Spectral dimension of simple random walk on a long-range percolation cluster[*]

V. H. Can[†‡]   D. A. Croydon[§]   T. Kumagai[¶]

Abstract

Consider the long-range percolation model on the integer lattice $\mathbb{Z}^d$ in which all nearest-neighbour edges are present and otherwise x and y are connected with probability $q_{x,y} := 1 - \exp(-|x - y|^{-s})$, independently of the state of other edges. Throughout the regime where the model yields a locally-finite graph, (i.e. for $s > d$,) we determine the spectral dimension of the associated simple random walk, apart from at the exceptional value $d = 1$, $s = 2$, where the spectral dimension is discontinuous. Towards this end, we present various on-diagonal heat kernel bounds, a number of which are new. In particular, the lower bounds are derived through the application of a general technique that utilises the translation invariance of the model. We highlight that, applying this general technique, we are able to partially extend our main result beyond the nearest-neighbour setting, and establish lower heat kernel bounds over the range of parameters $s \in (d, 2d)$. We further note that our approach is applicable to short-range models as well.

[†]Institute of Mathematics, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Cau Giay, Hanoi, Vietnam.
[‡]Department of Statistics and Data Science, National University of Singapore, 6 Science Drive 2 Singapore 117546. E-mail: cvhao89@gmail.com
[§]Research Institute for Mathematical Sciences, Kyoto University, Kyoto 606-8502, Japan.
E-mail: croydon@kurims.kyoto-u.ac.jp
[¶]Department of Mathematics, Faculty of Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan. E-mail: t-kumagai@waseda.jp

1 Introduction

The study of random walks on percolation clusters on the integer lattice $\mathbb{Z}^d$ goes back a long way, at least as far as de Gennes' 1976 description of such a process as an 'ant in a labyrinth' [27]. Mathematically, diffusive scaling limits were first established with respect to the so-called annealed/averaged law, under which both the random process and environment are integrated out [25]. More recently, building on the Gaussian heat kernel estimates of [2], scaling limits under the quenched law (that is, for typical realisations of the environment) have also been obtained [11, 34, 35]. When the random walk is strongly recurrent, some general theory has been established to obtain on-diagonal heat kernel estimates, and such methods have been used to identify the spectral dimension, which is the exponent governing the on-diagonal decay of the heat kernel, of the random walk on critical percolation clusters conditioned to be infinite (see for example [6, 7, 29, 31, 30]). Whilst the works cited so far have dealt with the nearest-neighbour case, in which only edges between points in $\mathbb{Z}^d$ a unit Euclidean distance apart are considered, it is natural to generalise the model to allow the possibility of edges spanning arbitrarily large distances. In the last decade, substantial progress has been made in understanding random walks on such long-range percolation models, most notably in [17, 18], some of the main results of which are recalled below. Our contribution in this paper is twofold:

(i) To give general sufficient conditions for an on-diagonal lower bound of the heat kernel for random walk under both the quenched and annealed laws on stationary random media (see Theorem 2.3 and Corollary 2.6 below);

(ii) To determine the spectral dimension of random walk on a long-range percolation cluster, under both the quenched and annealed laws, throughout (almost) the entire range of parameters for which the model is defined (see Theorems 1.3, 1.5 and Corollary 1.7).

Concerning (i), we note that the previous work in [6, 31, 30] applies only for strongly recurrent random walks, and the sufficient conditions the latter articles describe for heat kernel lower bounds are rather complicated. Instead, we use stationarity of the model and a useful estimate from [32, Theorem 3.7], which was motivated by the problem of understanding the behaviour of the random walk on certain planar random graphs, such as the uniform infinite planar triangulation/quadrangulation. Details are discussed in Section 2. Although the techniques we develop are principally targeted at understanding random walk on long-range percolation clusters, we note they are also applicable to short-range models. As a basic example of such, we discuss their use for studying the random walk on the integer lattice in Section 6.1.

To present the background and results concerning (ii) more precisely, we proceed to introduce the main application of interest in this paper. Specifically, we consider a long-range percolation model with vertex set $\mathbb{Z}^d$, where $d \geq 1$. For simplicity, in the introduction we suppose that all nearest-neighbour edges are present, although we will later discuss a generalisation of this. For any $x, y \in \mathbb{Z}^d$ with $|x - y| > 1$, we suppose the edge between them appears with probability

$$q_{x,y} = 1 - \exp(-|x - y|^{-s}), \tag{1.1}$$

independently of the state of other edges. (For most of the subsequent discussion, we could weaken the tail assumption on $q_{x,y}$. Indeed, for our heat kernel estimates in Theorems 1.3 and 1.5 below, it would be enough to assume that $c_1|x-y|^{-s} \leq q_{x,y} \leq c_2|x-y|^{-s}$. We choose to restrict to the specific choice of $q_{x,y}$ above simply for convenience.) The parameter s is called the exponent of the long-range percolation model. In order to ensure that each vertex is directly connected by an edge to only a finite number of
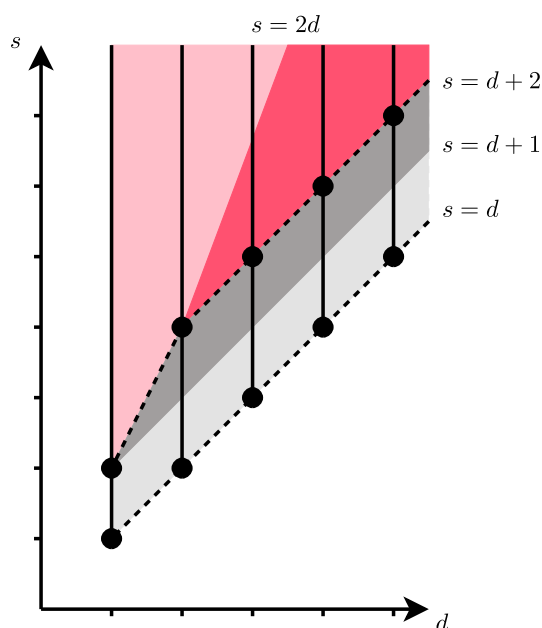
Figure 1: Heat kernel regimes for the long-range percolation model studied in this article. The two pink regions correspond to the Gaussian regime, that is $s > \min\{d + 2, 2d\}$. Gaussian scaling limits are known to hold in the lighter pink region ($s > 2d$). Our results provide new heat kernel lower bounds in the darker pink region ($d \geq 2$ and $s \in (\min\{d + 2, 2d\}, 2d]$), and establish that the spectral dimension is d throughout both regions. The two grey regions correspond to the stable regime, that is $s \in (d, \min\{d + 2, 2d\})$. Stable scaling limits are known to hold in the lighter grey region ($s \in (d, d + 1)$). We give new heat kernel lower bounds in the darker grey region ($s \in [d + 1, d + 2)$), and establish that the spectral dimension of the model is given by $2d/(s - d)$ throughout both regions. As confirmed by Corollary 1.7, when $d = 1$, there is a discontinuity in the spectral dimension at $s = 2$. Apart from when $d = 1$, we also provide estimates along the critical line $s = \min\{d + 2, 2d\}$.

vertices (as is required to define the associated discrete-time random walk), we assume that s takes a value strictly greater than d. We denote the resulting random graph by $G = (V, E)$, and take the root ρ to be the origin in $\mathbb{Z}^d$. Moreover, we will use the notation LRP(d, s) to represent this model, and suppose it is built on a probability space with probability measure $\mathbf{P}$ and expectation $\mathbf{E}$.

Providing some context for our results, the following summarises scaling limits that are known to hold for the discrete-time simple random walk on LRP(d,s). Given the environment G, this process, which we will denote by $(X_n)_{n \geq 0}$, jumps on each time step from its current location to a uniformly-chosen neighbour in the graph G. In the subsequent theorem, it is further assumed that $X_0 = \rho$. Part (a), which concerns the stable regime, was established in [18, Theorem 1.1]. As for the Gaussian regime of part (b), the $d = 1$ case was dealt with in [18, Theorem 1.2] (see also [36]), and the $d \geq 2$ case in [14]. (We give a new argument for $d = 1$, $s > 2$ in Section 6.3 below.) Note that both the stable and Gaussian regimes are thought to be incomplete (see discussion in [17, 18] and [14, Problem 2.9]), and our heat kernel bounds support conjectures about how they extend. Figure 1 gives a graphical overview of the situation.

Theorem 1.1 (Long-range percolation, scaling limits, [14, 18, 36])**.**
(a) If $d \geq 1$ and $s \in (d, d + 1)$, then for $\mathbf{P}$-a.e. realisation of LRP(d, s) and every $p \in [1, \infty)$,

the law of
$$\left(n^{-\frac{1}{s-d}}X_{nt}\right)_{t\in[0,1]}$$

on $L^p([0,1])$ converges weakly to the law of an isotropic α-stable Lévy process with $\alpha = s - d$.

(b) If $d \geq 1$ and $s > 2d$, then for $\mathbf{P}$-a.e. realisation of LRP(d, s), the law of
$$\left(n^{-\frac{1}{2}}X_{nt}\right)_{t\geq 0}$$

on $C([0, \infty))$ converges weakly to that of $(B_{\sigma^2 t})_{t\geq 0}$, where $(B_t)_{t\geq 0}$ is standard Brownian motion on $\mathbb{R}^d$, and $\sigma^2 \in (0, \infty)$ is a deterministic constant.

Remark 1.2. (i) To make the statement of the above theorem completely accurate, we need to describe a convention for determining the value of $(X_t)_{t\geq 0}$ between integer times. For both parts above, the result would hold if one were to do this by linear interpolation. Alternatively, using the J_1 topology on the Skorohod space $D([0, \infty))$ for part (b) above, one could consider $X_{\lfloor t \rfloor}$ in place of X_t. In this article, we will henceforth adopt the latter approach; that is, if we write a continuous variable, x say, where a discrete argument is required, we suppose it should be treated as $\lfloor x \rfloor$.

(ii) Note that part (a) above is proved in [18] without the assumption of nearest-neighbour edges being present, which requires a substantial amount of extra work to deal with the percolation issues involved.

We next set out our heat kernel estimates for LRP(d,s). Given G, the (quenched) heat kernel/transition density of X is defined by setting
$$p_t^G(x, y) := \frac{P_x^G(X_t = y)}{\deg_G(y)}, \qquad \forall x, y \in \mathbb{Z}^d, \ t \geq 0,$$

where P_x^G is the (quenched) law of X started from $X_0 = x$, and $\deg_G(y)$ is the usual graph degree of y in G. For typical realisations of the environment, we have the following bounds. The constants c_i and δ_i are deterministic, and the δ_i in particular are discussed in the subsequent remark. We further highlight that, in the parameter regimes where scaling limits are known, the lower heat kernel bounds follow from a general argument adapted from [13] (see Lemma 6.2 below). The main contribution of this article is in establishing the remaining lower bounds, which we do by developing [32, Theorem 3.7] (cited below as Proposition 2.1). As for the upper bounds, the result in the stable case was previously known from [17, Theorem 1]. This was based on a general argument for checking quenched heat kernel upper bounds on random media, which we believe would also be appropriate in the Gaussian case. However, we use another argument based on comparison with a simple random walk and a time change, which more easily adapts to the annealed case, and allows us to remove the logarithmic terms there. For discussion of the case $d = 1$, $s = 2$, see Remark 1.8 below.

Theorem 1.3 (Long-range percolation, quenched bounds)**.**
(a) If $d \geq 1$ and $s \in (d, \min\{d + 2, 2d\})$, then LRP$(d, s)$ satisfies, $\mathbf{P}$-a.s., for all $t \in \mathbb{N}$ large enough,
$$c_1 t^{-\frac{d}{s-d}}(\log t)^{-\delta_1} \leq p_{2t}^G(\rho, \rho) \leq c_2 t^{-\frac{d}{s-d}}(\log t)^{\delta_2}. \tag{1.2}$$

(b) If $d = 1$ and $s > 2$, then LRP(d, s) satisfies, $\mathbf{P}$-a.s., for all $t \in \mathbb{N}$ large enough,
$$c_3 t^{-\frac{1}{2}} \leq p_{2t}^G(\rho, \rho) \leq c_4 t^{-\frac{1}{2}}. \tag{1.3}$$

The upper bound holds for $d = 1$ and $s = 2$ as well. (c) If $d \geq 2$ and $s \geq d + 2$, then LRP(d, s) satisfies, $\mathbf{P}$-a.s., for all $t \in \mathbb{N}$ large enough,
$$c_5 t^{-\frac{d}{2}}(\log t)^{-\delta_3} \leq p_{2t}^G(\rho, \rho) \leq c_6 t^{-\frac{d}{2}}(\log t)^{\delta_4}. \tag{1.4}$$

Remark 1.4. In the above result, we can take

$$\delta_1 := \left(\frac{4s}{s-d} + \varepsilon \right) \mathbb{I}_{\{s \in [d+1, d+2)\}};$$

note that this means we do not need a log term in the regime where there is a scaling limit. As indicated above, the upper bound in (1.2) is essentially due to [17], with δ_2 being as given by the δ of [17, Theorem 1], which is not explicit. Similarly to Remark 1.2 above, the result of [17] does not require nearest-neighbour bonds to be present in the model. For $s \in (d, 2d)$, which includes the entire stable regime, we explain how to extend the lower bounds of Theorem 1.3 (and Theorem 1.5 below) to the non-nearest-neighbour setting in Section 6.4.

The article [31] gives upper and lower bounds with logs in the case $d = 1$ and $s > 2$; the above theorem improves on this. (In this case, the assumption of nearest-neighbour bonds is not purely for convenience – see the discussion preceding [18, Theorem 1.2].)

We further have

$$\delta_3 := \left(\frac{5d}{2} + 4 + \varepsilon \right) \mathbb{I}_{\{s=d+2\}} + (2d + 4 + \varepsilon) \mathbb{I}_{\{d+2<s\leq 2d\}};$$

again this is zero where we have a scaling limit. And also,

$$\delta_4 := \frac{d-1}{2}.$$

We do not expect any of the δ_is to be the best possible constants. For discussion concerning the removal of the log term in the upper bound of (1.4) in particular, see Remark 4.3 below.

As for the annealed heat kernel, which is obtained by integrating out the randomness of the environment, we have the following.

Theorem 1.5 (Long-range percolation, annealed bounds).
(a) If $d \geq 1$ and $s \in (d, \min\{d+2, 2d\})$, then LRP$(d,s)$ satisfies, for all $t \in \mathbb{N}$,

$$c_1 t^{-\frac{d}{s-d}} \leq \mathbf{E}\left(p_{2t}^G(\rho, \rho) \right) \leq c_2 t^{-\frac{d}{s-d}} (\log t)^{\delta_2}. \tag{1.5}$$

(b) If $d \geq 2$ and $s = d + 2$, then LRP(d,s) satisfies, for all $t \in \mathbb{N}$,

$$c_3 t^{-\frac{d}{2}} (\log t)^{-\delta_5} \leq \mathbf{E}\left(p_{2t}^G(\rho, \rho) \right) \leq c_4 t^{-\frac{d}{2}}. \tag{1.6}$$

(c) If $d \geq 1$ and $s > \min\{d+2, 2d\}$, then LRP(d,s) satisfies, for all $t \in \mathbb{N}$,

$$c_5 t^{-\frac{d}{2}} \leq \mathbf{E}\left(p_{2t}^G(\rho, \rho) \right) \leq c_6 t^{-\frac{d}{2}}. \tag{1.7}$$

The upper bound holds for $d = 1$ and $s = 2$ as well.

Remark 1.6. We can take the same δ_2 as in Theorem 1.3, and

$$\delta_5 := \frac{d}{2} + \varepsilon.$$

For $d = 1$ and $s > 2$, the bounds of (1.7) are obtained in [31, Section 2]. For (1.6), we conjecture that some log correction is necessary, i.e. the upper bound is not sharp. We also anticipate that the upper bound in (1.5) is not sharp, in that no log term is required in this case.
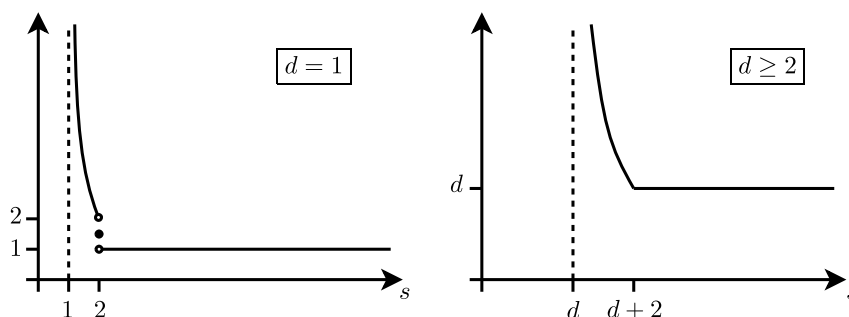
Figure 2: Spectral dimension as a function of s for $d = 1$ (left) and $d \geq 2$ (right). The existence and value of the spectral dimension is not yet known for $d = 1$, $s = 2$, see Remark 1.8 for discussion of this case.

As a straightforward consequence of Theorems 1.3 and 1.5, we can read off the spectral dimension of LRP(d, s) for $d \geq 1$, $s > d$, apart from at the value $d = 1$, $s = 2$. Precisely, the quenched spectral dimension is defined to be the $\mathbf{P}$-a.s. limit

$$d_s^{(q)}(d, s) := -\lim_{t \to \infty} \frac{2 \log p_{2t}^G(\rho, \rho)}{\log t},$$

and the corresponding annealed spectral dimension is the limit

$$d_s^{(a)}(d, s) := -\lim_{t \to \infty} \frac{2 \log \mathbf{E}\left(p_{2t}^G(\rho, \rho)\right)}{\log t}.$$

See Figure 2 for an illustration of the following result.

Corollary 1.7 (Long-range percolation, quenched and annealed spectral dimension)**.**
(a) If $d \geq 1$ and $s \in (d, \min\{d + 2, 2d\})$, then, $\mathbf{P}$-a.s.,

$$d_s^{(a)}(d, s) = d_s^{(q)}(d, s) = \frac{2d}{s - d}.$$

(b) If $d \geq 1$ and $s > \min\{d + 2, 2d\}$ or $d \geq 2$ and $s = d + 2$, then, $\mathbf{P}$-a.s.,

$$d_s^{(a)}(d, s) = d_s^{(q)}(d, s) = d.$$

Remark 1.8. As shown by Corollary 1.7 (and Figure 2), there is a discontinuity in the spectral dimension at $d = 1$, $s = 2$. Whilst it might be possible to argue from the techniques of this article that, if it exists, the spectral dimension lies in the interval [1,2], determining the exact value seems highly non-trivial. See the discussion of [8], [14, Problem 2.10], [26] and [31, Remark 2.3(2)] for further background on the difficulties found in this case.

The remainder of the article is organised as follows. In Section 2, we present our general approach for establishing quenched and annealed heat kernel lower bounds on random media, see Theorem 2.3 and Corollary 2.6, and also discuss the related upper bound of [17]. The various assumptions required to apply these results are checked for long-range percolation in Section 3. Then, in Section 4, we put the pieces together to deduce the lower heat kernel bounds of Theorems 1.3 and 1.5, and also give an argument for the corresponding upper heat kernel bounds. Finally, Section 5 lists some questions left open by this work, and Section 6 is an appendix in which we: explain how the lower heat kernel bound applies to the simpler setting of random walk on $\mathbb{Z}^d$; describe how a quenched scaling limit automatically implies a quenched lower heat

kernel bound; present an alternative proof of a quenched invariance principle for long-range percolation in the one-dimensional setting; and describe an extension of our lower heat kernel bounds to a long-range percolation model in which non-nearest-neighbour bonds are not necessarily present.

Concerning notational conventions, we write $x \wedge y := \min\{x, y\}$ and $x \vee y := \max\{x, y\}$. For non-negative sequences $f(n)$ and $g(n)$, we define $f(n) \asymp g(n)$ to mean that there exist strictly positive constants c_1 and c_2 such that $c_1 f(n) \leq g(n) \leq c_2 f(n)$, and $f(n) \preceq g(n)$ to mean that there exists a strictly positive constant c such that $f(n) \leq cg(n)$. We write c, C for deterministic constants that might change value from line to line.

2 Heat kernel estimates

2.1 Lower heat kernel bound

Towards establishing our lower heat kernel bounds, we present a result from [32] that shows a lower heat kernel bound must hold on some proportion of vertices in a graph, as determined by the sizes and capacities of the pieces in a suitable decomposition of the graph. The latter result is given for an arbitrary connected, finite graph $G = (V, E)$, where V is a set of vertices and E is a set of bonds. We set $\tilde{\pi}(x) = \deg_G(x)$ for all $x \in V$, and define π to be a version of $\tilde{\pi}$ normalised to be a probability measure, namely

$$\pi(x) = \frac{\deg_G(x)}{2|E|}, \qquad \forall x \in V,$$

where for a set A, we denote by $|A|$ the number of elements of A. Note that π is the stationary probability measure for the discrete-time simple random walk associated with G. We further write, for any $\varepsilon > 0$,

$$\pi^\star(\varepsilon) = \max\left\{\pi(W) : |W| \leq \varepsilon|V|\right\}. \tag{2.1}$$

The (on-diagonal part of the) natural Dirichlet form on G is defined for functions $\psi : V \to \mathbb{R}$ as follows:

$$\mathcal{E}(\psi) = \frac{1}{2} \sum_{x,y:\, x \sim y} (\psi(x) - \psi(y))^2, \tag{2.2}$$

where we write $x \sim y$ to mean that x and y are connected by an edge in E. As a final piece of notation needed to state the result of [32], let us call a pair (A, Ω) of subsets $A \subseteq \Omega \subseteq V$ a capacitor, and define the capacity of (A, Ω) by

$$\mathrm{cap}_\Omega(A) := \inf_{\psi:V:\to[0,1]} \left\{2\mathcal{E}(\psi) :\, \psi|_A = 1,\, \mathrm{supp}(\psi) \subseteq \Omega\right\},$$

where $\mathrm{supp}(\psi) := \{x \in V :\, \psi(x) \neq 0\}$. Note that our definition of capacity differs from the definition in [32] by a factor of $2|E|$.

Proposition 2.1 ([32, Theorem 3.7]). Let $G = (V, E)$ be a connected, finite graph, and $M \in (0, \infty)$ a constant. Suppose that for some $k \leq |V|$, there are capacitors $(A_1, \Omega_1), \ldots, (A_k, \Omega_k)$ such that $(\Omega_i)_{i=1}^k$ are pairwise disjoint and $|\Omega_i| \leq M$ for all $i = 1, \ldots, k$. Then, for all $\varepsilon > 0$ and $t \in \mathbb{N}$,

$$\pi\left(\left\{x \in V :\, p_{2t}^G(x,x) \geq \frac{\varepsilon|V|}{8M|E|}\right\}\right) \geq -2\pi^\star(\varepsilon) + \sum_{i=1}^k \pi(A_i) - \frac{t}{|E|} \sum_{i=1}^k \mathrm{cap}_{\Omega_i}(A_i).$$

Remark 2.2. The appearance of the above result in [32] is slightly different, as it is expressed in terms of the transition probability, rather than the heat kernel/transition density.

We will apply this result to sequences of graphs that converge in the sense of Benjamini-Schramm [9]. In particular, for the limit of the sequence, we take G to be a random connected, locally-finite graph, rooted at a distinguished vertex ρ. For each $n \geq 1$, G_n will be a random connected, finite graph, rooted at a uniformly chosen vertex ρ_n. Given any $k \geq 1$, we suppose that the ball in G_n of radius k (according to the usual shortest path graph distance) centred at the root ρ_n converges in distribution to the ball in G of radius k centred at the root ρ (see [9, Section 1.2] for details). To apply Proposition 2.1 to such a sequence, for given non-negative constants α, $(\delta_i)_{i=0}^{3}$ and γ, and deterministic function $\lambda : (0, \infty) \to [1, \infty)$, we consider the following conditions (that are in fact relevant to any sequence of connected, finite graphs).

(A1) For all $n \in \mathbb{N}$,

$$\mathbf{E}\left(\frac{|E_n|^2}{|V_n|^2}\right) \leq \alpha,$$

where E_n and V_n are the edge and vertex set of G_n, respectively.

(A2) For all $\varepsilon > 0$, if n is large enough, then

$$\mathbf{E}\left(\pi_n^\star(\varepsilon)^2\right) \leq \alpha\varepsilon,$$

where $\pi_n^\star$ is defined as at (2.1) from π_n, the stationary probability measure of simple random walk on G_n.

(A3) For each $t \in \mathbb{N}$, there exists an integer $n_0 = n_0(t)$ such that for each $n \geq n_0$, with probability at least $1 - \lambda(t)^{-\delta_0}$, there are $k \leq |V_n|$ capacitors $(A_1, \Omega_1), \ldots, (A_k, \Omega_k)$ for the graph G_n such that $(\Omega_i)_{i=1}^{k}$ are pairwise disjoint and also

    (a) $\max_{i=1,\ldots,k} |\Omega_i| \leq \alpha t^\gamma \lambda(t)^{\delta_1}$;
    (b) $\sum_{i=1}^{k} \pi_n(A_i) \geq 1 - \alpha\lambda(t)^{-\delta_2}$;
    (c) $\sum_{i=1}^{k} \mathrm{cap}_{\Omega_i}(A_i) \leq 2\alpha|E_n|t^{-1}\lambda(t)^{-\delta_3}$.

We are now ready to state the main result of this section. The assumptions (A1), (A2) and (A3) are clearly designed to feed into the bound of Proposition 2.1.

Theorem 2.3. Assume that (G, ρ) is the Benjamini-Schramm limit of the sequence (G_n, ρ_n), $n \geq 1$, which satisfies (A1), (A2), (A3) for some positive constants α, $(\delta_i)_{i=0}^{3}$ and γ, and deterministic function $\lambda : (0, \infty) \to [1, \infty)$. There then exists a constant C_α only depending on α such that, for all $t \in \mathbb{N}$ large enough,

$$\mathbf{P}\left(p_{2t}^G(\rho, \rho) \geq \frac{1}{t^\gamma \lambda(t)^{\delta_1 + 2(\delta_2 \wedge \delta_3)}}\right) \geq 1 - C_\alpha \lambda(t)^{-\frac{\delta_0 \wedge \delta_2 \wedge \delta_3}{2}}.$$

Proof. Since (G_n, ρ_n) converges in a Benjamini-Schramm sense to (G, ρ), and $p_{2t}^G(\rho, \rho)$ only depends on the ball of radius $2t$ about ρ, it holds that, for each fixed t,

$$\lim_{n \to \infty} \mathbf{P}\left(p_{2t}^{G_n}(\rho_n, \rho_n) < \frac{1}{t^\gamma \lambda(t)^{\delta_1 + 2(\delta_2 \wedge \delta_3)}}\right) = \mathbf{P}\left(p_{2t}^G(\rho, \rho) < \frac{1}{t^\gamma \lambda(t)^{\delta_1 + 2(\delta_2 \wedge \delta_3)}}\right).$$

Hence, for each fixed t and δ, if n is large, then

$$\left|\mathbf{P}\left(p_{2t}^{G_n}(\rho_n, \rho_n) < \frac{1}{t^\gamma \lambda(t)^{\delta_1 + 2(\delta_2 \wedge \delta_3)}}\right) - \mathbf{P}\left(p_{2t}^G(\rho, \rho) < \frac{1}{t^\gamma \lambda(t)^{\delta_1 + 2(\delta_2 \wedge \delta_3)}}\right)\right| \leq \lambda(t)^{-\delta}.$$

$$(2.3)$$

Moreover, writing $\mathbb{I}$ for an indicator function,

$$
\begin{aligned}
& \mathbf{P}\left(p_{2t}^{G_n}(\rho_n, \rho_n) < \frac{1}{t^\gamma \lambda(t)^{\delta_1 + 2(\delta_2 \wedge \delta_3)}}\right) \\
&= \mathbf{E}\left[\mathbf{P}\left(p_{2t}^{G_n}(\rho_n, \rho_n) < \frac{1}{t^\gamma \lambda(t)^{\delta_1 + 2(\delta_2 \wedge \delta_3)}}\Big| G_n\right)\right] \\
&= \mathbf{E}\left[\frac{1}{|V_n|} \sum_{x \in V_n} \mathbb{I}\left(\left\{p_{2t}^{G_n}(x, x) < \frac{1}{t^\gamma \lambda(t)^{\delta_1 + 2(\delta_2 \wedge \delta_3)}}\right\}\right)\right] \\
&\leq \mathbf{E}\left[\frac{2|E_n|}{|V_n|} \sum_{x \in V_n} \frac{\deg_{G_n}(x)}{2|E_n|} \mathbb{I}\left(p_{2t}^{G_n}(x, x) < \frac{1}{t^\gamma \lambda(t)^{\delta_1 + 2(\delta_2 \wedge \delta_3)}}\right)\right] \\
&= 2\mathbf{E}\left[\frac{|E_n|}{|V_n|} \pi_n\left(\left\{x \in V_n : p_{2t}^{G_n}(x, x) < \frac{1}{t^\gamma \lambda(t)^{\delta_1 + 2(\delta_2 \wedge \delta_3)}}\right\}\right)\right] \\
&\leq 2\mathbf{E}\left[\frac{|E_n|^2}{|V_n|^2}\right]^{1/2} \mathbf{E}\left[\pi_n\left(\left\{x \in V_n : p_{2t}^{G_n}(x, x) < \frac{1}{t^\gamma \lambda(t)^{\delta_1 + 2(\delta_2 \wedge \delta_3)}}\right\}\right)^2\right]^{1/2} \\
&\leq 2\sqrt{\alpha}\mathbf{E}\left[\pi_n\left(\left\{x \in V_n : p_{2t}^{G_n}(x, x) < \frac{1}{t^\gamma \lambda(t)^{\delta_1 + 2(\delta_2 \wedge \delta_3)}}\right\}\right)^2\right]^{1/2},
\end{aligned}
\tag{2.4}
$$

where we have applied (A1) to deduce the final inequality. Now, suppose $n_0 = n_0(t)$ is an integer as in (A3). Applying (A3) and Proposition 2.1 with $M = \alpha t^\gamma \lambda(t)^{\delta_1}$ and $\varepsilon = 8\alpha\lambda(t)^{-(\delta_2 \wedge \delta_3)}$, we obtain that, for $n \geq n_0$, on an event of probability at least $1 - \lambda(t)^{-\delta_0}$,

$$
\begin{aligned}
& \pi_n\left(\left\{x \in V_n : p_{2t}^{G_n}(x, x) < \frac{1}{t^\gamma \lambda(t)^{\delta_1 + 2(\delta_2 \wedge \delta_3)}}\right\}\right) \mathbb{I}\left(\frac{|E_n|}{|V_n|} \leq \lambda(t)^{\delta_2 \wedge \delta_3}\right) \\
&\leq 2\pi_n^\star(\varepsilon) + C_\alpha \lambda(t)^{-(\delta_2 \wedge \delta_3)},
\end{aligned}
$$

where C_α is a constant that only depends on α. Therefore, by (A1) and (A2), for large n,

$$
\begin{aligned}
& \mathbf{E}\left[\pi_n\left(\left\{x \in V_n : p_{2t}^{G_n}(x, x) < \frac{1}{t^\gamma \lambda(t)^{\delta_1 + 2(\delta_2 \wedge \delta_3)}}\right\}\right)^2\right]^{1/2} \\
&\leq C_\alpha'\left(\mathbf{E}[\pi_n^\star(\varepsilon)^2] + \lambda(t)^{-2(\delta_2 \wedge \delta_3)} + \lambda(t)^{-\delta_0}\right)^{1/2} \\
&\leq C_\alpha'' \lambda(t)^{-\frac{\delta_0 \wedge \delta_2 \wedge \delta_3}{2}}.
\end{aligned}
\tag{2.5}
$$

Combining (2.3), (2.4) and (2.5), we get that, for all t large enough,

$$
\mathbf{P}\left(p_{2t}^G(\rho, \rho) < \frac{1}{t^\gamma \lambda(t)^{\delta_1 + 2(\delta_2 \wedge \delta_3)}}\right) \leq C_\alpha''' \lambda(t)^{-\frac{\delta_0 \wedge \delta_2 \wedge \delta_3}{2}},
$$

which implies the desired result. □

Remark 2.4. When the assumption (A3) is hard to check for *all* integers n, we could replace (A3) by (A3'): There exists an integer $n_1 = n_1(t)$ such that for $n = n_1$, the conditions (A3)(a)-(c) hold and, moreover,

$$
\mathbf{P}\left(p_{2t}^{G_n}(\rho_n, \rho_n) < \frac{1}{t^\gamma \lambda(t)^{\delta_1 + 2(\delta_2 \wedge \delta_3)}}\right) \leq \mathbf{P}\left(p_{2t}^G(\rho, \rho) < \frac{2}{t^\gamma \lambda(t)^{\delta_1 + 2(\delta_2 \wedge \delta_3)}}\right) + \lambda(t)^{-\delta_0}.
$$

The above condition gives a replacement of (2.3); the other parts of the proof of Lemma 2.3 are exactly the same.

Remark 2.5. We note that Benjamini-Schramm convergence is a key input into the proof, allowing us to transfer a heat kernel estimate from a large, but unspecified, set to a single specified point. This is somewhat analogous to the approach used to understand the on-diagonal part of the annealed heat kernel of Brownian motion on stable trees in [22, 23], whereby random re-rooting was used to obtain point-wise asymptotics for the heat kernel from the asymptotics of the trace of the heat semigroup, which is typically a smoother object.

To complete the section, we give a corollary that explains how the distributional bound of Theorem 2.3 can be applied to yield quenched and annealed lower heat kernel bounds.

Corollary 2.6. Suppose that (G, ρ) is the Benjamini-Schramm limit of the sequence (G_n, ρ_n), $n \geq 1$.
(a) Assume that (A1),(A2), (A3) hold for some positive constants α, $(\delta_i)_{i=0}^3$ and γ, and deterministic function $\lambda : (0, \infty) \to [1, \infty)$ satisfying $\sum_{i=1}^{\infty} \lambda(e^i)^{-\frac{\delta_0 \wedge \delta_2 \wedge \delta_3}{2}} < \infty$ and $c_1 \leq \lambda(t)/\lambda(e^{i+1})$ for all $e^i \leq t \leq e^{i+1}$ and all $i \in \mathbb{N}$. Then, **P**-a.s., for all $t \in \mathbb{N}$ large enough

$$p_{2t}^G(\rho, \rho) \geq \frac{c_2}{t^\gamma \lambda(t)^{\delta_1 + 2(\delta_2 \wedge \delta_3)}}.$$

(b) Assume that (A1),(A2), (A3) hold for some positive constants α, $(\delta_i)_{i=0}^3$ and γ, and constant function $\lambda : (0, \infty) \to [1, \infty)$ given by $\lambda(t) = \lambda_0$, where $1 - C_\alpha \lambda_0^{-\frac{\delta_0 \wedge \delta_2 \wedge \delta_3}{2}} > 0$. Then

$$\mathbf{E}\left(p_{2t}^G(\rho, \rho)\right) \geq \frac{c(\alpha, \lambda_0)}{t^\gamma}, \qquad \forall t \in \mathbb{N},$$

where $c(\alpha, \lambda_0)$ is a constant depending on the values of α and λ_0.

Proof. (a) Given Theorem 2.3, this follows from the Borel-Cantelli lemma and the monotonicity of the on-diagonal part of the heat kernel $p_{2t}^G(\rho, \rho)$. (That the latter quantity is decreasing in t can be seen from [4, Lemma 4.1] and [21, Equation (15)], for example.)

(b) This is an obvious consequence of Theorem 2.3. □

2.2 Upper heat kernel bound

In [17, Lemma 3.1], a general heat kernel upper bound was given. For comparison with the approach of the previous subsection, we summarize it here. In both [17] and this article, a key aspect of the required input is that we can decompose a large part of the graph in question into suitably-sized pieces that behave well in some way. Here, the focus is on the capacity of the pieces; in [17], it is their spectral gap that plays a central role.

To present the setting of [17, Section 3], let G be a connected, locally-finite graph, and $(Y_t)_{t \geq 0}$ be the continuous-time random walk on G with unit mean holding times. For any connected, finite subgraph $H \subseteq G$ and any vertex $x \in H$, we denote by $\deg_H(x)$ the degree of x within H. The stationary measure of the random walk on H is given by

$$\pi_H(x) = \frac{\deg_H(x)}{\sum_{y \in H} \deg_H(y)}.$$

Moreover, the spectral gap of H is defined as

$$\mathrm{Gap}_H = \inf \left\{ \frac{\sum_{x,y \in H : x \sim y} (f(x) - f(y))^2}{2 \sum_{x \in H} f(x)^2 \pi_H(x)} : f : H \to \mathbb{R} \text{ non-constant}, \sum_{x \in H} f(x) \pi_H(x) = 0 \right\}.$$

For given constants $T_1 \leq T_2$ and $\gamma > 0$, and distinguished vertex $\rho \in G$, the following assumption is then considered in [17]. There exist

- two positive functions $\lambda_s, V_s : s \in [T_1/2, T_2/2] \mapsto \mathbb{R}_+$ such that λ_s is decreasing and V_s is increasing,

- a family of universal constants $\{c_i, C_i\}_{1 \leq i \leq 4}$ and δ_1,

- for each $s \in [T_1/2, T_2/2]$, a distinguished connected set $B(s)$ containing ρ and a partition $\mathcal{P}_s$ of $B(s)$ into connected sets $\{H : H \in \mathcal{P}_s\}$,

such that the following holds:

(B1) for all $H \in \mathcal{P}_s$,
$$\mathrm{Gap}_H \geq \lambda_s,$$

(B2) for all $H \in \mathcal{P}_s$,
$$c_1 V_s \leq |H| \leq C_1 V_s,$$

(B3)
$$c_2 V_s^{-1/\gamma} \log^{-\delta_1} V_s \leq \lambda_s,$$

(B4)
$$\Delta_{\mathcal{P}_s} := \inf_{x \in B(s)} \frac{\deg_H(x)}{\deg_G(x)} \geq c_3 \log^{-1} V_s,$$

(B5)
$$P_\rho^G(Y_s \in B(s)^c) \leq C_3 V_s^{-1} \log V_s,$$

(B6)
$$1 + C_3 + 1/(c_1 c_3) \leq \psi_s(\rho) V_s \log^{-1} V_s \leq C_4,$$

where $\psi_s(\rho) = P_\rho^G(Y_{2s} = \rho)/\deg_G(\rho)$.

Applying these, the following result is proved in [17].

Theorem 2.7. [17, Lemma 3.1] Assume the conditions (B1)–(B6) are satisfied. Then for $\delta = \delta_1 + 1/\gamma$ and $t \in [T_1, T_2]$, we have that

$$\psi_t(\rho) \leq \psi_{T_1}(\rho) \wedge C_5(1 + C_6(t - T_1))^{-\gamma} |\log(1 + C_6(t - T_1))|^{\delta\gamma},$$

where C_5 and C_6 are universal constants.

Roughly speaking, the above theorem states that if, for all s large enough, we can find a connected subgraph $B(s) \subseteq G$ containing ρ and a partition of this, $\{H : H \in \mathcal{P}_s\}$, such that:

(B1') for all H,
$$|H| \asymp s^\gamma \log^{-\delta_1} s,$$

(B2') for all H,
$$\mathrm{Gap}_H \succeq 1/(s \log^{\delta_2} s),$$

(B3')
$$\Delta_{\mathcal{P}_s} \succeq 1/\log s,$$

(B4')
$$P_\rho^G(Y_s \in B(s)^c) \preceq s^{-\gamma} \log^{\delta_3} s,$$

then

$$\psi_s(\rho) \preceq s^{-\gamma} \log^{\delta_4} s.$$

We observe that, in the choice of exponents and the quantities of interest, the condition (B1') is similar to (A3)(a) and the condition (B2') is related to (A3)(c). The remaining conditions, (B3') and (B4'), are more technical, and ensure the behaviour of the random walk on the subgraph suitably captures that on the entire graph. As (a simplification of) the main result of [17], we state the following. We highlight that, although set-out here in terms of the continuous-time random walk, the conclusion is readily transferred to the discrete-time random walk by applying the argument used in the proof of [4, Theorem 5.14], for example.

Theorem 2.8. [17, Theorem 1] If $d \geq 1$ and $s \in (d, \min\{d+2, 2d\})$, then the continuous-time simple random walk on the long-range percolation cluster described in the introduction satisfies the conditions (B1)–(B6) with the exponent $\gamma = d/(s-d)$. As a consequence, with probability one, the upper heat kernel bound

$$\psi_T(\rho) \leq CT^{-d/(s-d)} \log^{\delta} T$$

holds for all large T, where C, δ are deterministic constants.

3 Application to long-range percolation

In this section, we prepare the ground for deriving the lower heat kernel bounds for the long-range percolation model LRP(d,s). Given a parameter $q \in [0, 1]$, we slightly generalise the setting presented in the introduction around (1.1) by supposing: for $x, y \in \mathbb{Z}^d$, the edge between them appears with probability

$$q_{x,y} = \begin{cases} q & \text{if } |x - y| = 1, \\ 1 - \exp(-|x-y|^s) & \text{if } |x - y| > 1, \end{cases}$$

independently of the state of other edges. In particular, when $q = 1$, we assume nearest-neighbour edges are present. Note that we only allow $q < 1$ in this section, where we discuss the verification of the conditions (A1)–(A3) for the more general model.

It was shown in [1] that LRP(d,s) admits an infinite cluster with probability 0 or 1, and has at most one infinite cluster almost-surely. We will suppose that LRP(d,s) percolates, i.e. it has an unique infinite cluster, which we denote by $G = (V, E)$; clearly this includes the $q = 1$ setting, and indeed any q above the nearest-neighbour percolation threshold for $\mathbb{Z}^d$. For each n, we let $G_n = (V_n, E_n)$ be the largest connected component of $G \cap [-n, n]^d$, and assume the following.

(V) There exists a universal constant $c > 0$, such that for all $n \geq 1$,

$$\mathbf{P}(|V_n| \geq cn^d) \geq 1 - \exp(-c(\log n)^2).$$

When $q = 1$, the condition (V) is trivial, since $V_n = [-n, n]^d$. Applying an estimate from [12], we can also check it in the case $s \in (d, 2d)$, see Lemma 6.6 below.

In the subsequent three subsections, we take the condition (V) as given, and proceed to check each of the assumptions (A1), (A2) and (A3). Thus we reduce the problem of obtaining heat kernel lower bounds to that of checking Benjamini-Schramm convergence and (V).

3.1 Checking (A1) for long-range percolation under (V)

The assumption (A1) is straightforward to handle for all cases simultaneously.

Lemma 3.1. For any $d \geq 1$ and $s > d$, the random graph G_n satisfies (A1).

Proof. By applying the Cauchy-Schwarz inequality and (V), we have for some $c > 0$,

$$
\begin{aligned}
\mathbf{E}\left(\left(\frac{|E_n|}{|V_n|}\right)^2\right) &\leq c^{-2}n^{-2d}\mathbf{E}[|E_n|^2] + \mathbf{E}[|E_n|^2\mathbb{I}(|V_n| < cn^d)] \\
&\leq c^{-2}n^{-2d}\mathbf{E}[|E_n|^2] + (\mathbf{E}[|E_n|^4])^{1/2}(\mathbf{P}[|V_n| < cn^d])^{1/2}. \quad (3.1)
\end{aligned}
$$

In addition, writing $V_n^* = [-n, n]^d$,

$$
\begin{aligned}
\mathbf{E}[|E_n|^2] &\leq \mathbf{E}\left(\left(\sum_{x,y\in V_n^*}\mathbb{I}(x \sim y)\right)^2\right) = \sum_{x,y,u,v\in V_n^*}\mathbf{P}[x \sim y, u \sim v] \\
&\preceq \sum_{x,y,u,v\in V_n^*}(1 + |x-y|^s)^{-1}(1 + |u-v|^s)^{-1} \preceq n^{2d}. \quad (3.2)
\end{aligned}
$$

Similarly,

$$
\mathbf{E}[|E_n|^4] \preceq n^{4d}, \quad (3.3)
$$

Combining (3.1), (3.2) and (3.3) with (V), we obtain that $\mathbf{E}[(|E_n|/|V_n|)^2] \preceq 1$, as desired.
$\square$

3.2 Checking (A2) for long-range percolation under (V)

For assumption (A2), we can also deal with all cases simultaneously.

Lemma 3.2. For any $d \geq 1$ and $s > d$, the random graph G_n satisfies (A2).

Proof. Since $|E_n| \geq |V_n| - 1$, it follows from (V) that

$$
\mathbf{P}[|E_n| \geq cn^d] \geq 1 - \exp(-c(\log n)^2),
$$

for some $c > 0$. Therefore,

$$
\begin{aligned}
\mathbf{E}\left[\pi_n^\star(\varepsilon)^2\right] &= \mathbf{E}\left[\sup_{|W|\leq\varepsilon|V_n|}\left(\frac{\sum_{x\in W}\deg_{G_n}(x)}{2|E_n|}\right)^2\right] \\
&\leq c^{-2}\varepsilon + \exp(-c(\log n)^2) + \mathbf{P}\left(\sup_{|W|\leq\varepsilon|V_n|}\sum_{x\in W}\deg_{G_n}(x) \geq \sqrt{\varepsilon}n^d\right). \quad (3.4)
\end{aligned}
$$

Since $V_n \subseteq V_n^* := [-n, n]^d$, applying the union bound and Markov's inequality yields that

$$
\begin{aligned}
\mathbf{P}&\left(\sup_{|W|\leq\varepsilon|V_n|}\sum_{x\in W}\deg_{G_n}(x) \geq \sqrt{\varepsilon}n^d\right) \\
&\leq \sum_{|W|\leq\varepsilon|V_n^*|}\mathbf{P}\left(\sum_{x\in W}\deg_{G_n}(x) \geq \sqrt{\varepsilon}n^d\right) \\
&\leq \sum_{|W|\leq\varepsilon|V_n^*|}e^{-\sqrt{\varepsilon}n^d}\mathbf{E}\left[\exp\left(\sum_{x\in W}\deg(x)\right)\right], \quad (3.5)
\end{aligned}
$$

where $\deg(x) := \sum_{y \in \mathbb{Z}^d} \mathbb{I}(x \sim y)$. Uniformly in n and W with $|W| \leq \varepsilon|V_n^*|$, we have

$$
\begin{aligned}
\mathbf{E}\left[\exp\left(\sum_{x \in W} \deg(x)\right)\right] &= \mathbf{E}\left[\exp\left(\sum_{x \in W}\sum_{y \in \mathbb{Z}^d} \mathbb{I}(x \sim y)\right)\right] \\
&= \prod_{y \in \mathbb{Z}^d}\prod_{x \in W} \mathbf{E}\left(\exp\left(\mathbb{I}(x \sim y)\right)\right) = \prod_{y \in \mathbb{Z}^d}\prod_{x \in W}\left(1 - q_{x,y} + q_{x,y}e\right) \\
&\leq \prod_{y \in \mathbb{Z}^d}\prod_{x \in W} \exp(2q_{x,y}) = \exp\left(2\sum_{x \in W}\sum_{y \in \mathbb{Z}^d} \mathbf{P}(x \sim y)\right) \\
&= \exp\left(2\sum_{x \in W} \mathbf{E}(\deg(x))\right) \leq \exp\left(C\varepsilon n^d\right). \qquad (3.6)
\end{aligned}
$$

Combining (3.5) and (3.6), and applying Stirling's formula, we obtain, for all n large,

$$
\begin{aligned}
\mathbf{P}\left(\sup_{|W| \leq \varepsilon|V_n|}\sum_{x \in W} \deg_G(x) \geq \sqrt{\varepsilon}n^d\right) &\leq \binom{|V_n^*|}{\varepsilon|V_n^*|} \exp\left(-\sqrt{\varepsilon}n^d + C\varepsilon n^d\right) \\
&\leq \exp\left(CI(\varepsilon)n^d - \sqrt{\varepsilon}n^d + C\varepsilon n^d\right), \qquad (3.7)
\end{aligned}
$$

where $I(\varepsilon) := -\varepsilon\log(\varepsilon) - (1-\varepsilon)\log(1-\varepsilon)$. Now, if $\varepsilon \in (0, \varepsilon_0)$ for suitably small $\varepsilon_0 > 0$, then $C(\varepsilon + I(\varepsilon)) - \sqrt{\varepsilon} < 0$, and we assume that this is the case. Putting (3.4) and (3.7) together, we thus deduce that, for all n large, $\mathbf{E}[\pi_n^\star(\varepsilon)^2] \leq C\varepsilon$, as desired. Since $\mathbf{E}[\pi_n^\star(\varepsilon)^2] \leq 1$, the result is obviously true for $\varepsilon \in [\varepsilon_0, \infty)$, and so the proof is complete. $\qquad \square$

3.3 Checking (A3) for long-range percolation under (V)

Recall that to verify the assumption (A3), we have to describe within the random graph G_n a sequence of disjoint capacitors $\{(A_i, \Omega_i)\}_{i=1}^k$. Specifically, for (A3)(c), we need to establish an upper bound for the sum of the associated capacities in terms of $|E_n|$. Hence we should find an upper bound for the capacities and a lower bound for $|E_n|$. The condition (V) guarantees a lower bound on the number of edges. For the upper bound on the capacities, we will use the following observation. By the definition of the capacity and the Dirichlet form on the finite graph G_n, we have: for any function $\varphi^{(i)}$ taking value 0 outside Ω_i and value 1 inside A_i,

$$
\mathrm{cap}_{\Omega_i}(A_i) \leq 2\mathcal{E}_{G_n}(\varphi^{(i)}) \leq 2\mathcal{E}(\varphi^{(i)}),
$$

where $\mathcal{E}$ is defined by setting

$$
\mathcal{E}(f) := \frac{1}{2}\sum_{\substack{x,y \in \mathbb{Z}^d \\ x \sim y}}(f(x) - f(y))^2
$$

for $f : \mathbb{Z}^d \to \mathbb{R}$. Since its definition does not involve summing over vertices in the largest percolation cluster in $[-n, n]^d$, computing with respect to $\mathcal{E}$ is more convenient than doing so for $\mathcal{E}_{G_n}$. In the next two results, Lemma 3.3 and Proposition 3.4, we obtain some useful estimates on the covariance between $\mathcal{E}(f)$ and $\mathcal{E}(g)$ for suitable functions f, g, and the expectation of $\mathcal{E}(f)$ for suitable f, respectively.

Lemma 3.3. Let $N \geq 1$ be an integer and $a, b \in \mathbb{Z}^d$ satisfy $|a - b| \geq 2N$. Suppose that f_a, f_b are two bounded functions with $\mathrm{supp}(f_a) \subseteq \Omega_a$ and $\mathrm{supp}(f_b) \subseteq \Omega_b$, where $\Omega_a = a + [-N, N]^d$ and $\Omega_b = b + [-N, N]^d$. Then there exists an universal constant C such that

$$
\mathrm{Cov}\left(\mathcal{E}(f_a), \mathcal{E}(f_b)\right) \leq \frac{CN^{4d}|f_a|^2|f_b|^2}{(|a-b| - 2N)^{2s} + 1},
$$

where we write $|f| := \sup_{x \in \mathbb{Z}^d} |f(x)|$ for $f : \mathbb{Z}^d \to \mathbb{R}$.

Proof. Since $\mathrm{supp}(f_a) \subseteq \Omega_a$,

$$
\begin{aligned}
\mathcal{E}(f_a) &= \sum_{\substack{\{x,y\}:\, x \in \Omega_a,\, y \in \mathbb{Z}^d, \\ x \sim y}} (f_a(x) - f_a(y))^2 \\
&= \sum_{\substack{\{x,y\}:\, x \in \Omega_a,\, y \in \Omega_b \\ x \sim y}} (f_a(x) - f_a(y))^2 + \sum_{\substack{\{x,y\}:\, x \in \Omega_a,\, y \in \Omega_b^c \\ x \sim y}} (f_a(x) - f_a(y))^2 \\
&=: \ \mathcal{E}_{a,1} + \mathcal{E}_{a,2}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathcal{E}(f_b) &= \sum_{\substack{\{x,y\}:\, x \in \Omega_b,\, y \in \Omega_a \\ x \sim y}} (f_b(x) - f_b(y))^2 + \sum_{\substack{\{x,y\}:\, x \in \Omega_b,\, y \in \Omega_a^c \\ x \sim y}} (f_b(x) - f_b(y))^2 \\
&=: \ \mathcal{E}_{b,1} + \mathcal{E}_{b,2}.
\end{aligned}
$$

By construction, $\mathcal{E}_{a,2}$ is independent of $\mathcal{E}_{b,1}$ and $\mathcal{E}_{b,2}$, and $\mathcal{E}_{b,2}$ is independent of $\mathcal{E}_{a,1}$ and $\mathcal{E}_{a,2}$. Therefore,

$$
\mathrm{Cov}\left(\mathcal{E}(f_a), \mathcal{E}(f_b)\right) = \mathrm{Cov}\left(\mathcal{E}_{a,1}, \mathcal{E}_{b,1}\right) \le \mathbf{E}\left(\mathcal{E}_{a,1}\mathcal{E}_{b,1}\right) \le |f_a|^2 |f_b|^2 \mathbf{E}\left(e(\Omega_a, \Omega_b)^2\right), \qquad (3.8)
$$

where

$$
e(\Omega_a, \Omega_b) := \sum_{\{x,y\}:\, x \in \Omega_b,\, y \in \Omega_a} \mathbb{I}(x \sim y).
$$

Writing $\lesssim$ for stochastic domination and $\mathrm{Bin}(m,p)$ for a binomial random variable with parameters $m \in \mathbb{N}$ and $p \in [0,1]$, we clearly have that

$$
e(\Omega_a, \Omega_b) \lesssim \mathrm{Bin}(|\Omega_a||\Omega_b|, p_{a,b}),
$$

where

$$
p_{a,b} := \max_{x \in \Omega_a, y \in \Omega_b} q_{x,y}.
$$

Since $|\Omega_a||\Omega_b| \asymp N^{2d}$ and $p_{a,b} \asymp ((|a-b| - 2N)^s + 1)^{-1}$,

$$
\mathbf{E}\left(e(\Omega_a, \Omega_b)^2\right) \le \mathbf{E}\left(\mathrm{Bin}(|\Omega_a||\Omega_b|, p_{a,b})^2\right) \preceq \frac{N^{4d}}{(|a-b| - 2N)^{2s} + 1}. \qquad (3.9)
$$

Combining (3.8) and (3.9), we arrive at the desired result. $\qquad \square$

For the next step, we will need to be careful about the distinction between the stable, Gaussian, and critical settings. In particular, the capacity estimate we require differs between the three cases. Towards deriving this, we introduce a linear cut-off function, and estimate the expected value of its energy. For $N \ge M$, set

$$
\varphi_{N,M}(x) = \begin{cases} 1, & \text{if } |x| \le N - M, \\ \frac{N - |x|}{M}, & \text{if } N - M < |x| \le N, \\ 0, & \text{if } |x| > N. \end{cases}
$$

Moreover, for a given sequence $(\beta_N)_{N \ge 1}$ in $[2, \infty)$, define $\varphi_N := \varphi_{N,M}$, where $M := N/\beta_N$. Note that the following result does not cover the part of the Gaussian regime corresponding to parameters $d = 1$ and $s > 2$, for which we deduce the lower heat kernel bound from the scaling limit of Theorem 1.1(b).

Proposition 3.4. (a) Fix $d \geq 1$ and $s \in (d, \min\{d + 2, 2d\})$. There exists a positive constant c such that, for any sequence $(\beta_N)_{N \geq 1}$ in $[2, \infty)$ such that $\beta_N = o(N)$, for all large N,

$$cN^{2d-s} \leq \mathbf{E}\left(\mathcal{E}(\varphi_N)\right) \leq N^{2d-s}\beta_N/c.$$

(b) Fix $d \geq 2$ and $s > d + 2$. The statement of part (a) holds with the following bound:

$$cN^{d-2}\beta_N \leq \mathbf{E}\left(\mathcal{E}(\varphi_N)\right) \leq N^{d-2}\beta_N/c.$$

(c) Fix $d \geq 2$ and $s = d + 2$. The statement of part (a) holds with the following bound:

$$cN^{d-2}\beta_N \log(N/\beta_N) \leq \mathbf{E}\left(\mathcal{E}(\varphi_N)\right) \leq N^{d-2}\beta_N \log(N/\beta_N)/c.$$

Proof. (a) By definition of φ_N,

$$
\begin{aligned}
2\mathcal{E}(\varphi_N) \;=\;& \sum_{\substack{|x| \leq N-M \\ |y| > N}} \mathbb{I}(x \sim y) + \sum_{\substack{|x| \leq N-M \\ N-M < |y| \leq N}} \mathbb{I}(x \sim y)\left(\frac{N - M - |y|}{M}\right)^2 \\
&+ \sum_{\substack{|x| > N \\ N-M < |y| \leq N}} \mathbb{I}(x \sim y)\left(\frac{N - |y|}{M}\right)^2 + \sum_{N-M < |x|, |y| \leq N} \mathbb{I}(x \sim y)\left(\frac{|x| - |y|}{M}\right)^2 \\
=:\;& S_1 + S_2 + S_3 + S_4,
\end{aligned}
$$

Using $q_{x,y} \asymp |x - y|^{-s}$ for $|x - y| \geq 1$, we have

$$
\begin{aligned}
\mathbf{E}(S_1) \;\asymp\;& \sum_{\substack{|x| \leq N-M \\ |y| > N}} |x - y|^{-s} = \sum_{|x| \leq N-M} \sum_{k > N-|x|} k^{-s} \#\{y : |y| > N, |x - y| = k\} \\
\asymp\;& \sum_{|x| \leq N-M} \sum_{k \in [N-|x|, 2N-|x|]} k^{d-1-s} \asymp \sum_{|x| \leq N-M} (N - |x|)^{d-s} \\
\asymp\;& \sum_{\ell=0}^{N-M} (N - \ell)^{d-s}\ell^{d-1} \asymp \int_1^{N-M} (N - \ell)^{d-s}\ell^{d-1}d\ell \\
\asymp\;& N^{2d-s} \int_{1/\beta_N}^{1-(1/N)} u^{d-s}(1 - u)^{d-1}du, \quad\quad\quad (3.10)
\end{aligned}
$$

where we used the fact that $\#\{y : |y - x| = k\} \asymp k^{d-1}$ in the second and third lines. Observe that

$$\int_{1/\beta_N}^{1-(1/N)} u^{d-s}(1 - u)^{d-1}du \asymp \begin{cases} 1, & \text{if } d + 1 - s > 0, \\ \log(\beta_N), & \text{if } d + 1 - s = 0, \\ \beta_N^{s-d-1}, & \text{if } d + 1 - s < 0. \end{cases} \quad (3.11)$$

Combining (3.10) and (3.11) yields that

$$\mathbf{E}(S_1) \asymp \begin{cases} N^{2d-s}, & \text{if } d + 1 - s > 0, \\ N^{2d-s}\log(\beta_N), & \text{if } d + 1 - s = 0, \\ N^{2d-s}\beta_N^{s-d-1}, & \text{if } d + 1 - s < 0. \end{cases} \quad (3.12)$$

We next estimate $\mathbf{E}(S_4)$. Using the inequality $||x| - |y|| \leq |x - y|$, we have, for all $N - M \leq |x|, |y| \leq N$,

$$\left(\frac{|x| - |y|}{M}\right)^2 \leq \frac{|x - y|^2}{M^2} \wedge 1. \quad (3.13)$$

Therefore,

$$
\begin{aligned}
\mathbf{E}(S_4) &= \frac{1}{2} \sum_{N-M<|x|,|y|\le N} q_{x,y} \left(\frac{|x|-|y|}{M} \right)^2 \\
&\preceq \sum_{N-M<|x|\le N} \left[\frac{1}{M^2} \sum_{y:|y-x|\le M} |x-y|^{2-s} + \sum_{y:|y-x|>M} |x-y|^{-s} \right] \\
&\preceq \sum_{N-M<|x|\le N} \left[\frac{1}{M^2} \sum_{k=1}^{M} k^{2-s} k^{d-1} + \sum_{k>M} k^{-s} k^{d-1} \right] \\
&\preceq \sum_{N-M<|x|\le N} M^{d-s} \preceq N^{d-1} M^{d+1-s} = N^{2d-s} \beta_N^{s-d-1}. \quad (3.14)
\end{aligned}
$$

Note that for the second line we used the fact that $\#\{y : |y-x| = k\} \asymp k^{d-1}$ and for the last line we used $\#\{x : N - M < |x| \le N\} \asymp N^{d-1} M$. To estimate $\mathbf{E}(S_3)$, we observe that for $|x| > N$ and $N - M < |y| \le N$,

$$
\left(\frac{N-|y|}{M} \right)^2 \le \frac{|x-y|^2}{M^2} \wedge 1. \quad (3.15)
$$

Hence, using the same argument for as (3.14), with (3.15) playing the role of (3.13), we can prove that

$$
\mathbf{E}(S_3) \preceq N^{2d-s} \beta_N^{s-d-1}. \quad (3.16)
$$

The term $\mathbf{E}(S_2)$ can be also bounded in the same way as $\mathbf{E}(S_3)$ or $\mathbf{E}(S_4)$. Indeed, for $|x| \le N - M$ and $N - M < |y| \le N$,

$$
\left(\frac{N-M-|y|}{M} \right)^2 \le \frac{|x-y|^2}{M^2} \wedge 1.
$$

Hence, using the argument that was used to obtain (3.14) and (3.16), we find

$$
\mathbf{E}(S_2) \preceq N^{2d-s} \beta_N^{s-d-1}. \quad (3.17)
$$

Using (3.12), (3.14), (3.16) and (3.17), we have $\max\{\mathbf{E}(S_2), \mathbf{E}(S_3), \mathbf{E}(S_4)\} \preceq \mathbf{E}(S_1)$, and so

$$
\mathbf{E}\left(\mathcal{E}(\varphi_N) \right) \asymp \mathbf{E}(S_1),
$$

which together with (3.12) implies that

$$
N^{2d-s} \preceq \mathbf{E}\left(\mathcal{E}(\varphi_N) \right) \preceq N^{2d-s} \beta_N.
$$

(b) Repeating the above argument yields

$$
\mathbf{E}(S_1) \asymp N^{2d-s} \beta_N^{s-d-1}.
$$

However, this is no longer the dominant term. Specifically, for $s > d + 2$, we have that

$$
\begin{aligned}
\mathbf{E}(S_4) &\asymp \sum_{N-M<|x|,|y|\le N} |x-y|^{-s} \left(\frac{|x|-|y|}{M} \right)^2 \\
&\preceq \sum_{N-M<|x|\le N} \sum_{k\ge 1} k^{-s} \left(\frac{k^2}{M^2} \wedge 1 \right) k^{d-1} \\
&\preceq \sum_{N-M<|x|\le N} \left(\sum_{k=1}^{M} M^{-2} k^{d+1-s} + \sum_{k>M} k^{d-1-s} \right) \\
&\preceq N^{d-1} M^{-1} = N^{d-2} \beta_N.
\end{aligned}
$$

Similar considerations yield a lower bound of the same form, i.e. $\mathbf{E}(S_4) \succeq N^{d-2}\beta_N$. Moreover, again arguing similarly, we find that

$$\mathbf{E}(S_2) + \mathbf{E}(S_3) \preceq N^{d-2}\beta_N.$$

Therefore,

$$\mathbf{E}\left(\mathcal{E}(\varphi_N)\right) \asymp \mathbf{E}(S_4) \asymp N^{d-2}\beta_N.$$

(c) The final case $d \geq 2$ and $s = d + 2$. The proof is similar to that of (b). The additional log term appears since for $s = d + 2$ it holds that $\sum_{k=1}^{M} k^{d+1-s} \asymp \log(M)$. □

We are now ready to check (A3), and we start with the stable case.

Lemma 3.5. (a) Fix $d \geq 1$ and $s \in (d, \min\{d + 2, 2d\})$. For $\kappa > 0$, LRP(d, s) satisfies (A3) with $\delta_1 = \frac{2d\kappa}{s-d}$, $\delta_2 = \delta_3 = \kappa$, δ_0 arbitrarily large, and $\lambda(t) = \max\{1, \log(t)\}$.
(b) Fix $d \geq 1$ and $s \in (d, \min\{d+2, 2d\})$. LRP$(d, s)$ satisfies (A3) with $\delta_1 = \frac{2d}{s-d}$, $\delta_2 = \delta_3 = 1$, δ_0 arbitrarily large, and $\lambda(t) = \lambda_0$. Moreover, the constants α and λ_0 can be chosen so that, taking C_α as the constant of Theorem 2.3, it holds that $1 - C_\alpha \lambda_0^{-1/2} > 0$.

Proof. (a) We will verify the condition with $n_0 := n_0(t) = t^\Delta$ suitably large Δ. To do so, first define

$$N := t^{\frac{1}{s-d}}(\log t)^{\frac{2\kappa}{s-d}}, \tag{3.18}$$

where $\kappa > 0$ is some constant. For $n \geq n_0$, we then cover $[-n + N, n - N]^d$ by disjoint boxes of side-length $2N$, the intersections of which with V_n we will denote by $(\Omega_i)_{i=1}^k$. Writing x_i for the center of the box containing Ω_i, so $\Omega_i = B_{x_i}(N) \cap V_n$ (where $B_x(r)$ is the ℓ_∞-ball of radius r centred at x), we also introduce $A_i := B_{x_i}(N - M) \cap V_n$, where $M := N/(\log N)^\kappa$. Now, by construction, the sets $(\Omega_i)_{i=1}^k$ are disjoint, and moreover satisfy

$$\max_{i=1,\ldots,k} |\Omega_i| \preceq N^d = t^{\frac{d}{s-d}}(\log t)^{\frac{2d\kappa}{s-d}} = t^\gamma (\log t)^{\delta_1},$$

where $\delta_1 := 2d\kappa/(s - d)$. Thus (A3)(a) is satisfied for capacitors (A_i, Ω_i), $i = 1, \ldots, k$.

For verifying (A3)(b), define

$$W = \{x : n - N \leq |x| \leq n\} \cup (\cup_{i=1}^k W_i),$$

where $W_i = \{x : N - M \leq |x - x_i| \leq N\}$, and observe that

$$\sum_{i=1}^{k} \pi_n(A_i) \geq 1 - \pi_n(W \cap V_n).$$

Now, for any C, ε_n and suitably small constant c, we have

$$
\begin{aligned}
\mathbf{P}\left(\pi_n(W \cap V_n) \geq C\varepsilon_n\right) &\leq \mathbf{P}\left(\sum_{x \in W} \deg_{G_n}(x) \geq Cc\varepsilon_n n^d\right) + \mathbf{P}(|E_n| \leq cn^d) \\
&\leq \mathbf{P}\left(\sum_{x \in W} \deg(x) \geq Cc\varepsilon_n n^d\right) + \mathbf{P}(|V_n| \leq cn^d + 1) \\
&\leq \exp(-Cc\varepsilon_n n^d)\mathbf{E}\left(\exp\left(\sum_{x \in W} \deg(x)\right)\right) + \exp(-c(\log n)^2) \\
&\leq \exp\left(-Cc\varepsilon_n n^d + 2|W|\mathbf{E}(\deg(\rho))\right) + \exp(-c(\log n)^2),
\end{aligned}
$$

where for the second and third inequalities we use $|V_n| \leq |E_n| + 1$ and (V), and for the last one we follow the argument leading to (3.6). Furthermore, noting that the number of boxes satisfies $k \asymp \left(\frac{n}{N}\right)^d$, we have

$$|W| = |\{x : \, n - N \leq |x| \leq n\}| + \sum_{i=1}^{k} |W_i| \asymp n^{d-1}N + kN^{d-1}M \asymp n^d (\log N)^{-\kappa},$$

where we have used that $N = o(n^{1/2})$ (for suitably large Δ). Hence, by taking $\varepsilon_n = (\log N)^{-\kappa}$ and C sufficiently large, we find that

$$\mathbf{P}\left(\pi_n(W \cap V_n) \geq C(\log N)^{-\kappa}\right) \leq \exp(-n^d (\log N)^{-\kappa}) + \exp(-c(\log n)^2) \leq (\log t)^{-\delta}$$

for any $\delta > 0$ (once t is large). Reformulating this bound, we find that

$$\mathbf{P}\left(\sum_{i=1}^{k} \pi_n(A_i) \geq 1 - \alpha(\log t)^{-\kappa}\right) \geq 1 - (\log t)^{-\delta},$$

which confirms (A3)(b) with $\delta_2 = \kappa$.

Finally, to check (A3)(c), we start by defining a collection of functions $(\varphi^{(i)})_{i=1}^{k}$ by setting

$$\varphi^{(i)}(x) = \varphi_N(x - x_i).$$

Uniformly in $1 \leq i \leq k$,

$$\mathbf{E}\left(\mathcal{E}(\varphi^{(i)})^2\right) \leq \mathbf{E}\left(\left(\sum_{x \in \Omega_i} \deg(x)\right)^2\right) \leq |\Omega_i| \mathbf{E}\left(\sum_{x \in \Omega_i} (\deg(x))^2\right) \preceq |\Omega_i|^2 \preceq N^{2d}.$$

By Lemma 3.3, uniformly in all pairs $i \neq j$,

$$\mathrm{Cov}\left(\mathcal{E}(\varphi^{(i)}), \mathcal{E}(\varphi^{(j)})\right) \preceq \frac{N^{4d}}{(|x_i - x_j| - 2N)^{2s} + 1}.$$

Therefore,

$$\mathrm{Var}\left(\sum_{i=1}^{k} \mathcal{E}(\varphi^{(i)})\right) \preceq kN^{2d} + \sum_{1 \leq i \neq j \leq k} \frac{N^{4d}}{(|x_i - x_j| - 2N)^{2s} + 1}.$$

In addition,

$$\sum_{1 \leq i \neq j \leq k} \frac{1}{(|x_i - x_j| - 2N)^{2s} + 1} = \sum_{1 \leq i \neq j \leq k} \frac{(2N)^{-2s}}{\left(\left|\frac{x_i - x_j}{2N}\right| - 1\right)^{2s} + (2N)^{-2s}}$$

$$\asymp \sum_{x,y \in [-\frac{n}{2N}, \frac{n}{2N}]^d \cap \mathbb{Z}^d : \, x \neq y} \frac{(2N)^{-2s}}{(|x - y| - 1)^{2s} + (2N)^{-2s}} \asymp \left(\frac{n}{N}\right)^d \asymp k,$$

and so

$$\mathrm{Var}\left(\sum_{i=1}^{k} \mathcal{E}(\varphi^{(i)})\right) \preceq kN^{4d}. \tag{3.19}$$

By Proposition 3.4 (a), we also have that

$$kN^{2d-s} \preceq \mathbf{E}\left(\sum_{i=1}^{k} \mathcal{E}(\varphi^{(i)})\right) \preceq kN^{2d-s}\beta_N \asymp n^d N^{d-s}(\log N)^{\kappa} \asymp n^d t^{-1}(\log t)^{-\kappa}, \tag{3.20}$$

where $\beta_N = N/M = (\log N)^\kappa$ and $N^{d-s}(\log N)^\kappa \asymp t^{-1}(\log t)^{-\kappa}$ by (3.18). Using the variance bound (3.19), the lower bound for the expectation of (3.20), and Chebyshev's inequality, we thus obtain

$$\mathbf{P}\left(\sum_{i=1}^k \mathcal{E}(\varphi^{(i)}) \geq 2\sum_{i=1}^k \mathbf{E}[\mathcal{E}(\varphi^{(i)})]\right) \preceq \frac{kN^{4d}}{(kN^{2d-s})^2} = \frac{N^{2s}}{k} \preceq k^{-1/2},$$

provided that $N^{4s} \preceq k$, or equivalently $N^{1+4s/d} \preceq n$ (which holds for Δ large). Since it is the case that $\mathrm{supp}(\varphi^{(i)}) = \Omega_i$, $\varphi^{(i)}|_{A_i} = 1$ and $0 \leq \varphi^{(i)} \leq 1$, we have $\mathrm{cap}_{\Omega_i}(A_i) \leq 2\mathcal{E}_{G_n}(\varphi^{(i)}) \leq 2\mathcal{E}(\varphi^{(i)})$. Hence,

$$\sum_{i=1}^k \mathrm{cap}_{\Omega_i}(A_i) \leq 2\sum_{i=1}^k \mathcal{E}(\varphi^{(i)}).$$

Furthermore, by (V),

$$\mathbf{P}(|E_n| \leq cn^d) \leq \exp(-c(\log n)^2),$$

for c a positive constant. Combining the last four displayed equations, we obtain that, for α suitably large,

$$\mathbf{P}\left(\sum_{i=1}^k \mathrm{cap}_{\Omega_i}(A_i) > 2\alpha|E_n|t^{-1}(\log t)^{-\kappa}\right)$$

$$\leq \mathbf{P}\left(\sum_{i=1}^k \mathcal{E}(\varphi^{(i)}) > \alpha|E_n|t^{-1}(\log t)^{-\kappa}\right)$$

$$\leq \mathbf{P}\left(\sum_{i=1}^k \mathcal{E}(\varphi^{(i)}) > \alpha cn^d t^{-1}(\log t)^{-\kappa}\right) + \mathbf{P}(|E_n| \leq cn^d)$$

$$\leq \mathbf{P}\left(\sum_{i=1}^k \mathcal{E}(\varphi^{(i)}) \geq 2\sum_{i=1}^k \mathbf{E}[\mathcal{E}(\varphi^{(i)})]\right) + \exp(-c(\log n)^2)$$

$$\preceq 2k^{-1/2}. \tag{3.21}$$

Hence we obtain (A3)(c) with $\delta_3 = \kappa$.

(b) Replacing $(\log t)^\kappa$ and $(\log N)^\kappa$ by λ_0 in the above argument, it follows that (A3) holds with the given constants. Furthermore, since the constant α that comes out of the argument does not depend on λ_0, by increasing the value of the latter quantity if needed, we can ensure that $1 - C_\alpha \lambda_0^{-1/2} > 0$. $\qquad\square$

As for the Gaussian case, we check the following version of (A3). We underline that, although we include $d = 2$ in the following result, we will handle the $d = 1$ and $d = 2$ cases separately in the proofs of our heat kernel estimates, using the quenched invariance principle that is known to hold throughout the Gaussian regime for those.

Lemma 3.6. (a) Fix $d \geq 2$ and $s > d + 2$. For $\kappa \geq 1$, LRP(d, s) satisfies (A3) with $\delta_1/d = \delta_2 = \delta_3 = \kappa$, δ_0 arbitrarily large, and $\lambda(t) = \max\{1, \log(t)\}$.

(b) Fix $d \geq 2$ and $s > d + 2$. LRP(d, s) satisfies (A3) with $\delta_1/d = \delta_2 = \delta_3 = 1$, δ_0 arbitrarily large, and $\lambda(t) = \lambda_0$. Moreover, the constants α and λ_0 can be chosen so that, taking C_α as the constant of Theorem 2.3, it holds that $1 - C_\alpha \lambda_0^{-1} > 0$.

Proof. (a) The argument is again similar to Lemma 3.5(a), but now we take

$$N := t^{\frac{1}{2}}(\log t)^\kappa.$$

In this case, we then get that

$$\max_{i=1,\dots,k} |\Omega_i| \preceq N^d = t^{\frac{d}{2}}(\log t)^{\kappa d},$$

and so (A3)(a) holds with $\delta_1 = \kappa d$. Proceeding as in the previous proof with $\beta_N := (\log N)^\kappa$, we further have

$$\mathbf{P}\left(\pi_n(W \cap V_n) \geq C(\log N)^{-\kappa}\right) \leq (\log t)^{-\delta}$$

for any $\delta > 0$ (once t is large), which verifies (A3)(b) with $\delta_2 = \kappa$. By using similar arguments as for (3.21), noting that the variance bound (3.19) holds for all d, s, and now with the help of Proposition 3.4(b), we can also prove that

$$\mathbf{P}\left(\sum_{i=1}^k \mathrm{cap}_{\Omega_i}(A_i) > 2\alpha|E_n|t^{-1}(\log t)^{-\kappa}\right) \leq 2k^{-1/2},$$

for $n \geq n_0(t) = t^\Delta$, for α, Δ chosen suitably large.

(b) Making appropriate adaptations to the proof of (a), the proof is similar to that of Lemma 3.5(b). □

Finally, in the critical case, we have the following. Note that in this case we do not provide a separate bound with $\lambda(t)$ constant as we do in the previous two lemmas, since the additional log term in Proposition 3.4(c) means that we are unable to avoid incorporating a log in the estimates somewhere.

Lemma 3.7. Fix $d \geq 2$ and $s = d + 2$. For $\kappa > 0$, LRP(d,s) satisfies (A3) with $\delta_1/d - 1/2 = \delta_2 = \delta_3 = \kappa$, δ_0 arbitrarily large, and $\lambda(t) = \max\{1, \log(t)\}$.

Proof. The argument is again similar to Lemma 3.5, but now we take

$$N := t^{\frac{1}{2}}(\log t)^{\kappa + \frac{1}{2}},$$

and have to take into account the additional $\log$ term that appears in the critical case, i.e. apply Proposition 3.4(c), rather than Proposition 3.4(b). □

4 Proof of Theorems 1.3 and 1.5

4.1 Proof of lower bounds

Given the preparations of the previous section concerning (A1)–(A3), the main outstanding issue when it comes to the application of Theorem 2.3 is to check the assumption of Benjamini-Schramm convergence for the LRP(d,s) model. Defining $G_n = (V_n, E_n)$, $n \geq 1$, as at the start of Section 3, we make precise the desired condition as follows.

(BS) Let ρ be the origin of $\mathbb{Z}^d$ and ρ_n be a uniformly chosen vertex in G_n. Then the random graphs (G_n, ρ_n) Benjamini-Schramm converge to (G, ρ), conditioned that $\rho \in G$.

In Section 6.4, we will explain how to verify (BS) in the non-nearest-neighbour setting when s lies in the restricted range $(d, 2d)$. Unfortunately, for the full range of parameters, $s > d$, we are only able to verify (BS) in the nearest-neighbour case, i.e. taking $q = 1$ in the more general model of Section 3. (See Remark 4.2 for further discussion of how our heat kernel estimates apply in general when both (V) and (BS) hold.)

Lemma 4.1. The LRP(d,s) model with $q = 1$ satisfies (BS).

Proof. We write B_G for balls in G with respect to the graph distance, B_{G_n} for balls in G_n with respect to the graph distance, and B_∞ for ℓ_∞-balls in $\mathbb{Z}^d$. We need to prove that for any rooted, finite graph H and finite r,

$$\lim_{n\to\infty} \mathbf{P}(B_{G_n}(\rho_n, r) = H) = \mathbf{P}(B_G(\rho, r) = H). \tag{4.1}$$

We have that

$$\begin{aligned}
\mathbf{P}\left(B_{G_n}(\rho_n, r) = H\right) &\leq \mathbf{P}\left(B_{G_n}(\rho_n, r) = H,\ B_G(\rho_n, r) \subseteq B_\infty(\rho_n, N) \subseteq B_\infty(\rho, n)\right) \\
&\quad + \mathbf{P}\left(B_G(\rho_n, r) \nsubseteq B_\infty(\rho_n, N)\right) + \mathbf{P}\left(B_\infty(\rho_n, N) \nsubseteq B_\infty(\rho, n)\right).
\end{aligned}$$

Now, since ρ_n is uniformly chosen on V_n,

$$\mathbf{P}\left(B_\infty(\rho_n, N) \nsubseteq B_\infty(\rho, n)\right) \leq \frac{cn^{d-1}N}{n^d} = \frac{cN}{n} \to 0,$$

as $n \to \infty$ (for each fixed N). Moreover, by the translation invariance of the model,

$$\mathbf{P}\left(B_G(\rho_n, r) \nsubseteq B_\infty(\rho_n, N)\right) = \mathbf{P}\left(B_G(\rho, r) \nsubseteq B_\infty(\rho, N)\right).$$

Since G is $\mathbf{P}$-a.s. locally-finite when $s > d$, $B_G(\rho, r)$ is $\mathbf{P}$-a.s. a finite set, and so the above probability converges to 0 as $N \to \infty$. Finally, on the event $B_G(\rho_n, r) \subseteq B_\infty(\rho_n, N) \subseteq B_\infty(\rho, n)$, it holds that $B_{G_n}(\rho_n, r) = B_G(\rho_n, r)$, and so

$$\begin{aligned}
&\mathbf{P}\left(B_{G_n}(\rho_n, r) = H,\ B_G(\rho_n, r) \subseteq B_\infty(\rho_n, N) \subseteq B_\infty(\rho, n)\right) \\
&\quad \leq \mathbf{P}\left(B_G(\rho_n, r) = H\right) = \mathbf{P}\left(B_G(\rho, r) = H\right).
\end{aligned}$$

In particular, it follows from what we have so far established that

$$\limsup_{n\to\infty} \mathbf{P}\left(B_{G_n}(\rho_n, r) = H\right) \leq \mathbf{P}\left(B_G(\rho, r) = H\right).$$

By noting that

$$\begin{aligned}
&\mathbf{P}\left(B_{G_n}(\rho_n, r) = H\right) \\
&\quad \geq \mathbf{P}\left(B_{G_n}(\rho_n, r) = H,\ B_G(\rho_n, r) \subseteq B_\infty(\rho_n, N) \subseteq B_\infty(\rho, n)\right) \\
&\quad \geq \mathbf{P}\left(B_G(\rho, r) = H\right) - \mathbf{P}\left(B_G(\rho, r) \nsubseteq B_\infty(\rho, N)\right) - \mathbf{P}\left(B_\infty(\rho_n, N) \nsubseteq B_\infty(\rho, n)\right),
\end{aligned}$$

and applying the results of the preceding discussion, one may similarly conclude that

$$\liminf_{n\to\infty} \mathbf{P}\left(B_{G_n}(\rho_n, r) = H\right) \geq \mathbf{P}\left(B_G(\rho, r) = H\right),$$

which is enough to complete the proof of (4.1). □

Proof of lower bounds of Theorem 1.3. Combining Theorem 1.1(a) and Lemma 6.2 gives the lower bound of (1.2) in the case $s \in (d, d+1)$ with $\delta_1 = 0$. Similarly, combining Theorem 1.1(b) and Lemma 6.2 gives the lower bounds of (1.4) in the case $s > 2d$ with $\delta_3 = 0$ and (1.3). In the remaining cases, we have from Lemmas 3.1, 3.2, 3.5(a), 3.6(a), 3.7 that (A1), (A2) and (A3) hold for the $\delta_0, \delta_1, \delta_2, \delta_3$ and $\lambda(t)$ given by the latter three results. Taking $\kappa > 2$ in Lemmas 3.5, 3.6, 3.7 enables us to apply Corollary 2.6(a) (and Lemma 4.1) to derive quenched lower heat kernel bounds in each case, with the δ_is of (1.2) and (1.4) as in Remark 1.4. □

Proof of lower bounds of Theorem 1.5. Again we appeal to Lemmas 3.1, 3.2 to confirm that (A1) and (A2) hold in all three cases. In conjunction with Lemmas 3.5(b) and 3.6(b), we have that (A3) holds in the sense required by Corollary 2.6(b) in the stable and Gaussian cases. Putting this together with Lemma 4.1, we thus obtain the lower bounds of (1.5) and (1.7). The lower bound of (1.6) readily follows by applying Lemmas 3.1, 3.2 and 3.7, with $\kappa > 0$ chosen arbitrarily small, in conjunction with Theorem 2.3 (and Lemma 4.1). □

Remark 4.2. The lower bounds of Theorems 1.3 and 1.5 hold for the more general LRP(d,s) model of Section 3 (i.e. with $q \in [0,1]$) whenever the conditions (V) and (BS) are satisfied. Indeed, we proved in Section 3 that under (V) the assumptions (A1)–(A3) are valid and hence, by arguments in the proof of Theorem 2.3,

$$\mathbf{P}\left(p_{2t}^{G_n}(\rho_n, \rho_n) < t^{-\gamma}\lambda(t)^{-\delta}\right) \preceq \lambda(t)^{-\delta'}$$

with $\gamma = \frac{d}{s-d}$ and $\lambda(t) = \log(t)$ and suitable $\delta, \delta' \geq 0$. Furthermore, the condition (BS) assures the Benjamini-Schramm convergence of the random graphs G_n, which in particular implies that

$$\left|\mathbf{P}\left(p_{2t}^{G_n}(\rho_n, \rho_n) < t^{-\gamma}\lambda(t)^{-\delta_4}\right) - \mathbf{P}\left(p_{2t}^{G}(\rho, \rho) < 2t^{-\gamma}\lambda(t)^{-\delta_4} \mid \rho \in G\right)\right| \leq \lambda(t)^{-\delta_0},$$

for all n sufficiently large. The lower bounds of the heat kernel follow from the above two estimates; see Section 6.4, and Corollary 6.9 in particular, for our application of this argument to the non-nearest-neighbour long-range percolation model of Section 3 with $s \in (d, 2d)$.

4.2 Proof of upper bounds

Proof of upper bounds of Theorem 1.3. As noted in Remark 1.4, the upper bound of (1.2) follows from [17, Theorem 1] (which did not require the assumption of nearest-neighbour edges being present), using the argument in the proof of [4, Theorem 5.14] to transfer to discrete time. As for (1.3), this is an immediate consequence of the general bound of [5, Theorem 2.1], which implies that there exists a universal constant C such that $p_{2t}^{G}(x, x) \leq Ct^{-1/2}$ for the simple random walk on any infinite connected graph G.

It remains to establish the upper bound of (1.4). To this end, we will first consider the continuous-time Markov process $(Z_t)_{t \geq 0}$, which has jump chain given by X, but the jump rate at site x is equal to $\deg_G(x)$ (i.e. the holding time is exponential with this parameter). The idea of the following proof comes from the unpublished version of [14]. Note that the measure m on V placing mass 1 on each vertex is invariant for $(Z_t)_{t \geq 0}$. We let

$$A_t := \int_0^t \deg_G(Z_s)ds,$$

and define $(Y_t)_{t \geq 0}$ by setting $Y_t = Z_{A_t^{-1}}$, where A_t^{-1} is the right continuous inverse of the non-decreasing additive functional $(A_t)_{t \geq 0}$; the process $(Y_t)_{t \geq 0}$ has the same jump chain as Z (and X), but mean one exponential holding times. We claim that there exists a deterministic constant c_1 such that, for any realisation of G,

$$P_x^G(Z_t = x) \leq c_1 t^{-d/2}, \qquad \forall x \in \mathbb{Z}^d. \tag{4.2}$$

Indeed, for the (constant speed) simple symmetric random walk on $\mathbb{Z}^d$, the Nash inequality

$$\|f\|_{L^2(m)}^{2+4/d} \leq c_2 \, \mathcal{E}^{\text{SRW}}(f, f) \, \|f\|_{L^1(m)}^{4/d}$$

holds for some constant $c_2 = c_2(d)$, see for instance [4, Lemma 3.13]. Here, we have written $\|f\|_{L^p(m)}$ for the L^p-norm with respect to m, and

$$\mathcal{E}^{\text{SRW}}(f, f) := \frac{1}{2} \sum_{x,y \in \mathbb{Z}^d: \, |x-y|=1} (f(x) - f(y))^2.$$

Since nearest-neighbour edges are present in G, it holds that $\mathcal{E}^{\text{SRW}}(f, f) \leq \mathcal{E}(f, f)$, where $\mathcal{E}$ was defined at (2.2), and so the same inequality holds with $\mathcal{E}^{\text{SRW}}$ replaced by $\mathcal{E}$. By [15, Theorem (2.1)], we thus obtain (4.2).

We next estimate $P_x^G(Y_{2t} = x)$ by controlling the time change. Using the monotonicity of $s \mapsto P_x^G(Y_{2s} = x)$, we get

$$
\begin{aligned}
P_x^G(Y_{2t} = x) &\leq \frac{1}{t} E_x^G \left(\int_t^{2t} \mathbb{I}(Y_s = x) ds \right) \\
&= \frac{1}{t} E_x^G \left(\int_{A_t^{-1}}^{A_{2t}^{-1}} \mathbb{I}(Z_u = x) A'_u du \right) \\
&\leq \frac{\deg_G(x)}{t} \int_0^{2t} P_x^G \left(Z_u = x, t \leq A_u \leq 2t \right) du,
\end{aligned}
\tag{4.3}
$$

where we first changed variables using $s := A_u$, and then used that the derivative A'_u satisfies $A'_u = \deg_G(Z_u) = \deg_G(x)$ on the event $Z_u = x$. In the last inequality, we also used the fact that $A_{2t} \geq 2t$, which holds because $\deg_G(x) \geq 2d \geq 1$ (since all nearest-neighbour edges are present).

Now, let $D_t := \sup_{0 \leq s \leq t} |Y_s|$. By [17, Lemma 4.1], there exist $c_3, c_4, c_5 > 0$ such that, for any $T, \lambda > 0$ and any $p > (s - d)^{-1}$, $r < s - d$,

$$
\mathbf{P} \left(\{ P_\rho^G \left(D_t \geq c_3 t^{p+1} \right) > c_4 t^{-\lambda} \} \right) \leq c_5 t^{\lambda + 1 - pr}.
\tag{4.4}
$$

Hence, taking p large enough so that $\sum_{t=1}^{\infty} t^{d/2 + 1 - pr} < \infty$, by applying the Borel-Cantelli lemma, one can deduce that, $\mathbf{P}$-a.s., for all large $t \in \mathbb{N}$,

$$
P_\rho^G(E_t^c) \leq c_4 t^{-d/2},
\tag{4.5}
$$

where $E_t := \{ D_t < c_3 t^{p+1} \}$. Since

$$
D_{2t} = \sup_{0 \leq s \leq 2t} |Y_s| = \sup_{0 \leq s \leq 2t} |Z_{A_s^{-1}}| = \sup_{0 \leq s \leq A_{2t}^{-1}} |Z_s|,
$$

it further holds that $E_{2t} \cap \{ A_u \leq 2t \} \subseteq F_{u,t}$, where $F_{u,t} := \{ \sup_{0 \leq s \leq u} |Z_s| \leq c_3 (2t)^{p+1} \}$. In particular, applying (4.5), we deduce from (4.3) that, $\mathbf{P}$-a.s., for all large $t \in \mathbb{N}$,

$$
\frac{P_x^G(Y_{2t} = x)}{\deg_G(x)} \leq 2c_4 (2t)^{-d/2} + \frac{1}{t} \int_0^{2t} P_x^G \left(Z_u = x, t \leq A_u \leq 2t, F_{u,t} \right) du.
\tag{4.6}
$$

Using the Markov property, we moreover have

$$
\begin{aligned}
&P_x^G(Z_u = x, t \leq A_u \leq 2t, F_{u,t}) \\
&\leq \sum_{y \in \mathbb{Z}^d} P_x^G(Z_{u/2} = y, Z_u = x, t/2 \leq A_{u/2} \leq 2t, F_{u,t}) \\
&\qquad + \sum_{y \in \mathbb{Z}^d} P_x^G(Z_{u/2} = y, Z_u = x, t/2 \leq A_u - A_{u/2} \leq 2t, F_{u,t}) \\
&\leq \sum_{y \in \mathbb{Z}^d} E_x^G \left(\mathbb{I}(Z_{u/2} = y, t/2 \leq A_{u/2} \leq 2t, F_{u/2,t}) P_y^G \left(Z_{u/2} = x \right) \right) \\
&\qquad + \sum_{y \in \mathbb{Z}^d} E_x^G \left(\mathbb{I}(Z_{u/2} = y) P_y^G \left(Z_{u/2} = x, t/2 \leq A_{u/2} \leq 2t, F_{u/2,t} \right) \right).
\end{aligned}
$$

Noting that $P_y^G(Z_{u/2} = x) \leq c_6 u^{-d/2}$, which is due to (4.2) and the Cauchy-Schwarz inequality, and

$$
P_y^G \left(Z_{u/2} = x, t/2 \leq A_{u/2} \leq 2t, F_{u/2,t} \right) = P_x^G \left(Z_{u/2} = y, t/2 \leq A_{u/2} \leq 2t, F_{u/2,t} \right),
$$

which is due to the symmetry of $(Z_t)_{t\geq 0}$, we obtain

$$P_x^G(Z_u = x, t \leq A_u \leq 2t, F_{u,t})$$
$$\leq\ 2 \sum_{y\in\mathbb{Z}^d} P_x^G\left(Z_{u/2} = y\right) P_x^G\left(Z_{u/2} = y, t/2 \leq A_{u/2} \leq 2t, F_{u/2,t}\right)$$
$$\leq\ c_7 u^{-d/2} P_x^G\left(t/2 \leq A_{u/2} \leq 2t, F_{u/2,t}\right).$$

Plugging this into (4.6), we have, $\mathbf{P}$-a.s., for all large $t\in\mathbb{N}$,

$$\frac{P_x^G(Y_{2t} = x)}{\deg_G(x)} \leq c_8 t^{-d/2} + \frac{c_9}{t} \int_0^{2t} u^{-d/2} P_x^G\left(t/2 \leq A_{u/2} \leq 2t, F_{u/2,t}\right) du. \qquad (4.7)$$

Note, on $F_{u/2,t}$, it holds that

$$A_{u/2} = \int_0^{u/2} \deg_G(Z_s) ds \leq \frac{u}{2} \max_{|x|\leq c_3(2t)^{p+1}} \deg_G(x),$$

and so

$$P_x^G\left(t/2 \leq A_{u/2} \leq 2t, F_{u/2,t}\right) \ \leq\ \left(\frac{u}{t}\right)^{(d-1)/2} E_x^G\left(\left(\frac{A_{u/2}}{u/2}\right)^{(d-1)/2} \mathbb{I}(F_{u/2,t})\right) \quad (4.8)$$
$$\leq\ \left(\frac{u}{t}\right)^{(d-1)/2} \max_{|x|\leq c_3(2t)^{p+1}} \deg_G(x)^{(d-1)/2}.$$

In particular, together with (4.7), this implies, $\mathbf{P}$-a.s., for all large $t\in\mathbb{N}$,

$$\frac{P_x^G(Y_{2t} = x)}{\deg_G(x)} \leq c_{10} t^{-d/2} \max_{|x|\leq c_3(2t)^{p+1}} \deg_G(x)^{(d-1)/2}.$$

For bounding the max term, we note that

$$\mathbf{P}\left(\max_{|x|\leq c_3(2t)^{p+1}} \deg_G(x) \geq c_{11}\log t\right) \leq c_3(2t)^{d(p+1)} e^{-c_{11}\log t} \mathbf{E}[e^{\deg_G(\rho)}] \leq c_{12} t^{d(p+1)-c_{11}},$$

which, by taking $c_{11} > d(p+1) + 1$, is summable over $t\in\mathbb{N}$. Consequently, on applying the Borel-Cantelli lemma, we obtain an estimate of the desired form for $(Y_t)_{t\geq 0}$.

To complete the proof, we need to transfer the estimate to discrete time. Note that we can write $Y_t = X_{T_t}$, where $(T_t)_{t\geq 0}$ is a unit rate Poisson process on $[0,\infty)$, independent of X. Hence we have that

$$P_x^G(Y_t = x) \ =\ P_x^G\left(X_{T_t} = x\right)$$
$$=\ \sum_{s\geq 0} P_x^G\left(X_{2s} = x\right) P_x^G(T_t = 2s)$$
$$\geq\ \min_{s\in[t/4,t]} P_x^G\left(X_{2s} = x\right) P_x^G(T_t \in [t/2, 2t])$$
$$\geq\ P_x^G(X_{2t} = x)\left(1 - P_t^G(|T_t - t| > t/2)\right).$$

By Chebyshev's inequality, it holds that $P_t^G(|T_t - t| > t/2) \leq \frac{4}{t^2}\mathrm{Var}_x^G(T_t) = \frac{4}{t}$. Consequently, for $t\geq 8$, it holds that $P_x^G(X_{2t} = x) \leq 2P_x^G(Y_t = x)$, and so the result follows from the continuous-time estimate. $\qquad\square$

Remark 4.3. By Jensen's inequality and Fubini's theorem, we have that

$$E_x^G\left(\left(\frac{A_{u/2}}{u/2}\right)^{(d-1)/2}\right) \leq \frac{2}{u}\int_0^{u/2} E_x^G\left(\deg_G(Z_s)^{(d-1)/2}\right) ds. \qquad (4.9)$$

Hence if we could prove a bound of the form $E_x^G(\deg_G(Z_s)^{(d-1)/2}) \leq C_G$ for all $s\geq 0$, where C_G is a random constant depending only on the environment G, then we would obtain the quenched upper bound without a logarithm.

Proof of upper bounds of Theorem 1.5. Similarly to (1.2), the upper bound of (1.5) follows from [17, Theorem 1]. Indeed, it is proved there that there exist deterministic constants c, δ such that, $\mathbf{P}$-a.s.,

$$p_{2t}^G(\rho, \rho) \le ct^{-\frac{d}{s-d}} (\log t)^\delta$$

holds for $t \ge T$, where T is a random variable that satisfies: for any $\eta > 0$, there exists a constant C such that

$$\mathbf{P}(T > t) \le Ct^{-\eta}.$$

(Again, we highlight that, although the general bound of [17] is given for the continuous-time random walk, this is readily transferred to discrete time by applying the argument used in the proof of [4, Theorem 5.14].) Hence taking $\eta \ge d/(s - d)$ yields

$$\mathbf{E}\left(p_{2t}^G(\rho, \rho)\right) \le ct^{-\frac{d}{s-d}} (\log t)^\delta + \mathbf{P}(T > t) \le ct^{-\frac{d}{s-d}} (\log t)^\delta.$$

The proof of (1.6) and (1.7) can be obtained using the estimates in the quenched cases. Indeed, for $d = 1$, one just takes the expectation under $\mathbf{E}$ of (1.3), recalling from the proof of the latter result that the constant in the upper bound is deterministic, and the bound holds for all $t \in \mathbb{N}$. For $d \ge 2$, we return to (4.7), replacing the first term in the upper bound by the probability that it is bounding:

$$\frac{P_x^G(Y_{2t} = x)}{\deg_G(x)} \le 2P_x^G(E_{2t}^c) + \frac{c_9}{t} \int_0^{2t} u^{-d/2} P_x^G\left(t/2 \le A_{u/2} \le 2t, F_{u/2,t}\right) du.$$

For the expectation of the first term, we have from (4.4) with $\lambda = d/2$ and p chosen suitably large that

$$\mathbf{E}\left(P_x^G(E_{2t}^c)\right) \le c_4 t^{-d/2} + c_5 t^{d/2+1-pr} \le c_{13} t^{-d/2}.$$

For the expectation of the second term, we apply (4.8) and (4.9) to deduce that

$$\mathbf{E}\left(P_x^G\left(t/2 \le A_{u/2} \le 2t, F_{u/2,t}\right)\right) \le \left(\frac{u}{t}\right)^{(d-1)/s} \frac{2}{u} \int_0^{u/2} \mathbf{E}\left(E_x^G\left(\deg_G(Z_s)^{(d-1)/2}\right)\right) ds.$$

To obtain the desired bound in the continuous-time setting, it follows that it is enough to prove that there exists a constant C, independent of s, such that

$$\mathbf{E}\left(E_x^G\left(\deg_G(Z_s)^{(d-1)/2}\right)\right) \le C.$$

To prove this, note that the environment process $(G_{Z_s})_{s\ge 0}$, where G_x is the graph G translated by $-x$, is invariant and reversible with respect to the measure m (as introduced in the previous proof) under the annealed measure, see, for instance, [17, Section 4.1]. Thus, setting $f(G) := \deg_G(\omega(\rho))^{(d-1)/2}$, we have

$$
\begin{aligned}
\mathbf{E}\left(E_x^G\left(\deg_G(Z_s)^{(d-1)/2}\right)\right) &= \mathbf{E}\left(E_x^G\left(f(G_{Z_s})^{(d-1)/2}\right)\right) \\
&= \mathbf{E}\left(E_x^G\left(f(G_{Z_0})^{(d-1)/2}\right)\right) \\
&= \mathbf{E}\left(\deg_G(x)^{(d-1)/2}\right),
\end{aligned}
$$

and we further have that the right hand side is finite when $s > d$. In particular, this establishes that

$$\mathbf{E}\left(\frac{P_x^G(Y_{2t} = x)}{\deg_G(x)}\right) \le ct^{-d/2}.$$

We can transfer this result to the discrete-time process $(X_t)_{t\ge 0}$ exactly as in the quenched case. $\qquad\square$

5   Open questions

Now we have completed the proofs of our main results, we collate a number of issues left open by the present work. (Some of these are discussed in more detail elsewhere.)

1. In the quenched results of (1.2) and (1.4), and the annealed results of (1.5) and (1.6), it is natural to optimise the log exponents. For the annealed bounds, one would expect a log term in the case $d \geq 2$, $s = d + 2$ only. One might conjecture that this is also the case for the quenched bounds.

2. As in [17, 18], one might seek to derive similar heat kernel bounds to ours when the assumption that nearest-neighbour bonds are present is dropped. At least for the lower bounds, we have reduced the problem to checking the conditions (V) and (BS) (recall Remark 4.2), and verified these in the case $s \in (d, 2d)$ (see Section 6.4 below). Is it also possible to check (V) and (BS) in the case $s \geq 2d$?

3. Our results support the extension of the Gaussian regime of Theorem 1.1(b) to $d \geq 1$, $s > \min\{d + 2, 2d\}$, and the extension of the stable scaling regime of Theorem 1.1(a) to $d \geq 1$, $s \in (d, \min\{d + 2, 2d\})$. Can this be proved? Some discussion of the latter case is provided in [18, Section 3].

4. In what sense is it possible to determine the walk dimension of $(X_t)_{t \geq 0}$, that is, the exponent governing the space-time scaling of this process? A related problem is to establish bounds that satisfactorily describe the off-diagonal decay of the heat kernel, for which the techniques of the current article are insufficient. (As noted in the introduction, for nearest-neighbour bond percolation, quenched and annealed off-diagonal Gaussian heat kernel estimates are established in [2].)

5. All questions remain open in the case $d = 1$, $s = 2$. What can be said here?

6. Throughout this paper, we consider unweighted random graphs partly because Proposition 2.1 (which is [32, Theorem 3.7]) is stated in this setting. It would be interesting to extend our results to weighted random graphs, including those arising in random conductance models. Such an extension would potentially be applicable to the model of [16]. In particular, the latter paper established heat kernel estimates for long-range random conductance models on integer lattices when the conductance between x and y is given by $w_{x,y}/|x - y|^{d+\alpha}$, where $\{w_{x,y} = w_{y,x} \geq 0 : x, y \in \mathbb{Z}^d\}$ are independent and satisfy some moment condition. Is it possible to extend our approach cover this random conductance model when the $\{w_{x,y}\}$ have a translation-invariant distribution?

6   Appendix

We finish with a miscellany of results related to the heat kernel estimation techniques and long-range percolation model of this article. Lemma 6.2 in particular is required for our lower heat kernel estimates.

6.1   Heat kernel lower bounds on $\mathbb{Z}^d$

In this section, we explain how to check the assumptions of Section 2 for the graph $\mathbb{Z}^d$. For the statement of the next result, we write B_n for the ℓ_∞-ball of radius n, centred at 0, and R for the effective resistance on the integer lattice (see [4, Chapter 2], for example).

Lemma 6.1. For $0 \leq m < n$,

$$R(B_m, B_n^c) \geq \begin{cases} c_1(n - m), & \text{for } d = 1; \\ c_2(\log(n) - \log(m)), & \text{for } d = 2; \\ c_d(m^{2-d} - n^{2-d}), & \text{for } d \geq 3. \end{cases}$$

Proof. By applying the Nash-Williams inequality (see [33, Proposition 9.15], for example), we have the following:

$$R\left(B_m, B_n^c\right) \geq \sum_{l=m}^{n-1} \frac{1}{l^{d-1}},$$

from which the result readily follows. $\qquad\square$

Now, tile B_n, $n \geq t^{1/2}$, by boxes (Ω_i) that are each translations of $B_{t^{1/2}}$. Let A_i be the central part of Ω_i, as given by a translation of $B_{(1-\lambda^{-1})t^{1/2}}$. It is then the case that

$$\sum_i \text{cap}_{\Omega_i}(A_i) \preceq \begin{cases} \frac{n}{t^{1/2}} \times t^{-1/2}\lambda = C_\lambda n t^{-1}, & \text{for } d = 1; \\ \left(\frac{n}{t^{1/2}}\right)^2 \times \log(1/(1-\lambda^{-1}))^{-1} = C_\lambda n^2 t^{-1}, & \text{for } d = 2; \\ \left(\frac{n}{t^{1/2}}\right)^d \times \frac{\left(t^{1/2}\right)^{d-2}}{(1-\lambda^{-1})^{2-d}-1} = C_\lambda n^d t^{-1}, & \text{for } d \geq 3. \end{cases}$$

This is enough to check (A3) in this setting. The remaining assumptions are straightforward to check.

6.2 Quenched lower bound from simple random walk scaling limit

The following bound is adapted from [13, Lemma 5.1], and implies that a scaling limit for a random walk on a random graph of an appropriate form immediately yields a quenched heat kernel lower bound.

Lemma 6.2. Let X be a simple random walk on a connected, locally-finite graph $G = (V, E)$, started at root vertex ρ, and $p_t^G(x, y)$ be its heat kernel (with respect to the measure $\tilde{\pi}(x) = \deg_G(x)$). Let d_G be a metric on V, and suppose that: for some constants $d_w, d_f \in (0, \infty)$, the laws of

$$\left(n^{-\frac{1}{d_w}} d_G(\rho, X_{nt})\right)_{t \geq 0}, \qquad n \geq 1, \tag{6.1}$$

form a tight sequence in $L^1([0, 1])$, and also

$$\sup_{n \geq 1} n^{-d_f} \tilde{\pi}\left(\{x : d_G(\rho, x) \leq n\}\right) < \infty. \tag{6.2}$$

It is then the case that there exists a constant $c > 0$ such that: for all $t \geq 1$,

$$p_{2t}^G(\rho, \rho) \geq c t^{-\frac{d_f}{d_w}}.$$

Proof. We have that

$$
\begin{aligned}
p_{2t}^G(\rho, \rho) &= \sum_{x \in V} p_t^G(\rho, x) p_t^G(x, \rho) \tilde{\pi}(x) \\
&\geq \sum_{x \in V : d_G(\rho,x) \leq n} p_t^G(\rho, x)^2 \tilde{\pi}(x) \\
&\geq \frac{P_\rho^G\left(d_G(\rho, X_t) \leq Cn\right)^2}{\tilde{\pi}\left(\{x \in V : d_G(\rho, x) \leq Cn\}\right)} \\
&\geq c(Cn)^{-d_f} P_\rho^G\left(d_G(\rho, X_t) \leq Cn\right)^2
\end{aligned}
$$

where for the first inequality we use the symmetry of the heat kernel for the second we apply the Cauchy-Schwarz inequality, and for the third we appeal to (6.2). Now, by

applying the monotonicity of the on-diagonal part of the heat kernel, it follows that

$$
\begin{aligned}
p_{2t}^G(\rho,\rho) \;&\geq\; \left(t^{-1}\sum_{s=t}^{2t-1} p_{2s}^G(\rho,\rho)^{\frac{1}{2}}\right)^2 \\[2mm]
&\geq\; c(Cn)^{-d_f}\left(t^{-1}\sum_{s=t}^{2t-1} P_\rho^G\left(d_G(\rho,X_s)\leq Cn\right)\right)^2 \\[2mm]
&=\; c(Cn)^{-d_f}\left(E_\rho^G\left(t^{-1}\sum_{s=t}^{2t-1}\mathbb{I}_{\{d_G(\rho,X_s)\leq Cn\}}\right)\right)^2.
\end{aligned}
$$

Setting $n=t^{\frac{1}{d_w}}$, this yields

$$
p_{2t}^G(\rho,\rho)\geq cC^{-d_f}t^{-\frac{d_f}{d_w}}\left(E_\rho^G\left(t^{-1}\sum_{s=t}^{2t-1}\mathbb{I}_{\{d_G(\rho,X_s)\leq Ct^{\frac{1}{d_w}}\}}\right)\right)^2.
$$

Finally, given (6.1), the Kolmogorov-Riesz compactness theorem (see [28, Theorem 5], for example) implies that, by taking a suitably large value of C,

$$
E_\rho^G\left(t^{-1}\sum_{s=t}^{2t-1}\mathbb{I}_{\{d_G(\rho,X_s)\leq Ct^{\frac{1}{d_w}}\}}\right)=1-E_\rho^G\left(t^{-1}\sum_{s=t}^{2t-1}\mathbb{I}_{\{t^{-\frac{1}{d_w}}d_G(\rho,X_s)>C\}}\right)\geq\frac{1}{2},
$$

uniformly in t. In conjunction with the previous bound, this completes the proof. □

Remark 6.3. In examples, d_w will typically represent the so-called walk dimension of X, which is the exponent governing the space-time scaling of the random walk (with respect to the metric d_G). This is also sometimes called the escape time exponent. Moreover, d_f will be the volume growth exponent (again, with respect to the metric d_G). In the case when d_G is the usual shortest path graph distance (on a suitably regular graph), discussion of the possible values of d_w and d_f appears in [3].

Remark 6.4. If in place of condition (6.1) one had that the laws of

$$
n^{-\frac{1}{d_w}}d_G(\rho,X_n),\qquad n\geq 1,
$$

form a tight sequence, then one would be able to deduce the same result by an easier proof. In particular, the integration over time would not be necessary. Whilst this would be enough for us in the Gaussian case, we need the above L^1 version to deal with the weaker convergence statement that is known to hold in the stable case.

6.3 Quenched invariance principle in one-dimension via resistance scaling

In this section, via the resistance scaling techniques of [19] (see also [24]), we establish a quenched invariance principle for simple random walk on LRP(d,s) for $d=1$ and $s>2$ (cf. Theorem 1.1(b)), which, in conjunction with Lemma 6.2, gives an on-diagonal lower bound for the heat kernel of the model in question. We assume nearest-neighbour edges are present, i.e. $q=1$. Our proof gives an alternative viewpoint to the arguments of [18], which used a martingale approach, and [36], which applied the corrector method. (Note that it was also the case in both [18] and [36] that nearest-neighbour edges were assumed to be present, which ensures percolation occurs; see [10] for discussion of percolation for long-range percolation with $d=1$.) The particular form of our proof is closely related to that used to understand the scaling of the Mott random walk in [20]. Since it is not an original result, we are brief with the details.

Proposition 6.5. If $d = 1$ and $s > 2$, then for **P**-a.e. realisation of LRP(d, s), the law of

$$\left(n^{-\frac{1}{2}} X_{nt}\right)_{t \geq 0}$$

on $C([0, \infty))$ converges weakly to that of $(B_{\sigma^2 t})_{t \geq 0}$, where $(B_t)_{t \geq 0}$ is standard Brownian motion on $\mathbb{R}^d$, and $\sigma^2 \in (0, \infty)$ is a deterministic constant.

Proof. Let $\tilde{\pi}$ be the measure on $\mathbb{Z}$ given by $\tilde{\pi}(\{x\}) = \deg_G(x)$. Since $C := \mathbf{E}(\tilde{\pi}(\{0\})) \in (0, \infty)$, we readily deduce from the ergodic theorem that, **P**-a.s.,

$$n^{-1} \tilde{\pi}\left(\{an, \ldots, bn\}\right) \to C(b - a), \qquad \forall a, b \in \mathbb{R}, \ a < b. \tag{6.3}$$

Writing R for the effective resistance on $\mathbb{Z}$, we further claim that there exists a deterministic constant $R_\infty \in (0, \infty)$ such that, **P**-a.s.,

$$\left(n^{-1} R(xn, yn)\right)_{x, y \in \mathbb{R}} \to (R_\infty |x - y|)_{x, y \in \mathbb{R}}, \tag{6.4}$$

uniformly on compacts. To check this, we first apply the triangle inequality for the effective resistance and Kingman's subadditive ergodic theorem to deduce that, **P**-a.s.

$$\left(n^{-1} R(0, xn)\right)_{x \in \mathbb{R}} \to (R_\infty |x|)_{x \in \mathbb{R}},$$

uniformly on compacts, where $R_\infty := \inf_{n \geq 1} n^{-1} \mathbf{E} R(0, n)$. We note that $R_\infty \leq 1$, because the presence of nearest-neighbour edges ensures that $R(0, n) \leq n$. Now, from [18, Lemma 10.1], we have that the cut-points of the underlying graph are dense on the appropriate scale, where we say that x is a cut-point for G if $\{x, x + 1\}$ is the only edge in E that crosses this interval. In particular, it is an elementary consequence of [18, Lemma 10.1] that if C_x is the closest cut-point to x that lies on the left-hand side of x, then, **P**-a.s., $n^{-1} C_{xn} \to x$ uniformly on compacts. It follows that, **P**-a.s., uniformly over compact regions of $0 \leq x \leq y$,

$$\begin{aligned}
&\left|n^{-1} R(xn, yn) - R_\infty |x - y|\right| \\
&\leq \quad \left|n^{-1} R(C_{xn}, C_{yn}) - R_\infty |x - y|\right| + 2 \max_{z \in \{x, y\}} n^{-1} |zn - C_{zn}| \\
&= \quad \left|n^{-1} R(C_0, C_{yn}) - n^{-1} R(C_0, C_{xn}) - R_\infty |x - y|\right| + 2 \max_{z \in \{x, y\}} n^{-1} |zn - C_{zn}| \\
&\leq \quad \left|n^{-1} R(0, yn) - n^{-1} R(0, xn) - R_\infty |x - y|\right| + 6 \max_{z \in \{0, x, y\}} n^{-1} |zn - C_{zn}| \\
&\to \quad 0,
\end{aligned}$$

where to deduce the equality, we apply the series law for resistors, which clearly holds at cut-points. To complete the proof of (6.4), it remains to check that $R_\infty > 0$. Since $R(0, n)$ is bounded below by the number of cut-points between 0 and n, this can be deduced by another application of [18, Lemma 10.1].

Moreover, we have that, **P**-a.s.,

$$\lim_{r \to \infty} \liminf_{n \to \infty} n^{-1} R\left(0, \{-rn, \ldots, rn\}^c\right) = \infty. \tag{6.5}$$

Indeed, if we define C_0 and C_{rn} as above, and let $\tilde{C}_{-rn}$ be the closest cut-point to $-rn$ that lies on the right-hand side of $-rn$, then

$$R\left(0, \{-rn, \ldots, rn\}^c\right) \geq R\left(C_0, \{C_{rn}, \tilde{C}_{-rn}\}\right) - R(0, C_0) \geq \frac{1}{2} \min_{z \in \{C_{rn}, \tilde{C}_{-rn}\}} R(0, z) - R(0, C_0),$$

where we have applied the parallel law to deduce the second inequality. Hence applying the conclusion of the previous paragraph for point-wise resistances yields that

$$\liminf_{n\to\infty} n^{-1} R\left(0, \{-rn, \dots, rn\}^c\right) \geq \frac{R_\infty r}{2},$$

which clearly implies (6.5).

Putting (6.3) and (6.4) together similarly to the argument of [20, Theorem A.1], we obtain that

$$\left(\mathbb{Z}, n^{-1}R, n^{-1}\tilde{\pi}, 0, n^{-1}I_{\mathbb{Z}}\right),$$

where $I_{\mathbb{Z}}$ is the identity map on $\mathbb{Z}$, $\mathbf{P}$-a.s. converges in the spatial Gromov-Hausdorff-vague topology (see [19, Section 7] and [24, Section 2.2] for details) to

$$\left(\mathbb{R}, R_\infty d_E, C\mathcal{L}, 0, I_{\mathbb{R}}\right),$$

where d_E is the Euclidean metric, $\mathcal{L}$ is the Lebesgue measure on $\mathbb{R}$, and $I_{\mathbb{R}}$ is the identity map on $\mathbb{R}$. Together with the resistance divergence of (6.5), this enables us to apply [19, Theorem 7.1] to deduce a Brownian motion scaling limit for the continuous-time version of X, with mean one exponential holding times. The result for the discrete-time process readily follows. $\qquad\square$

6.4 Long-range percolation beyond the nearest-neighbour case

As we noted in Remark 4.2, to go beyond the nearest-neighbour case in establishing heat kernel lower bounds via the approach of this article, it will suffice to check the conditions (V) and (BS). In this section, we describe some progress in this direction, which allows us to consider the non-nearest-neighbour model of Section 3 for $d \geq 1$ and $s \in (d, 2d)$.

For (V), the essential work was completed by Biskup in [12], where estimates on the size of a largest percolation cluster in a box were given. In the following lemma, we transfer the desired estimate to the vertex set V_n. (We recall that G is the infinite cluster of the long-range percolation model, and $G_n = (V_n, E_n)$ is the largest connected component of $G \cap [-n, n]^d$.)

Lemma 6.6. If $d \geq 1$ and $s \in (d, 2d)$, then LRP(d,s) satisfies (V).

Proof. Let $\mathcal{C}_1(n)$ be the largest connected component of LRP(d,s) inside $[-n, n]^d$. It is proved as [12, Theorem 3.2] that, for any $s' \in (s, 2d)$, there exists a constant $\varepsilon > 0$ such that

$$\mathbf{P}\left(|\mathcal{C}_1(n)| < \varepsilon n^d\right) \leq e^{-\varepsilon n^{2d-s'}} \tag{6.6}$$

for all large n. We seek to replace $\mathcal{C}_1(n)$ by V_n in this estimate, which we will do by showing that $\mathcal{C}_1(n)$ is a part of the infinite component with suitably high probability (cf. the proof of [12, Corollary 3.3]). To this end, for a given n, define $r_n^k := 2^k n$, and let x_n^k be a sequence of points on the first coordinate axis such that $x_n^0 = 0$ and $|x_n^{k+1} - x_n^k| = 3r_n^k$. In particular, the ℓ_∞-balls $B_\infty(x_n^k, r_n^k)$, $k \geq 1$ are disjoint, but consecutive elements of the sequence touch. Write $\mathcal{C}_n^k$ for the largest connected component of LRP(d,s) inside $B_\infty(x_n^k, r_n^k)$, and define $\mathcal{C}_n^k \not\leftrightarrow \mathcal{C}_n^{k+1}$ to be the event that the two components in question are not connected by a direct edge. Applying (6.6), we then have that, for large n,

$$
\begin{aligned}
\mathbf{P}\left(\mathcal{C}_n^k \not\leftrightarrow \mathcal{C}_n^{k+1}\right) &\leq 2e^{-\varepsilon(r_n^k)^{2d-s'}} + \mathbf{P}\left(\mathcal{C}_n^k \not\leftrightarrow \mathcal{C}_n^{k+1},\ |\mathcal{C}_n^k| \geq \varepsilon(r_n^k)^d,\ |\mathcal{C}_n^{k+1}| \geq \varepsilon(r_n^{k+1})^d\right) \\
&\leq 2e^{-\varepsilon(r_n^k)^{2d-s'}} + e^{-\varepsilon^2 (r_n^k)^d (r_n^{k+1})^d / (2r_n^k + 2r_n^{k+1})^s} \\
&\leq Ce^{-2c(n2^k)^{2d-s'}} \\
&\leq Ce^{-c(n^{2d-s'} + 2^{k(2d-s')})},
\end{aligned}
$$

where for the second inequality we have used the fact that the maximal distance between points in $\mathcal{C}_n^k$ and $\mathcal{C}_n^{k+1}$ is $2r_n^k + 2r_n^{k+1}$, and for the last, we use that $a + b \leq 2ab$ for $a, b \geq 1$. Hence, writing $\mathcal{C}_n^k \leftrightarrow \mathcal{C}_n^{k+1}$ for the event that the two components in question are connected by a direct edge, we find that

$$
\begin{aligned}
\mathbf{P}\left(V_n = \mathcal{C}_1(n)\right) &= \mathbf{P}\left(\mathcal{C}_1(n) \subseteq G\right) \\
&\geq \mathbf{P}\left(\mathcal{C}_n^k \leftrightarrow \mathcal{C}_n^{k+1} \text{ for all } k \geq 0\right) \\
&\geq 1 - \sum_{k=0}^{\infty} C e^{-c(n^{2d-s'} + 2^{k(2d-s')})} \\
&= 1 - C e^{-cn^{2d-s'}}.
\end{aligned}
\tag{6.7}
$$

Putting this bound together with (6.6), we readily obtain (V). $\qquad \square$

We next present sufficient conditions for the LRP(d,s) model of Section 3 to satisfy the Benjamini-Schramm convergence condition (BS). Roughly speaking, the first of the conditions we introduce means that there is only a small probability that a long path avoids the largest connected component in a box, and the second one implies a weak law of large numbers for the size of the component. (Clearly, both conditions are trivial in the nearest-neighbour case.)

Lemma 6.7. Suppose LRP(d,s) (as defined in Section 3) satisfies the volume condition (V). Let $(a_n)_{n \geq 1}$ be a divergent sequence of positive integers such that $a_n = o(n)$, and suppose that

(a) $\sup_{x \in W_n^*} \mathbf{P}(x \in G \setminus W_n) = o(1)$,
(b) $\mathrm{Var}(|W_n|) = o(n^{2d})$,

where $W_n^* = [-n + a_n, n - a_n]^d \cap \mathbb{Z}^d$ and $W_n = W_n^* \cap V_n$. Then the random rooted graphs $(G_n, \rho_n)_{n \geq 1}$, with ρ_n uniformly chosen in V_n, Benjamini-Schramm converge to (G, ρ), conditioned that $\rho \in G$.

Proof. Assume that (V), (a) and (b) hold. We need to show that for any finite graph H and $r \in \mathbb{N}$,

$$
\lim_{n \to \infty} \mathbf{P}(B_{G_n}(\rho_n, r) = H) = \mathbf{P}(B_G(\rho, r) = H \mid \rho \in G).
\tag{6.8}
$$

First, we have

$$
\begin{aligned}
\mathbf{P}(\rho_n \notin W_n) = \mathbf{E}(|V_n \setminus W_n| / |V_n|) &\preceq n^{-d} \mathbf{E}(|V_n \setminus W_n|) + \mathbf{P}(|V_n| \leq cn^d) \\
&\preceq n^{-d} n^{d-1} a_n + \exp(-c(\log n)^2) = o(1).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathbf{P}[B_{G_n}(\rho_n, r) = H] &= \mathbf{P}[B_{G_n}(\rho_n, r) = H \mid \rho_n \in W_n] + o(1) \\
&= \mathbf{E}\left(\frac{1}{|W_n|} \sum_{x \in W_n} \mathbb{I}(B_{G_n}(x, r) = H)\right) + o(1) \\
&= \mathbf{E}\left(\frac{1}{|W_n|} \sum_{x \in W_n^*} \mathbb{I}(B_{G_n}(x, r) = H; x \in W_n)\right) + o(1).
\end{aligned}
$$

Moreover, by translation invariance,

$$
\begin{aligned}
\mathbf{P}[B_G(\rho, r) = H \mid \rho \in G] &= \frac{\mathbf{P}\left(B_G(\rho, r) = H; \rho \in G\right)}{\mathbf{P}(\rho \in G)} \\
&= \frac{1}{|W_n^*| \mathbf{P}(\rho \in G)} \mathbf{E}\left(\sum_{x \in W_n^*} \mathbb{I}(B_G(x, r) = H; x \in G)\right).
\end{aligned}
$$

It follows from the last two equations that

$$|\mathbf{P}[B_{G_n}(\rho_n, r) = H] - \mathbf{P}[B_G(\rho, r) = H \mid \rho \in G]| \leq \left| \mathbf{E}\left(\frac{A_n}{B_n} - \frac{A'_n}{B'_n} \right) \right| + o(1), \qquad (6.9)$$

where

$$A_n = \sum_{x \in W_n^*} \mathbb{I}(B_G(x, r) = H; \ x \in G), \qquad B_n = |W_n^*| \mathbf{P}(\rho \in G),$$
$$A'_n = \sum_{x \in W_n^*} \mathbb{I}(B_{G_n}(x, r) = H; \ x \in W_n), \qquad B'_n = |W_n|.$$

Observe that

$$\left| \mathbf{E}\left(\frac{A_n}{B_n} - \frac{A'_n}{B'_n} \right) \right| \leq \frac{1}{B_n} \mathbf{E}[|A_n - A'_n|] + \mathbf{E}\left(\frac{A'_n |B_n - B'_n|}{B_n B'_n} \right). \qquad (6.10)$$

To bound the first term, note that

$$\begin{aligned}
\mathbf{E}|A_n - A'_n| &\leq \mathbf{E} \sum_{x \in W_n^*} \left(\mathbb{I}(x \in G \setminus W_n) + \mathbb{I}(x \in G, \ B_G(x, r) \not\subseteq [-n, n]^d) \right) \\
&\preceq n^d \max_{x \in W_n^*} \left(\mathbf{P}(x \in G \setminus W_n) + \mathbf{P}(x \in G, \ B_G(x, r) \not\subseteq [x - a_n, x + a_n]^d) \right) \\
&= o(n^d) + n^d \mathbf{P}(\rho \in G, \ B_G(\rho, r) \not\subseteq [-a_n, a_n]^d) \\
&= o(n^d),
\end{aligned}$$

where we have used that $|W_n^*| \asymp n^d$ for the second inequality, (a) and translation invariance for the first equality, and the almost-sure finiteness of $B_G(\rho, r)$ (and the divergence of $(a_n)_{n \geq 1}$) for the second equality. Since $B_n \asymp n^d$, it follows that

$$\frac{1}{B_n} \mathbf{E}[|A_n - A'_n|] = o(1). \qquad (6.11)$$

Using that $A'_n \leq |W_n^*| \asymp B_n$, we further obtain that

$$\begin{aligned}
\mathbf{E}&\left(\frac{A'_n |B_n - B'_n|}{B_n B'_n} \right) \\
&\preceq \mathbf{E}\left(\frac{|B_n - B'_n|}{B'_n} \right) = \mathbf{E}\left(\frac{||W_n| - |W_n^*| \mathbf{P}(\rho \in G)|}{|W_n|} \right) \\
&\preceq n^{-d} \mathbf{E}[||W_n| - |W_n^*| \mathbf{P}(\rho \in G)|] + n^d \mathbf{P}[|W_n| \leq cn^d] \\
&\preceq n^{-d} \left(\mathbf{E}\left((|W_n| - |W_n^*| \mathbf{P}(\rho \in G))^2 \right) \right)^{1/2} + \exp(-c(\log n)^2), \qquad (6.12)
\end{aligned}$$

by using the Cauchy-Schwarz inequality and (V) (and the fact that $|V_n| \preceq |W_n| + n^{d-1} a_n = |W_n| + o(n^d)$). Additionally,

$$\begin{aligned}
\mathbf{E}&\left((|W_n| - |W_n^*| \mathbf{P}(\rho \in G))^2 \right) = \mathbf{E}\left(\left(\sum_{x \in W_n^*} (\mathbb{I}(x \in W_n) - \mathbf{P}(x \in G)) \right)^2 \right) \\
&= \mathbf{E}\left(\left(\sum_{x \in W_n^*} (\mathbb{I}(x \in W_n) - \mathbf{P}(x \in W_n) - \mathbf{P}(x \in G \setminus W_n)) \right)^2 \right) \\
&\leq 2\mathbf{E}\left(\left(\sum_{x \in W_n^*} \mathbb{I}(x \in W_n) - \mathbf{P}(x \in W_n) \right)^2 \right) + 2|W_n^*|^2 \max_{x \in W_n^*} \mathbf{P}[x \in G \setminus W_n] \\
&= 2\mathrm{Var}[|W_n|] + 2|W_n^*|^2 \max_{x \in W_n^*} \mathbf{P}[x \in G \setminus W_n] \\
&= o(n^{2d}), \qquad (6.13)
\end{aligned}$$

by using (a) and (b). Combining the estimates (6.9)–(6.13), and using (a) again, we obtain (6.8). □

In the subsequent lemma, we apply Lemma 6.7 for $d \geq 1$ and $s \in (d, 2d)$. We highlight that the proof depends on the estimates (6.6) and (6.7) from the proof of Lemma 6.6, and also two further statements from [17, Theorem 2]. Inspecting the latter reference, one would find that [17, Theorem 2] is stated for the smaller range $s \in (d, \min\{d + 2, 2d\})$. However, as is commented below [17, Theorem 2] (and a careful checking of the argument establishes), this restriction, which is principally due to the main focus of the paper being on the stable regime, is not essential, and the percolation estimates of [17, Theorem 2] extend to the range $s \in (d, 2d)$.

Lemma 6.8. If $d \geq 1$ and $s \in (d, 2d)$, then LRP(d,s) satisfies (BS).

Proof. We will check the two conditions of Lemma 6.7.

Towards verifying (a), let $x \in W_n^*$, and $\mathcal{C}_1(x, a_n)$ be the largest connected component of LRP(d,s) in the ℓ_∞-ball $B_\infty(x, a_n)$. We first claim that if $a_n = (\log n)^\Delta$ for suitably large Δ, then

$$\max_{x \in W_n^*} \mathbf{P}\left(\mathcal{C}_1(x, a_n) \not\subseteq V_n\right) = o(1). \tag{6.14}$$

Letting $\mathcal{C}_1(n)$ and $\mathcal{C}_2(n)$ be the first and second largest components of LRP(d,s) inside $[-n, n]^d$, respectively, we have that, for any $\varepsilon > 0$,

$$\begin{aligned}
\mathbf{P}\left(\mathcal{C}_1(x, a_n) \not\subseteq V_n\right) &\leq \mathbf{P}\left(|\mathcal{C}_1(x, a_n)| < \varepsilon a_n^d\right) + \mathbf{P}\left(\mathcal{C}_1(x, a_n) \not\subseteq V_n, |\mathcal{C}_1(x, a_n)| \geq \varepsilon a_n^d\right) \\
&\leq \mathbf{P}\left(|\mathcal{C}_1(a_n)| < \varepsilon a_n^d\right) + \mathbf{P}\left(|\mathcal{C}_2(n)| \geq \varepsilon a_n^d\right) + \mathbf{P}\left(V_n \neq \mathcal{C}_1(n)\right),
\end{aligned}$$

where to obtain the second inequality we use that on the event $\{V_n = \mathcal{C}_1(n), \mathcal{C}_1(x, a_n) \not\subseteq V_n\}$, it holds that $|\mathcal{C}_2(n)| \geq |\mathcal{C}_1(x, a_n)|$. Moreover, we highlight that we have used the translation invariance of the model to derive a bound that does not depend on the choice of $x \in W_n^*$. By (6.6) and (6.7), the first and third probabilities above are $o(1)$ as $n \to \infty$. As for the second term, this is shown to be $o(1)$ in [17, Theorem 2] for Δ chosen large enough. Hence we have established (6.14).

Applying the estimate of previous paragraph, we find that

$$\begin{aligned}
\max_{x \in W_n^*} \mathbf{P}\left(x \in G \backslash W_n\right) &\leq o(1) + \max_{x \in W_n^*} \mathbf{P}\left(x \in G \backslash \mathcal{C}_1(x, a_n)\right) \\
&\leq o(1) + \max_{x \in W_n^*} \mathbf{P}\left(x \leftrightsquigarrow B_\infty(x, a_n)^c, x \notin \mathcal{C}_1(x, a_n)\right) \\
&= o(1) + \mathbf{P}\left(0 \leftrightsquigarrow B_\infty(0, a_n)^c, 0 \notin \mathcal{C}_1(a_n)\right)
\end{aligned}$$

where $x \leftrightsquigarrow B_\infty(x, a_n)^c$ means that x is connected to the complement of $B_\infty(x, a_n)$, but not necessarily by a single edge. The probability in the final line above is shown to be $o(1)$ in [17, Theorem 2]. This confirms (a).

For proving condition (b) of Lemma 6.7, we first observe

$$\mathrm{Var}(|W_n|) \preceq (n a_n)^d + \sum_{\substack{x, y \in W_n^*: \\ |x-y| > 2a_n}} \mathrm{Cov}(\mathbb{I}(x \in W_n), \mathbb{I}(y \in W_n)).$$

Now, some elementary manipulation of probabilities allows it to be checked that, for $x, y \in W_n^*$,

$$\begin{aligned}
\mathrm{Cov}(\mathbb{I}(x \in W_n), \mathbb{I}(y \in W_n)) &\leq \mathrm{Cov}(\mathbb{I}(x \in \mathcal{C}_1(x, a_n)), \mathbb{I}(y \in \mathcal{C}_1(y, a_n))) \\
&\quad + 2 \max_{z \in W_n^*} \mathbf{P}\left(z \in \mathcal{C}_1(z, a_n) \backslash W_n\right) \\
&\quad + 2 \max_{z \in W_n^*} \mathbf{P}\left(z \in W_n \backslash \mathcal{C}_1(z, a_n)\right).
\end{aligned}$$

Moreover, if $|x - y| > 2a_n$, then the first term in the upper bound is zero. To bound the first of the probabilities, note that if the event $z \in \mathcal{C}_1(z, a_n)\backslash W_n$ occurs, then it must be the case that $\mathcal{C}_1(z, a_n) \nsubseteq V_n$. Thus, by (6.14),

$$\max_{z \in W_n^*} \mathbf{P}\left(z \in \mathcal{C}_1(z, a_n)\backslash W_n\right) \leq \max_{z \in W_n^*} \mathbf{P}\left(z \in \mathcal{C}_1(z, a_n) \nsubseteq V_n\right) = o(1).$$

Additionally, since $W_n \subseteq V_n \subseteq G$, we have that

$$\max_{z \in W_n^*} \mathbf{P}\left(z \in W_n\backslash\mathcal{C}_1(z, a_n)\right) \leq \max_{z \in W_n^*} \mathbf{P}\left(z \in G\backslash\mathcal{C}_1(z, a_n)\right).$$

That the upper bound is $o(1)$ was established earlier in the proof. Combining the previous estimates yields (b), as desired. $\qquad\square$

From Remark 4.2 and Lemmas 6.6 and 6.8 (and the heat kernel upper bound of [17]), we obtain the following result for the long-range percolation model of Section 3, which we underline does not require nearest-neighbour connections. In particular, this result verifies that the on-diagonal heat kernel estimates of [17] are sharp (up to logarithmic factors) throughout the stable regime $s \in (d, \min\{d + 2, 2d\})$ (and not just when $s \in (d, d + 1)$), which is what is implied by the invariance principle of [18]).

Corollary 6.9. If $d \geq 1$ and $s \in (d, 2d)$, then LRP(d,s) satisfies the relevant lower heat kernel bounds of Theorems 1.3 and 1.5. If $d \geq 1$ and $s \in (d, \min\{d + 2, 2d\})$, then it also satisfies the spectral dimension result of Corollary 1.7.

References

[1] M. Aizenman, H. Kesten, and C. M. Newman, *Uniqueness of the infinite cluster and continuity of connectivity functions for short and long range percolation*, Comm. Math. Phys. **111** (1987), no.4, 505–531. MR901151

[2] M. T. Barlow, *Random walks on supercritical percolation clusters*, Ann. Probab. **32** (2004), no. 4, 3024–3084. MR2094438

[3] M. T. Barlow, *Which values of the volume growth and escape time exponent are possible for a graph?*, Rev. Mat. Iberoamericana **20** (2004), no. 1, 1–31. MR2076770

[4] M. T. Barlow, *Random walks and heat kernels on graphs*, London Mathematical Society Lecture Note Series, vol. 438, Cambridge University Press, Cambridge, 2017. MR3616731

[5] M. T. Barlow, T. Coulhon, and A. Grigor'yan, *Manifolds and graphs with slow heat kernel decay*, Invent. Math. **144** (2001), no. 3, 609–649. MR1833895

[6] M. T. Barlow, A. A. Járai, T. Kumagai, and G. Slade, *Random walk on the incipient infinite cluster for oriented percolation in high dimensions*, Comm. Math. Phys. **278** (2008), no. 2, 385–431. MR2372764

[7] M. T. Barlow and T. Kumagai, *Random walk on the incipient infinite cluster on trees*, Illinois J. Math. **50** (2006), no. 1-4, 33–65. MR2247823

[8] I. Benjamini and N. Berger, *The diameter of long-range percolation clusters on finite cycles*, Random Struct. Alg. **19** (2001), no. 2, 102–111. MR1848786

[9] I. Benjamini and O. Schramm, *Recurrence of distributional limits of finite planar graphs*, Electron. J. Probab. **6** (2001), no. 23, 13. MR1873300

[10] N. Berger, *Transience, recurrence and critical behavior for long-range percolation*, Comm. Math. Phys. **226** (2002), no. 3, 531–558. MR1896880

[11] N. Berger and M. Biskup, *Quenched invariance principle for simple random walk on percolation clusters*, Probab. Theory Rel. Fields **137** (2007), no. 1-2, 83–120. MR2278453

[12] M. Biskup, *On the scaling of the chemical distance in long-range percolation models*, Ann. Probab. **32** (2004), no. 4, 2938–2977. MR2094435

[13] M. Biskup, *Recent progress on the random conductance model*, Probab. Surv. **8** (2011), 294–373. MR2861133

[14] M. Biskup, X. Chen, T. Kumagai, and J. Wang, *Quenched invariance principle for a class of random conductance models with long-range jumps*, Probab. Theory Rel. Fields **180** (2021), no. 3-4, 847–889. MR4288333

[15] E. A. Carlen, S. Kusuoka, and D. W. Stroock, *Upper bounds for symmetric Markov transition functions*, Ann. Inst. H. Poincaré Probab. Statist. **23** (1987), no. 2, suppl., 245–287. MR898496

[16] X. Chen, T. Kumagai, and J. Wang, *Random conductance models with stable-like jumps: heat kernel estimates and Harnack inequalities*, J. Funct. Anal. **279** (2020), no. 7, 108656, 51. MR4108401

[17] N. Crawford and A. Sly, *Simple random walk on long range percolation clusters I: heat kernel bounds*, Probab. Theory Rel. Fields **154** (2012), no. 3-4, 753–786. MR3000561

[18] N. Crawford and A. Sly, *Simple random walk on long-range percolation clusters II: scaling limits*, Ann. Probab. **41** (2013), no. 2, 445–502. MR3077517

[19] D. A. Croydon, *Scaling limits of stochastic processes associated with resistance forms*, Ann. Inst. Henri Poincaré Probab. Stat. **54** (2018), no. 4, 1939–1968. MR3865663

[20] D. A. Croydon, R. Fukushima, and S. Junk, *Anomalous scaling regime for one-dimensional mott variable-range hopping*, preprint available at arXiv:2010.01779, 2020.

[21] D. A. Croydon and B. M. Hambly, *Local limit theorems for sequences of simple random walks on graphs*, Potential Anal. **29** (2008), no. 4, 351–389. MR2453564

[22] D. A. Croydon and B. M. Hambly, *Self-similarity and spectral asymptotics for the continuum random tree*, Stochastic Process. Appl. **118** (2008), no. 5, 730–754. MR2411518

[23] D. A. Croydon and B. M. Hambly, *Spectral asymptotics for stable trees*, Electron. J. Probab. **15** (2010), no. 57, 1772–1801. MR2738338

[24] D. A. Croydon, B. M. Hambly, and T. Kumagai, *Time-changes of stochastic processes associated with resistance forms*, Electron. J. Probab. **22** (2017), Paper No. 82, 41. MR3718710

[25] A. De Masi, P. A. Ferrari, S. Goldstein, and W. D. Wick, *An invariance principle for reversible Markov processes. Applications to random motions in random environments*, J. Statist. Phys. **55** (1989), no. 3-4, 787–855. MR1003538

[26] J. Ding and A. Sly, *Distances in critical long range percolation*, preprint available at arXiv:1303.3995, 2013.

[27] P.-G. de Gennes, *La percolation: un concept unificateur*, La Recherche **7** (1976), no. 72, 919–927.

[28] H. Hanche-Olsen and H. Holden, *The Kolmogorov-Riesz compactness theorem*, Expo. Math. **28** (2010), no. 4, 385–394. MR3494284

[29] G. Kozma and A. Nachmias, *The Alexander-Orbach conjecture holds in high dimensions*, Invent. Math. **178** (2009), no. 3, 635–654. MR2551766

[30] T. Kumagai, *Random walks on disordered media and their scaling limits*, Lecture Notes in Mathematics, vol. 2101, Springer, Cham, 2014, Lecture notes from the 40th Probability Summer School held in Saint-Flour, 2010, École d'Été de Probabilités de Saint-Flour. [Saint-Flour Probability Summer School]. MR3156983

[31] T. Kumagai and J. Misumi, *Heat kernel estimates for strongly recurrent random walk on random media*, J. Theoret. Probab. **21** (2008), no. 4, 910–935. MR2443641

[32] J. R. Lee, *Conformal growth rates and spectral geometry on distributional limits of graphs*, Ann. Probab. **49** (2021), no. 6, 2671–2731. MR4348678

[33] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, Providence, RI, 2009, With a chapter by James G. Propp and David B. Wilson. MR2466937

[34] P. Mathieu and A. Piatnitski, *Quenched invariance principles for random walks on percolation clusters*, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **463** (2007), no. 2085, 2287–2307. MR2345229

[35] V. Sidoravicius and A.-S. Sznitman, *Quenched invariance principles for walks on clusters of percolation or among random conductances*, Probab. Theory Rel. Fields **129** (2004), no. 2, 219–244. MR2063376

[36] Z. Zhang and L. Zhang, *Scaling limits for one-dimensional long-range percolation: using the corrector method*, Statist. Probab. Lett. **83** (2013), no. 11, 2459–2466. MR3144026

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: http://www.imstat.org/publications/open.htm