The Annals of Applied Probability
2017, Vol. 27, No. 3, 1588–1645
DOI: 10.1214/16-AAP1239

MINIMAL SPANNING TREES AND STEIN'S METHOD

BY SOURAV CHATTERJEE[1] AND SANCHAYAN SEN[2]

Stanford University and McGill University

Kesten and Lee [*Ann. Appl. Probab.* **6** (1996) 495–527] proved that the total length of a minimal spanning tree on certain random point configurations in $\mathbb{R}^d$ satisfies a central limit theorem. They also raised the question: how to make these results quantitative? Error estimates in central limit theorems satisfied by many other standard functionals studied in geometric probability are known, but techniques employed to tackle the problem for those functionals do not apply directly to the minimal spanning tree. Thus, the problem of determining the convergence rate in the central limit theorem for Euclidean minimal spanning trees has remained open. In this work, we establish bounds on the convergence rate for the Poissonized version of this problem by using a variation of Stein's method. We also derive bounds on the convergence rate for the analogous problem in the setup of the lattice $\mathbb{Z}^d$.

The contribution of this paper is twofold. First, we develop a general technique to compute convergence rates in central limit theorems satisfied by minimal spanning trees on sequences of weighted graphs, including minimal spanning trees on Poisson points inside a sequence of growing cubes. Second, we present a way of quantifying the Burton–Keane argument for the uniqueness of the infinite open cluster. The latter is interesting in its own right and based on a generalization of our technique, Duminil-Copin, Ioffe and Velenik [*Ann. Probab.* **44** (2016) 3335–3356] have recently obtained bounds on probability of two-arm events in a broad class of translation-invariant percolation models.

CONTENTS

Received March 2015; revised May 2016.
[1]Supported in part by NSF Grant DMS-10-05312.
[2]Supported in part by NSF Grant DMS-10-07524 and Netherlands Organisation for Scientific Research (NWO) through the Gravitation Networks Grant 024.002.003.
MSC2010 subject classifications. 60D05, 60F05, 60B10.
Key words and phrases. Minimal spanning tree, central limit theorem, Stein's method, Burton–Keane argument, two-arm event.

1. Introduction. Consider a finite, connected weighted graph (V, E, w) where (V, E) is the underlying graph and $w : E \to [0, \infty)$ is the weight function. A spanning tree of (V, E) is a tree which is a connected subgraph of (V, E) with vertex set V. A minimal spanning tree (MST) T of (V, E, w) satisfies

$$\sum_{e \in T} w(e) = \min\left\{ \sum_{e \in T'} w(e) : T' \text{ is a spanning tree of } (V, E) \right\}.$$

In this paper, whenever (V, E) is a graph on some random point configuration in $\mathbb{R}^d$, the weight function will map every edge to its Euclidean length.

Minimal spanning trees and other related functionals are of great interest in geometric probability. For an account of law of large numbers and related asymptotics for these functionals; see, for example, [5, 6, 10, 17, 55, 57]. One of the early successes in the direction of proving distributional convergence of such functionals came with the paper of Avram and Bertsimas [11] in 1993 where the authors proved central limit theorems (CLT) for three such functionals, namely the lengths of the kth nearest neighbor graph, the Delaunay triangulation and the Voronoi diagram on Poisson point configurations in $[0, 1]^2$. Central limit theorems for minimal spanning trees were first proven by Kesten and Lee [36] and by Alexander [8] in 1996. This was a long-standing open question at the time of its solution. In [36], the CLT for the total weight of an MST on both the complete graph on Poisson points inside $[0, n^{1/d}]^d$ and the complete graph on n i.i.d. uniformly distributed points inside $[0, 1]^d$ were established when $d \geq 2$. (Their results included the case of more general weight functions and not just Euclidean distances.) Alexander [8] proved the CLT for the Poissonized problem in two dimensions. Later certain other CLTs related to MSTs were proven in [40] and [41].

Studies related to Euclidean MSTs in several other directions were undertaken in [12, 18, 44, 45, 48]. An account of the structural properties of minimal spanning forests (in both Euclidean and non-Euclidean setting) can be found in [7, 9, 34, 42] and the references therein. For an account of the scaling limit of minimal spanning trees see, for example, [2, 22, 51].

Minimal spanning trees on the complete graph and on the hypercube have been studied extensively as well and we refer the reader to [4, 29, 35, 46, 56] for such results. In the recent preprint [1], existence of a scaling limit of the minimal spanning tree on the complete graph viewed as a metric space has been established. Our primary focus in this paper, however, will be on minimal spanning trees on Poisson points and subsets of $\mathbb{Z}^d$.

The methods of [8] and [36] cannot be used to get bounds on the rate of convergence to normality in the CLT for Euclidean MSTs. Indeed, Kesten and Lee remark that

> *"... [A] drawback of our approach is that it is not quantitative. Further ideas are needed to obtain an error estimate in our central limit theorem."*

A general method for tackling such a problem is to show that the function of interest satisfies certain "stabilizing" properties [50]. In [49] (see also [40]), it was shown that Euclidean MSTs do satisfy a stabilizing property but there was no quantitative bound on how fast this stabilization occurs. Quoting Penrose and Yukich [50],

> *"Some functionals, such as those defined in terms of the minimal spanning tree, satisfy a weaker form of stabilization but are not known to satisfy exponential stabilization. In these cases univariate and multivariate central limit theorems hold... but our [main theorem] does not apply and explicit rates of convergence are not known."*

This poses the major difficulty in obtaining an error estimate in the CLT and the problem has remained open since the work of Kesten and Lee.

In this paper, we use a variation of Stein's method, given by approximation theorems from [24, 38], to connect the problem of bounding the convergence rate in this CLT to the problem of getting upper bounds on the probabilities of certain events in the setup of continuum percolation driven by a Poisson process, and thus obtaining an error estimate in this CLT (Theorem 2.1). Using a similar approach, we also obtain error estimates in the CLT for the total weights of the MSTs on subgraphs of $\mathbb{Z}^d$ under various assumptions on the edge weights (Theorem 2.4). In Theorem 2.6, we present a general CLT satisfied by the MSTs on subgraphs of a vertex-transitive graph. The percolation theoretic estimates used in the proofs are given in Section 5. Our techniques for proving these percolation theoretic estimates are of independent interest.

This paper is organized as follows. In Section 2, we state our results about convergence rates in CLTs satisfied by MSTs. In Section 3, we give a brief survey of literature on Stein's method and state the theorems used for Gaussian approximation. In Section 4, we introduce the necessary notation. In Section 5, we state the percolation theoretic estimates we will be using. In Section 7, we briefly discuss the idea in the proof and how to connect the problem of getting convergence rates in the CLT to a problem in percolation. Section 8 lists some properties and preliminary results about minimal spanning trees. Sections 9–13 are devoted to proofs of the central limit theorems and the percolation theoretic estimates.

2. Main results. We summarize our main results in this section. Define the distance $\mathcal{D}(\mu_1, \mu_2)$ between two probability measures μ_1 and μ_2 on $\mathbb{R}$ by the sup norm of the difference between their distribution functions, or equivalently

$$\mathcal{D}(\mu_1, \mu_2) := \sup_{x\in\mathbb{R}} \bigl|\mu_1(-\infty, x] - \mu_2(-\infty, x]\bigr|. \tag{2.1}$$

This metric is sometimes called the "Kolmogorov distance." A bound on the Kolmogorov distance between two probability measures is sometimes called a "Berry–Esseen bound."

Recall also that the Kantorovich–Wasserstein distance between two probability measures μ_1 and μ_2 on $\mathbb{R}$ is given by

$$\begin{aligned} &\mathcal{W}(\mu_1, \mu_2) \\ &\quad := \sup\left\{\left|\int f\, d\mu_1 - \int f\, d\mu_2\right| : f \text{ Lipschitz with } \|f\|_{\mathrm{Lip}} \le 1\right\}. \end{aligned} \tag{2.2}$$

Convergence in this metric implies weak convergence.

Our result on Euclidean minimal spanning trees is the following.

THEOREM 2.1. *Let $\mathcal{P}$ be a Poisson process with intensity one in $\mathbb{R}^d$. Let (V_n, E_n, w_n) be the complete graph on $\mathcal{P} \cap [-n, n]^d$ with each edge weighted by its Euclidean length. Let μ_n be the law of $(M_n - \mathbb{E}(M_n))/\sqrt{\mathrm{Var}(M_n)}$, where M_n is the total weight of an MST of (V_n, E_n, w_n). Let γ denote the standard normal distribution on $\mathbb{R}$.*

(i) *When $d = 2$, there exist positive constants ξ and c_1 such that, for every $n \ge 1$,*

$$\max\bigl\{\mathcal{W}(\mu_n, \gamma), \mathcal{D}(\mu_n, \gamma)\bigr\} \le c_1 n^{-\xi}. \tag{2.3}$$

(ii) *When $d \ge 3$, for every $p > 1$ and every $n \ge 2$,*

$$\max\bigl\{\mathcal{W}(\mu_n, \gamma), \mathcal{D}(\mu_n, \gamma)\bigr\} \le c_2 (\log n)^{-\frac{d}{4p}} \tag{2.4}$$

for a positive constant c_2 depending only on p and d.

REMARK 2.2. If $\mathcal{P}_\lambda$ is a Poisson process with intensity $\lambda > 0$ in $\mathbb{R}^d$ and $M_n(\lambda)$ is the weight of a minimal spanning tree of the complete graph on $\mathcal{P}_\lambda \cap [-n/\lambda^{\frac{1}{d}}, n/\lambda^{\frac{1}{d}}]^d$, then $(M_n(\lambda) - \mathbb{E}M_n(\lambda))/\sqrt{\operatorname{Var}(M_n(\lambda))}$ is distributed as μ_n where μ_n is as defined in the statement of Theorem 2.1. For this reason, it is enough to consider only Poisson processes with intensity one.

Our next theorem deals with the case of minimal spanning trees on subsets of $\mathbb{Z}^d$. To state the theorem conveniently, we first make a definition. In what follows, $p_c = p_c(\mathbb{Z}^d)$ denotes the critical probability of bond percolation in $\mathbb{Z}^d$ (see, e.g., [19, 33]).

DEFINITION 2.3. A probability measure μ on $[0, \infty)$ satisfies:

(A) *Property* A_δ (for some $\delta > 0$) if μ has unbounded support and $\int_0^\infty x^{4+\delta}\mu(dx) < \infty$;
(B) *Property* B if μ has bounded support;
(C) *Property* C if either $\mu[0, x] = p_c(\mathbb{Z}^d)$ for some unique $x \in \mathbb{R}$, or $\mu[0, x) = p_c(\mathbb{Z}^d)$ for some unique $x \in \mathbb{R}$;
(D) *Property* D if $\mu[0, x] > p_c(\mathbb{Z}^d) > \mu[0, x)$ for some $x \in \mathbb{R}$.

THEOREM 2.4. *Let $d \geq 2$ and assume that the edges of the lattice $\mathbb{Z}^d$ have been given i.i.d. nonnegative weights having some nondegenerate distribution μ. Let M_n denote the total weight of an MST of the weighted subgraph of $\mathbb{Z}^d$ within the cube $[-n, n]^d$, and let ν_n be the distribution of $(M_n - \mathbb{E}(M_n))/\sqrt{\operatorname{Var}(M_n)}$. Let γ be the standard normal distribution on $\mathbb{R}$.*

(i) *If μ satisfies either Property B or Property A_δ for some $\delta > 0$, then, for every $n \geq 2$,*

$$\mathcal{W}(\nu_n, \gamma) \leq \varepsilon_n (\log n)^{\frac{1}{4(1+3\xi)}} / n^{\frac{1}{6(1+2\xi)}}, \tag{2.5}$$

where

$$\xi = \begin{cases} 1/\delta, & \text{if } \mu \text{ satisfies Property } A_\delta, \\ 0, & \text{if } \mu \text{ satisfies Property } B, \end{cases}$$

and $\varepsilon_n \to 0$ if μ satisfies Property C, and is a bounded sequence otherwise.

If μ satisfies either Property B or Property A_δ for some $\delta \geq 2$, then (2.5) *holds if we replace $\mathcal{W}(\nu_n, \gamma)$ by $\mathcal{D}(\nu_n, \gamma)$.*

(ii) *If μ satisfies Property D and either Property B or Property A_δ for some $\delta > 0$, then for every $\eta < d/2$,*

$$\mathcal{W}(\nu_n, \gamma) \leq c_3 n^{-\eta} \qquad \text{for } n \geq 1, \tag{2.6}$$

where c_3 is a positive constant depending on μ, d and η.

If μ satisfies Property D and either Property B or Property A_δ for some $\delta \geq 2$, then (2.6) *holds if we replace $\mathcal{W}(\nu_n, \gamma)$ by $\mathcal{D}(\nu_n, \gamma)$.*

REMARK 2.5. It is very likely that the bounds are suboptimal. However, the question of optimal error bounds is probably very difficult. Improving the bounds stated in Theorems 2.1 and 2.4 can be thought of as an independent problem in percolation (see Remark 5.5).

Our approach can be used to give a simple proof of asymptotic normality of the total weight of the minimal spanning tree under a very general assumption on the underlying graph. We present this result in the following theorem. *The advantage of this approach is that we can get a convergence rate in the central limit theorem whenever we can prove the percolation theoretic estimates analogous to the ones used in the proofs of Theorems 2.1 and 2.4.*

Before stating the theorem, let us recall the definition of a vertex-transitive graph. A graph $G = (V, E)$ is said to be vertex-transitive if for any $v_1, v_2 \in V$, there exists a graph automorphism f of G such that $f(v_1) = v_2$.

For a graph $G = (V, E)$ and a vertex $v \in V$, we will write $S_G(v, r)$ to denote the subgraph of G spanned by the set of all vertices $v' \in V$ such that $d_G(v', v) \leq r$ where d_G denotes the graph distance of G.

THEOREM 2.6. *Let $G = (V, E)$ be a*:

(I) *connected*, *infinite*, *locally finite*, *vertex-transitive graph*.

Consider a sequence of finite connected subgraphs $G_n = (V_n, E_n)$ such that

(II) $|V_n| \to \infty$, *and*

(III) $|\{v \in V_n : S_G(v, r) \not\subset G_n\}| = o(|V_n|)$ *for every* $r > 0$.

Consider i.i.d. nonnegative weights associated with the edges of G where the weights follow some non-degenerate distribution μ that satisfies either Property B or Property A_δ for some $\delta > 0$. Let M_n be the total weight of a minimal spanning tree of G_n. Then:

(i) $\mathrm{Var}(M_n) = \Theta(|V_n|)$ *and*

(ii) $(M_n - \mathbb{E}(M_n))/\sqrt{\mathrm{Var}(M_n)} \xrightarrow{d} Z$, *where Z follows a $N(0, 1)$ distribution.*

REMARK 2.7. Note that G in Theorem 2.6 is necessarily amenable [because of Conditions (II) and (III)].

3. Stein's method. In 1972, Charles Stein [58] proposed a radically different approach to proving convergence to normality. Stein's observation was that the standard normal distribution is the only probability distribution that satisfies the equation

$$\mathbb{E}\big(Zf(Z)\big) = \mathbb{E}f'(Z)$$

for all absolutely continuous f with a.e. derivative f' such that $\mathbb{E}|f'(Z)| < \infty$. From this, one might expect that if W is a random variable that satisfies the above equation in an approximate sense, then the distribution of W should be close to the standard normal distribution. The key to Stein's implementation of his idea was the method of exchangeable pairs, devised by Stein in [58]. A notable success story of Stein's method was authored by Bolthausen [20] in 1984 when he used a sophisticated version of the method of exchangeable pairs to obtain an error bound in a famous combinatorial central limit theorem of Hoeffding. Stein's 1986 monograph [59] was the first book-length treatment of Stein's method. After the publication of [59], the field was given a boost by the popularization of the method of dependency graphs by Baldi and Rinott [13], a striking application to the number of local maxima of random functions by Baldi, Rinott and Stein [14], and central limit theorems for random graphs by Barbour, Karoński and Ruciński [16], all in 1989.

The new surge of activity that began in the late 1980s continued through the nineties, with important contributions coming from Barbour [15] in 1990, who introduced the diffusion approach to Stein's method; Avram and Bertsimas [11] in 1993, who applied Stein's method to solve an array of important problems in geometric probability; Goldstein and Rinott [32] in 1996, who developed the method of size-biased couplings for Stein's method, improving on earlier insights of Baldi, Rinott and Stein [14]; Goldstein and Reinert [31] in 1997, who introduced the method of zero-bias couplings; and Rinott and Rotar [52] in 1997, who solved a well-known open problem related to the antivoter model using Stein's method. Sometime later, in 2004, Chen and Shao [27] did an in-depth study of the dependency graph approach, producing optimal Berry–Esséen type error bounds in a wide range of problems. The 2003 monograph of Penrose [47] gave extensive applications of the dependency graph approach to problems in geometric probability.

A new version of Stein's method with potentially wider applicability was introduced for discrete systems [24], and a corresponding continuous version in [25]. This new approach was used to solve a number of questions in geometric probability in [24], random matrix central limit theorems in [25] and number theoretic central limit theorems in [26]. The main result of [24] gives convergence rates in terms of the Kantorovich–Wasserstein distance. Very recently, this approach has been generalized in [38], Theorem 4.2, to give convergence rates in the Kolmogorov distance. These two results are our main tools for normal approximation.

As mentioned before in Section 1, MSTs on Poisson points exhibit a stabilization property; but no tail bound on the radius of stabilization (in the sense of [49]) is known. If such a tail bound were known, then there would be a number of ways of obtaining a convergence rate in the CLT satisfied by MSTs on Poisson points (e.g., using the results of [24] or [39] or [50]). However, [24], Theorem 2.2, and [38], Theorem 4.2, allow us to circumvent this problem and instead reduce the problem to finding upper bounds on probability of two-arm events. We will state these theorems in the following section.

3.1. *Main approximation theorems.* To state the theorems, we need some notation; we will use them repeatedly in this paper.

Let $\mathcal{X}$ be a Polish space. For every $A \subset [n] := \{1, \dots, n\}$, define the "replacement" operator $\mathcal{R}^A : \mathcal{X}^n \times \mathcal{X}^n \to \mathcal{X}^n$ as follows: for $y = (y_1, \dots, y_n) \in \mathcal{X}^n$, and $y' = (y'_1, \dots, y'_n) \in \mathcal{X}^n$, the ith component of $\mathcal{R}^A(y, y')$ is given by

$$(\mathcal{R}^A(y, y'))_i = \begin{cases} y'_i, & \text{if } i \in A, \\ y_i, & \text{if } i \notin A. \end{cases}$$

Suppose $f : \mathcal{X}^n \to \mathbb{R}$ is a measurable function. For $j \in [n]$, define $\Delta_j f : \mathcal{X}^n \times \mathcal{X}^n \to \mathbb{R}$ by

$$\Delta_j f(y, y') := f(y) - f(\mathcal{R}^{\{j\}}(y, y')).$$

Let $X_1, \dots, X_n$ be independent $\mathcal{X}$ valued random variables and set $X = (X_1, \dots, X_n)$. Let $X' = (X'_1, \dots, X'_n)$ be an independent copy of X. To simplify notation, we will write X^A to denote the random vector $\mathcal{R}^A(X, X')$. We will simply write X^j instead of $X^{\{j\}}$. With this convention, for every $A \subset [n]$,

$$\Delta_j f(X^A, X') = f(X^A) - f(X^{A \cup \{j\}}).$$

For every $A \subset [n]$, let

$$T_A := \sum_{j \notin A} \Delta_j f(X, X') \Delta_j f(X^A, X') \quad \text{and}$$
$$T'_A := \sum_{j \notin A} \Delta_j f(X, X') |\Delta_j f(X^A, X')|.$$

Finally, define

$$T = \frac{1}{2} \sum_{A \subsetneq [n]} \frac{T_A}{\binom{n}{|A|}(n - |A|)} \quad \text{and} \quad T' = \frac{1}{2} \sum_{A \subsetneq [n]} \frac{T'_A}{\binom{n}{|A|}(n - |A|)}.$$

Recall the definitions of the Kantorovich–Wasserstein distance [see (2.2)] and the Kolmogorov distance [see (2.1)].

THEOREM 3.1 ([24], Theorem 2.2). *Let all terms be defined as above and let $W = f(X)$ with $\sigma^2 := \mathrm{Var}(W) < \infty$. Then $\mathbb{E}T = \sigma^2$ and*

$$\mathcal{W}(\mu, \gamma) \le \frac{1}{\sigma^2} [\mathrm{Var}(\mathbb{E}(T | W))]^{1/2} + \frac{1}{2\sigma^3} \sum_{j=1}^{n} \mathbb{E}|\Delta_j f(X, X')|^3, \tag{3.1}$$

where μ is the law of $(W - \mathbb{E}W)/\sigma$.

THEOREM 3.2 ([38], Theorem 4.2). *Let all terms be defined as above and let $W = f(X)$ with $\sigma^2 := \mathrm{Var}(W) < \infty$. Then*

$$\begin{aligned}\mathcal{D}(\mu,\gamma) \le{}& \frac{1}{\sigma^2}\big[\mathrm{Var}\big(\mathbb{E}(T|X)\big)\big]^{1/2} + \frac{1}{\sigma^2}\big[\mathrm{Var}\big(\mathbb{E}(T'|X)\big)\big]^{1/2} \\ &+ \frac{1}{4\sigma^3}\sum_{j=1}^{n}\big(\mathbb{E}\big|\Delta_j f(X,X')\big|^6\big)^{1/2} + \frac{\sqrt{2\pi}}{16\sigma^3}\sum_{j=1}^{n}\mathbb{E}\big|\Delta_j f(X,X')\big|^3,\end{aligned} \tag{3.2}$$

where μ is the law of $(W-\mathbb{E}W)/\sigma$.

Note that

$$\mathrm{Var}\big(\mathbb{E}(T|W)\big) \le \mathrm{Var}(T) \quad \text{and} \quad \mathrm{Var}\big(\mathbb{E}(T|X)\big) \le \mathrm{Var}(T)$$

and

$$\begin{aligned}\mathrm{Var}(T) &= \frac{1}{4}\mathrm{Var}\Bigg[\sum_{A\subsetneq[n]}\sum_{j\in[n]\setminus A}\frac{\Delta_j f(X)\Delta_j f(X^A)}{\binom{n}{|A|}(n-|A|)}\Bigg] \\ &= \frac{1}{4}\sum_{\substack{A\subsetneq[n]\\ j\in[n]\setminus A}}\sum_{\substack{A'\subsetneq[n]\\ j'\in[n]\setminus A'}}\frac{\mathrm{Cov}(\Delta_j f(X)\Delta_j f(X^A), \Delta_{j'} f(X)\Delta_{j'} f(X^{A'}))}{\binom{n}{|A|}(n-|A|)\binom{n}{|A'|}(n-|A'|)}.\end{aligned} \tag{3.3}$$

We will make repeated use of this identity.

The expression of the upper bound in Theorem 3.2 is very similar to the bound in Theorem 3.1. We will give detailed proofs of bounds in the Kantorovich–Wasserstein distance using Theorem 3.1, and then briefly sketch how to adapt the proof using Theorem 3.2 to get a bound of the same order in the Kolmogorov distance.

4. Notation. We will use some notation frequently throughout this paper. For convenience, we collect them together in this section.

4.1. *Euclidean setup.* If x is a point in $\mathbb{R}^d$ and $A \subset \mathbb{R}^d$, then we define $x + A := \{x+y : y \in A\}$. If $r > 0$, $S_{\mathbb{R}^d}(x,r)$ will denote the closed L^2 ball of radius r centered at x, and $B_{\mathbb{R}^d}(x,r)$ will denote the closed L^∞ ball of radius r centered at x, that is, $B_{\mathbb{R}^d}(x,r) = x + [-r,r]^d$. When x is the origin, we will simply write $B_{\mathbb{R}^d}(r)$ instead of $B_{\mathbb{R}^d}(0,r)$. For any cube B, we refer to its center as $c(B)$. We will denote by $d_{\mathbb{R}^d}(\cdot,\cdot)$, the metric induced by the L^2 norm in $\mathbb{R}^d$. When the underlying space is clear from the context, we will drop the subscript $\mathbb{R}^d$ and simply write $S(\cdot,\cdot)$, $B(\cdot,\cdot)$, and $d(\cdot,\cdot)$.

For a finite subset X of $\mathbb{R}^d$, $M_{\mathbb{R}^d}(X)$ will denote the sum of edge weights of the minimal spanning tree on the complete graph on X having Euclidean distance as edge weights. When the ambient space is clear, we will drop the subscript and simply write $M(X)$.

For $A \subset \mathbb{R}^d$ and $r > 0$, we define

$$A^{(r)} := \{x \in \mathbb{R}^d : d_{\mathbb{R}^d}(x, A) \le r\}.$$

[With this notation $S_{\mathbb{R}^d}(x, r) = \{x\}^{(r)}$.] Let us also define

$$A_{(r)} := \{x \in A : d_{\mathbb{R}^d}(x, \partial A) \le r\}.$$

Let $\mathcal{P}$ be a Poisson process in $\mathbb{R}^d$ and let A be a subset of $\mathbb{R}^d$. Then $\mathcal{C} \subset \mathcal{P} \cap A$ will be called an r-cluster in A (or just r-cluster if A is clear) if $\mathcal{C}^{(r)}$ is a connected component of $(\mathcal{P} \cap A)^{(r)}$; $\mathcal{C}^{(r)}$ should be thought of as the region occupied by the cluster $\mathcal{C}$. We say that two r-clusters $\mathcal{C}_1$ and $\mathcal{C}_2$ in $A \subset \mathbb{R}^d$ are disjoint if $\mathcal{C}_1^{(r)}$ and $\mathcal{C}_2^{(r)}$ are. We emphasize that the occupied regions must be disjoint in $\mathbb{R}^d$, and it is not enough to have their restrictions to A to be disjoint. We will write *configuration* to mean a locally finite subset of $\mathbb{R}^d$. For $A \subset \mathbb{R}^d$, $\mathfrak{X}(A)$ will denote the space of all locally finite subsets of A.

For two compact sets $K_1, K_2 \subset \mathbb{R}^d$ with $K_1 \subset K_2$, a positive integer k and a positive real r, we write $K_1 \overset{k}{\underset{r}{\longleftrightarrow}} K_2$ if there exists a collection of k disjoint r-clusters $\mathcal{C}_1, \dots, \mathcal{C}_k$ in $K_2 \setminus K_1$ such that

$$\mathcal{C}_j \cap K_1^{(r)} \ne \varnothing \quad \text{and} \quad \mathcal{C}_j \cap (K_2)_{(2r)} \ne \varnothing \qquad \text{for } j = 1, \dots, k.$$

For $x \in \mathbb{R}^d$ and $b > a > 0$, we call $\{B(x, a) \overset{2}{\underset{r}{\longleftrightarrow}} B(x, b)\}$ a two-arm event at level r.

We will write $K_1 \xrightarrow[r]{k} K_2$, if there exists a collection of k pairwise disjoint r-clusters $\mathcal{C}_1, \dots, \mathcal{C}_k$ in $(K_2 \setminus K_1)$ such that

$$\mathcal{C}_j \cap K_1^{(2r)} \ne \varnothing \quad \text{and} \quad \mathcal{C}_j \cap (K_2)_{(2r)} \ne \varnothing \qquad \text{for } j = 1, \dots, k. \tag{4.1}$$

4.2. *Discrete setup*. Consider a graph $G = (V, E)$. Recall from Section 2 that $d_G(\cdot, \cdot)$ denotes the graph distance on G, and

$$S_G(v, r) := \{v' \in V : d_G(v', v) \le r\}.$$

Assume that each $e \in E$ has a nonnegative weight x_e attached to it. Let $\boldsymbol{x} = (x_e : e \in E)$. Then for any finite connected subgraph $H = (V_1, E_1)$ of G, $M_G(H, \boldsymbol{x})$ will denote the total weight of an MST on the weighted graph H, where $e_1 \in E_1$ has weight x_{e_1}. When the underlying graph G is clear, we will drop the subscripts and simply write $d(\cdot, \cdot)$, $S(v, r)$, and $M(H, \boldsymbol{x})$.

For any $e \in E$, $G - e$ will denote the graph $(V, E - e)$. If $G_i = (V_i, E_i)$, $i = 1, 2$ are two subgraphs of G, then $G_1 \cap G_2$ will denote the subgraph $(V_1 \cap V_2, E_1 \cap E_2)$.

When working with the lattice $\mathbb{Z}^d$, $B_{\mathbb{Z}^d}(x, r)$ will denote the set of all lattice points inside $x + [-r, r]^d$ and $B_{\mathbb{Z}^d}(r)$ will stand for $B_{\mathbb{Z}^d}(0, r)$. We will simply write $B(x, r)$ and $B(r)$ when the ambient space is clear from the context.

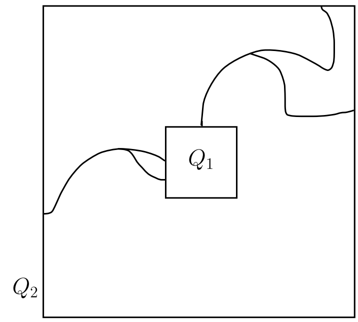

FIG. 1. $Q_1 \overset{2}{\underset{p}{\leftrightsquigarrow}} Q_2$.

For a subset V of $\mathbb{Z}^d$, let $G(V)$ denote the subgraph of $\mathbb{Z}^d$ induced by V. We will sometimes make abuse of notation by referring to $G(V)$ as V. With this convention $B_{\mathbb{Z}^d}(x,r)$ will sometimes mean $G(B_{\mathbb{Z}^d}(x,r))$ and the meaning will be clear from the context. For a cube Q in $\mathbb{Z}^d$, $\partial^{\mathrm{in}} Q$ will denote the "inner vertex boundary" of Q, that is, the set of all vertices in Q that are adjacent to at least one vertex not in Q.

For $p \in [0,1]$, consider i.i.d. Bernoulli(p) random variables $\{X_e\}_{e \in \mathbb{Z}^d}$ associated with edges of $\mathbb{Z}^d$, that is, $\mathbb{P}(X_e = 1) = p = 1 - \mathbb{P}(X_e = 0)$. We call an edge e open (resp., closed) at level p if $X_e = 1$ (resp., $X_e = 0$). Given a subgraph $G = (V, E)$ of $\mathbb{Z}^d$ and $V' \subset V$, we say that V' forms a p-cluster in G if there is a path consisting of open edges in E between any two vertices in V' and V' is a maximal subset of V in this regard.

For two cubes $Q_1 \subset Q_2$ in $\mathbb{Z}^d$, denote by $Q_2 - Q_1$ the subgraph (V, E) of Q_2 with

$$E = \{\text{all edges in } Q_2 \text{ except the ones with both endpoints in } Q_1\} \quad \text{and}$$

$$V = \{v : v \text{ is an endpoint of } e \text{ for some } e \in E\}.$$

For two cubes $Q_1 \subset Q_2$ in $\mathbb{Z}^d$ and $p \in [0,1]$, $Q_1 \overset{k}{\underset{p}{\leftrightsquigarrow}} Q_2$ will mean that there exist at least k disjoint p-clusters in $Q_2 - Q_1$ that intersect both $\partial^{\mathrm{in}} Q_1$ and $\partial^{\mathrm{in}} Q_2$ (see Figure 1). If Q_1, Q_2, Q_3 are cubes in $\mathbb{Z}^d$ such that (i) $Q_1 \subset Q_2 \cap Q_3$, and (ii) $\partial^{\mathrm{in}} Q_2$ has a vertex in Q_3, then we will write "$Q_1 \overset{k}{\underset{p}{\leftrightsquigarrow}} Q_2$ in Q_3" if there exist k disjoint p-clusters in $(Q_2 - Q_1) \cap Q_3$ each intersecting $\partial^{\mathrm{in}} Q_1$ and $\partial^{\mathrm{in}} Q_2$ (see Figure 2).

For an edge $\{x, y\}$ in $\mathbb{Z}^d$ and a cube Q containing both x and y, $\{x, y\} \overset{2}{\underset{p}{\leftrightsquigarrow}} Q$ will mean that the p-clusters in Q containing x and y are disjoint and that they both intersect $\partial^{\mathrm{in}} Q$. Similarly, we can define $\{x, y\} \overset{2}{\underset{p}{\leftrightsquigarrow}} Q - \{x, y\}$ to be the event that the p-clusters in $Q - \{x, y\}$ containing x and y intersect $\partial^{\mathrm{in}} Q$ and are disjoint.

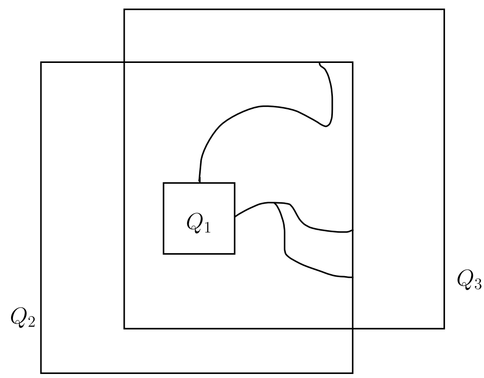

FIG. 2. $Q_1 \overset{2}{\underset{p}{\leftrightsquigarrow}} Q_2$ *in* Q_3.

Assume that $\{x, y\}$ is an edge of $\mathbb{Z}^d$ and $n \geq 2$. Analogous to the continuum setup, we call $\{\{x, y\} \overset{2}{\underset{p}{\leftrightsquigarrow}} B_{\mathbb{Z}^d}(x, n)\}$ or $\{B_{\mathbb{Z}^d}(x, 1) \overset{2}{\underset{p}{\leftrightsquigarrow}} B_{\mathbb{Z}^d}(x, n)\}$ a two-arm event at level p.

4.3. *Convention about constants*. To ease notation, most constants in this paper will be denoted by c, c', C, etc. and their values may change from line to line. These constants may depend on parameters like the dimension and often we will not mention this dependence explicitly; none of these constants will depend on the quantity "n," used to index infinite sequences. Specific constants will have a subscript as, for example, c_1, c_2, etc.

5. Two-arm event: Quantification of the Burton–Keane argument. The key ingredients in the proofs of Theorems 2.1 and 2.4 are some percolation theoretic estimates which are of independent interest. We state them in the following lemmas.

LEMMA 5.1. *Assume $d \geq 3$ and let $\mathcal{P}$ be a Poisson process having intensity one in $\mathbb{R}^d$. Let $0 < r_1 < r_2 < \infty$. Then there exist constants c_6 and c_7 depending only on r_1, r_2 and d such that for every $r \in [r_1, r_2]$, every $n \geq 2$ and every $a \in (1/2, (\log\log n)^{1/(d-1/2)})$, we have*

$$\mathbb{P}\big(B_{\mathbb{R}^d}(a) \overset{2}{\underset{r}{\longleftrightarrow}} B_{\mathbb{R}^d}(n)\big) \leq \frac{c_6 \exp(c_7 a^{d-1})}{(\log n)^{\frac{d}{2}}}. \tag{5.1}$$

The same bound holds if we replace $B_{\mathbb{R}^d}(a)$ by $B_{\mathbb{R}^d}(a)^{(r)}$ or $B_{\mathbb{R}^d}(a)^{(r)} \cup S_{\mathbb{R}^d}(x, r)$ for some $x \in B_{\mathbb{R}^d}(a)^{(r)}$.

The proof of this lemma is given in Section 9. Lemma 5.1 deals with the case $d \geq 3$. The case $d = 2$ is simpler and will be handled in Lemma 9.5. The next lemma states a similar result for the lattice case.

LEMMA 5.2 ([23], Proposition 5.3). *Consider the lattice* $\mathbb{Z}^d$ *where* $d \geq 2$ *and let* $e_1, \dots, e_{2d}$ *be as in Lemma 5.7. Then for any* $0 < p_1 < p_2 < 1$, *there exists a constant* c_9 *depending only on* p_1, p_2 *and* d *such that for any* $p \in [p_1, p_2]$ *and* $n \geq 2$,

$$\mathbb{P}\big(\{0, e_i\} \overset{2}{\underset{p}{\leftrightsquigarrow}} B_{\mathbb{Z}^d}(n)\big) \leq c_9 \left(\frac{\log n}{n}\right)^{1/2} \qquad \text{for } 1 \leq i \leq 2d. \tag{5.2}$$

The same bound holds if we replace the edge $\{0, e_i\}$ *by the cube* $B_{\mathbb{Z}^d}(1)$.

REMARK 5.3. Let $r_c = r_c(d)$ be the critical radius for continuum percolation in $\mathbb{R}^d$ driven by a Poisson process with intensity one (see, e.g., [8] or [19], Chapter 8). Note that we can actually get an exponentially decaying bound in (5.1) when $r_2 < r_c$. It is also possible to prove exponential decay in (5.1) if $r_1 > r_c$. So the bound in (5.1) is really useful when $r_c \in (r_1, r_2)$.

The same is true for Lemma 5.2. Exponential decay in (5.2) is standard when $p_c(\mathbb{Z}^d) \notin [p_1, p_2]$.

REMARK 5.4. Proposition 5.3 of [23] was actually proved for site percolation on $\mathbb{Z}^d$. However, the proof can be easily generalized to bond percolation. Also, the bound given in Proposition 5.3 of [23] is of the form $O(\log n/\sqrt{n})$, but it is straightforward to modify the proof to get a bound of the form $O(\sqrt{(\log n/n)})$. Indeed, in Section 5 of [23], we can modify the definition of the event $\mathcal{E}$ as follows:

$$\mathcal{E} := \big\{\forall C \in \mathcal{C}, \big|h\big(\overline{C} \cap \Lambda(n)\big)\big| < \alpha(\log n)^{1/2}\big|\overline{C} \cap \Lambda(n)\big|^{1/2}\big\},$$

where $\alpha > 0$ is a large constant. Then it will follow that

$$\mathbb{P}(\mathcal{E}^c) \leq 2\big|\Lambda(n)\big|^2 \exp\big(-2\alpha^2(\log n)p^2(1-p)^2\big).$$

We can choose α sufficiently large and follow the rest of analysis in [23] to get a bound of the form $O(\sqrt{(\log n/n)})$.

REMARK 5.5. In the proof of Theorem 2.4, we need a bound on the probability of two-arm events *which is uniform in* p *over an open interval containing* $p_c(\mathbb{Z}^d)$. Lemma 5.2 serves this purpose. It is, however, possible that the estimate in Lemma 5.2 is sub-optimal. In [23], Cerf improves the bound given in Lemma 5.2 but only at $p = p_c$. In the recent preprint [28], the authors prove a bound of the form $O(1/n)$ for bond percolation in $\mathbb{Z}^2$ (in fact, their result is true for the more general random cluster model), and Kozma and Nachmias [37] prove a bound of the form $O(1/n^4)$ for bond percolation in $\mathbb{Z}^d$ when $d \geq 19$ but again, these bounds hold only at $p = p_c$. For site percolation on the triangular lattice, a bound of the form $O(n^{-5/4+o(1)})$ is known to hold at criticality [54], but an analogous result is not known for the square lattice $\mathbb{Z}^2$.

To the best of our knowledge, the bound in (5.2) is the best-known estimate valid uniformly over an interval around p_c. Any improvement over Lemma 5.2 can be used in the proof of Theorem 2.4 to get better bounds in (2.5). Similarly, any improvement over Lemma 5.1 will yield a sharper upper bound in (2.4).

REMARK 5.6. The arguments used in the proof of Lemma 5.1 can be used in the lattice setup to get the following result.

LEMMA 5.7. *Consider the lattice* $\mathbb{Z}^d$ *where* $d \geq 3$. *Denote the vertices adjacent to the origin by* $e_1, \dots, e_{2d}$. *Then for any* $0 < p_1 < p_2 < 1$, *there exists a constant* c_8 *depending only on* p_1, p_2 *and* d *such that for any* $p \in [p_1, p_2]$ *and* $n \geq 2$,

$$\mathbb{P}\big(\{0, e_i\} \overset{2}{\underset{p}{\leftrightsquigarrow}} B_{\mathbb{Z}^d}(n)\big) \leq c_8 (\log n)^{-\frac{d}{2}} \qquad \textit{for } 1 \leq i \leq 2d. \tag{5.3}$$

The same bound holds if we replace the edge $\{0, e_i\}$ *by the cube* $B_{\mathbb{Z}^d}(1)$.

The proof of this lemma is outlined briefly in the Appendix. Lemmas 5.1 and 5.7 may be seen as quantifications of the statement that the infinite open cluster is unique. This uniqueness theorem was first proved by Aizenman, Kesten and Newman [3] for percolation on lattices (see also [30]). A very elegant proof was given by Burton and Keane [21], which has now become the standard textbook proof of the theorem. Unlike the original argument of Aizenman, Kesten and Newman, the Burton–Keane argument admits a wide array of applications and generalizations due to its simplicity and robustness.

The AKN argument is known to have a quantitative version *in the lattice setup* (Lemma 5.2), while the Burton–Keane argument, due to its use of translation-invariance, is not expected to be quantifiable. The argument used in the proofs of Lemmas 5.1 and 5.7 show that it is actually possible to quantify the Burton–Keane argument. Thus, the technique used in the proofs of Lemmas 5.1 and 5.7 is expected to have wider applicability in other contexts, where the Burton–Keane argument works but the AKN argument does not. As mentioned earlier, using a generalization of the arguments used in the proof of Lemma 5.7, Duminil-Copin, Ioffe and Velenik [28] have recently obtained bounds on the probability of two-arm events in a broad class of translation-invariant percolation models on $\mathbb{Z}^d$. Due to this recent development, we have included a brief sketch of the proof of Lemma 5.7 in the Appendix even though in the proof of Theorem 2.4 we will use Lemma 5.2 which gives a sharper bound.

6. Two standard facts about minimal spanning trees. We collect two well-known facts about minimal spanning trees in this section.

6.1. *Minimax property of paths in MST.*

LEMMA 6.1. *Consider a finite, connected and weighted graph* $G = (V, E, w)$. *Let* T *be a minimal spanning tree of* G. *Then any path* $(x_0, \dots, x_n)$ *with* $x_i \in V$ *and* $\{x_i, x_{i+1}\} \in T$ *satisfies*

$$\max_i w(\{x_i, x_{i+1}\}) \le \max_j w(\{x'_j, x'_{j+1}\})$$

for any path $(x'_0, \dots, x'_m)$ *with* $\{x'_j, x'_{j+1}\} \in E$ *and* $x_0 = x'_0$ *and* $x_n = x'_m$.

PROOF. This is just a restatement of [36], Lemma 2. $\square$

In words, Lemma 6.1 states that any path in the MST is minimax, that is, for any two vertices x and y, the path in the MST that connects x and y minimizes the maximum edge-weight among all paths in the graph that connect x and y.

6.2. *Add and delete algorithm.* We now state an algorithm from [36] for constructing an MST on a connected graph starting from an MST on a connected subgraph:

(i) *Addition of an edge*: Suppose $G_1 = (V, E_1, w)$ is a finite connected weighted graph and $G_0 = (V, E_0, w)$ is a connected subgraph of G_1 such that $E_1 = E_0 \cup \{e_0\}$, that is, G_1 has the same vertex set and one extra edge e_0. Suppose T_0 is an MST on G_0. Consider the graph $T_0 \cup \{e_0\}$, that is, add the edge e_0 to T_0. Then $T_0 \cup \{e_0\}$ has a unique cycle C. Let e be an edge in C such that $w(e) = \max_{e' \in C} w(e')$, and set $T_1 = T_0 \cup \{e_0\} \setminus e$. (Thus, we are removing an edge in C that has the maximal edge-weight in C.)

(ii) *Addition of a vertex*: Suppose $G_1 = (V_1, E_1, w)$ is a finite connected weighted graph and $G_0 = (V_0, E_0, w)$ is a connected subgraph of G_1 such that $V_1 = V_0 \cup \{v_0\}$ and $E_1 = E_0 \cup \{e_0\}$. (Thus G_1 has one extra vertex v_0 and one extra edge e_0. Since G_1 is connected, v_0 is necessarily an endpoint of e_0.) Suppose T_0 is an MST on G_0. Set $T_1 = T_0 \cup \{e_0\}$.

PROPOSITION 6.2 ([36], Proposition 2). *The tree* T_1 *constructed in* (i) *or* (ii) *is an MST on* G_1.

We can start from an MST on a connected graph and use the add and delete algorithm inductively to construct an MST on any larger finite connected graph.

7. Outline of proof. We briefly sketch here the main ideas in the proof. For simplicity, let us consider the case where *the edges of* $\mathbb{Z}^d$ *have been weighted by i.i.d.* Uniform[0, 1] *random variables*. Let X_f denote the weight associated with an edge f of $\mathbb{Z}^d$, and let $X = (X_f : f \text{ is an edge of } B_{\mathbb{Z}^d}(n))$. Heuristically, we expect $M(B_{\mathbb{Z}^d}(n), X)$ to satisfy a CLT if the change in $M(B_{\mathbb{Z}^d}(n), X)$ due to the replacement of X_f by an independent identically distributed observation X'_f "is not observed far away from f." A quantitative formulation of this vague statement will give us a convergence rate in the CLT.

To this end, fix $\alpha \in (0, 1)$ and take an edge $e = \{x_1, x_2\}$ in $B_{\mathbb{Z}^d}(n)$ such that $d(x_1, \partial^{\text{in}} B_{\mathbb{Z}^d}(n)) \geq \lceil n^\alpha \rceil$. Let X' be an independent copy of X. Recall the notation X^e from Section 3.1. Define

$$\Delta_e M = M(B_{\mathbb{Z}^d}(n), X) - M(B_{\mathbb{Z}^d}(n), X^e) \quad \text{and}$$
$$\tilde{\Delta}_e M = M(B_{\mathbb{Z}^d}(x_1, n^\alpha), X) - M(B_{\mathbb{Z}^d}(x_1, n^\alpha), X^e).$$

Then an application of Theorem 3.1 reduces the problem to getting an upper bound on $\mathbb{E}|\Delta_e M - \tilde{\Delta}_e M|$. The actual calculations are given in Section 12.2. This is the precise formulation of the heuristics explained above.

Noting that

$$\begin{aligned}\Delta_e M = {} & [M(B_{\mathbb{Z}^d}(n), X) - M(B_{\mathbb{Z}^d}(n) - e, X)] \\ & - [M(B_{\mathbb{Z}^d}(n), X^e) - M(B_{\mathbb{Z}^d}(n) - e, X^e)],\end{aligned}$$

and a similar identity holds for $\tilde{\Delta}_e M$, it is easily seen that getting a bound on $\mathbb{E}|\Delta_e M - \tilde{\Delta}_e M|$ amounts to proving an upper bound on $\mathbb{E}|\delta_e M|$, where

$$\begin{aligned}\delta_e M := {} & [M(B_{\mathbb{Z}^d}(n), X) - M(B_{\mathbb{Z}^d}(n) - e, X)] \\ & - [M(B_{\mathbb{Z}^d}(x_1, n^\alpha), X) - M(B_{\mathbb{Z}^d}(x_1, n^\alpha) - e, X)].\end{aligned}$$

It follows from Proposition 6.2 that

$$M(B_{\mathbb{Z}^d}(n), X) - M(B_{\mathbb{Z}^d}(n) - e, X) = X_e - \max\{X_e, Y\} \quad \text{and}$$
$$M(B_{\mathbb{Z}^d}(x_1, n^\alpha), X) - M(B_{\mathbb{Z}^d}(x_1, n^\alpha) - e, X) = X_e - \max\{X_e, \tilde{Y}\},$$

where Y (resp., $\tilde{Y}$) is the maximum weight associated with the edges in the path, Γ_1 (resp. Γ_2) connecting x_1 and x_2 in an MST of $B_{\mathbb{Z}^d}(n) - e$ [resp., $B_{\mathbb{Z}^d}(x_1, n^\alpha) - e$]. Thus, $\mathbb{E}|\delta_e M| \leq \mathbb{E}|\tilde{Y} - Y|$.

By the minimax property of paths in MST (Lemma 6.1), $(\tilde{Y} - Y)$ is always nonnegative. Further,

$$\mathbb{E}(\tilde{Y} - Y) = \int_0^1 \mathbb{P}(Y < u < \tilde{Y})\, du. \tag{7.1}$$

Note that $\{\mathbb{I}_{X_f \leq u} : f \text{ is an edge of } B_{\mathbb{Z}^d}(n)\}$ is a collection of i.i.d. Bernoulli(u) random variables. Declare the edge f to be open at level u if $X_f \leq u$, and consider the corresponding u-clusters. On the set $\{Y < u < \tilde{Y}\}$, the u-clusters in $B_{\mathbb{Z}^d}(x_1, n^\alpha) - e$ containing x_1 and x_2 are disjoint (since $\tilde{Y} > u$). However, x_1 and x_2 are connected in $B_{\mathbb{Z}^d}(n) - e$ by a path open at level u (since $Y < u$). Hence, the

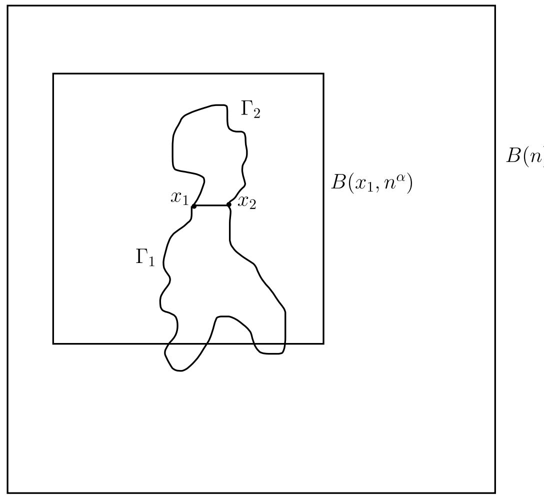

FIG. 3. *The minimax paths connecting x_1 and x_2 when $\tilde{Y} > Y$.*

u-clusters in $B_{\mathbb{Z}^d}(x_1, n^\alpha) - e$ containing x_1 and x_2 both intersect $\partial^{\text{in}} B_{\mathbb{Z}^d}(x_1, n^\alpha)$. [In this case, part of Γ_1 lies outside $B_{\mathbb{Z}^d}(x_1, n^\alpha)$; see Figure 3.] Thus,

$$\mathbb{P}(Y < u < \tilde{Y}) \le \mathbb{P}\big(e \overset{2}{\underset{u}{\leftrightsquigarrow}} B_{\mathbb{Z}^d}(x_1, n^\alpha) - e\big).$$

We can now use estimates on probability of two-arm events to bound $\mathbb{E}(\tilde{Y} - Y)$. Thus, for any small positive ε, the integrand in (7.1) is bounded by $c(\log(n)/n)^{1/2}$ for $u \in (p_c - \varepsilon, p_c + \varepsilon)$ (Lemma 5.2), and benefits from the exponential decay when $u \notin (p_c - \varepsilon, p_c + \varepsilon)$.

For Euclidean MST, we start by dividing $B_{\mathbb{R}^d}(n)$ into cubes $\{Q \in \mathcal{Q}\}$ with disjoint interiors having side length $s \in [1, 2]$. Consider a Poisson process $\mathcal{P}$ in $\mathbb{R}^d$ of intensity one and let $X_Q := \mathcal{P} \cap Q$ for any cube Q. Set $X = (X_Q : Q \in \mathcal{Q})$, and let X' be an independent copy of Q. Consider a cube $Q_0 \in \mathcal{Q}$ with $d(c(Q_0), \partial B_{\mathbb{R}^d}(n)) \ge n^\alpha$. In line with the notation in Section 3.1, X^{Q_0} denotes the configuration in $B_{\mathbb{R}^d}(n)$ when the configuration inside Q_0 is X'_{Q_0}, and the configuration in $B_{\mathbb{R}^d}(n) \setminus Q_0$ is given by $\bigcup_{Q \in \mathcal{Q} \setminus Q_0} X_Q$. Similar to the discrete case, our aim then is to get a bound on $\mathbb{E}|\Delta_{Q_0} M_n - \tilde{\Delta}_{Q_0} M_n|$, where

$$\Delta_{Q_0} M_n = M_{\mathbb{R}^d}(X) - M_{\mathbb{R}^d}\big(X^{Q_0}\big) \quad \text{and}$$
$$\tilde{\Delta}_{Q_0} M_n = M_{\mathbb{R}^d}\big(X \cap B_{\mathbb{R}^d}\big(c(Q_0), n^\alpha\big)\big) - M_{\mathbb{R}^d}\big(X^{Q_0} \cap B_{\mathbb{R}^d}\big(c(Q_0), n^\alpha\big)\big).$$

This can also be reduced to getting a bound on the probability of the two-arm event in the setup of continuum percolation. However, since all possible edges between points are permitted, this step requires a little work. We achieve this by introducing the concept of a "wall" (Definition 8.1) and then using the add and delete algorithm. We will omit the details of these steps from the proof sketch.

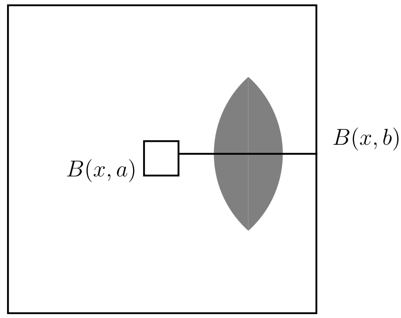

FIG. 4. *For a wall to exist around $B(x,a)$ in $B(x,b)$, the shaded region must contain a point.*

8. Some results about Euclidean minimal spanning trees. In this section, the underlying space will always be $\mathbb{R}^d$, and we will simply write $B(\cdot,\cdot)$, $d(\cdot,\cdot)$ and $M(\cdot)$ instead of $B_{\mathbb{R}^d}(\cdot,\cdot)$, $d_{\mathbb{R}^d}(\cdot,\cdot)$, and $M_{\mathbb{R}^d}(\cdot)$.

When dealing with Euclidean minimal spanning trees, we would like to have a criterion which ensures that if we fix a small cube, then there are no "long" edges in the MST with one endpoint inside that cube. Kesten and Lee [36] used the idea of a "separating set" to meet this purpose. (We will not define separating sets since we do not use them in this paper.) We generalize their ideas to define a "wall" (see Definition 8.1 below). The reason behind this is that using the notion of separating sets in our proof will yield a weaker convergence rate than the one stated in Theorem 2.1.

DEFINITION 8.1. Suppose that $b > a$ are positive numbers and $x \in \mathbb{R}^d$ and let K be a cube containing $B(x,a)$. Further assume that $K \cap \partial B(x,b) \neq \varnothing$. We say that a subset $\mathfrak{W}$ of $\mathbb{R}^d$ contains a K-wall around $B(x,a)$ in $B(x,b)$ if the following holds:

$$\text{For any } p_1 \in \partial B(x,a) \text{ and } p_2 \in K \cap \partial B(x,b), \text{ the set}$$
$$K \cap \mathfrak{W} \cap S\big(p_1, 3d(p_1,p_2)/4\big) \cap S\big(p_2, 3d(p_1,p_2)/4\big) \cap \big\{B(x,b) \setminus B(x,a)\big\}$$
$$\text{is nonempty.}$$

If $B(x,b) \subset K$, we will simply say $\mathfrak{W}$ contains a wall around $B(x,a)$ in $B(x,b)$ (see Figure 4).

The importance of this definition will be clear from the following lemma.

LEMMA 8.2. *Let a, b, x, K be as in Definition 8.1. Let ω be a finite set of points in K and consider the complete graph (V,E) on ω with edge weights being the Euclidean length of edges. If ω contains a K-wall around $B(x,a)$ in $B(x,b)$, then no edge in E with one endpoint in $B(x,a)$ and other endpoint in $B(x,b)^c$ is included in any MST of (V,E).*

PROOF. Let y_1, y_2 be two points in ω such that $y_1 \in B(x,a)$ and $y_2 \in B(x,b)^c$. Assume that $p_1 \in \partial B(x,a)$ and $p_2 \in \partial B(x,b)$ are points on the line segment $\overline{y_1 y_2}$. Since ω contains a K-wall around $B(x,a)$ in $B(x,b)$, we can find a point z such that

$$z \in \bigl[\omega \cap S\bigl(p_1, 3d(p_1, p_2)/4\bigr) \cap S\bigl(p_2, 3d(p_1, p_2)/4\bigr) \cap \bigl(B(x,b) \setminus B(x,a)\bigr)\bigr].$$

Then

$$\begin{aligned} d(y_1, z) &\le d(y_1, p_1) + d(p_1, z) \le d(y_1, p_1) + 3d(p_1, p_2)/4 \\ &< d(y_1, p_1) + d(p_1, y_2) = d(y_1, y_2). \end{aligned}$$

Similarly, $d(z, y_2) < d(y_1, y_2)$. Hence, it follows from Lemma 6.1 that $\overline{y_1 y_2}$ will not be included in any minimal spanning tree of (V, E). $\square$

Next, we show that a wall exists in a large annulus with high probability.

LEMMA 8.3. *Let $d \ge 2$ and $x \in \mathbb{R}^d$. As always we let $\mathcal{P}$ be a Poisson process of intensity one in $\mathbb{R}^d$. Then for any $a_0 > 0$, there exist constants c and c' depending only on a_0 and d such that the following holds: for every $a \le a_0$ and $b > a$,*

$$\begin{aligned} &\mathbb{P}\bigl(\mathcal{P} \textit{ does not contain a } B(n)\textit{-wall around } B(x,a) \textit{ in } B(x,b)\bigr) \\ &\qquad \le c \exp\bigl(-c' b^d\bigr) \end{aligned}$$

for any n for which $B(x,a) \subset B(n)$ and $B(n) \cap \partial B(x,b) \ne \varnothing$.

PROOF. It suffices to prove the claim for large values of b, so let us start with the assumption $b > 4a_0 + 16$.

Cover $B(n) \cap \partial B(x,b)$ by $(d-1)$ dimensional cubes, $\{Q_i^1\}_{i \le m_1}$ of diameter one. This can be done in a way so that the total number of cubes, m_1, is at most cb^{d-1}. Similarly, cover $\partial B(x,a)$ by $(d-1)$ dimensional cubes $\{Q_i^2\}_{i \le m_2}$ of diameter $\min(1, 2a\sqrt{d-1})$ so that the total number of cubes, m_2, is at most $c \max(1, a_0^{d-1})$.

Let p_1', p_2' be two points on $\partial B(x,a)$ and $B(n) \cap \partial B(x,b)$, respectively, and let $z' = (p_1' + p_2')/2$ be the midpoint of $\overline{p_1' p_2'}$. Let p_1 and p_2 be the centers of the cubes Q_i^1 and Q_j^2 such that $p_1' \in Q_i^1$ and $p_2' \in Q_j^2$. Let $z = (p_1 + p_2)/2$.

Consider $y' \in S(z', b/8)$. Then $\|y' - z'\|_\infty \le b/8$, and hence

$$\|z' - x\|_\infty - b/8 \le \|y' - x\|_\infty \le \|z' - x\|_\infty + b/8.$$

Now,

$$\begin{aligned} \|z' - x\|_\infty + \frac{b}{8} &= \left\| \frac{1}{2}(p_1' + p_2' - 2x) \right\|_\infty + \frac{b}{8} \\ &\le \frac{a_0 + b}{2} + \frac{b}{8} < b. \end{aligned}$$

Also

$$\|z' - x\|_\infty - \frac{b}{8} \geq \frac{b - a_0}{2} - \frac{b}{8} > a.$$

Hence, $S(z', b/8) \subset B(x, b) \setminus B(x, a)$. Further, if $y \in S(z, b/16)$, then

$$d(y, z') \leq \frac{b}{16} + d(z, z') = \frac{b}{16} + \left\| \frac{p_1 + p_2}{2} - \frac{p_1' + p_2'}{2} \right\|_{L^2} \leq \frac{b}{16} + 1 \leq \frac{b}{8}.$$

So $S(z, b/16) \subset S(z', b/8) \subset B(x, b) \setminus B(x, a)$.

If $y' \in S(z', b/8)$, then

$$d(y', p_1') \leq d(y', z') + d(z', p_1') \leq \frac{b}{8} + \frac{d(p_1', p_2')}{2} \leq \frac{3d(p_1', p_2')}{4}.$$

The last inequality holds since

$$d(p_1', p_2') \geq b - a \geq b - a_0 \geq b/2.$$

By a similar argument, $d(y', p_2') \leq 3d(p_1', p_2')/4$. Hence,

$$S(z', b/8) \subset S(p_1', 3d(p_1', p_2')/4) \cap S(p_2', 3d(p_1', p_2')/4) \cap (B(x, b) \setminus B(x, a)).$$

Letting $\mathfrak{Leb}$ denote the Lebesgue measure, we note that $\mathfrak{Leb}(S(z, b/16) \cap B(n)) \geq c' b^d$. So we can conclude that

$$\begin{aligned}
&\mathbb{P}(\mathcal{P} \text{ does not contain a } B(n)\text{-wall around } B(x, a) \text{ in } B(x, b)) \\
&\quad \leq \mathbb{P}\biggl(\text{For some } i \leq m_1, j \leq m_2, \mathcal{P} \cap B(n) \cap S\left(\frac{p_1 + p_2}{2}, \frac{b}{16}\right) = \varnothing \\
&\qquad\qquad \text{where } p_1 \text{ and } p_2 \text{ are the centers of } Q_i^1 \text{ and } Q_j^2 \text{ respectively}\biggr) \\
&\quad \leq c \max(1, a_0^{d-1}) b^{d-1} \exp(-c' b^d),
\end{aligned}$$

where the last inequality follows from union bound. This proves the claim. □

The next lemma puts an upper bound on how much the weight of the MST changes when some points are removed.

LEMMA 8.4. *Let a, b, x, K be as in Definition 8.1. Let $\mathcal{A}$ and $\mathcal{B}$ be finite sets of points in $\mathbb{R}^d$ such that $\mathcal{A} \subset B(x, a)$ and $\mathcal{B} \subset K \setminus B(x, a)$. If $\mathcal{B}$ contains a K-wall around $B(x, a)$ in $B(x, b)$, then*

$$|M(\mathcal{A} \cup \mathcal{B}) - M(\mathcal{B})| \leq c|\mathcal{A}| b$$

for some constant c depending only on d. If such a wall does not exist, then

$$|M(\mathcal{A} \cup \mathcal{B}) - M(\mathcal{B})| \leq c|\mathcal{A}| \operatorname{diameter}(K).$$

The proof of Lemma 8.4 is similar to the proof of [36], Lemma 7. We include this argument for the reader's convenience. The proof depends on an auxiliary lemma.

LEMMA 8.5 ([5], Lemma 4). *Consider an MST $\mathcal{T}$ on a finite subset ω of $\mathbb{R}^d$. Then there exists a constant $D_{\max}$ depending only on d such that the degree, in $\mathcal{T}$, of any point in ω is bounded by $D_{\max}$.*

PROOF OF LEMMA 8.4. First, we assume that $\mathcal{B}$ contains a K-wall around $B(x,a)$ in $B(x,b)$. Then $\mathcal{B}$ has a point, say p, in $B(x,b) \setminus B(x,a)$. Thus, we can start from an MST on $\mathcal{B}$ and connect the points in $\mathcal{A}$ to p to get a spanning tree on $\mathcal{A} \cup \mathcal{B}$. This gives

$$M(\mathcal{A} \cup \mathcal{B}) \le M(\mathcal{B}) + |\mathcal{A}| b\sqrt{d}.$$

To get the other inequality, we start from an MST on $\mathcal{A} \cup \mathcal{B}$ and delete the points in $\mathcal{A}$ and all edges incident to them. By Lemma 8.2, each of these edges is contained in $B(x,b)$. By Lemma 8.5, we have deleted at most $D_{\max}|\mathcal{A}|$ many edges and this can create at most $(D_{\max}|\mathcal{A}| + 1)$ many components. Each of these components has a point in $B(x,b)$. We can then connect these points to get a spanning tree on $\mathcal{B}$. This gives

$$M(\mathcal{B}) \le M(\mathcal{A} \cup \mathcal{B}) + D_{\max}|\mathcal{A}| b\sqrt{d}.$$

The proof is similar when a wall does not exist. □

Lemma 8.4 gives us control over the tails of $|M_{\mathbb{R}^d}(\mathcal{A} \cup \mathcal{B}) - M_{\mathbb{R}^d}(\mathcal{B})|$. Using this, we can show that all moments of this quantity are finite when the configuration comes from a Poisson process.

LEMMA 8.6. *For $x \in \mathbb{R}^d$, $0 < a \le a_0$ and $n \ge \max(2a_0, 1)$ for which $B(x,a) \subset B(n)$, we have*

$$\mathbb{E}\big(\big|M\big(\mathcal{P} \cap B(n)\big) - M\big(\mathcal{P} \cap \big[B(n) \setminus B(x,a)\big]\big)\big|^q\big) \le C_q \qquad \text{for every } q \ge 1.$$

The constant C_q depends only on a_0, d and q.

PROOF. Define a random variable Z as follows: if there does not exist a $b \ge a$ such that $\partial B(x,b) \cap B(n) \ne \varnothing$ and $\mathcal{P}$ contains a $B(n)$-wall around $B(x,a)$ in $B(x,b)$, set $Z = 2\sqrt{d}n$; otherwise define Z to be the infimum of all such b. From Lemma 8.3,

$$\begin{aligned}
\mathbb{E}(Z^q) &= \int_0^{2\sqrt{d}n} q u^{q-1} \mathbb{P}(Z > u)\, du \\
&\le a_0^q + c\int_a^n q u^{q-1} \exp(-c' u^d)\, du + c(2\sqrt{d}n)^q \exp(-c' n^d).
\end{aligned}$$

The last expression is bounded by a constant depending only on a_0, d and q. Now, from Lemma 8.4

$$\mathbb{E}(|M(\mathcal{P}\cap B(n)) - M(\mathcal{P}\cap[B(n)\setminus B(x,a)])|^q)$$
$$\leq c\mathbb{E}(Z\cdot|\mathcal{P}\cap B(x,a)|)^q \leq \frac{c}{2}\mathbb{E}[Z^{2q} + (|\mathcal{P}\cap B(x,a)|)^{2q}],$$

and this completes the proof. $\square$

9. Proofs of percolation estimates in the Euclidean setup. In this section, the underlying space will always be $\mathbb{R}^d$, and all Poisson processes will have intensity one. We will simply write $B(\cdot,\cdot)$ and $d(\cdot,\cdot)$ without referring to the ambient space. Recall form Remark 5.3 that $r_c(d)$ denotes the critical radius for continuum percolation in $\mathbb{R}^d$ driven by a Poisson process with intensity one. When the dimension d is clear, we will simply write r_c instead of $r_c(d)$.

Before beginning the proof of Lemma 5.1, we collect two simple facts in the following lemma.

LEMMA 9.1. (i) *Let $X_1,\ldots,X_n$ be independent random variables defined on $(\Omega,\mathcal{A},\mathbb{P})$ taking values in some measurable space $(\mathcal{X},\mathcal{S})$. Let $f:\mathcal{X}^n\to\mathbb{R}$ be a bounded measurable function. Then for any $A_1,\ldots,A_k\subset\{1,\ldots,n\}$ such that A_i are pairwise disjoint*,

$$\operatorname{Var}(f(X_1,\ldots,X_n)) \geq \sum_{i=1}^k \operatorname{Var}[\mathbb{E}(f(X_1,\ldots,X_n)|\{X_j\}_{j\in A_i})]. \tag{9.1}$$

(ii) *If Y_1 and Y_2 are independent and identically distributed real valued random variables such that $\mathbb{E}(Y_1^2)<\infty$, then*

$$\operatorname{Var}(Y_1) = \frac{1}{2}\mathbb{E}(Y_1-Y_2)^2. \tag{9.2}$$

PROOF. Equation (9.2) is a basic identity whose proof we will omit. To prove (9.1), without loss of generality, we can assume $\mathbb{E}(f(X_1,\ldots,X_n))=0$. Let

$$H = \left\{g\in L^2(\Omega,\mathcal{A},\mathbb{P}) : \int g = 0\right\} \quad \text{and}$$
$$H_i = \{g\in H : g \text{ is } \sigma(\{X_j\}_{j\in A_i}) \text{ measurable}\}.$$

Then under the natural inner product, H is a Hilbert space and the H_i are closed orthogonal subspaces of H. Equation (9.1) follows upon observing that $\mathbb{E}(f(X_1,\ldots,X_n)|\{X_j\}_{j\in A_i})$ is the projection of $f(X_1,\ldots,X_n)$ on H_i. $\square$

The following lemma plays a crucial role in the proof of Lemma 5.1.

LEMMA 9.2. *Let $0 < r_1 < r_2 < \infty$. Fix two nonnegative numbers s and t such that $s + t > 2r_2$. Then there exist positive constants c and c' depending only on r_1, r_2 and the dimension d such that for every $m > 100(s + t)$ and $r \in [r_1, r_2]$*

$$\mathbb{P}\big(B(s)^{(t)} \xleftrightarrow[r]{3} B(m)\big) \le c \cdot \exp\big(c'(s+t)\big)/m. \tag{9.3}$$

For $z_1, z_2 \in B(s)^{(t)}$, the same bound holds for $\mathbb{P}(B(s)^{(t)} \cup S(z_1, r) \xleftrightarrow[r]{3} B(m))$ and $\mathbb{P}(B(s)^{(t)} \cup S(z_1, r) \cup S(z_2, r) \xleftrightarrow[r]{3} B(m))$.

The proof of Lemma 9.2 will be given in Section 9.2. We now proceed with the following.

9.1. *Proof of Lemma* 5.1. Let us first prove the bound for $\mathbb{P}(B(a) \xleftrightarrow[r]{2} B(n))$. The arguments are similar when we replace $B(a)$ by the other sets. Fix $r \in [r_1, r_2]$. We write $\mathbb{R}^d$ as a union of cubes

$$\mathbb{R}^d = \bigcup_{k \in \mathbb{Z}^d} B_k \qquad \text{where } B_k = 2ak + B(a).$$

Since $\mathbb{P}(\mathcal{P} \cap \partial B_k \neq \varnothing \text{ for some } k \in \mathbb{Z}^d) = 0$, we will assume that no Poisson point lies in any of the common interfaces shared by two cubes.

Consider a sequence $a_n \to \infty$ such that $a_n = o(n)$ but $a_n = \Omega((\log\log n)^2)$ (so that a_n is large compared to a). We will fix the sequence a_n later. Define

$$E = \big\{\exists \text{ exactly one } r\text{-cluster } \mathcal{C} \text{ in } B(n) \text{ such that}$$
$$\mathcal{C}^{(r)} \text{ intersects both } \partial\big(B(n)_{(r)}\big) \text{ and } \partial B(a_n)\big\}.$$

Let

$$\mathcal{L} = \big\{k \in \mathbb{Z}^d : B_k \cap B(n) \neq \varnothing\big\} \quad \text{and} \quad \mathcal{I} = \big\{k \in \mathbb{Z}^d : B_k \cap B(a_n/3) \neq \varnothing\big\}.$$

Define $f : \prod_{k \in \mathcal{L}} \mathfrak{X}(B_k) \to \mathbb{R}$ by

$$f\big((\omega_k : k \in \mathcal{L})\big) = \mathbb{I}_E\bigg(\bigcup_{k \in \mathcal{L}} \omega_k\bigg).$$

Write

$$X_k = \mathcal{P} \cap B_k \quad \text{and} \quad X = (X_k : k \in \mathcal{L}).$$

It then follows from Lemma 9.1 that

$$\operatorname{Var}\big(f(X)\big) \ge \sum_{i \in \mathcal{I}} \operatorname{Var}\big[\mathbb{E}\big(f(X)|X_i\big)\big]. \tag{9.4}$$

Consider another Poisson process $\mathcal{P}'$ independent of $\mathcal{P}$, and set

$$X'_k = \mathcal{P}' \cap B_k \quad \text{and} \quad X' = \big(X'_k : k \in \mathcal{L}\big).$$

Recall the notation X^j from Section 3.1. Define

$$\mathcal{S}_i := \{\omega_i \in \mathfrak{X}(B_i) : B_{i\,(2r)} \subset \omega_i^{(r)}\} \quad \text{and} \quad \mathcal{G}_i := \{\omega_i \in \mathfrak{X}(B_i) : B_{i\,(2r)} \cap \omega_i^{(r)} = \varnothing\}.$$

Then, for any fixed $i \in \mathcal{I}$,

$$\begin{aligned}\operatorname{Var}[\mathbb{E}(f(X)|X_i)] &= \frac{1}{2}\mathbb{E}[(\mathbb{E}(f(X)|X_i) - \mathbb{E}(f(X^i)|X_i'))^2] \\ &\geq \frac{1}{2}\mathbb{E}[(\mathbb{E}(f(X) - f(X^i)|X_i, X_i'))^2 \cdot \mathbb{I}(X_i \in \mathcal{S}_i, X_i' \in \mathcal{G}_i)],\end{aligned} \tag{9.5}$$

where the first step uses (9.2) and the fact that $\mathbb{E}(f(X)|X_i)$ and $\mathbb{E}(f(X^i)|X_i')$ are independent and identically distributed.

Consider $i \in \mathcal{I}$, $\omega \in \mathfrak{X}(B(n) \setminus B_i)$ and $\omega_i' \in \mathcal{G}_i$. Then $\omega^{(r)}$ and $(\omega_i')^{(r)}$ are disjoint. Thus, if E holds when the configuration in B_i is ω_i' and the configuration in $B(n) \setminus B_i$ is ω, then E continues to hold when B_i is empty and the configuration in $B(n) \setminus B_i$ is ω. Further, if the event E holds with some configuration in $B(n)$, then E continues to hold with the configuration obtained by adding extra points inside $B(a_n/3)$. Thus, for any $\omega_i \in \mathcal{S}_i$ and $\omega_i' \in \mathcal{G}_i$,

$$\{\omega \in \mathfrak{X}(B(n) \setminus B_i) : \mathbb{I}_E(\omega \cup \omega_i') = 1\} \subset \{\omega \in \mathfrak{X}(B(n) \setminus B_i) : \mathbb{I}_E(\omega \cup \omega_i) = 1\}.$$

Therefore, if $X_i \in \mathcal{S}_i$ and $X_i' \in \mathcal{G}_i$, then

$$f(X) - f(X^i) \geq 0. \tag{9.6}$$

Now, for any $\omega \in \mathfrak{X}(B(n) \setminus B_i)$ for which the event

$$\begin{aligned}A_i := \{B_i \overset{2}{\underset{r}{\longleftrightarrow}} B(n),\ &\text{every } r\text{-cluster } \mathcal{C} \text{ in } B(n) \setminus B_i \text{ for which } \mathcal{C}^{(r)} \\ &\text{intersects both } \partial B(a_n) \text{ and } \partial(B(n)_{(r)}) \text{ has a point in } B_i^{(r)}\}\end{aligned}$$

is true, $\mathbb{I}_E(\omega \cup \omega_i) = 1$ when $\omega_i \in \mathcal{S}_i$ and $\mathbb{I}_E(\omega \cup \omega_i') = 0$ when $\omega_i' \in \mathcal{G}_i$. Consequently, if $X_i \in \mathcal{S}_i$ and $X_i' \in \mathcal{G}_i$ and $\mathbb{I}_{A_i}(\bigcup_{k \neq i} X_k) = 1$, then

$$f(X) - f(X^i) = 1.$$

Hence, from (9.5) and (9.6),

$$\begin{aligned}\operatorname{Var}[\mathbb{E}(f(X)|X_i)] &\geq \frac{1}{2}\mathbb{P}(A_i)^2 \cdot \mathbb{P}(X_i \in \mathcal{S}_i) \cdot \mathbb{P}(X_i' \in \mathcal{G}_i) \\ &\geq \frac{1}{2}\mathbb{P}(A_i)^2 \exp(-ca^{d-1}).\end{aligned} \tag{9.7}$$

The constant depends on d and r_1 only.

For $i \in \mathcal{I}$, we also have

$$\begin{aligned}\mathbb{P}(A_i) &\geq \mathbb{P}\big(B_i \xleftrightarrow[r]{2} B(n);\ \text{any } r\text{-cluster } \mathcal{C} \text{ in } B(n) \setminus B_i \\ &\qquad \text{for which } \mathcal{C}^{(r)} \text{ intersects both } \partial B\big(c(B_i), 2a_n\big) \\ &\qquad \text{and } \partial\big(B(n)_{(r)}\big) \text{ has a point in } B_i^{(r)}\big) \\ &\geq \mathbb{P}\big(B_i \xleftrightarrow[r]{2} B\big(c(B_i), 2n\big);\ \text{if } \mathcal{C} \text{ is an } r\text{-cluster in} \\ &\qquad B\big(c(B_i), 2n\big) \setminus B_i \text{ then every connected component} \\ &\qquad \text{of } \big(\mathcal{C} \cap B\big(c(B_i), n/2\big)\big)^{(r)} \text{ that intersects both} \\ &\qquad \partial B\big(c(B_i), 2a_n\big) \text{ and } \partial\big(B\big(c(B_i), n/2\big)_{(r)}\big) \text{ also intersects } \partial B_i\big).\end{aligned} \tag{9.8}$$

Define the event

$$\begin{aligned}F = \big\{&B_0 \xleftrightarrow[r]{2} B(2n),\ \text{if } \mathcal{C} \text{ is an } r\text{-cluster in } B(2n) \setminus B_0 \\ &\text{then every connected component of } \big(\mathcal{C} \cap B(n/2)\big)^{(r)} \text{ that} \\ &\text{intersects both } \partial B(2a_n) \text{ and } \partial\big(B(n/2)_{(r)}\big) \text{ also intersects } \partial B_0\big\}.\end{aligned}$$

From (9.4), (9.7), (9.8) and translational invariance, we get

$$\mathbb{P}(F) \leq \frac{c \exp(c' a^{d-1})}{\sqrt{|\mathcal{I}|}} \leq \frac{c'' \exp(c' a^{d-1}) a^{d/2}}{a_n^{d/2}}. \tag{9.9}$$

Here, we have used the fact that $\mathrm{Var}(f(X)) \leq 1/4$ and $|\mathcal{I}| = \Theta((a_n/a)^d)$.

On the event $\{B_0 \xleftrightarrow[r]{2} B(2n)\} \cap F^c$, we can find two disjoint r-clusters $\mathcal{C}_1, \mathcal{C}_2$ in $B(2n) \setminus B_0$ and an r-cluster $\overline{\mathcal{C}}$ (which may be the same as one of the r-clusters $\mathcal{C}_1, \mathcal{C}_2$) in $B(2n) \setminus B_0$ such that:

(i) each of $\mathcal{C}_1$ and $\mathcal{C}_2$ has a point in $B_0^{(r)}$ and a point in $B(2n)_{(2r)}$,

(ii) there is an r-cluster in $B(n/2) \setminus B_0$, call it $\overline{\mathcal{C}}'$, which is contained in $\overline{\mathcal{C}} \cap B(n/2)$, such that $\overline{\mathcal{C}}'$ has a point in $B(2a_n)^{(r)}$ and a point in $B(n/2)_{(2r)}$ but does not have a point in $B_0^{(r)}$.

So we can find two disjoint r-clusters $\mathcal{C}_1'$ and $\mathcal{C}_2'$ in $B(n/2) \setminus B_0$ that are contained in $\mathcal{C}_1 \cap B(n/2)$ and $\mathcal{C}_2 \cap B(n/2)$, respectively, such that $\mathcal{C}_1'$ and $\mathcal{C}_2'$ satisfy the requirements for $\{B_0 \xleftrightarrow[r]{2} B(n/2)\}$ to be true. Further, $\overline{\mathcal{C}}'$ is different from $\mathcal{C}_1'$ and $\mathcal{C}_2'$ since $\overline{\mathcal{C}}'$ does not have a point in $B_0^{(r)}$. Hence, the restrictions of $\overline{\mathcal{C}}', \mathcal{C}_1'$ and $\mathcal{C}_2'$ to $B(n/2) \setminus B(2a_n)$ will contain three disjoint r-clusters satisfying the requirements for $\{B(2a_n) \xrightarrow[r]{3} B(n/2)\}$ to be true.

Hence, we have

$$\mathbb{P}\big(B_0 \overset{2}{\underset{r}{\longleftrightarrow}} B(2n)\big) \le \mathbb{P}(F) + \mathbb{P}\big(B(2a_n) \overset{3}{\underset{r}{\longrightarrow}} B(n/2)\big). \tag{9.10}$$

All we need now is an upper bound for the second term on the right-hand side. We would like to apply a Burton–Keane type argument to get a bound for this term.

Assume that $\mathcal{C}_1, \mathcal{C}_2$ and $\mathcal{C}_3$ are three disjoint r-clusters in $B(n/2) \setminus B(2a_n)$ such that $\mathcal{C}_j^{(r)}$ intersects both $B(n/2)_{(r)}$ and $B(2a_n)^{(r)}$ and let x_j be the point in $\mathcal{C}_j$ closest to $B(2a_n)$ for $j = 1, 2, 3$.

If $x_j \in B(2a_n)^{(r)}$ for every j, then $B(2a_n) \overset{3}{\underset{r}{\longleftrightarrow}} B(n/2)$ holds true, and if $x_j \in B(2a_n)^{(2r)} \setminus B(2a_n)^{(r)}$ for every j, then $B(2a_n)^{(r)} \overset{3}{\underset{r}{\longleftrightarrow}} B(n/2)$ holds true.

Assume now that the event

$$\big\{B(2a_n) \overset{3}{\underset{r}{\longrightarrow}} B(n/2)\big\} \cap \big(\big\{B(2a_n) \overset{3}{\underset{r}{\longleftrightarrow}} B(n/2)\big\} \cup \big\{B(2a_n)^{(r)} \overset{3}{\underset{r}{\longleftrightarrow}} B(n/2)\big\}\big)^c$$

is true. Then the number of x_i's in $B(2a_n)^{(r)} \setminus B(2a_n)$ is one or two.

Let us assume that $x_1, x_2 \in B(2a_n)^{(r)}$ and $x_3 \in B(2a_n)^{(2r)} \setminus B(2a_n)^{(r)}$ (the other possibilities can be handled similarly). We can find a sequence of points $z_1^{(j)}, \dots, z_{k_j}^{(j)}$ in $\mathcal{C}_j$ for $j = 1, 2$ such that:

(i) $z_1^{(j)} \in B(2a_n)^{(r)}$ and $z_i^{(j)} \notin B(2a_n)^{(r)}$ if $i \ge 2$,

(ii) $z_{k_j}^{(j)} \in B(n/2)_{(2r)}$,

(iii) $d\big(z_i^{(j)}, z_{i+1}^{(j)}\big) \le 2r$ for $1 \le i \le k_j - 1$ and

(iv) $d\big(z_i^{(j)}, z_{i'}^{(j)}\big) > 2r$ whenever $i' \ge i + 2$.

Let $\mathcal{C}_j'(\subset \mathcal{C}_j)$ be the r-cluster in $B(n/2) \setminus B(2a_n)^{(r)}$ containing $\{z_2^{(j)}, \dots, z_{k_j}^{(j)}\}$. Note that

$$\max_{j=1,2} d\big(z_1^{(j)}, z_2^{(j)}\big) > r,$$

because otherwise the event $\{B(2a_n)^{(r)} \overset{3}{\underset{r}{\longleftrightarrow}} B(n/2)\}$ will be true (the r-clusters $\mathcal{C}_1', \mathcal{C}_2'$ and $\mathcal{C}_3$ will satisfy the requirements). If $\min_{j=1,2} d(z_1^{(j)}, z_2^{(j)}) \le r$ then $E_1(z_1^{(1)}) \cup E_1(z_1^{(2)})$ holds, where

$$E_1(x) := \big\{B(2a_n)^{(r)} \cup S(x, r) \overset{3}{\underset{r}{\longleftrightarrow}} B(n/2)\big\}$$

for $x \in B(2a_n)^{(r)}$ and if $\min_{j=1,2} d(z_1^{(j)}, z_2^{(j)}) > r$ then the event

$$E_2\big(z_1^{(1)}, z_1^{(2)}\big) := \big\{B(2a_n)^{(r)} \cup S\big(z_1^{(1)}, r\big) \cup S\big(z_1^{(2)}, r\big) \overset{3}{\underset{r}{\longleftrightarrow}} B(n/2)\big\},$$

holds; in each case, $\mathcal{C}_3$ and the appropriate r-clusters containing the points $\{z_2^{(j)}, \ldots, z_{k_j}^{(j)}\}$ $(j = 1, 2)$ satisfying the requirements. Hence,

$$\begin{aligned}
&\mathbb{P}\big(B(2a_n) \xrightarrow[r]{3} B(n/2)\big) \\
&\quad\le \mathbb{P}\big(B(2a_n) \xleftrightarrow[r]{3} B(n/2)\big) + \mathbb{P}\big(B(2a_n)^{(r)} \xleftrightarrow[r]{3} B(n/2)\big) \\
&\qquad + \mathbb{P}\big(\exists x, y \in \mathcal{P} \cap \big(B(2a_n)^{(r)} \setminus B(2a_n)\big) \\
&\qquad \text{such that } x \neq y \text{ and } E_2(x, y) \text{ holds}\big) \\
&\qquad + \mathbb{P}\big(\exists x \in \mathcal{P} \cap \big(B(2a_n)^{(r)} \setminus B(2a_n)\big) \text{ such that } E_1(x) \text{ holds}\big).
\end{aligned} \tag{9.11}$$

This gives

$$\begin{aligned}
&\mathbb{P}\big(B(2a_n) \xrightarrow[r]{3} B(n/2)\big) \\
&\quad\le \mathbb{P}\big(B(2a_n) \xleftrightarrow[r]{3} B(n/2)\big) + \mathbb{P}\big(B(2a_n)^{(r)} \xleftrightarrow[r]{3} B(n/2)\big) \\
&\qquad + \mathbb{E}\big|\mathcal{P} \cap \big(B(2a_n)^{(r)} \setminus B(2a_n)\big)\big|^2 \sup\nolimits_1 \mathbb{P}\big(E_2(x, y)\big) \\
&\qquad + \mathbb{E}\big|\mathcal{P} \cap \big(B(2a_n)^{(r)} \setminus B(2a_n)\big)\big| \sup\nolimits_2 \mathbb{P}\big(E_1(x)\big),
\end{aligned} \tag{9.12}$$

where $\sup_1$ (resp., $\sup_2$) is supremum taken over all x, y (resp., x) in $B(2a_n)^{(r)} \setminus B(2a_n)$. Lemma 9.2 helps us in estimating $\mathbb{P}(E_2(x, y))$ and $\mathbb{P}(E_1(x))$.

From (9.9), (9.10), (9.12) and Lemma 9.2, we get

$$\mathbb{P}\big(B_0 \xleftrightarrow[r]{2} B(2n)\big) \le c\left(\exp\big(c' a^{d-1}\big)\frac{a^{d/2}}{a_n^{d/2}} + \exp\big(c'' a_n\big)\frac{a_n^{3d-2}}{n}\right). \tag{9.13}$$

We choose a_n so that $c'' a_n = \frac{1}{2} \log n$, plug this into (9.13) and finally replace n by $n/2$ to get (5.1).

If we replace $B(a)$ in (5.1) by, say, $K = B(a)^{(r)} \cup S(x, r)$, then define $B_k := 2(a + 2r)k + K$ so that the sets B_k remain disjoint. Define $\mathcal{I}$ as before and think of f as a function of the configurations inside $\{B_k\}_{k \in \mathcal{I}}$ and the configuration in the complement of $\bigcup_{k \in \mathcal{I}} B_k$. The rest of the proof can be carried out by following the same arguments as before. This concludes the proof of Lemma 5.1.

9.2. *Proof of Lemma* 9.2. We start with some auxiliary lemmas. The following lemma is a restatement of Lemma 3.2 in [43].

LEMMA 9.3. *Let R be a finite non empty subset of a set S. Assume further that*:

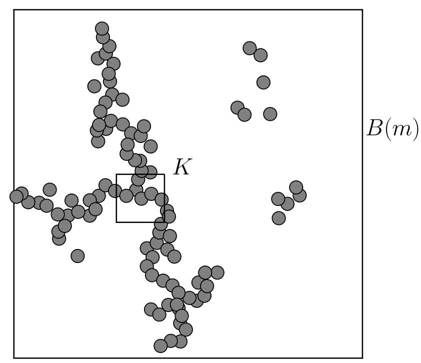

FIG. 5. *K is a trifurcation box in $B(m)$.*

(I) *for every $r \in R$, there exist pairwise disjoint subsets* (*which we call "branches"*) $C_r^{(1)}, \dots, C_r^{(m_r)}$ *of S and a positive integer k such that*

$$\text{(Ia)} \quad m_r \geq 3,$$

$$\text{(Ib)} \quad r \notin C_r^{(i)} \qquad \textit{for } i \leq m_r \quad \textit{and}$$

$$\text{(Ic)} \quad |C_r^{(i)}| \geq k \qquad \textit{for } i \leq m_r;$$

(II) *for all $r, r' \in R$, either*

$$\text{(IIa)} \quad \Big(\bigcup_{j \leq m_r} C_r^{(j)} \cup \{r\}\Big) \cap \Big(\bigcup_{i \leq m_{r'}} C_{r'}^{(i)} \cup \{r'\}\Big) = \varnothing \quad \textit{or}$$

$$\text{(IIb)} \quad \Big(\bigcup_{j \leq m_r} C_r^{(j)} \cup \{r\}\Big) \setminus C_r^{(j_0)} \subset C_{r'}^{(i_0)} \quad \textit{and}$$

$$\Big(\bigcup_{i \leq m_{r'}} C_{r'}^{(i)} \cup \{r'\}\Big) \setminus C_{r'}^{(i_0)} \subset C_r^{(j_0)} \qquad \textit{for some } i_0 \leq m_{r'} \textit{ and } j_0 \leq m_r.$$

Then $|S| \geq k|R|$.

Let $K \subset B(m)$ be a translate of $B(s)^{(t)}$ where s, t, m are as in the statement of Lemma 9.2.

We will say that K is a trifurcation box in $B(m)$ [in short "K T-box in $B(m)$"] at level r (Figure 5) if:

(i) there is an r-cluster $\mathcal{C}$ in $B(m)$ with $\mathcal{C} \cap K \neq \varnothing$ and

(ii) $\mathcal{C} \cap K^c$ contains at least three disjoint r-clusters in $B(m) \setminus K$ each having a point in $B(m)_{(2r)}$.

Let us define

$$\mathcal{T} := \{j \in \mathbb{Z}^d : 4(s+t)j + B(s)^{(t)} \subset B(m/4)\}$$

and denote $4(s+t)j + B(s)^{(t)}$ by K_j for $j \in \mathcal{T}$. Then we have the following.

LEMMA 9.4. *There exists a positive constant c depending only on r_2 such that*

$$\big|\{\mathcal{P} \cap B(m/2)\}\big| \ge cm\big|\{j \in \mathcal{T} : K_j \text{ T-box in } B(m/2)\}\big|. \tag{9.14}$$

PROOF. Set $S = \mathcal{P} \cap B(m/2)$. If K_j is a trifurcation box in $B(m/2)$ for some $j \in \mathcal{T}$, then there is an r-cluster $\mathcal{C}_j$ in $B(m/2)$ such that there is a point r_j in $\mathcal{C}_j \cap K_j$. Further, $\mathcal{C}_j \cap B(m/2) \setminus K_j$ contains $m_j(\ge 3)$ disjoint r-clusters, say $\mathcal{C}_j^{(1)}, \dots, \mathcal{C}_j^{(m_j)}$, each having a point in $B(m/2)_{(2r)}$. Call these clusters the "branches" of r_j. Set $R = \{r_j : j \in \mathcal{T}, K_j \text{ T-box in } B(m/2)\}$.

For any $r_j, r_{j'}$ in R, condition (IIa) of Lemma 9.3 holds if $\mathcal{C}_j$ and $\mathcal{C}_{j'}$ are disjoint and condition (IIb) holds otherwise. Also

$$|\mathcal{C}_{r_j}^{(i)}| \ge \frac{m/4 - 2r_2}{2r_2} \ge cm$$

for every $r_j \in R$ and $i \le m_j$. Hence, an application of Lemma 9.3 yields the result. □

We are now ready to prove Lemma 9.2. Note that

$$\begin{aligned}
&\mathbb{P}\big(B(s)^{(t)} \text{ T-box in } B(m)\big) \\
&\quad \ge \mathbb{P}\big(B(s)^{(t)} \overset{3}{\underset{r}{\longleftrightarrow}} B(m)\big) \\
&\qquad \times \mathbb{P}\big(B(s)^{(t)} \text{ T-box in } B(m) | B(s)^{(t)} \overset{3}{\underset{r}{\longleftrightarrow}} B(m)\big).
\end{aligned} \tag{9.15}$$

Now, given any $\eta \in \mathfrak{X}(B(m) \setminus B(s)^{(t)})$ for which the event

$$A := \big\{B(s)^{(t)} \overset{3}{\underset{r}{\longleftrightarrow}} B(m)\big\}$$

is true, we can ensure that the event $\{B(s)^{(t)} \text{ T-box in } B(m)\}$ happens just by placing enough Poisson points inside $B(s)^{(t)}$ so that at least three of the r-clusters in $B(m) \setminus B(s)^{(t)}$ satisfying the requirements for A to be true get connected to form a single component. Since this can be done by placing at least one Poisson point in each of at most $6d^{3/2}(s+t)/r_1$ cubes (of side length $r_1/\sqrt{d}$) inside $B(s)^{(t)}$,

$$\mathbb{P}\big(B(s)^{(t)} \text{ T-box in } B(m) | B(s)^{(t)} \overset{3}{\underset{r}{\longleftrightarrow}} B(m)\big) \ge \exp\big(-c(s+t)\big)$$

for a positive universal constant c depending only on r_1 and d. Plugging this into (9.15), we get

$$\mathbb{P}\big(B(s)^{(t)} \overset{3}{\underset{r}{\longleftrightarrow}} B(m)\big) \le \exp\big(c(s+t)\big) \cdot \mathbb{P}\big(B(s)^{(t)} \text{ T-box in } B(m)\big). \tag{9.16}$$

Taking expectation in (9.14), we get

$$\frac{m^{d-1}}{c2^d} \geq \sum_{j\in\mathcal{T}} \mathbb{P}\big(K_j \text{ T-box in } B(m/2)\big)$$
$$\geq \sum_{j\in\mathcal{T}} \mathbb{P}\big(K_j \text{ T-box in } 4(s+t)j + B(m)\big).$$

By translational invariance and the fact that $|\mathcal{T}| \cdot (s+t)^d = \Theta(m^d)$, we get

$$c' m^{d-1} \geq \frac{m^d}{(s+t)^d} \mathbb{P}\big(B(s)^{(t)} \text{ T-box in } B(m)\big) \tag{9.17}$$

and (9.3) follows if we plug this in (9.16).

The same type of arguments work when $B(s)^{(t)}$ is replaced by the other sets, so we do not repeat them.

9.3. *Estimates in different regimes.* We now collect the estimates on $\mathbb{P}(B(a) \xrightarrow[r]{2} B(n))$ in different regimes together in the following lemma.

LEMMA 9.5. *For positive numbers r_1, r_2 satisfying $r_1 < r_c(d) < r_2$ and $n \geq 2$, we have the following estimates*:

(i) *When $d = 2$ and $a \in [1/2, \log n]$,*

$$\mathbb{P}\big(B(a) \xrightarrow[r]{2} B(n)\big) \leq \begin{cases} c_{10}\exp(-c_{11}n), & \text{if } r \leq r_1, \\ c_{12}/n^{\beta}, & \text{if } r_1 < r \leq (\log n)^2, \end{cases}$$

where c_{10} and c_{11} depend only on r_1, and c_{12} and β are universal positive constants.

(ii) *When $d \geq 3$ and $a \in (1/2, (\log\log n)^{1/(d-1/2)})$,*

$$\mathbb{P}\big(B(a) \xrightarrow[r]{2} B(n)\big) \leq \begin{cases} c_{13}\exp(-c_{14}n), & \text{if } r \leq r_1, \\ c_{15}\dfrac{\exp(c_{16}a^{d-1})}{(\log n)^{d/2}}, & \text{if } r \in [r_1, r_2], \\ c_{17}\exp(-c_{18}n), & \text{if } r_2 \leq r \leq n/8. \end{cases} \tag{9.18}$$

The constants appearing here depend only on r_1, r_2 and d.

PROOF. The proof can be divided into different parts.

(A) *$r \leq r_1$ and $d \geq 2$*: Note that for any $r > 0$ and $d \geq 2$,

$$\begin{aligned} \big\{B(a) \xrightarrow[r]{2} B(n)\big\} &\subset \big\{B(a) \xrightarrow[r]{1} B(n)\big\} \\ &\subset \big\{B(a) \xleftrightarrow[r]{1} B(n)\big\} \cup \big\{B(a)^{(r)} \xleftrightarrow[r]{1} B(n)\big\}. \end{aligned} \tag{9.19}$$

That the last inclusion holds can be seen as follows. Consider an r-cluster $\mathcal{C}$ in $B(n) \setminus B(a)$ which has a point in both $B(a)^{(2r)}$ and $B(n)_{(2r)}$ and let $x \in \mathcal{C}$ be the point closest to $B(a)$. If $x \in B(a)^{(r)}$, then $\{B(a) \xleftrightarrow[r]{1} B(n)\}$ is true and if $x \in B(a)^{(2r)} \setminus B(a)^{(r)}$ then $\{B(a)^{(r)} \xleftrightarrow[r]{1} B(n)\}$ is true.

For any $r \leq r_1$ and $d \geq 2$, $\{B(a) \xleftrightarrow[r]{1} B(n)\} \subset \{B(a) \xleftrightarrow[r_1]{1} B(n)\}$ and a similar statement holds if we replace $B(a)$ by $B(a)^{(r)}$. If we fix a configuration in $B(n) \setminus B(a)$ [resp., $B(n) \setminus B(a)^{(r)}$] for which $\{B(a) \xleftrightarrow[r_1]{1} B(n)\}$ [resp., $\{B(a)^{(r)} \xleftrightarrow[r_1]{1} B(n)\}$] holds, we can connect any of the corresponding clusters to the origin by placing at least one Poisson point in at most $c(a + r_2)/r_1$ many cubes inside $B(a)$ [resp., $B(a)^{(r)}$] each of side length $\min(2a, r_1/\sqrt{d})$. Thus, if $\mu_{\mathcal{P}}$ is the probability measure corresponding to a Poisson process of intensity one with an extra point added at the origin, then

$$\mu_{\mathcal{P}}\bigl(\text{diameter}(\mathcal{C}_0) \geq n \text{ at level } r_1 | B(a) \xleftrightarrow[r_1]{1} B(n)\bigr) \geq c \exp\bigl(-c' a\bigr),$$

$\mathcal{C}_0$ being the occupied component containing the origin. A similar inequality holds for $\mu_{\mathcal{P}}(\text{diameter}(\mathcal{C}_0) \geq n$ at level $r_1 | B(a)^{(r)} \xleftrightarrow[r_1]{1} B(n))$. Hence, from (9.19), we get

$$\begin{aligned}
\mathbb{P}\bigl(B(a) \xrightarrow[r]{2} B(n)\bigr) &\leq \mathbb{P}\bigl(B(a) \xleftrightarrow[r_1]{1} B(n)\bigr) + \mathbb{P}\bigl(B(a)^{(r)} \xleftrightarrow[r_1]{1} B(n)\bigr) \\
&\leq c \exp\bigl(c' a\bigr) \cdot \mu_{\mathcal{P}}\bigl(\text{diameter}(\mathcal{C}_0) \geq n \text{ at level } r_1\bigr) \\
&\leq c \exp\bigl(c' a\bigr) \exp\bigl(-c'' n\bigr).
\end{aligned}$$

The last inequality is just an application of [43], equation (3.60).

(B) $r \in [r_1, r_c]$ *and* $d = 2$: In this case,

$$\begin{aligned}
\mathbb{P}\bigl(B(a) \xrightarrow[r]{2} B(n)\bigr) &\leq \mathbb{P}\bigl(B(a) \xleftrightarrow[r_c]{1} B(n)\bigr) + \mathbb{P}\bigl(B(a)^{(r)} \xleftrightarrow[r_c]{1} B(n)\bigr) \\
&\leq c/n^{\theta} \qquad \text{for some } \theta > 0.
\end{aligned}$$

The last inequality holds because of the following reason. First, note that

$$\begin{aligned}
g_{\ell}(r_c) &:= \mathbb{P}\bigl(\exists \text{ a vacant left-right crossing of } [0, \ell] \times [0, 3\ell] \text{ at level } r_c\bigr) \geq \kappa_0 \\
&:= \frac{1}{(9e)^{122}}
\end{aligned} \tag{9.20}$$

for every $\ell \geq r_c$. (This is true since otherwise there exists $\ell^{\star} \geq r_c$ for which (9.20) fails. By continuity of the function $g_{\ell^{\star}}$, we will be able to find $r < r_c$ such that $g_{\ell^{\star}}(r) < (9e)^{-122}$. Then by Lemma 4.1 of [43], the vacant component containing the origin is bounded almost surely which leads to a contradiction since $r < r_c$.)

Now (9.20) together with Lemma 4.4 of [43] and the RSW lemma for vacant crossings (see [53] or Theorem 4.2 in [43]) will yield

$$\mathbb{P}\big(\exists \text{ a vacant left-right crossing of } [0,3\ell]\times[0,\ell] \text{ at level } r_c\big) \geq \delta$$

for a positive constant δ and every ℓ bigger than a fixed threshold ℓ_0. It then follows from standard arguments that with probability at least $1-c/n^\theta$, a vacant circuit around $B(a+r_c)$ exists in $B(n)$ at level r_c. Hence, we get the desired upper bound on $\mathbb{P}(B(a) \xrightarrow[r]{2} B(n))$ for $r \in [r_1, r_c]$.

(C) $r \geq r_c$ and $d = 2$: In this case, the polynomial decay of $\mathbb{P}(B(a) \xrightarrow[r]{2} B(n))$ follows from the existence of occupied "circuits" at level r_c around $B(a)$. The argument for this is also standard. We will give an outline in the Appendix.

(D) $r \in [r_1, r_2]$ and $d \geq 3$: Fix $r \in [r_1, r_2]$ and assume that $\{B(a) \xrightarrow[r]{2} B(n)\}$ holds. Take any two disjoint clusters $\mathcal{C}_1$ and $\mathcal{C}_2$ in $B(n) \setminus B(a)$ each having a point in $B(a)^{(2r)}$ and $B(n)_{(2r)}$ and let $x_j \in \mathcal{C}_j$ be the point closest to $B(a)$. If $x_j \in B(a)^{(r)}$ for $j = 1, 2$, then the event $\{B(a) \xleftrightarrow[r]{2} B(n)\}$ is true, and if $x_j \in B(a)^{(2r)} \setminus B(a)^{(r)}$ for $j = 1, 2$ then the event $\{B(a)^{(r)} \xleftrightarrow[r]{2} B(n)\}$ is true.

Now, assume that the event

$$\big\{B(a) \xrightarrow[r]{2} B(n)\big\} \cap \big[\big\{B(a) \xleftrightarrow[r]{2} B(n)\big\} \cup \big\{B(a)^{(r)} \xleftrightarrow[r]{2} B(n)\big\}\big]^c$$

is true. Then each of the sets $B(a)^{(r)}$ and $B(a)^{(2r)} \setminus B(a)^{(r)}$ contain exactly one of the points x_1 and x_2.

By arguments similar to the ones leading to (9.12), we can show that in this case the event

$$E := \big\{\exists x \in \mathcal{P} \cap \big(B(a)^{(r)} \setminus B(a)\big) \text{ such that } S(x,r) \cup B(a)^{(r)} \xleftrightarrow[r]{2} B(n)\big\}$$

is true. For any realization $\eta = \{\eta_1, \dots, \eta_\ell\}$ of $\mathcal{P} \cap (B(a)^{(r)} \setminus B(a))$, we have

$$E \subset \bigcup_{j=1}^{\ell} \big\{S(\eta_j, r) \cup B(a)^{(r)} \xleftrightarrow[r]{2} B(n)\big\}.$$

Hence, from Lemma 5.1,

$$\begin{aligned}
\mathbb{P}(E) &\leq \frac{c_6 \exp(c_7 a^{d-1})}{(\log n)^{\frac{d}{2}}} \mathbb{E}\big|\mathcal{P} \cap \big(B(a)^{(r)} \setminus B(a)\big)\big| \\
&\leq \frac{c \exp(c_7 a^{d-1}) a^{d-1}}{(\log n)^{\frac{d}{2}}}.
\end{aligned}$$

From our earlier discussion and another application of Lemma 5.1,

$$\mathbb{P}\big(B(a) \xrightarrow[r]{2} B(n)\big) \le \mathbb{P}\big(B(a) \xleftrightarrow[r]{2} B(n)\big) + \mathbb{P}\big(B(a)^{(r)} \xleftrightarrow[r]{2} B(n)\big) + \mathbb{P}(E)$$
$$\le c_{15} \exp\big(c_{16} a^{d-1}\big)/(\log n)^{\frac{d}{2}}.$$

(E) $r_2 \le r \le n/8$ and $d \ge 3$: The exponential decay in this regime can be proven using standard slab technology; see, for example, the proof of Lemma 10.12 in [47]. (Lemma 10.12 in [47] is stated in the setup where each pair of Poisson points are connected if they are at distance at most one and the intensity of the Poisson process determines sub- or super-criticality. This result translated to our setup where the parameter r varies and the intensity of the Poisson process is kept fixed gives an upper bound for $\mathbb{P}(B(a) \xrightarrow[r]{2} B(n))$ for every fixed $r > r_c$; whereas the bound in (9.18) is uniform for $r_2 \le r \le n/8$. This can be justified as follows. While using slab technology in the supercritical regime, $\mathbb{P}(B(a) \xrightarrow[r]{2} B(n))$ is bounded by probability of an event which is decreasing in r. Thus, the bound on $\mathbb{P}(B(a) \xrightarrow[r_2]{2} B(n))$ obtained from the proof of Lemma 10.12 in [47] works for each $r \in [r_2, n/8]$.) We omit the details. □

10. Rate of convergence in the CLT for Euclidean MST.

Our goal in this section is to prove Theorem 2.1. As before, d will denote the dimension of the ambient space. Choose an integer K such that $(n-1)/2 \ge K \ge (n-2)/4$ and let $s = n/(2K+1)$. Thus, $s \in [1, 2]$. Write $\mathbb{R}^d$ as the union of cubes,

$$\mathbb{R}^d = \bigcup_{j \in \mathbb{Z}^d} B_j \qquad \text{where } B_j := 2sj + B(s).$$

Let $B(n) = \bigcup_{j \in \mathcal{L}} B_j$. Clearly, $\ell := |\mathcal{L}| = \Theta(n^d)$. Fix $\alpha \in (0, 1)$ and let

$$\tilde{B}_j := B\big(2sj, n^\alpha\big). \tag{10.1}$$

Further, define

$$B_j^\star := \begin{cases} B(2sj, a_n), & \text{if } d \ge 3, \\ B(2sj, \alpha \log n), & \text{if } d = 2, \end{cases}$$

where a_n is a sequence increasing to infinity in a way so that $a_n \le (\log \log n)^{1/(d-1/2)}$. (We will choose the sequence a_n appropriately later in the proof.) We first prove a result that will be crucial in the proof.

10.1. *Preliminary estimates.*

Let $\mathcal{P}$ be a Poisson process in $\mathbb{R}^d$ having intensity one, and let B_j, $\tilde{B}_j$ and $B_j^\star$ be as above. Define the event E_j as follows:

$$E_j := \big\{\mathcal{P} \text{ contains a wall around } B_j \text{ in } B_j^\star\big\}. \tag{10.2}$$

PROPOSITION 10.1. *For any bounded subset A of $\mathbb{R}^d$, set $\mathcal{H}(A) = M(\mathcal{P} \cap A)$. Then the following hold*:

(i) *For every j with $\|2sj\|_\infty \le n - n^\alpha$,*

$$\begin{aligned}&\mathbb{E}\big[\mathbb{I}_{E_j} \cdot \big|\big(\mathcal{H}\big(B(n)\big) - \mathcal{H}\big(B(n) \setminus B_j\big)\big) - \big(\mathcal{H}(\tilde{B}_j) - \mathcal{H}(\tilde{B}_j \setminus B_j)\big)\big|\big] \\ &\qquad \le \begin{cases} c\exp\big(c' a_n^{d-1}\big)(\log n)^{-d/2}, & \text{if } d \ge 3, \\ c(\log n)^3 n^{-\alpha\beta}, & \text{if } d = 2,\end{cases}\end{aligned} \tag{10.3}$$

where β is as in Lemma 9.5.

(ii) *Lower bound on variance*:

$$\liminf_n \frac{1}{n^d} \mathbb{E}\big(\mathcal{H}\big(B(n)\big) - \mathbb{E}\mathcal{H}\big(B(n)\big)\big)^2 > 0. \tag{10.4}$$

PROOF OF (10.3). We first deal with the case $d \ge 3$. Note that

$$\begin{aligned}&\mathbb{E}\big[\mathbb{I}_{E_j} \cdot \big|\big(\mathcal{H}\big(B(n)\big) - \mathcal{H}\big(B(n) \setminus B_j\big)\big) - \big(\mathcal{H}(\tilde{B}_j) - \mathcal{H}(\tilde{B}_j \setminus B_j)\big)\big|\big] \\ &\quad = \mathbb{E}\mathbb{E}_\eta\big[\mathbb{I}_{E_j} \cdot \big|\big(\mathcal{H}\big(B(n)\big) - \mathcal{H}\big(B(n) \setminus B_j\big)\big) - \big(\mathcal{H}(\tilde{B}_j) - \mathcal{H}(\tilde{B}_j \setminus B_j)\big)\big|\big],\end{aligned} \tag{10.5}$$

where $\mathbb{E}_\eta$ denotes expectation conditional on the event $\{\mathcal{P} \cap B_j^\star = \eta\}$.

Fix realizations η, ω_1 and ω_2 of $\mathcal{P}$ in $B_j^\star$, $\tilde{B}_j \setminus B_j^\star$ and $B(n) \setminus \tilde{B}_j$, respectively, for which the event E_j is true. If $|\eta \cap B_j| = 0$, then $\mathcal{H}(B(n)) - \mathcal{H}(B(n) \setminus B_j)$ and $\mathcal{H}(\tilde{B}_j) - \mathcal{H}(\tilde{B}_j \setminus B_j)$ are both zero. So let us assume $|\eta \cap B_j| > 0$, and write

$$\eta \cap B_j = \{v_1, \dots, v_m\} \quad \text{and} \quad \eta \cap \big(B_j^\star \setminus B_j\big) = \{p_1, \dots, p_r\}.$$

Let $\mathfrak{J}_0 = \varnothing$ and $\mathfrak{J}_i = \{v_1, \dots, v_i\}$ for $1 \le i \le m$. Then

$$\big(\mathcal{H}\big(B(n)\big) - \mathcal{H}\big(B(n) \setminus B_j\big)\big) - \big(\mathcal{H}(\tilde{B}_j) - \mathcal{H}(\tilde{B}_j \setminus B_j)\big) = \sum_{i=1}^m \delta_i, \tag{10.6}$$

where

$$\begin{aligned}\delta_i := &\big[M\big(\mathfrak{J}_i \cup \big(\mathcal{P} \cap \big(B(n) \setminus B_j\big)\big)\big) - M\big(\mathfrak{J}_{i-1} \cup \big(\mathcal{P} \cap \big(B(n) \setminus B_j\big)\big)\big)\big] \\ &- \big[M\big(\mathfrak{J}_i \cup \big(\mathcal{P} \cap (\tilde{B}_j \setminus B_j)\big)\big) - M\big(\mathfrak{J}_{i-1} \cup \big(\mathcal{P} \cap (\tilde{B}_j \setminus B_j)\big)\big)\big].\end{aligned}$$

To keep the notation simple, let us focus on δ_1. Note that since η contains a wall around B_j in $B_j^\star$, by Lemma 8.2, an MST on the complete graph on $\{v_1, p_1, \dots, p_r\} \cup \{\omega_1 \cup \omega_2\}$ (resp. $\{v_1, p_1, \dots, p_r\} \cup \omega_1$) cannot contain an edge of the form $\{v_1, p\}$ with $p \in \omega_1 \cup \omega_2$ (resp. $p \in \omega_1$). Thus, an MST on the complete graph on $\{v_1, p_1, \dots, p_r\} \cup \{\omega_1 \cup \omega_2\}$ (resp., $\{v_1, p_1, \dots, p_r\} \cup \omega_1$) can be obtained from an MST on $\{p_1, \dots, p_r\} \cup \{\omega_1 \cup \omega_2\}$ (resp., $\{p_1, \dots, p_r\} \cup \omega_1$) by introducing the edges $\{v_1, p_j\}$ one by one and deleting the edge with maximum weight in the resulting cycle to make sure all paths in the new tree are minimax, that is, by repeatedly using the add and delete algorithm (Section 6.2). We start with an MST T_0 (resp., $\tilde{T}_0$) on $\{p_1, \dots, p_r\} \cup \{\omega_1 \cup \omega_2\}$ (resp., $\{p_1, \dots, p_r\} \cup \omega_1$) with edge set E (resp., $\tilde{E}$) and proceed in the following manner.

Set $E_0 = E$ (resp., $\tilde{E}_0 = \tilde{E}$), $Y_0 = d(v_1, p_1)$ [resp., $\tilde{Y}_0 = d(v_1, p_1)$] and let w_0 (resp., $\tilde{w}_0$) be the weight of T_0 (resp., $\tilde{T}_0$). For $k = 1, \dots, r$:

(i) Introduce the edge $\{v_1, p_k\}$. If $k = 1$, there will be no cycles in $E_0 \cup \{v_1, p_1\}$ (resp., $\tilde{E}_0 \cup \{v_1, p_1\}$). In this case, set $E_1 = E_0 \cup \{v_1, p_1\}$ (resp., $\tilde{E}_1 = \tilde{E}_0 \cup \{v_1, p_1\}$). Otherwise, there will be a unique cycle in $E_{k-1} \cup \{v_1, p_k\}$ (resp., $\tilde{E}_{k-1} \cup \{v_1, p_k\}$) having $\{v_1, p_k\}$ as one of its edges. Delete the edge in this cycle with maximum weight and set E_k (resp., $\tilde{E}_k$) to be the resulting set of edges. If $k \le r - 1$, let Y_k (resp., $\tilde{Y}_k$) be the maximum edge weight in the path connecting v_1 and p_{k+1} in the resulting tree, T_k (resp., $\tilde{T}_k$) and let w_k (resp. $\tilde{w}_k$) be the total weight of T_k (resp., $\tilde{T}_k$).

(ii) If $k = r$, stop. Otherwise increase k by one and repeat step (i).

A consequence of Proposition 6.2 is that the tree we get at the end of this process is an MST on the graph which has $\{v_1, p_1, \dots, p_r\} \cup \{\omega_1 \cup \omega_2\}$ (resp., $\{v_1, p_1, \dots, p_r\} \cup \omega_1$) as its vertex set and contains every possible edge between these vertices except the ones of the form $\{v_1, p\}$ with $p \in \omega_1 \cup \omega_2$ (resp. $p \in \omega_1$). It is easy to see that the resulting tree is actually an MST on the complete graph on $\{v_1, p_1, \dots, p_r\} \cup \{\omega_1 \cup \omega_2\}$ (resp., $\{v_1, p_1, \dots, p_r\} \cup \omega_1$), because as argued before, an edge of the form $\{v_1, x\}$ with $x \notin B_j^\star$ cannot be present in an an MST since η contains a wall around B_j in $B_j^\star$.

Hence,

$$\delta_1 = (w_r - w_0) - (\tilde{w}_r - \tilde{w}_0) = \sum_{k=1}^{r} \bigl[(w_k - w_{k-1}) - (\tilde{w}_k - \tilde{w}_{k-1})\bigr]. \tag{10.7}$$

Now,

$$w_k - w_{k-1} = \begin{cases} d(v_1, p_1), & \text{if } k = 1, \\ d(v_1, p_k) - \max\bigl(Y_{k-1}, d(v_1, p_k)\bigr), & \text{if } 2 \le k \le r. \end{cases} \tag{10.8}$$

A similar statement holds for $\tilde{w}_k$ with $\tilde{Y}_{k-1}$ replacing Y_{k-1}. Proposition 6.2 shows that T_{k-1} (resp. $\tilde{T}_{k-1}$) is an MST on the graph with vertex set $\mathcal{V} = (\mathcal{P} \cap (B(n) \setminus B_j)) \cup \{v_1\}$ [resp., $\tilde{\mathcal{V}} = (\mathcal{P} \cap (\tilde{B}_j \setminus B_j)) \cup \{v_1\}$] and edge set $\mathcal{E}_{k-1} = \bigcup_{i=1}^{k-1} \{v_1, p_i\} \cup$ {edges in the complete graph on $\mathcal{P} \cap (B(n) \setminus B_j)$} [resp., $\tilde{\mathcal{E}}_{k-1} = \bigcup_{i=1}^{k-1} \{v_1, p_i\} \cup$ {edges in the complete graph on $\mathcal{P} \cap (\tilde{B}_j \setminus B_j)$}] for $k \ge 2$. Hence, Y_{k-1} (resp., $\tilde{Y}_{k-1}$) is the maximum edge-weight in a minimax path connecting v_1 and p_k in $(\mathcal{V}, \mathcal{E}_{k-1})$ [resp., $(\tilde{\mathcal{V}}, \tilde{\mathcal{E}}_{k-1})$]. This gives $Y_{k-1} \le \tilde{Y}_{k-1}$. From (10.8),

$$0 \le (w_k - w_{k-1}) - (\tilde{w}_k - \tilde{w}_{k-1}) \le \tilde{Y}_{k-1} - Y_{k-1}. \tag{10.9}$$

Consider a random variable U uniformly distributed on $(0, 2\sqrt{d}a_n)$ which is independent of $\mathcal{P}$. We have

$$\begin{aligned}
&\mathbb{E}_\eta\big|(w_k - w_{k-1}) - (\tilde{w}_k - \tilde{w}_{k-1})\big| \\
&\quad \le \mathbb{E}_\eta(\tilde{Y}_{k-1} - Y_{k-1}) = 2\sqrt{d}a_n \cdot \mathbb{P}_\eta(Y_{k-1} < U < \tilde{Y}_{k-1}) \qquad (10.10)\\
&\quad = \int_0^{2\sqrt{d}a_n} \mathbb{P}_\eta(Y_{k-1} < u < \tilde{Y}_{k-1})\,du \le \int_0^{2\sqrt{d}a_n} \mathbb{P}\big(B_j^\star \xrightarrow[u/2]{2} \tilde{B}_j\big)\,du.
\end{aligned}$$

The last inequality holds because of the following reason. Assume that $Y_{k-1} < u < \tilde{Y}_{k-1}$ and let $(v_1 = z_0, z_1, \dots, z_\ell = p_k)$ be a minimax path connecting v_1 and p_k in $(\mathcal{V}, \mathcal{E}_{k-1})$. Since $Y_{k-1} < \tilde{Y}_{k-1}$, $z_i \in \tilde{B}_j^c$ for some $i \le \ell$. Let $k_1 + 1 := \min\{i \le \ell : z_i \in \tilde{B}_j^c\}$ and $k_2 - 1 := \max\{i \le \ell : z_i \in \tilde{B}_j^c\}$. Then the $u/2$-clusters in $\tilde{B}_j \setminus B_j$ containing $\{z_1, \dots, z_{k_1}\}$ and $\{z_{k_2}, \dots, z_\ell\}$ are disjoint, since otherwise we could find a path $(z_i = y_0, y_1, \dots, y_t = z_{i'})$ for some $i \le k_1$, $i' \ge k_2$ such that $y_p \in \tilde{\mathcal{V}} \setminus \{v_1\}$ and $d(y_p, y_{p+1}) \le u$ for every $p \le t-1$. But this would mean that $(z_0, \dots, z_i, y_1, \dots, y_{t-1}, z_i', \dots, z_\ell)$ is a path in $(\tilde{\mathcal{V}}, \tilde{\mathcal{E}}_{k-1})$ connecting v_1 and p_k with maximum edge-weight strictly smaller than $\tilde{Y}_{k-1}$, a contradiction. Then the restrictions of the (disjoint) $u/2$-clusters in $\tilde{B}_j \setminus B_j$ containing $\{z_1, \dots, z_{k_1}\}$ and $\{z_{k_2}, \dots, z_\ell\}$ to $\tilde{B}_j \setminus B_j^\star$ will contain two disjoint $u/2$-clusters which will satisfy the criteria for $\{B_j^\star \xrightarrow[u/2]{2} \tilde{B}_j\}$ to hold.

Combining (10.7) and (10.10),

$$\mathbb{E}_\eta[\delta_1] \le 2\sqrt{d}a_n \sup_{0<u<2\sqrt{d}a_n} \mathbb{P}\big(B_j^\star \xrightarrow[u/2]{2} \tilde{B}_j\big) \cdot \big(|\mathcal{P} \cap B_j^\star|\big). \qquad (10.11)$$

Inductively, having obtained an MST on $\mathfrak{J}_i \cup (\mathcal{P} \cap (B(n) \setminus B_j))$ [resp., $\mathfrak{J}_i \cup (\mathcal{P} \cap (\tilde{B}_j \setminus B_j))$], $1 \le i \le m-1$, an MST on $\mathfrak{J}_{i+1} \cup (\mathcal{P} \cap (B(n) \setminus B_j))$ [resp., $\mathfrak{J}_{i+1} \cup (\mathcal{P} \cap (\tilde{B}_j \setminus B_j))$] can be obtained by introducing the edges $\{v_{i+1}, p_j\}$, $1 \le j \le r$, and $\{v_{i+1}, v_s\}$, $1 \le s \le i$, one by one and again using the *add and delete algorithm*. Thus, δ_{i+1} will have a decomposition similar to (10.7) that has $r + i \le |\mathcal{P} \cap B_j^\star|$ terms, and each of these terms will obey the bound on the right side of (10.10). Hence, for each $i \le m$, (10.11) will continue to hold for $\mathbb{E}_\eta[\delta_i]$.

Combining this observation with (10.5) and (10.6), we get

$$\begin{aligned}
&\mathbb{E}\big[\mathbb{I}_{E_j} \cdot \big|\big(\mathcal{H}(B(n)) - \mathcal{H}(B(n) \setminus B_j)\big) - \big(\mathcal{H}(\tilde{B}_j) - \mathcal{H}(\tilde{B}_j \setminus B_j)\big)\big|\big] \\
&\quad \le 2\sqrt{d}a_n \sup_{0<u<2\sqrt{d}a_n} \mathbb{P}\big(B_j^\star \xrightarrow[u/2]{2} \tilde{B}_j\big) \cdot \mathbb{E}\big(|\mathcal{P} \cap B_j| \cdot |\mathcal{P} \cap B_j^\star|\big) \qquad (10.12)\\
&\quad \le c a_n^{d+1} \sup_{0<u<2\sqrt{d}a_n} \mathbb{P}\big(B_j^\star \xrightarrow[u/2]{2} \tilde{B}_j\big) \le c\frac{\exp(c' a_n^{d-1})}{(\log n)^{d/2}},
\end{aligned}$$

where the last step follows from Lemma 9.5. This completes the proof for the case $d \geq 3$.

When $d = 2$, we can proceed in the exact same manner and the only difference is the percolation estimate from Lemma 9.5. Thus, when $d = 2$,

$$\mathbb{E}\big[\mathbb{I}_{E_j} \cdot \big|\big(\mathcal{H}\big(B(n)\big) - \mathcal{H}\big(B(n) \setminus B_j\big)\big) - \big(\mathcal{H}(\tilde{B}_j) - \mathcal{H}(\tilde{B}_j \setminus B_j)\big)\big|\big]$$

$$\leq 2\sqrt{2} \cdot \alpha \log n \sup_{0<u<2\sqrt{2}\cdot\alpha \log n} \mathbb{P}\big(B_j^\star \xrightarrow[u/2]{2} \tilde{B}_j\big) \cdot \mathbb{E}\big(|\mathcal{P} \cap B_j| \cdot \big|\mathcal{P} \cap B_j^\star\big|\big) \tag{10.13}$$

$$\leq c(\log n)^3 \sup_{0<u<2\sqrt{2}\cdot\alpha \log n} \mathbb{P}\big(B_j^\star \xrightarrow[u/2]{2} \tilde{B}_j\big) \leq c' \frac{(\log n)^3}{n^{\alpha\beta}}.$$

This completes the proof of (10.3). □

PROOF OF (10.4). This is implicit in the work of Kesten and Lee in [36]. Let us write $\mathcal{L} = \{j_1, \dots, j_l\}$ [recall the definition of $\mathcal{L}$ from around (10.1)]. Define the sigma-fields $\mathcal{F}_k := \sigma\{\mathcal{P} \cap B_{j_i} : i \leq k\}$ for $k = 1, \dots, \ell$, and let $\mathcal{F}_0$ be the trivial sigma-field. Then we can express $\mathcal{H}(B(n)) - \mathbb{E}\mathcal{H}(B(n))$ as a sum of martingale differences:

$$\mathcal{H}\big(B(n)\big) - \mathbb{E}\mathcal{H}\big(B(n)\big) = \sum_{k=1}^{\ell} Z_k,$$

$$\text{where } Z_k := \mathbb{E}\big(\mathcal{H}\big(B(n)\big)|\mathcal{F}_k\big) - \mathbb{E}\big(\mathcal{H}\big(B(n)\big)|\mathcal{F}_{k-1}\big).$$

From [36], equation (4.27), it will follow that

$$\frac{1}{\ell} \sum_{k=1}^{\ell} Z_k^2 \xrightarrow{P} \zeta,$$

for a positive constant ζ. An application of Fatou's lemma together with fact $\ell = \Theta(n^d)$ yields

$$\liminf_n \frac{1}{n^d} \mathbb{E}\big(\mathcal{H}\big(B(n)\big) - \mathbb{E}\mathcal{H}\big(B(n)\big)\big)^2 > 0,$$

as desired. □

10.2. *Proof of Theorem* 2.1. At this point, we ask the reader to recall the notation used in Section 3.1. Consider two independent Poisson process $\mathcal{P}$ and $\mathcal{P}'$ having intensity one in $\mathbb{R}^d$. We will apply (3.1) and (3.2) with

$$X_j := \mathcal{P} \cap B_j, \qquad X'_j := \mathcal{P}' \cap B_j,$$
$$X := (X_j : j \in \mathcal{L}), \qquad X' := (X'_j : j \in \mathcal{L}),$$

and the function $f : \prod_{i \in \mathcal{L}} \mathfrak{X}(B_i) \to \mathbb{R}$ given by

$$f(\{\omega_i : i \in \mathcal{L}\}) = M\Big(\bigcup_{i \in \mathcal{L}} \omega_i\Big).$$

By definition, for any $A \subset \mathcal{L}$, X^A is a random vector whose ith coordinate is a configuration in B_i, $i \in \mathcal{L}$, but there is also a natural way of identifying X^A with a configuration in $B(n)$, and we will often blur the distinction between the two to simplify notation. In particular, with this convention, $X \cap R$ will represent a configuration in R for any $R \subset \mathbb{R}^d$, and $M(X)$ will be synonymous with $M(\bigcup_{i \in \mathcal{L}} X_i)$. We will use the shorthand $\Delta_j f(X^A) := \Delta_j f(X^A, X')$. Thus,

$$\Delta_j f(X^A) := f(X^A) - f(X^{A \cup \{j\}}),$$

for $A \subset \mathcal{L}$.

We first focus on proving the bounds on the Kantorovich–Wasserstein distance. Bounds of the same order in the Kolmogorov distance can be obtained in an almost identical fashion, and we will briefly comment on this at the end.

Bounds on the Kantorovich–Wasserstein distance. We will use Theorem 3.1 to prove bounds on the Kantorovich–Wasserstein distance. Note that $X \cap (B(n) \setminus B_j) = X^j \cap (B(n) \setminus B_j)$, and hence

$$\Delta_j f(X) = [M(X) - M(X \cap (B(n) \setminus B_j))] - [M(X^j) - M(X^j \cap (B(n) \setminus B_j))]$$

for every $j \in \mathcal{L}$. Lemma 8.6 and the fact $s \in [1, 2]$ imply that for every $j \in \mathcal{L}$ and $q \geq 1$,

$$\mathbb{E}|\Delta_j f(X)|^q \leq C'_q, \tag{10.14}$$

for constants C'_q depending only on d and q. Here, we make note of two direct consequences of (10.14). First,

$$|\mathrm{Cov}(\Delta_j f(X)\Delta_j f(X^A), \Delta_{j'} f(X)\Delta_{j'} f(X^{A'}))| \leq C_{10.15} \tag{10.15}$$

for any $j, j' \in \mathcal{L}$ and $A, A' \subset \mathcal{L}$, where $C_{10.15}$ is a finite constant. Second, (10.14) combined with (10.4) and the fact $\ell = |\mathcal{L}| = \Theta(n^d)$ yields

$$\frac{1}{\mathrm{Var}(f(X))^{3/2}} \sum_{j=1}^{\ell} \mathbb{E}|\Delta_j f(X)|^3 \leq \frac{c}{n^{d/2}}. \tag{10.16}$$

This gives us control over the second term on the right-hand side of (3.1). Our aim in the remainder of the proof is to bound $\mathrm{Var}(\mathbb{E}(T|W))$. We first focus on the case $d \geq 3$.

PROOF OF (2.4). We plan to show that the covariance term appearing in the numerator on the right side of (3.3) is small when j and j' are "far away." With this in mind, we break up the sum on the right-hand side of (3.3) into two parts $\sum_1$

and $\sum_2$; $\sum_1$ denotes the sum over all $(j, j', A, A') \in \mathfrak{E}^{(\alpha)}$ [for some $\alpha \in (0,1)$], where

$$\begin{aligned}\mathfrak{E}^{(\alpha)} := \{(j, j', A, A') : A, A' \subsetneq \mathcal{L};\ j \in \mathcal{L} \setminus A,\ j' \in \mathcal{L} \setminus A' \text{ and either}\\ \|j - j'\|_\infty \le n^\alpha \text{ or } \|2sj\|_\infty > (n - n^\alpha) \text{ or } \|2sj'\|_\infty > (n - n^\alpha)\}\end{aligned} \tag{10.17}$$

and $\sum_2$ denotes the sum over the remaining terms, that is, all $(j, j', A, A') \in \mathfrak{F}^{(\alpha)}$ where

$$\mathfrak{F}^{(\alpha)} := \{(j, j', A, A') : A, A' \subsetneq \mathcal{L};\ j \in \mathcal{L} \setminus A,\ j' \in \mathcal{L} \setminus A'\} \setminus \mathfrak{E}^{(\alpha)}.$$

Let $\mathfrak{E}_{1,2}^{(\alpha)}$ be the collection of all (j, j') for which $(j, j', \varnothing, \varnothing) \in \mathfrak{E}^{(\alpha)}$. Then, from (10.15),

$$\begin{aligned}&\sum\nolimits_1 \frac{\operatorname{Cov}(\Delta_j f(X)\Delta_j f(X^A), \Delta_{j'} f(X)\Delta_{j'} f(X^{A'}))}{\binom{\ell}{|A|}(\ell - |A|)\binom{\ell}{|A'|}(\ell - |A'|)}\\ &\quad\le C_{10.15} \sum_{(j,j') \in \mathfrak{E}_{1,2}^{(\alpha)}} \sum_{\substack{A \not\ni j\\ A' \not\ni j'}} \left(\binom{\ell}{|A|}(\ell - |A|)\binom{\ell}{|A'|}(\ell - |A'|)\right)^{-1}\\ &\quad= C_{10.15} \sum_{(j,j') \in \mathfrak{E}_{1,2}^{(\alpha)}} \sum_{k,k'=0}^{\ell-1} \sum_{\substack{A \not\ni j, A' \not\ni j'\\ |A|=k, |A'|=k'}} \left(\binom{\ell}{|A|}(\ell - |A|)\binom{\ell}{|A'|}(\ell - |A'|)\right)^{-1}\\ &\quad= C_{10.15}|\mathfrak{E}_{1,2}^{(\alpha)}| \le c(n^{2d-1} \cdot n^\alpha + n^d \cdot n^{\alpha d}) \le c' n^{2d-1+\alpha}.\end{aligned} \tag{10.18}$$

We now turn to the sum $\sum_2$. Note that for $(j, j') \notin \mathfrak{E}_{1,2}^{(\alpha)}$, $\|2sj - 2sj'\|_\infty > 2sn^\alpha \ge 2n^\alpha$, and so the cubes $\tilde{B}_j$ and $\tilde{B}_{j'}$ are disjoint [recall the definition from (10.1)]. As a result, the restrictions of X (and of X') to these cubes are independent. Let us now define

$$\tilde{\Delta}_j f(X^A) := M(X^A \cap \tilde{B}_j) - M(X^{A \cup \{j\}} \cap \tilde{B}_j) \tag{10.19}$$

for every j with $\|2sj\|_\infty \le (n - n^\alpha)$ and $A \subset \mathcal{L}$. Whenever $(j, j', A, A') \in \mathfrak{F}^{(\alpha)}$, we have

$$\begin{aligned}&\operatorname{Cov}(\Delta_j f(X)\Delta_j f(X^A), \Delta_{j'} f(X)\Delta_{j'} f(X^{A'}))\\ &\quad= \operatorname{Cov}([\Delta_j f(X) - \tilde{\Delta}_j f(X)]\Delta_j f(X^A), \Delta_{j'} f(X)\Delta_{j'} f(X^{A'}))\\ &\qquad+ \operatorname{Cov}(\tilde{\Delta}_j f(X)[\Delta_j f(X^A) - \tilde{\Delta}_j f(X^A)], \Delta_{j'} f(X)\Delta_{j'} f(X^{A'}))\\ &\qquad+ \operatorname{Cov}(\tilde{\Delta}_j f(X)\tilde{\Delta}_j f(X^A), [\Delta_{j'} f(X) - \tilde{\Delta}_{j'} f(X)]\Delta_{j'} f(X^{A'}))\\ &\qquad+ \operatorname{Cov}(\tilde{\Delta}_j f(X)\tilde{\Delta}_j f(X^A), \tilde{\Delta}_{j'} f(X)[\Delta_{j'} f(X^{A'}) - \tilde{\Delta}_{j'} f(X^{A'})]).\end{aligned} \tag{10.20}$$

We will give an upper bound for the first term on the right-hand side of (10.20). The other terms can be dealt with in a similar fashion. Note that

$$
\begin{aligned}
&\mathrm{Cov}\big([\Delta_j f(X) - \tilde{\Delta}_j f(X)]\Delta_j f(X^A), \Delta_{j'} f(X)\Delta_{j'} f(X^{A'})\big) \\
&\quad \le \mathbb{E}\big(\big|(\Delta_j f(X) - \tilde{\Delta}_j f(X))\Delta_j f(X^A)\Delta_{j'} f(X)\Delta_{j'} f(X^{A'})\big|\big) \\
&\qquad + \mathbb{E}\big(\big|(\Delta_j f(X) - \tilde{\Delta}_j f(X))\Delta_j f(X^A)\big|\big) \cdot \mathbb{E}\big(\big|\Delta_{j'} f(X)\Delta_{j'} f(X^{A'})\big|\big) \\
&\quad =: T_1 + T_2.
\end{aligned} \tag{10.21}
$$

Then for any $p, q > 1$ satisfying $p^{-1} + q^{-1} = 1$, we have from (10.14) that

$$
\begin{aligned}
T_2 &\le C_2'\big(\mathbb{E}\big|(\Delta_j f(X) - \tilde{\Delta}_j f(X))\big|\big)^{1/p} \\
&\qquad \times \big(\mathbb{E}\big|(\Delta_j f(X) - \tilde{\Delta}_j f(X))\big|\Delta_j f(X^A)\big|^q\big|\big)^{1/q} \\
&\le c\big(\mathbb{E}\big|(\Delta_j f(X) - \tilde{\Delta}_j f(X))\big|\big)^{1/p}.
\end{aligned}
$$

A similar bound holds for T_1. We plug all these estimates into (10.20) to get

$$
\begin{aligned}
&\mathrm{Cov}\big(\Delta_j f(X)\Delta_j f(X^A), \Delta_{j'} f(X)\Delta_{j'} f(X^{A'})\big) \\
&\quad \le c\big(\big(\mathbb{E}\big|(\Delta_j f(X) - \tilde{\Delta}_j f(X))\big|\big)^{\frac{1}{p}} \\
&\qquad + \big(\mathbb{E}\big|(\Delta_{j'} f(X) - \tilde{\Delta}_{j'} f(X))\big|\big)^{\frac{1}{p}}\big),
\end{aligned} \tag{10.22}
$$

for $(j, j', A, A') \in \mathfrak{F}^{(\alpha)}$. Let E_j be the event in (10.2). Then (10.14) and Lemma 8.6 yield

$$
\begin{aligned}
\mathbb{E}\big(\mathbb{I}_{E_j^c}\big|(\Delta_j f(X) - \tilde{\Delta}_j f(X))\big|\big) &\le \big(\mathbb{E}\big|(\Delta_j f(X) - \tilde{\Delta}_j f(X))\big|^2\big)^{\frac{1}{2}}\mathbb{P}\big(E_j^c\big)^{\frac{1}{2}} \\
&\le c\exp\big(-c_{10.23}a_n^d\big)
\end{aligned} \tag{10.23}
$$

for every j with $\|2sj\| \le n - n^\alpha$. Hence, for every j with $\|2sj\| \le n - n^\alpha$,

$$
\begin{aligned}
&\mathbb{E}\big|(\Delta_j f(X) - \tilde{\Delta}_j f(X))\big| \\
&\quad \le c\exp\big(-c_{10.23}a_n^d\big) + \mathbb{E}\big(\mathbb{I}_{E_j} \cdot \big|(\Delta_j f(X) - \tilde{\Delta}_j f(X))\big|\big).
\end{aligned} \tag{10.24}
$$

Since the restrictions of the vectors X and X^j to $B(n) \setminus B_j$ (resp., $\tilde{B}_j \setminus B_j$) are the same,

$$
M\big(X \cap \big(B(n) \setminus B_j\big)\big) = M\big(X^j \cap \big(B(n) \setminus B_j\big)\big) \quad \text{and}
$$
$$
M\big(X \cap (\tilde{B}_j \setminus B_j)\big) = M\big(X^j \cap (\tilde{B}_j \setminus B_j)\big).
$$

Hence, we can write, for every j with $\|2sj\| \le n - n^\alpha$,

$$
\begin{aligned}
&\Delta_j f(X) - \tilde{\Delta}_j f(X) \\
&\quad = \big[\big(M(X) - M\big(X \cap \big(B(n) \setminus B_j\big)\big)\big) - \big(M(X \cap \tilde{B}_j) - M\big(X \cap (\tilde{B}_j \setminus B_j)\big)\big)\big]
\end{aligned}
$$

$$
\begin{aligned}
&- \big[\big(M\big(X^j\big) - M\big(X^j \cap \big(B(n) \setminus B_j\big)\big)\big) \\
&- \big(M\big(X^j \cap \tilde{B}_j\big) - M\big(X^j \cap (\tilde{B}_j \setminus B_j)\big)\big)\big].
\end{aligned}
$$

Therefore, for every j with $\|2sj\| \le n - n^\alpha$,

$$
\begin{aligned}
&\mathbb{E}\big[\mathbb{I}_{E_j}\big|\Delta_j f(X) - \tilde{\Delta}_j f(X)\big|\big] \\
&\quad \le 2\mathbb{E}\big[\mathbb{I}_{E_j} \cdot \big|\big(M(X) - M\big(X \cap \big(B(n) \setminus B_j\big)\big)\big) \\
&\qquad - \big(M(X \cap \tilde{B}_j) - M\big(X \cap (\tilde{B}_j \setminus B_j)\big)\big)\big|\big].
\end{aligned}
\tag{10.25}
$$

Using (10.3), we conclude that

$$
\mathbb{E}\big[\mathbb{I}_{E_j}\big|\Delta_j f(X) - \tilde{\Delta}_j f(X)\big|\big] \le c\frac{\exp(c' a_n^{d-1})}{(\log n)^{d/2}}. \tag{10.26}
$$

In view of (10.24), we choose a_n so that $c_{10.23} a_n^d = \frac{d}{2} \log\log n$ to get

$$
\mathbb{E}\big|\Delta_j f(X) - \tilde{\Delta}_j f(X)\big| \le c\frac{\exp(c''(\log\log n)^{\frac{d-1}{d}})}{(\log n)^{d/2}} \tag{10.27}
$$

for every j with $\|2sj\| \le n - n^\alpha$. Hence,

$$
\begin{aligned}
&\sum\nolimits_2 \frac{\mathrm{Cov}(\Delta_j f(X)\Delta_j f(X^A), \Delta_{j'} f(X)\Delta_{j'} f(X^{A'}))}{\binom{\ell}{|A|}(\ell - |A|)\binom{\ell}{|A'|}(\ell - |A'|)} \\
&\quad \le cn^{2d} \max_{\mathfrak{F}^{(\alpha)}} \mathrm{Cov}\big(\Delta_j f(X)\Delta_j f\big(X^A\big), \Delta_{j'} f(X)\Delta_{j'} f\big(X^{A'}\big)\big) \\
&\quad \le cn^{2d} \frac{\exp(c'' \cdot (\log\log n)^{\frac{d-1}{d}}/p)}{(\log n)^{\frac{d}{2p}}},
\end{aligned}
\tag{10.28}
$$

where the last inequality is a consequence of (10.22) and (10.27). Combining (3.3), (10.18) and (10.28) and observing that (10.28) is true for any $p > 1$, we get

$$
\mathrm{Var}\big(\mathbb{E}(T|W)\big) \le cn^{2d}(\log n)^{-\frac{d}{2p}}. \tag{10.29}
$$

Combining (3.1), (10.4), (10.29) and (10.16), we see that there exists a positive constant c depending on p and d such that

$$
\mathcal{W}(\mu_n, \gamma) \le c(\log n)^{-\frac{d}{4p}}, \tag{10.30}
$$

which is the bound claimed in (2.4). $\square$

Let us now turn to the case $d = 2$.

PROOF OF (2.3). Let $\mathfrak{E}^{(\alpha)}, \mathfrak{F}^{(\alpha)}, \sum_1$ and $\sum_2$ be as defined around (10.17). (Later we will make a suitable choice of α.) The calculation in (10.18) gives

$$\sum\nolimits_1 \frac{\operatorname{Cov}(\Delta_j f(X)\Delta_j f(X^A), \Delta_{j'} f(X)\Delta_{j'} f(X^{A'}))}{\binom{\ell}{|A|}(\ell-|A|)\binom{\ell}{|A'|}(\ell-|A'|)} \le cn^{3+\alpha}. \tag{10.31}$$

We now bound the sum $\sum_2$. First, recall the definition of $\tilde{B}_j$ from (10.1) and let E_j be the event in (10.2). With $\tilde{\Delta}_j f(X)$ as in (10.19) [defined for j with $\|2sj\|_\infty \le (n-n^\alpha)$], (10.20), (10.21) and (10.22) continue to hold. Further, the bound (10.23) now reads

$$\mathbb{E}\big(\mathbb{I}_{E_j{}^c}\big|(\Delta_j f(X) - \tilde{\Delta}_j f(X))\big|\big) \le c\exp\big(-c'(\log n)^2\big) \tag{10.32}$$

for every j with $\|2sj\|_\infty \le (n-n^\alpha)$, and (10.3) combined with (10.25) gives

$$\mathbb{E}\big[\mathbb{I}_{E_j}\big|\Delta_j f(X) - \tilde{\Delta}_j f(X)\big|\big] \le c(\log n)^3 n^{-\alpha\beta}. \tag{10.33}$$

Combining (10.32) and (10.33), we arrive at

$$\mathbb{E}\big|\Delta_j f(X) - \tilde{\Delta}_j f(X)\big| \le c(\log n)^3 n^{-\alpha\beta}$$

for every j with $\|2sj\|_\infty \le (n-n^\alpha)$. Arguments similar to the ones used previously for $d \ge 3$ [see (10.28) and (10.22)] now yield

$$\sum\nolimits_2 \frac{\operatorname{Cov}(\Delta_j f(X)\Delta_j f(X^A), \Delta_{j'} f(X)\Delta_{j'} f(X^{A'}))}{\binom{\ell}{|A|}(\ell-|A|)\binom{\ell}{|A'|}(\ell-|A'|)} \le cn^4/n^{\frac{\alpha\beta}{p}}. \tag{10.34}$$

Combining (10.31) and (10.34) and taking $\alpha = p/(\beta+p)$, we get

$$\operatorname{Var}\big(\mathbb{E}(T|W)\big) \le cn^{\frac{4p+3\beta}{\beta+p}}. \tag{10.35}$$

Combining (10.4), (10.16) (with $d=2$), (10.35) and (3.1), and noting that $(4p+3\beta)/(\beta+p) < 4$, we get the bound in (2.3) in the Kantorovich–Wasserstein distance. $\square$

Bounds on the Kolmogorov distance. We can use Theorem 3.2 to prove bounds on the Kolmogorov distance in an almost identical fashion. The difference in the bound in (3.2) comes from the terms T'_A, which are sums of terms of the form $\Delta_j f(X,X')|\Delta_j f(X^A,X')|$ [instead of $\Delta_j f(X,X')\Delta_j f(X^A,X')$ as in Theorem 3.1]. To take this into account, we modify (10.20) as follows:

$$\begin{aligned}
&\operatorname{Cov}\big(\Delta_j f(X)|\Delta_j f(X^A)|, \Delta_{j'} f(X)|\Delta_{j'} f(X^{A'})|\big)\\
&\quad= \operatorname{Cov}\big((\Delta_j f(X) - \tilde{\Delta}_j f(X))|\Delta_j f(X^A)|, \Delta_{j'} f(X)|\Delta_{j'} f(X^{A'})|\big)\\
&\qquad+ \operatorname{Cov}\big(\tilde{\Delta}_j f(X)(|\Delta_j f(X^A)| - |\tilde{\Delta}_j f(X^A)|), \Delta_{j'} f(X)|\Delta_{j'} f(X^{A'})|\big)\\
&\qquad+ \operatorname{Cov}\big(\tilde{\Delta}_j f(X)|\tilde{\Delta}_j f(X^A)|, (\Delta_{j'} f(X) - \tilde{\Delta}_{j'} f(X))|\Delta_{j'} f(X^{A'})|\big)\\
&\qquad+ \operatorname{Cov}\big(\tilde{\Delta}_j f(X)|\tilde{\Delta}_j f(X^A)|, \tilde{\Delta}_{j'} f(X)(|\Delta_{j'} f(X^{A'})| - |\tilde{\Delta}_{j'} f(X^{A'})|)\big),
\end{aligned}$$

whenever $(j, j', A, A') \in \mathfrak{F}^{(\alpha)}$. Noting that

$$||\Delta_j f(X^A)| - |\tilde{\Delta}_j f(X^A)|| \leq |\Delta_j f(X^A) - \tilde{\Delta}_j f(X^A)|,$$

it is easy to see that a bound similar to (10.22) continues to hold:

$$\begin{aligned}
&\mathrm{Cov}\big(\Delta_j f(X)\big|\Delta_j f(X^A)\big|, \Delta_{j'} f(X)\big|\Delta_{j'} f(X^{A'})\big|\big) \\
&\qquad \leq c\big(\big(\mathbb{E}\big|\big(\Delta_j f(X) - \tilde{\Delta}_j f(X)\big)\big|\big)^{\frac{1}{p}} + \big(\mathbb{E}\big|\big(\Delta_{j'} f(X) - \tilde{\Delta}_{j'} f(X)\big)\big|\big)^{\frac{1}{p}}\big).
\end{aligned}$$

The rest of the analysis can be carried out in the exact same way to get bounds on the Kolmogorov distance. This completes the proof of Theorem 2.1.

11. Percolation estimates in the lattice setup. We will now give an analogue of Lemma 9.5.

LEMMA 11.1. *Assume that $d \geq 2$, $p_1 \in (0, p_c(\mathbb{Z}^d))$, $p_2 \in (p_c(\mathbb{Z}^d), 1)$ and $n \geq 1$. Then we have the following estimates*:

$$\mathbb{P}\big(\{0, e_1\} \underset{p}{\overset{2}{\leftrightsquigarrow}} B(n)\big) \leq \begin{cases} c_{19}\exp(-c_{20}n), & \text{if } p \leq p_1, \\ c_9(\log n/n)^{1/2}, & \text{if } p \in [p_1, p_2], \\ c_{21}\exp(-c_{22}n), & \text{if } p \geq p_2. \end{cases} \tag{11.1}$$

The constants appearing here depend only on p_1, p_2 and d. The same bounds hold for $\mathbb{P}(B(1) \underset{p}{\overset{2}{\longleftrightarrow}} B(n))$. Further,

$$\mathbb{P}\big(B(1) \underset{p}{\overset{2}{\leftrightsquigarrow}} B(n) \text{ in } Q\big) \leq \begin{cases} c_{19}\exp(-c_{20}n), & \text{if } p \leq p_1, \\ c_{21}\exp(-c_{22}n), & \text{if } p \geq p_2, \end{cases} \tag{11.2}$$

whenever Q is a cube containing the origin and $\partial^{\mathrm{in}} B(n)$ has a vertex in Q.

PROOF. The bounds in the subcritical regime follow from Menshikov's theorem (see, e.g., [33]). When $d \geq 3$ and $p \geq p_2$, exponential decay will follow from the proof of [33], Lemma 7.89. When $d = 2$ and $p \geq p_2$, the stated bound follows from arguments similar to the ones used in the proof of Proposition 10.13 in [47]. The bound for $p \in [p_1, p_2]$ is just the content of Lemma 5.2. $\square$

12. Rate of convergence in the CLT in the lattice setup. We will prove Theorem 2.4 in this section. Let $u_1, \ldots, u_\ell$ be the edges of $\mathbb{Z}^d$ having both endpoints in $B(n)$, and let $X_1, \ldots, X_\ell$ be the weights associated with them. Define $X = (X_1, \ldots, X_\ell)$ and let $X' = (X'_1, \ldots, X'_\ell)$ be an independent copy of X. Write F_μ for the distribution function of X_1. Fix $\alpha \in (0, 1)$. We will make an appropriate choice of α later. Let

$$\begin{aligned}
&\mathcal{J} := \{j \mid \text{both endpoints of } u_j \text{ are in } B(n - n^\alpha)\} \quad \text{and} \\
&\mathcal{L} := \{1, \ldots, \ell\}.
\end{aligned} \tag{12.1}$$

For each $j \in \mathcal{L}$, choose and fix an endpoint x_j of u_j, and let

$$B_j := B(x_j, 1) \cap B(n) \quad \text{and} \quad \tilde{B}_j := B(x_j, n^\alpha) \cap B(n). \tag{12.2}$$

Thus, $\tilde{B}_j = B(x_j, n^\alpha)$ if $j \in \mathcal{J}$.

We will apply (3.1) with

$$f(X) = M(B(n), X).$$

As in the proof of Theorem 2.1, we will use the shorthand

$$\Delta_j f(X^A) := \Delta_j f(X^A, X')$$

for any $A \subset \mathcal{L}$ and $j \in \mathcal{L}$. We further define

$$\tilde{\Delta}_j f(X^A) := M(\tilde{B}_j, X^A) - M(\tilde{B}_j, X^{A \cup \{j\}}) \tag{12.3}$$

for every $j \in \mathcal{L}$ and $A \subset \mathcal{L}$.

12.1. *Preliminary estimates*. In this section, we give an analogue of Proposition 10.1.

PROPOSITION 12.1. *The following hold*:

(i) *Let Z_j be the maximum of the weights associated with the edges of $B_j - u_j$. Let $\mathbb{P}_1$ denote probability conditional on the weights associated with the edges of $B_j - u_j$. Then for $j \in \mathcal{L}$*,

$$\mathbb{E}\left|\Delta_j f(X) - \tilde{\Delta}_j f(X)\right| \le 2\mathbb{E}\left[Z_j \cdot \int_0^1 \mathbb{P}_1\left(B_j \overset{2}{\underset{F_\mu(uZ_j)}{\leftrightsquigarrow}} \tilde{B}_j \textit{ in } B(n)\right) du\right]. \tag{12.4}$$

(ii) *Order of variance*:

$$\operatorname{Var}\big(M(B(n), X)\big) = \Theta(n^d). \tag{12.5}$$

PROOF. For $j \in \mathcal{L}$, define Y_j to be the maximum edge-weight in the path connecting the two endpoints of u_j in an MST of $B(n) - u_j$, when the edge-weights are given by the appropriate subvector of X. From the add and delete algorithm (Section 6.2), it follows that

$$M(B(n), X) = M(B(n) - u_j, X) + X_j - \max(X_j, Y_j) \qquad \text{for } j \in \mathcal{L}, \tag{12.6}$$

and a similar assertion is true when X is replaced by X^j. Similarly, define $\tilde{Y}_j$ to be the maximum edge-weight in the path connecting the two endpoints of u_j in an MST of $\tilde{B}_j - u_j$. Then (12.6) holds if we replace $B(n)$ by $\tilde{B}_j$ and Y_j by $\tilde{Y}_j$.

Note also that for $j \in \mathcal{L}$

$$\begin{aligned} M(B(n) - u_j, X) &= M(B(n) - u_j, X^j) \quad \text{and} \\ M(\tilde{B}_j - u_j, X) &= M(\tilde{B}_j - u_j, X^j). \end{aligned} \tag{12.7}$$

Hence,

$$
\begin{aligned}
&\left|\Delta_j f(X) - \tilde{\Delta}_j f(X)\right| \\
&\quad = \left|\max(X_j, Y_j) - \max(X_j, \tilde{Y}_j) - \max(X'_j, Y_j) + \max(X'_j, \tilde{Y}_j)\right| \qquad (12.8)\\
&\quad \le 2|Y_j - \tilde{Y}_j|.
\end{aligned}
$$

From Lemma 6.1, it follows that $Y_j \le \tilde{Y}_j$. Combining this with the definition of Z_j, we get $Y_j \le \tilde{Y}_j \le Z_j$. Thus,

$$
\mathbb{E}|Y_j - \tilde{Y}_j| = \mathbb{E}\left(Z_j \cdot \mathbb{E}_1 \frac{(\tilde{Y}_j - Y_j)}{Z_j}\right), \tag{12.9}
$$

where $\mathbb{E}_1$ denotes expectation conditional on the weights associated with the edges of $B_j - u_j$. Then for a random variable U following Uniform[0, 1] distribution that is independent of X, X',

$$
\begin{aligned}
\mathbb{E}_1 \frac{(\tilde{Y}_j - Y_j)}{Z_j} &= \mathbb{P}_1(Y_j < U Z_j < \tilde{Y}_j) \\
&= \int_0^1 \mathbb{P}_1(Y_j < u Z_j < \tilde{Y}_j)\, du \qquad (12.10)\\
&\le \int_0^1 \mathbb{P}_1\big(B_j \underset{F_\mu(uZ_j)}{\overset{2}{\leftrightsquigarrow}} \tilde{B}_j \text{ in } B(n)\big)\, du,
\end{aligned}
$$

where the last inequality follows from an argument identical to the one given right after (10.10). Equation (12.4) follows upon combining (12.8), (12.9) and (12.10).

The conclusion in (12.5) is included in the more general Theorem 2.6 whose proof will be given in Section 13. $\square$

12.2. *Proof of Theorem* 2.4. The proof can be divided into two parts.

PROOF OF (2.5). Recall the definition of the sets $\mathcal{J}$ and $\mathcal{L}$ from (12.1). Recall also that for every $j \in \mathcal{L}$, we have chosen and fixed an endpoint x_j of u_j. Mimicking the proof of Theorem 2.1, we define the sets

$$
\begin{aligned}
\mathfrak{E}^{(\alpha)} = \{&(j, j', A, A') : j, j' \in \mathcal{L}, A, A' \subsetneq \mathcal{L}; j \notin A, j' \notin A' \text{ and} \\
&\text{either } j \notin \mathcal{J}, \text{ or } j' \notin \mathcal{J}, \text{ or } \|x_j - x_{j'}\|_\infty \le 2n^\alpha\}
\end{aligned}
$$

and

$$
\mathfrak{F}^{(\alpha)} = \{(j, j', A, A') : j, j' \in \mathcal{L}, A, A' \subsetneq \mathcal{L}; j \notin A, j' \notin A'\} \setminus \mathfrak{E}^{(\alpha)}.
$$

From (12.6) and (12.7), it is clear that under the assumption of finite $(4+\delta)$th moment on μ,

$$
\mathbb{E}\left|\Delta_j f(X)\right|^{(4+\delta)} \le C \qquad \text{for every } j \le \ell. \tag{12.11}
$$

Hence, (10.15) remains true in our present setup. In view of (12.5), (10.16) continues to hold as well.

If we split the sum appearing in (3.3) into two parts $\sum_1$ (the sum over $\mathfrak{E}^{(\alpha)}$) and $\sum_2$ (the sum over $\mathfrak{F}^{(\alpha)}$), then (10.18) continues to hold, that is,

$$\sum\nolimits_1 \frac{\operatorname{Cov}(\Delta_j f(X)\Delta_j f(X^A), \Delta_{j'} f(X)\Delta_{j'} f(X^{A'}))}{\binom{\ell}{|A|}(\ell-|A|)\binom{\ell}{|A'|}(\ell-|A'|)} \le c' n^{2d-1+\alpha}. \tag{12.12}$$

Further, (10.20) and (10.21) apply whenever $(j, j', A, A') \in \mathfrak{F}^{(\alpha)}$. Let T_1 and T_2 be as in (10.21). If μ has bounded support, then

$$T_1 + T_2 \le c\mathbb{E}\bigl|\Delta_j f(X) - \tilde{\Delta}_j f(X)\bigr|, \tag{12.13}$$

and if μ has unbounded support and finite $(4+\delta)$th moment, then

$$\begin{aligned} T_1 &\le \bigl(\mathbb{E}\bigl|\Delta_j f(X) - \tilde{\Delta}_j f(X)\bigr|\bigr)^{1/q'} \\ &\quad \times \bigl[\mathbb{E}\bigl(\bigl|\Delta_j f(X) - \tilde{\Delta}_j f(X)\bigr|\bigl(\bigl|\Delta_j f\bigl(X^A\bigr)\Delta_{j'} f(X)\Delta_{j'} f\bigl(X^{A'}\bigr)\bigr|\bigr)^q\bigr)\bigr]^{1/q} \\ &= C_{12.14}\bigl(\mathbb{E}\bigl|\Delta_j f(X) - \tilde{\Delta}_j f(X)\bigr|\bigr)^{1/q'}, \end{aligned} \tag{12.14}$$

where $q = 1 + \delta/3$ and $q' = 1 + 3/\delta$. That $C_{12.14}$ is finite is ensured by (12.11). An application of Hölder's inequality will give a similar bound for T_2. Let

$$\overline{q} = \begin{cases} 1, & \text{if } \mu \text{ satisfies Property } B, \\ 1 + 3/\delta, & \text{if } \mu \text{ satisfies Property } A_\delta. \end{cases} \tag{12.15}$$

We have thus shown

$$\begin{aligned} &\operatorname{Cov}\bigl(\Delta_j f(X)\Delta_j f\bigl(X^A\bigr), \Delta_{j'} f(X)\Delta_{j'} f\bigl(X^{A'}\bigr)\bigr) \\ &\qquad \le c\bigl(\mathbb{E}\bigl|\Delta_j f(X) - \tilde{\Delta}_j f(X)\bigr|\bigr)^{1/\overline{q}}, \end{aligned} \tag{12.16}$$

whenever $(j, j', A, A') \in \mathfrak{F}^{(\alpha)}$. $\square$

We now consider two possibilities separately.

If μ satisfies Property C. If there exists a unique $x \in \mathbb{R}$ such that $\mu[0, x] = p_c(\mathbb{Z}^d)$, then there are two possibilities. The first possibility is that the distribution function of μ, namely F_μ, is continuous at x, and the second possibility is $F_\mu(x-) < F_\mu(x) = p_c$.

Assume first that F_μ is continuous at x where x is the unique point such that $F_\mu(x) = p_c(\mathbb{Z}^d)$. Choose a small enough positive ε_0 so that $F_\mu(x - \varepsilon_0) > 0$ and $F_\mu(x + \varepsilon_0) < 1$. For $\varepsilon > 0$, define the functions

$$p_1(\varepsilon) = F_\mu(x - \varepsilon) \quad \text{and} \quad p_2(\varepsilon) = F_\mu(x + \varepsilon).$$

Note that when $j \in \mathcal{J}$, the integral on the right-hand side of (12.4) can be written as $\int_0^1 \mathbb{P}_1(B_j \overset{2}{\underset{F_\mu(uZ_j)}{\leftrightsquigarrow}} \tilde{B}_j)\,du$. For any $\varepsilon \in (0, \varepsilon_0)$, we can break up this integral into

$$\int_0^{\min((x-\varepsilon)/Z_j,1)}, \qquad \int_{\min((x-\varepsilon)/Z_j,1)}^{\min((x+\varepsilon)/Z_j,1)} \quad \text{and} \quad \int_{\min((x+\varepsilon)/Z_j,1)}^1,$$

to get

$$\int_0^1 \mathbb{P}_1\big(B_j \overset{2}{\underset{F_\mu(uZ_j)}{\leftrightsquigarrow}} \tilde{B}_j\big)\,du \le c_{23}\exp(-c_{24}n^\alpha) + \frac{2\varepsilon}{Z_j}\cdot c_9\left(\frac{\alpha\log n}{n^\alpha}\right)^{1/2} \tag{12.17}$$

by an application of Lemma 11.1. The constants c_{23} and c_{24} depend on c_{19}, c_{20}, c_{21} and c_{22} as in Lemma 11.1 corresponding to the choices $p_i = p_i(\varepsilon), i = 1, 2$, and the constant c_9 is the one from Lemma 11.1 corresponding to the choices $p_i = p_i(\varepsilon_0), i = 1, 2$.

From (12.4) and (12.17), we get

$$\begin{aligned} &\mathbb{E}\big|\Delta_j f(X) - \tilde{\Delta}_j f(X)\big| \\ &\qquad \le 2c_{23}\mathbb{E}(Z_j)\exp(-c_{24}n^\alpha) + 4\varepsilon c_9\left(\frac{\alpha\log n}{n^\alpha}\right)^{1/2} \end{aligned} \tag{12.18}$$

for every $j \in \mathcal{J}$. Combining (12.16) with (12.18), we get

$$\begin{aligned} &\sum\nolimits_2 \frac{\operatorname{Cov}(\Delta_j f(X)\Delta_j f(X^A), \Delta_{j'} f(X)\Delta_{j'} f(X^{A'}))}{\binom{\ell}{|A|}(\ell - |A|)\binom{\ell}{|A'|}(\ell - |A'|)} \\ &\qquad \le c\cdot n^{2d}\left(2c_{23}\mathbb{E}(Z_j)\exp(-c_{24}n^\alpha) + 4\varepsilon c_9\left(\frac{\alpha\log n}{n^\alpha}\right)^{1/2}\right)^{1/\bar{q}}. \end{aligned}$$

The last inequality combined with (12.12), (12.5), (3.1), (3.3) and (10.16) (we have already observed that the last inequality holds in our present setup) yields

$$\begin{aligned} &\mathcal{W}(\nu_n, \gamma) \\ &\quad \le c'\left[\frac{1}{n^{(1-\alpha)/2}} + \left(2c_{23}\mathbb{E}(Z_j)\exp(-c_{24}n^\alpha) + 4\varepsilon c_9\left(\frac{\alpha\log n}{n^\alpha}\right)^{1/2}\right)^{\frac{1}{2\bar{q}}} + \frac{1}{n^{d/2}}\right], \end{aligned}$$

where c' is a constant free of ε. We take $\alpha = 2\bar{q}/(1+2\bar{q})$ in the last inequality. It then follows that

$$\limsup_n \frac{n^{\frac{1}{2(1+2\bar{q})}}}{(\log n)^{1/(4\bar{q})}}\mathcal{W}(\nu_n, \gamma) \le c'\left(4\varepsilon c_9\left(\frac{2\bar{q}}{1+2\bar{q}}\right)^{1/2}\right)^{\frac{1}{2\bar{q}}}.$$

This inequality is true for any $\varepsilon > 0$, and recall that c' and c_9 $(= c_9(p_1(\varepsilon_0), p_2(\varepsilon_0)))$ do not depend on ε. This shows that (2.5) holds in this case.

The argument is similar if (i) $\mu[0, x] = p_c(\mathbb{Z}^d)$ for some unique $x \in \mathbb{R}$ and $F_\mu(x-) < F_\mu(x)$ or (ii) $\mu[0, x) = p_c(\mathbb{Z}^d)$ for some unique $x \in \mathbb{R}$, so we do not repeat it.

If μ does not satisfy Property C. Combining the bound in (12.4) with (12.16), and Lemma 11.1, we get

$$\begin{aligned}\operatorname{Cov}\bigl(\Delta_j f(X)\Delta_j f\bigl(X^A\bigr), \Delta_{j'} f(X)\Delta_{j'} f\bigl(X^{A'}\bigr)\bigr)\\ \le \sup_{0<p<1} c\bigl[\mathbb{P}(B_j \overset{2}{\underset{p}{\leftrightsquigarrow}} \tilde{B}_j)\bigr]^{1/\bar{q}} \le c'\biggl(\frac{\log n}{n^\alpha}\biggr)^{1/2\bar{q}},\end{aligned} \tag{12.19}$$

whenever $(j, j', A, A') \in \mathfrak{F}^{(\alpha)}$, and hence

$$\sum\nolimits_2 \frac{\operatorname{Cov}(\Delta_j f(X)\Delta_j f(X^A), \Delta_{j'} f(X)\Delta_{j'} f(X^{A'}))}{\binom{\ell}{|A|}(\ell-|A|)\binom{\ell}{|A'|}(\ell-|A'|)} \le c\cdot n^{2d}\biggl(\frac{\log n}{n^\alpha}\biggr)^{1/2\bar{q}}.$$

Combining this inequality with (12.12), and taking $\alpha = 2\bar{q}/(1+2\bar{q})$, we get

$$\operatorname{Var}\bigl(\mathbb{E}(T\,|\,W)\bigr) \le cn^{2d}\frac{(\log n)^{1/2\bar{q}}}{n^{1/(1+2\bar{q})}}. \tag{12.20}$$

The last bound together with (12.5), (3.1) and (10.16) yields the bound in (2.5).

PROOF OF (2.6). We introduce

$$\begin{aligned}\overline{\mathfrak{E}}^{(\alpha)} &:= \bigl\{\bigl(j, j', A, A'\bigr) : j, j' \in \mathcal{L}, A, A' \subsetneq \mathcal{L}; j \notin A, j' \notin A'\\ &\qquad \text{and } \|x_j - x_{j'}\|_\infty \le 2n^\alpha\bigr\} \quad \text{and}\\ \overline{\mathfrak{F}}^{(\alpha)} &:= \bigl\{\bigl(j, j', A, A'\bigr) : j, j' \in \mathcal{L}, A, A' \subsetneq \mathcal{L}; j \notin A, j' \notin A'\bigr\} \setminus \bar{\mathfrak{E}}^{(\alpha)}\end{aligned}$$

for $0 < \alpha < 1$. We split the sum appearing in (3.3) into $\overline{\sum}_1$, the sum over $(j, j', A, A') \in \overline{\mathfrak{E}}^{(\alpha)}$ and $\overline{\sum}_2$, the sum over $(j, j', A, A') \in \overline{\mathfrak{F}}^{(\alpha)}$. Then similar to (10.18),

$$\begin{aligned}\overline{\sum}_1 \frac{\operatorname{Cov}(\Delta_j M(X)\Delta_j M(X^A), \Delta_{j'} M(X)\Delta_{j'} M(X^{A'}))}{\binom{\ell}{|A|}(\ell-|A|)\binom{\ell}{|A'|}(\ell-|A'|)}\\ \le c\bigl|\bigl\{\bigl(j, j'\bigr) : \bigl(j, j', \varnothing, \varnothing\bigr) \in \overline{\mathfrak{E}}^{(\alpha)}\bigr\}\bigr| \le cn^{d+\alpha d}.\end{aligned} \tag{12.21}$$

Further, the argument leading to (12.19) yields

$$\begin{aligned}\operatorname{Cov}\bigl(\Delta_j f(X)\Delta_j f\bigl(X^A\bigr), \Delta_{j'} f(X)\Delta_{j'} f\bigl(X^{A'}\bigr)\bigr)\\ \le \sup_{0<u<1} c\bigl[\mathbb{P}_1\bigl(B_j \overset{2}{\underset{F_\mu(uZ_j)}{\leftrightsquigarrow}} \tilde{B}_j \text{ in } B(n)\bigr)\bigr]^{1/\bar{q}},\end{aligned} \tag{12.22}$$

whenever $(j, j', A, A') \in \overline{\mathfrak{F}}^{(\alpha)}$, where $\bar{q}$ is as in (12.15). Since μ satisfies Property *D* by assumption, $\operatorname{Range}(F_\mu) \subset (p_c - \varepsilon, p_c + \varepsilon)^c$ for some $\varepsilon > 0$. It thus follows from (12.22) and Lemma 11.1 that

$$\begin{aligned}
&\mathrm{Cov}\big(\Delta_j f(X)\Delta_j f\big(X^A\big), \Delta_{j'} f(X)\Delta_{j'} f\big(X^{A'}\big)\big) \\
&\quad\le \sup_{p\notin(p_c-\varepsilon,p_c+\varepsilon)} c\big[\mathbb{P}\big(B_j \underset{F_\mu(uZ_j)}{\overset{2}{\leftrightsquigarrow}} \tilde{B}_j \text{ in } B(n)\big)\big]^{1/\bar{q}} \qquad (12.23)\\
&\quad\le c' \exp\big(-c'' n^\alpha\big)
\end{aligned}$$

whenever $(j, j', A, A') \in \overline{\mathfrak{F}}^{(\alpha)}$. Hence,

$$\begin{aligned}
&\overline{\sum}_2 \frac{\mathrm{Cov}(\Delta_j M(X)\Delta_j M(X^A), \Delta_{j'} M(X)\Delta_{j'} M(X^{A'}))}{\binom{\ell}{|A|}(\ell-|A|)\binom{\ell}{|A'|}(\ell-|A'|)} \qquad (12.24)\\
&\quad\le cn^{2d}\exp\big(-c'n^\alpha\big).
\end{aligned}$$

As before, we combine (12.5), (12.21), (12.24), (10.16) and (3.1) to conclude that

$$\mathcal{W}(\nu_n,\gamma) \le c/n^{\frac{d(1-\alpha)}{2}}.$$

We get the bound in (2.6) once we replace $d(1-\alpha)/2$ by η. $\square$

Bounds on the Kolmogorov distance. Bounds on the Kolmogorov distance can be obtained by using Theorem 3.2 and following the same line of arguments (see the discussion at the end of Section 10.2). Note the presence of the term $\mathbb{E}|\Delta_j f(X,X')|^6$ in (3.2). We require μ to satisfy either Property B or Property A_δ with $\delta \ge 2$ to show that $\mathbb{E}|\Delta_j f(X,X')|^6 < \infty$. The rest of the argument goes through verbatim.

This completes the proof of Theorem 2.4.

13. General graphs: Proof of Theorem 2.6. To fix ideas, we first assume that G is symmetric, that is, for every two pairs of adjacent vertices v_1, v_2 and v_1', v_2', there exists a graph automorphism f of G such that $f(v_i) = v_i'$, $i = 1, 2$.

If G is symmetric and deletion of an edge of G creates two components, then G is a regular tree. Hence all our claims follow trivially. So we can assume that this is not the case.

Let $E_n = \{u_1, \dots, u_{\ell_n}\}$ and let $X_1, \dots, X_{\ell_n}$ be the associated edge weights. Let $X = (X_1, \dots, X_{\ell_n})$ and let $X' = (X_1', \dots, X_{\ell_n}')$ be an independent copy of X. Then $M_n = M(G_n, X)$. As before, we want to apply Theorem 3.1 with

$$f(X) := M(G_n, X).$$

As in the proof of Theorem 2.1, we will use the shorthand

$$\Delta_j f\big(X^A\big) := \Delta_j f\big(X^A, X'\big),$$

for any $A \subset [\ell_n]$ and $1 \le j \le \ell_n$.

Define

$$\mathcal{I}_r^n := \{i \le \ell_n : S(v, r) \subset G_n \text{ for each endpoint } v \text{ of } u_i\}.$$

For large r and $i \in \mathcal{I}_r^n$, fix an endpoint v_i of u_i and let Y_i^n [resp., $Y_i^n(r)$] be the maximum edge weight in a path connecting the endpoints of u_i in an MST of $G_n - u_i$ [resp., $S(v_i, r) - u_i$] with the edge weights being the appropriate subvector of X. We will suppress the dependence on n and simply write $\mathcal{I}_r$, Y_i and $Y_i(r)$.

An application of Lemma 9.1 yields, with our usual notation,

$$\begin{aligned}
\operatorname{Var}(f(X)) &\ge \sum_{i=1}^{\ell_n} \operatorname{Var}(\mathbb{E}(f(X)|X_i)) \\
&= \frac{1}{2}\sum_{i=1}^{\ell_n} \mathbb{E}[\mathbb{E}(f(X)|X_i) - \mathbb{E}(f(X^i)|X_i')]^2 \\
&\ge \frac{1}{2}\sum_{i\in\mathcal{I}_r} \mathbb{E}[\mathbb{E}(f(X)|X_i) - \mathbb{E}(f(X^i)|X_i')]^2 \qquad (13.1)\\
&= \frac{1}{2}\sum_{i\in\mathcal{I}_r} \mathbb{E}[\mathbb{E}(f(X) - f(X^i)|X_i, X_i')]^2.
\end{aligned}$$

By the add and delete algorithm (Section 6.2), for $i \in \mathcal{I}_r$,

$$f(X) = M(G_n - u_i, X) + X_i - \max(X_i, Y_i),$$

and hence

$$f(X) - f(X^i) = \min(X_i, Y_i) - \min(X_i', Y_i). \qquad (13.2)$$

Since μ is nondegenerate, we can find real numbers $b > a$ such that $\mu[0, a] > 0$ and $\mu[b, \infty] > 0$. Going back to (13.1),

$$\begin{aligned}
\operatorname{Var}(f(X)) &\ge \frac{1}{2}\sum_{i\in\mathcal{I}_r} \mathbb{E}[\mathbb{E}(\min(X_i, Y_i) - \min(X_i', Y_i)|X_i, X_i')]^2 \\
&\ge \frac{1}{2}\sum_{i\in\mathcal{I}_r} \mathbb{E}[((b-a)\mathbb{P}(Y_i \ge b))^2 \mathbb{I}\{X_i \le a, X_i' \ge b\}] \qquad (13.3)\\
&\ge \frac{1}{2}|\mathcal{I}_r|(b-a)^2 \cdot p^2 \cdot \mu[0, a] \cdot \mu[b, \infty),
\end{aligned}$$

where

$$p := \mathbb{P}(\text{The weight associated with each edge sharing one vertex with } u_i \text{ is at least } b).$$

Note that p does not depend on the edge u_i since G is symmetric. By assumption (III),

$$|\mathcal{I}_r| = \Theta(|V_n|). \tag{13.4}$$

From (13.3) and (13.4), it follows that

$$\operatorname{Var}(f(X)) \geq c|V_n|.$$

The upper bound is a simple consequence of the Efron–Stein inequality:

$$\operatorname{Var}(f(X)) \leq \frac{1}{2}\sum_{j=1}^{\ell_n} \mathbb{E}(\Delta_j f(X))^2.$$

Thus, we have proven that $\operatorname{Var}(M_n) = \operatorname{Var}(f(X)) = \Theta(|V_n|)$.

Turning toward the proof of the central limit theorem, define, for large r,

$$\begin{aligned}\mathfrak{E}_n(r) :=\{(j, j', A, A') &: j, j' \leq \ell_n, A, A' \subsetneq \{1, \ldots, \ell_n\}, j \notin A, j' \notin A' \\ &\text{and either } d_G(x_j, x_{j'}) \leq 2r \text{ or } S(x_j, r) \not\subset G_n \text{ or } S(x_{j'}, r) \not\subset G_n \\ &\text{for some endpoints } x_j, x_{j'} \text{ of } u_j \text{ and } u_{j'} \text{ respectively}\} \quad \text{and} \\ \mathfrak{F}_n(r) = \{(j, j', A, A') &: j, j' \leq \ell_n, A, A' \subsetneq \{1, \ldots, \ell_n\}, j \notin A, j' \notin A'\} \setminus \mathfrak{E}_n(r).\end{aligned}$$

Proceeding as before, we split the sum in (3.3) into $\sum_1$, the sum over all $(j, j', A, A') \in \mathfrak{E}_n(r)$ and $\sum_2$, the sum over the rest of the terms. It follows from (13.2) that $|\Delta_j f(X)| \leq |X_j - X'_j|$. Further, $\mathbb{E}(X_j^4) < \infty$. Thus, a computation similar to (10.18) will yield

$$\begin{aligned}&\sum\nolimits_1 \frac{\operatorname{Cov}(\Delta_j f(X)\Delta_j f(X^A), \Delta_{j'} f(X)\Delta_{j'} f(X^{A'}))}{\binom{\ell_n}{|A|}(\ell_n - |A|)\binom{\ell_n}{|A'|}(\ell_n - |A'|)} \\ &\qquad \leq c|V_n|\big(a_r + |\{v \in V_n : S(v, r) \not\subset G_n\}|\big),\end{aligned} \tag{13.5}$$

where $a_r := |\{v' \in V : d_G(v, v') \leq 2r\}|$ for some (and hence all, by symmetry) $v \in V$.

For $j \in \mathcal{I}_r$, define

$$\tilde{\Delta}_j f(X) = M(S(v_j, r), X) - M(S(v_j, r), X^j).$$

With this definition of $\tilde{\Delta}_j f(X)$, (10.20) and (10.21) hold for $(j, j', A, A') \in \mathfrak{F}_n(r)$. As in (12.13) and (12.14), we get

$$T_1 \leq c\big(\mathbb{E}|\Delta_j f(X) - \tilde{\Delta}_j f(X)|\big)^{\frac{1}{1+3/\delta}}$$

for some $\delta \geq 0$, where $\delta = 0$ if μ satisfies Property B and $\delta > 0$ if μ satisfies Property A_δ. A similar bound holds for T_2. A calculation similar to (12.8) yields

$$|\Delta_j f(X) - \tilde{\Delta}_j f(X)| \leq 2(Y_j(r) - Y_j) \tag{13.6}$$

for $j \in \mathcal{I}_r$.

Fix a vertex v of G and let e be an edge incident to v. Let $Y(v,e,r)$ be the maximum edge weight in the path connecting the endpoints of e in an MST of $S(v,r)-e$, clearly $Y(v,e,r)$ is decreasing in r. Define

$$Y(v,e) := \lim_{r\to\infty} Y(v,e,r).$$

The above convergence also holds in L^1 as a consequence of dominated convergence theorem. Since G is symmetric, $Y_i^n(r)$ has the same distribution as $Y(v,e,r)$ and Y_i^n dominates $Y(v,e)$ stochastically for every $i \in \mathcal{I}_r^n$. Hence,

$$\begin{aligned}&\lim_{r\to\infty}\limsup_{n\to\infty}\Big[\max_{i\in\mathcal{I}_r^n}\mathbb{E}\big(Y_i^n(r)-Y_i^n\big)\Big]\\ &\qquad\le \lim_{r\to\infty}\mathbb{E}\big(Y(v,e,r)-Y(v,e)\big)=0.\end{aligned}\tag{13.7}$$

Thus, we have

$$\begin{aligned}&\lim_{r\to\infty}\limsup_{n\to\infty}\max_{\substack{(j,j',A,A')\\ \in\mathfrak{F}_n(r)}}\operatorname{Cov}\big(\Delta_j f(X)\Delta_j f\big(X^A\big),\Delta_{j'}f(X)\Delta_{j'}f\big(X^{A'}\big)\big)\\ &\qquad=0,\end{aligned}\tag{13.8}$$

which gives us control over $\sum_2$. Further,

$$\frac{1}{\operatorname{Var}(f(X))^{3/2}}\sum_{j=1}^{\ell_n}\mathbb{E}\big|\Delta_j f(X)\big|^3 \le \frac{c}{|V_n|^{1/2}}.$$

The last inequality, together with (13.5), (13.8), (3.1) and the fact that $\operatorname{Var}(M_n)=\Theta(|V_n|)$, yields

$$\limsup_n \mathcal{W}(\mu_n,\gamma)=0,$$

where μ_n is the law of $(M_n-\mathbb{E}M_n)/\sqrt{\operatorname{Var}(M_n)}$.

Assume now that G is vertex-transitive so that there are two kinds of edges. Call an edge $e\in E$ of type A if deletion of e results in the creation of two disjoint components. We say e is of type B if it is not of type A. Define $\tilde{\mathcal{I}}_r^n := \{i\in\mathcal{I}_r : i \text{ is of type B}\}$. Define Y_i^n and $Y_i^n(r)$ as before for each $i\in\tilde{\mathcal{I}}_r^n$. Then as in (13.1),

$$\operatorname{Var}\big(f(X)\big)\ge\frac{1}{2}\sum_{i\in\tilde{\mathcal{I}}_r}\mathbb{E}\big[\mathbb{E}\big(f(X)|X_i\big)-\mathbb{E}\big(f\big(X^i\big)|X_i'\big)\big]^2.$$

Note also that $|\tilde{\mathcal{I}}_r|=\Theta(|V_n|)$ if G is not a tree. So we can argue as before to conclude that $\operatorname{Var}(f(X))=\Theta(|V_n|)$.

Next, note that if $j\in\mathcal{I}_r^n$ and u_j is of type A, then $\Delta_j f(X)-\tilde{\Delta}_j f(X)=0$. Further, our previous arguments show that

$$\begin{aligned}&\lim_{r\to\infty}\limsup_{n\to\infty}\Big[\max_{i\in\tilde{\mathcal{I}}_r^n}\mathbb{E}\big(Y_i^n(r)-Y_i^n\big)\Big]\\ &\qquad\le\lim_{r\to\infty}\sum\nolimits_*\mathbb{E}\big(Y(v,e,r)-Y(v,e)\big)=0,\end{aligned}\tag{13.9}$$

where $\sum_*$ is the sum over all type B edges e incident to v. The rest of the arguments remain the same. This completes the proof of the central limit theorem.

APPENDIX

A.1. Completing the proof of Lemma 9.5. The following proposition fills in the gap in the proof of Lemma 9.5.

PROPOSITION. *Assume that* $n \geq 2$, $a \in [1/2, \log n]$, *and* $r_c \leq r \leq (\log n)^2$. *Then there exists positive universal constants* c_{12} *and* β *such that*

$$\mathbb{P}\big(B_{\mathbb{R}^2}(a) \xrightarrow[r]{2} B_{\mathbb{R}^2}(n)\big) \leq c_{12}/n^{\beta}.$$

PROOF. As usual $\mathcal{P}$ will denote a Poisson process of intensity one. Let $\sigma((a,b); r, j)$ denote the probability of an occupied crossing of the rectangle $[0,a] \times [0,b]$ at level r in the jth direction, $j = 1, 2$; that is,

$$\begin{aligned}\sigma\big((a,b); r, 1\big) = \mathbb{P}(&\mathcal{P}^{(r)} \text{ contains a curve } \gamma \subset [0,a] \times [0,b] \\ &\text{such that } \gamma \text{ intersects both } S_1 \text{ and } S_2\big),\end{aligned}$$

where $S_1 = \{0\} \times [0,b]$ and $S_2 = \{a\} \times [0,b]$ and define $\sigma((a,b); r, 2)$ similarly. First, we note that

$$\sigma\big((m, 3m); r_c, 1\big) \geq \kappa_0 := (9e)^{-122} \qquad \text{whenever } m > r_c.$$

(We can prove this assertion by observing that $\sigma((m, 3m); r, 1)$ is a continuous function of r and then using arguments similar to the ones given right after (9.20) and [43], Lemma 3.3.)

Now, the proof of Lemma 4.4 of [43] applies to occupied crossings as well. Since $\sigma((m, 3m); r_c, 1) \geq \kappa_0$ for $m > r_c$, the arguments of Lemma 4.4 of [43] would furnish positive constants $f(t)$ for each $t > 0$ such that

$$\sigma\big((m, (1+t)m); r_c, 1\big) \geq f(t).$$

Applying Theorem 2.1 of [8] with the parameters $h = \ell/(1+t)$ and $b = \ell/(1+t)^2$ with t small enough so that $2/(1+t)^2 - 1/2 > 1 + \varepsilon$ (for some positive ε) and $(1+t)^2 < 4/3$ and ℓ large so that $h > 4r_c$ and $b > \ell/2 + 2r_c$, we get

$$\begin{aligned}&\sigma\left(\left(\ell\left[\frac{2}{(1+t)^2} - \frac{1}{2}\right] + r_c, \frac{\ell}{1+t} - 2r_c\right); r_c, 1\right) \\ &\quad \geq c\sigma\left(\left(\frac{\ell}{(1+t)^2} + r_c, \frac{\ell}{1+t} - 4r_c\right); r_c, 1\right)^4 \times \sigma\left(\left(\ell, \frac{\ell}{1+t} + 3r_c\right); r_c, 2\right)^2\end{aligned}$$

for large ℓ. Hence,

$$\begin{aligned}
&\sigma\big((\ell(1+\varepsilon),\ell);r_c,1\big)\\
&\quad\ge c\sigma\bigg(\bigg(\frac{\ell}{(1+3t/4)^2}+r_c,\frac{\ell}{1+5t/4}\bigg);r_c,1\bigg)^4\times\sigma\bigg(\bigg(\ell,\frac{\ell}{1+t/2}\bigg);r_c,2\bigg)^2\\
&\quad\ge cf\bigg(\frac{(1+3t/4)^2}{1+5t/4}-1\bigg)^4\times f(t/2)^2
\end{aligned}$$

for every ℓ bigger than a fixed threshold ℓ_0. Hence, Lemma 3.1 of [8] yields

$$\sigma\big((3\ell,\ell);r_c,1\big)\ge\kappa_1 \tag{A.1}$$

for a positive constant κ_1 and $\ell\ge\ell_0$.

Let A_k be the event that there is an occupied circuit at level r_c in the annulus $B_{\mathbb{R}^2}(3\ell_k/2)\setminus B_{\mathbb{R}^2}(\ell_k/2)$, where $\ell_k=3\ell_{k-1}+4r_c$ and $\ell_1=\max(2a+2r,\ell_0)$. FKG inequality and (A.1) gives $\mathbb{P}(A_k)\ge\kappa_1^4$. Hence,

$$\mathbb{P}\big(B_{\mathbb{R}^2}(a)\xrightarrow[r]{2}B_{\mathbb{R}^2}(n)\big)\le\mathbb{P}\big(A_1^c\cap\cdots\cap A_t^c\big)=\prod_{k=1}^{t}\mathbb{P}\big(A_k^c\big)\le\big(1-\kappa_1^4\big)^t,$$

where $3\ell_t/2+r_c\le n-r<3\ell_{t+1}/2+r_c$. This yields the desired bound. $\square$

A.2. Proof of Lemma 5.7. Fix $p\in[p_1,p_2]$. Let $u_1,\dots,u_m$ be the edges of $\mathbb{Z}^d$ both of whose endpoints lie in $B(n)$ and let $X_1,\dots,X_m$ be i.i.d. Bernoulli(p) random variables [i.e., $\mathbb{P}(X_1=1)=p=1-\mathbb{P}(X_1=0)$] associated to them. Let $X:=(X_1,\dots,X_m)$ and let $X':=(X_1',\dots,X_m')$ be an independent copy of X. As earlier we define the event

$$\begin{aligned}
E:=\big\{&\text{there is exactly one } p\text{-cluster in } B_{\mathbb{Z}^d}(n) \text{ that intersects both } B_{\mathbb{Z}^d}(a_n)\\
&\text{and } \partial^{\mathrm{in}}B_{\mathbb{Z}^d}(n)\big\}
\end{aligned}$$

for some $a_n\to\infty$ in a way so that $a_n=o(n)$. Define the function f by $f(X):=\mathbb{I}_E(X)$. Then an application of Lemma 9.1 yields

$$\operatorname{Var}\big(f(X)\big)\ge\sum_{i\in\mathcal{I}}\operatorname{Var}\big[\mathbb{E}\big(f(X)|X_i\big)\big], \tag{A.2}$$

where $\mathcal{I}:=\{i\le m:\text{both endpoints of } u_i \text{ lie in } B(a_n/3)\}$. Fix $i\in\mathcal{I}$, denote the endpoints of u_i by v_1 and v_2. With our usual notation,

$$\begin{aligned}
\operatorname{Var}\big[\mathbb{E}\big(f(X)|X_i\big)\big]&=\frac{1}{2}\mathbb{E}\big[\big(\mathbb{E}\big(f(X)|X_i\big)-\mathbb{E}\big(f(X^i)|X_i'\big)\big)^2\big]\\
&\ge\frac{1}{2}\mathbb{E}\big[\mathbb{P}(A_i)^2\mathbb{I}\{X_i=1,X_i'=0\}\big],
\end{aligned} \tag{A.3}$$

where

$$A_i = \big\{u_i \overset{2}{\underset{p}{\leftrightsquigarrow}} B_{\mathbb{Z}^d}(n) - u_i,\ \text{any } p\text{-cluster in } B_{\mathbb{Z}^d}(n) - u_i \text{ that intersects both } \partial^{\text{in}} B_{\mathbb{Z}^d}(n) \text{ and } B_{\mathbb{Z}^d}(a_n) \text{ contains either } v_1 \text{ or } v_2\big\}.$$

Now,

$$\begin{aligned}\mathbb{P}(A_i) \geq \mathbb{P}\big(&u_i \overset{2}{\underset{p}{\leftrightsquigarrow}} B_{\mathbb{Z}^d}(v_1, 2n) - u_i,\ \text{if } \mathcal{C} \text{ is a } p\text{-cluster in } B_{\mathbb{Z}^d}(v_1, 2n) - u_i \\ &\text{then every connected component of } \mathcal{C} \cap B_{\mathbb{Z}^d}(v_1, n/2) \\ &\text{that intersects both } \partial^{\text{in}} B_{\mathbb{Z}^d}(v_1, n/2) \text{ and } B_{\mathbb{Z}^d}(v_1, 2a_n) \\ &\text{contains either } v_1 \text{ or } v_2\big) \\ = \mathbb{P}&(F)/(1-p),\end{aligned}$$

where

$$\begin{aligned}F := \big\{&\{0, e_1\} \overset{2}{\underset{p}{\leftrightsquigarrow}} B_{\mathbb{Z}^d}(2n),\ \text{if } \mathcal{C} \text{ is a } p\text{-cluster in } B_{\mathbb{Z}^d}(2n) - \{0, e_1\} \\ &\text{then every connected component of } \mathcal{C} \cap B_{\mathbb{Z}^d}(n/2) \\ &\text{that intersects both } \partial^{\text{in}} B_{\mathbb{Z}^d}(n/2) \text{ and } B_{\mathbb{Z}^d}(2a_n) \text{ contains either } 0 \text{ or } e_1\big\}.\end{aligned}$$

From (A.2) and (A.3), we conclude that

$$\mathbb{P}(F) \leq c/a_n^{d/2}. \tag{A.4}$$

Note that

$$\mathbb{P}\big(\{0, e_1\} \overset{2}{\underset{p}{\leftrightsquigarrow}} B_{\mathbb{Z}^d}(2n)\big) \leq \mathbb{P}(F) + \mathbb{P}\big(B_{\mathbb{Z}^d}(2a_n) \overset{3}{\underset{p}{\leftrightsquigarrow}} B_{\mathbb{Z}^d}(n/2)\big). \tag{A.5}$$

We now define a cube $Q \subset B_{\mathbb{Z}^d}(n/2)$ to be a trifurcation box in $B_{\mathbb{Z}^d}(n/2)$ at level p, if:

(i) there is a p-cluster $\mathcal{C}$ in $B_{\mathbb{Z}^d}(n/2)$ with $\mathcal{C} \cap Q \neq \varnothing$, and
(ii) the vertices of $\mathcal{C}$ contained in $B_{\mathbb{Z}^d}(n/2) - Q$ contain at least three p-clusters in $B_{\mathbb{Z}^d}(n/2) - Q$ each of which intersects $\partial^{\text{in}} B_{\mathbb{Z}^d}(n/2)$.

We can then apply the arguments in the proof of Lemma 9.2 [see the arguments leading up to (9.17)] to show that

$$\mathbb{P}\big(B_{\mathbb{Z}^d}(2a_n) \text{ is a trifurcation box in } B_{\mathbb{Z}^d}(n/2) \text{ at level } p\big) \leq \frac{ca_n^d}{n},$$

from which it will follow that

$$\mathbb{P}\big(B_{\mathbb{Z}^d}(2a_n) \overset{3}{\underset{p}{\leftrightsquigarrow}} B_{\mathbb{Z}^d}(n/2)\big) \leq c \exp\big(c' a_n\big) \frac{a_n^d}{n}. \tag{A.6}$$

Combining (A.4), (A.5) and (A.6), we choose $c' a_n = \log n/2$ to get the desired bound.

Acknowledgments. The authors are indebted to Larry Goldstein and Ümit Işlak for their valuable help in improving the manuscript and checking the proof. The authors also thank two anonymous referees whose careful reading and detailed reports improved the presentation considerably.

REFERENCES

[1] ADDARIO-BERRY, L., BROUTIN, N., GOLDSCHMIDT, C. and MIERMONT, G. (2013). The scaling limit of the minimum spanning tree of the complete graph. Preprint. Available at http://arxiv.org/abs/1301.1664.

[2] AIZENMAN, M., BURCHARD, A., NEWMAN, C. M. and WILSON, D. B. (1999). Scaling limits for minimal and random spanning trees in two dimensions. *Random Structures Algorithms* **15** 319–367. MR1716768

[3] AIZENMAN, M., KESTEN, H. and NEWMAN, C. M. (1987). Uniqueness of the infinite cluster and continuity of connectivity functions for short and long range percolation. *Comm. Math. Phys.* **111** 505–531. MR0901151

[4] ALDOUS, D. (1990). A random tree model associated with random graphs. *Random Structures Algorithms* **1** 383–402. MR1138431

[5] ALDOUS, D. and STEELE, J. M. (1992). Asymptotics for Euclidean minimal spanning trees on random points. *Probab. Theory Related Fields* **92** 247–258. MR1161188

[6] ALEXANDER, K. S. (1994). Rates of convergence of means for distance-minimizing subadditive Euclidean functionals. *Ann. Appl. Probab.* **4** 902–922.

[7] ALEXANDER, K. S. (1995). Percolation and minimal spanning forests in infinite graphs. *Ann. Probab.* **23** 87–104.

[8] ALEXANDER, K. S. (1996). The RSW theorem for continuum percolation and the CLT for Euclidean minimal spanning trees. *Ann. Appl. Probab.* **6** 466–494. MR1398054

[9] ALEXANDER, K. S. and MOLCHANOV, S. A. (1994). Percolation of level sets for two-dimensional random fields with lattice symmetry. *J. Stat. Phys.* **77** 627–643. MR1301459

[10] AVRAM, F. and BERTSIMAS, D. (1992). The minimum spanning tree constant in geometrical probability and under the independent model: A unified approach. *Ann. Appl. Probab.* **2** 113–130. MR1143395

[11] AVRAM, F. and BERTSIMAS, D. (1993). On central limit theorems in geometrical probability. *Ann. Appl. Probab.* **3** 1033–1046. MR1241033

[12] BAI, Z. D., LEE, S. and PENROSE, M. D. (2006). Rooted edges of a minimal directed spanning tree on random points. *Adv. in Appl. Probab.* **38** 1–30. MR2213961

[13] BALDI, P. and RINOTT, Y. (1989). On normal approximations of distributions in terms of dependency graphs. *Ann. Probab.* **17** 1646–1650. MR1048950

[14] BALDI, P., RINOTT, Y. and STEIN, C. (1989). A normal approximation for the number of local maxima of a random function on a graph. In *Probability, Statistics, and Mathematics* 59–81. Academic Press, Boston, MA. MR1031278

[15] BARBOUR, A. D. (1990). Stein's method for diffusion approximations. *Probab. Theory Related Fields* **84** 297–322.

[16] BARBOUR, A. D., KAROŃSKI, M. and RUCIŃSKI, A. (1989). A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B* **47** 125–145. MR1047781

[17] BEARDWOOD, J., HALTON, J. H. and HAMMERSLEY, J. M. (1959). The shortest path through many points. *Math. Proc. Cambridge Philos. Soc.* **55** 299–327.

[18] BHATT, A. G. and ROY, R. (2004). On a random directed spanning tree. *Adv. in Appl. Probab.* **36** 19–42. MR2035772

[19] BOLLOBÁS, B. and RIORDAN, O. (2006). *Percolation*. Cambridge Univ. Press, Cambridge. MR2283880

[20] BOLTHAUSEN, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. *Probab. Theory Related Fields* **66** 379–386. MR0751577

[21] BURTON, R. M. and KEANE, M. (1989). Density and uniqueness in percolation. *Comm. Math. Phys.* **121** 501–505. MR0990777

[22] CAMIA, F., FONTES, L. R. and NEWMAN, C. M. (2006). Two-dimensional scaling limits via marked nonsimple loops. *Bull. Braz. Math. Soc.* (*N.S.*) **37** 537–559.

[23] CERF, R. (2013). A lower bound on the two-arms exponent for critical percolation on the lattice. *Ann. Probab.* **43** 2458–2480.

[24] CHATTERJEE, S. (2008). A new method of normal approximation. *Ann. Probab.* **36** 1584–1610.

[25] CHATTERJEE, S. (2009). Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields* **143** 1–40.

[26] CHATTERJEE, S. and SOUNDARARAJAN, K. (2012). Random multiplicative functions in short intervals. *Int. Math. Res. Not. IMRN* **2012** 479–492.

[27] CHEN, L. H. Y. and SHAO, Q.-M. (2004). Normal approximation under local dependence. *Ann. Probab.* **32** 1985–2028. MR2073183

[28] DUMINIL-COPIN, H., IOFFE, D. and VELENIK, Y. (2016). A quantitative Burton–Keane estimate under strong FKG condition. *Ann. Probab.* **44** 3335–3356. MR3551198

[29] FRIEZE, A. M. (1985). On the value of a random minimum spanning tree problem. *Discrete Appl. Math.* **10** 47–56. MR0770868

[30] GANDOLFI, A., GRIMMETT, G. and RUSSO, L. (1988). On the uniqueness of the infinite cluster in the percolation model. *Comm. Math. Phys.* **114** 549–552. MR0929129

[31] GOLDSTEIN, L. and REINERT, G. (1997). Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* **7** 935–952. MR1484792

[32] GOLDSTEIN, L. and RINOTT, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33** 1–17. MR1371949

[33] GRIMMETT, G. (1999). *Percolation*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* **321**. Springer, Berlin. MR1707339

[34] HÄGGSTRÖM, O. (1995). Random-cluster measures and uniform spanning trees. *Stochastic Process. Appl.* **59** 1–75. MR1357655

[35] JANSON, S. (1995). The minimal spanning tree in a complete graph and a functional limit theorem for trees in a random graph. *Random Structures Algorithms* **7** 337–355. MR1369071

[36] KESTEN, H. and LEE, S. (1996). The central limit theorem for weighted minimal spanning trees on random points. *Ann. Appl. Probab.* **6** 495–527. MR1398055

[37] KOZMA, G. and NACHMIAS, A. (2010). Arm exponents in high dimensional percolation. *J. Amer. Math. Soc.* **24** 375–409. MR2748397

[38] LACHIÉZE-REY, R. and PECCATI, G. (2015). New Kolmogorov bounds for functionals of binomial point processes. Preprint. Available at http://arxiv.org/abs/1505.04640.

[39] LAST, G., PECCATI, G. and SCHULTE, M. (2014). Normal approximation on Poisson spaces: Mehler's formula, second order Poincaré inequalities and stabilization. Preprint. Available at http://arxiv.org/abs/1401.7568.

[40] LEE, S. (1997). The central limit theorem for Euclidean minimal spanning trees. I. *Ann. Appl. Probab.* **7** 996–1020. MR1484795

[41] LEE, S. (1999). The central limit theorem for Euclidean minimal spanning trees II. *Adv. in Appl. Probab.* **31** 969–984. MR1747451

[42] LYONS, R., PERES, Y. and SCHRAMM, O. (2006). Minimal spanning forests. *Ann. Probab.* **34** 1665–1692. MR2271476

[43] MEESTER, R. and ROY, R. (1996). *Continuum Percolation*. Cambridge Univ. Press, Cambridge. MR1409145

[44] PENROSE, M. D. (1996). The random minimal spanning tree in high dimensions. *Ann. Probab.* **24** 1903–1925. MR1415233

[45] PENROSE, M. D. (1997). The longest edge of the random minimal spanning tree. *Ann. Appl. Probab.* **7** 340–361. MR1442317

[46] PENROSE, M. D. (1998). Random minimal spanning tree and percolation on the N-cube. *Random Structures Algorithms* **12** 63–82. MR1637395

[47] PENROSE, M. D. (2003). *Random Geometric Graphs*. Oxford University Press, Oxford. MR1986198

[48] PENROSE, M. D. and WADE, A. R. (2004). Random minimal directed spanning trees and Dickman-type distributions. *Adv. in Appl. Probab.* **36** 691–714. MR2079909

[49] PENROSE, M. D. and YUKICH, J. E. (2003). Weak laws of large numbers in geometric probability. *Ann. Appl. Probab.* **13** 277–303. MR1952000

[50] PENROSE, M. D. and YUKICH, J. E. (2005). Normal approximation in geometric probability. In *Stein's Method and Applications* (A. D. Barbour and L. H. Y. Chen, eds.). *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **5** 37–58. Singapore Univ. Press, Singapore. MR2201885

[51] PETE, G., GARBAN, C. and SCHRAMM, O. (2013). The scaling limits of the Minimal Spanning Tree and Invasion Percolation in the plane. Preprint. Available at http://arxiv.org/abs/1309.0269.

[52] RINOTT, Y. and ROTAR, V. (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. *Ann. Appl. Probab.* **7** 1080–1105. MR1484798

[53] ROY, R. (1990). The Russo–Seymour–Welsh theorem and the equality of critical densities and the "dual" critical densities for continuum percolation on $\mathbb{R}^2$. *Ann. Probab.* **18** 1563–1575. MR1071809

[54] SMIRNOV, S. and WERNER, W. (2001). Critical exponents for two-dimensional percolation. *Math. Res. Lett.* **8** 729–744. MR1879816

[55] STEELE, J. M. (1981). Subadditive Euclidean functionals and nonlinear growth in geometric probability. *Ann. Probab.* **9** 365–376. MR0626571

[56] STEELE, J. M. (1987). On Frieze's $\zeta(3)$ limit for lengths of minimal spanning trees. *Discrete Appl. Math.* **18** 99–103. MR0905183

[57] STEELE, J. M. (1988). Growth rates of Euclidean minimal spanning trees with power weighted edges. *Ann. Probab.* **16** 1767–1787. MR0958215

[58] STEIN, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proc. of the Sixth Berkeley Sympos. Math. Statist. Probab., Vol. II: Probability Theory* 583–602. Univ. California Press, Berkeley, CA. MR0402873

[59] STEIN, C. (1986). *Approximate Computation of Expectations*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **7**. IMS, Hayward, CA. MR0882007

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
390 SERRA MALL, SEQUOIA HALL
STANFORD, CALIFORNIA 94305
USA
E-MAIL: souravc@stanford.edu

DEPARTMENT OF MATHEMATICS AND STATISTICS
MCGILL UNIVERSITY
805 SHERBROOKE STREET WEST
MONTRÉAL, QUÉBEC H3A 0B9
CANADA
E-MAIL: sanchayan.sen1@gmail.com