it holds

$$\begin{aligned} &E_x^{\mathcal{G}_n}\left[\Psi_{\mathcal{G}_n}(\overline{X}_{H_A})\mathbf{1}_{\{H_A \geq T_{F_A}+t_n\}}\right] \\ &\quad = \sum_{z \in B_{\mathcal{G}_n}(A,s_n)^{\mathsf{c}}} P_x^{\mathcal{G}_n}[\overline{X}_{T_{F_A}} = z, H_A > T_{F_A}] E_z^{\mathcal{G}_n}\left[\Psi_{\mathcal{G}_n}(\overline{X}_{H_A})\mathbf{1}_{\{H_A \geq t_n\}}\right]. \end{aligned}$$

This combined with (2.34) implies $Z_{A,x}^{\mathcal{G}_n} \leq \sup_{z \in B_{\mathcal{G}_n}(A,s_n)^{\mathsf{c}}} \left|E_z^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(\overline{X}_{H_A})\mathbf{1}_{\{H_A \geq t_n\}}] - E_\pi^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(\overline{X}_{H_A})]\right|$. Now for $z \in B_{\mathcal{G}_n}(A, s_n)^{\mathsf{c}}$, by the Markov property applied at time t_n and the definition of $E_\pi^{\mathcal{G}_n}$,

$$\begin{aligned} &\left|E_z^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(\overline{X}_{H_A})\mathbf{1}_{\{H_A \geq t_n\}}] - E_\pi^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(\overline{X}_{H_A})]\right| \\ &\leq \sum_{w \in \mathcal{G}_n} \left|E_w^{\mathcal{G}_n}[\Psi_{\mathcal{G}_n}(\overline{X}_{H_A})]\right| \cdot \left|P_z^{\mathcal{G}_n}[\overline{X}_{t_n} = w, H_A \geq t_n] - \tfrac{1}{N_n}\right| \overset{(2.36)}{\leq} c_{b,b'}(\ln(N_n))^{-4}, \end{aligned}$$

where in the last inequality we also use the assumption on the supremum of $\Psi_{\mathcal{G}_n}$ on A. All in all we have shown $Z_{A,x}^{\mathcal{G}_n} \leq c_{b,b'}(\ln(N_n))^{-4}$. Thus by (2.46) we deduce $W_{A,x}^{\mathcal{G}_n} \leq c_{b,b'}(\ln(N_n))^{-2}$ and the proof of (2.42) is complete.

We come to the proof of (2.43) for which we bound the three terms $\overline{U}_{A,x}^{\mathcal{G}_n}$, $\overline{V}_{A,x}^{\mathcal{G}_n}$ and $\overline{W}_{A,x}^{\mathcal{G}_n}$ separately. For $\overline{U}_{A,x}^{\mathcal{G}_n}$ we first note that one has $E_x^{\mathcal{G}_n}\left[G_{\mathcal{G}_n}(X_{H_A}, x)\mathbf{1}_{\{H_A = T_{F_A}\}}\right] = E_x^{\mathcal{G}_n}\left[G_{\mathcal{G}_n}(X_{T_{F_A}}, x)\mathbf{1}_{\{H_A = T_{F_A}\}}\right] = E_x^{\mathcal{G}_n}\left[G_{\mathcal{G}_n}(X_{T_{F_A}}, x)\right] - E_x^{\mathcal{G}_n}\left[G_{\mathcal{G}_n}(X_{T_{F_A}}, x)\mathbf{1}_{\{H_A > T_{F_A}\}}\right]$. By (2.38), on the event $\{H_A > T_{F_A}\}$ the simple random walk started at x is at distance s_n from x when it leaves F_A. Therefore

$$E_x^{\mathcal{G}_n}\left[G_{\mathcal{G}_n}(X_{T_{F_A}}, x)\mathbf{1}_{\{H_A > T_{F_A}\}}\right] \leq \sup_{z \in S_{\mathcal{G}_n}(x,s_n)} G_{\mathcal{G}_n}(z,x)\, \underbrace{P_x^{\mathcal{G}_n}[H_A > T_{F_A}]}_{\leq 1} \underset{(2.25)}{\overset{(1.23)}{\leq}} c(\ln(N_n))^{-8}.$$

Thus we have

$$\begin{aligned} \overline{U}_{A,x}^{\mathcal{G}_n} &\leq \left|G_{\mathcal{G}_n}(x,x) - E_x^{\mathcal{G}_n}\left[G_{\mathcal{G}_n}(X_{T_{F_A}}, x)\right] - \tfrac{d}{d-1}\right| + c(\ln(N_n))^{-8} \\ &\underset{(2.33)}{\overset{(1.18)}{\leq}} \left|g_{\mathcal{G}_n}^{F_A}(x,x) - \tfrac{d}{d-1}\right| + \tfrac{c\ln(N_n)}{N_n} + c(\ln(N_n))^{-8}. \end{aligned} \tag{2.47}$$

Note that by assumption $\mathtt{tx}(F_A) = 0$. So if we define $B := B_{\mathbb{T}_d}^+(\text{o}, s_n) \setminus \{\text{o}\} \subseteq \mathbb{T}_d$ and take $x_1 \in S_{\mathbb{T}_d}^+(\text{o}, 1)$, then by definition we have $g_{\mathcal{G}_n}^{F_A}(x,x) = g_{\mathbb{T}_d}^B(x_1, x_1)$. From (1.7) we see that

$$\begin{aligned} g_{\mathbb{T}_d}^B(x_1,x_1) &= g_{\mathbb{T}_d}(x_1,x_1) - E_{x_1}^{\mathbb{T}_d}[g_{\mathbb{T}_d}(X_{T_B}, x_1)] \\ &= g_{\mathbb{T}_d}(x_1,x_1) - g_{\mathbb{T}_d}(\text{o},x_1)P_{x_1}^{\mathbb{T}_d}[H_{\text{o}} = T_B] - g_{\mathbb{T}_d}(z,x_1)P_{x_1}^{\mathbb{T}_d}[H_{\text{o}} > T_B] \end{aligned}$$

for any fixed $z \in S_{\mathbb{T}_d}^+(\text{o}, s_n + 1)$. By (1.6) this shows that

$$g_{\mathbb{T}_d}^B(x_1,x_1) = \frac{d-1}{d-2} - \frac{d-1}{d-2}\frac{1}{d-1}P_x^{\mathcal{G}_n}[H_A = T_{F_A}] - \frac{d-1}{d-2}\Big(\frac{1}{d-1}\Big)^{s_n} P_x^{\mathcal{G}_n}[H_A > T_{F_A}].$$

So we have obtained

$$\begin{aligned}\left|g^{F_A}_{\mathcal{G}_n}(x,x)-\tfrac{d}{d-1}\right| &\le \left|\tfrac{d-1}{d-2}-\tfrac{1}{d-2}P^{\mathcal{G}_n}_x[H_A=T_{F_A}]-\tfrac{d}{d-1}\right|+\tfrac{d-1}{d-2}\left(\tfrac{1}{d-1}\right)^{s_n}P^{\mathcal{G}_n}_x[H_A>T_{F_A}]\\ &\underset{(2.25)}{\overset{(2.33)}{\le}} \tfrac{1}{d-2}c(\ln(N_n))^{-7}+c(\ln(N_n))^{-8}.\end{aligned}$$

This, together with (2.47) shows $\overline{U}^{\mathcal{G}_n}_{A,x}\le c(\ln(N_n))^{-7}$.

We turn to $\overline{V}^{\mathcal{G}_n}_{A,x}$. By (1.22) there exists $c>0$ such that $\sup_{y,z\in\mathcal{G}_n}G_{\mathcal{G}_n}(y,z)\le c$. Therefore $\overline{V}^{\mathcal{G}_n}_{A,x}\le c\,P^{\mathcal{G}_n}_x[T_{F_A}<H_A<T_{F_A}+t_n]$ and so (2.35) implies $\overline{V}^{\mathcal{G}_n}_{A,x}\le c_b(\ln(N_n))^{-5}$.

Finally, we consider $\overline{W}^{\mathcal{G}_n}_{A,x}$. Let us define

$$\overline{Z}^{\mathcal{G}_n}_{A,x}:=\left|E^{\mathcal{G}_n}_x\left[G_{\mathcal{G}_n}(X_{H_A},x)\mathbf{1}_{\{H_A\ge T_{F_A}+t_n\}}\right]-P^{\mathcal{G}_n}_x[H_A>T_{F_A}]E^{\mathcal{G}_n}_\pi[G_{\mathcal{G}_n}(X_{H_A},x)]\right|$$

and recall $Y^{\mathcal{G}_n}_{A,x}$ from (2.44). Inside $\overline{W}^{\mathcal{G}_n}_{A,x}$ we can add and subtract $\frac{1}{P^{\mathcal{G}_n}_x[H_A>T_{F_A}]}\frac{E^{\mathcal{G}_n}_x[H_A]}{E^{\mathcal{G}_n}_\pi[H_A]}\cdot E^{\mathcal{G}_n}_x\left[G_{\mathcal{G}_n}(X_{H_A},x)\mathbf{1}_{\{H_A\ge T_{F_A}+t_n\}}\right]$ to obtain

$$\overline{W}^{\mathcal{G}_n}_{A,x}\le\frac{E^{\mathcal{G}_n}_x[G_{\mathcal{G}_n}(X_{H_A},x)\mathbf{1}_{\{H_A\ge T_{F_A}+t_n\}}]}{P^{\mathcal{G}_n}_x[H_A>T_{F_A}]}Y^{\mathcal{G}_n}_{A,x}+\frac{1}{P^{\mathcal{G}_n}_x[H_A>T_{F_A}]}\frac{E^{\mathcal{G}_n}_x[H_A]}{E^{\mathcal{G}_n}_\pi[H_A]}\overline{Z}^{\mathcal{G}_n}_{A,x}. \tag{2.48}$$

To the first term on the right hand side of (2.48) we apply $P^{\mathcal{G}_n}_x[H_A>T_{F_A}]\ge\frac{d-2}{d-1}$ (by (2.33)) as well as (2.37) and $\sup_{y,z\in\mathcal{G}_n}G_{\mathcal{G}_n}(y,z)\le c$ (by (1.22)). For the second term we first observe (2.37) and then again use $P^{\mathcal{G}_n}_x[H_A>T_{F_A}]\ge\frac{d-2}{d-1}$. In this way we obtain

$$\overline{W}^{\mathcal{G}_n}_{A,x}\le c_b(\ln(N_n))^{-3}+(1+c_b(\ln(N_n))^{-3})\overline{Z}^{\mathcal{G}_n}_{A,x}. \tag{2.49}$$

We proceed to bound $\overline{Z}^{\mathcal{G}_n}_{A,x}$. By (2.38) and the strong Markov property it holds

$$\begin{aligned}&E^{\mathcal{G}_n}_x\left[G_{\mathcal{G}_n}(X_{H_A},x)\mathbf{1}_{\{H_A\ge T_{F_A}+t_n\}}\right]\\ &\qquad=\sum_{z\in B_{\mathcal{G}_n}(A,s_n)^{\mathsf{c}}}P^{\mathcal{G}_n}_x[\overline{X}_{T_{F_A}}=z,H_A>T_{F_A}]E^{\mathcal{G}_n}_z[G_{\mathcal{G}_n}(X_{H_A},x)\mathbf{1}_{\{H_A\ge t_n\}}].\end{aligned}$$

This combined with (2.34) gives $\overline{Z}^{\mathcal{G}_n}_{A,x}\le\sup_{z\in B_{\mathcal{G}_n}(A,s_n)^{\mathsf{c}}}\left|E^{\mathcal{G}_n}_z[G_{\mathcal{G}_n}(X_{H_A},x)\mathbf{1}_{\{H_A\ge t_n\}}]-E^{\mathcal{G}_n}_\pi[G_{\mathcal{G}_n}(X_{H_A},x)]\right|$. Now for $z\in B_{\mathcal{G}_n}(A,s_n)^{\mathsf{c}}$, by the Markov property applied at time t_n and the definition of $E^{\mathcal{G}_n}_\pi$,

$$\begin{aligned}&\left|E^{\mathcal{G}_n}_z[G_{\mathcal{G}_n}(X_{H_A},x)\mathbf{1}_{\{H_A\ge t_n\}}]-E^{\mathcal{G}_n}_\pi[G_{\mathcal{G}_n}(X_{H_A},x)]\right|\\ &\le\sum_{w\in\mathcal{G}_n}E^{\mathcal{G}_n}_w[G_{\mathcal{G}_n}(X_{H_A},x)]\cdot\left|P^{\mathcal{G}_n}_z[\overline{X}_{t_n}=w,H_A\ge t_n]-\tfrac{1}{N_n}\right|\overset{(2.36)}{\le}c_b(\ln(N_n))^{-5},\end{aligned}$$

where in the last inequality we again use $\sup_{y,w\in\mathcal{G}_n}G_{\mathcal{G}_n}(y,w)\le c$ by (1.22). All in all we have shown $\overline{Z}^{\mathcal{G}_n}_{A,x}\le c_b(\ln(N_n))^{-5}$. Thus by (2.49) we deduce $\overline{W}^{\mathcal{G}_n}_{A,x}\le c_b(\ln(N_n))^{-3}$ and (2.43) is shown. This concludes the proof of Proposition 2.7 and Section 2.2. □

3 Microscopic components in the subcritical phase

We start the analysis of level-set percolation of the zero-average Gaussian free field $\Psi_{\mathcal{G}_n}$ on $\mathcal{G}_n$. The goal of this section is to show (0.8) in the form of Theorem 3.1 below, i.e. the existence of a subcritical phase in which, with high probability for large n, level sets of $\Psi_{\mathcal{G}_n}$ only have connected components of cardinality at most logarithmic in the size of the graph. To precisely state the result, we recall from the introduction the critical value $h_\star$ for level-set percolation of the Gaussian free field $\varphi_{\mathbb{T}_d}$ on $\mathbb{T}_d$ (see (0.7)) and also the notation $E^{\geq h}_{\Psi_{\mathcal{G}_n}}$ for the level set of $\Psi_{\mathcal{G}_n}$ above level $h \in \mathbb{R}$ (see (0.5)). For $h \in \mathbb{R}$ we further denote by $\mathcal{C}^{\mathcal{G}_n,h}_{\max}$ an arbitrary connected component of $E^{\geq h}_{\Psi_{\mathcal{G}_n}}$ with maximal number of vertices. We will only be interested in its cardinality. Moreover, for $x \in \mathcal{G}_n$ and $h \in \mathbb{R}$ we define $\mathcal{C}^{\mathcal{G}_n,h}_x$ to be the connected component of $E^{\geq h}_{\Psi_{\mathcal{G}_n}}$ containing x. The main result of this section is

Theorem 3.1. *Let $h > h_\star$. Then for all $\kappa > 0$ there exist $c_{h,\kappa} > 0$ and $K_{h,\kappa} > 0$ such that for all $n \geq 1$*

$$\mathbb{P}^{\mathcal{G}_n}\left[|\mathcal{C}^{\mathcal{G}_n,h}_{\max}| \geq K_{h,\kappa} \ln(N_n)\right] \leq c_{h,\kappa} N_n^{-\kappa}.$$

In particular, for some $K_h > 0$ one has $\lim_{n\to\infty} \mathbb{P}^{\mathcal{G}_n}\left[|\mathcal{C}^{\mathcal{G}_n,h}_{\max}| \leq K_h \ln(N_n)\right] = 1$.

Before explaining the details of the proof of Theorem 3.1, let us make the basic observation that a union bound reduces the problem to show that for $h > h_\star$ and for all $\kappa > 0$ there exist $c_{h,\kappa} > 0$ and $K_{h,\kappa} > 0$ such that for all $n \geq 1$ and $x \in \mathcal{G}_n$

$$\mathbb{P}^{\mathcal{G}_n}\left[|\mathcal{C}^{\mathcal{G}_n,h}_x| \geq K_{h,\kappa} \ln(N_n)\right] \leq c_{h,\kappa} N_n^{-1-\kappa}. \tag{3.1}$$

So it remains to show (3.1). We will make use of a certain exploration process exploring $\mathcal{C}^{\mathcal{G}_n,h}_x$ for a fixed $x \in \mathcal{G}_n$. This will enable us to control $\mathbb{P}^{\mathcal{G}_n}\left[|\mathcal{C}^{\mathcal{G}_n,h}_x| \geq K_{h,\kappa} \ln(N_n)\right]$. A similar approach has for example been followed in [ČTW11] to prove a result analogous to the above Theorem 3.1 but for the vacant set of simple random walk on $\mathcal{G}_n$ in place of the level set of the zero-average Gaussian free field.

We now give the idea of the proof of (3.1). The details of the exploration process itself are given afterwards. A crucial ingredient is the precise understanding of the conditional distribution of the zero-average Gaussian free field on non-explored vertices given its value on already explored vertices. As we have seen in Proposition 2.7 in Section 2.2, under certain geometric conditions the conditional distribution of $\Psi_{\mathcal{G}_n}$ shows strong similarities with the conditional distribution of the Gaussian free field $\varphi_{\mathbb{T}_d}$ on $\mathbb{T}_d$. While exploring $\mathcal{C}^{\mathcal{G}_n,h}_x$, the exploration process will separate the vertices found in $\mathcal{C}^{\mathcal{G}_n,h}_x$ into a union of rooted disjoint subtrees of $\mathcal{G}_n$ in which all vertices except for the root satisfy the aforementioned geometric conditions. In this way we reduce the proof of (3.1) to a control of the number of vertices contained in these union of subtrees (Proposition 3.2). As a result from [ČTW11] shows (see also Lemma 3.3), the number of steps the exploration process encounters a situation in which the geometric assumptions fail to be satisfied is not too large. This controls the number of distinct subtrees created by the exploration process because in each subtree there is exactly one vertex which does not satisfy the conditions (its root). Since the other vertices of a subtree satisfy the geometric conditions, we can employ the similarity between the conditional distribution of $\Psi_{\mathcal{G}_n}$ and $\varphi_{\mathbb{T}_d}$ to couple the zero-average Gaussian free field on each distinct subtree separately with an independent copy of the Gaussian free field $\varphi_{\mathbb{T}_d}$ on $\mathbb{T}_d$ (Lemma 3.4). This translates the question about the number of vertices contained in the disjoint subtrees into the number of vertices contained in connected components of the level set of $\varphi_{\mathbb{T}_d}$ (Corollary 3.5). A result from [AČ19] (recalled in (1.14)) about exponential moments of the size of these connected components then ultimately leads to the proof of Proposition 3.2 and hence of (3.1).

We now describe the exploration process exploring $\mathcal{C}_x^{\mathcal{G}_n,h}$ for a fixed $x \in \mathcal{G}_n$ and to facilitate the discussion we include a concrete algorithm implementing it (Algorithm 1). The exploration process is a *modified* breadth-first-search that discovers the field $\Psi_{\mathcal{G}_n}$ on the graph step by step. It employs *two* queues (a primary and a secondary one) that work in the usual first-in-first-out manner and store the vertices to be explored. The exploration process starts by revealing $\Psi_{\mathcal{G}_n}(x)$. The vertices where $\Psi_{\mathcal{G}_n}$ has been revealed are called explored and they can be either part of $\mathcal{C}_x^{\mathcal{G}_n,h}$ or not. If a vertex is explored and is revealed to be part of $\mathcal{C}_x^{\mathcal{G}_n,h}$, then its neighbours which are neither already explored nor already in one of the two queues are added to the primary queue. To avoid ambiguity, we suppose that the vertices of $\mathcal{G}_n$ are equipped with some ordering and that they are added to the queue following this ordering. Vertices taken out of the primary queue are first checked to be *good vertices at the boundary of the so far explored vertices* (recall (2.29) and above it for the definition): if they are, the exploration process proceeds with their exploration; if they are not, they are transferred to the secondary queue and their exploration is postponed. The first vertex in the secondary queue is only taken out to be explored if the primary queue is empty.

To formalise this exploration process we now give an algorithm implementing it (see Algorithm 1 below). The algorithm constructs on some auxiliary probability space $(\Omega, \mathcal{A}, \mathbb{P})$ a family of random variables $(\psi(z))_{z\in B}$ such that $(\psi(z))_{z\in B}$ under $\mathbb{P}$ has the same distribution as $(\Psi_{\mathcal{G}_n}(z))_{z\in B}$ under $\mathbb{P}^{\mathcal{G}_n}$. Here $B \subseteq \mathcal{G}_n$ is some (random) connected set of vertices containing x. We use PQ, SQ and E to denote the evolving sets of vertices in the primary queue, vertices in the secondary queue and explored vertices during the run of the algorithm. Furthermore, we also keep track of the explored vertices $z \in \mathsf{E}$ for which $\psi(z) \geq h$ using the set $\mathsf{C} \subseteq \mathsf{E}$. Additionally to the exploration, the algorithm aggregates the vertices discovered to be in C into disjoint subtrees $(\mathsf{T}^y)_y$ of $\mathcal{G}_n$ indexed by *bad vertices* $y \in \mathcal{G}_n$ (meaning they were in SQ at some point of the algorithm). Moreover, the algorithm stops for one of two reasons: either because both the primary and secondary queue are empty, or because it already discovered that C has at least size $K_{h,\kappa} \ln(N_n)$ for some $K_{h,\kappa}$ to be specified later (below (3.22)).

We need some more notation for the algorithm. Let $(\xi_z)_{z\in\mathcal{G}_n}$ be i.i.d. standard normal random variables on the auxiliary probability space $(\Omega, \mathcal{A}, \mathbb{P})$. For $A \subseteq \mathcal{G}_n$ non-empty and $u \in \mathcal{G}_n$ we abbreviate by $a(u, \psi, A)$ the right hand side of (2.16) where x and $\Psi_{\mathcal{G}_n}$ are replaced by u and ψ. In particular, $a(u, \psi, A)$ is a random variable measurable with respect to $\sigma(\psi(w), w \in A)$. By $b(u, A)$ we abbreviate the right hand side of (2.17) where x is replaced by u. For $A = \emptyset$ and $u \in \mathcal{G}_n$ we define $a(u, \psi, \emptyset) := 0$ and $b(u, \emptyset) := G_{\mathcal{G}_n}(u, u)$. By Lemma 2.6 and the fact that ψ is a Gaussian field, we have that

$$\text{for } A \subseteq \mathcal{G}_n \text{ and } u \in \mathcal{G}_n \text{ the random variable } a(u,\psi,A)+\xi_u \cdot b(u,A)^{\frac{1}{2}} \text{ under } \mathbb{P} \text{ has the same distribution as } \Psi_{\mathcal{G}_n}(u) \text{ conditional on } \sigma(\Psi_{\mathcal{G}_n}(w), w \in A) \text{ under } \mathbb{P}^{\mathcal{G}_n}. \tag{3.2}$$

The algorithm is as follows:

Algorithm 1

1: set $\mathsf{PQ} := \emptyset$, $\mathsf{SQ} := \{x\}$, $\mathsf{E} := \emptyset$, $\mathsf{C} := \emptyset$ and also $\mathsf{T}^w := \emptyset$ for all $w \in \mathcal{G}_n$
2: **while** secondary queue SQ is not empty **do**
3: take vertex y out of SQ
4: generate the random variable $\psi(y) := a(y, \psi, \mathsf{E}) + \xi_y \cdot b(y, \mathsf{E})^{\frac{1}{2}}$
5: add y to the set E of explored vertices
6: **if** $\psi(y) \geq h$ **then**
7: add y to the subtree T^y and to the set C
8: **if** $|\mathsf{C}| \geq K_{h,\kappa} \ln(N_n)$ **then** stop the algorithm
9: **end if**
10: add all neighbours of y which are neither already explored nor in any of the two queues to the primary queue PQ
11: **while** primary queue PQ is not empty **do**
12: take vertex z out of PQ
13: **if** z is not a good vertex at the boundary of E, that is, $z \notin G_{\mathsf{E}}$, **then**
14: add z to the secondary queue SQ
15: **else**
16: generate the random variable $\psi(z) := a(z, \psi, \mathsf{E}) + \xi_z \cdot b(z, \mathsf{E})^{\frac{1}{2}}$
17: add z to the set E of explored vertices
18: **if** $\psi(z) \geq h$ **then**
19: add z to the subtree T^y and to the set C
20: **if** $|\mathsf{C}| \geq K_{h,\kappa} \ln(N_n)$ **then** stop the algorithm
21: **end if**
22: add all neighbours of z which are neither already explored nor in any of the two queues to the primary queue PQ
23: **end if**
24: **end if**
25: **end while**
26: **end if**
27: **end while**

Let E_{end}, C_{end} and $\mathsf{T}^w_{\text{end}}$, $w \in \mathcal{G}_n$, denote the sets E, C and T^w, $w \in \mathcal{G}_n$, at the end of the algorithm. By that moment we have constructed $(\psi(z))_{z\in\mathsf{E}_{\text{end}}}$ and (see (3.2))

$$(\psi(z))_{z\in\mathsf{E}_{\text{end}}} \text{ under } \mathbb{P} \text{ has the same distribution as } (\Psi_{\mathcal{G}_n}(z))_{z\in\mathsf{E}_{\text{end}}} \text{ under } \mathbb{P}^{\mathcal{G}_n}. \tag{3.3}$$

By construction of the algorithm one has $|\mathsf{C}_{\text{end}}| \leq K_{h,\kappa} \ln(N_n) + 1$ and so by (0.1) also

$$|\mathsf{E}_{\text{end}}| \leq d(K_{h,\kappa} \ln(N_n) + 1). \tag{3.4}$$

This is due to $\mathsf{E} \subseteq B_{\mathcal{G}_n}(\mathsf{C}, 1)$ with $B_{\mathcal{G}_n}(\emptyset, 1) := \{x\}$ holding at any moment of the algorithm since a vertex can only get explored (except for x) if at some point it was added to a queue, meaning it was a neighbour of a vertex added to C.

Note that, whenever some $y \in \mathcal{G}_n$ is taken out of SQ on line 3 of the algorithm (a bad vertex), one has $\mathsf{PQ} = \emptyset$ at that moment by construction. Until the next bad vertex is taken out of SQ, all vertices $z \in \mathcal{G}_n$ considered by the algorithm and which are found to be good and in C will be part of $\mathsf{T}^y_{\text{end}}$. So if $y_1, \dots, y_{k_{\text{end}}}$ denote the successive vertices that were taken out of SQ during the algorithm, then $\mathsf{C}_{\text{end}} = \bigcup_{i=1}^{k_{\text{end}}} \mathsf{T}^{y_i}_{\text{end}}$. In particular, $y_1 = x$ and k_{end} is the total number of bad vertices encountered by the algorithm.

Furthermore, on the event that the algorithm terminates because both queues become empty (and not because at some point $|\mathsf{C}_{\text{end}}| \geq K_{h,\kappa} \ln(N_n)$), note that $|\mathsf{C}_{\text{end}}|$ has the same distribution as $|\mathcal{C}_x^{\mathcal{G}_n,h}|$ under $\mathbb{P}^{\mathcal{G}_n}$ by (3.3). Therefore $\mathbb{P}\big[|\mathsf{C}_{\text{end}}| < K_{h,\kappa} \ln(N_n)\big] = \mathbb{P}^{\mathcal{G}_n}\big[|\mathcal{C}_x^{\mathcal{G}_n,h}| < K_{h,\kappa} \ln(N_n)\big]$.

We want to distinguish the situation in which the field ψ produced by the algorithm has anomalous values, meaning $|\psi(z)| \geq M_n$ for some $z \in \mathsf{E}_{\text{end}}$ and $M_n > 0$. We are going to specify this value now. Note that for any $\kappa > 0$ there is $c_\kappa > 0$ such that

$$\mathbb{P}^{\mathcal{G}_n}\Big[\sup_{z\in\mathcal{G}_n} |\Psi_{\mathcal{G}_n}(z)| \geq c_\kappa \sqrt{\ln(N_n)}\Big] \leq 2N_n^{-1-\kappa} \quad \text{for all } n \geq 1. \tag{3.5}$$

This can be shown by the same computations as in [RS13], equations (2.35)–(2.38), replacing $g(0)$ therein with $\sup_{z\in\mathcal{G}_n} G_{\mathcal{G}_n}(z,z)$, which is bounded by $3\frac{d-1}{d-2}$ (see (1.23)). Use also use $\mathbb{P}^{\mathcal{G}_n}\big[\sup_{z\in\mathcal{G}_n} |\Psi_{\mathcal{G}_n}(z)| \geq a\big] \leq 2\mathbb{P}^{\mathcal{G}_n}\big[\sup_{z\in\mathcal{G}_n} \Psi_{\mathcal{G}_n}(z) \geq a\big]$ for (3.5). We set

$$M_n \coloneqq c_\kappa \sqrt{\ln(N_n)}.$$

So one has

$$\begin{aligned}
&\mathbb{P}^{\mathcal{G}_n}\big[|\mathcal{C}_x^{\mathcal{G}_n,h}| \geq K_{h,\kappa} \ln(N_n)\big] = \mathbb{P}\big[|\mathsf{C}_{\text{end}}| \geq K_{h,\kappa} \ln(N_n)\big] \\
&\quad\leq\ \mathbb{P}\Big[|\mathsf{C}_{\text{end}}| \geq K_{h,\kappa} \ln(N_n)\,,\ \sup_{z\in\mathsf{E}_{\text{end}}} |\psi(z)| < M_n\Big] + \mathbb{P}\big[|\psi(z)| \geq M_n \text{ for some } z \in \mathsf{E}_{\text{end}}\big] \\
&\quad\underset{(3.5)}{\overset{(3.3)}{\leq}}\ \mathbb{P}\Big[\sum_{i=1}^{k_{\text{end}}} |\mathsf{T}_{\text{end}}^{y_i}| \geq K_{h,\kappa} \ln(N_n)\,,\ \sup_{z\in\mathsf{E}_{\text{end}}} |\psi(z)| < M_n\Big] + 2N_n^{-1-\kappa}.
\end{aligned}$$

Thus in order to show (3.1) and ultimately Theorem 3.1 we need to show

Proposition 3.2. *Let $h > h_\star$. Then for all $\kappa > 0$ there exist $c_{h,\kappa} > 0$ and $K_{h,\kappa} > 0$ such that for all $n \geq 1$ and $x \in \mathcal{G}_n$ one has for the Algorithm 1 above*

$$\mathbb{P}\Big[\sum_{i=1}^{k_{\text{end}}} |\mathsf{T}_{\text{end}}^{y_i}| \geq K_{h,\kappa} \ln(N_n)\,,\ \sup_{z\in\mathsf{E}_{\text{end}}} |\psi(z)| < M_n\Big] \leq c_{h,\kappa} N_n^{-1-\kappa}. \tag{3.6}$$

The proof of Proposition 3.2 relies on the following two lemmas. The first one (Lemma 3.3, already proven in [ČTW11]) bounds the number of bad vertices k_{end} encountered by Algorithm 1, that is, the number of vertices of $\mathcal{G}_n$ that at some point during the run of the algorithm were in the secondary queue SQ. The second one (Lemma 3.4) constructs for each $i = 1, \ldots, k_{\text{end}}$ a coupling of ψ on $\mathsf{T}_{\text{end}}^{y_i}$ with an independent copy of $\varphi_{\mathbb{T}_d}$, showing that ψ on $\mathsf{T}_{\text{end}}^{y_i}$ can be approximated by $\varphi_{\mathbb{T}_d}$. This makes use of Proposition 2.7. Via Corollary 3.5 of Lemma 3.4 we then prove Proposition 3.2.

Lemma 3.3. *There exists $c_1 > 0$ such that for all $n \geq 1$ and $x \in \mathcal{G}_n$ one has for the above Algorithm 1 that $k_{\text{end}} \leq c_1 K_{h,\kappa} s_n^2 =: k_{\max}$ (recall that s_n is given in (2.25)).*

Proof. This follows from [ČTW11], Proposition 5.4. Although the algorithm employed there does not exactly match our algorithm, the proof does not rely on a specific algorithm (as explained in the proof of Proposition 5.4 in [ČTW11]). It is purely deterministic and only uses the properties (0.1)–(0.3) of $\mathcal{G}_n$. □

Lemma 3.4. *Let $h \in \mathbb{R}$ and $\varepsilon > 0$. Consider Algorithm 1 and recall $k_{\max}$ from Lemma 3.3. Then on the same auxiliary space $(\Omega, \mathcal{A}, \mathbb{P})$ as ψ one can define centred Gaussian fields $\phi^1, \dots, \phi^{k_{\max}}$ on $\mathbb{T}_d$ such that, conditionally on $\psi(y_1), \dots, \psi(y_{k_{\text{end}}})$, the following properties hold (see (1.11) for notation):*

- *for all n large enough and all $i = 1, \dots, k_{\text{end}}$ there exists a set B^i with $\mathsf{T}^{y_i}_{\text{end}} \subseteq B^i \subseteq B_{\mathcal{G}_n}(\mathsf{T}^{y_i}_{\text{end}}, 1)$ and an injection $\tau^i : B^i \to \mathbb{T}_d$ such that $\tau^i(\mathsf{T}^{y_i}_{\text{end}})$ is a connected subset of $\mathbb{T}_d$ containing the root $\mathrm{o} \in \mathbb{T}_d$ and on the event $\{\sup_{z \in \mathsf{E}_{\text{end}}} |\psi(z)| < M_n\}$ one has $\left|\psi(z) - \phi^i(\tau^i(z))\right| \le \varepsilon$ for all $z \in B^i$* (3.7)
- *ϕ^i has the same distribution as $\varphi_{\mathbb{T}_d}$ under $\mathbb{P}^{\mathbb{T}_d}_{\psi(y_i)}$ for all $i = 1, \dots, k_{\text{end}}$,* (3.8)
  ϕ^i has the same distribution as $\varphi_{\mathbb{T}_d}$ under $\mathbb{P}^{\mathbb{T}_d}_{M_n}$ for all $i = k_{\text{end}} + 1, \dots, k_{\max}$
- *$\phi^1, \dots, \phi^{k_{\max}}$ are independent.* (3.9)

Proof. Let Y^i_x for $x \in \mathbb{T}_d \setminus \{\mathrm{o}\}$ and $1 \le i \le k_{\max}$ be a sequence of i.i.d. random variables of distribution $\mathcal{N}(0, \frac{d}{d-1})$ defined on the auxiliary probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $i \in \{1, \dots, k_{\text{end}}\}$. As explained below (3.4), the subtree $\mathsf{T}^{y_i}_{\text{end}}$ of $\mathcal{G}_n$ is constructed between line 3 (when y_i is taken out of SQ) and line 26 of the algorithm (after which the next bad vertex y_{i+1} is taken out of SQ or the algorithm terminates because $i = k_{\text{end}}$). The injection τ^i and the random field ϕ^i will be defined according to the behaviour of the algorithm during this time.

On line 4 of the algorithm we generate $\psi(y_i)$. If $\psi(y_i) < h$, then the algorithm continues back on line 2 and $\mathsf{T}^{y_i}_{\text{end}} = \emptyset$. In this case define recursively $\phi^i(\mathrm{o}) \coloneqq \psi(y_i)$ and $\phi^i(z) \coloneqq \frac{1}{d-1}\phi^i(\overline{z}) + Y^i_x$ for $z \in \mathbb{T}_d \setminus \{\mathrm{o}\}$. Then (3.8) holds for ϕ^i by (1.9)–(1.11). Moreover (3.7) is trivially satisfied since $\mathsf{T}^{y_i}_{\text{end}} = \emptyset$ (set $B^i \coloneqq \emptyset$). Otherwise we have $\psi(y_i) \ge h$ and y_i is added to T^{y_i}. If the algorithm terminates on line 8, then $\mathsf{T}^{y_i}_{\text{end}} = \{y_i\}$. In this case set $B^i \coloneqq \{y_i\}$ and $\tau^i(y_i) \coloneqq \mathrm{o} \in \mathbb{T}_d$ and again recursively define $\phi^i(\mathrm{o}) \coloneqq \psi(y_i)$ and $\phi^i(z) \coloneqq \frac{1}{d-1}\phi^i(\overline{z}) + Y^i_x$ for $z \in \mathbb{T}_d \setminus \{\mathrm{o}\}$. Then (3.8) holds for ϕ^i by (1.9)–(1.11) and also (3.7) is satisfied since $\left|\psi(y_i) - \phi^i(\tau^i(y_i))\right| = 0$. If the algorithm does not terminate on line 8, then on line 10 we now add all neither explored nor already queuing neighbours of y_i to PQ (which before that was empty). Consider the **while**-loop on line 11. During this **while**-loop, if z is taken out of PQ and $z \notin G_{\mathsf{E}}$, then it is transferred to SQ and it will not be part of $\mathsf{T}^{y_i}_{\text{end}}$. Let $z_1, \dots, z_m$ be the successive vertices taken out of PQ during the **while**-loop which are in G_{E} at the moment they are checked (on line 13). Possibly there are no such vertices, so we might have $\{z_1, \dots, z_m\} = \emptyset$ and $\mathsf{T}^{y_i}_{\text{end}} = \{y_i\}$. In any case, $\mathsf{T}^{y_i}_{\text{end}} = \{y_i\} \cup \{z_1, \dots, z_m \,|\, \psi(z_i) \ge h\} \subseteq \{y_i, z_1, \dots, z_m\} \subseteq B_{\mathcal{G}_n}(\mathsf{T}^{y_i}_{\text{end}}, 1)$. The injection τ^i we are going to construct now, will map $B^i \coloneqq \{y_i, z_1, \dots, z_m\}$ to $\mathbb{T}_d$. By definition one has $\overline{z}_1 = y_i$ whereas for $j = 2, \dots, m$ one has $\overline{z}_j = z$ for some $z \in \{y_i, z_1, \dots, z_{j-1}\}$. More precisely, $\overline{z}_j$ is the unique neighbour of z_j in E at the moment z_j was added to PQ (which happened on line 10 or line 22). There cannot be more than one since at a later point $z_j \in G_{\mathsf{E}}$ and the set of explored vertices only grows. Since there are at most $d-1$ not explored neighbours that can be added on line 10 or 22 (except if $i = 1$ when $y_1 = x$ and on line 10 there are added exactly d neighbours), this shows that for $z \in B^i$ there are at most $d-1$ elements $w \in \{z_1, \dots, z_m\}$ such that $\overline{w} = z$ (exactly d elements if $i = 1$ and $z = y_1$). Therefore, we can define $\tau^i : B^i \to \mathbb{T}_d$ inductively by $\tau^i(y_i) \coloneqq \mathrm{o}$ and such that τ^i restricted to $\{w \in \{z_1, \dots, z_m\} \,|\, \overline{w} = z\}$ is an injective map to $S_{\mathbb{T}_d}(\mathrm{o}, 1)$ for $z = y_i$ and an injective map to $S_{\mathbb{T}_d}(\tau^i(z), 1) \setminus \{\tau^i(\overline{z})\}$ for $z \in \{z_1, \dots, z_m\}$. Note that $\tau^i(B^i)$ is a connected subset of $\mathbb{T}_d$ containing the root $\mathrm{o} \in \mathbb{T}_d$. By construction also $\tau^i(\mathsf{T}^{y_i}_{\text{end}})$

is a connected subset of $\mathbb{T}_d$ containing $\mathrm{o} \in \mathbb{T}_d$ because $y_i \in \mathsf{T}_{\text{end}}^{y_i}$ with $\tau^i(y_i) = \mathrm{o}$ and $B^i \subseteq B_{\mathcal{G}_n}(\mathsf{T}_{\text{end}}^{y_i}, 1)$. We now define ϕ^i on $\tau^i(B^i) \subseteq \mathbb{T}_d$ and check the remaining properties in (3.7) and (3.8) for this case.

Set $\phi^i(\mathrm{o}) = \phi^i(\tau^i(y_i)) \coloneqq \psi(y_i)$ and for $j = 1, \dots, m$ define inductively $\phi^i(\tau^i(z_j)) \coloneqq \frac{1}{d-1}\phi^i(\tau^i(\overline{z}_j)) + \xi_{z_j} \cdot (\frac{d}{d-1})^{\frac{1}{2}}$. Recall that here ξ_{z_j} for $j = 1, \dots, m$ are the i.i.d. standard Gaussian random variables used to define $\psi(z_1), \dots, \psi(z_m)$ at the respective moments on line 16 of the algorithm. Note that conditionally on $\psi(y_i)$, the field $(\phi^i(\tau^i(z)))_{z \in B^i}$ has the same distribution as $(\varphi_{\mathbb{T}_d}(\tau^i(z)))_{z \in B^i}$ under $\mathbb{P}^{\mathbb{T}_d}_{\psi(y_i)}$. This follows by (1.9)–(1.11) since τ^i is defined in such a way that $\tau^i(\overline{z})$ (in the notation of (2.26)) for $z \in \{z_1, \dots, z_m\}$ is equal to $\overline{\tau^i(z)}$ (in the notation above (1.2)). We extend ϕ^i to all $w \in U \coloneqq \mathbb{T}_d \setminus B^i$ by recursively defining $\phi^i(w) \coloneqq \frac{1}{d-1}\phi^i(\overline{w}) + Y^i_w$. Then (3.8) holds for ϕ^i by (1.9)–(1.11). We proceed to show the remaining claim of (3.7). Note that $\big|\psi(y_i) - \phi^i(\tau^i(y_i))\big| = \big|\psi(y_i) - \phi^i(\mathrm{o})\big| = \big|\psi(y_i) - \psi(y_i)\big| = 0$ by definition. For $j = 1, \dots, m$ one has

$$\begin{aligned}\big|\psi(z_j) - \phi^i(\tau^i(z_j))\big| &= \big|a(z_j, \psi, \mathsf{E}) + \xi_{z_j} \cdot b(z_j, \mathsf{E})^{\frac{1}{2}} - \tfrac{1}{d-1}\phi^i(\tau^i(\overline{z}_j)) - \xi_{z_j} \cdot (\tfrac{d}{d-1})^{\frac{1}{2}}\big| \\ &\le \big|a(z_j, \psi, \mathsf{E}) - \tfrac{1}{d-1}\psi(\overline{z}_j)\big| + |\xi_{z_j}| \cdot \big|b(z_j, \mathsf{E})^{\frac{1}{2}} - (\tfrac{d}{d-1})^{\frac{1}{2}}\big| + \tfrac{1}{d-1}\big|\psi(\overline{z}_j) - \phi^i(\tau^i(\overline{z}_j))\big|.\end{aligned}$$

To the first two differences on the right hand side we can apply (2.30) and (2.31) (on the event $\{\sup_{z \in \mathsf{E}_{\text{end}}} |\psi(z)| < M_n\}$) since at any moment of the algorithm $|\mathsf{E}| \le |\mathsf{E}_{\text{end}}| \le c_{h,\kappa} \ln(N_n)$ by (3.4). We also use the inequality $|\sqrt{s} - \sqrt{t}| = \frac{|s-t|}{\sqrt{s}+\sqrt{t}} \le \frac{1}{\sqrt{t}}|s-t|$. So by Proposition 2.7 (for $b \coloneqq c_{h,\kappa}$ and $b' \coloneqq c_\kappa$) we find $c'_{h,\kappa} > 0$ such that for all $j = 1, \dots, m$

$$\big|\psi(z_j) - \phi^i(\tau^i(z_j))\big| \le (1 + |\xi_{z_j}|)c'_{h,\kappa}(\ln(N_n))^{-2} + \tfrac{1}{d-1}\big|\psi(\overline{z}_j) - \phi^i(\tau^i(\overline{z}_j))\big|. \tag{3.10}$$

Now note that on the event $\{\sup_{z \in \mathsf{E}_{\text{end}}} |\psi(z)| < M_n\}$ one has, again by using Proposition 2.7 for the same $b \coloneqq c_{h,\kappa}$ and $b' \coloneqq c_\kappa$, that for $j = 1, \dots, m$

$$\begin{aligned}M_n \;&>\; |\psi(z_j)| = \big|a(z_j, \psi, \mathsf{E}) + \xi_{z_j} \cdot b(z_j, \mathsf{E})^{\frac{1}{2}}\big| \ge |\xi_{z_j}| \cdot |b(z_j, \mathsf{E})^{\frac{1}{2}}| - |a(z_j, \psi, \mathsf{E})| \\ &\underset{(2.31)}{\overset{(2.30)}{\ge}} |\xi_{z_j}| \cdot \big(\tfrac{d}{d-1} - c'_{h,\kappa}(\ln(N_n))^{-3}\big)^{\frac{1}{2}} - \big(\tfrac{1}{d-1}\psi(\overline{z}_j) + c'_{h,\kappa}(\ln(N_n))^{-2}\big) \ge |\xi_{z_j}| - M_n,\end{aligned}$$

where the last inequality holds if n is large enough. Combine this with (3.10) to obtain that for n large enough, $j = 1, \dots, m$ and on the event $\{\sup_{z \in \mathsf{E}_{\text{end}}} |\psi(z)| < M_n\}$

$$\big|\psi(z_j) - \phi^i(\tau^i(z_j))\big| \le (1 + 2M_n)c'_{h,\kappa}(\ln(N_n))^{-2} + \tfrac{1}{d-1}\big|\psi(\overline{z}_j) - \phi^i(\tau^i(\overline{z}_j))\big|. \tag{3.11}$$

This is the main ingredient to show the remainder of (3.7). By induction we will now show that for n large enough and on the event $\{\sup_{z \in \mathsf{E}_{\text{end}}} |\psi(z)| < M_n\}$ one has

$$\big|\psi(z_j) - \phi^i(\tau^i(z_j))\big| \le j(1 + 2M_n)c'_{h,\kappa}(\ln(N_n))^{-2} \quad \text{for } j = 1, \dots, m. \tag{3.12}$$

For $j = 1$ one has $\overline{z}_1 = y_i$ and the last summand on the right hand side of (3.11) vanishes by definition of $\phi^i(\tau^i(y_i))$. Assume the statement holds for $1 \le j < m$. Then either $\overline{z}_{j+1} = y_i$ and therefore the last summand on the right hand side of (3.11) vanishes again, or $\overline{z}_{j+1} = z_k$ for some $k \in \{1, \dots, j\}$ and the induction hypothesis implies $\big|\psi(\overline{z}_{j+1}) - \phi^i(\tau^i(\overline{z}_{j+1}))\big| \le k(1 + 2M_n)c'_{h,\kappa}(\ln(N_n))^{-2} \le j(1 + 2M_n)c'_{h,\kappa}(\ln(N_n))^{-2}$. In any case by (3.11), $\big|\psi(z_{j+1}) - \phi^i(\tau^i(z_{j+1}))\big| \le (1 + 2M_n)c'_{h,\kappa}(\ln(N_n))^{-2} + \frac{1}{d-1}j(1 + 2M_n)c'_{h,\kappa}(\ln(N_n))^{-2} \le (j+1)(1 + 2M_n)c'_{h,\kappa}(\ln(N_n))^{-2}$. This concludes the induction.

Let $\varepsilon > 0$. Since $\{z_1, \dots, z_m\} \subseteq \mathsf{E}_{\text{end}}$, one has $m \le |\mathsf{E}_{\text{end}}| \le c_{h,\kappa} \ln(N_n)$ by (3.4). Moreover, $M_n = c_\kappa \sqrt{\ln(N_n)}$. So by (3.12) we obtain that for n large enough and on the event $\{\sup_{z \in \mathsf{E}_{\text{end}}} |\psi(z)| < M_n\}$ one has $\big|\psi(z_j) - \phi^i(\tau^i(z_j))\big| \le c_{h,\kappa} \ln(N_n)(1 + 2c_\kappa \sqrt{\ln(N_n)}) c'_{h,\kappa} (\ln(N_n))^{-2} \le \varepsilon$ for $j = 1, \dots, m$. We deduce (3.7).

It remains to show (3.8) for $i = k_{\text{end}}+1, \dots, k_{\max}$ and (3.9). For $i = k_{\text{end}}+1, \dots, k_{\max}$ define recursively $\phi^i(\mathrm{o}) := M_n$ and $\phi^i(z) := \frac{1}{d-1}\phi^i(\overline{z}) + Y_x^i$ for $z \in \mathbb{T}_d \setminus \{\mathrm{o}\}$, so that (3.8) holds for ϕ^i by (1.9)–(1.11). Finally, note that for each $i = 1, \dots, k_{\max}$ the field ϕ^i is constructed using $(Y_x^i)_{x \in \mathbb{T}_d \setminus \{\mathrm{o}\}}$ and possibly the i.i.d. random variables $(\xi_z)_{z \in B^i}$ and $\Psi(y_i)$. Since for $j = 1, \dots, k_{\max}$ with $j \ne i$ one has that $(Y_x^i)_{x \in \mathbb{T}_d \setminus \{\mathrm{o}\}}$ is independent of $(Y_x^j)_{x \in \mathbb{T}_d \setminus \{\mathrm{o}\}}$ and $B^i \cap B^j = \emptyset$, this shows (3.9) conditionally on $\Psi(y_1), \dots, \Psi(y_{k_{\text{end}}})$ (all random variables are Gaussian). The proof is complete. □

Corollary 3.5. *Let $h \in \mathbb{R}$ and $\varepsilon > 0$. Consider Algorithm 1 and recall $k_{\max}$ from Lemma 3.3. Then on the same auxiliary space $(\Omega, \mathcal{A}, \mathbb{P})$ as ψ one can define random variables $Z^1, \dots, Z^{k_{\max}}$ such that, conditionally on $\psi(y_1), \dots, \psi(y_{k_{\text{end}}})$, the following properties hold (see (1.11) and below (0.7) for notation):*

- *for all n large enough and all $i = 1, \dots, k_{\text{end}}$ one has $Z^i \ge |\mathsf{T}_{\text{end}}^{y_i}|$ on the event $\{\sup_{z \in \mathsf{E}_{\text{end}}} |\psi(z)| < M_n\}$* (3.13)
- *Z^i is distributed as $|\mathcal{C}_{\mathrm{o}}^{\mathbb{T}_d, h-\varepsilon}|$ under $\mathbb{P}_{\psi(y_i)}^{\mathbb{T}_d}$ for all $i = 1, \dots, k_{\text{end}}$,*
  Z^i is distributed as $|\mathcal{C}_{\mathrm{o}}^{\mathbb{T}_d, h-\varepsilon}|$ under $\mathbb{P}_{M_n}^{\mathbb{T}_d}$ for all $i = k_{\text{end}} + 1, \dots, k_{\max}$ (3.14)
- *$Z^1, \dots, Z^{k_{\max}}$ are independent.* (3.15)

Proof. We consider Lemma 3.4 and define Z^i as the size of the connected component of $\{w \in \mathbb{T}_d \,|\, \phi^i(w) \ge h - \varepsilon\}$ containing the root $\mathrm{o} \in \mathbb{T}_d$. Then (3.14) and (3.15) follow from (3.8) and (3.9). We turn to (3.13). Note that for $i = 1, \dots, k_{\text{end}}$ and $z \in \mathsf{T}_{\text{end}}^{y_i}$ one has $\big|\psi(z) - \phi^i(\tau^i(z))\big| \le \varepsilon$ by (3.7) under the assumptions of (3.13). This shows that $\phi^i(\tau^i(z)) \ge h - \varepsilon$ since $\psi(z) \ge h$ due to $z \in \mathsf{T}_{\text{end}}^{y_i}$. Hence $\tau^i(\mathsf{T}_{\text{end}}^{y_i}) \subseteq \{w \in \mathbb{T}_d \,|\, \phi^i(w) \ge h - \varepsilon\}$. As $\tau^i(\mathsf{T}_{\text{end}}^{y_i})$ is also a connected subset of $\mathbb{T}_d$ containing $\mathrm{o} \in \mathbb{T}_d$, we conclude $|\tau^i(\mathsf{T}_{\text{end}}^{y_i})| \le Z^i$ by definition of Z^i. The proof of (3.13) follows since τ^i is an injection and hence $|\tau^i(\mathsf{T}_{\text{end}}^{y_i})| = |\mathsf{T}_{\text{end}}^{y_i}|$. □

We are now ready to show Proposition 3.2, which as explained above its statement implies Theorem 3.1 and thereby concludes Section 3.

Proof of Proposition 3.2. Let $h > h_\star$ and $\kappa > 0$. Choose $\varepsilon > 0$ small enough such that $h - \varepsilon > h_\star$. Moreover, let $\delta_{h-\varepsilon} > 0$ be such that $g_{h-\varepsilon}$ defined in (1.14) has the properties explained therein. Let $K = K_{h,\kappa} > 0$ to be fixed later (below (3.22)). By conditioning on $\sigma(\psi(y_1), \dots, \psi(y_{k_{\text{end}}}))$ and then applying (3.13), one has for n large enough

$$
\begin{aligned}
&\mathbb{P}\Big[\sum_{i=1}^{k_{\text{end}}} |\mathsf{T}_{\text{end}}^{y_i}| \ge K \ln(N_n)\,,\ \sup_{z \in \mathsf{E}_{\text{end}}} |\psi(z)| < M_n\Big] \\
&\le \mathbb{E}\Bigg[\mathbb{P}\Big[\sum_{i=1}^{k_{\max}} Z^i \ge K \ln(N_n)\,,\ \sup_{z \in \mathsf{E}_{\text{end}}} |\psi(z)| < M_n \,\Big|\, \sigma(\psi(y_1), \dots, \psi(y_{k_{\text{end}}}))\Big]\Bigg] \qquad (3.16) \\
&\le \mathbb{E}\Bigg[\mathbf{1}_{\{|\psi(y_i)| < M_n \text{ for all } i = 1, \dots, k_{\text{end}}\}} \mathbb{P}\Big[\sum_{i=1}^{k_{\max}} Z^i \ge K \ln(N_n) \,\Big|\, \sigma(\psi(y_1), \dots, \psi(y_{k_{\text{end}}}))\Big]\Bigg].
\end{aligned}
$$

Since $\{\sum_{i=1}^{k_{\max}} Z^i \geq K\ln(N_n)\} = \{\prod_{i=1}^{k_{\max}}(1+\delta_{h-\varepsilon})^{Z^i} \geq (1+\delta_{h-\varepsilon})^{K\ln(N_n)}\}$, the conditional Markov inequality leads to $\mathbb{P}$-almost surely

$$\begin{aligned}
&\mathbb{P}\Big[\sum_{i=1}^{k_{\max}} Z^i \geq K\ln(N_n) \,\Big|\, \sigma(\psi(y_1),\dots,\psi(y_{k_{\mathrm{end}}}))\Big] \\
&\leq\; (1+\delta_{h-\varepsilon})^{-K\ln(N_n)}\mathbb{E}\Big[\prod_{i=1}^{k_{\max}}(1+\delta_{h-\varepsilon})^{Z^i} \,\Big|\, \sigma(\psi(y_1),\dots,\psi(y_{k_{\mathrm{end}}}))\Big] \\
&\overset{(3.15)}{\underset{(3.14)}{=}} (1+\delta_{h-\varepsilon})^{-K\ln(N_n)} \prod_{i=1}^{k_{\mathrm{end}}} \mathbb{E}^{\mathbb{T}_d}_{\psi(y_i)}\Big[(1+\delta_{h-\varepsilon})^{|\mathcal{C}_{\mathrm{o}}^{\mathbb{T}_d,h-\varepsilon}|}\Big] \cdot \prod_{i=k_{\mathrm{end}}+1}^{k_{\max}} \mathbb{E}^{\mathbb{T}_d}_{M_n}\Big[(1+\delta_{h-\varepsilon})^{|\mathcal{C}_{\mathrm{o}}^{\mathbb{T}_d,h-\varepsilon}|}\Big].
\end{aligned} \tag{3.17}$$

For $i=1,\dots,k_{\mathrm{end}}$ one has $\mathbb{E}^{\mathbb{T}_d}_{\psi(y_i)}\big[(1+\delta_{h-\varepsilon})^{|\mathcal{C}_{\mathrm{o}}^{\mathbb{T}_d,h-\varepsilon}|}\big] \leq \mathbb{E}^{\mathbb{T}_d}_{M_n}\big[(1+\delta_{h-\varepsilon})^{|\mathcal{C}_{\mathrm{o}}^{\mathbb{T}_d,h-\varepsilon}|}\big]$ on the event $\{\sup_{z\in\mathsf{E}_{\mathrm{end}}}|\psi(z)| < M_n\}$ by (1.9)–(1.11). This combined with (3.16) and (3.17) shows that for n large enough

$$\begin{aligned}
&\mathbb{P}\Big[\sum_{i=1}^{k_{\mathrm{end}}} |\mathsf{T}^{y_i}_{\mathrm{end}}| \geq K\ln(N_n)\,,\, \sup_{z\in\mathsf{E}_{\mathrm{end}}}|\psi(z)| < M_n\Big] \\
&\qquad\qquad \leq (1+\delta_{h-\varepsilon})^{-K\ln(N_n)}\mathbb{E}^{\mathbb{T}_d}_{M_n}\Big[(1+\delta_{h-\varepsilon})^{|\mathcal{C}_{\mathrm{o}}^{\mathbb{T}_d,h-\varepsilon}|}\Big]^{k_{\max}}.
\end{aligned} \tag{3.18}$$

Let us write $S_{\mathbb{T}_d}(\mathrm{o},1) =: \{x_1,\dots,x_d\}$ so that $\mathbb{T}_d = \{\mathrm{o}\}\cup\bigcup_{i=1}^d U_{x_i}$ (see (1.2)). Note that for n large enough one has $M_n \geq h$. Therefore [AČ19], equation (1.11), implies that

$$\begin{aligned}
\mathbb{E}^{\mathbb{T}_d}_{M_n}\Big[(1+\delta_{h-\varepsilon})^{|\mathcal{C}_{\mathrm{o}}^{\mathbb{T}_d,h-\varepsilon}|}\Big] &= (1+\delta_{h-\varepsilon})\,\mathbb{E}^{\mathbb{T}_d}_{M_n}\Big[\prod_{i=1}^{d}(1+\delta_{h-\varepsilon})^{|\mathcal{C}_{\mathrm{o}}^{\mathbb{T}_d,h-\varepsilon}\cap U_{x_i}|}\Big] \\
&= (1+\delta_{h-\varepsilon})\mathbb{E}^{Y}\Big[\mathbb{E}^{\mathbb{T}_d}_{\frac{M_n}{d-1}+Y}\Big[(1+\delta_{h-\varepsilon})^{|\mathcal{C}_{\mathrm{o}}^{\mathbb{T}_d,h-\varepsilon}\cap\mathbb{T}_d^+|}\Big]\Big]^d,
\end{aligned} \tag{3.19}$$

where $Y\sim\mathcal{N}(0,\frac{d}{d-1})$ and the expectation $\mathbb{E}^Y$ is taken with respect to Y. The inner expectation on the right hand side of (3.19) is equal to $g_{h-\varepsilon}(\frac{M_n}{d-1}+Y)$, see (1.14). Thus (3.19) shows that for n large enough

$$\mathbb{E}^{\mathbb{T}_d}_{M_n}\Big[(1+\delta_{h-\varepsilon})^{|\mathcal{C}_{\mathrm{o}}^{\mathbb{T}_d,h-\varepsilon}|}\Big] \leq \Big((1+\delta_{h-\varepsilon})\,\mathbb{E}^Y\Big[g_{h-\varepsilon}(\tfrac{M_n}{d-1}+Y)\Big]^{d-1}\Big)^{\frac{d}{d-1}}. \tag{3.20}$$

For n large enough (so $M_n \geq h$) one has that $(1+\delta_{h-\varepsilon})\,\mathbb{E}^Y\big[g_{h-\varepsilon}(\frac{M_n}{d-1}+Y)\big]^{d-1} = g_{h-\varepsilon}(M_n)$ by (1.14). Hence (3.20) and (3.18) imply that for n large enough

$$\mathbb{P}\Big[\sum_{i=1}^{k_{\mathrm{end}}} |\mathsf{T}^{y_i}_{\mathrm{end}}| \geq K\ln(N_n)\,,\, \sup_{z\in\mathsf{E}_{\mathrm{end}}}|\psi(z)| < M_n\Big] \leq (1+\delta_{h-\varepsilon})^{-K\ln(N_n)}\big(g_{h-\varepsilon}(M_n)\big)^{\frac{d}{d-1}k_{\max}}. \tag{3.21}$$

By (1.14) we know that there exist $c_h > 0$ and $c'_h > 0$ such that for n large enough one has $g_{h-\varepsilon}(M_n) \leq c_h\exp(c'_h M_n^{3/2}) = c_h\exp\big(c'_h(c_\kappa\sqrt{\ln(N_n)})^{3/2}\big) \leq c_h\exp(c_{h,\kappa}(\ln(N_n))^{3/4})$ for some $c_{h,\kappa} > 0$. Now recall that $k_{\max} = c_1 K s_n^2$. Therefore due to (2.25), we can find $c_h, c_{h,\kappa} > 0$ for which $\big(g_{h-\varepsilon}(M_n)\big)^{\frac{d}{d-1}k_{\max}} \leq c_h\exp(c_{h,\kappa}K(\ln(N_n))^{7/8})$. So for some $c_h > 0$

and $c_{h,\kappa} > 0$ we obtain $(1+\delta_{h-\varepsilon})^{-K\ln(N_n)}\big(g_{h-\varepsilon}(M_n)\big)^{\frac{d}{d-1}k_{\max}} \le c_h \exp\big(-c_{h,\kappa}K\ln(N_n)\big)$ for all n large enough. Hence by (3.21), for $n \ge 1$,

$$\mathbb{P}\Big[\sum_{i=1}^{k_{\text{end}}} |\mathsf{T}_{\text{end}}^{y_i}| \ge K\ln(N_n)\,,\ \sup_{z\in \mathsf{E}_{\text{end}}} |\psi(z)| < M_n\Big] \le c_{h,\kappa} N_n^{-c'_{h,\kappa}K}. \tag{3.22}$$

Take $K = K_{h,\kappa} > 0$ large enough such that $c_{h,\kappa}N_n^{-c'_{h,\kappa}K_{h,\kappa}} \le N_n^{-1-\kappa}$. Then by (3.22) we can find $c_{h,\kappa} > 0$ large enough such that (3.6) holds for all $n \ge 1$. This concludes the proof of Proposition 3.2 and ultimately of Theorem 3.1. □

4 Mesoscopic components in the supercritical phase

The last section of this article concerns the proof of (0.9) in the form of Theorem 4.1 below, that is, the existence of a supercritical phase (complementary to the subcritical situation in Section 3) in which the connected components of the levels sets of $\Psi_{\mathcal{G}_n}$ of at least mesoscopic size contain a non-negligible fraction of the vertices of $\mathcal{G}_n$. By mesoscopic size we mean that the number of vertices contained is a fractional power of the total number of vertices of $\mathcal{G}_n$. To be more specific, we recall the critical value $h_\star$ (see (0.7)) and the notation $\mathcal{C}_x^{\mathcal{G}_n,h}$ for the connected component of the level set of $\Psi_{\mathcal{G}_n}$ above level $h \in \mathbb{R}$ containing $x \in \mathcal{G}_n$ (see beginning of Section 3). Similarly, we denote by $\mathcal{C}_x^{\mathbb{T}_d,h}$ for $x \in \mathbb{T}_d$ and $h \in \mathbb{R}$ the connected component of the level set of $\varphi_{\mathbb{T}_d}$ above level h containing x. We also remind of the function η^+ given in (1.13). The main result of this section is the following

Theorem 4.1. *Let $h < h_\star$. Then there exist $c_h > 0$ (see beginning of the proof of Lemma 4.4) such that*

$$\lim_{n\to\infty} \mathbb{P}^{\mathcal{G}_n}\Big[\sum_{x\in\mathcal{G}_n} \mathbf{1}_{\{|\mathcal{C}_x^{\mathcal{G}_n,h}|\ge N_n^{c_h}\}} \ge \frac{\eta^+(h)}{2}\,N_n\Big] = 1. \tag{4.1}$$

As explained in the introduction below (0.9), it remains open whether in the supercritical phase $h < h_\star$, as the size of the graphs tends to infinity, one actually observes the emergence of a (unique) giant connected component of the level set above level h (see also Remark 4.7).

We now give the idea of the proof of Theorem 4.1. Roughly, the strategy is to control the expectation and variance of the sum in (4.1) and then to deduce Theorem 4.1 via a second moment inequality. Now recall that by (0.2) all vertices of $\mathcal{G}_n$ have an almost tree-like neighbourhood. One can also show that only a negligible fraction does not have an exactly tree-like neighbourhood of smaller size (Remark 4.3). So essentially we can consider only vertices with a tree-like neighbourhood in the sum in (4.1). Moreover, instead of counting the vertices $x \in \mathcal{G}_n$ with $\big|\mathcal{C}_x^{\mathcal{G}_n,h}\big| \ge N_n^\gamma$ for some fixed $\gamma > 0$ (i.e. contained in a mesoscopic connected component of $E_{\Psi_{\mathcal{G}_n}}^{\ge h}$), it will be easier to only consider the vertices $x \in \mathcal{G}_n$ for which the connected component $\mathcal{C}_x^{\mathcal{G}_n,h}$ is already mesoscopic when intersected with the tree-like neighbourhood of x (see (4.3)). We show that the expected number of such vertices grows linearly in the total number of vertices of the graph $\mathcal{G}_n$ as n tends to infinity (Lemma 4.4). A variance computation then implies that the number of vertices contained in mesoscopic components concentrates around its expectation as n goes to infinity (Lemma 4.6). The computations concerning the expectation and variance rely on the local approximation of $\Psi_{\mathcal{G}_n}$ by $\varphi_{\mathbb{T}_d}$ around vertices with tree-like neighbourhood that we developed in Section 2.1, which allows us to reduce the computations about $\Psi_{\mathcal{G}_n}$ to computations about $\varphi_{\mathbb{T}_d}$ and apply results from Section 1.1 on $\varphi_{\mathbb{T}_d}$. With a second moment inequality Theorem 4.1 promptly follows. The section ends with open questions in the supercritical regime $h < h_\star$ (Remark 4.7).

It will be convenient to introduce some additional notation. For $x \in \mathcal{G}_n$, $n \geq 1$ and $R \geq 0$ we set $S^+_{\mathcal{G}_n}(x,R) := \pi_{n,x}\big(S^+_{\mathbb{T}_d}(\text{o},R)\big)$ (see below (1.2) for the notation). We also define

$$r_n := \max\{1, \lfloor \tfrac{c_0}{18} \log_{d-1}(N_n)\rfloor\} \qquad \text{and} \qquad R_n := \max\{1, \lfloor \tfrac{c_0}{6} \log_{d-1}(N_n)\rfloor\}. \tag{4.2}$$

For $n \geq 1$, $h \in \mathbb{R}$ and $\gamma > 0$ we define the events

$$\begin{aligned} A_x^{\mathcal{G}_n,h,\gamma} &:= \big\{ \big|\mathcal{C}_x^{\mathcal{G}_n,h} \cap S^+_{\mathcal{G}_n}(x,r_n)\big| \geq N_n^\gamma \big\} \quad \text{for } x \in \mathcal{G}_n, \\ A_x^{\mathbb{T}_d,h,\gamma} &:= \big\{ \big|\mathcal{C}_x^{\mathbb{T}_d,h} \cap S^+_{\mathbb{T}_d}(x,r_n)\big| \geq N_n^\gamma \big\} \quad \text{for } x \in \mathbb{T}_d. \end{aligned} \tag{4.3}$$

Note that the dependency on n in the definition of $A_x^{\mathbb{T}_d,h,\gamma}$ in (4.3) does not appear in the notation. Finally, we define (with c_0 as in (1.21))

$$\gamma_h := \frac{c_0}{20} \log_{d-1}(\lambda_h) \quad \text{for } h \in \mathbb{R}. \tag{4.4}$$

By (1.12) note that γ_h is decreasing in h and $\gamma_h > 0$ for $h < h_\star$.

In the remainder of this section we will apply several times Theorem 2.1 for $r = r_n$ and $R = R_n$ given in (4.2). Note that, for n large enough, $1 \leq r_n < R_n \leq \frac{c_0}{6}\log_{d-1}(N_n)$ as required by Theorem 2.1 and furthermore $r_n \leq \frac{c_0}{18}\log_{d-1}(N_n)$ and $R_n - 2r_n \geq \big(\frac{c_0}{6}\log_{d-1}(N_n) - 1\big) - 2\frac{c_0}{18}\log_{d-1}(N_n) = \frac{c_0 \log_{d-1}(N_n)}{18} - 1$. Therefore Theorem 2.1 directly implies (with the notation from the beginning of Section 2.1)

Lemma 4.2 (Corollary of Theorem 2.1)**.** *There exist $c, c' > 0$ such that for all $n \geq 1$ and $x, x' \in \mathcal{G}_n$ with $\texttt{tx}(B_{\mathcal{G}_n}(x,2R_n)) = 0$, $\texttt{tx}(B_{\mathcal{G}_n}(x',2R_n)) = 0$ and $B_{\mathcal{G}_n}(x,2R_n) \cap B_{\mathcal{G}_n}(x',2R_n) = \emptyset$, there is a coupling $\mathbb{Q}_n$ of $\Psi_{\mathcal{G}_n}$ and $\varphi_{\mathbb{T}_d}$ satisfying for all $\varepsilon > 0$*

$$\mathbb{Q}_n\Big[\sup_{y \in B_{\mathcal{G}_n}(x,r_n) \cup B_{\mathcal{G}_n}(x',r_n)} \big|\Psi_{\mathcal{G}_n}(y) - \varphi_{\mathbb{T}_d}(\rho_{x,x',2R_n}(y))\big| > \varepsilon\Big] \leq c \exp\big(-c'\varepsilon^2 N_n^{\frac{c_0}{18}}\big). \tag{4.5}$$

In particular, there exist $c, c' > 0$ such that for all $n \geq 1$, $x \in \mathcal{G}_n$ with $\texttt{tx}(B_{\mathcal{G}_n}(x,2R_n)) = 0$, there is a coupling $\mathbb{Q}_n$ of $\Psi_{\mathcal{G}_n}$ and $\varphi_{\mathbb{T}_d}$ such that for all $\varepsilon > 0$ the same bound as in (4.5) applies to $\mathbb{Q}_n\big[\sup_{y \in B_{\mathcal{G}_n}(x,r_n)} \big|\Psi_{\mathcal{G}_n}(y) - \varphi_{\mathbb{T}_d}(\rho_{x,2R_n}(y))\big| > \varepsilon\big]$.

As the following remark explains, the assumptions on the vertices in the statement of Lemma 4.2 are typical.

Remark 4.3. Recall R_n from (4.2). For n large enough the number of vertices $x \in \mathcal{G}_n$ that do not satisfy $\texttt{tx}(B_{\mathcal{G}_n}(x,2R_n)) = 0$ is negligible when compared to the total number of vertices of $\mathcal{G}_n$. Indeed, for n large enough one has $2R_n \leq \lfloor \alpha \log_{d-1}(N_n)\rfloor$ by (1.21) and (1.1) and thus $\texttt{tx}(B_{\mathcal{G}_n}(x,2R_n)) \leq 1$ for all $x \in \mathcal{G}_n$ by assumption (0.2). Now by [ČTW11], Lemma 6.1, we have for n large enough

$$\big|\{x \in \mathcal{G}_n \,|\, \texttt{tx}(B_{\mathcal{G}_n}(x,2R_n)) = 1\}\big| \leq (d-1)^{-(\lfloor \alpha \log_{d-1}(N_n)\rfloor - 2R_n)} N_n \overset{(*)}{\leq} (d-1) N_n^{1-\frac{2\alpha}{3}}, \tag{4.6}$$

where in $(*)$ we use that $\lfloor \alpha \log_{d-1}(N_n) \rfloor - 2R_n \geq \alpha \log_{d-1}(N_n) - 1 - \frac{c_0}{3}\log_{d-1}(N_n) \geq \frac{2\alpha}{3}\log_{d-1}(N_n) - 1$ (because $c_0 \leq \alpha$ by (1.21) and (1.1)). Moreover, for n large enough, also the number of pairs of vertices $x, x' \in \mathcal{G}_n$ for which $B_{\mathcal{G}_n}(x, 2R_n) \cap B_{\mathcal{G}_n}(x', 2R_n) \neq \emptyset$ is negligible when compared to the total number N_n^2 of pairs of vertices of $\mathcal{G}_n$. Indeed, for n large enough and for such $x, x' \in \mathcal{G}_n$ one has $x' \in B_{\mathcal{G}_n}(x, 4R_n)$ and hence

$$\begin{aligned}
&\left|\{x, x' \in \mathcal{G}_n \,|\, B_{\mathcal{G}_n}(x, 2R_n) \cap B_{\mathcal{G}_n}(x', 2R_n) \neq \emptyset\}\right| \leq \sum_{x \in \mathcal{G}_n} |B_{\mathcal{G}_n}(x, 4R_n)| \\
&\overset{(0.1)}{\leq} N_n |B_{\mathbb{T}_d}(\mathrm{o}, 4R_n)| = N_n \frac{d(d-1)^{4R_n} - 2}{d-2} \overset{(4.2)}{\leq} N_n\, d(d-1)^{\frac{2c_0}{3}\log_{d-1}(N_n)} \leq dN_n^{\frac{5}{3}},
\end{aligned} \tag{4.7}$$

where the last inequality follows because $c_0 \leq 1$ (see (1.21)). □

We are now ready to proceed with the expectation and variance computation announced after the statement of Theorem 4.1.

Lemma 4.4. *Let $h < h_\star$. There exists $c_h > 0$ such that for all $0 < \varepsilon < \frac{h_\star - h}{2}$ and $\zeta > 0$ one has for n large enough*

$$\mathbb{P}^{\mathcal{G}_n}\left[A_x^{\mathcal{G}_n, h, c_h}\right] \geq \eta^+(h+\varepsilon) - \zeta \quad \text{for } x \in \mathcal{G}_n \text{ with } \mathtt{tx}(B_{\mathcal{G}_n}(x, 2R_n)) = 0. \tag{4.8}$$

As a consequence, one has

$$\liminf_{n\to\infty} \frac{1}{N_n} \mathbb{E}^{\mathcal{G}_n}\left[\sum_{x \in \mathcal{G}_n} \mathbf{1}_{A_x^{\mathcal{G}_n, h, c_h}}\right] \geq \eta^+(h) > 0. \tag{4.9}$$

Proof. Let $h < h_\star$ and take $\delta := \frac{h_\star - h}{2} > 0$ so that $h + \delta < h_\star$. Let $\varepsilon < \delta$. Set also $c_h := \gamma_{h+\delta}$. For $x \in \mathcal{G}_n$ as in the statement of (4.8) we can apply Lemma 4.2 and obtain that for $n \geq 1$ one has (recall that $\rho_{x,2R_n}$ is a graph isomorphism from $B_{\mathcal{G}_n}(x, 2R_n)$ to $B_{\mathbb{T}_d}(\mathrm{o}, 2R_n)$, see beginning of Section 2.1)

$$\begin{aligned}
\mathbb{P}^{\mathcal{G}_n}\left[A_x^{\mathcal{G}_n, h, c_h}\right] &\geq \mathbb{Q}_n\Big[A_x^{\mathcal{G}_n, h, \gamma_{h+\delta}}, \sup_{y \in B_{\mathcal{G}_n}(x, r_n)} \left|\Psi_{\mathcal{G}_n}(y) - \varphi_{\mathbb{T}_d}(\rho_{x,2R_n}(y))\right| \leq \varepsilon\Big] \\
&\overset{(4.3)}{\geq} \mathbb{Q}_n\Big[A_{\mathrm{o}}^{\mathbb{T}_d, h+\varepsilon, \gamma_{h+\delta}}, \sup_{y \in B_{\mathcal{G}_n}(x, r_n)} \left|\Psi_{\mathcal{G}_n}(y) - \varphi_{\mathbb{T}_d}(\rho_{x,2R_n}(y))\right| \leq \varepsilon\Big] \\
&\geq \mathbb{P}^{\mathbb{T}_d}\left[A_{\mathrm{o}}^{\mathbb{T}_d, h+\varepsilon, \gamma_{h+\delta}}\right] - c\exp\left(-c'\varepsilon^2 N_n^{\frac{c_0}{18}}\right).
\end{aligned} \tag{4.10}$$

Note that since $0 < \varepsilon < \delta$ one has $\gamma_{h+\delta} < \gamma_{h+\varepsilon}$ and hence

$$\begin{aligned}
&\liminf_{n\to\infty} \mathbb{P}^{\mathbb{T}_d}\left[A_{\mathrm{o}}^{\mathbb{T}_d, h+\varepsilon, \gamma_{h+\delta}}\right] \geq \liminf_{n\to\infty} \mathbb{P}^{\mathbb{T}_d}\left[A_{\mathrm{o}}^{\mathbb{T}_d, h+\varepsilon, \gamma_{h+\varepsilon}}\right] \\
&\overset{(4.3)}{\underset{(4.4)}{=}} \liminf_{n\to\infty} \mathbb{P}^{\mathbb{T}_d}\Big[\left|\mathcal{C}_{\mathrm{o}}^{\mathbb{T}_d, h+\varepsilon} \cap S_{\mathbb{T}_d}^+(\mathrm{o}, r_n)\right| \geq \lambda_{h+\varepsilon}^{\frac{c_0}{20}\log_{d-1}(N_n)}\Big] \\
&\overset{(*)}{\geq} \liminf_{n\to\infty} \mathbb{P}^{\mathbb{T}_d}\Big[\left|\mathcal{C}_{\mathrm{o}}^{\mathbb{T}_d, h+\varepsilon} \cap S_{\mathbb{T}_d}^+(\mathrm{o}, r_n)\right| \geq \frac{\lambda_{h+\varepsilon}^{r_n}}{r_n^2}\Big] \overset{(1.15)}{=} \eta^+(h+\varepsilon),
\end{aligned} \tag{4.11}$$

where in $(*)$ we use that $\lambda_{h+\varepsilon} > 1$ (see (1.12)) and the definition of r_n (see (4.2)). By combining (4.10) and (4.11) we find (4.8). For (4.9) we only need to notice that, for n large enough, $\left|\{x \in \mathcal{G}_n \,|\, \mathtt{tx}(B_{\mathcal{G}_n}(x, 2R_n)) = 0\}\right| \geq N_n - (d-1)N_n^{1-\frac{2\alpha}{3}}$ by (4.6). So (4.9) follows from (4.8) by summing only over $x \in \mathcal{G}_n$ with $\mathtt{tx}(B_{\mathcal{G}_n}(x, 2R_n)) = 0$ and applying (1.13). □

As a next step we want to show that $\sum_{x\in\mathcal{G}_n} \mathbf{1}_{A_x^{\mathcal{G}_n,h,c_h}}$ for $h<h_\star$ and the $c_h>0$ from Lemma 4.4 concentrates around its expectation. A variance computation will be enough. The main ingredient is contained in the next lemma.

Lemma 4.5. *Let $h<h_\star$. There exist $c,c'>0$ such that for all $n\geq 1$, $x,x'\in\mathcal{G}_n$ with $\mathtt{tx}(B_{\mathcal{G}_n}(x,2R_n))=0$, $\mathtt{tx}(B_{\mathcal{G}_n}(x',2R_n))=0$ and $B_{\mathcal{G}_n}(x,2R_n)\cap B_{\mathcal{G}_n}(x',2R_n)=\emptyset$ one has for all $\gamma>0$ and $\varepsilon>0$*

$$\mathbb{P}^{\mathcal{G}_n}\big[A_x^{\mathcal{G}_n,h,\gamma}\,,\,A_{x'}^{\mathcal{G}_n,h,\gamma}\big]\;\leq\;\mathbb{P}^{\mathbb{T}_d}\big[A_{\mathrm{o}}^{\mathbb{T}_d,h-\varepsilon,\gamma}\big]^2+c\exp\big(-c'\varepsilon^2 N_n^{\frac{c_0}{18}}\big). \tag{4.12}$$

Proof. Let us abbreviate $V:=B_{\mathcal{G}_n}(x,r_n)\cup B_{\mathcal{G}_n}(x',r_n)$. For $x,x'\in\mathcal{G}_n$ as in the assumptions we can apply Lemma 4.2 and obtain that for all $n\geq 1$, $\gamma>0$ and $\varepsilon>0$ one has (recall the notation $\rho_{x,x',2R_n}$ and $z_{x,x'}$ from the beginning of Section 2.1)

$$\begin{aligned}
&\mathbb{P}^{\mathcal{G}_n}\big[A_x^{\mathcal{G}_n,h,\gamma}\,,\,A_{x'}^{\mathcal{G}_n,h,\gamma}\big]\\
&\overset{(4.5)}{\leq}\;\mathbb{Q}_n\Big[A_x^{\mathcal{G}_n,h,\gamma},A_{x'}^{\mathcal{G}_n,h,\gamma},\sup_{y\in V}\big|\Psi_{\mathcal{G}_n}(y)-\varphi_{\mathbb{T}_d}(\rho_{x,x',2R_n}(y))\big|\leq\frac{\varepsilon}{2}\Big]+c\exp\big(-c'\varepsilon^2N_n^{\frac{c_0}{18}}\big)\\
&\leq\;\mathbb{P}^{\mathbb{T}_d}\big[A_{\mathrm{o}}^{\mathbb{T}_d,h-\frac{\varepsilon}{2},\gamma}\,,\,A_{z_{x,x'}}^{\mathbb{T}_d,h-\frac{\varepsilon}{2},\gamma}\big]+c\exp\big(-c'\varepsilon^2N_n^{\frac{c_0}{18}}\big).
\end{aligned}\tag{4.13}$$

To further bound the probability on the right hand side of (4.13) we apply the decoupling inequality [PR15], Corollary 1.3, with

$$\begin{aligned}
&\delta:=\frac{\varepsilon}{2},\ K_1:=B_{\mathbb{T}_d}(\mathrm{o},r_n),\ K_2:=B_{\mathbb{T}_d}(z_{x,x'},r_n)\text{ and }f_1,f_2:\mathbb{R}^{\mathbb{T}_d}\to[0,1]\text{ such that}\\
&f_1\big((\varphi_{\mathbb{T}_d}(x))_{x\in\mathbb{T}_d}\big)=\mathbf{1}_{A_{\mathrm{o}}^{\mathbb{T}_d,h-\frac{\varepsilon}{2},\gamma}}\text{ and }f_2\big((\varphi_{\mathbb{T}_d}(x))_{x\in\mathbb{T}_d}\big)=\mathbf{1}_{A_{z_{x,x'}}^{\mathbb{T}_d,h-\frac{\varepsilon}{2},\gamma}}
\end{aligned}$$

(the decoupling inequality [PR15], Corollary 1.3, is stated for the Gaussian free field on $\mathbb{Z}^d$ but its proof directly applies also for the Gaussian free field $\varphi_{\mathbb{T}_d}$ on $\mathbb{T}_d$). We obtain that for all $n\geq 1$, $\gamma>0$ and $\varepsilon>0$

$$\begin{aligned}
&\mathbb{P}^{\mathbb{T}_d}\big[A_{\mathrm{o}}^{\mathbb{T}_d,h-\frac{\varepsilon}{2},\gamma}\,,\,A_{z_{x,x'}}^{\mathbb{T}_d,h-\frac{\varepsilon}{2},\gamma}\big]\\
&\leq\mathbb{P}^{\mathbb{T}_d}\big[A_{\mathrm{o}}^{\mathbb{T}_d,h-\frac{\varepsilon}{2},\gamma}\big]\mathbb{P}^{\mathbb{T}_d}\big[A_{z_{x,x'}}^{\mathbb{T}_d,h-\varepsilon,\gamma}\big]+2\,\mathbb{P}^{\mathbb{T}_d}\Big[\sup_{y\in K_2}\Big|E_y^{\mathbb{T}_d}\big[\varphi_{\mathbb{T}_d}(X_{H_{K_1}})\mathbf{1}_{\{H_{K_1}<\infty\}}\big]\Big|>\frac{\varepsilon}{4}\Big].
\end{aligned}\tag{4.14}$$

Note that, since we are on a tree and K_1 and K_2 are two disjoint connected sets, there is a unique pair of vertices $z_1\in K_1$, $z_2\in K_2$ with $d_{\mathbb{T}_d}(K_1,K_2):=\inf_{z\in K_1,z'\in K_2}d_{\mathbb{T}_d}(z,z')=d_{\mathbb{T}_d}(z_1,z_2)$. Moreover, on the event $\{H_{K_1}<\infty\}$ one $P_y^{\mathbb{T}_d}$-almost surely has $\varphi_{\mathbb{T}_d}(X_{H_{K_1}})=\varphi_{\mathbb{T}_d}(z_1)$ for $y\in K_2$. Therefore,

$$\begin{aligned}
&\mathbb{P}^{\mathbb{T}_d}\Big[\sup_{y\in K_2}\Big|E_y^{\mathbb{T}_d}\big[\varphi_{\mathbb{T}_d}(X_{H_{K_1}})\mathbf{1}_{\{H_{K_1}<\infty\}}\big]\Big|>\frac{\varepsilon}{4}\Big]=\mathbb{P}^{\mathbb{T}_d}\Big[\sup_{y\in K_2}\Big|P_y^{\mathbb{T}_d}\big[H_{z_1}<\infty\big]\varphi_{\mathbb{T}_d}(z_1)\Big|>\frac{\varepsilon}{4}\Big]\\
&\leq\;\mathbb{P}^{\mathbb{T}_d}\Big[|\varphi_{\mathbb{T}_d}(z_1)|>\frac{\varepsilon}{4}P_{z_2}^{\mathbb{T}_d}\big[H_{z_1}<\infty\big]^{-1}\Big]\overset{(*)}{\leq}2\exp\bigg(-\frac{(\varepsilon/4)^2}{2P_{z_2}^{\mathbb{T}_d}\big[H_{z_1}<\infty\big]^2g_{\mathbb{T}_d}(\mathrm{o},\mathrm{o})}\bigg)\\
&\underset{(1.6)}{\overset{(**)}{\leq}}\;2\exp\Big(-c\,\varepsilon^2(d-1)^{2d_{\mathbb{T}_d}(z_1,z_2)}\Big),
\end{aligned}\tag{4.15}$$

where in $(*)$ we use the exponential Markov inequality for the centred Gaussian random variable $\varphi_{\mathbb{T}_d}(z_1)$ and in $(**)$ we use that $P_{z_2}^{\mathbb{T}_d}\big[H_{z_1}<\infty\big]=(\frac{1}{d-1})^{d_{\mathbb{T}_d}(z_1,z_2)}$ (see e.g. [Woe00], proof of Lemma 1.24). Since $K_1\subseteq B_{\mathbb{T}_d}(\mathrm{o},2R_n)$, $K_2\subseteq B_{\mathbb{T}_d}(z_{x,x'},2R_n)$ and $B_{\mathbb{T}_d}(\mathrm{o},2R_n)\cap$

$B_{\mathbb{T}_d}(z_{x,x'}, 2R_n) = \emptyset$ by assumption, one has the estimate $d_{\mathbb{T}_d}(z_1, z_2) = d_{\mathbb{T}_d}(K_1, K_2) > 2(2R_n - r_n) \geq \frac{5c_0}{9}\log_{d-1}(N_n) - 4$ for n large enough. Hence $\exp\big(-c\,\varepsilon^2(d-1)^{2d_{\mathbb{T}_d}(z_1,z_2)}\big) \leq c\exp\big(-c'\varepsilon^2 N_n^{\frac{10c_0}{9}}\big)$. Therefore we can combine (4.13), (4.14) and (4.15) to obtain that for all $n \geq 1$, $\gamma > 0$ and $\varepsilon > 0$ (using also the symmetry of $\mathbb{T}_d$)

$$\mathbb{P}^{\mathcal{G}_n}\big[A_x^{\mathcal{G}_n,h,\gamma},\, A_{x'}^{\mathcal{G}_n,h,\gamma}\big] \leq \mathbb{P}^{\mathbb{T}_d}\big[A_{\mathrm{o}}^{\mathbb{T}_d,h-\frac{\varepsilon}{2},\gamma}\big]\mathbb{P}^{\mathbb{T}_d}\big[A_{\mathrm{o}}^{\mathbb{T}_d,h-\varepsilon,\gamma}\big] + c\exp\big(-c'\varepsilon^2 N_n^{\frac{c_0}{18}}\big).$$

This concludes the proof of (4.12) since by (4.3) it holds $A_{\mathrm{o}}^{\mathbb{T}_d,h-\frac{\varepsilon}{2},\gamma} \subseteq A_{\mathrm{o}}^{\mathbb{T}_d,h-\varepsilon,\gamma}$. □

We are now ready to for the variance computation. This is the last ingredient for the proof of Theorem 4.1.

Lemma 4.6. *Let $h < h_\star$. Then for the $c_h > 0$ from Lemma 4.4 one has*

$$\lim_{n\to\infty} \frac{1}{N_n^2}\mathrm{Var}_{\mathbb{P}^{\mathcal{G}_n}}\Big(\sum_{x\in\mathcal{G}_n} \mathbf{1}_{A_x^{\mathcal{G}_n,h,c_h}}\Big) = 0. \tag{4.16}$$

Proof. By expanding the variance one finds that for all $\gamma > 0$

$$\begin{aligned}\mathrm{Var}_{\mathbb{P}^{\mathcal{G}_n}}\Big(\sum_{x\in\mathcal{G}_n} \mathbf{1}_{A_x^{\mathcal{G}_n,h,\gamma}}\Big) &= \sum_{x,x'\in\mathcal{G}_n}\Big(\mathbb{P}^{\mathcal{G}_n}\big[A_x^{\mathcal{G}_n,h,\gamma}, A_{x'}^{\mathcal{G}_n,h,\gamma}\big] - \mathbb{P}^{\mathcal{G}_n}\big[A_x^{\mathcal{G}_n,h,\gamma}\big]\mathbb{P}^{\mathcal{G}_n}\big[A_{x'}^{\mathcal{G}_n,h,\gamma}\big]\Big)\\ &= \sum_{x,x'\in\mathcal{G}_n}\mathbb{P}^{\mathcal{G}_n}\big[A_x^{\mathcal{G}_n,h,\gamma}, A_{x'}^{\mathcal{G}_n,h,\gamma}\big] - \mathbb{E}^{\mathcal{G}_n}\Big[\sum_{x\in\mathcal{G}_n}\mathbf{1}_{A_x^{\mathcal{G}_n,h,\gamma}}\Big]^2. \end{aligned}\tag{4.17}$$

We define $W \subseteq \mathcal{G}_n \times \mathcal{G}_n$ to be the set of pairs $(x, x') \in \mathcal{G}_n \times \mathcal{G}_n$ with $\mathtt{tx}(B_{\mathcal{G}_n}(x, 2R_n)) = 0$, $\mathtt{tx}(B_{\mathcal{G}_n}(x', 2R_n)) = 0$ and $B_{\mathcal{G}_n}(x, 2R_n) \cap B_{\mathcal{G}_n}(x', 2R_n) = \emptyset$. For $x, x' \in \mathcal{G}_n$ such that $(x, x') \notin W$ we can bound the probability on the right hand side of (4.17) by one. This will be good enough since for n large enough $|(\mathcal{G}_n\times\mathcal{G}_n)\setminus W| \leq 2N_n\cdot(d-1)N_n^{1-\frac{2\alpha}{3}} + dN_n^{\frac{5}{3}} \leq dN_n\big(2N_n^{1-\frac{2\alpha}{3}} + N_n^{\frac{2}{3}}\big)$ by (4.6) and (4.7). For $x, x' \in \mathcal{G}_n$ such that $(x, x') \in W$ we use (4.12) instead. There are at most N_n^2 such pairs. Thus we obtain for all $n \geq 1$, $\gamma > 0$ and $\varepsilon > 0$

$$\begin{aligned}&\frac{1}{N_n^2}\mathrm{Var}_{\mathbb{P}^{\mathcal{G}_n}}\Big(\sum_{x\in\mathcal{G}_n}\mathbf{1}_{A_x^{\mathcal{G}_n,h,\gamma}}\Big)\\ &\leq \mathbb{P}^{\mathbb{T}_d}\big[A_{\mathrm{o}}^{\mathbb{T}_d,h-\varepsilon,\gamma}\big]^2 + c\exp\big(-c'\varepsilon^2 N_n^{\frac{c_0}{18}}\big) + d\big(2N_n^{-\frac{2\alpha}{3}} + N_n^{-\frac{1}{3}}\big) - \frac{1}{N_n^2}\mathbb{E}^{\mathcal{G}_n}\Big[\sum_{x\in\mathcal{G}_n}\mathbf{1}_{A_x^{\mathcal{G}_n,h,\gamma}}\Big]^2.\end{aligned}\tag{4.18}$$

Now we apply (4.18) to $\gamma := c_h > 0$ for the c_h from Lemma 4.4 and deduce that for all $0 < \varepsilon < \frac{h_\star - h}{2}$

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{N_n^2}\mathrm{Var}_{\mathbb{P}^{\mathcal{G}_n}}\Big(\sum_{x\in\mathcal{G}_n}\mathbf{1}_{A_x^{\mathcal{G}_n,h,c_h}}\Big)\\ &\overset{(4.18)}{\leq} \limsup_{n\to\infty}\mathbb{P}^{\mathbb{T}_d}\big[A_{\mathrm{o}}^{\mathbb{T}_d,h-\varepsilon,c_h}\big]^2 - \liminf_{n\to\infty}\frac{1}{N_n^2}\mathbb{E}^{\mathcal{G}_n}\Big[\sum_{x\in\mathcal{G}_n}\mathbf{1}_{A_x^{\mathcal{G}_n,h,c_h}}\Big]^2\\ &\underset{(4.9)}{\overset{(1.13)}{\leq}} \eta^+(h-\varepsilon)^2 - \eta^+(h)^2.\end{aligned}$$

The statement follows by letting ε tend to zero and applying (1.13). □

Proof of Theorem 4.1. We will show that the probability of the complementary event tends to zero. For $n \geq 1$ let us define $W_n^{\geq h} := \sum_{x\in\mathcal{G}_n} \mathbf{1}_{A_x^{\mathcal{G}_n,h,c_h}}$ with $c_h > 0$ as in Lemma 4.4. Then we can estimate

$$\mathbb{P}^{\mathcal{G}_n}\Big[\sum_{x\in\mathcal{G}_n} \mathbf{1}_{\{|\mathcal{C}_x^{\mathcal{G}_n,h}|\geq N_n^{c_h}\}} < \frac{\eta^+(h)}{2} N_n\Big] \leq \mathbb{P}^{\mathcal{G}_n}\Big[W_n^{\geq h} < \frac{\eta^+(h)}{2} N_n\Big]$$
$$= \mathbb{P}^{\mathcal{G}_n}\Big[\frac{1}{N_n}\mathbb{E}^{\mathcal{G}_n}[W_n^{\geq h}] - \frac{1}{N_n}W_n^{\geq h} > \frac{1}{N_n}\mathbb{E}^{\mathcal{G}_n}[W_n^{\geq h}] - \frac{\eta^+(h)}{2}\Big]$$

and therefore

$$\limsup_{n\to\infty} \mathbb{P}^{\mathcal{G}_n}\Big[\sum_{x\in\mathcal{G}_n} \mathbf{1}_{\{|\mathcal{C}_x^{\mathcal{G}_n,h}|\geq N_n^{c_h}\}} < \frac{\eta^+(h)}{2} N_n\Big]$$
$$\overset{(4.9)}{\leq} \limsup_{n\to\infty} \mathbb{P}^{\mathcal{G}_n}\Big[\frac{1}{N_n}\mathbb{E}^{\mathcal{G}_n}[W_n^{\geq h}] - \frac{1}{N_n}W_n^{\geq h} > \frac{\eta^+(h)}{2}\Big]$$
$$\overset{(*)}{\leq} \limsup_{n\to\infty} \frac{4}{\eta^+(h)^2}\mathrm{Var}_{\mathbb{P}^{\mathcal{G}_n}}\Big(\frac{1}{N_n}W_n^{\geq h}\Big) \overset{(4.16)}{=} 0,$$

where in $(*)$ we use Chebyshev's inequality. This concludes the proof of Theorem 4.1. □

Remark 4.7. It remains open whether in the supercritical phase $h < h_\star$, with high probability for large n, there actually is a macroscopic (giant) connected component of the level set above level h (i.e. containing a number of vertices comparable to $\mathcal{G}_n$), and whether this giant component is unique (meaning the size of the second-largest connected component is negligible compared to $\mathcal{G}_n$). For other probabilistic models on essentially the same class of graphs this has been shown. One example is the emergence of a unique giant connected component for Bernoulli bond percolation on d-regular expanders of large girth (see [ABS04] and also [KLS18]). A second example is the emergence of a unique giant connected component in the vacant set of simple random walk on the same graphs $(\mathcal{G}_n)_{n\geq 1}$ as considered here (see [ČTW11]). As briefly mentioned in the introduction below (0.9), such results are typically obtained by a sprinkling argument out of an intermediary result like Theorem 4.1. In our setting, the zero-average property of $\Psi_{\mathcal{G}_n}$ (see below (1.18)) prevents us from easily implementing such a strategy. In particular, due to the zero-average property, the field $\Psi_{\mathcal{G}_n}$ neither satisfies an FKG-inequality nor does it possess the domain Markov property of the Gaussian free field $\varphi_{\mathbb{T}_d}$ (compare (1.20) with (1.8)). In contrast, the sprinkling argument in [DR15] for constructing an infinite connected component for the Gaussian free field on $\mathbb{Z}^d$ for high-dimension d crucially relies on the domain Markov property of the Gaussian free field on $\mathbb{Z}^d$ for $d \geq 3$. □

References

[Abä19] Angelo Abächerli. Local picture and level-set percolation of the Gaussian free field on a large discrete torus. *Stochastic Process. Appl.*, 129(9):3527–3546, 2019.

[ABS04] Noga Alon, Itai Benjamini, and Alan Stacey. Percolation on finite graphs and isoperimetric inequalities. *Ann. Probab.*, 32(3):1727–1745, 2004.

[AČ19] Angelo Abächerli and Jiří Černý. Level-set percolation of the Gaussian free field on regular graphs I: regular trees. Preprint, available at *arXiv:1909.01973*, 2019.

[AS18] Angelo Abächerli and Alain-Sol Sznitman. Level-set percolation for the Gaussian free field on a transient tree. *Ann. Inst. H. Poincaré Probab. Statist.*, 54(1):173–201, 2018.

[BLM87] Jean Bricmont, Joel L. Lebowitz, and Christian Maes. Percolation in strongly correlated systems: the massless Gaussian field. *J. Statist. Phys.*, 48(5-6):1249–1268, 1987.

[CF13] Colin Cooper and Alan Frieze. Component structure of the vacant set induced by a random walk on a random graph. *Random Structures Algorithms*, 42(2):135–158, 2013.

[ČT13] Jiří Černý and Augusto Teixeira. Critical window for the vacant set left by random walk on random regular graphs. *Random Structures Algorithms*, 43(3):313–337, 2013.

[ČTW11] Jiří Černý, Augusto Teixeira, and David Windisch. Giant vacant component left by a random walk in a random d-regular graph. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(4):929–968, 2011.

[DPR18a] Alexander Drewitz, Alexis Prévost, and Pierre-François Rodriguez. Geometry of Gaussian free field sign clusters and random interlacements. Preprint, available at *arXiv:1811.05970*, 2018.

[DPR18b] Alexander Drewitz, Alexis Prévost, and Pierre-François Rodriguez. The sign clusters of the massless Gaussian free field percolate on $\mathbb{Z}^d$, $d \geq 3$ (and more). *Comm. Math. Phys.*, 362(2):513–546, 2018.

[DR15] Alexander Drewitz and Pierre-François Rodriguez. High-dimensional asymptotics for percolation of Gaussian free field level sets. *Electron. J. Probab.*, 20:1–39, 2015.

[Fel68] William Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons, Inc., New York-London-Sydney, 1968.

[KLS18] Michael Krivelevich, Eyal Lubetzky, and Benny Sudakov. Asymptotics in bond percolation on expanders. Preprint, available at *arXiv:1803.11553v1*, 2018.

[LS86] Joel L. Lebowitz and H. Saleur. Percolation in strongly correlated systems. *Phys. A*, 138(1-2):194–205, 1986.

[LS10] Eyal Lubetzky and Allan Sly. Cutoff phenomena for random walks on random regular graphs. *Duke Math. J.*, 153(3):475–510, 2010.

[MS83] S. A. Molchanov and A. K. Stepanov. Percolation in random fields. I. *Teoret. Mat. Fiz.*, 55(2):246–256, 1983.

[PR15] Serguei Popov and Balázs Ráth. On decoupling inequalities and percolation of excursion sets of the Gaussian free field. *J. Stat. Phys.*, 159(2):312–320, 2015.

[RS13] Pierre-François Rodriguez and Alain-Sol Sznitman. Phase transition and level-set percolation for the Gaussian free field. *Comm. Math. Phys.*, 320(2):571–601, 2013.

[SC97] Laurent Saloff-Coste. Lectures on finite Markov chains. In *Lectures on probability theory and statistics (Saint-Flour, 1996)*, volume 1665 of *Lecture Notes in Math.*, pages 301–413. Springer, Berlin, 1997.

[Szn15] Alain-Sol Sznitman. Disconnection and level-set percolation for the Gaussian free field. *J. Math. Soc. Japan*, 67(4):1801–1843, 2015.

[Szn16] Alain-Sol Sznitman. Coupling and an application to level-set percolation of the Gaussian free field. *Electron. J. Probab.*, 21:1–26, 2016.

[Szn19] Alain-Sol Sznitman. On coupling and "vacant set level set" percolation. *Electron. Commun. Probab.*, 24:1–12, 2019.

[Wil91] David Williams. *Probability with martingales.* Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.

[Woe00] Wolfgang Woess. *Random walks on infinite graphs and groups*, volume 138 of *Cambridge Tracts in Mathematics.* Cambridge University Press, Cambridge, 2000.