The Annals of Applied Probability
2016, Vol. 26, No. 5, 2597–2625
DOI: 10.1214/15-AAP1155

UNORIENTED FIRST-PASSAGE PERCOLATION ON THE n-CUBE

BY ANDERS MARTINSSON

Chalmers University of Technology and University of Gothenburg

The n-dimensional binary hypercube is the graph whose vertices are the binary n-tuples $\{0,1\}^n$ and where two vertices are connected by an edge if they differ at exactly one coordinate. We prove that if the edges are assigned independent mean 1 exponential costs, the minimum length T_n of a path from $(0,0,\dots,0)$ to $(1,1,\dots,1)$ converges in probability to $\ln(1+\sqrt{2}) \approx 0.881$. It has previously been shown by Fill and Pemantle [*Ann. Appl. Probab.* **3** (1993) 593–629] that this so-called first-passage time asymptotically almost surely satisfies $\ln(1+\sqrt{2}) - o(1) \le T_n \le 1 + o(1)$, and has been conjectured to converge in probability by Bollobás and Kohayakawa [In *Combinatorics, Geometry and Probability* (*Cambridge*, 1993) (1997) 129–137 Cambridge]. A key idea of our proof is to consider a lower bound on Richardson's model, closely related to the branching process used in the article by Fill and Pemantle to obtain the bound $T_n \ge \ln(1+\sqrt{2}) - o(1)$. We derive an explicit lower bound on the probability that a vertex is infected at a given time. This result is formulated for a general graph and may be applicable in a more general setting.

1. Introduction. The n-dimensional binary hypercube $\mathbb{Q}_n$ is the graph with vertex set $\{0,1\}^n$ where two vertices share an edge if they differ at exactly one coordinate. We let $\hat{\mathbf{0}}$ and $\hat{\mathbf{1}}$ denote the all zeroes and all ones vertices, respectively. For any vertex $v \in \mathbb{Q}_n$, we let $|v|$ denote the number of coordinates of v that are 1. A path $v_0 \to v_1 \to \cdots \to v_k$ in $\mathbb{Q}_n$ is called *oriented* if $|v_i|$ is strictly increasing along the path.

First-passage percolation is a random process on a graph G, which was introduced by Hammersley and Welsh. In this process, each edge e in the graph is assigned a random variable W_e called the *passage time* of e. In this paper, the passage times will always be independent exponentially distributed random variables with expected value 1. The usual way in which this process is described is that there exists some vertex $v_0 \in G$ which is assigned some property, usually either that it is infected (v_0 is the source of some disease) or wet (v_0 is connected to a water source), which then spreads throughout the graph. The passage time of an edge corresponds to the time it takes for an infection to spread in any direction along the edge, that is, when a vertex v gets infected the infection spreads to each

Received June 2014; revised January 2015.
MSC2010 subject classifications. Primary 60K35; secondary 60C05.
Key words and phrases. First-passage percolation, Richardson's model, hypercube, branching translation process, lower bound on Richardson's model.

neighbor w after $W_{\{v,w\}}$ time, assuming w is not already infected at that time. More concretely, we can let the edge weights generate a metric on G. For a path γ in G, we define the *passage time* of γ as the sum of passage times of the edges along γ. Moreover, for any two vertices $v, w \in G$, we say that the *first-passage time* from v to w, denoted by $d_W(v, w)$, is the infimum of passage times over all paths from v to w in G. Then for any $v \in G$, the time at which v is infected is given by $d_W(v_0, v)$.

An alternative way to formulate first-passage percolation with independent exponentially distributed passage times is to consider the process $\{R(\cdot, t)\}_{t\geq 0}$, where for each $t \geq 0$, $R(v, t)$ is the map from the vertex set of G to $\{0, 1\}$ given by

$$R(v,t) = \begin{cases} 1, & \text{if } d_W(v_0, v) \leq t, \\ 0, & \text{otherwise,} \end{cases} \tag{1.1}$$

that is, $R(v, t)$ is the indicator function for the event that v is infected at time t. When the edge passage times are independent exponentially distributed with mean one, the memory-less property implies that the process $\{R(\cdot, t)\}_{t\geq 0}$ is Markovian, and its distribution is given by the initial condition $R(\cdot, 0) = \delta_{v_0,\cdot}$ together with the transitions $\{R(\cdot) \to R(\cdot) + \delta_{v,\cdot}\}$ at rate equal to the number of infected neighbors of v if v is healthy, and 0 if v is infected, see [3]. Here, $\delta_{\cdot,\cdot}$ denotes the Kronecker delta function. This Markov process is known as Richardson's model.

First-passage percolation and Richardson's model on the hypercube have previously been studied by Fill and Pemantle [4], and later by Bollobás and Kohayakawa [2]; see also example G7 in [1]. For Richardson's model, we always assume that the original infected vertex is $\hat{\mathbf{0}}$. The primary concerns in this study are the questions "What is the first-passage time between antipodal vertices in the cube, say $\hat{\mathbf{0}}$ and $\hat{\mathbf{1}}$?" and "How long does it take for Richardson's model to cover the entire cube?" Here, we will denote the first-passage time between $\hat{\mathbf{0}}$ and $\hat{\mathbf{1}}$ in $\mathbb{Q}_n$ by T_n. Note that T_n is then also the time when $\hat{\mathbf{1}}$ gets infected in Richardson's model on $\mathbb{Q}_n$. A simplified version of this quantity, the *oriented first-passage time* from $\hat{\mathbf{0}}$ to $\hat{\mathbf{1}}$, which we denote by $T_n^{\mathcal{O}}$, also appears in the literature. This is the smallest passage time of any oriented path from $\hat{\mathbf{0}}$ to $\hat{\mathbf{1}}$. Lastly, the *covering time* of $\mathbb{Q}_n$, denoted C_n, is the random amount of time in Richardson's model until all vertices are infected. Equivalently, $C_n = \max_{v\in\mathbb{Q}_n} d_W(\hat{\mathbf{0}}, v)$, the maximum first-passage time from $\hat{\mathbf{0}}$ to any other vertex in $\mathbb{Q}_n$.

The oriented first-passage time was first proposed by Aldous in [1]. By considering the expected number of oriented paths with passage time at most t, he observed that $T_n^{\mathcal{O}}$ is asymptotically almost surely bounded from below by $1 - o(1)$, and consequently conjectured that $T_n^{\mathcal{O}} \xrightarrow{p} 1$ as $n \to \infty$. This was later proved by Fill and Pemantle in [4]. Their argument for the upper bound is essentially a second moment analysis on the number of such paths together with a "variance reduction trick".

For the unoriented first-passage time from $\hat{\mathbf{0}}$ to $\hat{\mathbf{1}}$, Fill and Pemantle showed that, as $n \to \infty$, we have

$$\ln(1+\sqrt{2}) - o(1) \le T_n \le 1 + o(1) \tag{1.2}$$

with probability $1 - o(1)$. The upper bound follows directly from the result for $T_n^{\mathcal{O}}$. Fill and Pemantle give no explicit conjectures about the further behavior of T_n, but state that they doubt the upper bound is sharp. Prior to this article, this seems to be the best known upper bound on T_n. For the lower bound, they relay an argument by Durrett. In this argument, one considers a random process on $\mathbb{Q}_n$, which Durrett calls a branching translation process (BTP). We will postpone the definition of this process to the next section, but this is basically Richardson's model with the modification that we allow each site to contain multiple instances of the infection at the same time. Durrett argues that this process stochastically dominates Richardson's model in the sense that it is possible to couple the models such that the infected vertices in Richardson's model are always a subset of the so-called occupied vertices in the BTP. He proves that the time at which $\hat{\mathbf{1}}$ becomes occupied tends to $\ln(1+\sqrt{2})$ in probability as $n \to \infty$. As BTP stochastically dominates Richardson's model, this directly implies that $T_n \ge \ln(1+\sqrt{2}) - o(1) = 0.881\cdots - o(1)$ with probability $1 - o(1)$.

Bollobás and Kohayakawa [2] show that many global first-passage percolation properties on $\mathbb{Q}_n$, such as the covering time and the graph diameter with respect to $d_W(\cdot,\cdot)$, can be bounded from above in terms of T_n. They define the quantity

$$T_\infty = \inf\{t \in \mathbb{R} \mid \mathbb{P}(T_n \le t) \to 1 \text{ as } n \to \infty\}. \tag{1.3}$$

Their main result is that asymptotically almost surely $C_n \le T_\infty + \ln 2 + o(1)$ and the graph diameter is at most $T_\infty + 2\ln 2 + o(1)$. Note that it follows from the results by Fill and Pemantle that $\ln(1+\sqrt{2}) \le T_\infty \le 1$. Furthermore, it is easy to see that if T_n converges in probability as $n \to \infty$, then it must converge to T_∞. Bollobás and Kohayakawa explicitly conjectured that this is the case, and consequently referred to T_∞ as simply the first-passage percolation time between two antipodal vertices in $\mathbb{Q}_n$. While their article does not prove that T_n converges in probability, their approach can be used to prove some weaker statements for T_n. For instance, with some small modifications of their proof it follows that if T_n has a limit distribution as $n \to \infty$, then that limit must be trivial, meaning that T_n converges in probability.

Besides first-passage percolation, percolation on the hypercube with restriction to oriented paths has also been considered in regards to Bernoulli percolation by Fill and Pemantle (in the same article), and more recently, accessibility percolation[1] by Hegarty and the author in [5]. Common for these three cases of oriented

[1]The name accessibility percolation is not mentioned in the cited article. The term was coined by Joachim Krug and Stefan Nowak after its writing.

percolation is that the proofs are based on second moment analyses. Arguably, this is made possible by the relatively simple combinatorial properties of oriented paths. We have $n!$ oriented paths from $\hat{\mathbf{0}}$ to $\hat{\mathbf{1}}$ in $\mathbb{Q}_n$, all of length n and all equivalent up to permutation of coordinates. Perhaps more importantly, one can derive good estimates on the number of pairs of oriented paths from $\hat{\mathbf{0}}$ to $\hat{\mathbf{1}}$ that intersect a given number of times, something which is made possible by the natural representation of oriented paths as permutations. In contrast, general paths from $\hat{\mathbf{0}}$ to $\hat{\mathbf{1}}$ do not seem to have a similar representation in any meaningful way, and in any case, there is certainly a lot more variation between general paths than oriented such. Hence, it seems that these type of ideas from oriented percolation on the hypercube cannot be transferred to unoriented percolation.

The most promising approach to improve the result by Fill and Pemantle for T_n seems to be the BTP. Comparing the BTP to path-counting arguments, on the hypercube the former has the advantage that a number of relevant quantities, such as moment estimates, can be expressed by explicit analytical expressions, hence circumventing the problem of counting paths. However, beyond the fact that the BTP stochastically dominates Richardson's model, the relation between the two models is fairly subtle. It is therefore not immediately clear how proving anything about the BTP could imply upper bounds on the first-passage time.

In this article, we propose a way to do precisely this. A central idea of our approach is to consider a subprocess of the BTP with two important properties: First, Richardson's model is stochastically sandwiched between the full BTP and this subprocess, and second, it is possible to derive an explicit lower bound on the probability that a vertex is occupied at a given time in this subprocess, expressed in tractable quantities for the BTP. Applying these ideas to the hypercube, we are able to resolve the problem of determining the limit of T_n. This result is summarized in the following theorem, which is the main result of this paper.

THEOREM 1.1. *Let T_n denote the first-passage time from $\hat{\mathbf{0}}$ to $\hat{\mathbf{1}}$ in $\mathbb{Q}_n$ with exponentially distributed edge costs with mean* 1. *Then $T_n \to \ln(1+\sqrt{2})$ in probability and in L^p-norm for any $p \in [1,\infty)$ as $n \to \infty$. More precisely, for any fixed such p we have*

$$\|T_n - \ln(1+\sqrt{2})\|_p = \Theta\left(\frac{1}{n}\right). \tag{1.4}$$

Furthermore, $\mathbb{E}T_n = \ln(1+\sqrt{2}) + O(\frac{1}{n})$ *and* $\mathrm{Var}(T_n) = \Theta(\frac{1}{n^2})$.

A direct consequence of this result is that $T_\infty = \ln(1+\sqrt{2})$, which in particular improves the best known upper bound on the covering time to $\ln(1+\sqrt{2}) + \ln 2 + o(1) = 1.574 + \cdots + o(1)$. One can compare this with the best known lower bound $\frac{1}{2}\ln(2+\sqrt{5}) + \ln 2 - o(1) = 1.414\cdots - o(1)$, as shown by Fill and Pemantle.

Given this result for T_n, the question naturally arises how the minimizing path from $\hat{\mathbf{0}}$ to $\hat{\mathbf{1}}$ typically behaves. In particular, how long is this path (here length means the number of edges along the path), and how are the "backsteps" distributed along it. Let us denote this path by Γ_n. This question may also be of interest from the point of view of accessibility percolation. Though strictly speaking not part of the mathematical formulation of accessibility percolation, shorter paths are considered more biologically feasible. Hence, an important question for unoriented accessibility percolation on the hypercube is how much longer typical accessible paths are than in the oriented case.

We propose the following way to describe the asymptotic properties of Γ_n: Run a continuous-time simple random walk on $\mathbb{Q}_n$ which starts at $\hat{\mathbf{0}}$ and takes steps at rate n for $\ln(1+\sqrt{2})$ time, and condition on the event that the walk stops at $\hat{\mathbf{1}}$. Let Σ_n denote the traversed path.

THEOREM 1.2. *For any sequence $\{A_n\}_{n=1}^{\infty}$ where, for each n, A_n denotes a set of paths from $\hat{\mathbf{0}}$ to $\hat{\mathbf{1}}$ in $\mathbb{Q}_n$, we have that*

$$\lim_{n\to\infty} \mathbb{P}(\Sigma_n \in A_n) = 1 \quad \Rightarrow \quad \lim_{n\to\infty} \mathbb{P}(\Gamma_n \in A_n) = 1. \tag{1.5}$$

That is, any asymptotically almost sure property of Σ_n is an asymptotically almost sure property of the minimizing path from $\hat{\mathbf{0}}$ to $\hat{\mathbf{1}}$. In particular, the length of Γ_n is asymptotically almost surely $\sqrt{2}\ln(1+\sqrt{2})n \pm o(n)$.

In applying Theorem 1.2, it is helpful to note that each coordinate of a simple random walk on $\mathbb{Q}_n$ taking steps at rate n is an independent simple random walk on $\{0, 1\}$ taking steps at rate 1.

The remainder of the paper will be structured in the following way: In Section 2, we define the BTP and describe our stochastical sandwiching of Richardson's model. At the end of this section, we give an outline of the proof of Theorem 1.1. This proof is divided into three steps, which are shown in Sections 3, 4 and 5, respectively. Lastly, in Section 6 we give a short proof of Theorem 1.2 based on ideas from the preceding section.

2. Richardson's model, the BTP and uncontested particles. We first give an overview of the technique used by Durrett to obtain the lower bound on T_n in [4]. To accommodate Theorem 2.2 below, we present this technique in terms of a general graph G rather than just the hypercube. We remark that though Durrett only defined the branching translation process for the hypercube, the process can be extended to a general graph unambiguously. We let v_0 denote a fixed vertex in G. For simplicity, we will assume that G is finite, connected and simple.

The *branching translation process* (BTP), as introduced by Durrett, is a branching process on G defined in the following way: At time 0, we place a particle at v_0. After this, each existing particle generates offspring independently at rate equal to

the degree of the vertex it is placed at. Each offspring is then placed with uniform probability at any neighboring vertex. We stress that this occurs regardless of the number of previous offspring of the particle. Equivalently, each existing particle generates offspring at each neighboring vertex independently with rate 1. Note that the particles never move, and that each particle generates infinitely many children at each vertex adjacent to its location, including the location of its parent.

For a fixed G and a fixed location $v_0 \in G$ of the first particle, we let $Z(v,t)$ denote number of particles at vertex v at time t in the BTP and define

$$m(v,t) = \mathbb{E}Z(v,t). \tag{2.1}$$

One can observe that $\{Z(\cdot,t)\}_{t\geq 0}$ is a Markov process with the initial value $Z(v,0) = \delta_{v,v_0}$ and where, for each vertex v, the transition $\{Z(\cdot) \to Z(\cdot) + \delta_{\cdot,v}\}$ occurs at rate $\sum_{w\in N(v)} Z(w)$ where $N(v)$ denotes the neighborhood of v. It can be noted that, in [4], the BTP was formally defined as this Markov process. However, in our applications of the BTP, it is important that we are able to identify individual particles and discern their ancestry, which means that this representation is insufficient. We will return to the problem of defining a formal representation of the BTP in Section 3.

Below, we will use the terms *ancestor* and *descendant* of a particle to denote the natural partial order of particles generated by the BTP. For convenience, we will consider a particle to be both an ancestor and a descendant of itself. The terms *parent* and *child* are defined in the natural way. The location of a particle x is denoted by $v(x)$, and its birth time by $t(x)$. We define the *ancestral line* of a particle x as the ordered set

$$\mathrm{AL}(x) = \{x_0, x_1, \dots, x_l = x\} \tag{2.2}$$

of all ancestors of x (including x itself). If σ is the path obtained by following the locations of the vertices along the ancestral line of a particle x, then we say that the ancestral line of x *follows* σ, and we say that the ancestral line of x is *simple* if this path is simple. In certain parts of our proof, we will need to consider BTPs where the location of the initial particle can vary. In that case, we will refer to a BTP where the original particle is placed at v as the BTP *originating at* v.

As pointed out in [4], the BTP stochastically dominates Richardson's model in the sense that, for a common starting vertex v_0, the models can be coupled in such a way that $R(v,t) \leq Z(v,t)$ for all $v \in G$ and $t \geq 0$. This is clear from a comparison of the transition rates of Z and R. However, for our applications we need to consider this relation more closely. To this end, we imagine that we partition the particles in the BTP into two sets, which we call the set of *alive* particles and the set of *ghosts*. Which of these two sets a particle is placed in is decided at the time of its birth, and is then never changed. The original particle is placed in the set of alive particles. After this, whenever a new particle is born it is placed in the set of ghosts if its parent is a ghost or if its location is already

occupied by an alive particle, and placed in the set of alive particles otherwise. Clearly, the subprocess of the BTP consisting of all alive particles initially contains one particle, located at v_0, and it is straightforward to see that the rate at which alive particles are born at a given vertex v equals the number of adjacent vertices that contain alive particles if v does not currently contain an alive particle, and 0 if it does. As this is the same transition rate as for the corresponding transition in Richardson's model, we can consider Richardson's model as the subprocess of the BTP consisting of all alive particles. In a sense, for an observer not able to see the ghosts, the BTP will look like Richardson's model. Hence, with this coupling, the time at which a vertex gets infected is equal to one of the arrival times at the corresponding vertex in the full BTP, though not necessarily the first. We may here note that as at most one particle can be alive at each vertex, we can interpret $R(v,t)$ as the number of alive particles at v at time t.

A simplified version of the proof of the lower bound on T_n in [4] can now be summarized as follows: Consider a BTP on $\mathbb{Q}_n$ originating at $\hat{\mathbf{0}}$. Since the BTP dominates Richardson's model it suffices to show that with probability $1-o(1)$, no particle occupies $\hat{\mathbf{1}}$ at time $\ln(1+\sqrt{2})-\varepsilon$ for all $\varepsilon>0$ fixed. This is shown by a first moment method. It follows from standard methods in the theory of continuous-time Markov chains that $m(v,t)$, the expected number of particles at v at time t, is the unique solution to the initial value problem

$$\begin{aligned} \frac{d}{dt}m(v,t) &= \sum_{w\in N(v)} m(w,t), \qquad t>0, \\ m(v,0) &= \delta_{v,v_0}. \end{aligned} \tag{2.3}$$

In the case where $G=\mathbb{Q}_n$ and $v_0=\hat{\mathbf{0}}$, it is straightforward to check that the solution to (2.3) is

$$m(v,t) = (\sinh t)^{|v|}(\cosh t)^{n-|v|}. \tag{2.4}$$

Recall that $|v|$ denotes the number of ones in v. In particular, we have $m(\hat{\mathbf{1}},t) = (\sinh t)^n$. Clearly, this tends to 0 as $n\to\infty$ for any $t<\sinh^{-1}1=\ln(1+\sqrt{2})$, as desired.

We now turn to the upper bound for T_n. For particles x,z in a BTP, we say that z *contests* x if there exists a particle $y\in \mathrm{AL}(x)$ such that $v(y)=v(z)$ and $t(y)>t(z)$. Moreover, z contests x *with multiplicity* k if there are k such $y\in\mathrm{AL}(x)$. We say that a particle x in the BTP has index k, denoted by $I(x)=k$, if x is the kth particle to arrive at $v(x)$, and define the *degree of contestation of* x as

$$C(x) = \sum_{y\in \mathrm{AL}(x)} \big(I(y)-1\big). \tag{2.5}$$

It is easy to see that $C(x)$ gives the number of particles that contests x when the latter are counted with multiplicity. We say that x is *uncontested* if $C(x)=0$.

In analyzing the degree of contestation of a particle x, it is useful to partition the particles that contest x into two classes depending on whether or not they are ancestors of x. We consequently define

$$A(x) = \sum_{\substack{y,z \in \mathrm{AL}(x) \\ t(y) > t(z)}} \mathbb{1}_{v(y)=v(z)}, \tag{2.6}$$

the number of ancestors of x that contest it, and

$$B(x) = \sum_{\substack{y \in \mathrm{AL}(x), z \notin \mathrm{AL}(x) \\ t(y) > t(z)}} \mathbb{1}_{v(y)=v(z)}, \tag{2.7}$$

the number of other particles that contest x. Note that $A(x) + B(x) = C(x)$.

LEMMA 2.1. *We have the following properties*:

(i) $A(x) = 0$ *if and only if the ancestral line of* x *is simple*.

(ii) *If a particle is uncontested, then it is the first particle to be born at its location*.

(iii) *A particle* x *is a ghost if and only if an alive particle contests it. Hence, if a particle is uncontested, then it is alive*.

PROOF. (i) and (ii) are obvious from the definition of $A(x)$ and $C(x)$. (iii) For any ghost x in the BTP, there must exist an earliest ancestor y which is a ghost. As the original particle is alive, y must have a parent in the BTP. As the parent of y is alive but y is a ghost, the vertex occupied by y must already have been occupied by some alive particle z at the time of birth of y. Then z contests x. Conversely, if there is an alive particle z and $y \in \mathrm{AL}(x)$ such that $v(y) = v(z)$ and $t(y) > t(z)$, then y is a ghost, and hence so is x. $\square$

The third property is of particular interest as it allows us to express a lower bound on Richardson's model in terms of the BTP. Letting $Z_k(v,t)$ denote the number of particles x at vertex v at time t such that $C(x) = k$, we conclude that

$$Z_0 \overset{d}{\leq} \text{Richardson's model} \overset{d}{\leq} Z, \tag{2.8}$$

and with the proposed coupling between BTP and Richardson's model above we even have $Z_0 \leq R \leq Z$. However, it should be noted that, unlike Z and R, there is no reason why $Z_0(v)$ could not remain 0 forever. In fact, with the exception of the case where G is a chain of length 1, this occurs with positive probability. In order to see this, one can observe that if the first particle to arrive at a vertex is contested, which occurs with positive probability, then this particle will prevent all subsequent particles from being uncontested. On the other hand, in the event that $Z_0(v)$ is eventually nonzero, it follows from the second and third properties in Lemma 2.1 that the uncontested particle must have been the first particle at v and that this particle must have been alive. Hence, either $Z_0(v)$ remains 0 forever, or the time of the first arrival at v coincides in all three models.

2.1. *Outline of proof of Theorem 1.1.* For each vertex v and $t \geq 0$, we define

$$a(v,t) = \mathbb{E} \sum_{x \text{ at } v \text{ at time } t} A(x), \tag{2.9}$$

$$b(v,t) = \mathbb{E} \sum_{x \text{ at } v \text{ at time } t} B(x). \tag{2.10}$$

We similarly define

$$s(v,t) = \mathbb{E} \sum_{x \text{ at } v \text{ at time } t} \mathbb{1}_{A(x)=0}. \tag{2.11}$$

The core of finding upper bounds on the first-passage time using the BTP is the following theorem, which will be shown in Section 3.

THEOREM 2.2. *Let G be a finite connected simple graph. Consider the BTP on G originating at v_0, and let $Z_0(v,t)$, $b(v,t)$ and $s(v,t)$ be as above. Then, for any vertex v and $t \geq 0$ we have*

$$\mathbb{P}\big(Z_0(v,t) > 0\big) \geq s(v,t) \exp\left(-\frac{b(v,t)}{s(v,t)}\right). \tag{2.12}$$

In essence, Theorem 2.2 states that if, at a time t, the expected number of particles with simple ancestral line at v in the BTP is bounded away from 0, and if $b(v,t)$ is bounded, then with probability bounded away from 0 there is a particle at v at this time such that $A(x) = B(x) = 0$. Using the relation between the BTP and Richardson's model in (2.8), this immediately implies a lower bound on the probability that the first-passage time from v_0 to v in G is at most t. We remark that while the left-hand side of (2.12) certainly is increasing in t, the right-hand side is generally not.

We now apply this result to the hypercube. We let $G = \mathbb{Q}_n$, $v_0 = \hat{\mathbf{0}}$ and $t = \vartheta := \ln(1+\sqrt{2})$. In this case, the quantities $a(\hat{\mathbf{1}}, \vartheta)$ and $b(\hat{\mathbf{1}}, \vartheta)$ can be expressed analytically in a similar manner to the variance calculations for the BTP in [4]. This will be done in Section 4. The result of this can be summarized as follows.

PROPOSITION 2.3. *For $\vartheta = \ln(1+\sqrt{2})$, we have*

$$a(\hat{\mathbf{1}}, \vartheta) = \frac{\vartheta}{\sqrt{2}} + o(1) = 0.623 + \cdots + o(1), \tag{2.13}$$

$$b(\hat{\mathbf{1}}, \vartheta) = \vartheta + \frac{1}{3-2\sqrt{2}} + o(1) = 6.709 + \cdots + o(1). \tag{2.14}$$

In order to bound $s(\hat{\mathbf{1}}, \vartheta)$, we observe that $a(v,t)$ is an upper bound on the expected number of particles at v at time t whose ancestral lines are not simple.

This follows directly from the definition of $a(v,t)$ as $A(x)$ is an upper bound on the indicator function for the event that $A(x)$ is nonzero. We conclude that

$$m(v,t) - a(v,t) \le s(v,t) \le m(v,t), \tag{2.15}$$

and in particular, $1 - \frac{\vartheta}{\sqrt{2}} - o(1) = 0.376\cdots - o(1) \le s(\hat{\mathbf{1}}, \vartheta) \le 1$.

Plugging these values into Theorem 2.2, we conclude the following.

COROLLARY 2.4. *Let T_n denote the first-passage time from $\hat{\mathbf{0}}$ to $\hat{\mathbf{1}}$ in $\mathbb{Q}_n$ and let $\vartheta = \ln(1+\sqrt{2})$. There exists a constant $\varepsilon > 0$ such that $\mathbb{P}(T_n \le \vartheta) \ge \varepsilon$ for all n, and in particular* $\liminf_{n\to\infty} \mathbb{P}(T_n \le \vartheta) \ge 6.9 \cdot 10^{-9}$.

PROOF. The asymptotic lower bound on $\mathbb{P}(T_n \le \vartheta)$ is obtained directly from Theorem 2.2 and Proposition 2.3. From this, the uniform bound follows by the observation that $\mathbb{P}(T_n \le \vartheta)$ is nonzero for all n. $\square$

It may seem like Corollary 2.4 is far from our claimed result of convergence in L^p-norm, but given this result there are in fact a number of different ways to show that T_n converges to ϑ at least in probability. One could, for instance, apply the ideas of Bollobás and Kohayakawa in [2]. In this paper, we will instead employ a bootstrapping argument similar to one given in [5], which has the benefit of letting us get bounds on the L^p-norms of $T_n - \vartheta$ of optimal order. This will be shown in Section 5, completing the proof of Theorem 1.1.

3. Proof of Theorem 2.2. Before proceeding with the proof, we need to discuss the parametrization of the BTP more carefully. For a BTP originating at a vertex v, a particle is identified by a finite alternating sequence

$$\{e_1 t_1 e_2 t_2 \cdots e_k t_k\}, \tag{3.1}$$

of edges $e_1, e_2, \dots, e_k$ forming a path that starts at v, and positive real numbers $t_1, t_2, \dots, t_k$. The original particle is identified by $\{\}$, the empty sequence. For any other particle x, $e_1 e_2 \cdots e_k$ denotes the edges along the path followed by the ancestral line of x, and if $x_0, x_1, \dots, x_k = x$ are the ancestors of x in ascending order, then for each $1 \le i \le k$, we have $t_i = t(x_i) - t(x_{i-1})$. It is easy to see that such a sequence uniquely defines the location and birth time of x. Below will use $\oplus$ to denote concatenation of sequences. For a sequence a and a set of sequences B, we define $a \oplus B = \{a \oplus b \mid b \in B\}$.

By a BTP originating at a vertex v, we formally mean a random set of particles, which is interpreted as the set of all particles that will ever be born in the BTP, and, of course, whose distribution is given according to the transition rates as described above.

Because of the parametrization of the BTP described above, we now have a well-defined meaning of the event that a particle x exists in the BTP. Let $\mathbf{X}$ denote

a BTP originating at v_0. For any path σ with vertices $v_0, v_1, \dots, v_l$ and edges $\sigma_1, \sigma_2, \dots, \sigma_l$ and any $\mathbf{t} = (t_1, \dots, t_l) \in \mathbb{R}_+^l$, let

$$x(\sigma, \mathbf{t}) = \{\sigma_1, t_1, \sigma_2, t_2, \dots, \sigma_l, t_l\}. \tag{3.2}$$

Then $\{x(\sigma, \mathbf{t}) \in \mathbf{X}\}$ is the event that the original particle has given birth to a particle at v_1 at time t_1, which in turn has given birth to a particle at v_2 at time $t_1 + t_2$ and so on.

In treating the BTP, it turns out to be useful to condition on such events. A small caveat here is that it is a bit tricky to formally define the meaning of such a conditioning; it is tantamount to conditioning on an arrival in a Poisson process. It does however have an intuitive meaning as follows: Because of the independent increment property of the Poisson processes governing the birth times in the BTP, the conditioned process should evolve as usual at all times besides $t_1, t_1 + t_2, \dots,$ at which times the corresponding ancestor of $x(\sigma, \mathbf{t})$ is born with probability 1.

Let $\mathbf{X}_0, \mathbf{X}_1, \dots, \mathbf{X}_l$ be independent branching translation processes where $\mathbf{X}_i$ for $0 \le i \le l$ is a BTP originating at vertex v_i. Then the conditional distribution of $\mathbf{X}$ given $x(\sigma, \mathbf{t}) \in \mathbf{X}$ can be described formally as

$$\begin{aligned} \mathbf{X}(\sigma, \mathbf{t}) = \mathbf{X}_0 &\cup \big(\{\sigma_1 t_1\} \oplus \mathbf{X}_1\big) \cup \big(\{\sigma_1 t_1 \sigma_2 t_2\} \oplus \mathbf{X}_2\big) \\ &\cup \cdots \cup \big(\{\sigma_1 t_1 \sigma_2 t_2 \cdots \sigma_l t_l\} \oplus \mathbf{X}_l\big). \end{aligned} \tag{3.3}$$

That is, for each $0 \le i \le l$, $\{\sigma_1 t_1 \cdots \sigma_i t_i\} \oplus \mathbf{X}_i$ denotes the set consisting of the particle $\{\sigma_1 t_1 \cdots \sigma_i t_i\}$ and all descendants of it that are generated due to the usual transition rules for the BTP.

LEMMA 3.1. *Let σ be a path of length $l \ge 1$. We denote the vertices along the path by $v_0, \dots, v_l$ and the edges by $\sigma_1, \dots, \sigma_l$. Let $\mathbf{X}$ be a BTP originating at vertex v_0. Let f be a function taking pairs (X, x), X a realization of a BTP and x a particle in X, to nonnegative real numbers. Let $V_\sigma = V_\sigma(X)$ denote the set of particles at vertex v_l (no matter when they are born) whose ancestral line follows σ. Then we have*

$$\mathbb{E} \sum_{x \in V_\sigma(\mathbf{X})} f(\mathbf{X}, x) = \int_{s_1, s_2, \dots, s_l \ge 0} \mathbb{E} f\big(\mathbf{X}(\sigma, \mathbf{s}), x(\sigma, \mathbf{s})\big)\, d\mathbf{s}, \tag{3.4}$$

where $\mathbf{s} = (s_1, s_2, \dots, s_l)$ and $d\mathbf{s} = ds_1\, ds_2 \cdots ds_l$.

A proof of Lemma 3.1 is given in Appendix. As a comparison to this lemma it can be noted that if $\mathbf{T}$ is a random subset of a finite or even countable set, then the analogous statement

$$\mathbb{E} \sum_{t \in \mathbf{T}} f(\mathbf{T}, t) = \sum_t \mathbb{P}(t \in \mathbf{T}) \mathbb{E}\big[f(\mathbf{T}, t) \mid t \in \mathbf{T}\big] \tag{3.5}$$

is straightforward to show.

In the following lemma, we let φ denote any (appropriately measurable) indicator function taking as input a particle, x, encoded as a sequence of the form (3.1). This will be used later in the proof of Theorem 2.2, were the considered indicator functions will be specified.

LEMMA 3.2. *Let* $\mathbf{X}$ *be a BTP originating at a vertex* v, *and let* φ *be as above. If* $\varphi(\{\}) = 0$, *then*

$$\mathbb{P}\big(\varphi(x) = 0 \ \forall x \in \mathbf{X}\big) \geq \exp\Big(-\mathbb{E} \sum_{x \in \mathbf{X}} \varphi(x)\Big). \tag{3.6}$$

PROOF. For any particle $x \in \mathbf{X}$, let $\psi(\mathbf{X}, x)$ be the indicator function for the event that $\varphi(y) = 1$ for at least one descendant y of x. Clearly, we have

$$\sum_{x \text{ in gen } 1} \psi(\mathbf{X}, x) \leq \sum_{x \in \mathbf{X}} \varphi(x) \tag{3.7}$$

and

$$\sum_{x \text{ in gen } 1} \psi(\mathbf{X}, x) = 0 \quad \Leftrightarrow \quad \sum_{x \in \mathbf{X}} \varphi(x) = 0. \tag{3.8}$$

Let $\deg(v)$ denote the degree of the vertex v. Then the particles in generation one, that is, the children of the original particle, are born according to a Poisson process on $\mathbb{R}_+^{\deg(v)}$. Conditioned on these particles, the random variables $\psi(\mathbf{X}, x)$ for each such particle x are independent, and are one with probability only depending on the location and birth time of x. Hence, by the random selection property of a Poisson process, the particles in generation one that satisfy $\psi(\mathbf{X}, x) = 1$ are also born according to a Poisson process, and, in particular, the number of such particles is Poisson distributed. We conclude that

$$\begin{aligned} \mathbb{P}\big(\varphi(x) = 0 \ \forall x \in \mathbf{X}\big) &= \exp\Big(-\mathbb{E} \sum_{x \text{ in gen } 1} \psi(\mathbf{X}, x)\Big) \\ &\geq \exp\Big(-\mathbb{E} \sum_{x \in \mathbf{X}} \varphi(x)\Big). \end{aligned} \tag{3.9}$$

□

PROOF OF THEOREM 2.2. Denote the BTP originating at v_0 by $\mathbf{X}$. Before proceeding with the proof, we will revise our notation a bit. Strictly speaking, functions like $C(x)$ and $B(x)$ are functions not solely of x, but also of the realization of the BTP. Following the convention we have used earlier in this section, we will now denote these quantities by $C(\mathbf{X}, x)$ and $B(\mathbf{X}, x)$.

Recall that $Z_0(v, t)$ denotes the number of particles x at v at time t such that $C(\mathbf{X}, x) = 0$. As at most one particle at each vertex can be uncontested, we have

$\mathbb{P}(Z_0(v,t)>0)=\mathbb{E}Z_0(v,t)$. Furthermore, by Lemma 2.1, $C(\mathbf{X},x)=0$ if and only if $B(\mathbf{X},x)=0$ and the ancestral line of x is simple. By Lemma 3.1, we have

$$\begin{aligned}\mathbb{P}\big(Z_0(v,t)>0\big)&=\mathbb{E}\sum_{x\text{ at }v\text{ at time }t}\mathbb{1}_{C(\mathbf{X},x)=0}\\ &=\mathbb{E}\sum_{\sigma}\sum_{x\in V_\sigma(\mathbf{X})}\mathbb{1}_{t(x)\le t}\mathbb{1}_{\sigma\text{ simple}}\mathbb{1}_{B(\mathbf{X},x)=0}\\ &=\sum_{\sigma}\int_{s_1+\cdots+s_{|\sigma|}\le t}\mathbb{1}_{\sigma\text{ simple}}\mathbb{P}\big(B\big(\mathbf{X}(\sigma,\mathbf{s}),x(\sigma,\mathbf{s})\big)=0\big)\,d\mathbf{s},\end{aligned}\tag{3.10}$$

where σ goes over all paths from v_0 to v. Similarly, using Lemma 3.1 with $f_1(\mathbf{X},x)=\mathbb{1}_{t(x)\le t}\mathbb{1}_{\sigma\text{ simple}}$ and $f_2(\mathbf{X},x)=\mathbb{1}_{t(x)\le t}B(\mathbf{X},x)$, we get

$$s(v,t)=\sum_{\sigma}\int_{s_1+\cdots+s_{|\sigma|}\le t}\mathbb{1}_{\sigma\text{ simple}}\,d\mathbf{s}\tag{3.11}$$

and

$$b(v,t)=\sum_{\sigma}\int_{s_1+\cdots+s_{|\sigma|}\le t}\mathbb{E}B\big(\mathbf{X}(\sigma,\mathbf{s}),x(\sigma,\mathbf{s})\big)\,d\mathbf{s}.\tag{3.12}$$

Let σ be a simple path with vertices $v_0,v_1,\dots,v_{|\sigma|}=v$. Fix $s_1,\dots,s_{|\sigma|}>0$ such that $s_1+\cdots+s_{|\sigma|}\le t$ and consider the random variable $B(\mathbf{X}(\sigma,\mathbf{s}),x(\sigma,\mathbf{s}))$. As σ is simple, it follows that this quantity is equal to the number of particles $z\in\mathbf{X}(\sigma,\mathbf{s})$ that contest $x(\sigma,\mathbf{s})$, that is, such that $v(z)=v_i$ and $t(z)<\sum_{k=1}^{i}s_k$ for some $0\le i\le|\sigma|$. Let $\varphi(z)$ be the indicator function for z being such a particle, and for each $j=0,1,\dots,|\sigma|$ define $\varphi_j(z)=\varphi(\{\sigma_1s_1\cdots\sigma_js_j\}\oplus z)$. Then, by the definition of $\mathbf{X}(\sigma,\mathbf{s})$, we have

$$B\big(\mathbf{X}(\sigma,\mathbf{s}),x(\sigma,\mathbf{s})\big)=\sum_{z\in\mathbf{X}(\sigma,\mathbf{s})}\varphi(z)=\sum_{j=0}^{|\sigma|}\sum_{z\in\mathbf{X}_j}\varphi_j(z).\tag{3.13}$$

It follows by Lemma 3.2 that

$$\begin{aligned}\mathbb{P}\big(B\big(\mathbf{X}(\sigma,\mathbf{s}),x(\sigma,\mathbf{s})\big)=0\big)&=\prod_{j=0}^{|\sigma|}\mathbb{P}\big(\varphi_j(x)=0\ \forall x\in\mathbf{X}_j\big)\\ &\ge\prod_{j=0}^{|\sigma|}\exp\Big(-\mathbb{E}\sum_{x\in\mathbf{X}_j}\varphi_j(x)\Big)\\ &=\exp\big(-\mathbb{E}B\big(\mathbf{X}(\sigma,\mathbf{s}),x(\sigma,\mathbf{s})\big)\big).\end{aligned}\tag{3.14}$$

By convexity of the exponential function, for any $r,r_0\in\mathbb{R}$ we have $e^r\ge e^{r_0}(1-r_0)+e^{r_0}r$. Combining this inequality with (3.14), it follows that for any simple σ we have

$$\mathbb{P}\big(B\big(\mathbf{X}(\sigma,\mathbf{s}),x(\sigma,\mathbf{s})\big)=0\big)\ge e^{r_0}(1-r_0)-e^{r_0}\mathbb{E}B\big(\mathbf{X}(\sigma,\mathbf{s}),x(\sigma,\mathbf{s})\big)\tag{3.15}$$

and hence the inequality

$$\begin{aligned}\mathbb{1}_{\sigma\text{ simple}}\mathbb{P}\big(B\big(\mathbf{X}(\sigma,\mathbf{s}),x(\sigma,\mathbf{s})\big)=0\big)\\ \geq e^{r_0}(1-r_0)\mathbb{1}_{\sigma\text{ simple}}-e^{r_0}\mathbb{E}B\big(\mathbf{X}(\sigma,\mathbf{s}),x(\sigma,\mathbf{s})\big)\end{aligned}\tag{3.16}$$

holds for any path σ from v_0 to v and any $r_0\in\mathbb{R}$. By integrating both sides of (3.16) over $s_1,\dots,s_{|\sigma|}\geq 0$ such that $s_1+\cdots+s_{|\sigma|}\leq t$, summing over all paths from v_0 to v and comparing to (3.10), (3.11) and (3.12) we conclude that for any $r_0\in\mathbb{R}$ we have

$$\mathbb{P}\big(Z_0(v,t)>0\big)\geq e^{r_0}(1-r_0)s(v,t)-e^{r_0}b(v,t).\tag{3.17}$$

It is easy to verify that the right-hand side is maximized by $r_0=-\frac{b(v,t)}{s(v,t)}$, which yields the inequality $\mathbb{P}(Z_0(v,t)>0)\geq s(v,t)\exp(-\frac{b(v,t)}{s(v,t)})$ as desired. □

4. Proof of Proposition 2.3. Throughout this section, we assume that the underlying graph in the BTP is $\mathbb{Q}_n$, and, unless stated otherwise, the BTP is assumed to originate at $\hat{\mathbf{0}}$. We will accordingly let $m(v,t)$ denote the expected number of particles at v at time t for a BTP originating at $\hat{\mathbf{0}}$, as given by (2.4). In order to simplify notation, we will interpret the vertices of $\mathbb{Q}_n$ as the elements of the additive group $\mathbb{Z}_2^n$, the n-fold group product of $\mathbb{Z}_2$, and we let $e_1,e_2,\dots,e_n\in\mathbb{Z}_2^n$ denote the standard basis. Hence, adding e_i to a vertex v corresponds to flipping its ith coordinate. We note that for any fixed vertex $w\in\mathbb{Q}_n$, the map $v\mapsto v-w$ is a graph isomorphism taking w to $\hat{\mathbf{0}}$. Hence, for a BTP originating at w, the expected number of particles at v at time t is given by $m(v-w,t)$. While addition and subtraction are equivalent in $\mathbb{Z}_2^n$, we will sometimes make a formal distinction between them in order to indicate direction.

LEMMA 4.1. *For any $t>0$ and $v\in\mathbb{Q}_n$, we have*

$$\frac{d^2}{dt^2}m(v,t)=\sum_{i=1}^n\sum_{j=1}^n m(v+e_i+e_j,t)\tag{4.1}$$

and

$$\begin{aligned}\frac{1}{2}\frac{d^2}{dt^2}m(v,t)^2\\ =\sum_{i=1}^n\sum_{j=1}^n m(v+e_i+e_j,t)m(v,t)+m(v+e_i,t)m(v+e_j,t).\end{aligned}\tag{4.2}$$

PROOF. Recall that, by (2.3), $m(v,t)$ satisfies

$$\frac{d}{dt}m(v,t)=\sum_{i=1}^n m(v+e_i,t).\tag{4.3}$$

This directly implies that

$$\begin{aligned}\frac{d^2}{dt^2}m(v,t) &= \frac{d}{dt}\sum_{i=1}^{n} m(v+e_i,t)\\ &= \sum_{i=1}^{n}\frac{d}{dt}m(v+e_i,t)\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} m(v+e_i+e_j,t).\end{aligned}$$

The second equation now follows from $\frac{1}{2}\frac{d^2}{dt^2}m(v,t)^2 = m''(v,t)m(v,t) + m'(v,t)^2$. □

LEMMA 4.2. *Let $s,t \geq 0$ and $v \in \mathbb{Q}_n$. Then*

$$\sum_{w\in\mathbb{Q}_n} m(w,s)m(v+w,t) = m(v,s+t). \tag{4.4}$$

PROOF. If we condition on the state of the BTP at time s, then at subsequent times, the process can be described as a superposition of independent branching processes, originating from each particle alive at time s. For each such process originating from a particle at vertex w, we have, by symmetry of $\mathbb{Q}_n$, that the expected number of particles at vertex v at time $t+s$ is $m(v+w,t)$. Hence,

$$\mathbb{E}\big[Z(v,s+t) \mid Z(\cdot,s)\big] = \sum_{w\in\mathbb{Q}_n} Z(w,s)m(v+w,t). \tag{4.5}$$

Recall that $Z(v,t)$ denotes the number of particles at vertex v time t. The lemma follows by taking the expected value of this expression. □

In the following proposition, we derive explicit expressions for $a(\hat{\mathbf{1}},u)$ and $b(\hat{\mathbf{1}},u)$ in terms of $m(v,t)$. Our argument is reminiscent of the second moment calculation for $Z(\hat{\mathbf{1}},u)$ by Durrett in [4].

PROPOSITION 4.3. *For any $u > 0$, we have*

$$\begin{aligned}a(\hat{\mathbf{1}},u) = \sum_{v\in\mathbb{Q}_n}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_0^u\int_0^{u-t} m(v,s)\\ \times m(\hat{\mathbf{1}}-v,u-s-t)m(e_j+e_i,t)\,ds\,dt,\end{aligned} \tag{4.6}$$

$$\begin{aligned}a(\hat{\mathbf{1}},u)+b(\hat{\mathbf{1}},u) = \sum_{v\in\mathbb{Q}_n}\sum_{w\in\mathbb{Q}_n}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_0^u\int_0^{u-t} m(v,s)\\ \times m(\hat{\mathbf{1}}-w,u-s-t)\big(m(w-v,t)m(w-v-e_i+e_j,t)\\ +m(w-v-e_i,t)m(w-v+e_j,t)\big)\,ds\,dt.\end{aligned} \tag{4.7}$$

PROOF. Fix $u > 0$ and let $\mathbf{X}$ be a BTP on $\mathbb{Q}_n$ originating at $\hat{\mathbf{0}}$. Let $\mathcal{T}$ denote the set of triples (x, y, z) of particles in $\mathbf{X}$ such that:

- x is located at $\hat{\mathbf{1}}$ at time u,
- $y \in \mathrm{AL}(x)$,
- $v(y) = v(z)$,
- $t(y) > t(z)$.

We furthermore partition this set into $\mathcal{T}_A$, the set of all such triples where $z \in \mathrm{AL}(x)$, and $\mathcal{T}_B$, the set of all such triples where $z \notin \mathrm{AL}(x)$. For any x at $\hat{\mathbf{1}}$ at time u in $\mathbf{X}$, it is clear that $C(x)$ gives the number of triples in $\mathcal{T}$ where the first element is x. Hence,

$$|\mathcal{T}| = \sum_{x \text{ at } v \text{ at time } u} C(x). \tag{4.8}$$

Similarly, we see that

$$|\mathcal{T}_A| = \sum_{x \text{ at } v \text{ at time } u} A(x), \tag{4.9}$$

$$|\mathcal{T}_B| = \sum_{x \text{ at } v \text{ at time } u} B(x). \tag{4.10}$$

Hence, $a(\hat{\mathbf{1}}, u) = \mathbb{E}|\mathcal{T}_A|$, $b(\hat{\mathbf{1}}, u) = \mathbb{E}|\mathcal{T}_B|$ and $a(\hat{\mathbf{1}}, u) + b(\hat{\mathbf{1}}, u) = \mathbb{E}|\mathcal{T}|$.

Let us start by deriving the expression for $a(\hat{\mathbf{1}}, u)$. For any $(x, y, z) \in \mathcal{T}_A$ there are well-defined particles $c(x, z)$, the particle subsequent to z in the ancestral line of x, and $p(y)$, the parent of y. We note that y is not a child of z as then y and z would not be located at the same vertex, hence $c(x, z)$ must be an ancestor of $p(y)$. This means that the for each $(x, y, z) \in \mathcal{T}_A$, the particles $(x, y, z, c(x, z), p(y))$ must be related as illustrated in Graph 1 of Figure 1.

Each such quintuple can be described in the following way: A particle z at a vertex v gives birth to a particle c at $v + e_i$ at time s. A descendant p of c at $v - e_j$ gives birth to a particle y at v at time $s + t$, which then in turn has a descendant x at $\hat{\mathbf{1}}$ at time u. For a fixed vertex v, fixed $1 \le i, j \le n$, and fixed infinitesimal time intervals $(s, s + ds]$ and $(s + t, s + t + dt]$ where $0 \le s < s + t < u$, the expected number of such occurrences is

$$m(v, s)\, ds\, m(-e_j - e_i, t)\, dt\, m(v, u - s - t). \tag{4.11}$$

Equation (4.6) follows by integrating over all $s, t > 0$ such that $s + t < u$ and summing over all $v \in \mathbb{Q}_n$ and all $1 \le i, j \le n$.

We now turn to the formula for $a(\hat{\mathbf{1}}, u) + b(\hat{\mathbf{1}}, u)$. For each triple $(x, y, z) \in \mathcal{T}$, we define the following particles: $l(x, z)$, the last common ancestor of x and z, $c(x, z)$, the first particle which is an ancestor of exactly one of x and z, and $p(y)$, the parent of y. Note that $c(x, z)$ must be a child of $l(x, z)$. Similar to the case of $\mathcal{T}_A$, we note that $c(x, z) \neq y$. In order to see this, assume that equality holds. As

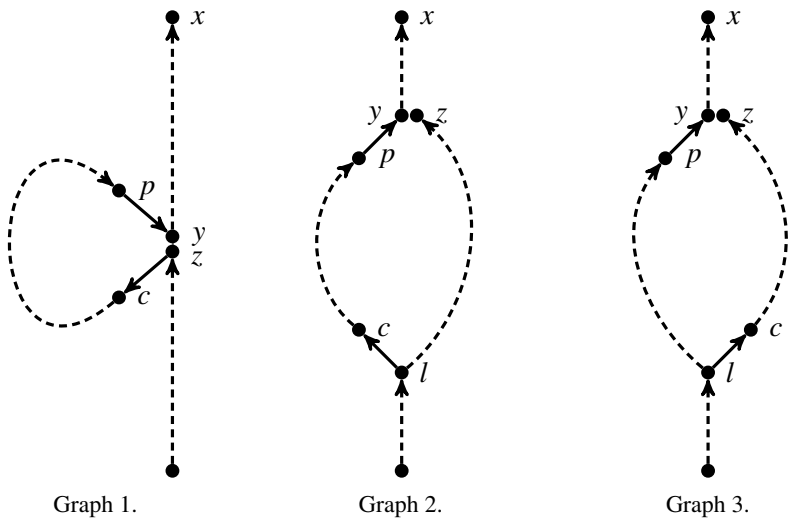

FIG. 1. *Illustration of the possible configurations of ancestral lines of elements in $\mathcal{T}_A$, as shown in Graph* 1, *and $\mathcal{T}$, as shown in Graphs* 2 *and* 3, *respectively. In all three graphs, $v(y) = v(z)$, $v(x) = \hat{\mathbf{1}}$ and $t(x) \le u$. A dashed respectively filled arrow from x_1 to x_2 means that x_2 is a descendant respectively child of x_1. The original particle in the BTP is drawn at the bottom of each graph. For Graphs* 2 *and* 3, *as $c(x, z)$ is the first particle to be an ancestor of exactly one of x and z, every particle after $l(x, z)$ in the ancestral line of z in Graph* 2 *and every particle after $l(x, z)$ in the ancestral line of $p(y)$ in Graph* 3 *is born after $c(x, z)$.*

$c(x, z)$ is the first particle to be an ancestor of precisely one of x and z, but z is older than y it follows that z must be an ancestor of x, and hence $l(x, z) = z$. But then, $y = c(x, z)$ and $z = l(x, z)$, which again means that y and z are located at adjacent vertices.

Here, we need to consider two cases depending on whether $c(x, z)$ is an ancestor of x or of z. In the former case, as $c(x, z) \neq y$, $c(x, z)$ must be an ancestor of $p(y)$ and so the particles $x, y, z, l(x, z), c(x, z)$ and $p(y)$ must be related as illustrated in Graph 2 of Figure 1. Similarly, it is clear that in the latter case, the particles must be related as illustrated in Graph 3 in Figure 1. Proceeding in a similar manner as for $\mathcal{T}_A$, we see that the expected number of triples in $\mathcal{T}$ corresponding to the former case and such that $v(l) = v$, $v(c(x, z)) = v + e_i$, $v(p(y)) = w - e_j$, $v(y) = v(z) = w$, where $c(x, z)$ is born during $(s, s + ds]$ and y during $(s + t, s + t + dt]$ is

$$m(v, s)\, ds\, m(w - v, t) m(w - e_j - v - e_i, t)\, dt\, m(\hat{\mathbf{1}} - w, u - s - t). \tag{4.12}$$

Analogously, for the latter case we get

$$m(v, s)\, ds\, m(w - e_j - v, t) m(w - v - e_i, t)\, dt\, m(\hat{\mathbf{1}} - w, u - s - t). \tag{4.13}$$

The proposition follows by summing these expressions over all $v, w \in \mathbb{Q}_n$, all $1 \le i, j \le n$ and integrating over all $s, t > 0$ such that $s + t < u$. $\square$

REMARK 4.4. In the proof of Proposition 4.3, the only crucial property of the underlying graph is that it should not contain loops. This is to guarantee that every particle has a different location than its parent. Hence, the proposition can directly be generalized to any simple graph by replacing the sums over i and j by sums over the corresponding neighborhoods.

PROPOSITION 4.5. *For $\vartheta = \ln(1+\sqrt{2})$, we have $a(\hat{\mathbf{1}}, \vartheta) \to \frac{\vartheta}{\sqrt{2}}$ as $n \to \infty$.*

PROOF. By reordering the sums and integrals in (4.6), we have

$$a(\hat{\mathbf{1}}, \vartheta) = \int_0^{\vartheta} \int_0^{\vartheta - t} \sum_{v \in \mathbb{Q}_n} m(v, s) m(\hat{\mathbf{1}} - v, \vartheta - s - t) \times \sum_{i=1}^{n} \sum_{j=1}^{n} m(e_j - e_i, t)\, ds\, dt. \tag{4.14}$$

Applying Lemmas 4.1 and 4.2, the right-hand side simplifies to

$$\begin{aligned} &\int_0^{\vartheta} \int_0^{\vartheta - t} m(\hat{\mathbf{1}}, \vartheta - t) \frac{d^2}{dt^2} m(\hat{\mathbf{0}}, t)\, ds\, dt \\ &\quad = \int_0^{\vartheta} (\vartheta - t) m(\hat{\mathbf{1}}, \vartheta - t) \frac{d^2}{dt^2} m(\hat{\mathbf{0}}, t)\, dt, \end{aligned} \tag{4.15}$$

and by plugging in the analytical formula (2.4) for $m(v, t)$ we get

$$\begin{aligned} a(\hat{\mathbf{1}}, \vartheta) &= \int_0^{\vartheta} (\vartheta - t)\big(\sinh(\vartheta - t)\big)^n \frac{d^2}{dt^2} (\cosh t)^n\, dt \\ &= \int_0^{\vartheta} (\vartheta - t)\big(\sinh(\vartheta - t)\big)^n \big(n + n(n-1)(\tanh t)^2\big)(\cosh t)^n\, dt \\ &= \int_0^{\vartheta} (\vartheta - t)\big(n + n(n-1)(\tanh t)^2\big) e^{nf(t)}\, dt, \end{aligned} \tag{4.16}$$

where $f(t) := \ln(\sinh(\vartheta - t)\cosh t)$.

What follows is a textbook application of the Lebesgue dominated convergence theorem. We begin examining the function f. The first and second derivatives of f are given by

$$f'(t) = -\coth(\vartheta - t) + \tanh t, \tag{4.17}$$

$$f''(t) = -\operatorname{csch}(\vartheta - t)^2 + \operatorname{sech}(t)^2. \tag{4.18}$$

As $\operatorname{sech} t \le 1$ for all $t \in \mathbb{R}$ and $\operatorname{csch} t \ge 1$ for $0 < t < \vartheta$, it follows that $f''(t) < 0$ for $0 < t < \vartheta$. Hence, f is concave in this interval, so in particular $f(t) \le f(0) + f'(0)t = -\sqrt{2}t$. Furthermore, we have $\tanh t \le Ct$ for some appropriate $C > 0$.

Substituting t by $s = nt$ in (4.16), we obtain

$$a(\hat{\mathbf{1}}, \vartheta) = \int_0^\infty \mathbb{1}_{s \le n\vartheta}\left(\vartheta - \frac{s}{n}\right)\left(1 + (n-1)\tanh\left(\frac{s}{n}\right)^2\right)e^{nf(s/n)}\,ds. \tag{4.19}$$

It is clear that the integrand is bounded for all n by $\vartheta(1 + Cs^2)e^{-\sqrt{2}s}$, which is integrable over $[0, \infty)$. Hence, by dominated convergence, it follows that

$$a(\hat{\mathbf{1}}, \vartheta) \to \int_0^\infty \vartheta e^{-\sqrt{2}s}\,ds = \frac{\vartheta}{\sqrt{2}} \qquad \text{as } n \to \infty. \tag{4.20}$$

□

PROPOSITION 4.6. *For $\vartheta = \ln(1 + \sqrt{2})$, we have*

$$a(\hat{\mathbf{1}}, \vartheta) + b(\hat{\mathbf{1}}, \vartheta) \to \frac{\vartheta e^{\vartheta}}{\sqrt{2}} + \frac{1}{3 - 2\sqrt{2}} \qquad \text{as } n \to \infty. \tag{4.21}$$

Hence, as $n \to \infty$ we have $b(\hat{\mathbf{1}}, \vartheta) \to \vartheta + \frac{1}{3-2\sqrt{2}}$.

PROOF. By reordering the sums in (4.7) and applying Lemma 4.1, we see that $a(\hat{\mathbf{1}}, \vartheta) + b(\hat{\mathbf{1}}, \vartheta)$ can be expressed as

$$\frac{1}{2}\sum_{v,w \in \mathbb{Q}_n} \int_0^\vartheta \int_0^{\vartheta - t} m(v, s)m(\hat{\mathbf{1}} - w, \vartheta - s - t)\frac{d^2}{dt^2}m(w - v, t)^2\,ds\,dt. \tag{4.22}$$

Letting $\Delta = w - v$, this sum can be rewritten as

$$\frac{1}{2}\sum_{v,\Delta \in \mathbb{Q}_n} \int_0^\vartheta \int_0^{\vartheta - t} m(v, s)m(\hat{\mathbf{1}} - \Delta + v, \vartheta - s - t)\frac{d^2}{dt^2}m(\Delta, t)^2\,ds\,dt, \tag{4.23}$$

which by Lemma 4.2 simplifies to

$$\frac{1}{2}\int_0^\vartheta (\vartheta - t)\sum_{\Delta \in \mathbb{Q}_n} m(\hat{\mathbf{1}} - \Delta, \vartheta - t)\frac{d^2}{dt^2}m(\Delta, t)^2\,dt. \tag{4.24}$$

To evaluate the sum in the above integral, we use a small trick. Let us replace $\vartheta - t$ in this sum by τ which we consider as a variable not depending on t. Then

$$\sum_{\Delta \in \mathbb{Q}_n} m(\hat{\mathbf{1}} - \Delta, \tau)\frac{d^2}{dt^2}m(\Delta, t)^2 = \frac{\partial^2}{\partial t^2}\sum_{\Delta \in \mathbb{Q}_n} m(\hat{\mathbf{1}} - \Delta, \tau)m(\Delta, t)^2.$$

By grouping all terms with $|\Delta| = k$, we get

$$\begin{aligned}\sum_{\Delta \in \mathbb{Q}_n} & m(\hat{\mathbf{1}} - \Delta, \tau)m(\Delta, t)^2 \\ &= \sum_{k=0}^{n}\binom{n}{k}(\sinh\tau)^k(\cosh\tau)^{n-k}(\sinh t)^{2n-2k}(\cosh t)^{2k}\end{aligned}$$

$$= \sum_{k=0}^{n} \binom{n}{k} \big(\sinh\tau(\cosh t)^2\big)^k \big(\cosh\tau(\sinh t)^2\big)^{n-k}$$
$$= \big(\sinh\tau(\cosh t)^2 + \cosh\tau(\sinh t)^2\big)^n$$
$$= \left(\frac{1}{2}e^{\tau}\cosh 2t - \frac{1}{2}e^{-\tau}\right)^n.$$

Note that $\frac{1}{2}e^{\tau}\cosh 2t - \frac{1}{2}e^{-\tau} > 0$ for any $t, \tau \geq 0$. Hence,

$$\sum_{\Delta\in\mathbb{Q}_n} m(\hat{\mathbf{1}} - \Delta, \tau)\frac{d^2}{dt^2}m(\Delta, t)^2$$
$$= \frac{\partial^2}{\partial t^2}\left(\frac{1}{2}e^{\tau}\cosh 2t - \frac{1}{2}e^{-\tau}\right)^n$$
$$= 2ne^{\tau}\cosh t\left(\frac{1}{2}e^{\tau}\cosh 2t - \frac{1}{2}e^{-\tau}\right)^{n-1}$$
$$+ n(n-1)e^{2\tau}(\sinh 2t)^2\left(\frac{1}{2}e^{\tau}\cosh 2t - \frac{1}{2}e^{-\tau}\right)^{n-2}.$$

Letting

$$f(t) = \ln\big(\tfrac{1}{2}e^{\vartheta - t}\cosh 2t - \tfrac{1}{2}e^{-\vartheta + t}\big), \tag{4.25}$$
$$g(t) = 2e^{\vartheta - t}\cosh t\, e^{-f(t)}, \tag{4.26}$$
$$h(t) = e^{2\vartheta - 2t}(\sinh t)^2 e^{-2f(t)} \tag{4.27}$$

we can write

$$a(\hat{\mathbf{1}}, \vartheta) + b(\hat{\mathbf{1}}, \vartheta) = \frac{1}{2}\int_0^{\vartheta}(\vartheta - t)\big(ng(t) + n(n-1)h(t)\big)e^{nf(t)}\,dt. \tag{4.28}$$

One can check that $f(0) = f(\vartheta) = 0$, $f(\frac{1}{2}) < -\frac{1}{5}$, and that f has derivatives

$$f'(t) = -1 + 2\frac{\sinh 2t - e^{-2\vartheta + 2t}}{\cosh 2t - e^{-2\vartheta + 2t}} \tag{4.29}$$

and

$$f''(t) = 4\frac{1 - 2e^{-2\vartheta}}{(\cosh 2t - e^{-2\vartheta + 2t})^2}. \tag{4.30}$$

Note that $\frac{1}{2}e^{\vartheta - t}\cosh 2t - \frac{1}{2}e^{-\vartheta + t} = \sinh(\vartheta - t)(\cosh t)^2 + \cosh(\vartheta - t)(\sinh t)^2 > 0$ for $t \in [0, \vartheta]$. Hence, it follows that $f(t)$ is convex. Furthermore, for $0 \leq t \leq \vartheta$, $g(t)$ and $h(t)$ are nonnegative bounded functions and $h(t) = O(t^2)$.

To evaluate the integral in equation (4.28), we divide it into two integrals, one over the interval $[0, \frac{1}{2}]$, and one over $[\frac{1}{2}, \vartheta]$:

$$\begin{aligned}\int_0^{1/2} &(\vartheta - t)\big(ng(t) + n(n-1)h(t)\big)e^{nf(t)}\,dt \\ &= \int_0^{n/2}\left(\vartheta - \frac{s}{n}\right)\left(g\left(\frac{s}{n}\right) + (n-1)h\left(\frac{s}{n}\right)\right)e^{nf(s/n)}\,ds\end{aligned} \tag{4.31}$$

and

$$\begin{aligned}\int_{1/2}^{\vartheta} &(\vartheta - t)\big(ng(t) + n(n-1)h(t)\big)e^{nf(t)}\,dt \\ &= \int_0^{(\vartheta-1/2)n} s\left(\frac{1}{n}g\left(\vartheta - \frac{s}{n}\right) + \frac{n-1}{n}h\left(\vartheta - \frac{s}{n}\right)\right)e^{nf(\vartheta - s/n)}\,ds.\end{aligned} \tag{4.32}$$

Now, using the convexity of $f(t)$ it is a standard calculation to show that the integrands of these expressions are uniformly dominated by $C(1+s^2)e^{-\lambda s}$ and $Cse^{-\lambda s}$, respectively, for appropriate positive constants λ and C. Hence, by the Lebesgue dominated convergence theorem, these integrals converge to

$$\int_0^{\infty} 2\vartheta e^{\vartheta + f'(0)s}\,ds = \frac{2\vartheta e^{\vartheta}}{-f'(0)} = \sqrt{2}\vartheta e^{\vartheta} \tag{4.33}$$

and

$$\int_0^{\infty} s(\sinh 2\vartheta)^2 e^{-f'(\vartheta)s}\,ds = \frac{8}{f'(\vartheta)^2} = \frac{2}{3-2\sqrt{2}} \tag{4.34}$$

respectively, as $n \to \infty$. We conclude that

$$a(\hat{\mathbf{1}}, \vartheta) + b(\hat{\mathbf{1}}, \vartheta) \to \frac{1}{2}\left(\sqrt{2}\vartheta e^{\vartheta} + \frac{2}{3-2\sqrt{2}}\right) \qquad \text{as } n \to \infty. \tag{4.35}$$

□

5. Proof of Theorem 1.1. In order to bound $\|T_n - \vartheta\|_p$, it is natural to treat the problems of bounding $T_n - \vartheta$ from above and below separately. To this end, we let $(T_n - \vartheta)_+$ denote the maximum of $T_n - \vartheta$ and 0, and let $(T_n - \vartheta)_-$ the maximum of $\vartheta - T_n$ and 0. We will begin by proving two simple propositions. The first shows that the variance of T_n and the L^p-norm of $T_n - \vartheta$ for any $1 \le p < \infty$ are $\Omega(\frac{1}{n})$. The second proposition uses the lower bound on T_n obtained by Durrett to prove that $\|(T_n - \vartheta)_-\|_p = O(\frac{1}{n})$. The remaining part of the section will be dedicated to bounding $\|(T_n - \vartheta)_+\|_p$.

PROPOSITION 5.1. *T_n has fluctuations of order at least $\frac{1}{n}$.*

PROOF. We can write T_n in terms of Richardson's model as the time until the first neighbor of $\hat{\mathbf{0}}$ gets infected plus the time from this event until $\hat{\mathbf{1}}$ gets infected. It is easy to see that these are independent, and the former is exponentially distributed with mean $\frac{1}{n}$. □

PROPOSITION 5.2. *Let $1 \le p < \infty$ be fixed. Then $\|(T_n - \vartheta)_-\|_p = O(\frac{1}{n})$.*

PROOF. We have

$$\begin{aligned}\mathbb{E}\big[((T_n - \vartheta)_-)^p\big] &= \mathbb{E}\int_0^\infty \mathbb{1}_{t \le (T_n - \vartheta)_-}\, pt^{p-1}\, dt \\ &= \int_0^\infty pt^{p-1}\mathbb{P}(T_n \le \vartheta - t)\, dt.\end{aligned} \tag{5.1}$$

To bound this, we use that $\mathbb{P}(T_n \le \vartheta - t) \le m(\hat{\mathbf{1}}, \vartheta - t) = (\sinh(\vartheta - t))^n$ for any $t \le \vartheta$ and $\mathbb{P}(T_n \le \vartheta - t) = 0$ for $t > \vartheta$ (naturally T_n is always nonnegative). It is straightforward to show that $\sinh(\vartheta - t) \le e^{-\sqrt{2}t}$ for any $t \ge 0$. Using this, we conclude that

$$\mathbb{E}\big[((T_n - \vartheta)_-)^p\big] \le \int_0^\vartheta pt^{p-1}e^{-\sqrt{2}nt}\, dt = O\left(\frac{1}{n^p}\right). \tag{5.2}$$

□

We now turn to the upper bound on T_n. Assume $n \ge 4$. Let $\{W_e\}_{e \in E(\mathbb{Q}_n)}$ be a collection of independent exponentially distributed random variables with expected value 1, denoting the passage times of the edges in $\mathbb{Q}_n$. For any vertex v adjacent to $\hat{\mathbf{0}}$, we will use W_v to denote the passage time of the edge between $\hat{\mathbf{0}}$ and v. Similarly, for any v adjacent to $\hat{\mathbf{1}}$, W_v denotes the passage time of the edge between v and $\hat{\mathbf{1}}$.

Condition on the weights of all edges connected to either $\hat{\mathbf{0}}$ or $\hat{\mathbf{1}}$. We pick vertices v_1 and v_2 adjacent to $\hat{\mathbf{0}}$ such that W_{v_1} and W_{v_2} have the smallest and second smallest edge weights respectively among all edges adjacent to $\hat{\mathbf{0}}$. Among all $n-2$ neighboring vertices of $\hat{\mathbf{1}}$ that are not antipodal to v_1 or v_2 we then pick w_1 and w_2 such that W_{w_1} and W_{w_2} have the smallest and second smallest values. Then W_{v_1}, $W_{v_2} - W_{v_1}$, W_{w_1} and $W_{w_2} - W_{w_1}$ are independent exponentially distributed random variables with respective expected values $\frac{1}{n}$, $\frac{1}{n-1}$, $\frac{1}{n-2}$ and $\frac{1}{n-3}$.

As v_1 and v_2 are adjacent to $\hat{\mathbf{0}}$ and w_1 and w_2 are adjacent to $\hat{\mathbf{1}}$, there is exactly one coordinate in each of v_1 and v_2 which is 1, and exactly one coordinate in w_1 and w_2 which is 0. Let the locations of these coordinates in v_1, v_2, w_1 and w_2 be denoted by i_1, i_2, j_1 and j_2, respectively. Note that the requirement on v_1, v_2, w_1 and w_2 not to be antipodal means that i_1, i_2, j_1 and j_2 are all distinct. For $k \in \{1, 2\}$, we define H_k as the induced subgraph of $\mathbb{Q}_n$ consisting of all vertices $v \in \mathbb{Q}_n$ such that the i_kth coordinate is 1 and the j_kth coordinate is 0. We furthermore define H_2' as the induced subgraph of $\mathbb{Q}_n$ whose vertex set is given by $H_2 \setminus H_1$. Note that H_1 and H_2' are vertex disjoint, and hence also edge disjoint.

The idea here is to bound T_n in terms of the minimum of the first-passage time from v_1 to w_1 in H_1 and the first-passage time from v_2 to w_2 in H_2', where the passage times for the edges are taken from $\{W_e\}_{e \in E(\mathbb{Q}_n)}$. As H_1 and H_2 are both isomorphic to $\mathbb{Q}_{n-2}$, where v_1 and w_1 are antipodal in H_1 and v_2 and w_2 are

antipodal in H_2, Corollary 2.4 implies that the corresponding first-passage times in H_1 and H_2 are at most ϑ with probability bounded away from 0. For our proof, we will make use of the slightly stronger statement that the same holds true for H_2', as stated in the following proposition. We postpone the proof of this to the end of the section.

PROPOSITION 5.3. *There exists a constant $\varepsilon_2 > 0$ such that for all $n \geq 4$, with probability at least ε_2 the first-passage time in H_2' from v_2 to w_2 is at most ϑ.*

Now, let ξ denote the indicator function for the event that the first-passage time from v_2 to w_2 in H_2' is at most ϑ. As H_1 is isomorphic to $\mathbb{Q}_{n-2}$ it is clear that the first-passage time from v_1 to w_1 in H_1 is distributed as T_{n-2}, and so we may couple T_{n-2} to $\{W_e\}_{e\in E(\mathbb{Q}_n)}$ such that T_{n-2} denotes this quantity. Note that this means that ξ and T_{n-2} are independent random variables. With this coupling it is clear that $T_n \leq W_{v_1} + W_{w_1} + T_{n-2}$. Furthermore, if $\xi = 1$ we similarly see that $T_n \leq W_{v_2} + W_{w_2} + \vartheta$. Combining these bounds we see that for any $n \geq 4$ we have

$$T_n \leq \xi(W_{v_2} + W_{w_2} + \vartheta) + (1-\xi)(W_{v_1} + W_{v_1} + T_{n-2}). \tag{5.3}$$

We now employ (5.3) to bound the L^p-norm of $(T_n - \vartheta)_+$. By subtracting ϑ and taking the positive part of both sides, we get

$$(T_n - \vartheta)_+ \leq \xi(W_{v_2} + W_{w_2}) + (1-\xi)\big(W_{v_1} + W_{w_1} + (T_{n-2} - \vartheta)_+\big). \tag{5.4}$$

As $W_{v_2} \geq W_{v_1}$ and $W_{w_2} \geq W_{w_1}$ we can replace $\xi(W_{v_2} + W_{w_2}) + (1-\xi)(W_{v_1} + W_{w_1})$ in the right-hand side of (5.4) by $W_{v_2} + W_{w_2}$. Taking the L^p-norm of both sides, we obtain the inequality

$$\big\|(T_n - \vartheta)_+\big\|_p \leq \|W_{v_2} + W_{w_2}\|_p + \big\|(1-\xi)(T_{n-2} - \vartheta)_+\big\|_p. \tag{5.5}$$

For each fixed p, it is straightforward to show that $\|W_{v_2} + W_{w_2}\|_p = O(\frac{1}{n})$. Furthermore, as ξ and $(T_{n-2} - \vartheta)_+$ are independent we have $\|(1-\xi)(T_{n-2} - \vartheta)_+\|_p = \|(1-\xi)\|_p \|(T_{n-2} - \vartheta)_+\|_p \leq (1-\varepsilon_2)^{1/p}\|(T_{n-2} - \vartheta)_+\|_p$. Hence, for any fixed p we have the inequality

$$\big\|(T_n - \vartheta)_+\big\|_p \leq O\left(\frac{1}{n}\right) + (1-\varepsilon_2)^{1/p}\big\|(T_{n-2} - \vartheta)_+\big\|_p. \tag{5.6}$$

As $(1-\varepsilon_2)^{1/p} < 1$, it follows that we must have $\|(T_n - \vartheta)_+\|_p = O(\frac{1}{n})$. Combining this with the corresponding bound on $\|(T_n - \vartheta)_-\|_p$ from Proposition 5.2, we have $\|T_n - \vartheta\|_p = O(\frac{1}{n})$, as desired.

It only remains to prove Proposition 5.3.

Loosely speaking, one can think of H_2' as half a hypercube. For instance, exactly half of the oriented paths from v_2 to w_2 in H_2 are contained in H_2', namely those that move in direction j_1 before direction i_1. Now, the paths from v_2 to w_2 in H_2 which are relevant for the early arrivals in the BTP are extremely unlikely to

be oriented, but they are not too far from being oriented either. Our approach to showing Proposition 5.3 is to show that H_2' is a sufficiently large subset of H_2 that, when considering a BTP on H_2 originating at v_2, if there is an uncontested particle at w_2 at time ϑ, then with probability bounded away from 0, its ancestral line is contained in H_2'.

In order to show this, we need a property of the BTP which was hinted at briefly in [4]. Let $\mathbf{X}$ denote a BTP on $\mathbb{Q}_n$ originating at $\hat{\mathbf{0}}$. For any set of paths A in $\mathbb{Q}_n$, let $m(A,t)$ denote the expected number of particles in the BTP at time t whose ancestral lines follow some path in A. Let $\{Y(t)\}_{t\geq 0}$ denote a continuous-time simple random walk on $\mathbb{Q}_n$ starting at $\hat{\mathbf{0}}$ with rate n, and for each $t \geq 0$ let Σ_t denote the path that the random walk has followed up to time t.

LEMMA 5.4. *For any set A of paths in $\mathbb{Q}_n$ from $\hat{\mathbf{0}}$ to $\hat{\mathbf{1}}$ and for any $t \geq 0$, we have*

$$\frac{m(A,t)}{m(\hat{\mathbf{1}},t)} = \mathbb{P}\big(\Sigma_t \in A \mid Y(t) = \hat{\mathbf{1}}\big). \tag{5.7}$$

PROOF. Let σ be any fixed path from $\hat{\mathbf{0}}$ to $\hat{\mathbf{1}}$ and let l denote the length of σ. By applying Lemma 3.1 with $f(\mathbf{X},x) = \mathbb{1}_{t(x)\leq t}$ where $t(x)$ denotes the birth time of x, we get

$$m\big(\{\sigma\},t\big) = \int_0^\infty \cdots \int_0^\infty \mathbb{1}_{s_1+\cdots+s_l\leq t}\, ds_1 \cdots ds_l = \frac{t^l}{l!}. \tag{5.8}$$

In comparison, it is straightforward to see that $\mathbb{P}(\Sigma_t = \sigma) = e^{-nt}\frac{t^l}{n^l l!}$. It follows that, for any set of paths A, we have $m(A,t) = \frac{e^{nt}}{n^l}\mathbb{P}(\Sigma_t \in A)$. In particular, by letting A be the set of all paths from $\hat{\mathbf{0}}$ to $\hat{\mathbf{1}}$ we have $m(\hat{\mathbf{1}},t) = \frac{e^{nt}}{n^l}\mathbb{P}(Y(t) = \hat{\mathbf{1}})$. Hence, for any set of paths A from $\hat{\mathbf{0}}$ to $\hat{\mathbf{1}}$ we have

$$\frac{m(A,t)}{m(\hat{\mathbf{1}},t)} = \frac{\mathbb{P}(\Sigma_t \in A)}{\mathbb{P}(Y(t) = \hat{\mathbf{1}})} = \mathbb{P}\big(\Sigma_t \in A \mid Y(t) = \hat{\mathbf{1}}\big), \tag{5.9}$$

as desired. $\square$

LEMMA 5.5. *Let $\mathbf{X}$ be a BTP on $\mathbb{Q}_n$ originating at $\hat{\mathbf{0}}$. Then with probability $1 - o(1)$, all particles at $\hat{\mathbf{1}}$ at time ϑ have ancestral lines of length $\sqrt{2}\vartheta n \pm o(n)$.*

PROOF. By applying Lemma 5.4, it suffices to show that the number of steps performed by $\{Y(t)\}_{t\geq 0}$ up to time ϑ, conditioned on the event that $Y(\vartheta) = \hat{\mathbf{1}}$, is concentrated around $\sqrt{2}\vartheta n$. In order to show this, we note that each coordinate $Y_i(t)$ of $Y(t)$ is an independent continuous-time simple random walk on $\{0,1\}$ with rate one. Hence, conditioned on the event that $Y(\vartheta) = \hat{\mathbf{1}}$, each coordinate

$Y_i(t)$ is an independent continuous-time simple random walk on $\{0, 1\}$ with rate 1 conditioned on the event that $Y_i(\vartheta) = 1$. It is easy to see that the expected number of steps taken by such a process up to time ϑ is

$$\frac{e^{-\vartheta}}{e^{-\vartheta}} \frac{\vartheta + \frac{\vartheta^3}{2!} + \frac{\vartheta^5}{4!} + \cdots}{\vartheta + \frac{\vartheta^3}{3!} + \frac{\vartheta^5}{5!} + \cdots} = \vartheta \coth \vartheta = \sqrt{2}\vartheta. \tag{5.10}$$

The lemma follows by the law of large numbers. □

PROOF OF PROPOSITION 5.3. Consider the BTP:s $\mathbf{X}$ and $\mathbf{X}'$ on H_2 and H_2', respectively, both originating at v_2. We may couple these processes such that $\mathbf{X}'$ consists of all particles in $\mathbf{X}$ whose ancestral lines are contained in H_2'. Note that any particle in $\mathbf{X}'$ is uncontested in $\mathbf{X}'$ if it is uncontested in $\mathbf{X}$.

As H_2 is graph isomorphic to $\mathbb{Q}_{n-2}$, we know from Theorem 2.2 and Proposition 2.3 together with Lemma 2.1 that, with probability bounded away from zero, there exists an uncontested particle in $\mathbf{X}$ at w_2 at time ϑ. Furthermore, by Lemma 5.5 we know that if such a particle exists, then with probability $1 - o(1)$ the length of its ancestral line is at most $1.25(n-2)$.

Let us now condition on the event that there exists an uncontested particle x in $\mathbf{X}$ at w_2 at time ϑ whose ancestral line is of length at most $1.25(n-2)$. As a path from $\hat{\mathbf{0}}$ to $\hat{\mathbf{1}}$ must traverse edges in each of the $n-2$ directions of $\mathbb{Q}_{n-2}$ an odd number of times, this bound on the length of the ancestral line implies that there are at least $\frac{7}{8}(n-2)$ directions in which the path followed by the ancestral line of x only traverses one edge. By the symmetry, the distribution of this path must be invariant under permutation of coordinates. Hence, with probability bounded from 0, this path only traverses one edge in direction i_1 and one in direction j_1, and traverses the edge in direction j_1 before that in direction i_1. Hence, with probability at least $\frac{49}{128} - o(1)$, this path is contained in H_2'.

We conclude that with probability bounded away from zero, there exists an uncontested particle at w_2 at time ϑ in $\mathbf{X}'$. The proposition follows from the fact that Richardson's model stochastically dominates the set of uncontested particles in a BTP. □

6. Proof of Theorem 1.2. In the following proof, we adopt the notation $m(A, t)$, $\{Y(t)\}_{t \geq 0}$ and Σ_t from the previous section. Hence, Σ_n in the statement of Theorem 1.2 will here be denoted by Σ_ϑ conditioned on $Y(\vartheta) = \hat{\mathbf{1}}$. Recall that Γ_n denotes the geodesic between $\hat{\mathbf{0}}$ and $\hat{\mathbf{1}}$ in $\mathbb{Q}_n$.

It is not hard to see that, with the coupling of the BTP and first-passage percolation described in this paper, the ancestral line of the alive particle at $\hat{\mathbf{1}}$ follows Γ_n. Hence, for any set A_n of paths from $\hat{\mathbf{0}}$ to $\hat{\mathbf{1}}$ in $\mathbb{Q}_n$ and any $c \in \mathbb{R}$, it is easy to show that

$$\mathbb{P}(\Gamma_n \in A_n) \leq m\left(A_n, \vartheta + \frac{c}{n}\right) + \mathbb{P}\left(T_n \geq \vartheta + \frac{c}{n}\right). \tag{6.1}$$

By the Cauchy–Schwarz inequality,

$$
\begin{aligned}
m\Big(A_n, \vartheta + \frac{c}{n}\Big) &= \sum_{\sigma \in A_n} \frac{\vartheta^{|\sigma|}}{|\sigma|!} 1 \cdot \Big(1 + \frac{c}{\vartheta n}\Big)^{|\sigma|} \\
&\le \sqrt{\sum_{\sigma \in A_n} \frac{\vartheta^{|\sigma|}}{|\sigma|!}} \cdot \sqrt{\sum_{\sigma \in A_n} \frac{\vartheta^{|\sigma|}}{|\sigma|!} \Big(1 + \frac{c}{\vartheta n}\Big)^{2|\sigma|}} \\
&= \sqrt{m(A_n, \vartheta)} \cdot \sqrt{m\Big(A_n, \vartheta\Big(1 + \frac{c}{\vartheta n}\Big)^2\Big)} \\
&\le \sqrt{m(A_n, \vartheta)} \cdot \sqrt{m\Big(\hat{\mathbf{1}}, \vartheta\Big(1 + \frac{c}{\vartheta} n\Big)^2\Big)}.
\end{aligned} \tag{6.2}
$$

Furthermore, by the analytical expression for $m(v, t)$ in (2.4) we have

$$
m\Big(\hat{\mathbf{1}}, \vartheta\Big(1 + \frac{c}{\vartheta n}\Big)^2\Big) = \Big(1 + \frac{2\sqrt{2}c}{n} + O\Big(\frac{1}{n^2}\Big)\Big)^n \to e^{2\sqrt{2}c} \tag{6.3}
$$

as $n \to \infty$. Hence, for each $c \in \mathbb{R}$ there exists a $K(c) > 0$ such that $m(\hat{\mathbf{1}}, \vartheta(1 + \frac{c}{\vartheta n})^2) \le K(c)^2$ for all n. Applying Lemma 5.4 to (6.2), it follows that

$$
\mathbb{P}(\Gamma_n \in A_n) \le K(c)\sqrt{\mathbb{P}\big(\Sigma_\vartheta \in A_n \mid Y(\vartheta) = \hat{\mathbf{1}}\big)} + \mathbb{P}\Big(T_n \ge \vartheta + \frac{c}{n}\Big). \tag{6.4}
$$

Consider any sequence A_n such that $\mathbb{P}(\Sigma_\vartheta \in A_n \mid Y(\vartheta) = \hat{\mathbf{1}})$ tends to 0 as $n \to \infty$, that is, A_n is the complement of the corresponding set in the statement of Theorem 1.2. Then, by taking lim sup of both sides in (6.4) we get

$$
\limsup_{n\to\infty} \mathbb{P}(\Gamma_n \in A_n) \le \limsup_{n\to\infty} \mathbb{P}\Big(T_n \ge \vartheta + \frac{c}{n}\Big). \tag{6.5}
$$

The general case of Theorem 1.2 follows from Theorem 1.1 by letting $c \to \infty$. For the special case of the length of Γ_n, see the proof of Lemma 5.5.

APPENDIX A: PROOF OF LEMMA 3.1

In order to prove Lemma 3.1, we need some ideas from Palm theory, and, in particular, the following special case of the Slivnyak–Mecke formula.

THEOREM A.1 (Slivnyak–Mecke formula). *Let* $\mathbf{T}$ *be a Poisson point process on the positive part of the real line with constant intensity* 1. *Let* G *be a function mapping pairs* (T, t) *where* T *is a discrete subset of* $\mathbb{R}_+$ *and* $t \in T$ *to nonnegative real numbers. Then*

$$
\mathbb{E} \sum_{t \in \mathbf{T}} G(\mathbf{T}, t) = \int_0^\infty \mathbb{E} G\big(\mathbf{T} \cup \{t\}, t\big)\, dt. \tag{A.1}
$$

PROOF. A proof of this can be found in various text books on point processes. See for instance Corollary 3.2.3 in [6]. $\square$

PROOF OF LEMMA 3.1. Below, we will use the term e-child of a particle x to denote a child of x that at the time of its birth was displaced along an edge e. For a vertex v and an edge e let $e \ni v$ denote that v is one of the end points of e. For each edge $e \ni v_0$, we let $\mathbf{T}_e$ denote the set of birth times of the e-children of the original particle in $\mathbf{X}_0$. Clearly, $\mathbf{T}_e$ for $e \ni v_0$ are independent Poisson processes on $\mathbb{R}_+$ with constant intensity 1.

An important property of the BTP is that, after a particle is born, the set of its descendants is itself distributed as a BTP. Furthermore, this subprocess is then independent of the behavior of any other particle. Hence, we can express $\mathbf{X}$ as

$$\mathbf{X} = \bigcup_{e \ni v_0} \bigcup_{t_i \in \mathbf{T}_e} \{et_i\} \oplus \mathbf{Y}_{e,i}, \tag{A.2}$$

where for each $e \ni v_0$ and each $i = 1, 2, \dots,$ we have $\mathbf{Y}_{e,i}$ independently distributed as a BTP originating at the vertex opposite to v_0 along e. Recall that $\{et_i\} \oplus \mathbf{Y}_{e,i}$ means adding the prefix $\{et_i\}$ to each particle in $\mathbf{Y}_{e,i}$.

Let $\mathbf{Z} = (\{\mathbf{T}_e\}_{e \ni v_0, e \neq \sigma_1}, \{\mathbf{Y}_{e,i}\}_{e \ni v_0, i \geq 1})$ denote the collection of all random variables in the right-hand side of (A.2) except $\mathbf{T}_{\sigma_1}$. We can consider $\mathbf{X}$ as a function $\mathbf{X} = X(\mathbf{T}_{\sigma_1}, \mathbf{Z})$ of the independent random variables $\mathbf{T}_{\sigma_1}$ and $\mathbf{Z}$ by, for each pair of realizations T_{σ_1} of $\mathbf{T}_{\sigma_1}$ and $Z = (\{T_e\}_{e \ni v_0, e \neq \sigma_1}, \{Y_{e,i}\}_{e \ni v_0, i \geq 1})$, letting

$$X(T_{\sigma_1}, Z) = \bigcup_{e \ni v_0} \bigcup_{t_i \in T_e} \{et_i\} \oplus Y_{e,i}. \tag{A.3}$$

For each such T and Z, we furthermore define

$$F(T, Z) = \sum_{x \in V_\sigma(X(T,Z))} f\big(X(T, Z), x\big), \tag{A.4}$$

$$F(T, Z, t) = \sum_{\substack{x \in V_\sigma(X(T,Z)) \\ x \geq \{\sigma_1 t\}}} f\big(X(T, Z), x\big), \tag{A.5}$$

where $x \geq \{\sigma_1 t\}$ denotes that x is a descendant of $\{\sigma_1 t\}$, or, equivalently, $\{\sigma_1 t\}$ is a prefix of x.

As $\mathbf{X} = X(\mathbf{T}_{\sigma_1}, \mathbf{Z})$ and as every particle in $V_\sigma(X(\mathbf{T}_{\sigma_1}, \mathbf{Z}))$ is a descendant of some σ_1-child of the original particle, it follows from equations (A.4) and (A.5) that

$$\mathbb{E} \sum_{x \in V_\sigma(\mathbf{X})} f(\mathbf{X}, x) = \mathbb{E} F(\mathbf{T}_{\sigma_1}, \mathbf{Z}) = \mathbb{E} \sum_{t \in \mathbf{T}_{\sigma_1}} F(\mathbf{T}_{\sigma_1}, \mathbf{Z}, t). \tag{A.6}$$

Furthermore, by conditioning on $\mathbf{Z} = Z$ the Slivnyak–Mecke formula implies that

$$\mathbb{E} \sum_{t \in \mathbf{T}_{\sigma_1}} F(\mathbf{T}_{\sigma_1}, Z, t) = \int_0^\infty \mathbb{E} F\big(\mathbf{T}_{\sigma_1} \cup \{s_1\}, Z, s_1\big)\, ds_1. \tag{A.7}$$

Hence, by combining (A.6) with (A.7) and plugging in the definition of $F(T, Z, t)$, we can conclude that

$$\begin{aligned}\mathbb{E}\sum_{x\in V_\sigma(\mathbf{X})} f(\mathbf{X}, x) &= \int_0^\infty \mathbb{E}F\big(\mathbf{T}_{\sigma_1}\cup\{s_1\}, \mathbf{Z}, s_1\big)\,ds_1 \\ &= \int_0^\infty \mathbb{E}\sum_{\substack{x\in V_\sigma(X_0(\mathbf{T}_{\sigma_1}\cup\{s_1\},\mathbf{Z}))\\ x\geq\{\sigma_1 s_1\}}} f\big(X_0(\mathbf{T}_{\sigma_1}\cup\{s_1\}, \mathbf{Z}), x\big)\,ds_1.\end{aligned} \tag{A.8}$$

By comparing the random process $X(\mathbf{T}_{\sigma_1}\cup\{s_1\}, \mathbf{Z})$ as defined in (A.3) to $\mathbf{X}$ as defined in (A.2) it is clear that the former has the same distribution as $\mathbf{X}_0\cup(\{\sigma_1 s_1\}\oplus\mathbf{X}_1)$, where $\mathbf{X}_0$ and $\mathbf{X}_1$ are independent BTP:s originating at v_0 and v_1, respectively. Hence, we can replace $X(\mathbf{T}_{\sigma_1}\cup\{s_1\}, \mathbf{Z})$ in (A.8) by this other random process. Letting $\tilde\sigma = \{\sigma_2, \sigma_3, \dots, \sigma_l\}$, we note that the subset of elements in $V_\sigma(\mathbf{X}_0\cup(\{\sigma_1 s_1\}\oplus\mathbf{X}_1))$ that are descendants of $\{\sigma_1 s_1\}$ is precisely the set $\{\sigma_1 s_1\}\oplus V_{\tilde\sigma}(\mathbf{X}_1)$. Hence, (A.8) simplifies to

$$\begin{aligned}&\mathbb{E}\sum_{x\in V_\sigma(\mathbf{X})} f(\mathbf{X}, x) \\ &\quad= \int_0^\infty \mathbb{E}\sum_{x\in V_{\tilde\sigma}(\mathbf{X}_1)} f\big(\mathbf{X}_0\cup\big(\{\sigma_1 s_1\}\oplus\mathbf{X}_1\big), \{\sigma_1 s_1\}\oplus x\big)\,ds_1.\end{aligned} \tag{A.9}$$

The lemma follows by induction on l, the length of σ. If $l = 1$, then the only particle in $V_{\tilde\sigma}(\mathbf{X}_1)$ is $\{\cdot\}$, the original particle in $\mathbf{X}_1$, and so equation (A.9) simplifies to

$$\mathbb{E}\sum_{x\in V_\sigma(\mathbf{X})} f(\mathbf{X}, x) = \int_0^\infty \mathbb{E}f\big(\mathbf{X}_0\cup\big(\{\sigma_1 s_1\}\oplus\mathbf{X}_1\big), \{\sigma_1 s_1\}\big)\,ds_1 \tag{A.10}$$

as desired.

Now assume $l > 1$. By the induction hypothesis we have for any nonnegative function $\tilde f$

$$\mathbb{E}\sum_{x\in V_{\tilde\sigma}(\mathbf{X}_1)} \tilde f(\mathbf{X}_1, x) = \int_{s_2, s_3, \dots, s_l\geq 0} \mathbb{E}\tilde f\big(\mathbf{X}(\tilde\sigma, \tilde{\mathbf{s}}), x(\tilde\sigma, \tilde{\mathbf{s}})\big)\,d\tilde{\mathbf{s}}, \tag{A.11}$$

where $\tilde{\mathbf{s}} = (s_2, s_3, \dots, s_l)$, $d\tilde{\mathbf{s}} = ds_2\,ds_3\cdots ds_l$ and $\mathbf{X}(\tilde\sigma, \tilde{\mathbf{s}})$ and $x(\tilde\sigma, \tilde{\mathbf{s}})$ are defined by (3.3) and (3.2), respectively. Let us consider the expression

$$\mathbb{E}\sum_{x\in V_{\tilde\sigma}(\mathbf{X}_1)} f\big(\mathbf{X}_0\cup\big(\{\sigma_1 s_1\}\oplus\mathbf{X}_1\big), \{\sigma_1 s_1\}\oplus x\big), \tag{A.12}$$

the integrand on the right-hand side of equation (A.9). If we fix $s_1 > 0$ and condition on $\mathbf{X}_0 = X_0$, then $f(\mathbf{X}_0\cup(\{\sigma_1 s_1\}\oplus\mathbf{X}_1), \{\sigma_1 s_1\}\oplus x)$ is a function of $\mathbf{X}_1$ and

x only. By the induction hypothesis,

$$\mathbb{E} \sum_{x \in V_{\tilde{\sigma}}(\mathbf{X}_1)} f\big(X_0 \cup (\{\sigma_1 s_1\} \oplus \mathbf{X}_1), \{\sigma_1 s_1\} \oplus x\big)$$
$$= \int_{s_2, s_3, \dots, s_l \geq 0} \mathbb{E} f\big(X_0 \cup \{\sigma_1 s_1\} \oplus \mathbf{X}(\tilde{\sigma}, \tilde{\mathbf{s}}), \{\sigma_1 s_1\} \oplus x(\tilde{\sigma}, \tilde{\mathbf{s}})\big)\, d\tilde{\mathbf{s}},$$

where $\mathbf{X}(\tilde{\sigma}, \tilde{\mathbf{s}})$ is generated independently of $\mathbf{X}_0$. Hence, by integrating this expression over s_1 and $\mathbf{X}_0$ we conclude that

$$\mathbb{E} \sum_{x \in V_{\sigma}(\mathbf{X})} f(\mathbf{X}, x)$$
$$= \int_{s_1, s_2, \dots, s_l \geq 0} \mathbb{E} f\big(\mathbf{X}_0 \cup \{\sigma_1 s_1\} \oplus \mathbf{X}(\tilde{\sigma}, \tilde{\mathbf{s}}), \{\sigma_1 s_1\} \oplus x(\tilde{\sigma}, \tilde{\mathbf{s}})\big)\, d\mathbf{s},$$

where it is easy to check that $x(\sigma, \mathbf{s}) = \{\sigma_1 s_1\} \oplus x(\tilde{\sigma}, \tilde{\mathbf{s}})$ and that $\mathbf{X}(\sigma, \mathbf{s})$ has the same distribution as $\mathbf{X}_0 \cup \{\sigma_1 s_1\} \oplus \mathbf{X}(\tilde{\sigma}, \tilde{\mathbf{s}})$. $\square$

Acknowledgments. I am very grateful to my supervisor, Peter Hegarty, for his constant support and for numerous insightful discussions and comments during this project. I would also like to thank Jeffrey Steif, Johan Tykesson and Johan Wästlund for valuable input on various aspects of this topic. I also thank the anonymous referee for his or her careful reading of the manuscript.

REFERENCES

[1] ALDOUS, D. (1989). *Probability Approximations Via the Poisson Clumping Heuristic*. *Applied Mathematical Sciences* **77**. Springer, New York. MR0969362

[2] BOLLOBÁS, B. and KOHAYAKAWA, Y. (1997). On Richardson's model on the hypercube. In *Combinatorics*, *Geometry and Probability* (*Cambridge*, 1993) 129–137. Cambridge Univ. Press, Cambridge. MR1476439

[3] DURRETT, R. (1988). *Lecture Notes on Particle Systems and Percolation*. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA. MR0940469

[4] FILL, J. A. and PEMANTLE, R. (1993). Percolation, first-passage percolation and covering times for Richardson's model on the n-cube. *Ann. Appl. Probab.* **3** 593–629. MR1221168

[5] HEGARTY, P. and MARTINSSON, A. (2014). On the existence of accessible paths in various models of fitness landscapes. *Ann. Appl. Probab.* **24** 1375–1395. MR3210999

[6] SCHNEIDER, R. and WEIL, W. (2008). *Stochastic and Integral Geometry*. Springer, Berlin. MR2455326

DEPARTMENT OF MATHEMATICAL SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
AND UNIVERSITY OF GOTHENBURG
41296 GOTHENBURG
SWEDEN
E-MAIL: andemar@chalmers.se