The Annals of Probability
2011, Vol. 39, No. 1, 327–368
DOI: 10.1214/10-AOP551

OCCUPATION STATISTICS OF CRITICAL BRANCHING RANDOM WALKS IN TWO OR HIGHER DIMENSIONS

BY STEVEN P. LALLEY[1] AND XINGHUA ZHENG[2]

University of Chicago and Hong Kong University of Science and Technology

Consider a critical nearest-neighbor branching random walk on the d-dimensional integer lattice initiated by a single particle at the origin. Let G_n be the event that the branching random walk survives to generation n. We obtain the following limit theorems, conditional on the event G_n, for a variety of occupation statistics: (1) Let V_n be the maximal number of particles at a single site at time n. If the offspring distribution has finite αth moment for some integer $\alpha \geq 2$, then, in dimensions 3 and higher, $V_n = O_p(n^{1/\alpha})$. If the offspring distribution has an exponentially decaying tail, then $V_n = O_p(\log n)$ in dimensions 3 and higher and $V_n = O_p((\log n)^2)$ in dimension 2. Furthermore, if the offspring distribution is nondegenerate, then $P(V_n \geq \delta \log n \mid G_n) \to 1$ for some $\delta > 0$. (2) Let $M_n(j)$ be the number of multiplicity-j sites in the nth generation, that is, sites occupied by exactly j particles. In dimensions 3 and higher, the random variables $M_n(j)/n$ converge jointly to multiples of an exponential random variable. (3) In dimension 2, the number of particles at a "typical" site (i.e., at the location of a randomly chosen particle of the nth generation) is of order $O_p(\log n)$ and the number of occupied sites is $O_p(n/\log n)$. We also show that, in dimension 2, there is particle clustering around a typical site.

1. Introduction. A *nearest-neighbor branching random walk* is a discrete-time particle system on the integer lattice $\mathbb{Z}^d$ that evolves according to the following rule: at each time $n = 0, 1, 2, \ldots,$ every particle generates a random number of offspring, with offspring distribution $\mathcal{Q} = \{Q_l\}_{l \geq 0}$; each of these then moves to a site randomly chosen from among the $2d + 1$ sites at distance ≤ 1 from the location of the parent.[3] We shall consider only the case where the branching random walk is *critical*, that is, where the mean number of offspring per particle is 1,

---

Received July 2007; revised November 2008.

[1]Supported in part by NSF Grant DMS-08-05755.

[2]Supported in part by the NSERC (Canada) and by research support from the Department of Information Systems, Business Statistics and Operations Management, Hong Kong University of Science and Technology.

AMS 2000 subject classifications. Primary 60J80; secondary 60G60; tertiary 60F05.

Key words and phrases. Critical branching random walks, limit theorems, occupation statistics.

[3]Allowing particles to remain at the same locations as their parents with positive probability eliminates some annoying periodicity problems that would require tedious, but routine, arguments to circumvent. Our main results could be proven under much less restrictive hypotheses on the jump distribution.

and we shall assume throughout that the offspring distribution has finite, positive variance σ^2.

By a well-known theorem of Kolmogorov [see Athreya and Ney (1972), Chapter 1], if the branching process is initiated by a single particle and if G_n is the event that the process survives to generation n, then

$$\pi_n := P(G_n) \sim \frac{2}{n\sigma^2}. \tag{1.1}$$

Therefore, if the branching random walk is started with n particles at time 0, then the number of initial particles whose families survive to time n follows, approximately for large n, a Poisson distribution with mean $2/\sigma^2$ and the number of particles Z_n alive at time n is of order $O_p(n)$. In fact, in this case, under suitable hypotheses on the initial distribution of particles, the measure-valued process associated with the branching random walk converges, after rescaling, to the super-Brownian motion X_t with variance parameter σ^2 [see, e.g., Etheridge (2000)]. In dimensions 2 and higher, the random measure X_t is, for each $t > 0$, almost surely singular with respect to the Lebesgue measure on $\mathbb{R}^d$. When $d \geq 3$, the measure X_t spreads its mass over the support in a fairly uniform manner [Perkins (1988)] and, in fact, can be recovered from its support [Perkins (1989)]. It is natural to conjecture that this uniformity also holds, in a suitable sense, for critical branching random walk and that the *maximal* number of particles at a single site at time n does not grow rapidly in n. Our main results show that this is indeed the case. For ease of exposition, we will state our results as conditional limit theorems, given the event G_n, of survival to generation n. Corresponding unconditional results for branching random walks started by n particles could easily be deduced.

We shall assume throughout the paper, unless otherwise specified, that the branching random walk is initiated by a single particle located at the origin at time 0. Define

$$\begin{aligned}
\mathcal{Z}_n &:= \text{set of particles in generation } n; \\
Z_n &:= |\mathcal{Z}_n| = \text{number of particles in generation } n; \\
U_n(x) &:= \text{number of particles at site } x \text{ in generation } n; \\
\Omega_n &:= \text{number of occupied sites in generation } n; \\
M_n(j) &:= \text{number of multiplicity-}j \text{ sites in generation } n \quad \text{and} \\
V_n &:= \max_{x \in \mathbb{Z}^d} U_n(x).
\end{aligned} \tag{1.2}$$

(A *multiplicity-j site* is a site with exactly j particles.)

DEFINITION 1. Let X_n be a sequence of random variables, $f(n)$ a sequence of positive real numbers and H_n a sequence of events of positive probability. We say that $X_n = O_P(f(n))$ given H_n if the conditional distributions of $X_n/f(n)$ given

H_n are tight. Similarly, we say that $X_n = o_P(f(n))$ given H_n if the conditional distributions of $X_n/f(n)$ given H_n converge weakly to the point mass at 0.

THEOREM 2. *Assume that the offspring distribution $\mathcal{Q}$ has finite αth moment for some integer $\alpha \geq 2$ and that $d \geq 3$. Then, conditional on G_n,*

$$V_n = O_P(n^{1/\alpha}). \tag{1.3}$$

In particular, if $\mathcal{Q}$ has finite moments of all orders, then $V_n = o_p(n^{\varepsilon})$ for all $\varepsilon > 0$.

THEOREM 3. *Assume that the offspring distribution $\mathcal{Q}$ has an exponentially decaying tail, that is, there exists $\delta > 0$ such that $\sum_l Q_l \exp(\delta l) < \infty$. Then, conditional on G_n,*

$$V_n = O_p(\log n) \qquad \text{if } d \geq 3; \tag{1.4}$$

$$V_n = O_p((\log n)^2) \qquad \text{if } d = 2. \tag{1.5}$$

In fact (see Corollary 16 below), for sufficiently large $C > 0$, the conditional probabilities $P(V_n \geq C \log n \mid G_n)$ in dimensions $d \geq 3$ and $P(V_n \geq C(\log n)^2 \mid G_n)$ in dimension $d = 2$ decay polynomially in n. For one-dimensional branching random walk, it is known that V_n is of order $\sqrt{n}$ [Theorem 7.10 in Révész (1994)]; stronger results are proved in Lalley (2009).

THEOREM 4. *Assume that $d \geq 2$. There then exists $\delta > 0$, depending on the offspring distribution $\mathcal{Q}$, such that*

$$\lim_{n\to\infty} P(V_n \geq \delta \log n \mid G_n) = 1. \tag{1.6}$$

Theorems 3 and 4 imply, in dimensions 3 and higher, that if the offspring distribution has an exponentially decaying tail, then V_n is of order $\log n$ on the event G_n of survival to generation n. In particular, the (conditional) distributions of $V_n/\log n$ are tight and any weak limit has support contained in $[\delta_1, \delta_2]$ for some $\delta_1, \delta_2 > 0$ (cf. Corollary 16). This partly settles an open question (Question 2, page 79) raised in Révész (1996).

THEOREM 5. *Assume that $d \geq 3$. Then, conditional on the event G_n, the joint distribution of the occupation statistics $M_n(j)/n$ converges as $n \to \infty$. In particular, for certain constants κ_j such that $\sum_{j=1}^{\infty} j \cdot \kappa_j = 1$,*

$$\mathcal{L}\left(\frac{Z_n}{n}, \left\{\frac{M_n(j)}{n}\right\}_{j\geq 1}, \frac{\Omega_n}{n} \,\Big|\, G_n\right) \Longrightarrow \left(1, \{\kappa_j\}_{j\geq 1}, \sum_j \kappa_j\right) \cdot Y, \tag{1.7}$$

where Y is exponentially distributed with mean $2/\sigma^2$.

This extends the classical theorem of Yaglom, according to which the conditional distribution of Z_n/n, given that the branching process survives to generation n, converges to the exponential law with mean $2/\sigma^2$. See Athreya and Ney [(1972), Chapter 1] for a discussion of Yaglom's theorem and related results, and Geiger (2000) for an interesting probabilistic proof.

Theorem 5 implies, in dimensions 3 and higher, that most occupied sites are occupied by only $O(1)$ particles. Ultimately, this is a consequence of the transience of random walk in dimensions $d \geq 3$. Since random walk in dimension $d = 2$ is recurrent, different behavior should be expected for the occupation statistics of branching random walk. In the following theorem, and throughout this article, we shall use the term *typical particle* to mean a particle chosen randomly from the nth generation $\mathcal{Z}_n$ of the branching process (with the choice made independently of the evolution of the branching random walk up to time n, according to the uniform distribution on $\mathcal{Z}_n$). By a *typical site*, we mean the location of a typical particle.

THEOREM 6. *In dimension $d = 2$, the number T_n of particles at a typical site at time n is, conditional on the event G_n, of order $O_p(\log n)$. Moreover, for some sufficiently small $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$\liminf_{n\to\infty} P(T_n \geq \varepsilon \log n \mid G_n) \geq \delta. \tag{1.8}$$

We conjecture that the conditional distributions of $T_n/\log n$ given G_n converge in distribution as $n \to \infty$. Fleischman (1978) has used the method of moments to establish a related result for the number of particles at a *fixed* site at distance $O(1)$ from the origin. Unfortunately, the calculation of higher moments for the number of particles at a *typical* site appears to be considerably more difficult and so the method of moments does not seem to be a feasible approach to the conjecture.

By Yaglom's theorem, conditional on the event of survival to generation n, there are $O_P(n)$ particles in total. Theorem 6 implies that at least a fraction δ of these are located at sites with (roughly) $\log n$ other particles. Thus, a substantial fraction of the particles can be found in just $O_P(n/\log n)$ sites. This does not logically rule out the possibility that many more sites are occupied; however, it does suggest that the number Ω_n of occupied sites is of order $o_p(n)$. This is consistent with the corresponding result for super-Brownian motion X_t, which states that for any $t > 0$, the random measure X_t is almost surely singular. The following is a sharp result concerning the number of occupied sites.

THEOREM 7. *For two-dimensional nearest-neighbor branching random walk, the number Ω_n of occupied sites is $O_p(n/\log n)$, given the event G_n.*

Theorem 6 implies that the number of occupied sites must be of order *at least* $n/\log n$. Combining this with Theorem 7, we see that $n/\log n$ is the true asymptotic rate. Révész [(1996), Theorem 3(ii)] asserts that a corresponding result is true for branching Brownian motion, but we believe that his proof has a serious flaw; see Section 7.2 for a detailed discussion of this matter.

The next theorem partially quantifies the degree of particle clustering around a typical site.

THEOREM 8. *Assume that* $d = 2$. *Let* $\{\ell_n\}$ *be any sequence of real numbers such that* $\lim_n \ell_n = \infty$ *and* $\lim_n \log \ell_n / \log n = 0$. *Let* S_n *be the location of a typical particle and let* $B(S_n; \ell_n)$ *be the ball of radius* ℓ_n *centered at* S_n. *Then, conditional on* G_n:

(A) *the number of unoccupied sites in* $B(S_n; \ell_n)$ *is* $o_P(\ell_n^2)$;
(B) *the number of particles in* $B(S_n; \ell_n)$ *is of order* $O_p(\log n \cdot \ell_n^2)$.

Theorems 2 and 3 are proved in Section 2, Theorem 4 in Section 3 and Theorem 5 in Section 4. Theorem 6 is proved in Section 5, Theorem 8 in Section 6 and Theorem 7 in Section 7. For each of the last three theorems, the calculations required for the proofs are considerably simpler in the special case of *binary fission*, where the offspring distribution $\mathcal{Q}$ is *double-or-nothing*, that is, $Q_0 = Q_2 = 1/2$. In the interests of clarity, we shall give complete arguments only for this special case. These arguments (as should be evident) can be extended to the general case of mean 1, finite-variance offspring distributions.

Fundamental to many of our arguments is the following elementary relation between the expected number of particles at a site x in generation n and the n-step transition probabilities $P_n(x)$ of the simple random walk:

$$EU_n(x) = P_n(x). \tag{1.9}$$

This is easily proved by induction on n, by conditioning on the first generation of the branching random walk. Here, and throughout the paper, the term *simple random walk* is used for the symmetric nearest-neighbor random walk on the lattice $\mathbb{Z}^d$ with holding probability $1/(2d+1)$—that is, each increment is uniformly distributed on the set $\mathcal{N}$ of $2d+1$ sites at distance ≤ 1 from the origin—and the notation $P_n(x)$ is reserved for the probability that a simple random walk started at the origin finds its way to site x in n steps. We use the notation $\mathbb{P}^n$ to denote the n-step transition probability kernel of simple random walk, that is, the nth iterate of the Markov operator $\mathbb{P}: \ell^\infty(\mathbb{Z}^d) \to \ell^\infty(\mathbb{Z}^d)$ associated with the random walk.

Notation. The following is a list of notation, in addition to that already established in equations (1.1), (1.2) and (1.9) above, that will be fixed throughout the paper:

- $\mathcal{N} = \{e_j\}_{-d \leq j \leq d}$ is the set of sites at distance 0 or 1 from the origin in $\mathbb{Z}^d$;
- $\mathcal{Q} = \{Q_l\}_{l \geq 0}$ is the offspring distribution and $\mathcal{Q}^i = \{Q_l^i\}$ its ith convolution power;

- $\mathcal{F}_n$ is the σ-algebra generated by the random variables $\{U_m(x)\}_{x\in\mathbb{Z}^d, m\le n}$;
- $A = 5/(4\pi)$ is the constant such that $P_n(0) \sim A/n$ in dimension 2; see, for instance, P7.9 on page 75 of Spitzer (1976).

In addition, we will follow the custom of writing $f \sim g$ to mean that the ratio f/g converges to 1 and $f \asymp g$ to mean that the ratio f/g remains bounded away from 0 and ∞. Throughout the paper, C, C_1, C', etc. denote generic constants whose values may change from line to line. Finally, we use a "local scoping rule" for notation: any notation introduced in a proof is local to the proof, unless otherwise indicated.

2. Proofs of Theorems 2 and 3.

2.1. *The case where the offspring distribution has finite moments*. The proof of Theorem 2 will rely on the following estimates for the moments of the occupation statistics $U_n(x)$.

PROPOSITION 9. *Suppose that the offspring distribution Q has finite αth moment for some integer $\alpha \ge 2$.*

(i) *If $d \ge 3$, then*

$$\sup_n \sum_x EU_n(x)^\alpha < \infty.$$

(ii) *If $d = 2$, then there exist $C_1, C_2 < \infty$ such that for all n,*

$$\sum_x EU_n(x)^\alpha \le C_1 n^{C_2 2^\alpha}.$$

PROOF. We will use the following inequality: for all $l \ge 2$ and all $b_i \ge 0$,

$$\left(\sum_{i=1}^l b_i\right)^\alpha \le \sum_{k=2}^\alpha \sum_{\mathcal{P}_k} \left(\sum_{\ell=1}^k b_{i_\ell} \cdot \mathbf{1}_{\{b_{i_1}>0,\dots,b_{i_k}>0\}}\right)^\alpha, \tag{2.1}$$

where $\mathcal{P}_k$ is the set of k-tuples $(i_1, \dots, i_k)$ of distinct positive integers no greater than l. Inequality (2.1) is obviously true for $l \le \alpha$. To see that it holds for $l > \alpha$, observe that, by the multinomial expansion, the left-hand side of (2.1) is a sum of terms of the form $t = \binom{\alpha}{j_1 j_2 \cdots j_l} b_1^{j_1} b_2^{j_2} \cdots b_l^{j_l}$, where the exponents j_i sum to α. Since at most α of these can be positive and t vanishes if any of the b_i with exponent $j_i > 0$ is zero, the term t is included in the sum on the right-hand side of (2.1).

Next, by the Hölder inequality, for each integer $k \ge 2$ and all real numbers $b_i \ge 0$,

$$\left(\sum_{i=1}^k b_i\right)^\alpha \le k^{\alpha-1} \sum_{i=1}^k b_i^\alpha. \tag{2.2}$$

This implies that if k independent branching random walks are started by particles $u_1, \dots, u_k$ located at sites $x_1, \dots, x_k$, respectively, and if $U_n^{u_i}(x)$ is the number of the nth generation descendants at site x of the particle u_i, then

$$
\begin{aligned}
&\sum_x E\left(\sum_{i=1}^k U_n^{u_i}(x) \cdot \mathbf{1}_{\{U_n^{u_1}(x)>0,\dots,U_n^{u_k}(x)>0\}}\right)^{\alpha} \\
&\quad \le k^{\alpha-1} \sum_{i=1}^k \sum_x E(U_n^{u_i}(x))^{\alpha} \cdot \prod_{j\ne i} P\bigl(U_n^{u_j}(x) > 0\bigr) \qquad (2.3)\\
&\quad \le k^{\alpha} \sum_x EU_n(x)^{\alpha} \cdot \left(C\frac{1}{\sqrt{n}^d}\right)^{k-1}.
\end{aligned}
$$

Here, we have used (2.2) in the first inequality; the second inequality follows by the local central limit theorem and the elementary observation that

$$P\bigl(U_n^{u_j}(x) > 0\bigr) \le EU_n^{u_j}(x) = P_n(x - x_j).$$

We are now prepared to estimate $\sum_x EU_n(x)^{\alpha}$. Conditioning on the first generation, we obtain

$$
\begin{aligned}
\sum_x EU_n(x)^{\alpha} &\le \sum_x EU_{n-1}(x)^{\alpha} \\
&\quad + \sum_{k=2}^{\alpha} \sum_x E\left[\sum_{\mathcal{P}_k}\left(\sum_{j=1}^k U_{n-1}^{u_j}(x) \cdot \mathbf{1}_{\{U_{n-1}^{u_1}(x)>0,\dots,U_{n-1}^{u_k}(x)>0\}}\right)^{\alpha}\right] \\
&\le \sum_x EU_{n-1}(x)^{\alpha} \cdot \left(1 + \sum_{k=2}^{\alpha}\sum_l Q_l \binom{l}{k} k^{\alpha} \cdot \left(C\frac{1}{\sqrt{n-1}^d}\right)^{k-1}\right),
\end{aligned}
$$

where $\mathcal{P}_k$ denotes the set of k-tuples $(u_1, \dots, u_k)$ of distinct particles in generation 1 and the first and second inequalities hold by (2.1) and (2.3), respectively. Therefore, for all n,

$$\sum_x EU_n(x)^{\alpha} \le \prod_{i=2}^n \left(1 + \sum_{k=2}^{\alpha}\sum_l Q_l \binom{l}{k} k^{\alpha} \cdot \left(C\frac{1}{\sqrt{i-1}^d}\right)^{k-1}\right) \cdot \sum_x EU_1(x)^{\alpha}.$$

Clearly, $\sum_x EU_1(x)^{\alpha} \le (2d+1)EZ_1^{\alpha} < \infty$. Furthermore, in dimensions $d \ge 3$,

$$
\begin{aligned}
&\prod_{i=2}^n \left(1 + \sum_{k=2}^{\alpha}\sum_l Q_l \binom{l}{k} k^{\alpha} \cdot \left(C\frac{1}{\sqrt{i-1}^d}\right)^{k-1}\right) \\
&\quad \le \exp\left(\sum_{i=2}^{\infty}\sum_{k=2}^{\alpha}\sum_l Q_l \binom{l}{k} k^{\alpha} \cdot \left(C\frac{1}{\sqrt{i-1}^d}\right)^{k-1}\right) \\
&\quad = \exp\left(C' \sum_{k=2}^{\alpha}\sum_l Q_l \binom{l}{k} k^{\alpha}\right),
\end{aligned}
$$

where $C' < \infty$ is independent of n, and in dimension $d = 2$,

$$\begin{aligned}\prod_{i=2}^{n}&\left(1+\sum_{k=2}^{\alpha}\sum_{l}Q_l\binom{l}{k}k^\alpha\cdot\left(\frac{C}{i-1}\right)^{k-1}\right)\\&\le\exp\left(C\sum_{l}Q_l\binom{l}{2}2^\alpha\cdot\sum_{i=2}^{n}\frac{1}{i-1}\right.\\&\qquad\left.+C\sum_{k=3}^{\alpha}\sum_{l}Q_l\binom{l}{k}k^\alpha\cdot\sum_{i=2}^{\infty}\left(\frac{1}{i-1}\right)^{k-1}\right)\\&\le\exp(C_2 2^\alpha\log n+C_3),\end{aligned}$$

where C_2 is a constant independent of both α and n, and C_3 is a constant independent of n. $\square$

PROOF OF THEOREM 2. By Kolmogorov's estimate (1.1), the probability that the process survives to time n is $O(1/n)$. By the Markov inequality,

$$P\{V_n \ge Cn^{1/\alpha}\} \le C^{-\alpha}n^{-1}EV_n^\alpha \le C^{-\alpha}n^{-1}E\sum_x U_n(x)^\alpha$$

and so the relation (1.3) follows from Proposition 9. $\square$

REMARK 10. Yaglom's limit theorem implies that, conditional on the event G_n, the number of particles at time $n-1$ is $O_p(n)$. For each of these, there is a small chance that the number of offspring will exceed $(2d+1)n^{1/(\alpha+\varepsilon)}$, in which case V_n will be at least $n^{1/(\alpha+\varepsilon)}$. If the tail of the offspring distribution decays like $m^{-(\alpha+\varepsilon)}$ as $m \to \infty$, then the chance that one of the $O_p(n)$ particles in generation $n-1$ will have more than $(2d+1)n^{1/(\alpha+\varepsilon)}$ offspring is of order one. Thus, the result in Theorem 2 is almost optimal. (This answers a question of Michael Stein.)

2.2. *The case where the offspring distribution has an exponentially decaying tail.* We begin with a stochastic comparison result for the random variables $U_n(x)$. First, observe that the law of the branching random walk (started by a single particle located at the origin) is invariant with respect to reflections in the coordinate axes and so $U_n(x) \stackrel{\mathcal{D}}{=} U_n(x')$ for any two sites x, x' at corresponding positions of different orthants. Now, define the usual partial order on the positive orthant $\mathbb{Z}_+^d$:

$$x \preceq y \qquad \text{if } x_i \le y_i \qquad \text{for all } 1 \le i \le d.$$

LEMMA 11. *If $x \preceq y$, then $U_n(x)$ stochastically dominates $U_n(y)$. In particular, $U_n(y)$ is stochastically dominated by $U_n(0)$ for every $y \in \mathbb{Z}^d$. Consequently, if $x \preceq y$, then for every $n \ge 0$,*

$$P_n(x) \ge P_n(y) \tag{2.4}$$

and

$$u_n(x) \geq u_n(y), \tag{2.5}$$

where $u_n(x) := P\{U_n(x) \geq 1\}$ *is the hitting probability function of the branching random walk.*

REMARK 12. The relation (2.4), which follows from the stochastic dominance $U_n(x) \geq_{\mathcal{D}} U_n(y)$ by taking expectations [recall the fundamental relation (1.9)], also follows more directly by the reflection principle for simple random walk.

PROOF OF LEMMA 11. Because the law of the branching random walk is invariant with respect to permutations of the coordinates, we may assume, without loss of generality, that $y = x + e_1$, where $e_1 = (1, 0, \dots, 0)$. Denote by L and L' the hyperplanes

$$L = \{z \in \mathbb{R}^d : z_1 = x_1\}$$

and

$$L' = \{z \in \mathbb{R}^d : z_1 = x_1 + 1/2\};$$

observe that y is the reflection of x in L'. We shall define a particle system with particles of three colors—red, blue and green—in such a way that:

(a) the subpopulation of all red and blue particles follows the law of the branching random walk started by one (red) particle at the origin;
(b) the subpopulation of all red and green particles follows the same law;
(c) there are no red particles to the right of the hyperplane L';
(d) at each time, the green and blue particles are paired (bijectively) in such a way that the green and blue particles in any pair are at symmetric locations on opposite sides of the hyperplane L'.

This will prove that $U_n(x) \geq_{\mathcal{D}} U_n(y)$ for each n, by the following reasoning: first, the distribution of $U_n(x)$ coincides with the distribution of the total number of *red* and *blue* particles at location x and time n, by (a); second, the number of *blue* particles at x equals the number of *green* particles at y, by (d), since x and y are at symmetric locations on opposite sides of the hyperplane L'; third, the number of *green* particles at y has the same distribution as $U_n(y)$, by (b) and (c).

The particle system is constructed as follows. To begin, color the initial particle at the origin red. Offspring of blue and green particles will always have the same color as their parents and each blue particle b will always be paired with a green particle g located at the mirror image (relative to reflection in the hyperplane L') of the site of b. Offspring of red particles will be red, except possibly when the parent red particle is located at a site on the hyperplane L. In this case—say, for definiteness, that the red parent particle ξ is at site $z \in L$—each offspring particle ζ first makes a jump according to the law of the nearest-neighbor random walk and then chooses a color as follows: (a) if the jump is to a site $z' \neq z$ to the *left* of hyperplane L', then ζ becomes *red*; (b) if the jump is either to the same site z as the parent or to its mirror image z^* on the *right* of L', then ζ chooses randomly between blue and green. In case (b), the offspring particle ζ generates a *doppelganger* (mirror particle) ζ' of the opposite color at the reflected site on the other side of L'. Note that the distribution of the position of ζ is the same as that of ζ'. The particle ζ generates an offspring branching random walk $\mathcal{G}_\zeta$ with all particles having the same color as ζ; the mirror image $\mathcal{G}_{\zeta'}$ of $\mathcal{G}_\zeta$ relative to L' (with particles colored oppositely) is attached to ζ'. Note that $\mathcal{G}_{\zeta'}$ is itself a branching random walk started at the location of ζ', by the symmetry of the nearest-neighbor random walk.

Properties (a)–(d) above are now readily apparent. Property (c) holds because, by construction, children of red particles on L that jump across L' are either green or blue and offspring of blue and green particles are either blue or green. Property (d) is inherent in the construction. Finally, (a) and (b) follow from the blue/green symmetry of the reproduction law for red particles located at sites on L. $\square$

PROPOSITION 13. *Assume that the offspring distribution $\mathcal{Q}$ has a finite moment generating function in some neighborhood of the origin. Then, in dimensions $d \geq 3$, there exist $\delta_d > 0$ and $C > 0$ such that for any $\theta \in [0, \delta_d]$, all $x \in \mathbb{Z}^d$ and all $n \geq 1$,*

$$E \exp\{\theta U_n(x)\} - 1 \leq C P_n(x)\theta. \tag{2.6}$$

In dimension $d = 2$, there exist $\delta_2 > 0$ and $C > 0$ such that for any $\theta \in [0, \delta_2]$, all $x \in \mathbb{Z}^2$ and all $n \geq 1$,

$$E \exp\{\theta U_n(x)/\log n\} - 1 \leq C P_n(x)\theta/\log n. \tag{2.7}$$

PROOF. Let $\Phi(z) = \sum_{l=1}^{\infty} Q_l z^l$ be the probability generating function of $\mathcal{Q}$. By hypothesis, $\Phi(z)$ is finite and analytic in a neighborhood of the closed disk $|z| \leq e^\delta$ for some $\delta > 0$ and, since the variance of $\mathcal{Q}$ is strictly positive, $\Phi(z)$ is strictly convex on $[0, e^\delta]$. Moreover, $\Phi'(1) = 1$ because the offspring distribution has mean 1.

Define

$$G_n(x) = G_n(x; \theta) = E \exp(\theta U_n(x)) - 1.$$

Clearly, $G_n(x; \theta) \to 0$ as $\theta \to 0$. Moreover, by Lemma 11, for each value of $\theta > 0$, the function $G_n(x)$ is maximal at $x = 0$. Since the random variables $U_1(x)$ are zero, except for $x \in \mathcal{N}$, and have the same distribution for $x \in \mathcal{N}$, the function

$G_1(x)$ is, for any fixed θ, a scalar multiple of the uniform distribution P_1 on $\mathcal{N}$. Conditioning on the first generation of the branching random walk shows that

$$G_{n+1}(x) + 1 = \Phi\big(\mathbb{P}G_n(x) + 1\big), \tag{2.8}$$

where $\mathbb{P}$ is the one-step Markov operator for the simple random walk, that is, $\mathbb{P}f(x) = Ef(x+Y)$, where Y is uniformly distributed on $\mathcal{N}$. Since $\Phi(1) = 1$ and $\Phi(z)$ is strictly convex for $z \in [0, e^\delta]$, equation (2.8) implies that

$$G_{n+1}(x) \le \mathbb{P}G_n(x)\Phi'\big(1 + \mathbb{P}G_n(x)\big). \tag{2.9}$$

Unfortunately, both relations (2.8) and (2.9) are nonlinear in G_n. For this reason, we introduce dominating functions $H_n(x) = H_n(x;\theta)$ that satisfy corresponding *linear* relations: set $H_1(x) = G_1(x)$ and define H_n inductively by

$$H_{n+1}(x) = \mathbb{P}H_n(x)\Phi'\big(1 + H_n(0)\big). \tag{2.10}$$

Note that H_{n+1} may take the value $+\infty$ if $H_n(0)$ exceeds the radius of convergence of Φ. Since $G_n(x) \le G_n(0)$, the inequality (2.9) implies that $H_2 \ge G_2$ and so, by induction, that $H_n \ge G_n$ for all $n \ge 1$. Thus, to prove inequalities (2.6) and (2.7), it suffices to prove analogous inequalities for the functions $H_n(x;\theta)$.

The advantage of working with the functions H_n is that the linear relation (2.10) can be iterated. In general, if functions f and g satisfy $g = a\mathbb{P}f$ for some scalar a, then $\mathbb{P}g = a\mathbb{P}^2 f$. Employing this identity in (2.10) and iterating yields

$$H_n(x) = \mathbb{P}^{n-1}H_1(x)\prod_{j=1}^{n-1}\Phi'\big(1 + H_j(0)\big).$$

Because the function $H_1 = G_1$ is itself a scalar multiple of P_1, it follows that

$$H_n(x;\theta) = P_n(x)H_1(0;\theta)(2d+1)\prod_{j=1}^{n-1}\Phi'\big(1 + H_j(0;\theta)\big). \tag{2.11}$$

Since $\Phi'(1) = 1$, the factors in the product are well approximated by $(1 + \Phi''(1) \times H_j(0;\theta))$, as long as $H_j(0;\theta)$ remains small. In particular, for suitable constants $C < \infty$ and $\varepsilon > 0$, if $H_j(0;\theta) < \varepsilon$ for all $j \le n-1$, then

$$H_n(x;\theta) \le (2d+1)P_n(x)H_1(0;\theta)\prod_{j=1}^{n-1}\big(1 + CH_j(0;\theta)\big). \tag{2.12}$$

Taking $x = 0$ yields

$$\frac{H_n(0;\theta)}{\prod_{j=1}^{n}(1 + CH_j(0;\theta))} \le (2d+1)P_n(0)H_1(0;\theta). \tag{2.13}$$

The large-n behavior of the products on the right-hand side of (2.12) will depend on whether or not the sequence $P_n(0)$ is summable, that is, on whether or not the simple random walk is transient. There are two cases to consider.

Dimensions $d \geq 3$. In dimensions $d \geq 3$, the return probabilities $P_n(0)$ are summable. Moreover, when $\theta > 0$ is small, the factor $(2d+1)H_1(0;\theta)$ on the right-hand side of (2.13) is also small because $H_1 = G_1$ is a continuous function of θ that takes the value 0 at $\theta = 0$. Hence, by choosing θ small, we can make the sum over n of the quantities on the right-hand side of inequality (2.13) arbitrarily small. Now, the fraction on the left-hand side of (2.13) is the nth term of the telescoping series

$$C^{-1} \sum \left(\frac{1}{\prod_{j=1}^{n-1}(1 + CH_j(0;\theta))} - \frac{1}{\prod_{j=1}^{n}(1 + CH_j(0;\theta))} \right); \tag{2.14}$$

consequently, (2.13) implies that, for all sufficiently small $\theta > 0$, the products

$$\prod_{j=1}^{n} \bigl(1 + CH_j(0;\theta)\bigr)$$

remain bounded for large n and, for small θ, remain close to 1. It now follows, by (2.12), that for a suitable constant $C' < \infty$ and all small θ, the functions $H_n(x;\theta)$ are all finite and satisfy

$$H_n(x;\theta) \leq C' P_n(x) H_1(0;\theta).$$

Finally, the differentiability of $H_1(0;\theta)$ in θ guarantees that $H_1(0;\theta) \leq C\theta$ for an appropriate constant $C < \infty$ for all small θ. This proves (2.6).

Dimension $d = 2$. It is still the case that the fraction on the left-hand side of (2.13) is the nth term of the telescoping series (2.14), but, since $\sum P_n(0)$ diverges, this no longer implies that the products on the right-hand side of (2.12) remain bounded. However, the local central limit theorem gives an explicit estimate for the partial sums of the return probabilities: in particular, for some $C' \geq A = 5/(4\pi)$,

$$\sum_{j=1}^{n} P_j(0) \leq C \log n \qquad \text{for all } n \geq 2.$$

Consequently, substituting $\theta/\log n$ for θ in inequality (2.13) and summing gives

$$1 - \prod_{j=1}^{n} \bigl(1 + CH_j(0;\theta/\log n)\bigr)^{-1} \leq C''\theta.$$

This, in turn, implies that

$$\prod_{j=1}^{n} \bigl(1 + CH_j(0;\theta/\log n)\bigr) \leq 1/(1 - C''\theta).$$

Using this upper bound for the product on the right-hand side of (2.12) and using the bound $H_1(0;\theta/\log n) \leq C\theta/\log n$ for small θ yields (2.7). $\square$

REMARK 14. In dimensions $d \geq 3$, the conclusion (2.6) cannot be extended to all $\theta > 0$, even for the double-or-nothing case. In fact, for sufficiently large θ, the sums $\sum_{x\in\mathbb{Z}^d}(E\exp\{\theta U_n(x)\} - 1)$ are not bounded in n.

REMARK 15. In dimension $d = 2$, the relation (2.7) does not hold for large θ; see Remark 26 below.

We are now prepared to prove Theorem 3. In fact, we will establish the following, stronger, result.

COROLLARY 16. *Under the hypotheses of Proposition 13, with the same notation*:

(i) *if $d \geq 3$, then for all $\theta \leq \delta_d$,*

$$P\biggl(V_n \geq \frac{\log n}{\theta} \Bigm| G_n\biggr) = O\biggl(\frac{1}{n^{\delta_d/\theta - 1}}\biggr)$$

and, in particular, conditional on G_n, $V_n = O_p(\log n)$;

(ii) *if $d = 2$, then for all $\theta \leq \delta_2$,*

$$P\biggl(V_n \geq \frac{(\log n)^2}{\theta} \Bigm| G_n\biggr) = O\biggl(\frac{1}{n^{\delta_2/\theta} - 1}\biggr)$$

and, in particular, conditional on G_n, $V_n = O_p((\log n)^2)$.

PROOF. We will prove this only for dimensions $d \geq 3$; the dimension $d = 2$ case can be handled similarly. By Markov's inequality,

$$\begin{aligned} P\biggl(V_n \geq \frac{\log n}{\theta} \Bigm| G_n\biggr) &\leq \frac{1}{\exp(\delta_d/\theta \cdot \log n)} \sum_x E\bigl(\exp(\delta_d U_n(x)) \cdot \mathbf{1}_{\{U_n(x)>0\}} \mid G_n\bigr) \\ &= O\biggl(\frac{1}{n^{\delta_d/\theta} - 1} \cdot \sum_x E\bigl(\exp(\delta_d U_n(x)) \cdot \mathbf{1}_{\{U_n(x)>0\}}\bigr)\biggr). \end{aligned}$$

For any random variable $X \geq 0$,

$$\begin{aligned} E\exp(X) &= E\exp(X) \cdot \mathbf{1}_{\{X>0\}} + E\exp(X) \cdot \mathbf{1}_{\{X=0\}} \\ &= E\exp(X) \cdot \mathbf{1}_{\{X>0\}} + P(X = 0) \\ &= E\exp(X) \cdot \mathbf{1}_{\{X>0\}} + 1 - P(X > 0). \end{aligned}$$

Hence,

$$\begin{aligned} &\sum_x E\bigl(\exp(\delta_d U_n(x)) \cdot \mathbf{1}_{\{U_n(x)>0\}}\bigr) \\ &\qquad = \sum_x \bigl(E\exp(\delta_d U_n(x)) - 1\bigr) + \sum_x P\bigl(U_n(x) > 0\bigr) \\ &\qquad \leq \sum_x \bigl(E\exp(\delta_d U_n(x)) - 1\bigr) + 1, \end{aligned}$$

so, by Proposition 13,

$$\sum_x E\big(\exp(\delta_d U_n(x)) \cdot \mathbf{1}_{\{U_n(x)>0\}}\big) \le C \qquad \text{for all } n \ge 1.$$

The conclusion then follows. $\square$

3. Proof of Theorem 4. The proof uses the following elementary lemma, whose proof is left to the reader.

LEMMA 17. *Suppose that on some probability space $(\Omega, \mathcal{F}, P)$, there are two events E_1, E_2 of positive probability such that*

$$\frac{P(E_1 \Delta E_2)}{P(E_1)} \le \varepsilon, \tag{3.1}$$

where $E_1 \Delta E_2$ is the symmetric difference of E_1 and E_2. Then,

$$\|P(\cdot|E_1) - P(\cdot|E_2)\|_{\mathrm{TV}} \le 2\varepsilon, \tag{3.2}$$

where $P(\cdot|E_i)$ denotes the conditional probability measure given the event E_i and $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation distance.

Lemma 17 will allow us to replace the event of conditioning G_n in Theorems 4 and 5 by asymptotically equivalent events of the form

$$H_n = \{Z_{m(n)} \ge n\varepsilon_n\}. \tag{3.3}$$

LEMMA 18. *Let $m(n) < n$ be integers and $\varepsilon_n > 0$ real numbers such that $m(n)/n \to 1$ and $\varepsilon_n \to 0$ as $n \to \infty$. Then,*

$$\lim_{n\to\infty} \frac{P(G_n \Delta H_n)}{P(G_n)} = 0. \tag{3.4}$$

PROOF. This is an easy consequence of Kolmogorov's estimate (1.1) and Yaglom's theorem for critical Galton–Watson processes. Let $K_n = \{Z_{m(n)} \ge 1\}$. Clearly, $H_n \subset K_n$ and so $P(K_n \mid H_n) = 1$. On the other hand, Yaglom's theorem implies that $P(H_n \mid K_n) \to 1$ since $m(n)/n \to 1$. Consequently,

$$\lim_{n\to\infty} \frac{P(H_n \Delta K_n)}{P(K_n)} = 0. \tag{3.5}$$

A similar argument shows that the symmetric difference $K_n \Delta G_n$ is an asymptotically negligible part of K_n. Obviously, $G_n \subset K_n$, so $P(K_n \mid G_n) = 1$. Yaglom's theorem implies that for any $\delta > 0$, there exists $\alpha > 0$ such that

$$P\big(Z_{m(n)} > \alpha n \mid K_n\big) \ge 1 - \delta.$$

However, on the event $\{Z_{m(n)} > \alpha n\}$, the event G_n of survival to generation n is nearly certain for large n because the $Z_{m(n)}$ particles in generation $m(n)$ initiate independent Galton–Watson processes, each of which survives to generation n with probability $\sim 2/(n - m(n))\sigma^2$, by Kolmogorov's estimate (1.1). Hence,

$$P(G_n \mid K_n) \geq 1 - 2\delta$$

for large n. Since $\delta > 0$ is arbitrary, it follows that $P(G_n \mid K_n) \to 1$. By Lemma 17 and (3.5), we get

$$P(G_n \mid H_n) \to 1. \tag{3.6}$$

Furthermore, since $G_n \subset K_n$,

$$\lim_{n\to\infty} \frac{P(G_n \Delta K_n)}{P(K_n)} = 0.$$

By Lemma 17, this implies that conditioning on G_n is asymptotically equivalent to conditioning on K_n and so the difference $P(H_n \mid K_n) - P(H_n \mid G_n) \to 0$. However, we have seen that $P(H_n \mid K_n) \to 1$, hence $P(H_n \mid G_n) \to 1$. This, along with (3.6), implies that (3.4). $\square$

PROOF OF THEOREM 4. The offspring distribution is nondegenerate, so there exists $l_0 > 1$ such that $Q_{l_0} > 0$. Let $p = Q_{l_0} \cdot (1/(2d+1))^{l_0}$ be the probability that the initial particle produces l_0 offspring and these offspring all stay at the origin. Then, for all $k \in \mathbb{N}$,

$$P(U_k(0) \geq l_0^k) \geq p \cdot p^{l_0} \cdot p^{l_0^2} \cdots p^{l_0^{k-1}} \geq p^{l_0^k/(l_0-1)}.$$

Our objective is to show that for some $\delta > 0$, $P(V_n \geq \delta \log n \mid Z_n > 0) \to 1$. By Lemmas 17 and 18, this will follow if we can show that for some $m(n) \leq n$ with $m(n)/n \to 1$ and some $\varepsilon_n \to 0$, the probability

$$P(V_n \geq \delta \log n \mid Z_{m(n)} > n\varepsilon_n) \to 1.$$

To do so, for $\delta > 0$ to be determined later and all n big enough, define k such that $l_0 \delta \log n > l_0^k \geq \delta \log n$ and $m(n) = n - k$. Then, $m(n)/n \to 1$. Fixing a sequence $\varepsilon_n = O(1/\log n)$, we then have

$$\begin{aligned}
&P\left(V_n \geq \delta \log n \,\middle|\, \frac{Z_{m(n)}}{n} \geq \varepsilon_n\right) \\
&\quad \geq 1 - (1 - P(U_k(0) \geq l_0^k))^{\varepsilon_n n} \\
&\quad \geq 1 - (1 - p^{l_0^k/(l_0-1)})^{\varepsilon_n n} \\
&\quad \geq 1 - \exp(\varepsilon_n n(-p^{l_0^k/(l_0-1)})) \\
&\quad \geq 1 - \exp(-\varepsilon_n n p^{l_0 \delta \log n/(l_0-1)}) \\
&\quad = 1 - \exp(-\varepsilon_n n^{1 + l_0 \delta \log p/(l_0-1)}) \to 1,
\end{aligned} \tag{3.7}$$

provided that $\delta < (l_0 - 1)/(-l_0 \log p)$. $\square$

4. Proof of Theorem 5.

4.1. *Strategy.* By Lemmas 17 and 18, the difference between conditioning on the event $G_n = \{Z_n > 0\}$ and conditioning on the event $H_n := \{Z_{m(n)} \ge n\varepsilon_n\}$ is asymptotically negligible if $m(n)/n \to 1$ and $\varepsilon_n \to 0$. Thus, it suffices to prove the weak convergence of the conditional distributions in (1.7) when the conditioning event is H_n rather than G_n. The advantage of this is that, conditional on the state of the branching random walk at time $m(n)$, the next $n - m(n)$ generations are obtained by running *independent* branching random walks for time $n - m(n)$, starting from the locations of the particles in generation $m(n)$. The argument will hinge on showing that if $m(n) < n$ is chosen appropriately, then these independent branching random walks will not overlap much at time n and so the total number $M_n(j)$ of multiplicity-j sites will be, approximately, the sum of $Z_{m(n)}$ independent copies of $M_{n-m(n)}(j)$.

4.2. *Overlapping.*

LEMMA 19. *Suppose that a critical branching random walk starts at time* 0 *with two particles* u, v *located at sites* $x_u, x_v \in \mathbb{Z}^d$, *respectively. Let* $D_n(u, v)$ *be the number of particles in generation* n *located at sites with descendants of both* u *and* v. *There then exists* $C > 0$ *such that for all generations* $n \ge 1$,

$$ED_n(u, v) \le 2P_{2n}(x_v - x_u) \le C\big(1/\sqrt{n}\big)^d. \tag{4.1}$$

PROOF. Denote by $U_n^\zeta(x)$ the number of descendants of particle ζ at site x in generation n. Since the progeny of particles u and v make up mutually independent branching random walks, the random variables $U_n^u(x)$ and $U_n^v(x)$ are independent. However,

$$\begin{aligned} ED_n(u, v) &= E \sum_{x \in \mathbb{Z}^d} \big(U_n^u(x) + U_n^v(x)\big)\mathbf{1}_{\{U_n^u(x) \ge 1\}}\mathbf{1}_{\{U_n^v(x) \ge 1\}} \\ &= 2 \sum_{x \in \mathbb{Z}^d} EU_n^u(x)\mathbf{1}_{\{U_n^v(x) \ge 1\}} \\ &\le 2 \sum_{x \in \mathbb{Z}^d} P_n(x - x_u)P_n(x - x_v) \\ &= 2P_{2n}(x_v - x_u) \\ &\le C\big(1/\sqrt{n}\big)^d. \end{aligned}$$

□

COROLLARY 20. *Let* $Y_{n;m}$ *be the number of particles in generation* n *located at sites with descendants of at least two distinct particles of generation* $m < n$. *Then*,

$$E(Y_{n;m} \mid \mathcal{F}_m) \le CZ_m^2/(n - m)^{d/2}. \tag{4.2}$$

4.3. *Convergence of means.*

PROPOSITION 21. *In dimensions $d \geq 3$,*

$$\lim_n EM_n(j) \triangleq \kappa_j \qquad \textit{exists for every } j \geq 1 \tag{4.3}$$

and

$$\sum_{j=1}^{\infty} j \cdot \kappa_j = 1. \tag{4.4}$$

PROOF. The random variable $M_n(j)$ counts the number of multiplicity-j sites in generation n. The particles at such a site will either all be descendants of a common first generation particle or not; hence, by conditioning on the first generation of the branching random walk, we may decompose $M_{n+1}(j)$ as

$$M_{n+1}(j) = \sum_{i=1}^{Z_1} M_n^i(j) + A_{n+1}(j) - B_{n+1}(j), \tag{4.5}$$

where: (a) the random variables $\{M_n^i(j)\}_{i \leq Z_1}$ are independent copies of $M_n(j)$; (b) the error term $A_{n+1}(j)$ is the number of multiplicity-j sites at time $n+1$ with descendants of different particles in generation 1; (c) the correction $B_{n+1}(j)$ equals

$$\sum_{x \in \mathcal{M}_{n+1}(j+)} \# \text{ particles in generation 1}$$
$$\text{with exactly } j \text{ descendants at } x \text{ in generation } n+1,$$

where $\mathcal{M}_{n+1}(j+)$ is the set of sites with $j+1$ or more particles in generation $n+1$. Obviously, $A_{n+1}(1) = 0$ because a site with only one particle cannot have descendants of distinct first generation particles and so it follows that $EM_{n+1}(1) \leq EM_n(1)$. This implies that $\lim_n EM_n(1)$ exists.

To see that $\lim_{n\to\infty} EM_n(j)$ exists for $j \geq 2$, observe that both $A_{n+1}(j)$ and $B_{n+1}(j)$ are bounded by the number of $(n+1)$th generation particles at sites with descendants of different particles of generation 1. Hence, by Lemma 19, writing $\mathcal{Z}(1) = \mathcal{Z}_1$ for the first generation of the branching process, we have

$$\begin{aligned} E\big(A_{n+1}(j) + B_{n+1}(j)\big) &\leq 2E \sum_{u,v \in \mathcal{Z}(1)} D_n(u,v) \\ &\leq 2 \sum_{l=2}^{\infty} Q_l \binom{l}{2} C n^{-d/2} \qquad (4.6)\\ &\leq C' n^{-d/2} \end{aligned}$$

for some $C' < \infty$ because the offspring distribution has finite second moment. Consequently, by (4.5),

$$|EM_{n+1}(j) - EM_n(j)| = O(n^{-d/2}).$$

Since the sequence $n^{-d/2}$ is summable for $d \geq 3$, the sequence $\{EM_n(j)\}_{n\geq 1}$ must converge. This proves the convergence of means (4.3).

Clearly, for each $n \geq 1$, it is the case that $\sum_j jEM_n(j) = EZ_n = 1$. Hence, to prove the equation (4.4), it suffices to show that for every $\varepsilon > 0$, there exists an integer $k = k(\varepsilon)$ such that for all $n \geq 1$,

$$EY_n(k) \leq \varepsilon \qquad \text{where } Y_n(k) = \sum_{j=k}^{\infty} j \cdot M_n(j) \tag{4.7}$$

is the number of particles in generation n located at sites with at least $k-1$ other particles. Since $Y_n(k) \leq Z_n I\{Z_n \geq k\}$ and $EZ_n = 1$, it is certainly the case that for any fixed $n \geq 1$ and $\varepsilon > 0$, there exists $k = k(n; \varepsilon)$, so inequality (4.7) holds; the problem is to prove that $k(\varepsilon)$ can be chosen independently of n. By the same reasoning as in relation (4.5) above, for all $n, k \geq 1$,

$$Y_{n+1}(k) = \sum_{u \in \mathcal{Z}(1)} Y_n^u(k) + C_{n+1}(k), \tag{4.8}$$

where the random variables $Y_n^u(k)$ are independent copies of $Y_n(k)$ and the error term $C_{n+1}(k)$ is bounded by the total number of particles in generation $n+1$ at sites with descendants of at least two distinct particles in $\mathcal{Z}(1)$. Since $EZ_1 = 1$, the decomposition (4.8) implies that

$$|EY_{n+1}(k) - EY_n(k)| \leq EC_{n+1}(k).$$

However, by the same logic as in relation (4.6) above, there exists $C' < \infty$, independent of k and n, such that $EC_{n+1}(k) \leq C'n^{-d/2}$ for all $n, k \geq 1$. It follows that for sufficiently large $n(\varepsilon)$ and all $k \geq 1$,

$$\sum_{n=n(\varepsilon)}^{\infty} EC_{n+1}(k) < \varepsilon.$$

Thus, if for some $k \geq 1$ and $n = n(\varepsilon)$, the inequality (4.7) holds, then $EY_n(k) < 2\varepsilon$ for all $n \geq n(\varepsilon)$. This proves (4.4). $\square$

REMARK 22. Since the error term $C_{n+1}(k)$ in (4.8) is nonnegative, the expectations $EY_n(k)$ are nondecreasing in n. Because the offspring distribution is nondegenerate, for every $k \geq 1$, there exists $n \geq 1$ such that $Y_n(k) \geq 1$ with positive probability, which forces $EY_n(k) > 0$. Therefore, there are infinitely many integers $j \geq 1$ such that $\kappa_j > 0$.

4.4. *Conditional weak convergence*: *Proof of Theorem* 5. In view of Kolmogorov's estimate (1.1), the inequality (4.7) can be rewritten as

$$E\biggl(\sum_{j\geq k} jM_n(j) \Bigm| G_n\biggr) \leq Cn\varepsilon$$

for some constant $C < \infty$ not depending on n. Since $\Omega_n = \sum_j M_n(j)$, it follows that to prove Theorem 5, it suffices to prove that for any finite $k \geq 1$,

$$\mathcal{L}\biggl(\biggl\{\frac{M_n(j)}{n}\biggr\}_{1\leq j\leq k} \Bigm| G_n\biggr) \Longrightarrow \mathcal{L}(\{\kappa_j Y\}_{1\leq j\leq k}), \tag{4.9}$$

where Y is exponentially distributed with mean $2/\sigma^2$. For this, we will use Yaglom's theorem, the convergence of moments (4.3) and a crude bound on the variance of $M_n(j)$:

$$\operatorname{Var}(M_n(j)) \leq EZ_n^2 = 1 + n\sigma^2. \tag{4.10}$$

Fix $1 \leq m < n$ and, for each particle $u \in \mathcal{Z}_m$, let $M_{n-m}^u(j)$ be the number of sites that have exactly j descendants of particle u in generation n. The random variables $M_{n-m}^u(j)$ are, conditional on $\mathcal{F}_m$, independent copies of $M_{n-m}(j)$. Now, $M_n(j)$ decomposes as

$$M_n(j) = \sum_{u\in\mathcal{Z}(m)} M_{n-m}^u(j) + R_{n;m} \triangleq M_{n;m}^*(j) + R_{n;m}, \tag{4.11}$$

where the remainder $R_{n;m}$ is bounded, in absolute value, by the number of particles in generation n located at sites with descendants of at least two distinct particles of generation $m < n$. By Corollary 20,

$$E(|R_{n;m}| \mid \mathcal{F}_m) \leq CZ_m^2/(n-m)^{d/2}. \tag{4.12}$$

By Yaglom's theorem, the conditional distribution of Z_m/m, given the event G_m of survival to generation m, converges to the exponential distribution with mean $2/\sigma^2$. Thus, if $m = m(n)$ is chosen so that $m/n \to 1$ and $n - m > n^{2/(d-\varepsilon)}$ for some $\varepsilon > 0$, then the bound in (4.12) will be of order $o_P(n)$. In view of (4.11) and Lemmas 17 and 18, it follows that to prove (4.9), it suffices to prove the corresponding statement in which the random variables $M_n(j)$ are replaced by the approximations $M_{n;m}^*(j)$ in (4.11) and the conditioning events G_n are replaced by the events $H_n = \{Z_m \geq \varepsilon_n n\}$. However, this follows routinely by first and second moment estimates: if the scalars ε_n are chosen so that $\varepsilon_n \to 0$ but $n\varepsilon_n/(n-m) \to \infty$, then, by relation (4.3) and the variance bound (4.10), we have

$$E\biggl(Z_m^{-1} \sum_{u\in\mathcal{Z}(m)} M_{n-m}^u(j) \Bigm| \mathcal{F}_m\biggr)\mathbf{1}_{H_n} \longrightarrow \kappa_j \mathbf{1}_{H_n}$$

and

$$\operatorname{Var}\biggl(Z_m^{-1} \sum_{u\in\mathcal{Z}(m)} M_{n-m}^u(j) \Bigm| \mathcal{F}_m\biggr)\mathbf{1}_{H_n} \leq \mathbf{1}_{H_n}(1 + (n-m)\sigma^2)/Z_m \longrightarrow 0.$$

Chebyshev's inequality now implies that the conditional distribution of $M^*_{n;m}(j)/Z_m$ given H_n is concentrated in a vanishingly small neighborhood of κ_j as $n \to \infty$. Since the conditional distribution of Z_m/n given H_n converges to the exponential distribution with mean $2/\sigma^2$ by Yaglom's theorem and Lemmas 17 and 18, the desired result follows.

5. Typical sites in dimension 2: Proof of Theorem 6.

5.1. *Embedded Galton–Watson tree.* For simplicity, we shall consider only the binary fission case, that is, the special case where the offspring distribution is the double-or-nothing distribution $Q_0 = Q_2 = 1/2$. The arguments can all be easily adapted to the general case, at the expense of notational complexity.

We begin with the simple observation that the branching random walk can be constructed by first generating a Galton–Watson tree τ according to the given offspring distribution, then *independently* attaching to the edges of this tree random steps, distributed uniformly on the set $\mathcal{N}$ of nearest neighbors of the origin. The vertices of τ at height n represent the particles of generation n; the location in $\mathbb{Z}^2$ of a particle α of the nth generation is obtained by summing the random steps on the edges of the path in τ leading from the root to α. Henceforth, we will distinguish between the underlying Galton–Watson tree τ and the *marked* tree τ^* obtained by attaching step variables to the edges of τ. Observe that the conditional distribution of the marks of τ^* given the tree τ is the product uniform measure on $\mathcal{N}^{\mathcal{E}(\tau)}$, where $\mathcal{E}(\tau)$ denotes the set of edges of τ.

A *typical particle* of the nth generation in a branching random walk conditioned to survive to the nth generation can be obtained by first choosing a tree τ randomly according to the conditional distribution F_n of the Galton–Watson tree given the event of survival to generation n, then randomly selecting one of the $Z_n \geq 1$ vertices at height n. For this random choice, we assume that the underlying probability space supports a Uniform-[0, 1] random variable γ independent of all other random variables used in the construction of the branching random walk. Since this procedure does not use information about the step variables attached to the edges of the tree, it follows directly that the trajectory of the typical particle, conditional on the underlying Galton–Watson tree, is a simple random walk started at the origin.

5.2. *Reduction to the size-biased case.* The strategy of the proof of Theorem 6 will be based on a change of measure. Denote by $P_H = P_H^n$ the probability measure that is absolutely continuous relative to P with Radon–Nikodym derivative

$$\frac{dP_H}{dP} = Z_n. \tag{5.1}$$

The measure P_H^n thus defined is a probability measure because $EZ_n = 1$. We call it the *size-biased* measure. In the arguments below, the value of n will be fixed, so we will generally omit the dependence of the measure on n and write $P_H = P_H^n$. Because the Radon–Nikodym derivative depends only on the underlying Galton–Watson tree τ, which, under P, is independent of the marks, it follows that the conditional distribution under P_H of the marks given the tree τ is the same as under P. Thus, to construct a version of the marked tree τ^* under P_H, one may first build a size-biased version of the underlying Galton–Watson tree, then attach edge marks independently according to the (product) distribution on $\mathcal{N}$. Henceforth, we will call such a marked tree a *size-biased marked tree* or a *size-biased branching random walk*.

Observe that P_H is also absolutely continuous relative to the *conditional* distribution P_n^* of P given the event G_n of survival to generation n; the Radon–Nikodym derivative is

$$\frac{dP_H}{dP_n^*} = Z_n \pi_n, \tag{5.2}$$

where $\pi_n = P(G_n) \sim (2/n\sigma^2)$. By Yaglom's theorem, under $P_n^* = P(\cdot|G_n)$, the distribution of dP_H^n/dP_n^* converges in law to the unit exponential distribution. This implies the following.

LEMMA 23. *To prove Theorem 6, it suffices to prove the analogous statements for the measure P_H, that is, to prove that*: (i) *for each $\varepsilon > 0$, there exists $K < \infty$ such that*

$$P_H\{T_n \geq K \log n\} < \varepsilon; \tag{5.3}$$

and (ii) *for all sufficiently small $\varepsilon > 0$, there exists $\delta > 0$ such that for all large n*,

$$P_H\{T_n \geq \varepsilon \log n\} \geq \delta. \tag{5.4}$$

PROOF. This is a direct consequence of the fact that the Radon–Nikodym derivatives dP_H/dP^* converge in law under P^* as $n \to \infty$ because this implies that the Radon–Nikodym derivatives dP^*/dP_H converge in law under P_H. $\square$

5.3. *Structure of the size-biased process.* The size-biased measure P_H on marked trees is especially well suited to studying typical points and has been used by a number of authors [see Lyons, Pemantle and Peres (1995) and the references therein] for similar purposes. First, consider the distribution of the *unmarked* genealogical tree τ under P_H. According to Lyons, Pemantle and Peres (1995), a version of this random tree can be obtained by running a certain *Galton–Watson process with immigration*. In the case of the double-or-nothing offspring distribution, the nature of this process is especially simple.

Procedure SB. Each generation j has a single distinguished particle v_j which gives rise to two particles in generation $j+1$, one being the distinguished particle v_{j+1}, the other being an undistinguished particle. All undistinguished particles reproduce according to the double-or-nothing law. For each n, the distinguished particle v_n is uniformly distributed on the particles in generation n.

Thus, a version of the size-biased branching random walk, together with a randomly chosen point v_n of the nth generation, can be built by attaching independent step random variables to the edges of the random tree built according to Recipe SB. Equivalently, this process can be constructed using three independent sequences of auxiliary random variables and a Bernoulli random variable $B_0 \sim \text{Bernoulli}(1/(2d+1))$:

(T_a) $\{S_n\}_{n\geq 0}$ is a simple random walk in $\mathbb{Z}^2$ with initial point $S_0 = 0$;
(T_b) $\{\xi_i\}_{i\geq 0}$ are independent and uniformly distributed on $\mathcal{N}$;
(T_c) $\{U_n^i(x)\}_{i\geq 0}$ are independent copies of the branching random walk $\{U_n(x)\}$ run according to the law P.

[We emphasize that the auxiliary branching random walks $\{U_n^i(x)\}_{i\geq 1}$ are run according to the original probability measure P, not the size-biased measure P_H.] The size-biased branching random walk is obtained by letting the "typical" particle follow the trajectory S_j, then attaching an additional particle to each point (j, S_j) visited by the typical particle, letting it make a step to $S_j + \xi_j$ and then attaching the jth copy of the branching random walk U^j to this particle.

COROLLARY 24. *The distribution of T_n under the size-biased measure P_H is the same as the distribution under P of the random variable*

$$T_n^* = 1 + B_0 + \sum_{j=0}^{n-2} U_{n-j-1}^j (S_n - S_j - \xi_j). \tag{5.5}$$

The Bernoulli random variable B_0 accounts for the possibility that the sibling of the typical particle jumps to the same site as the typical particle.

Reversing the random walk will not affect the distribution of the random variable T_n since the random walk is independent of all other component variables of the representation (5.5), nor will reversing the indices of the auxiliary branching random walks U^j. Thus, the following random variable has the same distribution as that given by (5.5):

$$T_n^{**} = 1 + B_0 + \sum_{j=2}^{n} U_{j-1}^{j-1}(S_j + \xi_{j-1}). \tag{5.6}$$

5.4. *Variances of the occupation random variables.* Next, we focus on the distribution of the random variable T_n^{**} defined by (5.6). To obtain concentration results for this distribution, we will need bounds on the second moments of the random variables $U_n(x)$; for this, we use an exact formula for the second moment of $U_n(x)$, valid in all dimensions.

PROPOSITION 25.

$$EU_n(x)^2 = P_n(x) + \sigma^2 \sum_{i=0}^{n-1} \sum_z P_i(z) P_{n-i}^2(x-z). \tag{5.7}$$

PROOF. This is a special case of equation (81) in Lalley (2009), which gives the mth moment for all integers $m \geq 1$. In the case $m = 2$, a simple proof can be given by conditioning on the first generation of the branching random walk. Set $f_n(x) = EU_n(x)^2$ and $g_n(x) = P_n(x)^2$. Conditioning on generation 1 then gives

$$f_n(x) = \mathbb{P} f_{n-1}(x) + \sigma^2 g_n(x).$$

Since the operator $\mathbb{P}$ is linear, this relation can be iterated $n-1$ times, yielding

$$f_n(x) = \mathbb{P}^{n-1} f_1(x) + \sigma^2 \sum_{i=0}^{n-2} \mathbb{P}^i g_{n-i}(x).$$

This is equivalent to the identity (5.7). $\square$

REMARK 26. If the offspring distribution has an exponentially decaying tail, then one can deduce from (2.7) that $\sum_x EU_n(x)^2 \leq C \log n/\theta$. However, formula (5.7) implies that $\sum_x EU_n(x)^2$ grows at rate $\log n$, so (2.7) cannot hold for large θ.

5.5. *Mean and variance estimates for T_n^{**}.* The sum in the representation (5.6) can be decomposed as $\Gamma_n + \Delta_n$, where

$$\Gamma_n := \sum_{i=2}^{n} P_i(S_i) \tag{5.8}$$

and

$$\Delta_n := \sum_{i=2}^{n} X_{i-1} \qquad \text{with } X_{i-1} := U_{i-1}^{i-1}(S_i + \xi_{i-1}) - P_i(S_i). \tag{5.9}$$

LEMMA 27. *Let S_n be simple random walk in $\mathbb{Z}^2$ and let Γ_n be defined by* (5.8). *Then,*

$$\lim_{n\to\infty} \frac{E\Gamma_n}{\log n} = \frac{A}{2} \tag{5.10}$$

and

$$\lim_{n\to\infty} \operatorname{Var}\left(\frac{\Gamma_n}{\log n}\right) = 0. \tag{5.11}$$

Recall that $A = 5/(4\pi)$ is the constant such that $P_n(0) \sim A/n$.

PROOF OF LEMMA 27. By the symmetry of the simple random walk, $EP_i(S_i) = P_{2i}(0) \sim A/(2i)$ and so the first convergence (5.10) follows routinely. To estimate the variance, first observe that

$$\begin{aligned} E\Gamma_n^2 &= 2\sum_{i<j} EP_i(S_i)P_j(S_j) + \sum_{i=2}^{n} EP_i(S_i)^2 \\ &= 2\sum_{i<j} EP_i(S_i)P_j(S_j) + O(1). \end{aligned} \tag{5.12}$$

The second equation follows from the local central limit theorem in $d = 2$, which guarantees that $P_i(z) \le C/i$ for some constant $C < \infty$ independent of i and z. Next, observe that for $i < j$, by the symmetry of the random walk and the fact that $P_i(z)$ is maximal at $z = 0$ (Lemma 11),

$$\begin{aligned} EP_i(S_i)P_j(S_j) &= E(E(P_i(S_i)P_j(S_j)|S_i)) \\ &= EP_i(S_i) \sum_{x\in\mathbb{Z}^2} P_j(S_i + x)P_{j-i}(x) \\ &= EP_i(S_i)P_{2j-i}(S_i) \\ &\le EP_i(S_i)P_{2j-i}(0) \\ &= P_{2i}(0)P_{2j-i}(0). \end{aligned} \tag{5.13}$$

Substituting this bound in (5.12) and applying the local central limit theorem [in the form $P_n(0) \sim A/n$] yields

$$\begin{aligned} \sum_{i<j} EP_i(S_i)P_j(S_j) &\le \sum_{j=2}^{n}\sum_{i<j} P_{2i}(0)P_{2j-i}(0) \\ &\le 2\sum_{j=2}^{n}\sum_{i<j} A^2/\bigl(2i(2j-i)\bigr) + \text{error} \\ &\sim \frac{A^2}{4}\log^2 n + \text{error}, \end{aligned}$$

where the error is of a smaller order of magnitude. Together with (5.12) and (5.10), this shows that

$$\operatorname{Var}(\Gamma_n) = E\Gamma_n^2 - (E\Gamma_n)^2 = o(\log n)^2. \qquad \square$$

LEMMA 28. *Let S_n, $U_n^i(x)$ and ξ_i be independent sequences of random variables satisfying the hypotheses (T_a)–(T_c) of Section 5.3. If Δ_n and X_i are defined as in (5.9), then*

$$EX_i = 0 \quad \text{and} \quad EX_iX_j = 0 \qquad \text{for all } i \neq j. \tag{5.14}$$

Consequently,

$$E\Delta_n = 0 \quad \text{and} \quad \lim_{n\to\infty} \operatorname{Var}\left(\frac{\Delta_n}{\log n}\right) = \frac{A^2}{8}. \tag{5.15}$$

PROOF. To show that $E\Delta_n = 0$, it suffices to show that $EX_i = 0$. This follows from the fundamental relation (1.9) by conditioning on S_{i+1} and ξ_i:

$$\begin{aligned} EX_i &= EE(U_i^i(S_{i+1} + \xi_i) \mid S_{i+1}, \xi_i) - EP_{i+1}(S_{i+1}) \\ &= EP_i(S_{i+1} + \xi_i) - EP_{i+1}(S_{i+1}) = 0. \end{aligned}$$

Now, consider the covariances EX_iX_j. To compute these expectations for $i < j$, condition on the random variables $S_{i+1}, S_{j+1}, \{U_i^i(x)\}_{x\in\mathbb{Z}^2}$ and ξ_i (but not ξ_j) and use the fundamental identity (1.9). This implies that $EU_j(x + \xi_j) = P_{j+1}(x)$ for each $x \in \mathbb{Z}^2$ and so

$$\begin{aligned} EX_iX_j &= EE(X_iX_j|\cdot) \\ &= EX_iE(U_j^j(S_{j+1} + \xi_j) - P_{j+1}(S_{j+1})|\cdot) \\ &= EX_i \cdot 0 = 0. \end{aligned}$$

It follows that the variance of the sum Δ_n is the sum of the variances of the increments X_i and so

$$\begin{aligned} \operatorname{Var}(\Delta_n) &= \sum_{i=2}^n EX_{i-1}^2 = \sum_{i=2}^n EX_{i-1}^2 \\ &= \sum_{i=2}^n (EU_{i-1}^{i-1}(S_i + \xi_{i-1})^2 - EP_i(S_i)^2) \\ &= \sum_{i=2}^n EU_{i-1}^{i-1}(S_i + \xi_{i-1})^2 + O(1). \end{aligned}$$

Now, by the second moment formula (5.7),

$$\begin{aligned} &EU_{i-1}^{i-1}(S_i + \xi_{i-1})^2 \\ &\quad = E\left(P_{i-1}(S_i + \xi_{i-1}) + \sum_{j=1}^{i-1} \sum_z P_j(z)^2 P_{i-j-1}(S_i + \xi_{i-1} - z)\right) \end{aligned}$$

$$= EP_i(S_i) + \sum_{j=1}^{i-1} \sum_z P_j(z)^2 \cdot EP_{i-j}(S_i - z)$$

$$= P_{2i}(0) + \sum_{j=1}^{i-1} \sum_z P_j(z)^2 \cdot P_{2i-j}(z).$$

The first term is of order $O(1/i)$. To estimate the second, observe that by the local central limit theorem, for large j,

$$P_j(z)^2 \sim \frac{A}{2j} P_{[j/2]}(z),$$

where $[\cdot]$ denotes integer part and the relation holds uniformly for $|z| \leq C\sqrt{j}$. Consequently, for large i,

$$\begin{aligned}\sum_{j=1}^{i-1} \sum_z P_j(z)^2 \cdot P_{2i-j}(z) &\sim \sum_{j=1}^{i-1} \frac{A}{2j} \sum_z P_{[j/2]}(z) P_{2i-j}(z) \\ &= \sum_{j=1}^{i-1} \frac{A}{2j} P_{2i-j+[j/2]}(0) \\ &\sim \sum_{j=1}^{i-1} \frac{A}{2j} \frac{A}{2i - j/2} \\ &\sim \frac{A^2 \log i}{4i}.\end{aligned}$$

Summing from $i = 1$ to n shows that $\mathrm{Var}(\Delta_n) \sim (A^2/8) \log^2 n$. This proves (5.15). □

5.6. *Proof of Theorem 6: Binary fission case.* By Lemma 23, it suffices to prove assertions (5.3) and (5.4). By Corollary 24, the distribution of T_n under the size-biased measure P_H is identical to the distribution of the random variable $T_n^{**} := 1 + B_0 + \tilde{T}_n$ under P, where T_n^{**} is defined by (5.6). Finally, by Lemmas 27 and 28 (note that $E\Delta_n \Gamma_n = 0$),

$$ET_n^{**} \sim \frac{A}{2} \log n \quad \text{and} \quad \mathrm{Var}(T_n^{**}) \sim \frac{A^2}{8} \log^2 n.$$

The first of these implies, by the Markov inequality, that $T_n^{**} = O_P(\log n)$. This proves assertion (i) of Lemma 23. Assertion (ii) is a consequence of the following elementary lemma [see, e.g., Lawler and Limic (2010), Lemma 12.6.1].

LEMMA 29. *If X is a nonnegative random variable with positive, finite second moment, then for any $\alpha \in [0, 1]$,*

$$P\{X \geq \alpha EX\} \geq (1 - \alpha)^2 (EX)^2 / EX^2. \tag{5.16}$$

6. Clustering in dimension 2: Proof of Theorem 8.

6.1. *Occupied sites in the ball* $B(S_n;\ell_n)$. We consider only the case of binary fission. The proof of Theorem 8 in this case, like that of Theorem 6, is based on the change-of-measure strategy outlined in Section 5.2. In particular, we shall prove the assertions corresponding to statements (A) and (B) of Theorem 8 for the *size-biased* process of Section 5.3. Thus, assume throughout this section that the random variables S_j, U_k^j and ξ_j are as in (T_a), (T_b), (T_c) of Section 5.3. Recall that the size-biased branching random walk is obtained by letting the "typical" particle follow the trajectory S_j, then attaching an additional particle to each point (j, S_j) visited by the typical particle, letting it make a step to $S_j+\xi_j$ and then attaching the jth copy of the branching random walk U^j to this particle. To prove Theorem 8, it suffices to prove the following proposition.

PROPOSITION 30. *Let* $\{\ell_n\}$ *be any sequence of real numbers such that* $\lim_n \ell_n=\infty$ *and* $\lim_n \log \ell_n/\log n=0$. *Let* $B(S_n;\ell_n)$ *be the ball of radius* ℓ_n *centered at* S_n. *Then, for the size-biased branching random walk*:

(A) *the number of unoccupied sites in* $B(S_n;\ell_n)$ *is* $o_P(\ell_n^2)$;
(B) *the number of particles in* $B(S_n;\ell_n)$ *is of order* $O_p(\log n\cdot\ell_n^2)$.

The construction of Section 5.3 shows [cf. formulas (5.5) and (5.6)] that the number of particles at location S_n+x in the nth generation of the size-biased branching random walk is distributed as

$$U_n^{**}(S_n+x):=\delta_0(x)+B_0\cdot\mathbf{1}_{\{|x|\le 1\}}+\sum_{j=1}^{n-1}U_j^j(S_{j+1}+x+\xi_j). \tag{6.1}$$

6.2. *Vacant sites*: *Proof of Theorem* 8(A). The representation (6.1) implies that the probability that the site $x+S_n$ is unoccupied, that is, that $U_n^{**}(x+S_n)=0$, is equal to the probability that none of the branching random walks U_i^i succeeds in placing a particle at location x at time n. Since the attached branching random walks are independent of the random walk trajectory $\{S_i\}_{i\le n}$ and the displacement random variables ξ_i, this probability is

$$P\{\text{site}(S_n+x)\text{vacant}\}=\prod_i\bigl(1-u_i(x+S_{i+1}+\xi_i)\bigr), \tag{6.2}$$

where u_n is the *hitting probability function*

$$u_n(x):=P\{U_n(x)\ge 1\}. \tag{6.3}$$

PROPOSITION 31. *There exists* $C>0$ *such that for all* $n\ge 1$ *and all sites* $x\in\mathbb{Z}^2$,

$$u_n(x)\ge\frac{P_n(x)}{C+A\log n}. \tag{6.4}$$

PROOF. By the fundamental identity, $EU_n(x) = P_n(x)$. By the second moment formula (5.7) of Proposition 25,

$$
\begin{aligned}
EU_n(x)^2 &= P_n(x) + \sum_{i=0}^{n-1} \sum_z P_i(z) P_{n-i}^2(x-z) \\
&\le P_n(x) + \sum_{i=0}^{n-1} \sum_z P_i(z) P_{n-i}(x-z) P_{n-i}(0) \\
&= P_n(x) + P_n(x) \sum_{i=0}^{n-1} P_{n-i}(0) \\
&\le P_n(x)(C + A \log n).
\end{aligned}
\tag{6.5}
$$

Here, we have used the fact (Lemma 11) that $P_{n-i}(x)$ is maximal at the origin $x = 0$, together with a strong form of the local central limit theorem [specifically, the fact that the error in the local limit approximation is of order $O(n^{-2})$, which is summable]. The result (6.4) now follows immediately from the Cauchy–Schwarz inequality $P\{X > 0\} \ge (EX)^2/EX^2$, valid for any nonnegative random variable X. $\square$

The lower bound (6.4) leads easily to a useful *upper* bound for the probability that site x is vacant. Partition the indices $i \le n$ into two sets, the *good* and the *bad* indices, as follows: fix a large constant $\kappa < \infty$ and say that index i is *good* if $|S_{i+1} + \xi_i| \le \kappa\sqrt{i}$ and that i is *bad* otherwise. By the local central limit theorem, there is a constant $C' > 0$, not depending on κ, such that for every good index $i \ge |x|^2$,

$$
P_i(x + S_{i+1} + \xi_i) \ge C' e^{-2\kappa^2}/i. \tag{6.6}
$$

Thus, relations (6.2)–(6.4) and the concavity of the logarithm function imply that for a suitable constant $C'' > 0$ not depending on κ,

$$
P\{\text{site } (S_n + x) \text{ vacant}\} \le \exp\left\{ -C'' e^{-2\kappa^2} \sum_{i \text{ good},\ |x|^2 \le i \le n} \frac{1}{i \log i} \right\}. \tag{6.7}
$$

LEMMA 32. *Let $\{\ell_n\}$ be any sequence of real numbers such that $\lim_n \ell_n = \infty$ and $\lim_n \log \ell_n / \log n = 0$. Then, for every $b > 0$ and every $\varepsilon > 0$, there exists κ sufficiently large such that*

$$
\limsup_n P\left\{ \sum_{i \text{ good},\ \ell_n^2 \le i \le n} \frac{e^{-2\kappa^2}}{i \log i} \le b \right\} < \varepsilon. \tag{6.8}
$$

PROOF. The hypotheses regarding the growth of ℓ_n ensure that

$$L_n := \sum_{i=\ell_n^2}^{n} 1/(i \log i) \longrightarrow \infty.$$

Hence, it suffices to show that for some $0 < \varrho < 1$, if κ is sufficiently large, then

$$P\left\{ \sum_{i \text{ bad}, \ \ell_n^2 \le i \le n} \frac{1}{i \log i} \ge \varrho L_n \right\} < \varepsilon \tag{6.9}$$

for all large n. Recall that an index i is *bad* if $|S_{i+1} + \xi_i| > \kappa\sqrt{i}$. Chebyshev's inequality implies that for any $\varepsilon > 0$, if κ is sufficiently large, then $P\{|S_{i+1} + \xi_i| > \kappa\sqrt{i}\} < \varepsilon^3$. Hence, for large n,

$$E \sum_{\ell_n^2 \le i \le n} \frac{\mathbf{1}_{\{|S_{i+1}+\xi_i| > \kappa\sqrt{i}\}}}{i \log i} \le \varepsilon^3 L_n.$$

It now follows, by the Markov inequality, that

$$P\left\{ \sum_{\ell_n^2 \le i \le n} \frac{\mathbf{1}_{\{|S_{i+1}+\xi_i| > \kappa\sqrt{i}\}}}{i \log i} \ge \varepsilon L_n \right\} \le \varepsilon^2. \tag{6.10}$$

The relations (6.10) clearly imply (6.9) and therefore prove (6.8). □

PROOF OF PROPOSITION 30(A). For any $\varepsilon > 0$, inequality (6.7) and Lemma 32 together imply that for all large n, for any displacement x of magnitude $\le \ell_n$, the probability that site $(x + S_n)$ is vacant is less than 2ε. Therefore, the expected number of vacant sites in the ball $B(S_n; \ell_n)$ given the event G_n is, for large n, no larger than $4\pi\varepsilon\ell_n^2$. Assertion (A) of Theorem 8 follows directly, by the Markov inequality. □

6.3. *Proof of Proposition* 30(B). Assertion (B) of Proposition 30 can be proven in virtually the same manner as Theorem 6. The following is a brief sketch. Define

$$\begin{aligned} W_n := \#\ &\text{particles of generation } n \text{ within distance } \ell_n \text{ of } S_n \\ &\text{in the size-biased branching random walk.} \end{aligned} \tag{6.11}$$

By representation (6.1),

$$W_n = 2 + \sum_{i=1}^{n-1} \sum_{|x| \le \ell_n} U_i^i(x + S_{i+1} + \xi_i), \tag{6.12}$$

where $U_j^i(x)$, S_n and ξ_i satisfy conditions (T_a)–(T_c) of Section 5.3. The distribution of the sum on the right-hand side is analyzed by decomposing it as $\Gamma_n + \Delta_n$, where, now,

$$\begin{aligned}\Gamma_n &:= \sum_{i=2}^{n} \sum_{|x| \le \ell_n} P_i(x + S_i) \quad \text{and} \\ \Delta_n &:= \sum_{i=2}^{n} \biggl(\sum_{|x| \le \ell_n} (U_{i-1}^{i-1}(x + S_i + \xi_{i-1}) - P_i(x + S_i)) \biggr).\end{aligned} \tag{6.13}$$

By calculations similar to those used in proving Lemma 27, one shows that

$$\begin{aligned}\lim_{n\to\infty} E\Gamma_n/(\pi \ell_n^2 \log n) &= A/2; \\ \lim_{n\to\infty} \operatorname{Var}(\Gamma_n)/(\pi \ell_n^2 \log n) &= 0; \\ \lim_{n\to\infty} \operatorname{Var}(\Delta_n)/(\pi \ell_n^2 \log n) &\le A^2/8; \\ E\Delta_n = 0 \qquad &\text{for all } n \ge 1.\end{aligned} \tag{6.14}$$

Given these estimates, one now obtains the desired conclusion, that W_n is of order $O_P(\ell_n^2 \log n)$, by the same simple argument as in Section 5.6.

7. Occupied sites in dimension 2.

7.1. *Hitting probability function.* For simplicity, we consider in this section only the binary fission case; the case of a general offspring distribution with mean 1 and finite variance can be handled similarly. The proof of Theorem 7 will be based on careful analysis of the hitting probability function $u_n(x)$ defined by (6.3) above. The connection with the total number Ω_n of occupied sites at time n is obvious: $E\Omega_n = \sum_x u_n(x)$. Thus, our goal will be to bound the function u_n from *above*. (A good *lower* bound has already been obtained in Proposition 31.) Our main result is the following proposition.

PROPOSITION 33. *There exist constants $C_1, C_2 < \infty$ such that for all $n \ge 2$ and all sites $x \in \mathbb{Z}^2$,*

$$u_n(x) \le \frac{C_1}{n \log n} \exp\biggl(-C_2 \frac{|x|^2}{n}\biggr) \tag{7.1}$$

and, hence, for some $C > 0$, we have that

$$E\Omega_n = \sum_x u_n(x) \le \frac{C}{\log n}. \tag{7.2}$$

Theorem 7 follows as a direct consequence of (7.2) and Kolmogorov's estimate (1.1).

To obtain upper bounds on the function $u_n(x)$, we will exploit the fact that it satisfies a parabolic nonlinear partial difference equation. Recall that $\mathbb{P}$ is the Markov operator for the simple random walk, that is, for any bounded function $w:\mathbb{Z}^2 \to \mathbb{R}$,

$$\mathbb{P}w(x) = \frac{1}{5} \sum_{z-x \in \mathcal{N}} w(z).$$

LEMMA 34. *Assume that the offspring distribution is double-or-nothing. Then, for each $n \geq 0$ and each $x \in \mathbb{Z}^d$,*

$$u_{n+1}(x) = \mathbb{P}u_n(x) - \tfrac{1}{2}(\mathbb{P}u_n(x))^2. \tag{7.3}$$

PROOF. The event $\{U_{n+1}(x) > 0\}$ can only occur if the first generation is nonempty and hence consists of two particles with locations in $\mathcal{N}$. This happens with probability $1/2$. One or both of these particles must then engender a descendant branching random walk that places a particle at site x in its nth generation. Since the two descendant branching random walks are independent, with starting points randomly chosen from $\mathcal{N}$, this happens with probability $2p(1-p)+p^2$, where $p = \mathbb{P}u_n(x)$. $\square$

To extract information from the nonlinear difference equation (7.3), we will use the following standard comparison principle. [Compare, e.g., Proposition 2.1 of Aronson and Weinberger (1975).]

LEMMA 35. *Let $u_n(x)$ and $v_n(x)$ be functions taking values between 0 and 1 that satisfy the conditions*

$$u_{n+1}(x) = \mathbb{P}u_n(x) - \tfrac{1}{2}(\mathbb{P}u_n(x))^2 \tag{7.4}$$

and

$$v_{n+1}(x) \geq \mathbb{P}v_n(x) - \tfrac{1}{2}(\mathbb{P}v_n(x))^2. \tag{7.5}$$

If $v_0(x) \geq u_0(x)$ for all x, then

$$v_n(x) \geq u_n(x) \qquad \text{for all } n \geq 0 \text{ and } x \in \mathbb{Z}^2. \tag{7.6}$$

PROOF. Define $\Delta_n(x) = v_n(x) - u_n(x)$. Then, by the hypotheses (7.4) and (7.5),

$$\Delta_{n+1}(x) \geq \mathbb{P}\Delta_n(x) - \tfrac{1}{2}\big(\mathbb{P}u_n(x) + \mathbb{P}v_n(x)\big)\mathbb{P}\Delta_n(x). \tag{7.7}$$

Since u_n and v_n take values between 0 and 1, so does the average $(\mathbb{P}u_n + \mathbb{P}v_n)/2$. Therefore, (7.7) and the induction hypothesis imply that

$$\Delta_{n+1}(x) \geq \mathbb{P}\Delta_n(x)\bigl(1 - \tfrac{1}{2}(\mathbb{P}u_n(x) + \mathbb{P}v_n(x))\bigr) \geq 0. \qquad \square$$

The trick is to find a function v_n that satisfies inequality (7.5) and dominates u_0. To this end, fix $\kappa > 0$ and define

$$v_n(x) = \frac{\kappa}{n \log n} \exp\left(-\frac{\beta_n |x|^2}{2n}\right), \tag{7.8}$$

where

$$\beta_n = \beta\left(1 - \frac{1}{\log n}\right) \quad \text{and} \quad \beta = 5/2.$$

LEMMA 36. *There exist $N_0 \in \mathbb{N}$ and κ_0* independent of N_0 *such that for all $\kappa \geq \kappa_0$ and $n \geq N_0$,*

$$v_{n+1}(x) \geq \mathbb{P}v_n(x)\bigl(1 - \tfrac{1}{2}\mathbb{P}v_n(x)\bigr). \tag{7.9}$$

The (rather technical) proof is deferred to Section 7.3 below. [See Bramson, Cox and Greven (1993) for a similar argument in the context of the KPP equation.] Given Lemma 36, Proposition 33 is an easy consequence.

COROLLARY 37. *There exist $N_1 \in \mathbb{N}$ and $\kappa > 0$ such that for all $n \geq 0$,*

$$u_n(x) \leq v_{N_1+n}(x) \leq 1. \tag{7.10}$$

PROOF. Choose $N_1 \geq N_0$ such that $\kappa := N_1 \log N_1 \geq \kappa_0$. For such a choice of (N_1, κ) we have

$$u_0(x) = \mathbf{1}_{\{x=0\}} \leq v_{N_1}(x) \leq 1.$$

Moreover, by Lemma 36, the function $\tilde{v}_n(x) := v_{n+N_1}(x)$ satisfies (7.9). The conclusion now follows from the comparison Lemma 35. $\square$

7.2. *Representation of the conditional distribution.* Révész (1996) considers a branching random walk on $\mathbb{R}^d$ that is identical to the branching random walk we have studied, *except* that the particle motion is by Gaussian $N(0, I)$ increments rather than Uniform-$\mathcal{N}$ increments. One of the main results of Révész's article asserts that, conditional on the event that there is at least one particle of the nth generation in the ball B of radius $\varrho = \pi^{-1/2}$ centered at the origin, the expected total number of such particles is of order $\Theta(\log n)$. His argument seems to rest on the (unproved) assertion (see the first two sentences of his *Proof of Theorem* 3) that, conditional on the event that a region C is occupied by at least one particle at time t, the branching random walk consists of a single pinned random walk from which independent branching random walks are thrown. There is no proof of this assertion (in fact, it is not even stated clearly, as far as we can see).

We believe that Révész's assertion is false. The purpose of this section is to give a representation related to that of Révész's for the conditional law of the occupation random variable $U_n(x)$ given the event

$$G_{n,x} := \{U_n(x) > 0\}.$$

This representation is similar to Révész's in that it consists of independent branching random walks thrown from a random path from $(0, 0)$ to (n, x); however, the distribution of the random path is *not* that of a pinned simple random walk, but rather that of a *u-transformed* simple random walk. This is defined as follows.

DEFINITION 38. For each site x and integer $n \geq 1$ such that $u_n(x) > 0$, the *u-transformed simple random walk with endpoint* (n, x) is the n-step, time-inhomogeneous Markov chain $\{X_m\}_{0\leq m\leq n}$ on $\mathbb{Z}^d$ with initial point 0 and transition probabilities

$$q_m(z, y) := P(X_m = y \mid X_{m-1} = z) = P_1(y - z)\frac{u_{n-m}(x - y)}{\mathbb{P}u_{n-m}(x - z)}. \tag{7.11}$$

REMARK 39. Except in the trivial case $n = 1$, a u-transformed random walk is *not* a Doob h-process because the hitting probability function $u_n(x)$ is not space–time harmonic for the simple random walk, by (7.3) above. However, a pinned random walk *is* an h-process; in particular, the one-step transition probabilities of a pinned random walk conditioned to end at x_n are given by

$$q^*_m(z, y) = P_1(y - z)\frac{P_{n-m}(x_n - y)}{P_{n-m+1}(x_n - z)}. \tag{7.12}$$

Since the function $P_{n-m}(z, x_n)$ is space–time harmonic, the transition probabilities q^* are not the same as those of the u-transformed random walk.

LEMMA 40. *If $u_n(x) > 0$, then the u-transformed simple random walk with endpoint (n, x) is well defined and, with probability* 1, *ends at $X_n = x$.*

PROOF. What must be shown is that the Markov chain with transition probabilities (7.11) will visit no states (m, z) at which the denominator $\mathbb{P}u_{n-m}(x - z)$ is zero. This is accomplished by noting that as long as X_{m-1} is at a site z such that $u_{n-m+1}(x - z) > 0$, then, by Lemma 34, the denominator $\mathbb{P}u_{n-m}(x - z) > 0$ and so there is at least one site y among the nearest neighbors of z such that $u_{n-m}(x - y) > 0$. By (7.11), the next state X_m will then be chosen from among the nearest neighbors such that $u_{n-m}(x - y) > 0$. This proves that the Markov chain is well defined. The path ends at $X_n = x$ because 0 is the only site at which $u_0 > 0$. $\square$

Our representation of the conditional distribution of the random variable $U_n(x)$ given the event $G_{n,x}$ requires four mutually independent sequences of random variables:

(U_a) $\{X_m\}_{0\le m\le n}$ is a u-transformed simple random walk with endpoint (n,x);
(U_b) $\{B_m(w)\}_{0\le m<n;w\in\mathbb{Z}^d}$ are independent Bernoulli$(\beta_m(w))$ random variables;
(U_c) $\{U_m^i(y)\}_{i\ge 0}$ are independent copies of the branching random walk $\{U_m(y)\}$;
(U_d) $\{\xi_i\}_{i\ge 0}$ are independent and uniformly distributed on $\mathcal{N}$.

The Bernoulli parameters are

$$\beta_m(w) = \frac{1}{2 - \mathbb{P}u_{n-m-1}(x-w)}; \tag{7.13}$$

note that for large values of $n-m$, the parameters $\beta_m(w)$ are uniformly close to $1/2$ because $u_{n-m}(x-w)$ is bounded by the probability that the branching random walk will survive for $n-m$ generations.

PROPOSITION 41. *Assume that the offspring distribution is double-or-nothing and let x be a site for which $u_n(x)>0$. Then,*

$$\begin{aligned} &\mathcal{L}\big(U_n(x) \mid U_n(x)\ge 1\big) \\ &\quad = \mathcal{L}\left(1 + \sum_{m=0}^{n-1} B_m(X_m) U_{n-m-1}^m (x - X_m - \xi_{m+1})\right). \end{aligned} \tag{7.14}$$

PROOF. The assertion (7.14) is equivalent to the assertion (Claim 42 below) that the conditional distribution can be simulated by the following *Method* A: (1) let a particle ζ execute a u-transformed simple random walk $\{X_m\}_{m\le n}$ with endpoint (n,x); (2) at each location (m,X_m) where $0\le m<n$, toss a $\beta_m(X_m)$-coin to determine whether or not to attach a descendant branching random walk; (3) on the event that the coin toss is a head, create a new particle ζ_m, let it make one jump ξ_{m+1} to a neighboring site and then attach an independent branching random walk starting from this new location; (4) count the total number of particles, including ζ, that land at site x at time n.

CLAIM 42. *This simulates the conditional distribution of the total number of particles at site x in generation n given the event $\{U_n(x)\ge 1\}$.*

This claim is proved by induction on n. The case $n=1$ is routine, but for the reader's convenience we shall present the argument in detail. First, the only sites x such that $u_1(x)>0$ are the nearest neighbors of the origin, so we assume that x is one of these five points. Since $u_0=\delta_0$ is the Kronecker delta function, $\mathbb{P}u_0(x) = 1/5$ and so $\beta_0(0) = 1/(2-1/5) = 5/9$. Now, consider the first generation $\mathcal{Z}_1$ of the branching random walk: this will be empty unless the initial particle fissions, in which case the two offspring are located at randomly chosen nearest neighbors of the origin. Consequently, the unconditional distribution of $U_1(x)$ is

$$P\{U_1(x) = 0\} = \tfrac{1}{2} + \tfrac{1}{2} \times \tfrac{4}{5} \times \tfrac{4}{5},$$
$$P\{U_1(x) = 1\} = \tfrac{1}{2} \times 2 \times \tfrac{4}{5} \times \tfrac{1}{5},$$
$$P\{U_1(x) = 2\} = \tfrac{1}{2} \times \tfrac{1}{5} \times \tfrac{1}{5}.$$

It follows that the *conditional* distribution of $U_1(x)$ given the event $\{U_1(x) > 0\}$ is that of 1 plus a Bernoulli$(1/9)$ random variable. This coincides with the distribution of the random variable produced by Method A because $B_0(0) = 1$ with probability $\beta_0(0) = 5/9$ and, on this event, the particle jumps to x with probability $1/5$, leaving a second particle at x.

Next, consider the branching random walk conditioned to have at least one particle at site x in generation $n \geq 2$. The first generation must consist of two particles, at least one of which produces a descendant branching random walk that places particles at x in its $(n-1)$th generation. Conditional on the event that two particles are produced by the initial particle (i.e., the event $\{Z_1 = 2\}$), each will have chance $p := \mathbb{P}u_{n-1}(x)$ of producing a descendant at site x in generation n; consequently, the conditional probability that *both* particles will do so, given that *at least one* does, is

$$\frac{p^2}{p^2 + 2p(1-p)} = p\beta_{n-1}(0).$$

Moreover, given that either one of the particles produces a particle at site x in generation n, the conditional probability that its first jump is to site $y \in \mathcal{N}$ is

$$P_1(y)\frac{u_{n-1}(x-y)}{\mathbb{P}u_{n-1}(x)}; \tag{7.15}$$

this is the distribution of the first step of a u-transformed random walk with endpoint (n, x). Thus, a version of the random variable $U_n(x)$, conditional on $\{U_n(x) \geq 1\}$, can be produced by the following two-step procedure.

(1) Place a particle η at a randomly chosen neighbor y of 0 according to the distribution (7.15) and attach to it a branching random walk conditioned to produce at least one descendant at site $x - y$ in its $(n-1)$th generation. By the induction hypothesis, the contribution of offspring of η to site x in generation n will be

$$1 + \sum_{m=1}^{n-1} B_m(X_m)U^m_{n-m-1}(x - X_m - \xi_{m+1}). \tag{7.16}$$

(2) With probability $p\beta_{n-1}(0)$, do the same with a second particle τ. Observe that, conditional on the event that this second particle τ is attached, the contribution to site x in the nth generation will have distribution

$$\mathcal{L}\big(U_{n-1}(x - X_1) \mid U_{n-1}(x - X_1) > 0\big).$$

Since the particle τ is attached with probability $p\beta_{n-1}(0)$, where p is the probability that a particle born at time 0 will put a descendant at site x in generation n, step (2) has the same effect as the following alternative: (2′) with probability $\beta_{n-1}(0)$, place a second particle τ at a randomly chosen (i.e., uniformly distributed) neighbor y of 0 and attach an independent copy of the branching random walk. This, together with the representation (7.16) of the number of offspring of η at x in generation n, shows that the total number of particles at x in generation n will be

$$1+\sum_{m=0}^{n-1} B_m(X_m)U^m_{n-m-1}(x-X_m-\xi_{m+1}) \tag{7.17}$$

as desired. This completes the induction argument and thus proves (7.14). $\square$

7.3. *Proof of Lemma* 36. Let $x=(x_1,x_2)$. Then,

$$\mathbb{P}v_n(x)=v_n(x)\cdot e^{-\beta_n/(2n)}\cdot \tfrac{1}{5}w_n(x),$$

where

$$w_n(x)=\bigl(e^{\beta_n/(2n)}+e^{-\beta_n x_1/n}+e^{\beta_n x_1/n}+e^{-\beta_n x_2/n}+e^{\beta_n x_2/n}\bigr).$$

Then,

$$\begin{aligned}
&v_{n+1}(x)-\mathbb{P}v_n(x)+\frac{1}{2}(\mathbb{P}v_n(x))^2\\
&\quad=v_n(x)e^{-\beta_n/(2n)}\frac{1}{5}\biggl(5e^{\beta_n/(2n)}\frac{n\log n}{(n+1)\log(n+1)}\exp\biggl(\frac{\theta_n|x|^2}{2}\biggr)\\
&\qquad\qquad -w_n(x)+\frac{e^{-\beta_n/(2n)}}{10}v_n(x)w_n(x)^2\biggr),
\end{aligned}$$

where

$$\theta_n=\frac{\beta_n}{n}-\frac{\beta_{n+1}}{n+1}.$$

Therefore, it suffices to show that there exist N_0 and κ_0, independent of N_0, such that for all $\kappa\ge\kappa_0$ and all $n\ge N_0$, the following holds:

$$\begin{aligned}
&5e^{\beta_n/(2n)}\frac{n\log n}{(n+1)\log(n+1)}\exp\biggl(\frac{\theta_n|x|^2}{2}\biggr)-w_n(x)\\
&\quad+\frac{e^{-\beta_n/(2n)}}{10}v_n(x)w_n(x)^2\ge 0 \qquad \text{for all } x\in\mathbb{Z}^2.
\end{aligned} \tag{7.18}$$

(a) *Estimate of* $e^{\beta_n/(2n)}\frac{n\log n}{(n+1)\log(n+1)}$. First, we have

$$\begin{aligned}\frac{n\log n}{(n+1)\log(n+1)} &= 1-\frac{(n+1)\log(n+1)-n\log n}{(n+1)\log(n+1)}\\ &= 1-\frac{1}{n+1}-\frac{n\log(1+1/n)}{(n+1)\log(n+1)}\\ &= 1-\frac{1}{n+1}-\frac{1}{(n+1)\log(n+1)}+o\Big(\frac{1}{n^2}\Big)\\ &= 1-\frac{1}{n}-\frac{1}{n\log n}+\frac{1}{n^2}+o\Big(\frac{1}{n^2}\Big).\end{aligned}$$

Therefore, recalling that $\beta_n=\beta(1-1/\log n)$, we have

$$\begin{aligned}&e^{\beta_n/(2n)}\frac{n\log n}{(n+1)\log(n+1)}\\ &\quad=\Big(1+\frac{\beta}{2n}-\frac{\beta}{2n\log n}+\frac{\beta_n^2}{8n^2}+O\Big(\frac{1}{n^3}\Big)\Big)\\ &\qquad\times\Big(1-\frac{1}{n}-\frac{1}{n\log n}+\frac{1}{n^2}+o\Big(\frac{1}{n^2}\Big)\Big)\\ &\quad=1+\frac{(\beta-2)}{2n}-\frac{\beta+2}{2n\log n}+\frac{\beta_n^2-4\beta+8}{8n^2}+o\Big(\frac{1}{n^2}\Big).\end{aligned}$$

Since $\beta_n\to\beta=5/2$, there exists $N_0\in\mathbb{N}$ such that for all $n\ge N_0$,

$$e^{\beta_n/(2n)}\frac{n\log n}{(n+1)\log(n+1)}\ge 1+\frac{\beta-2}{2n}-\frac{\beta+2}{2n\log n}+\frac{2}{8n^2}\ge 1. \tag{7.19}$$

(b) *Estimate of* θ_n. Since $\beta_n=\beta(1-1/\log n)$, we have

$$\begin{aligned}\theta_n &= \beta\Big(\frac{1-1/\log n}{n}-\frac{1-1/\log(n+1)}{n+1}\Big)\\ &= \beta\frac{1-(n+1)/\log n+n/\log(n+1)}{n(n+1)}.\end{aligned}$$

However,

$$\begin{aligned}\frac{n+1}{\log n}-\frac{n}{\log(n+1)} &= \frac{\log(n+1)+n\log(1+1/n)}{\log n\log(n+1)}\\ &= \frac{1}{\log n}+O\Big(\frac{1}{\log n\log(n+1)}\Big),\end{aligned}$$

so it follows that

$$\theta_n=\beta\frac{1-1/\log n+O(1/(\log n\log(n+1)))}{n(n+1)}. \tag{7.20}$$

CLAIM 43. *Enlarging N_0 if necessary, we have that for all $n \geq N_0$,*

$$\beta\theta_n - \frac{\beta_n^2}{n^2} \geq \frac{1}{n^2 \log n}. \tag{7.21}$$

PROOF. Since $\beta_n = \beta(1 - 1/\log n)$,

$$\begin{aligned}
n^2 \cdot \left(\beta\theta_n - \frac{\beta_n^2}{n^2}\right) \\
&= \beta^2\left\{\left(1 - \frac{1}{n+1}\right)\left(1 - \frac{1}{\log n} + O\left(\frac{1}{\log n \log(n+1)}\right)\right)\right. \\
&\qquad \left. - \left(1 - \frac{2}{\log n} + \frac{1}{(\log n)^2}\right)\right\} \\
&= \beta^2\left(\frac{1}{\log n} + o\left(\frac{1}{\log n}\right)\right).
\end{aligned}$$

The relation (7.21) follows since $\beta = 5/2 > 1$. $\square$

(c) *Proof of* (7.18) *for* $|x| \geq 3n$. Equation (7.21) implies, enlarging N_0 if necessary, that for all $n \geq N_0$, $\theta_n \geq 2/n^2$. Hence, when $|x| \geq 3n$,

$$\theta_n |x|^2/2 \geq \beta_n |x_i|/n, \qquad i = 1, 2,$$

and

$$5 \exp\left(\frac{\theta_n |x|^2}{2}\right) \geq w_n(x).$$

The relation (7.18) follows by noting (7.19).

(d) *Estimate of* $w_n(x)$. For $|x|/n$ sufficiently small, a Taylor expansion yields

$$\begin{aligned}
w_n(x) &= e^{\beta_n/(2n)} + (e^{-\beta_n x_1/n} + e^{\beta_n x_1/n}) + (e^{-\beta_n x_2/n} + e^{\beta_n x_2/n}) \\
&= 1 + \frac{\beta}{2n} - \frac{\beta}{2n \log n} + \frac{\beta_n^2}{8n^2} + O\left(\frac{1}{n^3}\right) \\
&\quad + 2 + \frac{\beta_n^2 x_1^2}{n^2} + \frac{\beta_n^4 x_1^4}{12n^4} + O\left(\left(\frac{x_1}{n}\right)^6\right) \\
&\quad + 2 + \frac{\beta_n^2 x_2^2}{n^2} + \frac{\beta_n^4 x_2^4}{12n^4} + O\left(\left(\frac{x_2}{n}\right)^6\right) \\
&= 5 + \left[\frac{\beta}{2n} - \frac{\beta}{2n \log n}\right] + \left[\frac{\beta_n^2}{8n^2} + O\left(\frac{1}{n^3}\right)\right] \\
&\quad + \frac{\beta_n^2 |x|^2}{n^2} + \left[\frac{\beta_n^4 (x_1^4 + x_2^4)}{12n^4} + O\left(\frac{|x|^6}{n^6}\right)\right].
\end{aligned} \tag{7.22}$$

(e) *Estimate of* $e^{\beta_n/(2n)} \cdot \frac{n\log n}{(n+1)\log(n+1)} \cdot \exp(\theta_n|x|^2/2)$. By (7.19), for all $n \geq N_0$,

$$\begin{aligned} &e^{\beta_n/(2n)} \cdot \frac{n\log n}{(n+1)\log(n+1)} \cdot \exp\left(\frac{\theta_n|x|^2}{2}\right) \\ &\geq \left(1 + \frac{\beta-2}{2n} - \frac{\beta+2}{2n\log n} + \frac{2}{8n^2}\right) \cdot \left(1 + \frac{\theta_n|x|^2}{2} + \frac{\theta_n^2|x|^4}{8}\right) \\ &\geq 1 + \left[\frac{\beta-2}{2n} - \frac{\beta+2}{2n\log n}\right] + \frac{2}{8n^2} + \frac{\theta_n|x|^2}{2} + \frac{\theta_n^2|x|^4}{8}. \end{aligned} \tag{7.23}$$

(f) *Their difference.* By (7.23) and (7.22),

$$\begin{aligned} &5e^{\beta_n/(2n)} \cdot \frac{n\log n}{(n+1)\log(n+1)} \cdot \exp\left(\frac{\theta_n|x|^2}{2}\right) - w_n(x) \\ &\geq 5\left[\frac{\beta-2}{2n} - \frac{\beta+2}{2n\log n}\right] - \left[\frac{\beta}{2n} - \frac{\beta}{2n\log n}\right] \\ &\quad + \frac{10}{8n^2} - \frac{\beta_n^2}{8n^2} + O\left(\frac{1}{n^3}\right) \\ &\quad + \left(\beta\theta_n - \frac{\beta_n^2}{n^2}\right)|x|^2 \\ &\quad + \frac{5\theta_n^2|x|^4}{8} - \frac{\beta_n^4(x_1^4 + x_2^4)}{12n^4} + O\left(\frac{|x|^6}{n^6}\right). \end{aligned} \tag{7.24}$$

Since $\beta = 5/2$,

$$5\left(\frac{\beta-2}{2n} - \frac{\beta+2}{2n\log n}\right) - \left(\frac{\beta}{2n} - \frac{\beta}{2n\log n}\right) = -\frac{10}{n\log n} \tag{7.25}$$

and, enlarging N_0 if necessary, we can assume that for all $n \geq N_0$,

$$\frac{10}{8n^2} - \frac{\beta_n^2}{8n^2} + O\left(\frac{1}{n^3}\right) \geq 0. \tag{7.26}$$

Moreover, $\theta_n \sim \beta/n^2$ and it follows that for all n sufficiently large,

$$\frac{5\theta_n^2|x|^4}{8} - \frac{\beta_n^4(x_1^4 + x_2^4)}{12n^4} \geq \left(\frac{5\theta_n^2}{8} - \frac{\beta_n^4}{12n^4}\right) \cdot |x|^4 > \frac{|x|^4}{2n^4}. \tag{7.27}$$

(g) *Proof of* (7.18) *for* $\delta n \geq |x| > \sqrt{10n}$, *where* $\delta > 0$ *is sufficiently small.* By (7.21), when $|x| > \sqrt{10n}$,

$$\left(\beta\theta_n - \frac{\beta_n^2}{n^2}\right)|x|^2 \geq \frac{10}{n\log n}.$$

Hence, by (7.24), (7.25), (7.26) and (7.27), the relation (7.18) holds for x such that $|x| > \sqrt{10n}$ and $|x|/n$ is sufficiently small.

(h) *Proof of* (7.18) *for* $3n \geq |x| \geq \delta n$. By (7.21), for all $n \geq N_0$,

$$\theta_n \geq \frac{\beta_n^2}{n^2\beta} - \frac{1}{\beta n^2 \log n}.$$

Hence, when $|x| \leq 3n$,

$$\exp\left(\frac{\theta_n |x|^2}{2}\right) \geq \frac{\exp(\beta_n^2 |x|^2/(5n^2))}{\exp(|x|^2/(5n^2 \log n))} \geq \exp\left(-\frac{2}{\log n}\right) \cdot \exp\left(\frac{\beta_n^2 |x|^2}{5n^2}\right).$$

By (7.19), to show (7.18), it is sufficient to show that for all n sufficiently large,

$$\begin{aligned}&\left(e^{\beta_n/(2n)} + e^{-\beta_n x_1/n} + e^{\beta_n x_1/n} + e^{-\beta_n x_2/n} + e^{\beta_n x_2/n}\right)\\ &\qquad \leq 5 \exp\left(-\frac{2}{\log n}\right) \cdot \exp\left(\frac{\beta_n^2 |x|^2}{5n^2}\right).\end{aligned}$$

Since $|x| \leq 3n$,

$$\begin{aligned}&\left(1 - \exp\left(-\frac{2}{\log n}\right)\right) \cdot \exp\left(\frac{\beta_n^2 |x|^2}{5n^2}\right)\\ &\qquad \leq \left(1 - \exp\left(-\frac{2}{\log n}\right)\right) \exp\left(\frac{9\beta_n^2}{5}\right) = o(1).\end{aligned}$$

Hence, it suffices to show that

$$\begin{aligned}\liminf_n \inf_{3n \geq |x| \geq \delta n} \biggl\{ 5 \exp\left(\frac{\beta_n^2 |x|^2}{5n^2}\right) &- (1 + e^{-\beta_n x_1/n} + e^{\beta_n x_1/n}\\ &+ e^{-\beta_n x_2/n} + e^{\beta_n x_2/n}) \biggr\} > 0.\end{aligned} \tag{7.28}$$

By elementary calculus,

$$e^{-\beta_n x_1/n} + e^{\beta_n x_1/n} + e^{-\beta_n x_2/n} + e^{\beta_n x_2/n} \leq 2 + e^{-\beta_n |x|/n} + e^{\beta_n |x|/n}$$

and thus

$$\begin{aligned}&5 \exp\left(\frac{\beta_n^2 |x|^2}{5n^2}\right) - (1 + e^{-\beta_n x_1/n} + e^{\beta_n x_1/n} + e^{-\beta_n x_2/n} + e^{\beta_n x_2/n})\\ &\qquad \geq 5 \exp\left(\frac{\beta_n^2 |x|^2}{5n^2}\right) - 3 - e^{-\beta_n |x|/n} - e^{\beta_n |x|/n}.\end{aligned}$$

Relation (7.28) now follows from the simple fact that

$$f(x) := 5e^{x^2/5} - 3 - e^x - e^{-x}$$

is strictly increasing for $x \geq 0$ and equals 0 only when $x = 0$.

(i) *Proof of* (7.18) *when* $|x| \le \sqrt{10n}$. Since $|x| \le \sqrt{10n} = o(n)$, by relations (7.21), (7.24), (7.25), (7.26) and (7.27), we need only show that there exists κ_0 such that if $\kappa \ge \kappa_0$ and $n \ge N_0$, then

$$\frac{e^{-\beta_n/(2n)}}{10} v_n(x) w_n(x)^2 \ge \frac{10}{n \log n}.$$

Since $w_n(x) \ge 5$, when $|x| \le \sqrt{10n}$,

$$\frac{e^{-\beta_n/(2n)}}{10} v_n(x) w_n(x)^2 \ge \frac{25\kappa}{10n \log n} \exp(-6\beta),$$

so κ_0 can be chosen as $4\exp(6\beta)$, which is independent of N_0.

Acknowledgments. We would like to thank Lenya Ryzhik for a helpful suggestion regarding the proof of Proposition 13 and Andrej Zlatos for suggesting the form of the super-solution (7.8). We also thank Michael Wichura for carefully reading an earlier draft and pointing out a mistake in our original proof of Lemma 11. Finally, we are grateful to a referee for some valuable suggestions regarding the exposition and for alerting us to the work of Révész.

REFERENCES

ARONSON, D. G. and WEINBERGER, H. F. (1975). Nonlinear diffusion in population genetics, combustion, and nerve pulse propagation. In *Partial Differential Equations and Related Topics* (*Program*, *Tulane Univ*., *New Orleans*, *La*., 1974). *Lecture Notes in Math*. **446** 5–49. Springer, Berlin. MR0427837

ATHREYA, K. B. and NEY, P. E. (1972). *Branching Processes*. Springer, New York. MR0373040

BRAMSON, M., COX, J. T. and GREVEN, A. (1993). Ergodicity of critical spatial branching processes in low dimensions. *Ann*. *Probab*. **21** 1946–1957. MR1245296

ETHERIDGE, A. M. (2000). *An Introduction to Superprocesses*. *University Lecture Series* **20**. Amer. Math. Soc., Providence, RI. MR1779100

FLEISCHMAN, J. (1978). Limiting distributions for branching random fields. *Trans*. *Amer*. *Math*. *Soc*. **239** 353–389. MR0478375

GEIGER, J. (2000). A new proof of Yaglom's exponential limit law. In *Mathematics and Computer Science* (*Versailles*, 2000) 245–249. Birkhäuser, Basel. MR1798303

LALLEY, S. P. (2009). Spatial epidemics: Critical behavior in one dimension. *Probab*. *Theory Related Fields* **144** 429–469. MR2496439

LAWLER, G. F. and LIMIC, V. (2010). *Random Walk*: *A Modern Introduction*. *Cambridge Studies in Advanced Mathematics* **123**. Cambridge Univ. Press, Cambridge.

LYONS, R., PEMANTLE, R. and PERES, Y. (1995). Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann*. *Probab*. **23** 1125–1138. MR1349164

PERKINS, E. (1989). The Hausdorff measure of the closed support of super-Brownian motion. *Ann*. *Inst*. *H*. *Poincaré Probab*. *Statist*. **25** 205–224. MR1001027

PERKINS, E. A. (1988). A space–time property of a class of measure-valued branching diffusions. *Trans*. *Amer*. *Math*. *Soc*. **305** 743–795. MR924777

RÉVÉSZ, P. (1994). *Random Walks of Infinitely Many Particles*. World Scientific, River Edge, NJ. MR1645302

RÉVÉSZ, P. (1996). Distribution of the particles of a critical branching Wiener process. *Bernoulli* **2** 63–80. MR1394052

SPITZER, F. (1976). *Principles of Random Walks*, 2nd ed. *Graduate Texts in Mathematics* **34**. Springer, New York. MR0388547

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637
USA
E-MAIL: lalley@galton.uchicago.edu

DEPARTMENT OF INFORMATION SYSTEMS
BUSINESS STATISTICS
AND OPERATIONS MANAGEMENT
HONG KONG UNIVERSITY
OF SCIENCE AND TECHNOLOGY
CLEAR WATER BAY
KOWLOON
HONG KONG
E-MAIL: xhzheng@ust.hk