Voter Model Perturbations and Reaction Diffusion Equations

J. Theodore Cox*, Syracuse U.
Richard Durrett†, Duke U.
and Edwin A. Perkins‡, U. British Columbia

March 8, 2012

Abstract

We consider particle systems that are perturbations of the voter model and show that when space and time are rescaled the system converges to a solution of a reaction diffusion equation in dimensions $d \geq 3$. Combining this result with properties of the PDE, some methods arising from a low density super-Brownian limit theorem, and a block construction, we give general, and often asymptotically sharp, conditions for the existence of non-trivial stationary distributions, and for extinction of one type. As applications, we describe the phase diagrams of four systems when the parameters are close to the voter model: (i) a stochastic spatial Lotka-Volterra model of Neuhauser and Pacala, (ii) a model of the evolution of cooperation of Ohtsuki, Hauert, Lieberman, and Nowak, (iii) a continuous time version of the non-linear voter model of Molofsky, Durrett, Dushoff, Griffeath, and Levin, (iv) a voter model in which opinion changes are followed by an exponentially distributed latent period during which voters will not change again. The first application confirms a conjecture of Cox and Perkins [8] and the second confirms a conjecture of Ohtsuki et al [41] in the context of certain infinite graphs. An important feature of our general results is that they do not require the process to be attractive.

AMS 2000 *subject classifications.* Primary 60K35. Secondary 35K57, 60J68, 60F17, 92D15, 92D40.

Keywords and phrases. Interacting particle systems, voter model, reaction diffusion equation, evolutionary game theory, Lotka-Volterra model.

*Supported in part by NSF Grant DMS-0505439 Part of the research was done while the author was visiting The University of British Columbia

†Partially supported by NSF Grant DMS-0704996 from the probability program

‡Supported in part by an NSERC Discovery Grant

1 Introduction and Statement of Results

We first describe a class of particle systems, called voter model perturbations in [7]. The state space will be $\{0,1\}^{\mathbb{Z}^d}$, where throughout this work we assume $d \geq 3$. The voter model part of the process will depend on a symmetric (i.e, $p(x) = p(-x)$), irreducible probability kernel $p : \mathbb{Z}^d \to [0,1]$ with $p(0) = 0$, covariance matrix $\sigma^2 I$, and exponentially bounded tails so that for some $\kappa \in (0,1]$,

$$p(x) \leq Ce^{-\kappa|x|}. \tag{1.1}$$

Here and in what follows $|x| = \sup_i |x_i|$. Let $f_i(x,\xi) = \sum_{y\in\mathbb{Z}^d} p(y-x)1(\xi(y) = i)$. The voter flip rates are given by

$$c^v(x,\xi) = [(1-\xi(x))f_1(x,\xi) + \xi(x)f_0(x,\xi)]. \tag{1.2}$$

The processes of interest, $\xi_t \in \{0,1\}^{\mathbb{Z}^d}$, are spin-flip systems with rates

$$c^o_\varepsilon(x,\xi) = c^v(x,\xi) + \varepsilon^2 c^*_\varepsilon(x,\xi) \geq 0, \tag{1.3}$$

where $c^*_\varepsilon(x,\xi)$ is a translation invariant, signed perturbation of the form

$$c^*_\varepsilon(x,\xi) = (1-\xi(x))h^\varepsilon_1(x,\xi) + \xi(x)h^\varepsilon_0(x,\xi).$$

We assume there is a law q of $(Y^1,\dots,Y^{N_0}) \in \mathbb{Z}^{dN_0}$, functions g^ε_i on $\{0,1\}^{N_0}$, $i = 0,1$, and $\varepsilon_1 \in (0,\infty], \varepsilon_0 \in (0,1]$ so that

$$g^\varepsilon_i \geq 0, \tag{1.4}$$

and for all $\xi \in \{0,1\}^{\mathbb{Z}^d}$, $x \in \mathbb{Z}^d$, and $\varepsilon \in (0,\varepsilon_0]$, we have

$$h^\varepsilon_i(x,\xi) = -\varepsilon_1^{-2} f_i(x,\xi) + E_Y(g^\varepsilon_i(\xi(x+Y^1),\dots\xi(x+Y^{N_0}))), i = 0,1. \tag{1.5}$$

Here E_Y is expectation with respect to q and in practice the first term in the above will allow us to take $g^\varepsilon_i \geq 0$. It is important to have g^ε_i non-negative as we will treat it as a rate in the construction of a dual process in Section 2. On the other hand, in two of the particular examples of interest (Examples 2 and 4 below), the h^ε_i will at times be negative.

We also suppose that (decrease $\kappa > 0$ if necessary)

$$P(Y^* \geq x) \leq Ce^{-\kappa x} \text{ for } x > 0, \tag{1.6}$$

where $Y^* = \max\{|Y^1|,\dots|Y^{N_0}|\}$, and there are limiting maps $g_i : \{0,1\}^{N_0} \to \mathbb{R}_+$ such that

$$\lim_{\varepsilon\downarrow 0} \|g^\varepsilon_i - g_i\|_\infty = 0,\ i = 0,1. \tag{1.7}$$

The conditions (1.4) and (1.5) with $\varepsilon_0 < \varepsilon_1$ (without loss of generality) easily imply the non-negativity in (1.3) for $\varepsilon \leq \varepsilon_0$.

We now show that the conditions (1.4)-(1.7) hold for general finite range convergent translation invariant perturbations without any explicit non-negativity condition on the g^ε_i.

Proposition 1.1. *Assume there are distinct points $y_1, \dots, y_{N_0} \in \mathbb{Z}^d$ and $\hat{g}_i^\varepsilon, \hat{g}_i : \{0,1\}^{N_0} \to \mathbb{R}$ such that*

$$\begin{aligned} &h_i^\varepsilon(x,\xi) = \hat{g}_i^\varepsilon(\xi(x+y_1), \dots, \xi(x+y_{N_0})),\ x \in \mathbb{Z}^d, \xi \in \{0,1\}^{N_0}, \\ &\{x : p(x) > 0\} \subset \{y_1, \dots, y_{N_0}\}, \quad \lim_{\varepsilon \downarrow 0} \|\hat{g}_i^\varepsilon - \hat{g}_i\|_\infty = 0\ i = 0, 1. \end{aligned} \tag{1.8}$$

Then (1.5)-(1.7) hold for appropriate non-negative g_i^ε, g_i satisfying $\|g_i^\varepsilon - g_i\|_\infty = \|\hat{g}_i^\varepsilon - \hat{g}_i\|_\infty$, and $Y^i = y_i$.

The elementary proof is given in Section 2.1. In terms of our original rates (1.3) this shows that our class of models include spin-flip systems $\xi_t \in \{0,1\}^{\mathbb{Z}^d}$, $t \geq 0$, with rates

$$c_\varepsilon^o(x,\xi) = c^v(x,\xi) + \varepsilon^2 c^*(x,\xi) + \varepsilon^2 o(\varepsilon) \geq 0, \tag{1.9}$$

where p (governing c^v) is now finite range, $c^*(x,\xi) = h(\xi(x), \xi(x+y_1), \dots, \xi(x+\xi_{N_0}))$ is a finite range, translation invariant perturbation and $o(\varepsilon)$ means this term goes to zero with ε uniformly in (x,ξ). However, most of the Theorems below are formulated without this finite range assumption (Theorem 1.5 being the notable exception). Working with the random Y^i's will allow us to include certain natural infinite range interactions and will also simplify some of the arithmetic to come.

We stress that conditions (1.3)-(1.7) are in force throughout this work, and call such a process ξ a voter model perturbation on $\mathbb{Z}^d$. To be precise about the meaning of our process, let ξ_t, $t \geq 0$ be the unique $\{0,1\}^{\mathbb{Z}^d}$-valued Feller process with translation invariant flip rates given by $c_\varepsilon^o(x,\xi)$ in (1.3) and initial state $\xi_0 \in \{0,1\}^{\mathbb{Z}^d}$. More formally (see Theorem B.3 in [36] and Section 2 of [14]) the generator of ξ is

$$\text{the closure of } \Omega^o g(\xi) = \sum_{x \in Z^d} c_\varepsilon^o(x,\xi)(g(\xi^x) - g(\xi)) \tag{1.10}$$

on the space of $g : \mathbb{Z}^d \to \mathbb{R}$, depending on finitely many coordinates. Here ξ^x is ξ with the coordinate at x flipped to $1 - \xi(x)$. The condition (B4) of Theorem B.3 in [36] is trivial to derive from (1.5).

To motivate this class of spin-flip systems we will now describe our four examples.

Example 1. Nonlinear voter models. Molofsky et al. [39] considered a discrete time particle system on $\mathbb{Z}^2$ in which each site is in state 0 or 1 and

$$P(\xi_{n+1}(x,y) = 1 | \xi_n) = p_k$$

if k of the sites $(x,y), (x+1,y), (x-1,y), (x,y+1), (x,y-1)$ are in state 1. They assumed that $p_0 = 0$ and $p_5 = 1$, so that all 0's and all 1's were absorbing states and $p_1 = 1 - p_4$ and $p_2 = 1 - p_3$, so that the model was symmetric under interchange of 0's and 1's.

Our goal is to study the phase diagram of a continuous time nonlinear voter model in a neighborhood of the ordinary linear voter model. Our perturbation is determined by four points chosen at random from $x + \mathcal{N}$ where $\mathcal{N}$ is the set of integer lattice points in $([-L,L]^d - \{0\})$. Let $a(i) \geq 0$ be the flip rate at a given site when i (randomly chosen) neighbors have a type disagreeing with that of the site and suppose $a(0) = 0$. Let $(Y_1, \dots Y_4)$ be chosen at random and without replacement from $\mathcal{N}$. Our perturbation has the form given in (1.5) with $\varepsilon_1 = \infty$,

$$g_1^\varepsilon(\xi_1, \dots, \xi_4) = a\Big(\sum_1^4 \xi_i\Big), \quad \text{and} \quad g_0^\varepsilon(\xi_1, \dots, \xi_4) = a\Big(4 - \sum_1^4 \xi_i\Big).$$

General models of this type were introduced and studied in [4]. The long range nature of the interaction will make computations simpler.

Example 2. Lotka-Volterra systems were introduced by Neuhauser and Pacala [40]. In addition to the kernel p for the voter model, the flip rates depend on two non-negative competition parameters, α_0 and α_1, and are given by

$$\begin{aligned} c_{LV}(x,\xi) =& f_1(f_0 + \alpha_0 f_1)(1 - \xi(x)) + f_0(f_1 + \alpha_1 f_0)\xi(x) \\ =& c^v(x,\xi) + (\alpha_0 - 1) f_1^2 (1 - \xi(x)) + (\alpha_1 - 1) f_0^2 \xi(x). \end{aligned} \tag{1.11}$$

In words, a plant of type i at x dies with rate $f_i(x,\xi) + \alpha_i f_{1-i}(x,\xi)$ and is immediately replaced by the type of a randomly chosen neighboring plant, which will be $1 - i$ with probability $f_{1-i}(x,\xi)$. Hence α_i represents the effect of competition on a type i from neighbors of the opposite type and the interspecies competition parameters have been set to one.

If $\theta_i \in \mathbb{R}$ and we let $\alpha_i = \alpha_i^\varepsilon = 1 + \varepsilon^2 \theta_i$ then the perturbation has the form

$$h_i(x,\xi) = \theta_{1-i} f_i(x,\xi)^2, \tag{1.12}$$

for $i = 0,1$ $x \in \mathbb{Z}^d$ and $\xi \in \{0,1\}^{\mathbb{Z}^d}$. To verify (1.5) take $0 < \varepsilon_1 \leq (\theta_0^-)^{-1/2} \wedge (\theta_1^-)^{-1/2}$, $N_0 = 2$, Y^1, Y^2 chosen independently according to p, and define for $i = 0, 1$,

$$g_i(\eta_1, \eta_2) = \varepsilon_1^{-2} \eta_1 (1 - \eta_2) + (\varepsilon_1^{-2} + \theta_{1-i}) 1(\eta_1 = \eta_2 = i) \geq 0. \tag{1.13}$$

Then we have the required form, where now h_i and g_i are independent of ε:

$$\begin{aligned} h_i(x,\xi) &= -\varepsilon_1^{-2} f_i(x,\xi) + \varepsilon_1^{-2}(f_i - f_i^2)(x,\xi) + (\varepsilon_1^{-2} + \theta_{1-i}) f_i(x,\xi)^2 \\ &= -\varepsilon_1^{-2} f_i(x,\xi) + E(g_i(\xi(x + Y^1), \xi(x + Y^2))). \end{aligned}$$

Example 3. The Latent voter was introduced by Lambiotte, Saramaki, and Blondel [33]. To motivate the dynamics, they consider the adoption of new technology, such as choosing between a Blu-ray and HD-DVD player, or between an IBM netbook and iPad. Once the customer has made her purchase it is unlikely that she will immediately switch. To model this latency, we introduce states 0^* and 1^* which indicate that the voter is not willing to switch, and we postulate that $i^* \to i$ at rate λ. While in state i the voter changes to state $(1-i)^*$ at a rate equal to the fraction of nearest neighbors with the opposite opinion, counting both active and latent voters of that type. At $t = 0$ we assume initially the particles are all in state 1 or 0.

To realize this as a voter model perturbation, we will take λ large. To construct the process we take the usual graphical representation for the voter model, and for each site introduce an independent Poisson process of "wake-up dots" with rate λ, where each "wake-up" corresponds to a potential flip from i^* to i. Voting events occur at rate 1, so the number of voter events N_λ between two wake up dots has

$$P(N_\lambda = k) = \left(\frac{1}{\lambda+1}\right)^k \frac{\lambda}{\lambda+1} \quad \text{for } k = 0, 1, 2, \ldots$$

If there is only one voter event between two wake-up dots at x then this is an ordinary voter event, x imitates her neighbor. If there are two or more, then this is Lambiotte and Redner's vacillating voter model [34]: the voter changes if at least one of the neighbors chosen is different. To see this note that the first neighbor that is different causes an opinion change and then all of the other voter events before the wake up dot are ignored.

To fit this model into the framework of this paper we will take $\lambda = \varepsilon^{-2} \geq 1$ and identify state i^* with state i for the spin-flip system ξ, although they are of course distinguished in the above graphical representation. Let $\{y_1, \ldots, y_{2d}\}$ be the nearest neighbors of 0 and let $\{Y^i : i \in \mathbb{N}\}$ be iid r.v.'s which are chosen at random from the above nearest neighbors. For each $k \in \mathbb{N}$ let $S_k = \{I_1, \ldots, I_k\}$ be chosen from $\{1, \ldots, 2d\}$ with replacement so that S_k has cardinality at most k. Below P_S and P_Y will denote averaging over these random quantities only. Then multiplying the above probabilities by the rate λ of the wake-up process, we see that our rates are

$$\begin{aligned} c^o_\varepsilon(x,\xi) &= \left(\frac{\lambda}{\lambda+1}\right)^2 c_v(x,\xi) + \sum_{k=2}^{\infty} \frac{\lambda^2}{(\lambda+1)^{k+1}} P_Y(\xi(x+Y^j) \neq \xi(x)\ \exists j \leq k) \\ &= c_v(x,\xi) + \varepsilon^2 \Big[\Big(\frac{-2\varepsilon^{-4} - \varepsilon^{-2}}{(\varepsilon^{-2}+1)^2} \Big) c_v(x,\xi) \\ &\qquad + \sum_{k=2}^{\infty} \frac{\varepsilon^{-6}}{(\varepsilon^{-2}+1)^{k+1}} P_S(\xi(x+y_j) \neq \xi(x)\ \exists j \in S_k) \Big]. \end{aligned}$$

The first term in the square brackets and the $k = 2$ term in the summation are both $O(1)$ while the rest of the summation, $\Sigma(\varepsilon)$, is $O(\varepsilon^2)$ and so are error terms. If

$$\varepsilon_1^{-2} = \sup_{0<\varepsilon\leq 1} \frac{2+\varepsilon^2}{(1+\varepsilon^2)^2},$$

and

$$\Sigma_i(\varepsilon) = \sum_{k=3}^{\infty} \frac{\varepsilon^{-6}}{(\varepsilon^{-2}+1)^{k+1}} P_S(\xi(x+y_j) = i)\ \exists j \in S_k),$$

then $\Sigma_i(\varepsilon) \le C\varepsilon^2$, and a bit of arithmetic shows (1.3) holds with

$$\begin{aligned} h_i^\varepsilon(x,\xi) &= \frac{-2-\varepsilon^2}{(1+\varepsilon^2)^2} f_i(x,\xi) + (1+\varepsilon^2)^{-3}\Big[2P_Y(\xi(x+Y_1)=i) \\ &\qquad\qquad - P(\xi(x+Y_1)=i, \xi(x+Y_2)=i)\Big] + \Sigma_i(\varepsilon) \\ &= -\varepsilon_1^{-2} f_i(x,\xi) + \Big[\Big(\varepsilon_1^{-2} - \frac{2+\varepsilon^2}{(1+\varepsilon^2)^2} + \frac{2}{(1+\varepsilon^2)^3}\Big) f_i(x,\xi) - \frac{f_i(x,\xi)^2}{(1+\varepsilon^2)^3} \\ &\qquad + \Sigma_i(\varepsilon)\Big] \\ &\equiv \varepsilon_1^{-2} f_i(x,\xi) + g_i^\varepsilon(\xi(x+y_1),\dots,\xi(x+y_{2d})). \end{aligned}$$

Hence we derive the required form (1.5) with $(Y^1,\dots,Y^{2d}) = (y_1,\dots,y_{2d})$ and $g_i^\varepsilon \ge 0$. Moreover it is clear from the above definition of g_i^ε that

$$\|g_i^\varepsilon - \tilde g_i\|_\infty \le C\varepsilon^2, \tag{1.14}$$

where

$$\tilde g_i(\eta_1,\dots,\eta_{2d}) = (2d)^{-2}\sum_{j=1}^{2d}\sum_{k=1}^{2d} g_i(\eta_j,\eta_k),$$

and g_i as is in the Lotka-Volterra model, i.e. as in (1.13), with the nearest neighbour kernel p and $\theta_i = -1$. Hence we again have a voter perturbation on $\mathbb{Z}^d$. Note also that the above implies

$$\tilde g_i(\xi(x+y_1),\dots,\xi(x+y_{2d})) = E(g_i(\xi(x+Y^1),\xi(x+Y^2)), \tag{1.15}$$

exactly as for the limiting value of the Lotka-Volterra model.

Example 4. Evolutionary games. Ohtsuki et al [41] considered a system in which each site of a large (N vertex) graph G is occupied by a cooperator (1) or a defector (0). Simplifying their setting a bit, we will assume that each vertex in G has k neighbors. The interaction between these two types is governed by a payoff matrix with real entries

	C	**D**
C	α	β
D	γ	δ

This means that a cooperator receives a payoff α from each neighboring cooperator and a payoff β from each neighboring defector, while for defectors the payoffs are γ and δ from each neighboring cooperator or defector, respectively.

If $n_i(y)$ is the number of neighboring i's for site $y \in G$, $i = 0,1$, and $\xi(y) \in \{0,1\}$ is the state at site y, then the fitness $\rho_i(y)$ of site y in state i is determined by its local payoffs through

$$\begin{aligned} \rho_1(y) &= 1 - w + w(\alpha n_1(y) + \beta n_0(y)) \text{ if } \xi(y) = 1 \\ \rho_0(y) &= 1 - w + w(\gamma n_1(y) + \delta n_0(y)) \text{ if } \xi(y) = 0. \end{aligned} \tag{1.16}$$

Here $w \in [0,1]$ is a parameter determining the selection strength. In [41] they focus largely on the weak selection (small w) regime. Clearly for some $w_0(\alpha,\beta,\gamma,\delta,k) > 0$, $\rho_i \geq 0$ for $w \in [0, w_0]$, which we assume in what follows.

For the death-birth dynamics in [41] a randomly chosen individual is eliminated at x and its neighbors compete for the vacated site with success proportional to their fitness. We consider the continuous time analogue which is the spin-flip system $\xi_t(x) \in \{0,1\},\ x \in G$, with rates (write $y \sim x$ if and only if y and x are neighbors)

$$\begin{aligned} c(x,\xi) &= (1-\xi(x))r_1(x,\xi) + \xi(x) r_0(x,\xi) \geq 0, \\ r_i(x,\xi) &= \frac{\sum_{y\sim x} \rho_i(y) 1(\xi(y)=i)}{\sum_{y \sim x} \rho_1(y)\xi(y) + \rho_0(y)(1-\xi(y))} \in [0,1]. \end{aligned} \tag{1.17}$$

More precisely, choose a symmetric (about 0) set $\mathcal{N}$ of neighbors of 0 of size k, not containing 0, and consider the graph with vertex set $\mathbb{Z}^d$ and $x \sim y$ if and only if $x - y \in \mathcal{N}$. Assume also that the additive group generated by $\mathcal{N}$ is $\mathbb{Z}^d$ and $\sum_{x\in\mathcal{N}} x_i x_j / k = \sigma^2 \delta_{ij}$, so that $p(x) = k^{-1}1(x \in \mathcal{N})$ satisfies the conditions on our kernel given in, and prior to, (1.1). Set $w = \varepsilon^2$. For $x \in \mathbb{Z}^d$ and $\xi \in \{0,1\}^{\mathbb{Z}^d}$, let

$$\begin{aligned} f_i^{(2)}(x,\xi) &= k^{-1} \sum_{y \sim x} 1(\xi(y)=i) f_i(y,\xi) = k^{-2} \sum_{y\sim x}\sum_{z \sim y} 1(\xi(y)=\xi(z)=i) \in [0,1], \\ \theta_1(x,\xi) &= (\beta k - 1) f_1(x,\xi) + k(\alpha-\beta) f_1^{(2)}(x,\xi), \\ \theta_0(x,\xi) &= (\gamma k - 1) f_0(x,\xi) + k(\delta-\gamma) f_0^{(2)}(x,\xi), \\ \phi(x,\xi) &= (\theta_0+\theta_1)(x,\xi). \end{aligned}$$

Using (1.16) in (1.17), we get

$$r_i(x,\xi) = \frac{f_i + \varepsilon^2\theta_i}{1+\varepsilon^2\phi}(x,\xi). \tag{1.18}$$

Expanding in powers of ε we have

$$\begin{aligned} \frac{f+\varepsilon^2\theta}{1+\varepsilon^2\phi} &= f + \varepsilon^2(\theta - f\phi) + \varepsilon^4 \phi(f\phi-\theta)\Big[\sum_0^\infty (-\varepsilon^2\phi)^k\Big] \\ &= f + \varepsilon^2(\theta - f\phi) + \varepsilon^4 \psi_\varepsilon(f,\phi,\theta), \end{aligned}$$

and so our perturbation has the form

$$h_i^\varepsilon(x,\xi) = \theta_i(x,\xi) - f_i(x,\xi)\phi(x,\xi) + \varepsilon^2 \psi_\varepsilon(f_i(x,\xi), \phi(x,\xi), \theta_i(x,\xi)). \tag{1.19}$$

Note that

$$|\theta_1| \vee |\theta_0| \vee |\phi|(x,\xi) \leq 2k(1+|\alpha|+|\beta|+|\gamma|+|\delta|) \equiv R, \tag{1.20}$$

If $\varepsilon^2 < (2R)^{-1}$, then (1.20) easily gives

$$|\psi_\varepsilon((f_i,\phi,\theta_i)(x,\xi))| \leq 2R(R+1). \tag{1.21}$$

From this and (1.19) it is clear that the hypotheses of Proposition 1.1 hold with

$$\|\hat{g}_i^\varepsilon - \hat{g}_i\|_\infty \leq \varepsilon^2 2R(R+1), \tag{1.22}$$

and therefore our spin-flip system is a voter model perturbation on $\mathbb{Z}^d$.

For a fifth example on the evolution of seed dispersal range see [22].

The goal of our analysis. Given a process taking values in $\{0,1\}^{\mathbb{Z}^d}$, or more generally in $\{0,1\}^{\varepsilon\mathbb{Z}^d}$, we say that *coexistence holds* if there is a stationary distribution ν with

$$\nu\left(\sum_x \xi(x) = \sum_x 1-\xi(x) = \infty\right) = 1. \tag{1.23}$$

For voter model perturbations it is easy to see this is equivalent to both types being present ν-a.s., see Lemma 6.1.

We say *the i's take over* if for all L,

$$P(\xi_t(x) = i \text{ for all } x \in [-L,L]^d \ \text{ for } t \text{ large enough}) = 1 \tag{1.24}$$

whenever the initial configuration has infinitely many sites in state i.

Our main results (Theorems 1.4 and 1.5 below) give (often asymptotically sharp) conditions under which coexistence holds or one type takes over, respectively, in a voter model perturbation for small enough ε.

1.1 Hydrodynamic limit

In this section we will prove that the voter model perturbation on the rescaled lattice $\varepsilon\mathbb{Z}^d$ run at rate ε^{-2} converges to a PDE. For $1 \geq \varepsilon > 0$, $x \in \mathbb{Z}^d$ and $\xi \in \{0,1\}^{\mathbb{Z}^d}$ define rescalings of p and ξ by $p_\varepsilon(\varepsilon x) = p(x)$, and $\xi_\varepsilon(\varepsilon x) = \xi(x)$, so that $\xi_\varepsilon \in \{0,1\}^{\varepsilon\mathbb{Z}^d}$. Also define rescaled *local densities* f_i^ε by

$$f_i^\varepsilon(\varepsilon x, \xi_\varepsilon) = \sum_{y\in\varepsilon\mathbb{Z}^d} p_\varepsilon(y-\varepsilon x)1\{\xi_\varepsilon(y) = i\}, \qquad i = 0,1. \tag{1.25}$$

For x, ξ as above, introduce the rapid voter flip rates given by

$$c_\varepsilon^v(\varepsilon x, \xi_\varepsilon) = \varepsilon^{-2} c^v(x,\xi). \tag{1.26}$$

Therefore the rescaled processes of interest, $\xi_{\varepsilon^{-2}t}(\varepsilon x) \in \{0,1\}^{\varepsilon\mathbb{Z}^d}$, will have rates

$$c_\varepsilon(\varepsilon x, \xi_\varepsilon) = c_\varepsilon^v(\varepsilon x, \xi_\varepsilon) + c_\varepsilon^*(x,\xi) \geq 0, \tag{1.27}$$

where the non-negativity is immediate from (1.3), and generator equal to

$$\text{the closure of } \Omega^\varepsilon g(\xi_\varepsilon) = \sum_{x\in\varepsilon Z^d} c_\varepsilon(x,\xi_\varepsilon)(g(\xi_\varepsilon^x) - g(\xi_\varepsilon)), \tag{1.28}$$

on the space of $g : \varepsilon\mathbb{Z}^d \to \mathbb{R}$, depending on finitely many coordinates. We call this rescaled process a voter model perturbation on $\varepsilon\mathbb{Z}^d$ and will often denote it by $\xi^\varepsilon(t)$.

As $d \geq 3$, we see from Theorem V.1.8 of [35] the voter model with flip rates $c^v(x,\xi) = c_1^v(x,\xi)$ has a one-parameter family of translation invariant extremal invariant distributions $\{\mathbb{P}_u : u \in [0,1]\}$ on $\{0,1\}^{\mathbb{Z}^d}$ such that $\mathbb{E}_u(\xi(x)) = u$. We write $\langle g \rangle_u$ for $\mathbb{E}_u(g(\xi))$. (1.5) and (1.7) imply

$$\lim_{\varepsilon \downarrow 0} \|h_i^\varepsilon - h_i\|_\infty = 0 \text{ where} \tag{1.29}$$
$$h_i(x,\xi) = -\varepsilon_1^{-2} f_i(x,\xi) + E(g_i(\xi(x+Y^1),\ldots,\xi(x+Y^{N_0}))).$$

Define

$$f(u) = \langle (1-\xi(0))h_1(0,\xi) - \xi(0)h_0(0,\xi)\rangle_u. \tag{1.30}$$

Then f is a polynomial of degree at most N_0+1 (see (1.40) and Section 1.8 below). The non-negativity condition (1.27), the fact that

$$\langle c_\varepsilon^v(0,\xi)\rangle_0 = \langle c_\varepsilon^v(0,\xi)\rangle_1 = 0, \tag{1.31}$$

and the convergence (1.29) show that

$$f(0) \geq 0, \quad f(1) \leq 0. \tag{1.32}$$

Our first goal is to show that under suitable assumptions on the initial conditions, as $\varepsilon \to 0$ the particle systems converges to the PDE

$$\frac{\partial u}{\partial t} = \frac{\sigma^2}{2}\Delta u + f(u), \quad u(0,\cdot) = v(\cdot). \tag{1.33}$$

The remark after Proposition 2.1 in [2] implies that for any continuous $v : R^d \to [0,1]$ the equation has a unique solution u, which necessarily takes values in $[0,1]$.

For a continuous v as above we will say that a family of probability measures $\{\lambda_\varepsilon\}$ on $\{0,1\}^{\varepsilon\mathbb{Z}^d}$ has local density v if the following holds:

There is an $r \in (0,1)$ such that if $a_\varepsilon = \lceil \varepsilon^{r-1} \rceil \varepsilon$, $Q_\varepsilon = [0,a_\varepsilon)^d \cap \varepsilon\mathbb{Z}^d$, $|Q_\varepsilon| = \text{card}\,(Q_\varepsilon)$, and

$$D(x,\xi) = \frac{1}{|Q_\varepsilon|}\sum_{y \in Q_\varepsilon} \xi(x+y) \text{ for } x \in a_\varepsilon\mathbb{Z}^d, \xi \in \{0,1\}^{\varepsilon Z^d}, \tag{1.34}$$

then for all $R, \delta > 0$,

$$\lim_{\varepsilon \to 0} \sup_{\substack{x \in a_\varepsilon \mathbb{Z}^d \\ |x| \leq R}} \lambda_\varepsilon(|D(x,\xi) - v(x)| > \delta) = 0\,. \tag{1.35}$$

The family of Bernoulli product measures $\bar{\lambda}_\varepsilon$ given by

$$\bar{\lambda}_\varepsilon(\xi(w_i) = 1, i = 1,\ldots,n) = \prod_{i=1}^n v(w_i) \text{ for all } n \in \mathbb{N} \text{ and } w_i \in \varepsilon\mathbb{Z}^d\,. \tag{1.36}$$

certainly satisfies (1.35) for all $r \in (0,1)$.

Theorem 1.2. *Assume $v : \mathbb{R}^d \to [0,1]$ is continuous, and the collection of initial conditions $\{\xi_0^\varepsilon\}$ have laws $\{\lambda_\varepsilon\}$ with local density v. Let $x^k \in \mathbb{R}^d$ and $x_\varepsilon^k \in \varepsilon\mathbb{Z}^d$, $k = 1, \dots K$ satisfy*

$$x_\varepsilon^k \to x^k \text{ and } \varepsilon^{-1}|x_\varepsilon^k - x_\varepsilon^{k'}| \to \infty \text{ as } \varepsilon \to 0 \text{ for any } k \neq k'. \tag{1.37}$$

If u is the solution of (1.33), then for any $\eta \in \{0,1\}^{\{0,\dots,L\}\times\{1,\dots K\}}$, $y_0, \dots, y_L \in \mathbb{Z}^d$ and $T > 0$,

$$\begin{aligned}\lim_{\varepsilon\to 0} P(\xi_T^\varepsilon(x_\varepsilon^k + \varepsilon y_i) = \eta_{i,k},\quad i = 0,\dots,L,\ k = 1,\dots K)&\\ = \prod_{k=1}^{K} \langle 1\{\xi(y_i) = \eta_{i,k}, i = 0,\dots,L\}\rangle_{u(T,x^k)}.&\end{aligned} \tag{1.38}$$

In particular, if $x_\varepsilon \in \varepsilon\mathbb{Z}^d$ satisfies $x_\varepsilon \to x$ as $\varepsilon \to 0$, then

$$\lim_{\varepsilon\to 0} P(\xi_T^\varepsilon(x_\varepsilon) = 1) = u(T,x) \quad \text{for all } T > 0, x \in \mathbb{R}^d. \tag{1.39}$$

De Masi, Ferrari and Lebowitz [11], Durrett and Neuhauser [20] and Durrett [15] have proved similar results for particle systems with rapid stirring. The local equilibrium for rapid stirring is a Bernoulli product measure, but in our setting it is the voter equilibrium. As a result there is now dependence between nearby sites on the microscopic scale. However, there is asymptotic independence between sites with infinite separation on the microscopic scale.

It is easy to carry out a variance calculation to improve Theorem 1.2 to the following L^2-convergence theorem (see the end of Section 3 for the proof). If $\delta > 0$ and $x \in \mathbb{R}^d$, let $I_\delta(x)$ be the unique semi-open cube $\prod_{i=1}^d [k_i\delta, (k_i+1)\delta)$, $k_i \in \mathbb{Z}$, which contains x.

Theorem 1.3. *Assume the hypotheses of Theorem 1.2. Assume also $\delta(\varepsilon) \in \varepsilon\mathbb{N}$ decreases to zero so that $\delta(\varepsilon)/\varepsilon \to \infty$ as $\varepsilon \downarrow 0$. If*

$$\tilde{u}^\varepsilon(t,x) = \sum_{y \in I_{\delta(\varepsilon)}(x)} \xi_t^\varepsilon(y)(\varepsilon/\delta(\varepsilon))^d,$$

then as $\varepsilon \to 0$, $\tilde{u}^\varepsilon(t,x) \to u(t,x)$ in L^2 uniformly for x in compacts, for all $t > 0$.

A low density version of this theorem, in which the limit is random (super-Brownian motion with drift), was proved in [7] and is discussed in Section 1.8.

To apply Theorem 1.2 to the voter perturbation we will have to evaluate $f(u)$. This is in principle straightforward thanks to the duality between the voter model and coalescing random walk which we now recall. Let $\{\hat{B}^x : x \in \mathbb{Z}^d\}$ denote a rate 1 coalescing random walk system on $\mathbb{Z}^d$ with step distribution p and $\hat{B}_0^x = x$. For $A, B \subset \mathbb{Z}^d$, let $\hat{\xi}_t^A = \{\hat{B}_t^x : x \in A\}$, $\tau(A) = \inf\{t : |\hat{\xi}_t^A| = 1\}$ and $\tau(A,B)$ be the first time $\hat{\xi}_t^A \cap \hat{\xi}_t^B \neq \emptyset$ (it is ∞ if either A or B is empty). The duality between $\hat{B}$ and the voter model (see (V.1.7) and Theorem V.1.8 in [35]) implies for finite $A, B \subset \mathbb{Z}^d$,

$$\langle \prod_{y\in A}\xi(y)\prod_{z\in B}(1-\xi(z))\rangle_u$$
$$= \sum_{j=0}^{|A|}\sum_{k=0}^{|B|} u^j(1-u)^k P(|\hat{\xi}^A_\infty| = j, |\hat{\xi}^B_\infty| = k, \tau(A,B) = \infty). \tag{1.40}$$

The $k = 0$ term is non-zero only if $B = \emptyset$ in which case the above probability is $P(|\hat{\xi}^A_\infty| = j)$, and similarly for the $j = 0$ term. It follows from (1.40) and the form of the perturbation in (1.5) that $f(u)$ is a polynomial of degree at most $N_0 + 1$ with coefficients given by certain coalescing probabilities of $\hat{B}$ (see (1.90) below).

1.2 General Coexistence and Extinction Results

Our results for the four examples will be derived from general results with hypotheses concerning properties of the solution $u(t, x)$ to the limiting PDE (1.33). The coexistence results for the models discussed in the opening section are obtained by verifying the next assumption in the particular cases.

Assumption 1. *Suppose that there are constants* $0 < v_0 < u_* \le u^* < v_1 < 1$*, and* $w, L_i > 0$*, so that*
(i) if $u(0, x) \ge v_0$ *when* $|x| \le L_0$*, then* $\liminf_{t\to\infty} \inf_{|x|\le wt} u(t, x) \ge u_*$.
(ii) if $u(0, x) \le v_1$ *when* $|x| \le L_1$*, then* $\limsup_{t\to\infty} \sup_{|x|\le wt} u(t, x) \le u^*$.

We also will need a rate of convergence in (1.7), namely for some $r_0 > 0$,

$$\sum_{i=0}^{1} \|g^\varepsilon_i - g_i\|_\infty \le c_{1.41}\varepsilon^{r_0}. \tag{1.41}$$

Assumption 1 shows that the limiting PDE in Theorem 1.3 will have solutions which stay away from 0 and 1 for large t. A "block construction" as in [14] will be employed in Section 6 to convert this information about the PDE into information about the particle systems. In effect, this allows us to interchange limits as $\varepsilon \to 0$ and $t \to \infty$ and conclude the existence of a nontrivial stationary distribution, and also show that any stationary distribution will have particle density restricted by the asymptotic behavior of the PDE solutions at $t = \infty$.

Theorem 1.4. *Consider a voter model perturbation on* $\mathbb{Z}^d$ *satisfying* (1.41)*. Suppose Assumption 1. If* $\varepsilon > 0$ *is small enough, then coexistence holds and the nontrivial stationary distribution* ν *may be taken to be translation invariant.*

If $\eta > 0$ *and* $\varepsilon > 0$ *is small enough, depending on* η*, then any stationary distribution* ν *such that*

$$\nu\Big(\sum_x \xi(x) = 0 \text{ or } \sum_x (1-\xi(x)) = 0\Big) = 0 \tag{1.42}$$

satisfies $\nu(\xi(x) = 1) \in (u_* - \eta, u^* + \eta)$ *for all* x.

Note that in the second assertion we do not require that ν be translation invariant.

We now turn to the complementary case when one type takes over. Results that assert 0's will take over will require a pde input which is naturally mutually exclusive from Assumption 1 but also stronger in that it prescribes exponential rates of convergence.

Assumption 2. *There are constants $0 < u_1 < 1$, $c_2, C_2, w > 0$, $L_0 \geq 3$ so that for all $L \geq L_0$, if $u(0,x) \leq u_1$ for $|x| \leq L$, then for all $t \geq 0$,*

$$u(t,x) \leq C_2 e^{-c_2 t} \text{ for all } |x| \leq L + 2wt.$$

Finally we need to assume that the constant configuration of all 0's is a trap for our voter perturbation, that is,

$$g_1^\varepsilon(0,\dots,0) = 0, \text{ or equivalently } h_1^\varepsilon(0,\underline{0}) = 0, \text{ for } 0 < \varepsilon \leq \varepsilon_0\,, \tag{1.43}$$

where $\underline{0}$ is the zero configuration in $\{0,1\}^{\mathbb{Z}^d}$. This clearly implies $f(0) = 0$ and is equivalent to it if g_1^ε does not depend on ε, as is the case in some examples. Recall the definition of "i"s take over" from (1.24) and that q is the law of $(Y^1, \dots, Y^{N_0})$.

Theorem 1.5. *Consider a voter model perturbation on $\mathbb{Z}^d$ satisfying (1.41), (1.43), and such that $p(\cdot)$ and $q(\cdot)$ have finite support. Suppose Assumption 2, and $f'(0) < 0$. Then for $\varepsilon > 0$ sufficiently small, the 0's take over.*

We believe the theorem holds without the finite range assumptions on p and q. By Proposition 1.1 in the above finite support setting, it suffices to assume (1.8) in place of (1.5), and also assume (1.41) holds for the $\hat{g}_i^\varepsilon, \hat{g}_i$ appearing in (1.8) instead of the g_i^ε, g_i. Of course by symmetry there is a corresponding result giving conditions for 1's to take over.

We have formulated Theorems 1.4 and 1.5 for the original voter model perturbations on $\mathbb{Z}^d$. These results are clearly equivalent to the corresponding results for the (rescaled) voter model perturbations on $\varepsilon\mathbb{Z}^d$. In the proofs below we will in fact prove the results in this rescaled setting where Theorem 1.2 provides a bridge with the solution of the limiting PDE.

1.3 PDE results

To prepare for the discussion of the examples, we will state the PDE results on which their analysis will be based. The reaction function $f : \mathbb{R} \to \mathbb{R}$ is a continuously differentiable function (as already noted, in our context it will be a polynomial). Assume now, as will be the case in the examples, that $f(0) = f(1) = 0$. We let $u(t,x)$ denote the unique solution of (1.33) with continuous initial data $v : \mathbb{R}^d \to [0,1]$. In what follows it is useful to note that if u is a solution to (1.33) with reaction function f and initial condition v, then $1-u$ is a solution to (1.33) with reaction function $-f$ and initial condition $1-v$.

We start with a modification of a result of Aronson and Weinberger [2].

Proposition 1.6. *Suppose $f(0) = f(\alpha) = 0$, $f'(0) > 0$, $f'(\alpha) < 0$ and $f(u) > 0$ for $u \in (0, \alpha)$ with $0 < \alpha \le 1$. There is a $w > 0$ so that if the initial condition v is not identically 0, then*

$$\liminf_{t\to\infty} \inf_{|x|\le 2wt} u(t,x) \ge \alpha.$$

We also will need an exponential rate of convergence in this case under a stronger condition on the initial condition. We formulate it for $f < 0$ on $(0, 1)$. The brief proofs of Propositions 1.6 and 1.7 are given at the beginning of Section 4.

Proposition 1.7. *Assume $f < 0$ on $(0, 1)$ and $f'(0) < 0$. There is a $w > 0$, and if $\delta > 0$ there are positive constants L_δ, $c = c_\delta$, and $C = C_\delta$ so that if $L \ge L_\delta$ and $v(x) \le 1 - \delta$ for $|x| \le L$, then*

$$u(t, x) \le Ce^{-ct} \quad \textit{for } |x| \le L + 2wt.$$

There are different cases depending on the number of solutions of $f(u) = 0$ in $(0, 1)$. In all cases, we suppose that $f'(0) \ne 0$ and $f'(1) \ne 0$. We will consider some examples with multiple zeros in $(0, 1)$ in our treatment of the nonlinear voter models in Section 1.7 below, for the moment we will restrict our attention to two simple cases.

Case I: f has zero roots in $(0, 1)$. In this case we can apply Propositions 1.6 (with $\alpha = 1$) and 1.7, and their obvious analogues for $-f$.

Case II: f has one root $\rho \in (0, 1)$. There are two possibilities here.

(i) $f'(0) > 0$ and $f'(1) < 0$ and so the interior fixed point $\rho \in (0, 1)$ is attracting. In this case we will also assume $f'(\rho) \ne 0$. Then two applications of Proposition 1.6 show that if $v \not\equiv 0$ and $v \not\equiv 1$

$$\lim_{t\to\infty} \sup_{|x|\le wt} |u(t, x) - \rho| = 0. \tag{1.44}$$

(ii) $f'(0) < 0$ and $f'(1) > 0$, so that 0 and 1 are locally attracting and $\rho \in (0, 1)$ is unstable. In this case the limiting behavior of the PDE is determined by the speed r of the traveling wave solutions, i.e., functions w with $w(-\infty) = \rho$ and $w(\infty) = 0$ so that $u(t, x) = w(x - rt)$ solves the PDE. The next result was first proved in $d = 1$ by Fife and McLeod [26]. See page 296 and the appendix of [20] for the extension to $d > 1$ stated below as Proposition 1.8. The assumption there on the non-degeneracy of the interior zeros are not necessary (see Fife and McLeod [26]). These references also show that

$$\mathrm{sgn}(r) = \mathrm{sgn}\Big(\int_0^1 f(u)du\Big). \tag{1.45}$$

$|x|_2$ will denote the Euclidean norm of x and the conditions of Case II(ii) will apply in the next two propositions.

Proposition 1.8. *Suppose $\int_0^1 f(u)du < 0$ and fix $\eta > 0$. If $\delta > 0$ there are positive constants L^0_δ, $c_0 = c_0(\delta)$, and $C_0 = C_0(\delta)$ so that if $L \ge L^0_\delta$ and $v(x) \le \rho - \delta$ when $|x|_2 \le L$, then*

$$u(t, x) \le Ce^{-ct} \quad \textit{for } |x|_2 \le (|r| - \eta)t.$$

For the block construction it is useful to have a version of the last result for the L^∞ norm, and which adds an L to the region in which the result is valid.

Proposition 1.9. *Suppose $\int_0^1 f(u)du < 0$. There is a $w > 0$, and if $\delta > 0$ there are positive constants L_δ, $c = c_\delta$ and $C = C_\delta$ so that if $L \geq L_\delta$ and $v(x) \leq \rho - \delta$ for $|x| \leq L$, then*

$$u(t,x) \leq Ce^{-ct} \quad \text{for } |x| \leq L + 2wt.$$

The short derivation of Proposition 1.9 from Proposition 1.8 is given at the beginning of Section 4.

1.4 Lotka-Volterra systems

We refer to the model defined in Example 2 as the $LV(\alpha_0, \alpha_1)$ process. Proposition 8.1 of [8] implies that

$$\text{if } \alpha_0 \wedge \alpha_1 \geq 1/2 \text{ then } LV(\alpha_0, \alpha_1) \text{ is monotone (or attractive).} \tag{1.46}$$

Write $LV(\alpha) \leq LV(\alpha')$ if $LV(\alpha')$ stochastically dominates $LV(\alpha)$, that is, if one can define these processes, ξ and ξ', respectively, on a common probability space so that $\xi \leq \xi'$ pointwise a.s. Then, as should be obvious from the above interpretation of α_i (see (1.3) of [8]),

$$\begin{aligned} &0 \leq \alpha_0' \leq \alpha_0, 0 \leq \alpha_1 \leq \alpha_1', \text{ and either } \alpha_0 \wedge \alpha_1 \geq 1/2 \\ &\text{or } \alpha_0' \wedge \alpha_1' \geq 1/2, \text{ implies } LV(\alpha') \leq LV(\alpha). \end{aligned} \tag{1.47}$$

If $\alpha_i < 1$ for both $i = 0, 1$, then individuals are better off surrounded by the opposite type and one may expect coexistence to hold. On the other hand if both $\alpha_i > 1$, competition between types is stronger than within a type and one may expect one type to take over, depending on the α_i values.

To calculate the limiting reaction function in this case consider the system of coalescing random walks $\{\hat{B}^x : x \in \mathbb{Z}^d\}$ used in the duality formula (1.40). Let $\{e_1, e_2\}$ be i.i.d. with law $p(\cdot)$ and independent of the $\{\hat{B}^x : x \in \mathbb{Z}^d\}$. If we abuse our earlier notation and let $\langle \cdot \rangle_u$ denote expectation on the product space where (e_1, e_2) and ξ are independent, and ξ is given the voter equilibrium with density u, then from (1.30), (1.12) and the fact that $f_i(0,\xi)^2 = P_e(\xi(e_1) = \xi(e_2) = i)$, we have

$$f(u) = \theta_0 \langle (1 - \xi(0))\xi(e_1)\xi(e_2)\rangle_u - \theta_1 \langle \xi(0)(1 - \xi(e_1))(1 - \xi(e_2))\rangle_u.$$

In view of (1.40) we will be interested in various coalescence probabilities. For example,

$$p(x|y,z) = P(\exists t \geq 0 \text{ such that } \hat{B}_t^y = \hat{B}_t^z, \text{ and } \forall t \geq 0,\ \hat{B}_t^y \neq \hat{B}_t^x \text{ and } \hat{B}_t^z \neq \hat{B}_t^x)$$

and

$$p(x|y|z) = P(\hat{B}_t^x, \hat{B}_t^y \text{ and } \hat{B}_t^z \text{ are all distinct for all } t).$$

In general, walks within the same group coalesce and those separated by at least one bar do not. If we define

$$p_2 = p(0|e_1, e_2), \quad p_3 = p(0|e_1|e_2), \tag{1.48}$$

where the expected value is taken over e_1, e_2, then by the above formula for f and (1.40),

$$\begin{aligned} f(u) &= \theta_0 u(1-u)p_2 + \theta_0 u^2(1-u)p_3 - \theta_1(1-u)up_2 - \theta_1(1-u)^2 up_3 \\ &= u(1-u)[\theta_0 p_2 - \theta_1(p_2+p_3) + up_3(\theta_0+\theta_1)]. \end{aligned} \tag{1.49}$$

To see what this might say about the Lotka-Volterra model introduce

$$u^*(\theta_1/\theta_0) = \frac{\theta_1(p_2+p_3) - \theta_0 p_2}{p_3(\theta_1+\theta_0)} \tag{1.50}$$

so that $f(u) = 0$ for $u = 0, 1$ or $u^*(\theta_1/\theta_0)$. If

$$m_0 \equiv \frac{p_2}{p_2+p_3}, \tag{1.51}$$

then $u^*(m)$ increases from 0 to 1 as m increases from m_0 to m_0^{-1}. We slightly abuse the notation and write u^* for $u^*(\theta_1/\theta_0)$.

To analyze the limiting PDE we decompose the $\theta_0 - \theta_1$ plane into 5 open sectors drawn in Figure 1 on which the above 0's are all simple. Set $\theta = (\theta_0, \theta_1)$.

- If $\theta \in R_1$, $f > 0$ on $(0, u^*)$, $f < 0$ on $(u^*, 1)$, so $u^* \in (0,1)$ is an attracting fixed point for the ODE. Then (1.44) shows the PDE solutions will converge to u^* given a non-trivial initial condition in $[0,1]$. As a result we expect coexistence in the particle system.
- If $\theta \in R_2$, $f < 0$ on $(0,1)$, 0 is an attracting fixed point for the ODE. Proposition 1.6 implies solutions of the PDE will converge to 0 given a non-trivial initial condition and we expect 0's to win.
- If $\theta \in R_3$, $f > 0$ on $(0,1)$, 1 is an attracting fixed point for the ODE and so by the reasoning from the previous case we expect 1's to win.
- On $R_4 \cup R_5$, $u^* \in (0,1)$ is an unstable fixed point, while 0 and 1 are attracting fixed points for the ODE. This is case 2 of Durrett and Levin [19], so we expect the winner of the competition to be predicted by the direction of movement of the speed of the decreasing traveling wave solution $u(x,t) = w(x - rt)$ with $w(-\infty) = 1$ and $w(\infty) = 0$. If $r > 0$ then 1's will win and if $r < 0$ then 0's will win. Symmetry dictates that the speed is 0 when $\theta_0 = \theta_1$, so this gives the dividing line between the two cases and the monotonicity from (1.47) predicts 0's win on R_4 while 1's win on R_5. Alternatively, by (1.45) r has the same sign as $\int_0^1 f(u)\,du$ which is positive in R_5 and negative in R_4.

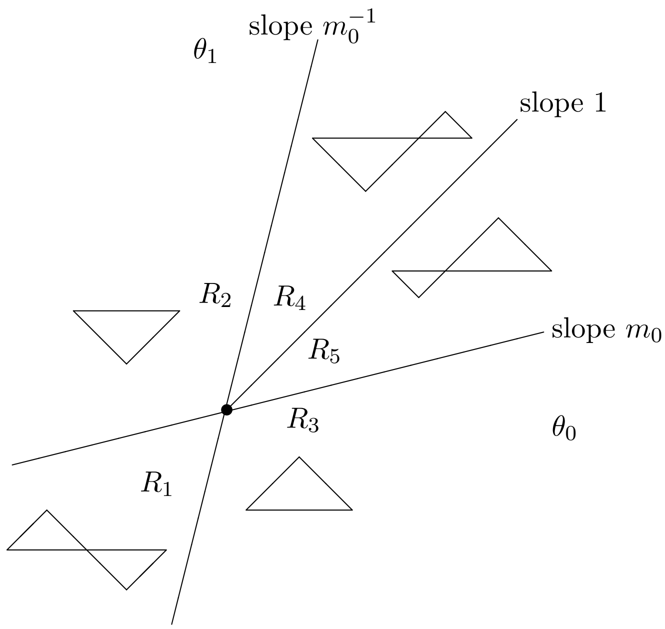


Figure 1: Phase diagram near (1,1) for the Lotka-Volterra model with the stylized shape of f in the regions.

Our next two results confirm these predictions for α close to $(1,1)$. For $0 \le \eta < 1$, define regions that are versions of R_1, $R_2 \cup R_4$ and $R_3 \cup R_5$ shrunken by changing the slopes of the boundary lines:

$$C^\eta = \Big\{(\alpha_0, \alpha_1) \in [0,1]^2 : \frac{(\alpha_0 - 1)(1-\eta)}{m_0} < \alpha_1 - 1 < \frac{m_0(\alpha_0 - 1)}{1-\eta}\Big\},$$

$$\Lambda_0^\eta = \Big\{(\alpha_0, \alpha_1) \in (0,\infty)^2 : 0 < \alpha_0 \le 1,\ m_0(1-\eta)(\alpha_0 - 1) < \alpha_1 - 1,$$
$$\text{or } 1 \le \alpha_0,\ (1+\eta)(\alpha_0 - 1) < \alpha_1 - 1\Big\},$$
$$\Lambda_1^\eta = \Big\{(\alpha_0, \alpha_1) \in (0,\infty)^2 : 0 < \alpha_0 \le 1,\ \alpha_1 - 1 < \frac{\alpha_0 - 1}{m_0(1-\eta)},$$
$$\text{or } 1 \le \alpha_0,\ \alpha_1 - 1 < (1-\eta)(\alpha_0 - 1)\Big\}.$$

Theorem 1.10. *For $0 < \eta < 1$ there is an $r_0(\eta) > 0$, non-decreasing in η, so that for the $LV(\alpha)$:*
(i) Coexistence holds for $(\alpha_0, \alpha_1) \in C^\eta$ and $1 - \alpha_0 < r_0(\eta)$.
(ii) If (α_0, α_1) is as in (i) and ν_α is a stationary distribution with $\nu_\alpha(\xi \equiv 0$ or $\xi \equiv 1) = 0$, then

$$\sup_x \Big|\nu_\alpha(\xi(x) = 1) - u^*\Big(\frac{\alpha_1 - 1}{\alpha_0 - 1}\Big)\Big| \le \eta.$$

(i) is a consequence of Theorem 4 of [8], which also applies to more general perturbations. The main conditions of that result translate into $f'(0) \ge \eta$ and $f'(1) \ge \eta$ in our present setting (see (1.92) in Section 1.8 below). (ii) sharpens (i) by showing that if η is small then the density of 1's in any nontrivial stationary distribution is close to the prediction of mean-field theory. Durrett and Neuhauser [20] prove results of this type for some systems with fast stirring and $f(1) < 0$. Neuhauser and Pacala [40] conjectured that coexistence holds for all $\alpha_0 = \alpha_1 < 1$ (see Conjecture 1 of that paper) and proved it for α_i sufficiently small. Hence (i) provides further evidence for the general conjecture.

Proof of Theorem 1.10. For $\eta \in (0, 1 - m_0)$ consider

$$\alpha_0^\varepsilon = 1 - \varepsilon^2, \alpha_1^\varepsilon = 1 + \varepsilon^2\theta_1^\varepsilon,\ \text{where } -\theta_1^\varepsilon \in \Big[\frac{m_0}{1-\eta}, \frac{1-\eta}{m_0}\Big],\ \lim_{\varepsilon \downarrow 0} \theta_1^\varepsilon = \theta_1. \tag{1.52}$$

Then the rescaled Lotka-Volterra model ξ^ε remains a voter model perturbation where g_0^ε may now depend on ε. From (1.49) we have

$$f(u) = u(1-u)[-p_2 - \theta_1(p_2 + p_3) + up_3(-1 + \theta_1)],$$

which has a zero, and attracting fixed point for the ODE, at

$$u^*(-\theta_1) = \frac{\theta_1(p_2 + p_3) + p_2}{p_3(\theta_1 - 1)} \in (0,1). \tag{1.53}$$

Proposition 1.6 and its mirror image, with 0 and 1 reversed, establish Assumption 1 with $u^* = u_* = u^*(-\theta_1)$ (see (1.44)). Theorem 1.4 therefore shows that for $0 < \varepsilon < \varepsilon_0(\eta)$

$$\begin{aligned}&\text{coexistence holds, and if } \nu \text{ is a stationary distribution satisfying} \\ &\nu(\xi \equiv 0 \text{ or } \xi \equiv 1) = 0, \text{ then } \sup_x |\nu(\xi(x) = 1) - u^*(-\theta_1)| < \eta.\end{aligned} \tag{1.54}$$

Suppose first that (ii) of Theorem 1.10 fails. Then there is a sequence $\varepsilon_n \downarrow 0$, $(\alpha_0^{\varepsilon_n}, \alpha_1^{\varepsilon_n})$ and $\delta_0 > 0$ so that (1.52) holds with $\varepsilon = \varepsilon_n$, and there is a stationary measure ν_n for ξ^{ε_n} satisfying $\nu_n(\xi \equiv 0 \text{ or } \xi \equiv 1) = 0$ and such that

$$\sup_x |\nu_n(\xi(x) = 1) - u^*(-\theta_1^{\varepsilon_n})| > \delta_0.$$

Since $u^*(-\theta_1^{\varepsilon_n}) \to u^*(-\theta_1)$, if we choose $\eta < \delta_0$ this contradicts (1.54) for large n, and so proves (ii). The proof of (i) is similar using the first part of (1.54). That is, if (i) fails, there is a sequence $\varepsilon_n \downarrow 0$ so that coexistence fails for $\alpha_i^{\varepsilon_n}$ as in (1.52), contradicting the first part of (1.54). $\square$

The next result is our main contribution to the understanding of Lotka-Volterra models. It shows that (i) of the previous result is asymptotically sharp, and verifies a conjecture in [8] (after Theorem 4 in that work). We assume p has finite support but believe this condition is not needed.

Theorem 1.11. *Assume $p(\cdot)$ has finite support. For $0 < \eta < 1$ there is an $r_0(\eta) > 0$, non-decreasing in η, so that for the LV(α):*
(i) 0's take over for $(\alpha_0, \alpha_1) \in \Lambda_0^\eta$ and $0 \le |\alpha_0 - 1| < r_0(\eta)$,
(ii) 1's take over for $(\alpha_0, \alpha_1) \in \Lambda_1^\eta$ and $0 \le |\alpha_0 - 1| < r_0(\eta)$.

Conjecture 2 of [40] states that 1's take over for $\alpha \in \Lambda_1^0$, $\alpha_0 > 1$ and 0's take over for $\alpha \in \Lambda_0^0$, $\alpha_0 > 1$. Theorem 1.11 establishes this result asymptotically as α gets close to $(1, 1)$, at least for $d \ge 3$.

Together, Theorems 1.10 and 1.11 give a fairly complete description of the phase diagram of the Lotka-Volterra model near the voter model. In Figure 2, C is the union over $\eta \in (0, 1)$ of the regions in Theorem 1.10 (i) on which there is coexistence, and Λ_i, $i = 0, 1$, is the union over η of the regions in Theorem 1.11 (i) and (ii), respectively, on which i's take over, as well as other parameter values for which the same result holds by monotonicity. For example, if $(\alpha_0, \alpha_1) \in \Lambda_1$, with $\alpha_0 \wedge \alpha_1 \ge 1/2$, and (α_0', α_1') has $\alpha_0' \ge \alpha_0$ and $\alpha_1' \le \alpha_1$, then by (1.47), $(\alpha_0', \alpha_1') \in \Lambda_1$. Theorem 1.10 (i) and Theorem 1.11 show that the three mutually exclusive classifications of coexistence, 0's take over, and 1's take over, occur on the three regions, C, Λ_0 and Λ_1, meeting at $(1, 1)$ along mutually tangential lines with slopes m_0, 1 and m_0^{-1}.

Proof of Theorem 1.11. It is enough to prove (i). Let $0 < \eta < 1$ and consider first

$$\alpha_0 = \alpha_0^\varepsilon = 1 - \varepsilon^2, \quad \alpha_1 = \alpha_1^\varepsilon = 1 - m_0(1 - \eta)\varepsilon^2,$$

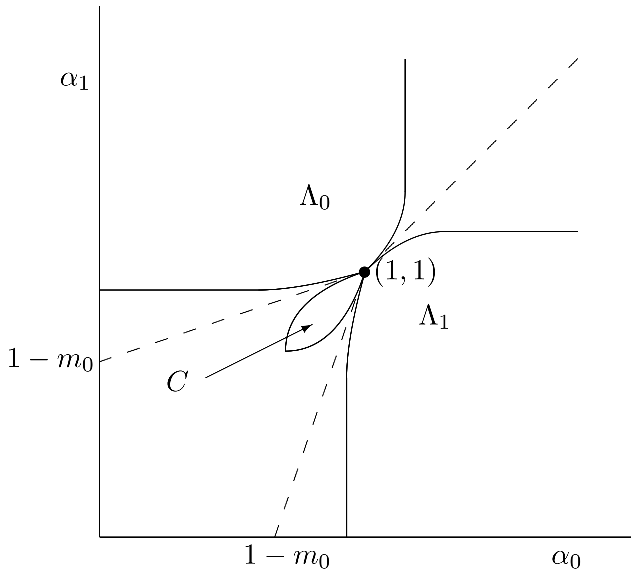


Figure 2: Coexistence on C, type i takes over on Λ_i.

so that in the notation of Example 2. we have set $\theta_0 = -1$, $\theta_1 = -m_0(1-\eta)$. The rescaled Lotka-Volterra process ξ^ε is a voter model perturbation on $\varepsilon\mathbb{Z}^d$, and from (1.49) we have

$$f(u) = -u(1-u)[\eta p_2 + up_3(1+m_0(1-\eta))] < 0 \text{ on } (0,1).$$

Proposition 1.7 verifies Assumption 2 in Theorem 1.5 and $f'(0) < 0$ is obvious. (1.41) is trivial ($g_i^\varepsilon = g_i$) and (1.43) is immediate from (1.13). The finite range assumption on $q = p \times p$ is immediate from that on p. Theorem 1.5 implies 0's take over for ε small. Therefore when $0 < 1-\alpha_0 \le r_0(\eta)$ and $\alpha_1 = 1+m_0(1-\eta)(\alpha_0-1)$, then 0's take over for $LV(\alpha)$. The monotonicity in (1.47) shows this is also the case for $\alpha_1 \ge 1+m_0(1-\eta)(\alpha_0-1)$ and α_0 as above.

Next consider

$$\alpha_0 = \alpha_0^\varepsilon = 1+\varepsilon^2,\ \alpha_1 = \alpha_1^\varepsilon = 1+(1+\eta)\varepsilon^2 \quad \text{that is, } \theta_0 = 1,\ \theta_1 = 1+\eta.$$

In this case we have

$$f(u) = u(1-u)[p_2 - (1+\eta)(p_2+p_3) + up_3(2+\eta)],$$

and so, assuming without loss of generality (by (1.47)) $1+\eta < m_0^{-1}$, from (1.50) f has a zero, and an unstable fixed point for the ODE, at

$$u^* = \frac{(1+\eta)(p_2+p_3)-p_2}{p_3(2+\eta)} \in \Big(\frac{1}{2}, 1\Big).$$

It follows that $\int_0^1 f(u)du < 0$ and Proposition 1.9 establishes Assumption 2. As above, Theorem 1.5 and (1.47) show that 0's take over if $0 < \alpha_0 - 1$ is sufficiently small and $\alpha_1 \geq 1 + (1+\eta)(\alpha_0 - 1)$. □

1.5 Latent voter model

By (1.14) and (1.15) our $f(u)$ is as for the Lotka-Volterra model, that is (1.49) holds, with $\theta_1 = \theta_2 = -1$ (which will put the Lotka-Volterra model in our coexistence regime C in Figure 2 for small ε), and (1.41) holds with $r_0 = 2$. Therefore we may apply Theorem 1.4 exactly as in the first part of the above proof of Theorem 1.10 for the voter model, with $u^*(-1) = 1/2$, to directly obtain (1.54) for small enough ε and hence prove:

Theorem 1.12. *Let $\eta > 0$. There is an $r_0(\eta) > 0$, non-decreasing in η so that for the latent voter model with $0 < \varepsilon \leq r_0(\eta)$:*
(i) Coexistence holds.
(ii) If ν is any stationary distribution for the latent voter model with $\nu(\xi \equiv 0$ or $\xi \equiv 1) = 0$, then

$$\sup_x |\nu(\xi(x) = 1) - 1/2| \leq \eta.$$

1.6 Evolution of cooperation

In their more general work Ohtsuki et al [41] singled out the special case of the evolutionary game defined in Example 4, in which each cooperator pays a benefit $b \geq 0$ to each neighbor at a cost $c \geq 0$ per neighbor, while each defector accepts the benefit but pays no cost. The resulting payoff matrix is then

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} b-c & -c \\ b & 0 \end{pmatrix}. \tag{1.55}$$

In this case the payoff for D always exceeds that for C. As a result in a homogeneously mixing population cooperators will die out. The fact that such cooperative behavior may nonetheless take over in a spatial competition is the reason for interest in these kind of models in evolutionary game theory. In a spatial setting the intuition is that it may be possible to the C's to form cooperatives which collectively have a selective advantage.

The authors used a non-rigorous pair approximation and diffusion approximation to argue that for the cooperator-defector model for large population size N and small selection w, cooperators are “favored” if and only if $b/c > k$. Here “favored” means that starting with a single cooperator the probability that cooperators take over is greater than $1/N$, the corresponding probability in a selectively neutral model. They also carried out a number of simulations which showed reasonable agreement for $N \gg k$ although they noted that $b/c > k$ appeared to be necessary but not sufficient in general. It is instructive for the reader to consider the nearest neighbor case on $\mathbb{Z}$ starting with cooperators to the right of 0 and defectors to the left. It is then easy to check that the C/D interface will drift to the left, and so cooperators take over, if and only if $b/c > 2$. This was noted in [41] as further evidence for their $b/c > k$ rule.

Our main result here (Corollary 1.14 below) is a rigorous verification of the $b/c > k$ rule for general symmetric translation invariant graphs with vertex set $\mathbb{Z}^d$ when w is small.

From (1.19) and (1.21) we have (recall h_i is as in (1.29))

$$h_i(x,\xi) = (\theta_i - f_i\phi)(x,\xi), \ \ i = 0, 1.$$

Since $h_0 + h_1 = \theta_0 + \theta_1 - \phi = 0$, we can write

$$\begin{aligned} h_0(x,\xi) &= (\gamma-\beta)kf_0f_1(x,\xi) + k(\delta-\gamma)f_0^{(2)}(x,\xi) \\ &\qquad\qquad - kf_0(x,\xi)[(\alpha-\beta)f_1^{(2)} + (\delta-\gamma)f_0^{(2)}](x,\xi), \\ h_1(x,\xi) &= -h_0(x,\xi). \end{aligned} \tag{1.56}$$

As before, let e_1, e_2, e_3 denote i.i.d. random variables with law p. If P_e denotes averaging over the e_i's then we have

$$\begin{aligned} &f_i(0,\xi) = P_e(\xi(e_1) = i), \ \ f_i^{(2)}(0,\xi) = P_e(\xi(e_1) = i, \xi(e_1+e_2) = i), \qquad (1.57) \\ &f_{i_1}(0,\xi)f_{i_2}^{(2)}(0,\xi) = P_e(\xi(e_1) = i_1, \xi(e_2) = i_2, \xi(e_2+e_3) = i_2), \end{aligned}$$

and similarly for higher order probabilities. We also continue to let $\langle\cdot\rangle_u$ denote expectation on the product space where (e_1, e_2, e_3) and the voter equilibrium ξ are independent. If $\hat{\xi} = 1 - \xi$, then starting with (1.30) we have,

$$\frac{f(u)}{k} = k^{-1}\langle \hat{\xi}(0)h_1(0,\xi) - \xi(0)h_0(0,\xi)\rangle_u = k^{-1}\langle h_1(0,\xi)\rangle_u,$$

where in the last equality (1.56) is used to see that what appears to be a quartic polynomial is actually a cubic. Using (1.57) and some arithmetic we obtain

$$\begin{aligned} \frac{f(u)}{k} =& (\beta-\gamma)\langle\hat{\xi}(e_1)\xi(e_2)\rangle_u + (\alpha-\beta)\langle\hat{\xi}(e_1)\xi(e_2)\xi(e_2+e_3)\rangle_u \\ &+ (\gamma-\delta)\langle\xi(e_1)\hat{\xi}(e_2)\hat{\xi}(e_2+e_3)\rangle_u. \end{aligned} \tag{1.58}$$

To simplify further we will use a simple lemma for coalescing random walk probabilities (Lemma 1.15 below) together with the duality formula (1.40) to establish the following more explicit expression for f, whose proof will be given below:

$$\begin{aligned} \frac{f(u)}{k} &= [(\beta-\delta) + k^{-1}(\gamma-\delta)]p(0|e_1)u(1-u) \\ &\quad + [(\alpha-\beta)-(\gamma-\delta)][u(1-u)(p(e_1|e_2,e_2+e_3) + up(e_1|e_2|e_2+e_3))]. \end{aligned} \tag{1.59}$$

Rather than try to analyze this cubic as in Section 1.3, assume $\alpha - \beta = \gamma - \delta$ (which holds in our motivating example) so that f becomes a quadratic with roots at 0 and 1. If $\beta - \delta > k^{-1}(\delta - \gamma)$, then f is strictly positive on $(0,1)$ and so Proposition 1.6 shows the PDE solutions will converge to 1. If $\beta - \delta < k^{-1}(\delta - \gamma)$, then f is strictly negative on $(0,1)$ and so by symmetry the PDE solutions will converge to 0. As a result for $w = \varepsilon^2$ small, in the former case we expect 1's to take over and in the latter case we expect 0's to take over, and this is in fact the case.

Theorem 1.13. *Consider the spin-flip system on $\mathbb{Z}^d$ ($d \geq 3$) with rates given by* (1.17) *where $\alpha - \beta = \gamma - \delta$. If $\gamma - \delta > k(\delta - \beta)$, then 1's take over for $w > 0$ sufficiently small; if $\gamma - \delta < k(\delta - \beta)$, then 0's take over for $w > 0$ sufficiently small.*

Proof of Theorem 1.13. This is now an easy application of Theorem 1.5. Assume $\gamma - \delta < k(\delta - \beta)$. Then (1.59) shows that $f(u) = c_1 u(1-u)$ for $c_1 < 0$. Proposition 1.7 shows that Assumption 2 of Theorem 1.5 is valid for any $u_1 \in (0,1)$. The condition (1.41) holds with $r_0 = 2$ by (1.22) (recall $\|g_i^\varepsilon - g_i\|_\infty = \|\hat{g}_i^\varepsilon - \hat{g}_i\|_\infty$ by Proposition 1.1). The condition (1.43) is clear from the expression for h_1^ε in (1.19). Since $f'(0) < 0$ is clear from the above, and $w = \varepsilon^2$, Theorem 1.5 completes the proof in this case. The case where the inequality is reversed follows by a symmetrical argument, or, if you prefer, just reverse the roles of 0 and 1. $\square$

The particular instance of (1.55) follows as a special case.

Corollary 1.14. *Consider the spin-flip system on $\mathbb{Z}^d$ ($d \geq 3$) with rates given by* (1.17) *where the payoff matrix is given by* (1.55)*. If $b/c > k$, then the cooperators take over for $w > 0$ sufficiently small, and if $b/c < k$, then the defectors take over for $w > 0$ sufficiently small.*

Proof. In this case $\alpha - \beta = \gamma - \delta = b$, $\delta - \beta = c$, and so we have $\gamma - \delta > k(\delta - \beta)$ iff $b > kc$ iff $b/c > k$. $\square$

Turning now to the missing details.

Lemma 1.15. *(a)* $p(e_1|e_2) = p(0|e_1)$.
(b) $p(e_1|e_2 + e_3) = \left(1 + \frac{1}{k}\right) p(0|e_1)$.

Proof. Let $\tilde{B}_t^x$ denote a rate 2 random walk with kernel p starting at x, and note that for $x \neq 0$, $\hat{B}^x - \hat{B}^0$ has the same law as $\tilde{B}^x$ until it hits 0. Note also that for $x \neq 0$, $P(\tilde{B}_t^x \neq 0 \ \forall\ t \geq 0) = \sum_y p(y) P(\tilde{B}_t^{x+y} \neq 0 \ \forall\ t \geq 0)$. For (a),

$$\begin{aligned} p(0|e_1) &= \sum_{x_1} p(x_1) P(\tilde{B}_t^{x_1} \neq 0 \text{ for all } t \geq 0) \\ &= \sum_{x_1} \sum_{x_2} p(x_1)p(x_2) P(\tilde{B}_t^{x_1+x_2} \neq 0 \text{ for all } t \geq 0) \quad (\text{use } x_1 \neq 0) \\ &= \sum_{x_1} \sum_{x_2} p(x_1)p(x_2) P(\tilde{B}_t^{x_1-x_2} \neq 0 \text{ for all } t \geq 0) \quad (\text{by symmetry}) \\ &= p(e_1|e_2). \end{aligned}$$

For (b), let $T_j(x)$ be the time of the jth jump of $\tilde{B}^x$. Then using symmetry,

$$\begin{aligned} p(e_1|e_2 + e_3) &= \sum_{x_1,x_2,x_3} p(x_1)p(x_2)p(x_3) P(\tilde{B}_t^{x_1+x_2+x_3} \neq 0 \text{ for all } t \geq 0) \\ &= \sum_{x_1} p(x_1) P(\tilde{B}_t^{x_1} \neq 0 \text{ for all } t \geq T_2(x_1)). \end{aligned}$$

Now using the above and first equality in the proof of (a),

$$
\begin{aligned}
p(e_1|e_2+e_3) - p(0|e_1) &= \sum_{x_1} p(x_1) P(\tilde{B}^{x_1}_{T_1} = 0,\ \tilde{B}^{x_1}_t \neq 0 \text{ for all } t \geq T_2(x_1)) \\
&= k^{-1} \sum_{x_1} p(x_1) P(\tilde{B}^0_t \neq 0 \text{ for all } t \geq T_1) \\
&= k^{-1} p(0|e_1).
\end{aligned}
$$

The result follows. □

Proof of (1.59). We first rewrite (1.58) as

$$
\begin{aligned}
\frac{f(u)}{k} =&(\beta-\gamma)\langle \hat{\xi}(e_1)\xi(e_2)\rangle_u + (\gamma-\delta)\langle \xi(e_1)\hat{\xi}(e_2)\hat{\xi}(e_2+e_3) + \hat{\xi}(e_1)\xi(e_2)\xi(e_2+e_3)\rangle_u \\
&+ ((\alpha-\beta)-(\gamma-\delta))\langle \hat{\xi}(e_1)\xi(e_2)\xi(e_2+e_3)\rangle_u \\
=&I + II + III.
\end{aligned}
\tag{1.60}
$$

Some elementary algebra shows that

$$
II = (\gamma-\delta)\langle \xi(e_1) - \xi(e_1)\xi(e_2) - \xi(e_1)\xi(e_2+e_3) + \xi(e_2)\xi(e_2+e_3)\rangle_u. \tag{1.61}
$$

Note that (1.40) and Lemma 1.15(a) imply

$$
\begin{aligned}
\langle \xi(e_1)\xi(e_2)\rangle_u =&u^2 p(e_1|e_2) + u(1-p(e_1|e_2)) \\
=&u^2 p(0|e_1) + u(1-p(0|e_1)) \\
=&\langle \xi(0)\xi(e_1)\rangle_u = \langle \xi(e_2)\xi(e_2+e_3)\rangle_u,
\end{aligned}
$$

the last by translation invariance. Using this in (1.61) and again applying (1.40), we get

$$
\begin{aligned}
I + II &= (\beta-\gamma)u(1-u)p(e_1|e_2) + (\gamma-\delta)[u - \langle \xi(e_1)\xi(e_2+e_3)\rangle_u] \\
&= (\beta-\gamma)u(1-u)p(e_1|e_2) \\
&\quad + (\gamma-\delta)[u - u^2 p(e_1|e_2+e_3) - u(1-p(e_1|e_2+e_3))] \\
&= u(1-u)[(\beta-\gamma)p(e_1|e_2) + (\gamma-\delta)p(e_1|e_2+e_3)] \\
&= u(1-u)p(0|e_1)[(\beta-\gamma) + (1+k^{-1})(\gamma-\delta)],
\end{aligned}
\tag{1.62}
$$

where Lemma 1.15 is used in the last equality. A straightforward application of (1.40) allows us to find the coefficients of the cubic III in (1.60) and we obtain the required expression for $f(u)/k$. □

1.7 Nonlinear voter models

If the states of adjacent sites were independent in the discrete time nonlinear voter model defined in Example 1, then the density would evolve according to the mean field dynamics

$$
\begin{aligned}
x_{t+1} = h(x_t) = &p_1 \cdot 5x_t(1-x_t)^4 + p_2 \cdot 10x_t^2(1-x_t)^3 \\
&+ (1-p_2)\cdot 10x_t^3(1-x_t)^2 + (1-p_1)\cdot 5x_t^4(1-x_t) + x_t^5
\end{aligned}
$$

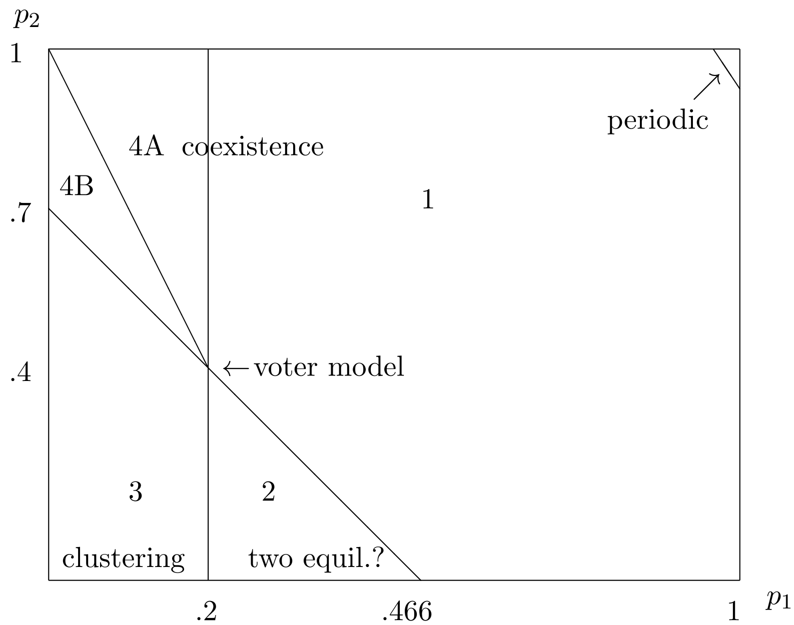


Figure 3: Conjectured phase diagram for the discrete time two-dimensional nonlinear voter model of [39].

Based on simulations and an analysis of the mean-field equation, Molofsky et al [39] predicted the phase diagram given in Figure 3. To explain this, $h(x) = x$ is a fifth degree equation with 0, 1/2, and 1 as roots. $h'(0) = h'(1) = 5p_1$ so 0 and 1 are locally attracting if $5p_1 < 1$ and unstable if $5p_1 > 1$. $h'(1/2) = (15 - 15p_1 - 10p_2)/8$, so 1/2 is locally attracting if $15p_1 + 10p_2 > 7$ and unstable if $15p_1 + 10p_2 < 7$. From the stability properties of 0, 1/2, and 1, it is easy to determine when there are additional roots α and $1 - \alpha$ in the unit interval and whether or not they are stable. The four shapes are given in Figure 4. To make the drawing easier we have represented the quintic as a piecewise linear function.

The implications of the shape of $f(u)$ ($= h(u) - u$ in the above) for the behavior for the system will be discussed below in the context of a similar system in continuous time. There we will see that the division between 4A and 4B is dictated by the speed of traveling waves for the PDE. Here we have drawn the "Levin line" $6p_1 + 2p_2 = 2$ which comes from computing the expected number of 1's at time 1 when we have two adjacent 1's at time 0. Simulations suggest that the true boundary curve exits the square at $(0.024, 1)$, see page 280 in [39].

We abuse our notation as before and incorporate expectation with respect to an independent copy of $Y = (Y^1, \dots, Y^4)$ in our voter equilibrium expectation $\langle\cdot\rangle_u$. If

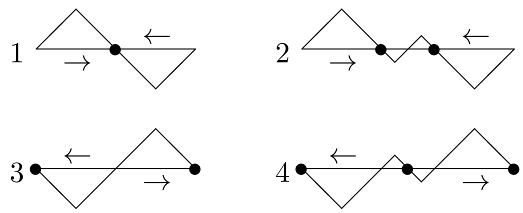


Figure 4: Four possible shapes of the symmetric quintic f. Black dots indicate the locations of stable fixed points.

$Y^0 \equiv 0$, then a short calculation shows that our reaction function in (1.30) is now

$$f(u) = \sum_{j=1}^{4} a(j)(q_j(u) - q_j(1-u)), \tag{1.63}$$

where

$$q_j(u) = \binom{4}{j} \langle \prod_{i=0}^{4-j} (1 - \xi(Y^i)) \prod_{i=5-j}^{4} \xi(Y^i) \rangle_u.$$

Clearly

$$f(0) = f(1) = f(1/2) = 0 \text{ and } f(u) = -f(1-u). \tag{1.64}$$

It does not seem easy to calculate f explicitly, but if L is large, most of the sum comes from Y^i that are well separated and so the ξ values at the above sites should be nearly independent. To make this precise let $A = \{Y^{5-j}, \dots, Y^4\}$, $B = \{Y^0, \dots, Y^{4-j}\}$, $(1 \le j \le 4)$, and note by (1.40) that

$$\begin{aligned} q_j(u) &= \binom{4}{j} \sum_{i=1}^{j} \sum_{k=1}^{5-j} u^i (1-u)^k P(|\hat{\xi}_\infty^A| = i, |\hat{\xi}_\infty^B| = k, \tau(A,B) = \infty) \\ &= \binom{4}{j} u^j (1-u)^{5-j} + \hat{q}_j(u), \end{aligned}$$

where

$$\begin{aligned} \hat{q}_j(u) = \binom{4}{j} \Big[&-u^j(1-u)^{5-j} P(|\hat{\xi}_\infty^{A\cup B}| < 5) \\ &+ \sum_{i=1}^{j} \sum_{k=1}^{5-j} 1(i+k<5) u^i (1-u)^k P(|\hat{\xi}_\infty^A| = i, |\hat{\xi}_\infty^B| = k, \tau(A,B) = \infty) \Big] \\ = \sum_{i=1}^{5} d_i(j,L) u^i. & \end{aligned}$$

If $\eta_0(L) = P(|\hat{\xi}_\infty^{A\cup B}| < 5)$, that is the probability that there is a coalescence among the random walks starting at $Y^0, \dots, Y^4$, then it follows easily from the above that

$|d_i(j,L)| \le c_0\eta_0(L)$. Use this in (1.63) to conclude that $f(u) = f_1(u) + f_2(u)$, where f_2 includes the (smaller) contributions from the $\hat{q}_j$'s. That is

$$f_2(u) = \sum_{j=1}^{5} e(j,L)u^j, \tag{1.65}$$

where

$$\sup_{1\le j\le 5} |e(j,L)| \le c_1\eta_0(L), \tag{1.66}$$

and

$$\begin{aligned} f_1(u) =& -u[a(4)(1-u)^4 + a(3)\cdot 4u(1-u)^3 + a(2)\cdot 6u^2(1-u)^2 + a(1)\cdot 4u^3(1-u)] \\ &+ (1-u)[a(4)u^4 + a(3)\cdot 4u^3(1-u) + a(2)\cdot 6u^2(1-u)^2 + a(1)\cdot 4u(1-u)^3] \\ =& b_1u(1-u)^4 + b_2u^2(1-u)^3 - b_2u^3(1-u)^2 - b_1u^4(1-u), \end{aligned} \tag{1.67}$$

where $b_1 = 4a(1) - a(4)$ and $b_2 = 6a(2) - 4a(3)$. By symmetry we have

$$f_1(0) = f_1(1) = f_1(1/2) = 0 \text{ and } f_1(u) = -f_1(1-u). \tag{1.68}$$

Clearly $\eta_0(L) \to 0$ as $L \to \infty$, in fact well-known return estimates (such as Lemma 2.6(a) below with $t_0 = 0$, $r_0 = 1$ and p large) and a simple optimization argument show that

$$\eta_0(L) \le C_\delta L^{-[d(d-2)/(2(d-1))]+\delta},\ \delta > 0. \tag{1.69}$$

To prepare for the next analysis we note that

$$\begin{aligned} f_1'(u) =& b_1[(1-u)^4 - 4u(1-u)^3] + b_2[2u(1-u)^3 - 3u^2(1-u)^2] \\ &- b_2[3u^2(1-u)^2 - 2u^3(1-u)] - b_1[4u^3(1-u) - u^4], \end{aligned}$$

and so we have $f_1'(0) = f_1'(1) = b_1$ and $f_1'(1/2) = -(6b_1 + 2b_2)/16$. A little calculus, left for the reader, shows

$$\int_0^{1/2} f_1(u)\,du = \frac{5b_1 + b_2}{192} = -\int_{1/2}^{1} f_1(u)\,du. \tag{1.70}$$

We are now ready to describe the phase diagram for the nonlinear voter. Consult Figure 5 for a picture. Note that in what follows when L is chosen large, it is understood that how large depends on $\bar{a} = (a(1), \dots, a(4))$.

(1) $f_1'(0) > 0$, $f_1'(1/2) < 0$ and so by (1.66) for large enough L the same is true for f. In this case 0, 1/2, and 1 are the only roots of f (all simple) and 1/2 is an attracting fixed point for the ODE. An application of Proposition 1.6 on $[0, 1/2]$ and a comparison principle, showing that solutions depend monotonically on their initial data (see Proposition 2.1 of [2]), to reduce to the case where $v \in [0, 1/2]$, shows that any non-trivial solution u of the PDE (1.33) satisfies $\liminf_{t\to\infty} \inf_{|x|\le 2wt} u(t,x) \ge 1/2$ for some $w > 0$. The same reasoning with 0 and 1 reversed shows the corresponding upper bound of 1/2. Therefore any non-trivial solution of (1.33) will converge to 1/2 and we expect coexistence.

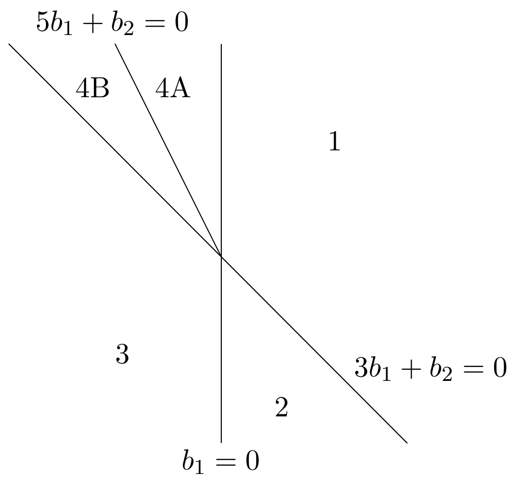


Figure 5: Phase diagram for the continuous time nonlinear voter model with large range in $d \geq 3$. For (b_1, b_2) in regions 1, 2 and $4A$, Theorem 1.16 gives coexistence for L sufficiently large and ε sufficiently small.

(2) $f_1'(0) > 0$, $f_1'(1/2) > 0$ and so by (1.66) for large enough L the same is true for f. In this case 0, 1/2, 1 are unstable fixed points for the ODE and there are attracting fixed points for the ODE at α and $1-\alpha$ for some $\alpha \in (0, 1/2)$. All are simple zeros of f. Another double application of Proposition 1.6 now shows that any non-trivial solution $u(t,x)$ to the PDE will have $\liminf_{t\to\infty} \inf_{|x|\leq 2wt} u(t,x) \geq \alpha$ and $\limsup_{t\to\infty} \sup_{|x|\leq 2wt} u(t,x) \leq 1-\alpha$, so we expect coexistence. Simulations in [39], see Figure 7 and the discussion on page 278 of that work, suggest that in this case there may be two nontrivial stationary distributions: one with density near α and the other with density near $1-\alpha$. The symmetry of f about 1/2 is essential for this last possibility as we note below (see Theorem 1.17).

(3) $f_1'(0) < 0$, $f_1'(1/2) > 0$ and so by (1.66) for large enough L the same is true for f. In this case 0, 1/2, and 1 are the only roots of f (all simple) and 1/2 is an unstable fixed point while 0 and 1 are attracting. In this bistable case the winner is dictated by the sign of the speed of the traveling wave, but by symmetry (recall (1.64)) the speed is 0. One would guess that clustering occurs in this case and there are only trivial stationary distributions, but our method yields no result.

(4) $f_1'(0) < 0$, $f_1'(1/2) < 0$ and so by (1.66) for large enough L the same is true for f. In this case 0, 1/2, 1 are attracting fixed points and there are unstable fixed points at α and $1-\alpha$ for some $\alpha \in (0, 1/2)$ (all simple zeros of f). By the discussion in Case II in Section 1.3 (with $[0, 1/2]$ and $[1/2, 1]$ in place of

the unit interval) there are traveling wave solutions $w_i(x - c_i t)$, $i = 1, 2$ with $w_1(-\infty) = 1$, $w_1(\infty) = w_2(-\infty) = 1/2$ and $w_2(\infty) = 0$. Symmetry implies $c_2 = -c_1$, but we can have $c_1 < 0 < c_2$ (Case 4A) in which case Proposition 1.9 and its mirror image show that solutions to the PDE will converge to 1/2 providing that the initial condition is bounded away from 0 and 1 on a large enough set. We again use the comparison principle as in Case 1 to assume the initial data takes values in the appropriate interval, $[0, 1/2]$ or $[1/2, 1]$, and assume L is large enough so that the integrals of f_1 and f on $[0, 1/2]$ (and hence on $[1/2, 1]$) have the same sign. Hence we expect coexistence in Case 4A and all invariant distributions to have density near 1/2. If $c_1 > 0 > c_2$ (Case 4B) and L is large enough, there is a standing wave solution $w_0(x)$ of the PDE in $d = 1$ with $w_0(-\infty) = 0$, $w_0(\infty) = 1$ (see p. 284 in [27]), and our method yields no result.

Theorem 1.16. *Assume (b_1, b_2) are as in Case 1, 2 or 4A. If L is sufficiently large (depending on $\bar{a}$) then:*
(a) Coexistence holds for ε small enough (depending on L and $\bar{a}$).
(b) In Case 1 or 4A if $\eta > 0$ there is an $\varepsilon_0(\eta, L, \bar{a})$ so that if $0 < \varepsilon \le \varepsilon_0$ and ν is any stationary distribution for the nonlinear voter model satisfying $\nu(\xi \equiv 0$ or $\xi \equiv 1) = 0$, then

$$\sup_x \Big| \nu(\xi(x) = 1) - \frac{1}{2} \Big| \le \eta.$$

Remark. Case 4A is of particular interest as there is coexistence even though $f'(0) < 0$. Here the low density limit theorem in [7] shows convergence to super-Brownian motion with drift $f'(0) < 0$ (see the discussion in Section 1.8 below). From this one might incorrectly guess (after an exchange of limits) that there is a.s. extinction of 1's for the nonlinear voter model, while our proof of coexistence will show that there is positive probability of survival of 1's even starting with a single 1.

Proof of 1.16. Consider Case 4A first. As pointed out above, in this case for L sufficiently large we may employ the mirror image of Proposition 1.9 on $[0, 1/2]$ with $\rho = a$, the unique root of f in $(0, 1/2)$, and Proposition 1.9 on $[1/2, 1]$ with $\rho = 1 - a$, along with the comparison principle (Proposition 2.1 in [2]), to see that Assumption 1 holds for $\varepsilon < \varepsilon_0(\eta)$ with $u_* = \frac{1}{2} - \eta$, $u^* = \frac{1}{2} + \eta$, $v_0 = \delta$, and $v_1 = 1 - \delta$. (1.41) is trivial because $g_i^\varepsilon = g_i$. Theorem 1.4 now implies (a) and (b) in this case. The proofs in Cases 1 and 2 are similar using Proposition 1.6 (note all the zeros are simple in these cases) to verify Assumption 1 (see the above discussion in these cases). $\square$

We next show that more can be said in Case 2 if we break the symmetry. This also demonstrates how one can handle higher degree reaction functions in the pde and still apply the general results in Section 1.2. Consider the nonlinear voter model ξ as before but now for $\lambda > 0$ replace g_1 with

$$g_{1,\lambda}(\xi_1, \dots, \xi_4)) = (1 + \lambda) a\Big(\sum_1^4 \xi_i\Big), \tag{1.71}$$

while g_0 is unchanged. To avoid trivialities we assume $\sum_1^4 a(j) > 0$. A short calculation now shows that if f is as in (1.63), then our reaction function in (1.30) becomes

$$f_{(\lambda)}(u) = f(u) + \lambda \sum_{j=1}^{4} a(j) q_j(u) \equiv f(u) + \lambda f_0(u) > f(u) \text{ on } (0,1). \tag{1.72}$$

Decomposing $q_j(u)$ as before we get

$$f_{(\lambda)}(u) = f_{1,\lambda}(u) + f_{2,\lambda}(u),$$

where

$$\begin{aligned} f_{1,\lambda}(u) &= (b_1 + 4\lambda a(1))u(1-u)^4 + (b_2 + 6\lambda a(2))u^2(1-u)^3 \\ &\quad - (b_2 - 4\lambda a(3))u^3(1-u)^2 - (b_1 - \lambda a(4))u^4(1-u) \\ &= f_1(u) + \lambda f_3(u) > f_1(u) \text{ on } (0,1), \end{aligned} \tag{1.73}$$

$$f_{2,\lambda}(u) = \sum_{j=1}^{5} e(j, L, \lambda) u^j, \text{ and } \sup_{1 \le j \le 5} |e(j, L, \lambda)| \le c_2(\lambda + 1)\eta_0(L), \tag{1.74}$$

and f_1 is as in (1.67). We also have

$$f'_{1,\lambda}(0) = b_1 + 4\lambda a(1), \quad f'_{1,\lambda}(1) = b_1 - \lambda a(4). \tag{1.75}$$

Theorem 1.17. *Suppose $b_1 > 0$ and $0 < \lambda < b_1/a(4)$.*
(a) Coexistence holds for large L and small enough ε (depending on L, λ and $\bar{a}$).
(b) Assume $3b_1 + b_2 < 0$ and let $1 - \alpha'$ denote the largest root of $f_1(u) = 0$ in $(0,1)$. If $\eta > 0$, $L > \varepsilon_1(\eta, \bar{a})^{-1}$, $0 < \lambda < \varepsilon_1(\eta, \bar{a})$, $0 < \varepsilon < \varepsilon_0(\eta, L, \lambda, \bar{a})$ and ν is any stationary distribution satisfying $\nu(\xi \equiv 0$ or $\xi \equiv 1) = 0$, then

$$\sup_x \Big| \nu(\xi(x) = 1) - (1 - \alpha') \Big| \le \eta.$$

For a concrete example, consider $a(1) = a(2) = 1$, $a(3) = a(4) = 3$, which is a version of the majority vote plus random flipping. Then $b_1 = 4a(1) - a(4) = 1$, $b_2 = 6a(2) - 4a(3) = -6$, and $3b_1 + b_2 = -3$, and so the hypotheses of (b) hold for small $\lambda > 0$. As L gets large and $\lambda \downarrow 0$, for small enough ε the density of any invariant measure approaches $1 - \alpha'$ the largest root of $f_1(u) = 0$. Of course the closer λ gets to 0, the smaller we must make ε to obtain the conclusion of Theorem 1.17, so we are not able to prove anything about the case $\lambda = 0$. Of course we could also consider $\lambda < 0$ in which case symmetric reasoning would show that the densities approach α'.

Proof of 1.17. (a) The conditions on b_1 and λ imply that $f'_{1,\lambda}(0) > 0$ and $f'_{1,\lambda}(1) > 0$. Coexistence for large L and small ε is now established as in Case 2 (or 1) of Theorem 1.16.

(b) By taking λ and L^{-1} small, depending on $(\eta, \bar{a})$, we see from (1.73), (1.74), (1.75), and our conditions on the b_i that $f_{(\lambda)}(u) = 0$ will have 3 simple roots in $(0, 1)$, $p_1(\lambda) < p_2(\lambda) < p_3(\lambda)$, within $\eta/4$ of the respective roots

$$\alpha' < 1/2 < 1 - \alpha'$$

of $f_1(u) = 0$. As (1.41) is again obvious, we now verify Assumption 1 of Theorem 1.4 with $u^* = 1-\alpha'+\frac{\eta}{2}$, $u_* = 1-\alpha'-\frac{\eta}{2}$, $v_0 \in (p_2, u_*)$, and $v_1 \in (u^*, 1)$ (η is small so these intervals are non-empty). The result would then follow by applying Theorem 1.4. The upper bound (ii) in Assumption 1 is an easy application of Proposition 1.7, with the interval $(p_3, 1)$ in place of $(0, 1)$, and the comparison principle.

For the lower bound (i) in Assumption 1, we use a result of Weinberger [45]. To state the result we need some definitions. His habitat $\mathcal{H}$ will be $\mathbb{R}^d$ in our setting and his space B is the set of continuous functions from $\mathcal{H}$ to $[0, \pi_+]$. In our case $\pi_+ = 1$. His result is for a discrete iteration $u_{n+1} = Q(u_n)$, where in our case $Q(u)$ is solution to the PDE at time 1 when the initial data is u. His assumption (3.1) has five parts:

(i) if $u \in B$ then $Q(u) \in B$.

(ii) If T_y is translation by y then $Q(T_y u) = T_y Q(u)$.

(iii) Given a number β, let $Q(\beta)$ be the constant value of $Q(u_\beta)$ for $u_\beta \equiv \beta$. There are $0 \le \pi_0 < \pi_1 \le \pi_+$ so that if $\beta \in (\pi_0, \pi_1)$ then $Q(\beta) > \beta$. $Q(\pi_0) = \pi_0$ and $Q(\pi_1) = \pi_1$.

(iv) $u \le v$ implies $Q(u) \le Q(v)$.

(v) If $u_n \in B$ and $u_n \to u$ uniformly on bounded sets then $Q(u_n)(x) \to Q(u)(x)$.

Clearly (i) and (ii) hold in our application. For (iii) we let $\pi_0 = p_2(\lambda)$ and $\pi_1 = p_3(\lambda)$. (iv) a consequence of PDE comparison principles, see, e.g., Proposition 2.1 in Aronson and Weinberger (1978). (v) follows from the representation of solutions of the PDE in terms of the dual branching Brownian motion (see Lemma 3.3).

The next ingredient for the result is

$$\mathcal{S} = \{x \in \mathbb{R}^d : x \cdot \xi \le c^*(\xi) \text{ for all } \xi \in S^{d-1}\},$$

where S^{d-1} is the unit sphere in $\mathbb{R}^d$. $c^*(\xi)$ is the wave speed in direction ξ defined in Section 5 of [45]. Due to the invariance of the PDE under rotation, all our speeds are the same, $c^* = \rho$, and $\mathcal{S}$ is a closed ball of radius ρ or the empty set. Here is Theorem 6.2 of [45].

Theorem 1.18. *Suppose (i)–(v) and that the interior of $\mathcal{S}$ is nonempty. Let $\mathcal{S}''$ be any closed and bounded subset of the interior of $\mathcal{S}$. For any $\gamma > \pi_0$, there is an r_γ so that if $u_0(x) \ge \gamma$ on a ball of radius r_γ and if $u_{n+1} = Q(u_n)$ then*

$$\liminf_{n\to\infty} \min_{x \in n\mathcal{S}''} u_n(x) \ge \pi_1. \tag{1.76}$$

To be able to use this result, we have to show that $\rho > 0$. Note that here we require lower bounds on the wave speed of solutions to the reaction diffusion equation in one spatial dimension. This is because traveling wave solutions in the direction ξ of the form $w(x \cdot \xi - \rho t)$ correspond to traveling waves w in one spatial dimension. Recall that in the decomposition (1.72) $f(u)$ is odd about $u = 1/2$, and for large L has $f'(1/2) > 0$ by $3b_1 + b_2 < 0$. The latter shows f has 3 simple zeros in $(0, 1)$ at $\alpha < 1/2 < 1 - \alpha$. The strict inequality in (1.72) on $(0, 1)$ now easily implies (compare the negative and positive humps separately)

$$\int_{p_1}^{p_3} f_{(\lambda)}(u)du > \int_{\alpha}^{1-\alpha} f(u)du = 0. \tag{1.77}$$

So by the discussion in part II(ii) of Section 1.3 there is a one-dimensional decreasing traveling wave solution to (1.33) (with $f = f_{(\lambda)}$) over (p_1, p_3) with positive wave speed $r_2(\lambda)$.

To consider traveling waves over $(0, p_1(\lambda))$, we note that Kolmogorov, Petrovsky, and Piscounov [32] have shown that if we consider

$$\frac{\partial u}{\partial t} = \frac{\sigma^2}{2}\Delta u + f(u)$$

in one dimension where f satisfies

$$f(0) = f(1) = 0, \quad f(u) > 0 \text{ for } 0 < u < 1, \quad f'(u) \le f'(0) \text{ for } 0 < u \le 1 \tag{1.78}$$

then there is a traveling wave solution with speed $\sqrt{2\sigma^2 f'(0)}$ and this is the minimal wave speed. For this fact one can consult Bramson [3] or Aronson and Weinberger [1]. However, the intuition behind the answer is simple: the answer is the same as for the linear equation

$$\frac{\partial u}{\partial t} = \frac{\sigma^2}{2}\Delta u + f'(0)u$$

which gives the mean of branching Brownian motion. For more on this connection, see McKean [38].

Now let $g_1 \le g_2$ be C^1 functions on $[0, 1]$ such that

$$0 < g_2 \le f_{(\lambda)} \text{ on } (0, p_1), \ g_1 = g_2 = f_{(\lambda)} \text{ on } [p_1, 1],$$

$$g_2'(0) \in \left(0, \frac{r_2(\lambda)^2}{2\sigma^2}\right), \ g_2'(u) \le g_2'(0) \text{ on } [0, p_1], \tag{1.79}$$

and for some $0 < p_0 < p_1$,

$$g_1(0) = 0, \ g_1 < 0 \text{ on } (0, p_0), g_1 > 0 \text{ on } (p_0, p_1), \int_0^{p_1} g_1(t)\, dt > 0, \ g_1'(0) < 0. \tag{1.80}$$

The existence of such functions is elementary. By the KPP result above, the minimal wave speed over $(0, p_1(\lambda))$ for the g_2 equation is

$$c_2 = \sqrt{2\sigma^2 g_2'(0)} < r_2(\lambda). \tag{1.81}$$

By Theorem 2.4 and Corollary 2.3 of [26] (or the discussion in part II(ii) of Section 1.3) there is a unique traveling wave solution $u(t,x) = w(x - c_1 t)$ (w decreasing) to the g_1 equation with unique wave speed $c_1 > 0$ (since the integral in (1.80) is positive) and range $(0, p_1)$. Note here and elsewhere that the traveling waves w in [26] are increasing and so our wave speeds have the opposite sign. By a comparison theorem for wave speeds (Proposition 5.5 of [45]) we may conclude that

$$c_2 \geq c_1. \tag{1.82}$$

The hypothesis of the above comparison result is easily verified using $g_1 \leq g_2$ and the standard comparison principle (e.g. Proposition 2.1 of [2]). It follows from (1.81) and (1.82) that $c_1 < r_2(\lambda)$ which are the wave speeds of the g_1 equation over $(0, p_1)$ and (p_1, p_3), respectively. We can therefore apply Theorem 2.7 of [26] to conclude the existence of a traveling wave over $(0, p_3(\lambda))$ for the g_1 equation with speed $r_1(\lambda) \in (c_1, r_2(\lambda))$. The wave and its speed are both unique by Corollary 3.3 of [26]. Since $f_{(\lambda)} \geq g_1$ on $[0, p_3]$, another application of Proposition 5.5 of [45] shows that $\rho \geq r_1(\lambda)$ and in particular $\rho > 0$.

Using (1.76), we have proved that for $0 < 2w = \rho$,

$$\liminf_{n\to\infty} \inf_{|x|\leq 2wn} u(n,x) \geq p_3(\lambda) \geq 1 - \alpha' - \frac{\eta}{4},$$

providing that $u(0,x) \geq v_0$ for $|x| \leq r_{v_0}$. The same reasoning gives the same conclusion with $n\tau$ in place of n for any $\tau > 0$. Taking τ small enough, a simple interpolation argument (use the weak form of the reaction diffusion equation and smoothing properties of the Brownian semigroup) now gives Assumption 1(i) with $u_* = 1 - \alpha' - \frac{\eta}{2}$ where the $2w$ in the above helps a bit in this last interpolation step. □

1.8 Comparison with low density superprocess limit theorem

To make a comparison between our hydrodynamic limit theorem (Theorem 1.3) and the superprocess limit theorem of Cox and Perkins [7] we will write our perturbation terms in a different form, which will also be useful in Section 7. Define

$$\Xi_S(\eta) = \prod_{i\in S} \eta_i \text{ for } \eta = (\eta_1, \dots, \eta_{N_0}) \in \{0,1\}^{N_0}, S \in \hat{\mathcal{P}}_{N_0} = \{\text{subsets of } \{1, \dots, N_0\}\},$$

and

$$\chi(A, x, \xi) = \prod_{y\in A} \xi(x+y),\ x \in \mathbb{Z}^d, \xi \in \{0,1\}^{\mathbb{Z}^d},$$
$$A \in \mathcal{P}_{N_0} = \{\text{subsets of } \mathbb{Z}^d \text{ of cardinality at most } N_0\}.$$

By adding an independent first coordinate to Y we may assume Y^1 has law p. If

$$\tilde{g}_i^\varepsilon(\xi_1, \dots, \xi_{N_0}) = -\varepsilon_1^{-2} 1(\xi_1 = i) + g_i^\varepsilon(\xi_1, \dots, \xi_{N_0}), \tag{1.83}$$

and $\tilde{g}_i$ is as above without the superscript ε, then

$$\lim_{\varepsilon\downarrow 0} \|\tilde{g}_i^\varepsilon - \tilde{g}_i\|_\infty = 0, \tag{1.84}$$

and we may rewrite (1.5) as

$$h_i^\varepsilon(x,\xi) = E_Y(\tilde{g}_i^\varepsilon(\xi(x+Y^1),\dots,\xi(x+Y^{N_0}))),\ i=0,1, \tag{1.85}$$

and similarly without the ε's. It is easy to check that $\{\Xi_S(\cdot) : S \in \hat{\mathcal{P}}_{N_0}\}$ is a basis for the vector space of functions from $\{0,1\}^{N_0}$ to $\mathbb{R}$ and so there are reals $\hat{\beta}_\varepsilon(S), \hat{\delta}_\varepsilon(S)$, $S \in \hat{\mathcal{P}}_{N_0}$, such that

$$\tilde{g}_1^\varepsilon(\eta) = \sum_{S\in\hat{\mathcal{P}}_{N_0}} \hat{\beta}_\varepsilon(S)\Xi_S(\eta), \quad \tilde{g}_0^\varepsilon(\eta) = \sum_{S\in\hat{\mathcal{P}}_{N_0}} \hat{\delta}_\varepsilon(S)\Xi_S(\eta), \tag{1.86}$$

and similarly without the ε's. If $S \in \hat{\mathcal{P}}_{N_0}$, let $Y^S = \{Y^i : i \in S\}$, where $Y \in \mathbb{Z}^{dN_0}$ has law q as usual. Let E_Y denote expectation with respect to Y. It is easy to use (1.85) to check that

$$h_1^\varepsilon(x,\xi) = \sum_{S\in\hat{\mathcal{P}}_{N_0}} \hat{\beta}_\varepsilon(S)E_Y(\chi(Y^S,x,\xi)) = \sum_{A\in\mathcal{P}_{N_0}} \beta_\varepsilon(A)\chi(A,x,\xi) \tag{1.87}$$

$$h_0^\varepsilon(x,\xi) = \sum_{S\in\hat{\mathcal{P}}_{N_0}} \hat{\delta}_\varepsilon(S)E_Y(\chi(Y^S,x,\xi)) = \sum_{A\in\mathcal{P}_{N_0}} \delta_\varepsilon(A)\chi(A,x,\xi), \tag{1.88}$$

where for $A \in \mathcal{P}_{N_0}$,

$$\beta_\varepsilon(A) = \sum_{S\in\hat{\mathcal{P}}_{N_0}} \hat{\beta}_\varepsilon(S)P(Y^S = A), \quad \delta_\varepsilon(A) = \sum_{S\in\hat{\mathcal{P}}_{N_0}} \hat{\delta}_\varepsilon(S)P(Y^S = A). \tag{1.89}$$

Analogous equations to (1.87), (1.88) and (1.89) hold without the ε's.

Now use (1.87) and (1.88) without the ε's, and (1.40) to see that

$$\begin{aligned} f(u) &\equiv \langle (1-\xi(0)h_1(0,\xi) - \xi(0)h_0(0,\xi)\rangle_u \\ &= \sum_{A\in\mathcal{P}_{N_0}} \Big[\beta(A)\Big[\sum_{j=1}^{|A|} u^j(1-u)P(|\hat{\xi}^A_\infty| = j, \tau(A,\{0\}) = \infty)\Big] && (1.90) \\ &\qquad + \beta(\emptyset)(1-u) - \delta(A)\Big[\sum_{j=1}^{|A\cup\{0\}|} u^j P(|\hat{\xi}^{A\cup\{0\}}_\infty| = j)\Big]\Big], && (1.91) \end{aligned}$$

which is a polynomial of degree at most N_0+1 as claimed in Section 1.1. If $\beta(\emptyset) = 0$, then $f(0) = 0$ and

$$f'(0) = \sum_{A\in\mathcal{P}_{N_0}} \beta(A)P(\tau(A) < \infty, \tau(A,\{0\}) = \infty) - \delta(A)P(\tau(A\cup\{0\}) < \infty). \tag{1.92}$$

From (1.86) one easily derives

$$\hat{\beta}_\varepsilon(S) = \sum_{V\subset S}(-1)^{|S|-|V|}\tilde{g}_1^\varepsilon(1_V), \quad \hat{\delta}_\varepsilon(S) = \sum_{V\subset S}(-1)^{|S|-|V|}\tilde{g}_0^\varepsilon(1_V), \tag{1.93}$$

and similarly without the ε's. Therefore

$$|\hat{\beta}_\varepsilon(S) - \hat{\beta}(S)| + |\hat{\delta}_\varepsilon(S) - \hat{\delta}(S)| \le 2^{N_0}(\|\tilde{g}_1^\varepsilon - \tilde{g}_1\|_\infty + \|\tilde{g}_0^\varepsilon - \tilde{g}_0\|_\infty), \tag{1.94}$$

and so

$$\sum_{A\in\mathcal{P}_{N_0}} |\beta_\varepsilon(A) - \beta(A)| + |\delta_\varepsilon(A) - \delta(A)| \le 2^{2N_0}(\|\tilde{g}_1^\varepsilon - \tilde{g}_1\|_\infty + \|\tilde{g}_0^\varepsilon - \tilde{g}_0\|_\infty). \tag{1.95}$$

Our spin-flips are now recast as

$$c_\varepsilon(\varepsilon x, \xi_\varepsilon) = \varepsilon^{-2}c^v(x,\xi) + \sum_{A\in\mathcal{P}_{N_0}} \chi(A,x,\xi)[\beta_\varepsilon(A)(1-\xi(x)) + \delta_\varepsilon(A)\xi(x)],$$

which is precisely (1.17) of [7] with $\varepsilon = N^{-1/2}$. If we assume

$$g_1^\varepsilon(0) = 0 \text{ (and hence } \tilde{g}_1^\varepsilon(0) = \hat{\beta}_\varepsilon(\emptyset) = 0) \text{ for small } \varepsilon, \tag{1.96}$$

and the voter kernel p has finite support, then using the fact that the right-hand side of (1.95) approaches 0 as $\varepsilon \to 0$ (by (1.84)), it is easy to check that all the hypotheses of Corollary 1.8 of [7] hold. Alternatively, in place of the finite support assumption on p one can assume the weaker hypothesis (P4) of Corollary 1.6 of [7], and then apply that result. These results state that for ε as above if $X_t^\varepsilon = \varepsilon^2\sum_{x\in\varepsilon\mathbb{Z}^d}\xi_t^\varepsilon(x)\delta_x$ and $X_0^\varepsilon \to X_0$ weakly in the space $M_F(\mathbb{R}^d)$ of finite measures on $\mathbb{R}^d$, then X^ε converges weakly in the Skorokhod space of $M_F(\mathbb{R}^d)$-valued paths to a super-Brownian motion with drift $\theta = f'(0)$ (as in (1.92)). In this result we are starting $O(\varepsilon^{-2})$ particles on a grid of ε^{-d} ($d \ge 3$) sites per unit volume, so it is a low density limit theorem producing a random limit, whereas Theorem 1.3 is a high density limit theorem producing a pde limit. The latter result gives a natural explanation for the drift θ in the super-Brownian limit which was defined by the right-hand side of (1.92) in [7]. Namely, under (1.96), in the low density limit we would expect a drift of $\lim_{u\to 0} f(u)/u = f'(0)$, which of course happens to equal the summation in (1.92).

1.9 Outline of the Proofs

In Section 2 we first introduce a family of Poisson processes ("graphical representation") which we use to define our voter model perturbation $\xi_t^\varepsilon(x)$ on $x \in \varepsilon\mathbb{Z}^d$. Using this and working backwards in time we define a "dual process" X which is a branching coalescing random walk with particles jumping at rate ε^{-2} according to $p^\varepsilon(x) = p(x/\varepsilon)$ and with a particle at x giving birth to particles at $x+\varepsilon Y^1, \dots x+\varepsilon Y^{N_0}$ when a reaction occurs at x. The ideas in the definition of the dual are a combination of those of Durrett and Neuhauser [20] for systems with fast stirring and those of

Durrett and Zähle [24] for biased voter models that are small perturbations of the voter model.

Duality allows us to compute the value at z at time T by running the dual process X^ε backwards from time T to time 0 starting with one particle at z at time T. Most of the work in Section 2 is to use coupling to show that for small ε, X^ε is close to a branching random walk $\hat{X}^\varepsilon$. Once this is done, it is straightforward to show (in Section 3) that as $\varepsilon \to 0$ the dual converges to a branching Brownian motion $\hat{X}^0$, and then derive Theorem 1.2 which includes convergence of $P(\xi_t^\varepsilon(x) = 1)$ to the solution $u(t,x)$ of a PDE.

Assumption 1 shows that the limiting PDE in Theorem 1.3 will have solutions which stay away from 0 and 1 for large t. A "block construction" as in [14] will be employed in Section 6 to convert this information about the PDE into information about the particle systems. In effect, this allows us to interchange limits as $\varepsilon \to 0$ and $t \to \infty$ and conclude the existence of a nontrivial stationary distribution, and also show that any stationary distribution will have particle density restricted by the asymptotic behavior of the PDE solutions at $t = \infty$.

In order to show that 0's take over, say, we will need several additional arguments. The pde results will only ensure we can get the particle density down to a low level (see Section 4) but clearly we cannot expect to do better than the error terms in this approximation. To then drive the population to extinction on a large region with high probability we will need to refine some coalescing random walk calculations from the low density setting in Cox and Perkins [8]–see Section 7 and especially Lemma 7.6. This is then used as input for another percolation argument of Durrett [13] to guarantee that there are no particles in a linearly growing region. Since there was an error in the original proof of the latter we give all the details here in Sections 5 and 7.3.

Quantifying the outline above, the first step in the proof of Theorem 1.5, taken in Section 4 is to use techniques of Durrett and Neuhauser [20] to show that if ξ_0^ε has density at most u_1 on $[-L,L]^d$ then at time $T_1 = c_1 \log(1/\varepsilon)$ the density is at most ε^β on $[-L-wT_1, L+wT_1]^d$ (see Lemma 4.2). Here β is a small parameter. The second step, taken in Section 7, is to show that if one waits an additional $T_2 = c_2 \log(1/\varepsilon)$ units of time then there will be no particles in $[-L-wT_1+AT_2, L+wT_1-A_2T_2]^d$ at time T_1+T_2. The first step here (Lemma 7.6) is to show that if we start a finite (rescaled) block of ones of density at most ε^β then with probability at least $1-\varepsilon^{\beta/2}$ it will be extinct by time $C\log(1/\varepsilon)$. Here it is convenient that arguments for the low density regime of [7] (density ε^{d-2}) continue to work all the way up to ε^β and also that the PDE arguments can be used to reduce the density down to ε^β. In short, although the precise limit theorems, Theorem 1.2 and Corollary 1.8 in [7] apply in disjoint regimes (particle densities of 1 and ε^{d-2}, respectively) the methods underlying these results apply in overlapping regimes which together allow us to control the underlying particle systems completely. Of course getting 1's to be extinct in a large block does not give us what we want. The block construction in [20] is suitably modified to establish complete extinction of 1's on a linearly growing set. A comparison result of [37], suitably modified to accommodate our percolation process (see Lemma 5.1), is used to simplify this construction.

2 Construction, Duality and Coupling

In this section, we first introduce a family of Poisson processes which we use to define ξ_t, our voter model perturbation on $\varepsilon\mathbb{Z}^d$, a "dual process" X and a "computation process" ζ. The duality equation (2.17) below gives a representation of $\xi_t(x)$ in terms of (X,ζ). Next we show that for small ε, (X,ζ) is close to the simpler $(\hat{X},\hat{\zeta})$, where $\hat{X}$ is a branching random walk system with associated computation process $\hat{\zeta}$. Finally we show by a strong invariance principle that for small ε, $(\hat{X},\hat{\zeta})$ is close to a branching Brownian motion and its associated computation process.

However, our first task will be to prove Proposition 1.1 and reduce to the case where $\varepsilon_1=\infty$ in (1.5).

2.1 Preliminaries

Proof of Proposition 1.1. Let $\underline{p}=\min\{p(y_i):p(y_i)>0\}$, choose $\varepsilon_0>0$ so that $M=\sup_{0<\varepsilon\le\varepsilon_0}\|\hat{g}_0^\varepsilon\|_\infty\vee\|\hat{g}_1^\varepsilon\|_\infty<\infty$ and then choose $\varepsilon_1>0$ so that

$$\varepsilon_1^{-2}\underline{p}>M. \tag{2.1}$$

For $0<\varepsilon<\varepsilon_0$ define g_i^ε on $\{0,1\}^{N_0}$ by

$$g_i^\varepsilon(\xi_1,\dots,\xi_{N_0})=\varepsilon_1^{-2}\sum_1^{N_0}1(\xi_j=i)p(y_j)+\hat{g}_i^\varepsilon(\xi_1,\dots,\xi_{N_0}),\ i=0,1, \tag{2.2}$$

and define g_i by the same equation without the ε's. Clearly $\|g_i^\varepsilon-g_i\|_\infty=\|\hat{g}_i^\varepsilon-\hat{g}_i\|_\infty\to0$ as $\varepsilon\to0$. We may assume $y_1=0$. By replacing $\hat{g}_i^\varepsilon$ with $\hat{g}_i^\varepsilon 1(\xi_1=1-i)$ and redefining h_i^ε analogously (this will not affect (1.27)), we may assume

$$\hat{g}_i^\varepsilon(\xi_1,\dots,\xi_{N_0})=0 \text{ if } \xi_1=i. \tag{2.3}$$

We now show that $g_1^\varepsilon\ge0$. Assume first

$$\sum_1^{N_0}\xi_i p(y_i)=0. \tag{2.4}$$

Choose $\xi\in\{0,1\}^{\mathbb{Z}^d}$ so that $\xi(y_i)=\xi_i$. If $\xi(0)=0$, then by (1.27), (1.8), (2.2), and (2.4) (the latter to make the voter part of c_ε in (1.27) and the first term in (2.2) vanish),

$$0\le c_\varepsilon(0,\xi_\varepsilon)=\hat{g}_1^\varepsilon(\xi(y_1),\dots,\xi(y_{N_0}))=g_1^\varepsilon(\xi_1,\dots,\xi_{N_0}).$$

If $\xi(0)=1$, then $\xi_1=\xi(0)=1$ and by (2.3), $g_1^\varepsilon(\xi_1,\dots,\xi_{N_0})=0$. Assume next that

$$\sum_1^{N_0}\xi_i p(y_i)>0.$$

Then the above sum is at least $\underline{p}$ and so

$$g_1^\varepsilon(\xi_1,\dots,\xi_{N_0})\ge\varepsilon_1^{-2}\underline{p}-\|\hat{g}_1^\varepsilon\|_\infty\ge\varepsilon_1^{-2}\underline{p}-M>0,$$

the last by (2.1). This proves $g_1^\varepsilon \geq 0$ and a similar argument shows $g_0^\varepsilon \geq 0$. Finally (1.5) with $Y^i = y_i$ is immediate from (1.8) and the definition of g_i^ε. $\square$

We claim we may assume without loss of generality that $\varepsilon_1 = \infty$ in (1.5), that is, the first term in the right-hand side of (1.5) is absent. To see why, let $\tilde{\varepsilon}^{-2} = \varepsilon^{-2} - \varepsilon_1^{-2}$ for $\varepsilon < \varepsilon_1$, and use (1.5) in (1.27) to rewrite the spin-flip rates of ξ^ε as

$$c^\varepsilon(\varepsilon x, \xi_\varepsilon) = \tilde{\varepsilon}^{-2} c^v(x,\xi) + (1-\xi(x))\tilde{h}_i^\varepsilon(x,\xi) + \xi(x)\tilde{h}_0^\varepsilon(x,\xi),$$

where

$$\tilde{h}_i^\varepsilon(x,\xi) = E_Y(g_i^\varepsilon(\xi(x+Y^1),\ldots,\xi(x+Y^{N_0}))). \tag{2.5}$$

So by working with $\tilde{h}_i^\varepsilon$ in place of h_i^ε throughout, we may use (2.5) in place of (1.5) and effectively set $\varepsilon_1 = \infty$. Note first that this does not affect the definition of the reaction term $f(u)$ in the PDE (1.33) since the terms involving $\varepsilon^{-2} f_i(x,\xi)$ cancel in (1.30). The only cost is that ε^{-2} is replaced with $\tilde{\varepsilon}^{-2}$. The ratio of these terms approaches 1 and so not surprisingly this only affects some of the proofs in a trivial manner. Rather than carry this $\tilde{\varepsilon}^{-2}$ with us throughout, we prefer to use ε and so

$$\textbf{henceforth set } \varepsilon_1 = \infty \textbf{ in } (1.5). \tag{2.6}$$

2.2 Construction of ξ_t

Define $c^* = c^*(g)$ by

$$c^* = \sup_{0<\varepsilon\leq\varepsilon_0/2} \|g_1^\varepsilon\|_\infty + \|g_0^\varepsilon\|_\infty + 1. \tag{2.7}$$

To construct the process, we use a graphical representation. For $x \in \varepsilon\mathbb{Z}^d$, introduce independent Poisson processes $\{T_n^x, n \geq 1\}$ and $\{T_n^{*,x}, n \geq 1\}$ with rates ε^{-2} and c^*, respectively. Recall $p_\varepsilon(y) = p(y/\varepsilon)$ for $y \in \varepsilon\mathbb{Z}^d$ and let $q_\varepsilon(y) = q(y/\varepsilon)$ for $y \in \varepsilon\mathbb{Z}^{dN_0}$. For $x \in \varepsilon\mathbb{Z}^d$ and $n \geq 1$, define independent random variables $Z_{x,n}$ with distribution p_ε, $Y_{x,n} = (Y_{x,n}^1, \ldots, Y_{x,n}^{N_0})$ with distribution q_ε, and $U_{x,n}$ uniform on $(0,1)$. These random variables are independent of the Poisson processes and all are independent of an initial condition $\xi_0 \in \{0,1\}^{\varepsilon\mathbb{Z}^d}$.

At times $t = T_n^x, n \geq 1$ (called voter times), we set $\xi_t(x) = \xi_{t-}(x + Z_{x,n})$. To facilitate the definition of the dual, we draw an arrow from $(x, T_n^x) \to (x + Z_{x,n}, T_n^x)$. At times $t = T_n^{*,x}$, $n \geq 1$ (called reaction times), if $\xi_{t-}(x) = i$ we set $\xi_t(x) = 1 - i$ if

$$U_{x,n} < g_{1-i}^\varepsilon(\xi_{t-}(x + Y_{x,n}^1), \ldots, \xi_{t-}(x + Y_{x,n}^{N_0}))/c^*, \text{ and otherwise } \xi_t(x) = \xi_{t-}(x).$$

At these times, we draw arrows from $(x, T_n^{*,x}) \to (x + Y_{x,n}^i, T_n^{*,x})$ for $1 \leq i \leq N_0$. We write a * next to $(x, T_n^{*,x})$ and call these *-arrows. It is not hard to use ideas of Harris [30] to show that under the exponential tail conditions on p and q, (1.1) and (1.6), this recipe defines a pathwise unique process. This reference assumes finite range interactions but the proof applies in our infinite range setting as there are still finitely many sites that need to be checked at each reaction time. To verify this construction and to develop a useful dual process we now show how to compute the state of x at time t by working backwards in time. It is easy to verify that ξ is the unique in law $\{0,1\}^{\mathbb{Z}^d}$-valued Feller process with rates given by (1.2) and (1.27). For example one can recast the graphical representation in terms of SDE's driven by Poisson point processes and use stochastic calculus as in Proposition 2.1(c) of [8] (it is easy to verify condition (2.3) of that reference in our current setting).

We use $B^{\varepsilon,x}$ to denote a continuous time random walk with jump rate ε^{-2} and jump distribution p_ε starting at $x \in \varepsilon\mathbb{Z}^d$ and drop dependence on x if $x = 0$. We also assume

$$\{B^{\varepsilon,x} : x \in \varepsilon\mathbb{Z}^d\} \text{ are independent random walks distributed as above.} \tag{2.8}$$

It will be convenient to extend the Poisson times to the negative time line indexed by non-positive integers, and hence have $\{T_n^x, n \in \mathbb{Z}\}$, $\{T_n^{*,x}, n \in \mathbb{Z}\}$ with the associated $\{Z_{x,n}, n \in \mathbb{Z}\}$ and $\{(Y_{x,n}, U_{x,n}), n \in \mathbb{Z}\}$, respectively. At times it is useful to work with the associated independent Poisson point processes of reaction events $\Lambda_r^x(dt, dy, du)$ ($x \in \varepsilon\mathbb{Z}^d$) on $\mathbb{R} \times \varepsilon\mathbb{Z}^{dN_0} \times [0,1]$) with points $\{(T_n^{*,x}, Y_{x,n}, U_{x,n})\}$ and intensity $c^* dt \times q_\varepsilon \times du$, and also the independent Poisson point processes of walk steps $\Lambda_w^x(dt, dz)$ ($x \in \varepsilon\mathbb{Z}^d$) on $\mathbb{R} \times \varepsilon\mathbb{Z}^d$ with points $\{(T_n^x, Z_{x,n})\}$ and intensity $\varepsilon^{-2} dt \times p_\varepsilon$.

2.3 The Dual X

Fix $T > 0$ and a vector of $M+1$ distinct sites $z = (z^0, \ldots, z^M)$, each $z_i \in \varepsilon\mathbb{Z}^d$. Our dual process $X = X^{z,T}$ starts from these sites at time T and works backwards in time to determine the values $\xi_T(z_i)$. X will be a *coalescing branching random walk* with $X_0 = (z_0, \ldots, z_M, \infty, \ldots)$ taking values in

$$\begin{aligned}\mathcal{D} = \{(X^0, X^1, \ldots) \in D([0,T], \mathbb{R}^d \cup \{\infty\})^{\mathbb{Z}_+} : \\ \exists K_0 \in \mathbb{Z}_+ \text{ s.t. } X_t^k = \infty \ \forall t \ \in [0,T] \ \text{ and } k > K_0\}.\end{aligned}$$

Here ∞ is added to $\mathbb{R}^d$ as a discrete point, $D([0,T], \mathbb{R}^d \cup \{\infty\})$ is given the Skorokhod J_1 topology, and $\mathcal{D}$ is given the product topology. The dependence of the dual on T is usually suppressed in our notation.

For $X = (X^0, X^1, \ldots) \in \mathcal{D}$, let $K(t) = \max\{i : X_t^i \neq \infty\}$, define $i \sim_t i'$ iff $X_t^i = X_t^{i'} \neq \infty$, and choose the minimal index j in each equivalence class in $\{0, \ldots K(t)\}$ to form the set $J(t)$. We also introduce

$$I(t) = \{X_t^i : i \in J(t)\} = \{X_t^i : X_t^i \neq \infty\}.$$

Durrett and Neuhauser [20] call $I(t)$ the influence set because it gives the locations of the sites we need to know at time $T - t$, to compute the values at $z^0, \ldots, z^M$ at time T.

To help digest the definitions, the reader should consult Figure 6, which shows a realization of the dual starting from a single site when $N_0 = 3$. If there were no reaction times $T_n^{*,x}$ then the coordinates $X_t^j, j \in J(t)$ follow the system of coalescing random walks dual to the voter part of the dynamics. Coalescing refers to the fact that if $X_s^j = X_s^{j'}$ for some s and j, j', then $X_t^j = X_t^{j'}$ for all $t \in [s,T]$. Jumps

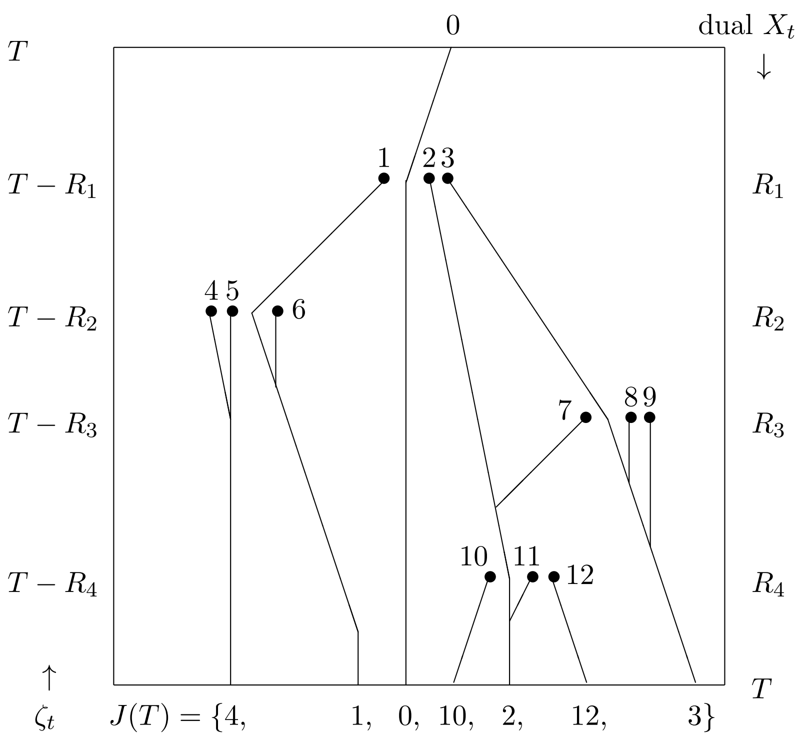


Figure 6: An example of the dual with $N_0 = 3$.

occur when a particle in the dual encounters the tail of an arrow in the graphical representation. That is, if $j \in J(s-)$ and $x = X^j_{s-}$ has $T-s = T^x_n$ then $X^j_s = x + Z^x_n$. It coalesces with $X^i_s = x + Z^x_n$ if such an i exists, meaning that $i \vee j$ is removed from $J(s-)$ to form $J(s)$. If B^ε is a rate ε^{-2} random walk on $\varepsilon\mathbb{Z}^d$ with step distribution p_ε then the coalescing random walks in the dual X follow coalescing copies of B^ε.

To complete the definition, we have to explain what happens at the reaction times. Put $R_0 = 0$, and for $m \geq 1$ let R_m be the first time $t > R_{m-1}$ that a particle in the dual encounters the tail of a *-arrow. If

$$j \in J(R_m-) \text{ and } x = X^j_{R_m-} \text{ has } T - R_m = T^{x,*}_n \text{ for some n,} \tag{2.9}$$

we let $\mu_m = j$ denote the parent site index. In the example in Figure 6 $\mu_1 = 0$, $\mu_2 = 1$, $\mu_3 = 3$, and $\mu_4 = 2$.

We create N_0 new walks by setting $Y^i_m = Y^i_{x,n}$, $1 \leq i \leq N_0$,

$$\begin{aligned} &K(R_m) = K(R_{m-1}) + N_0\,, \text{ and} \\ &X^{K(R_{m-1})+i}_{R_m} = x + Y^i_m,\ i = 1, \ldots, N_0\,. \end{aligned} \tag{2.10}$$

The values of the other coordinates $X^{j'}$, $j' \in J(R_m-)$, $j' \neq \mu_m$ remain unchanged. Each "new" particle immediately coalesces with any particle already at the site where it is born, and we make the resulting changes to $J(R_m-)$ to construct $J(R_m) \supset J(R_m-)$. To compute $\xi_T(z^i)$, we will also need the random variables

$$U_m = U_{x,n} \text{ where } x,\ m, \text{ and } n \text{ are as in (2.9).} \tag{2.11}$$

This computation is described in the next subsection.

$K(s)$ changes only at reaction times and always increases by exactly N_0, so

$$K(s) = M + mN_0, \text{ for } s \in [R_m, R_{m+1}). \tag{2.12}$$

Let $\mathcal{F}_t$ be the right-continuous (time reversed) filtration generated by the graphical representation restricted to $[T-t, T)$, but excluding the $\{U_{x,n}\}$. More precisely $\mathcal{F}_t$ is the right-continuous filtration generated by

$$\begin{aligned} &\{\Lambda^x_w([T-s,T) \times A) : s \leq t, x \in \varepsilon\mathbb{Z}^d, A \subset \varepsilon\mathbb{Z}^d\}, \\ &\{\Lambda^x_r([T-s,T) \times B \times [0,1]) : s \leq t, B \subset \varepsilon\mathbb{Z}^{dN_0}, x \in \varepsilon\mathbb{Z}^d\}. \end{aligned} \tag{2.13}$$

The $\{R_m\}$ are then $(\mathcal{F}_t)$-stopping times and X is $(\mathcal{F}_t)$-adapted. Since $P(R_{m+1} - R_m \in \cdot | \mathcal{F}_{R_m})$ is stochastically bounded below by an exponential random variable with mean $(c^*(M + mN_0))^{-1}$, $R_m \uparrow \infty$ a.s. (recall our graphical variables were extended to negative values of time) and the definition of X is complete.

Note that

$$\mu_m \text{ is } \mathcal{F}_{R_m} - \text{measurable and } \delta_{Y_m,U_m} = \Lambda_r^{X^{\mu_m}_{R_m-}}(\{T - R_m\} \times \cdot). \tag{2.14}$$

As the above time reversed Poisson point processes are also Poisson point processes, one may easily see that

$$\{Y_m\} \text{ are iid with law } q_\varepsilon \text{ and } Y_m \text{ is } \mathcal{F}_{R_m} - \text{measurable,} \tag{2.15}$$

and

$$\{U_m\} \text{ are iid uniform on } [0,1] \text{ and are independent of } \mathcal{F}_\infty. \tag{2.16}$$

2.4 The computation process ζ

Given an initial time $t_0 \in [0,T)$, the coalescing branching random walk $\{X_s, s \in [0, T-t_0]\}$, the sequence of parent indices $\{\mu_m\}$, the sequence of uniforms $\{U_m\}$, and a set of initial values in $\{0,1\}$, $\zeta_{t_0}(j) = \xi_{t_0}(X^j_{T-t_0})$, $j \in J(T-t_0)$, we will define $\{\zeta_r(k), r \in [0,T], 0 \le k \le K((T-r)-)\}$ so that

$$\zeta_r(k) = \xi_r(X^k_{T-r}) \text{ for all } r \in [t_0, T] \text{ and } k \le K((T-r)-)\,. \tag{2.17}$$

The left hand limits here reflect the fact that we have reversed the direction of time from that of X.

In general we consider a general initial state $\zeta_{t_0}(j) \in \{0,1\}$, $j \in J(T-t_0)$. First we complete this initial state by setting $\zeta_{t_0}(k) = \zeta_{t_0}(j)$ if $k \sim_{T-t_0} j \in J(T-t_0)$. Suppose that for some $m \ge 1$, R_m is the largest reaction time smaller than $T-t_0$. The values $\zeta_r(k)$ do not change except at times $T-R_n$, so $\zeta_r = \zeta_{t_0}$ for $r < T-R_m$. We decide whether or not to flip the value of ζ at μ_m at time $t-R_m$ as follows. Define $V_m \in \{0,1\}^{N_0}$ by

$$V^j_m = \zeta_{(T-R_m)-}(M + (m-1)N_0 + j)\,, \quad j = 1, \dots, N_0. \tag{2.18}$$

Letting $i = \zeta_{(T-R_m)-}(\mu_m)$ we set

$$\zeta_{(T-R_m)}(\mu_m) = \begin{cases} 1-i & \text{if } U_m \le g_{1-i}(V_m)/c^* \\ i & \text{otherwise.} \end{cases} \tag{2.19}$$

To update the dual now, for $k \le M + (m-1)N_0 = K((T-(T-R_m))-)$ and $k \ne \mu_m$,

$$\text{if } k \sim_{R_m} \mu_m \text{ set } \zeta_{T-R_m}(k) = \zeta_{T-R_m}(\mu_m)\,. \tag{2.20}$$

Otherwise we keep $\zeta_{(T-R_m)}(k) = \zeta_{(T-R_m)-}(k)$.

The values $\zeta_r(k)$ remain constant for $r \in [T-R_m, T-R_{m-1})$. Coming to $r = T-R_{m-1}$, if $m-1 \ge 1$ we proceed as above. When we reach $r = T-R_0 = T$ we end by setting $\zeta_T = \zeta_{T-}$. If $\xi_{t_0}(j) = \xi(X^j_{T-t_0})$ for $j \in J(T-t_0)$, the verification of (2.17) is an easy exercise from the definitions of X and ζ.

2.5 Branching random walk approximation $\hat{X}$

Due to the transience of random walk in dimensions $d \ge 3$, and the fact that the random walk steps are occurring at a fast rate in X when ε is small, any coalescing in X will occur soon after a branching event and close to the branching site. As in Durrett and Zähle [24], such births followed quickly by coalescing are not compatible with weak convergence of the dual. Thus we need a way to excise these events from X. As in [24] we define a *(non-coalescing) branching random walk* $\hat{X}$ and associated computation process $\hat{\zeta}$. Later we will couple (X, ζ) and $(\hat{X}, \hat{\zeta})$ so that they are close when ε is small.

For $m \in \mathbb{N}$, Π_m denotes the set of partitions of $\{0, \dots, m\}$ and for each $\pi \in \Pi_m$, $J_0(\pi)$ is the subset of $\{0, \dots, m\}$ obtained by selecting the minimal element of each cell of π. Write $i \sim_\pi j$ if i and j are in the same cell of π. Let $\{\hat{B}^{Y^i}, i = 0, \dots N_0\}$ be the rate one coalescing random walk system on $\mathbb{Z}^d$ with step distribution p and initial points at $Y^0 = 0, Y^1, \dots, Y^{N_0}$ where $(Y^1, \dots, Y^{N_0})$ has law q. Let ν_0 denote the law on Π_{N_0} of the random partition associated with the equivalence relation $i \sim j$ iff $\hat{B}^{Y^i}(t) = \hat{B}^{Y^j}(t)$ for some $t \geq 0$. For $\varepsilon > 0$ let ν_ε denote the law on Π_{N_0} of the random partition associated with the equivalence relation $i \sim^\varepsilon j$ iff $\hat{B}^{Y^i}(\varepsilon^{-3/2}) = \hat{B}^{Y^j}(\varepsilon^{-3/2})$. Note that $\varepsilon^{-3/2} = \varepsilon^{1/2}\varepsilon^{-2}$ so this is a short amount of time for the sped up process. For later use when we define the branching Brownian motion Z we note that since $\varepsilon^{-3/2} \to \infty$,

$$\nu_\varepsilon \text{ converges weakly to } \nu_0 \text{ as } \varepsilon \downarrow 0\,. \tag{2.21}$$

As before we will have a fixed $T > 0$ and distinct sites $z_0, \dots, z_M$ in $\varepsilon\mathbb{Z}^d$. Our branching random walk $\hat{X}$ will have paths in $\mathcal{D}$ and an associated set of indices $\hat{J}(t) = \{j : \hat{X}^j_t \neq \infty\}$. Let $\pi_0 \in \Pi_M$ be defined by the equivalence relation $i \sim j$ iff $\hat{B}^{\varepsilon^{-1}z_i}(\varepsilon^{-3/2}) = \hat{B}^{\varepsilon^{-1}z_j}(\varepsilon^{-3/2})$. In words, π_0 will be used to "mimic" the initial coalescence in X of the particles starting at z_i before any reaction events occur. We suppress the dependence of π_0 on ε in our notation.

For $n \geq 1$ let $\pi_n \in \Pi_{N_0}$ be iid with law ν_ε and independent of π_0. From $\{\pi_n\}$ we inductively define a sequence of nonempty subsets $\{\hat{J}_n\}$ of $\mathbb{Z}_+$ by $\hat{J}_0 = J_0(\pi_0)$ and for $n \geq 0$

$$\hat{J}_{n+1} = \hat{J}_n \cup \{M + nN_0 + j : j \in J_0(\pi_{n+1}) \setminus \{0\}\}. \tag{2.22}$$

Set $\hat{R}_0 = 0$ and conditional on $\{\pi_n\}$ let $\{\hat{R}_{n+1} - \hat{R}_n : n \geq 0\}$ be independent exponential random variables with means $(c^*|\hat{J}_n|)^{-1}$, and let $\{\hat{\mu}_n\}$ be an independent sequence of independent random variables where $\hat{\mu}_n$, $n \geq 1$, is uniformly distributed over $\hat{J}_{n-1}$. $\hat{\mu}_n$ is the index of the particle that gives birth at time $\hat{R}_n$.

To define $\hat{X}$ inductively we start with

$$\hat{X}^j_0 = z_j \text{ if } j \in \hat{J}_0 = \hat{J}(0) \text{ and } \infty \text{ otherwise.} \tag{2.23}$$

On $[\hat{R}_n, \hat{R}_{n+1})$, the $\hat{X}^j : j \in \hat{J}_n$ follow independent copies of B^ε starting at $\hat{X}^j_{\hat{R}_n}$. At $\hat{R}_{n+1}$ we define

$$\hat{X}^j_{\hat{R}_{n+1}} = \begin{cases} \hat{X}^j_{\hat{R}_{n+1}-} & \text{if } j \in \hat{J}_n = \hat{J}(\hat{R}_n)\,, \\ \hat{X}^{\hat{\mu}_{n+1}}_{\hat{R}_{n+1}-} & \text{if } j \in \hat{J}_{n+1} - \hat{J}_n\,, \\ \infty & \text{otherwise.} \end{cases}$$

Note that offspring are no longer displaced from their parents and that coalescence reduces the number of particles born at time $\hat{R}_{n+1}$, but otherwise no coalescence occurs as $\hat{J}(t) = \hat{J}_n$ on $[\hat{R}_n, \hat{R}_{n+1})$. Thus, conditional on the sequence $\{\pi_n\}$, $\hat{X}$ is a branching random walk starting with particles at z_j, $j \in J_0(\pi_0)$, with particle branching rate c^* and giving birth to $|\pi_n| - 1$ particles on top of the parent $\hat{X}^{\hat{\mu}_n}_{\hat{R}_n}$ (who also survives) at the nth branch time $\hat{R}_n$.

2.6 Computation process $\hat{\zeta}$

As we did for X, for $t_0 \in [0,T)$ we now define an computation process $\{\hat{\zeta}_r(k) : 0 \le k \le \hat{K}((T-r)-), r \in [t_0,T]\}$ for $\hat{X}$. Here $\hat{K}(s) = M + mN_0$ if $s \in [\hat{R}_m, \hat{R}_{m+1})$. Given are the branching random walks $\{\hat{X}_s, s \in [0, T-t_0]\}$, the associated sequence $\{\pi_n, \hat{R}_n, \hat{\mu}_n\}$, a sequence of iid random variables $\{\hat{U}_n\}$, uniformly distributed on $[0,1]$ and independent of $(\hat{X}, \{\pi_n, \hat{R}_n, \hat{\mu}_n : n \in \mathbb{N}\})$, and a set of initial values $\hat{\zeta}_{t_0}(j), j \in \hat{J}(T-t_0)$. In the next section when we couple (X,ζ) and $(\hat{X},\hat{\zeta})$ we will set $\hat{U}_n$ equal to U_n defined in (2.11). Define an equivalence relation $\approx_{\hat{R}_n}$ on $\{0,\dots,M+nN_0\}$ by

$$
\begin{aligned}
M+(m-1)N_0+j &\approx_{\hat{R}_n} M+(m-1)N_0+i \; (1 \le i,j \le N_0,\, 1 \le m \le n) \text{ iff } j \sim_{\pi_m} i,\\
M+(m-1)N_0+j &\approx_{\hat{R}_n} \hat{\mu}_m \qquad (1 \le j \le N_0,\, 1 \le m \le n) \quad \text{iff } j \sim_{\pi_m} 0,\\
j &\approx_{\hat{R}_n} i, \qquad (0 \le i,j \le M) \quad \text{iff } j \sim_{\pi_0} i.
\end{aligned} \tag{2.24}
$$

Finally if $\hat{R}_n \le t < \hat{R}_{n+1}$ define $i \approx_t j$ iff $i \approx_{\hat{R}_n} j$ for $0 \le i,j \le M+nN_0$. To prepare for the proof of Lemma 2.10, note that the definition of $\hat{\zeta}$ that follows is just the definition of ζ with hats added and $\approx_t$ used in place of $\sim_t$.

First we complete the initial state $\hat{\zeta}_{t_0}$ by setting $\hat{\zeta}_{t_0}(k) = \hat{\zeta}_{t_0}(j)$ if $k \approx_{T-t_0} j \in \hat{J}(T-t_0)$, $k \le K(T-t_0) = K((T-t_0)-)$ a.s. Suppose that for some $m \ge 1$, $\hat{R}_m$ is the largest branch time smaller than $T-t_0$. The values $\hat{\zeta}_r(k)$ do not change except at times $T-\hat{R}_n$, so $\hat{\zeta}_r = \hat{\zeta}_{t_0}$ for $t < T-\hat{R}_m$. We decide whether or not to flip the value of $\hat{\zeta}$ at $\hat{\mu}_m$ at time $t-\hat{R}_m$ as follows. Define $\hat{V}_m \in \{0,1\}^{N_0}$ by

$$\hat{V}_m^j = \hat{\zeta}_{(T-\hat{R}_m)-}(M+(m-1)N_0+j)\,, \quad j = 1,\dots,N_0. \tag{2.25}$$

Letting $i = \hat{\zeta}_{(T-\hat{R}_m)-}(\hat{\mu}_m)$ we set

$$\hat{\zeta}_{(T-\hat{R}_m)}(\hat{\mu}_m) = \begin{cases} 1-i & \text{if } \hat{U}_m \le g^{\varepsilon}_{1-i}(\hat{V}_m)/c^* \\ i & \text{otherwise.} \end{cases} \tag{2.26}$$

To update $\hat{\zeta}$ now, for $k \le M+(m-1)N_0$ and $k \ne \hat{\mu}_m$,

$$\text{if } k \approx_{\hat{R}_m} \hat{\mu}_m \text{ set } \hat{\zeta}_{T-\hat{R}_m}(k) = \hat{\zeta}_{(T-\hat{R}_m)}(\hat{\mu}_m)\,, \tag{2.27}$$

and for the remaining values of $k \le M+(m-1)N_0$ keep

$$\hat{\zeta}_{T-\hat{R}_m}(k) = \hat{\zeta}_{(T-\hat{R}_m)-}(k).$$

The values $\hat{\zeta}_r(k)$ remain constant for $r \in [T-\hat{R}_m, T-\hat{R}_{m-1})$. Coming to $r = T-\hat{R}_{m-1}$, if $m-1 \ge 1$ we proceed as above. When we reach $r = T-\hat{R}_0 = T$ we end by setting $\hat{\zeta}_T = \hat{\zeta}_{T-}$.

2.7 Coupling of (X, ζ) and $(\hat{X}, \hat{\zeta})$

We now give a construction of $\hat{X}, \hat{\zeta}$ which will have the property that with high probability for small ε, (i) X and $\hat{X}$ are close and (ii) given identical inputs, ζ and $\hat{\zeta}$ will compute the same result. As before, $T > 0$ and $z = (z_0, \dots, z_M)$, $z_i \in \varepsilon\mathbb{Z}^d$ are fixed. Recall the reaction times R_m, the uniform random variables U_m from (2.11), and the natural time-reversed filtration $\mathcal{F}_t$ used in the construction of the dual X given in (2.10).

The following general definition will be used to construct the partitions $\{\pi_n : n \in \mathbb{Z}_+\}$ needed to define $\hat{X}$, distributed as in Section 2.5, in terms of the graphical representation. Let V be an $\mathcal{F}_t$-stopping time (think of $V = R_m$), and let $\gamma_0 \dots, \gamma_{M'} \in \varepsilon\mathbb{Z}^d \in \varepsilon\mathbb{Z}^d$ be $\mathcal{F}_V$-measurable. Let $\{\hat{B}^{\varepsilon,\gamma_i} : i = 0, \dots, M'\} \subset \varepsilon\mathbb{Z}^d$ be the rescaled coalescing random walk system determined by the $\{T_n^x\}$ in the graphical representation, and starting at (dual) time V at locations $\gamma_0, \dots, \gamma_{M'}$. That is, $\{\hat{B}^{\varepsilon,\gamma_i} : i = 0, \dots, M'\}$ are as described in Figure 6 but now starting at time $T - V$ at sites $(\gamma_0, \dots, \gamma_{M'}) = (\hat{B}_0^{\varepsilon,\gamma_0}, \dots, \hat{B}_0^{\varepsilon,\gamma_{M'}})$. For each $t > 0$ let $\pi_{V,\gamma}(t) \in \Pi_{M'}$ be the random partition of $\{0, \dots, M'\}$ associated with the equivalence relation $i \sim i'$ iff $\hat{B}^{\varepsilon,\gamma_i}(t) = \hat{B}^{\varepsilon,\gamma_{i'}}(t)$. We call $\pi_{V,\gamma}(t)$ the random partition at time $V + t$ with initial condition $\gamma = (\gamma_0, \dots, \gamma_{M'})$ at time V.

Let $\pi_0 = \pi_{0,z}(\sqrt{\varepsilon}) \in \Pi_M$ be the random partition of $\{0, \dots, M\}$ at time $\sqrt{\varepsilon}$ with initial condition $z = (z_0, \dots, z_M)$ at time 0, and note that its law is the same as the law, ν_ε, of the π_0 described just before (2.21). For $m \geq 1$ let

$$\gamma_m = (X_{R_m}^{\mu_m}, X_{R_m}^{\mu_m} + Y_m^1, \dots, X_{R_m}^{\mu_m} + Y_m^{N_0})$$

and $\{\pi'_m, m \in \mathbb{N}\}$ be an iid sequence with law ν_ε and chosen independent of $\mathcal{F}_\infty$. For $m \in \mathbb{N}$, define

$$\pi_m = \begin{cases} \pi_{R_m,\gamma_m}(\sqrt{\varepsilon}) & \text{if } R_n > R_{n-1} + \sqrt{\varepsilon} \text{ for all } 1 \leq n \leq m \\ \pi'_m & \text{otherwise.} \end{cases} \tag{2.28}$$

By the translation invariance and independent increments properties of the Poisson point processes used in the graphical representation and also (2.15), π_m is independent of $\mathcal{F}_{R_{m-1}+\sqrt{\varepsilon}} \vee \sigma(\pi'_n, n < m) \equiv \bar{\mathcal{F}}_{m-1}$, and has law ν_ε defined just before (2.21). It is also easy to check that π_m is $\bar{\mathcal{F}}_m$-measurable ($m \geq 0$) and so $\{\pi_m, m \geq 0\}$ are independent and distributed as in Section 2.5.

For $m \in \mathbb{N}$ let

$$\begin{aligned} \tau'_m &= \inf\{s \geq R_{m-1} : \exists i \neq j \text{ both in } J(R_{m-1}-), \text{ or} \\ &\quad i \in J(R_{m-1}-) \setminus \{\mu_{m-1}\}, j \in J(R_{m-1}) \setminus J(R_{m-1}-), \text{ so that } X_s^i = X_s^j\}, \\ \tau_m &= \inf\{s \geq R_{m-1} + \sqrt{\varepsilon} : \inf_{i \neq j \in J(s)} |X_s^i - X_s^j| \leq \varepsilon^{7/8}\}, \\ Y_m^* &= \max\{|Y_m^i| : i = 1, \dots, N_0\}. \end{aligned}$$

We introduce the time, T_b, that one of four possible "bad events" occurs:

$$\begin{aligned} T_b = &\min\{R_m : m \geq 1, R_m \leq R_{m-1} + \sqrt{\varepsilon} \text{ or } Y_m^* \geq \frac{\varepsilon}{\kappa}\log(1/\varepsilon)\} \\ &\wedge \min\{\tau_m : m \geq 1, \tau_m < R_m\} \wedge \min\{\tau'_m : m \geq 2, \tau'_m \leq R_{m-1} + \sqrt{\varepsilon}\} \end{aligned}$$

(recall κ from (1.1)). Here $\min\emptyset = \infty$. To see why the last two minima should be large, note that after a birth of N_0 particles from particle μ_m at time R_m, we expect some coalescence to occur between the parent and its children. After time $\sqrt{\varepsilon}$, particles should all be separated by at least $\varepsilon^{7/8}$ and remain that way until the next reaction time when again there may be coalescing within the family producing offspring but no other coalescing events. The qualifier $m \geq 2$ is needed in the last minimum because we have no control over the spacings between particles at time 0. The collision of particles 2 and 7 in Figure 6, supposing it occurs before time $R_3 + \sqrt{\varepsilon}$, is an example of a bad event that enters into the definition of τ'_4. We assume throughout that

$$0 < \varepsilon < \varepsilon_1(\kappa) \text{ so that } \frac{\varepsilon}{\kappa}\log(1/\varepsilon) < \varepsilon^{7/8}/2. \tag{2.29}$$

Given $\{\pi_m\}$ we now construct $\hat{X}$ and $\hat{A}(s) = ((\hat{\mu}_n, \hat{R}_n)1(\hat{R}_n \leq s))_{n\in\mathbb{N}}$ (with the joint law described in Section 2.5) initially up to time $\hat{T} = T_b \wedge \hat{T}_b$, where

$$\hat{T}_b = \min\{\hat{R}_m : m \geq 1, \hat{R}_m - \hat{R}_{m-1} \leq \sqrt{\varepsilon}\}.$$

Once one of the five bad events (implicit in the definition of $\hat{T}$) occurs, we will give up and continue the definition of the branching random walk using independent information. The coupling of X and $\hat{X}$ will be through our definition of $\{\pi_n\}$ and also through the use of the random walks steps of X^j to define corresponding random walk steps in $\hat{X}^j$ whenever possible, as will be described below.

We begin our inductive construction by setting $\hat{R}_0 = 0$, $\hat{J}(0) = J_0(\pi_0)$, and define $\hat{X}_0$ as in (2.23). Note that

$$\hat{J}(0) = J(\sqrt{\varepsilon}) = J_0(\pi_0) \text{ if } R_1 > \sqrt{\varepsilon}. \tag{2.30}$$

Assume now that $(\hat{X}, \hat{A})$ has been defined on $[0, R_m \wedge \hat{T}]$. Assume also that $R_m < \hat{T}$ implies the following for all $1 \leq i \leq m$:

$$\hat{R}_i = R_i, \quad \hat{\mu}_i = \mu_i, \tag{2.31}$$

$$\hat{J}(R_i) = \hat{J}(R_{i-1}) \cup \{M + (i-1)N_0 + j : j \in J_0(\pi_i) \setminus \{0\}\}. \tag{2.32}$$

$$\hat{J}(R_{i-1}) = \hat{J}(s) \subset J(s) \text{ for all } s \in [R_{i-1}, R_i). \tag{2.33}$$

$$\hat{J}(s) = J(s) = J(R_{i-1} + \sqrt{\varepsilon}) \text{ for all } s \in [R_{i-1} + \sqrt{\varepsilon}, R_i), \tag{2.34}$$

The $m = 0$ case of the induction is slightly different, due for example to the special nature of π_0, so let us assume $m \geq 1$ first. To define $(\hat{X}, \hat{A})$ on $(R_m \wedge \hat{T}, R_{m+1} \wedge \hat{T}]$ we may assume $R_m(\omega) < \hat{T}(\omega)$ and so (2.31)-(2.34) hold by induction. On $(R_m, (R_m + \sqrt{\varepsilon}) \wedge R_{m+1} \wedge \hat{T}]$ let $(\hat{X}, \hat{A})$ evolve as in Section 2.5 conditionally independent of $\mathcal{F}_\infty \vee \sigma(\{U_m\})$ given $\{\pi_n\}$. Here it is understood that the unused partitions $\{\pi_i : i > m\}$ are used to define the successive branching events as in (2.22).

Next, to define $(\hat{X}, \hat{A})$ on $((R_m + \sqrt{\varepsilon}) \wedge R_{m+1} \wedge \hat{T}, R_{m+1} \wedge \hat{T}]$ we may assume $R_m(\omega) + \sqrt{\varepsilon} < R_{m+1} \wedge \hat{T}(\omega)$. By the definition of $\hat{T}_b$ this implies $\hat{R}_{m+1} > R_m + \sqrt{\varepsilon}$ and so for all $s \in [R_m, R_m + \sqrt{\varepsilon}]$,

$$\begin{aligned}\hat{J}(s) = \hat{J}(R_m) &= \hat{J}(R_{m-1}) \cup \{M + (m-1)N_0 + j : j \in J_0(\pi_m) \setminus \{0\}\} \\ &= J(R_{m-1} + \sqrt{\varepsilon}) \cup \{M + (m-1)N_0 + j : j \in J_0(\pi_m) \setminus \{0\}\}.\end{aligned} \tag{2.35}$$

In the first equality we used (2.32) and in the second we used (2.33) and (2.34) with $s = R_{m-1} + \sqrt{\varepsilon}$. The fact that $\tau_m \le R_m$ (since $T_b > R_m + \sqrt{\varepsilon}$) shows there are no coalescings of X on $[R_{m-1} + \sqrt{\varepsilon}, R_m)$ and so

$$J(R_{m-1} + \sqrt{\varepsilon}) = J(R_m-). \tag{2.36}$$

Again use $T_b > R_m + \sqrt{\varepsilon}$ together with (2.29) to see that $Y^*_m \le \frac{\varepsilon}{\kappa} \log(1/\varepsilon) \le \frac{\varepsilon^{7/8}}{2}$, and so the spacings of the previously existing particles at time $R_m \le \tau_m$ ensures that none of the N_0 new particles at time R_m will land on a previously occupied site. Therefore if

$$J_1(Y_m) = \{1 \le j \le N_0 : Y^j_m \notin \{Y^i_m : 0 \le i < j\}\},$$

then

$$J(R_m) = J(R_m-) \cup \{M + (m-1)N_0 + j : j \in J_1(Y_m)\}.$$

The fact that $R_{m+1} \wedge \tau'_{m+1} > R_m + \sqrt{\varepsilon}$ means that X has no branching events in $(R_m, R_m + \sqrt{\varepsilon}]$ and X has no particles coalescing on $[R_m, R_m + \sqrt{\varepsilon}]$ except those involving $X^{\mu_m}_{R_m} + Y^i_m, i = 0, \ldots, N_0$. Therefore, the definition of π_m ensures that

$$\begin{aligned}J(R_m + \sqrt{\varepsilon}) &= J(R_m-) \cup \{M + (m-1)N_0 + j : j \in J_0(\pi_m) \setminus \{0\}\}. \\ &= \hat{J}(s) \text{ for all } s \in [R_m, R_m + \sqrt{\varepsilon}],\end{aligned} \tag{2.37}$$

where in the last line we have used (2.35) and (2.36). For $s \in [R_m + \sqrt{\varepsilon}, R_{m+1} \wedge \hat{T})$ we have $s < \tau_{m+1}$ and so

$$|X^j_s - X^k_s| > \varepsilon^{7/8} \text{ for all } j \ne k \text{ both in } J(s), \text{ for all } s \in [R_m + \sqrt{\varepsilon}, R_{m+1} \wedge \hat{T}). \tag{2.38}$$

In particular X can have no coalescings on the above interval and so $J(s) = J(R_m + \sqrt{\varepsilon})$ for $s \in [R_m + \sqrt{\varepsilon}, R_{m+1} \wedge \hat{T})$. On $(R_m + \sqrt{\varepsilon}, R_{m+1} \wedge \hat{T}]$ let $(\hat{X}^j_s, j \in \hat{J}(s))$ follow the random walk steps and branching events of $\{X^j : j \in J(s)\}$ (of course there is at most one of the latter at time R_{m+1} providing $R_{m+1} \le \hat{T}$). In particular we are setting

$$\hat{J}(s) = J(s) = J(R_m + \sqrt{\varepsilon}) \text{ for } s \in [R_m + \sqrt{\varepsilon}, R_{m+1} \wedge \hat{T}) \text{ or } s = \hat{T} < R_{m+1}. \tag{2.39}$$

(2.38) shows that the random walk steps and branching events for distinct particles of X on $(R_m + \sqrt{\varepsilon}, R_{m+1} \wedge \hat{T}]$ are independent. In addition, these steps and branching events are independent of the random walk increments used to define $\{\pi_n\}$. This shows that $\hat{X}$ evolves like the branching random walk described in Section 2.5 on

$(R_m, R_{m+1}\wedge\hat T)$, and on $(R_m, R_{m+1}\wedge\hat T]$ if either $\hat T < R_{m+1}$, or $R_m+\sqrt{\varepsilon} \ge R_{m+1}\wedge\hat T$. (In the latter case the first part of the above construction did the job and in the former case there is no reaction event to define at $\hat T\wedge R_{m+1} = \hat T$.) So to complete the construction at $t = R_{m+1}\wedge\hat T$ we may assume

$$R_m + \sqrt{\varepsilon} < R_{m+1} \le \hat T. \tag{2.40}$$

The above definition shows that $\hat R_{m+1} = R_{m+1}$, we use (2.32) with $i = m+1$ to define $\hat J(R_{m+1})$ and we set $\hat\mu_{m+1} = \mu_{m+1}$. Clearly $\hat\mu_{m+1}$ is uniform on $\hat J(R_m) = J(R_m+\sqrt{\varepsilon})$ (given $\{\pi_n\}$) and is independent of $\{\hat\mu_n : n < m\}$. In addition the branching events used to define $\{\hat\mu_n\}$ are independent of the random walk steps used to define $\{\pi_n\}$. This completes our inductive definition of $(\hat X, \hat A)$ on $[0, R_{m+1}\wedge\hat T]$.

Next we complete the inductive step of the derivation of (2.31)-(2.34) for $m+1$ under (2.40) which is in fact weaker than the $R_{m+1} < \hat T$ condition. (2.39) implies (2.34) for $i = m+1$, and (2.31) and (2.32) hold by definition. On $\{R_m+\sqrt{\varepsilon} < R_{m+1} \le \hat T\}$ J can only decrease on $[R_m, R_m+\sqrt{\varepsilon}]$ due to coalescings of the random walks, while $\hat J$ is constant on this interval by (2.37). The inclusion (2.33) therefore follows from the equality in (2.34).

To complete the inductive construction of $(\hat X, \hat A)$ on each $[0, R_m\wedge\hat T]$ and proof of (2.31)-(2.34) it remains to give the $m = 0$ step of the construction and verify the $m = 1$ case of the induction. Both follow by making only minor changes in the above induction step. For example, (2.30) is used in place of the (now non-existent) induction hypothesis (2.32) both in defining $\hat X$ on the initial interval and in obtaining (2.37) for $m = 0$.

Since $R_m \uparrow \infty$ a.s. we have defined $(\hat X, \hat A)(s)$ on $[0, \hat T]$ and to complete the definition we let it evolve conditionally independently (given $\{\pi_n\}$) for $s \ge \hat T$.

The above construction and (2.16) show that

$$(X, \{\pi_n\}, \{\mu_n\}, \hat X, \{\hat\mu_n\}, \{\hat R_n\}) \text{ is independent of } \{U_n\}, \tag{2.41}$$

where $\{U_n\}$ are the uniforms from (2.11). Therefore the computation process $\hat\zeta$ for the above $\hat X$ may be defined as in Section 2.6 but with $\hat U_n = U_n$.

Lemma 2.1. *(a) For all $m \in \mathbb{Z}_+$, $R_m < T_b$ and $\hat R_m < \hat T_b$ imply $R_m = \hat R_m < \hat T$.*
(b) For all $m \in \mathbb{N}$, if

$$\begin{aligned} G_m = \{\omega : &\Big(\wedge_{i=1}^m (R_i - R_{i-1})\Big) \wedge \Big(\wedge_{i=2}^{m+1}(\tau_i' - R_{i-1})\Big) > \sqrt{\varepsilon},\\ &R_i \le \tau_i \ \forall i \le m, \max_{i\le m} Y_i^* < \frac{\varepsilon}{\kappa}\log(1/\varepsilon), \wedge_{i=1}^m(\hat R_i - \hat R_{i-1}) > \sqrt{\varepsilon}\},\end{aligned}$$

then $G_m \subset \{\hat R_m = R_m < \hat T\}$.

Proof. (a) The implication is trivial for $m = 0$ so assume it for m and assume also $R_{m+1} < T_b$, $\hat R_{m+1} < \hat T_b$. By induction we have $R_m = \hat R_m < \hat T$. Since $R_{m+1}\wedge\hat R_{m+1} > R_m + \sqrt{\varepsilon}$, we also know $\hat T > R_m + \sqrt{\varepsilon}$. The construction of $\hat X$ on $(R_m+\sqrt{\varepsilon}, R_{m+1}\wedge\hat T]$ shows that the next reaction time of $\hat X$ on this interval must be R_{m+1} (if it exists) and so $\hat T_b \ge R_{m+1}$. Since $T_b > R_{m+1}$ by hypothesis we get

$(R_m + \sqrt{\varepsilon}, R_{m+1} \wedge \hat{T}] = (R_m + \sqrt{\varepsilon}, R_{m+1}]$. Hence our construction of $\hat{X}$ on this interval shows $\hat{R}_{m+1} = R_{m+1}$ and so the result follows for $m+1$.
(b) The first four conditions in the definition of G_m imply

$$T_b \geq R_{m+1} \wedge \tau'_{m+2} \wedge \tau_{m+1} > R_m.$$

The last condition implies $\hat{T}_b > \hat{R}_m$. Now apply (a).

□

As an immediate consequence of the above and our inductive proof of (2.31)-(2.34) we get the following:

Lemma 2.2.

$$G_m \Rightarrow R_m < \hat{T} \Rightarrow \text{ for all } 1 \leq i \leq m \text{ (2.31)} - \text{(2.34) hold.}$$

On G_m and on the intervals $[R_{m-1}+\sqrt{\varepsilon}, R_m)$ our definition of $\hat{X}$ and Lemma 2.1(b) shows that the movement of particles in X and $\hat{X}$ are coupled (they take identical steps) but on $[R_{m-1}, R_{m-1}+\sqrt{\varepsilon})$ they move independently. To bound the discrepancies that accumulate during these intervals we use:

Lemma 2.3. *If $\omega \in G_m$, then*

$$\sup\{|\hat{X}^j_s - X^j_s| : j \in \hat{J}(s),\ s \in [0, R_m)\} \tag{2.42}$$
$$\leq (m-1)\frac{\varepsilon}{\kappa}\log(1/\varepsilon) + \sum_{l=0}^{m-1} \sup_{j\in\hat{J}(R_l), s\in[R_l, R_l+\sqrt{\varepsilon}]} |\hat{X}^j_s - \hat{X}^j_{\hat{R}_l}| + |X^j_s - X^j_{R_l}|.$$

Proof. Suppose first that $m > 1$ and we are on G_m. By the coupling of the spatial motions noted above, for $j \in \hat{J}(R_{m-1})$

$$\begin{aligned}
\sup_{s\in[R_{m-1}, R_m)} |\hat{X}^j_s - X^j_s| &= \sup_{s\in[R_{m-1}, R_{m-1}+\sqrt{\varepsilon}]} |\hat{X}^j_s - X^j_s| \\
&\leq |\hat{X}^j_{R_{m-1}} - X^j_{R_{m-1}}| + \sup_{s\in[R_{m-1}, R_{m-1}+\sqrt{\varepsilon}]} |\hat{X}^j_s - \hat{X}^j_{R_{m-1}}| \\
&\qquad + \sup_{s\in[R_{m-1}, R_{m-1}+\sqrt{\varepsilon}]} |X^j_s - X^j_{R_{m-1}}|.
\end{aligned}$$

On G_m, a newly born particle to $X^j_{R_{m-1}-}$ may jump a distance at most $\frac{\varepsilon}{\kappa}\log(1/\varepsilon)$ from its parent, while for $\hat{X}^j_{R_{m-1}-}$ it will be born on its parent site, so the above is at most

$$\begin{aligned}
&\sup_{k\in J(R_{m-1}-)} |\hat{X}^k_{R_{m-1}-} - X^k_{R_{m-1}-}| + \frac{\varepsilon}{\kappa}\log(1/\varepsilon) \\
&\quad + \sup_{s\in[R_{m-1}, R_{m-1}+\sqrt{\varepsilon}]} |\hat{X}^j_s - \hat{X}^j_{R_{m-1}}| + \sup_{s\in[R_{m-1}, R_{m-1}+\sqrt{\varepsilon}]} |X^j_s - X^j_{R_{m-1}}|.
\end{aligned}$$

Things are simpler when $m = 1$ because there are no initial jumps to worry about and so the second term in the above is absent. The required bound now follows by induction in m and the fact that G_m is decreasing in m. □

2.8 Bounding the probability of bad events

Here and in what follows it is useful to dominate X with a branching random walk $\bar{X}$, also with paths in $\mathcal{D}$ and with the same initial state. Particles in $\bar{X}$ follow independent copies of B^ε and with rate c^* give birth to N_0 particles located at $B^\varepsilon_t + Y^i_m$, $i = 1, \dots, N_0$, where B^ε_t is the location of the parent particle. At the mth birth time $\bar{R}_m$ we use $X^{M+(m-1)N_0+i}$, $i = 1, \dots, N_0$ to label the new particles, so that if $\bar{J}(t) = \{j : X^j_t \neq \infty\}$, then $\bar{J}(\bar{R}_m) = \{0, \dots, M + mN_0\}$. Coalescence is avoided in $\bar{X}$ by having the coalescing particle with the larger index have its future steps and branching events dictated by an independent copy of the graphical representation. This will ensure that $J(t) \subset \bar{J}(t)$ and $\{X^j(t) : j \in J(t)\} \subset \{\bar{X}^j(t) : j \in \bar{J}(t)\}$ for all $t \geq 0$.

Let $N_T = \min\{m : R_m > T\}$ and define $\bar{N}_T$ in the same way, using the branching times $\{\bar{R}_m\}$. Let

$$c_b = c^* N_0 \geq 1\,. \tag{2.43}$$

We will also need to separate the particles in $\hat{X}$ and so define

$$\hat{\tau}_m = \inf\{s \geq \hat{R}_{m-1} + \sqrt{\varepsilon} : \inf_{i \neq j \in \hat{J}(s)} |\hat{X}^i_s - \hat{X}^j_s| \leq \varepsilon^{7/8}\}, m \in \mathbb{N}. \tag{2.44}$$

Lemma 2.4. *There is a constant $c_{2.4}$ so that for all $T > 0$ and $n \in \mathbb{N}$*
(a) $P(N_T > n) \leq P(\bar{N}_T > n) \leq e^{c_b T}(M+1)(nN_0)^{-1}$.
(b) $P(\min_{1 \leq m \leq N_T} R_m - R_{m-1} \leq \sqrt{\varepsilon}$ or $\min_{1 \leq m \leq N_T} \hat{R}_m - \hat{R}_{m-1} \leq \sqrt{\varepsilon})$
$\leq c_{2.4} c_b e^{c_b T}(M+1)\varepsilon^{1/6}$.

Proof. (a) The first inequality follows from the domination of X by $\bar{X}$. For the second one note that $E(\bar{J}(T)) = (M+1)e^{c_b T}$ and conclude

$$\begin{aligned} P(\bar{N}_T > n) &\leq P(|\bar{J}(T)| \geq M + 1 + nN_0) \\ &\leq (M + 1 + nN_0)^{-1}(M+1)e^{c_b T}. \end{aligned}$$

(b) Let Z be a mean $1/c^*$ exponential random variable. The domination of X by $\bar{X}$ shows that for any $n \geq 1$,

$$\begin{aligned} &P\left(\min_{1 \leq m \leq N_T} R_m - R_{m-1} \leq \sqrt{\varepsilon}\right) \leq P\left(\min_{1 \leq m \leq \bar{N}_T} \bar{R}_m - \bar{R}_{m-1} \leq \sqrt{\varepsilon}\right) \\ &\leq P(\bar{N}_T > n) + \sum_{m=1}^{n} P\Big(\frac{Z}{M + 1 + (m-1)N_0} \leq \sqrt{\varepsilon}\Big) \\ &\leq e^{c_b T}(M+1)(nN_0)^{-1} + \sum_{m=1}^{n} c^*(M + 1 + (m-1)N_0)\sqrt{\varepsilon}, \end{aligned}$$

by (a). Note that the sum is at most $c^*(M+1)n\sqrt{\varepsilon} + c_b n^2\sqrt{\varepsilon}$ and set $n = \lceil \varepsilon^{-1/6} \rceil$. A similar calculation gives the same upper bound for the $\hat{R}_m$'s. ◻

Lemma 2.5. *There is a constant $c_{2.5}$ so that for all $T > 0$*

$$P(Y^*_m > \frac{\varepsilon}{\kappa}\log(1/\varepsilon) \text{ for some } m \leq N_T) \leq c_{2.5} e^{c_b T}(M+1)\varepsilon^{1/2}$$

Proof. By (a) of Lemma 2.4, $P(N_T > n) \le e^{c_b T}(M+1)(nN_0)^{-1}$. Using (1.6) gives

$$P(Y_m^* > \frac{\varepsilon}{\kappa}\log(1/\varepsilon) \text{ for some } m \le n) \le nC\varepsilon.$$

Taking $n = \lceil \varepsilon^{-1/2} \rceil$ now gives the desired result. $\square$

The following facts about random walks will be used frequently.

Lemma 2.6. *Let Z_s denote a continuous time rate 2 random walk on $\mathbb{Z}^d$ jumping with kernel p, and starting at $x \in \mathbb{Z}^d$ under P^x, and B^ε be our continuous time rescaled copy of Z, starting at $z \in \varepsilon\mathbb{Z}^d$ under P_z.*
(a) For any $t_0 \ge 0$, $r_0 \ge 1$, $x \in \mathbb{Z}^d$ and $p \ge 2$,

$$P^x(|Z_s| \le r_0 \text{ for some } s \ge t_0) \le c_{2.6} \int_{t_0}^\infty \Big[[(|x| - r_0)^+]^{-p}(s^{p/2} \vee s)\Big] \wedge \Big[(s \vee 1)^{-d/2} r_0^d\Big]\, ds.$$

(b) $\sup_x P^x(|Z_s| \le \varepsilon^{-1/8}$ *for some* $s \ge \varepsilon^{-3/2}) \le c_{2.6}\varepsilon^{3/8}$.
(c) For any $z \in \varepsilon\mathbb{Z}^d$, $r_0 \ge 1$

$$P_z(|B_s^\varepsilon| \le r_0\varepsilon \text{ for some } s \ge 0) \le c_{2.6}(|z|\varepsilon^{-1})^{-(2/3)(d-2)} r_0^{2(d+1)/3}.$$

Proof. (a) Use $T(t_0, y) \le \infty$ to denote the time of the first visit of Z to y after time t_0, and let

$$G = \int_0^\infty P^0(Z_s = 0)\, ds$$

be the expected time at 0 (which is finite since $d \ge 3$). Then

$$\begin{aligned}\int_{t_0}^\infty P^x(Z_s = y)\, ds &= E^x\Big(1\{T(t_0,y)(\omega) < \infty\} \int_{T(t_0,y)(\omega)}^\infty P^y(Z_{s - T(t_0,y)(\omega)} = y)\, ds\Big) \\ &= GP^x(T(t_0,y) < \infty).\end{aligned}$$

Summing over $|y| \le r_0$ for $r_0 \ge 1$ and rearranging, we get

$$\begin{aligned}P^x(|Z_s| \le r_0 \text{ for some } s \ge t_0) &\le G^{-1} \sum_{|y| \le r_0} \int_{t_0}^\infty P^x(Z_s = y)\, ds \\ &= G^{-1} \int_{t_0}^\infty P^x(|Z_s| \le r_0) ds. \qquad (2.45)\end{aligned}$$

A martingale square function inequality shows that for $p \ge 2$,

$$P^x(|Z_s| \le r_0) \le P^0(|Z_s| \ge (|x| - r_0)^+) \le c((|x| - r_0)^+)^{-p})(s^{p/2} \vee s). \qquad (2.46)$$

A local central limit theorem (see, e.g. (A.7) in [6]) shows that

$$P^x(|Z_s| \le r_0) \le c(s \vee 1)^{-d/2} r_0^d. \qquad (2.47)$$

Use the above two inequalities to bound the integrand in (2.45) and derive (a).

(b) Set $r_0 = \varepsilon^{-1/8}$ and $t_0 = \varepsilon^{-3/2}$ in (a) and use only the second term in the infimum inside the integral. The right-hand side is $c\varepsilon^{-(d/8)-(3/2)+(3d/4)}$. To complete the proof we note that exponent is smallest when $d = 3$.

(c) We may assume without loss of generality that $r_0 \le |z|\varepsilon^{-1}/2 = M/2$ (or the bound is trivial) and so $t_1 = M^{4/3} r_0^{2/3} \ge 1$. Apply (a) with $p = 2d$ and break the integral at t_1 to see that the probability in (c) is

$$\begin{aligned} P^{z\varepsilon^{-1}}(|Z_s| \le r_0 \text{ for some } s \ge 0) &\le c\Big[\int_0^{t_1} M^{-2d}(s^d \vee s)ds + \int_{t_1}^{\infty} s^{-d/2} r_0^d ds\Big] \\ &\le c(M^{-2d} t_1^{d+1} + t_1^{1-(d/2)} r_0^d) \\ &\le cM^{-(2/3)(d-2)} r_0^{2(d+1)/3}. \end{aligned}$$

□

Lemma 2.7. *$P(\tau_m < R_m$ or $\hat{\tau}_m < \hat{R}_m$ for some $1 \le m \le N_T)$*
$\le c_{2.7} e^{c_b T}(M+1)^2 \varepsilon^{3/32}$.

Proof. To bound $P(\tau_m < R_m$ for some $1 \le m \le N_T)$, we start with

$$\begin{aligned} P(\tau_m < R_m | \mathcal{F}_{R_{m-1}}) \le P(&R_m > R_{m-1} + \sqrt{\varepsilon},\ \exists i \ne j \text{ both in } J(R_{m-1} + \sqrt{\varepsilon}), \text{ s.t.} \\ &\inf_{\sqrt{\varepsilon}+R_{m-1} \le s \le R_m} |X_s^i - X_s^j| \le \varepsilon^{7/8} | \mathcal{F}_{R_{m-1}}). \end{aligned}$$

Now $i \ne j$ both in $J(R_{m-1} + \sqrt{\varepsilon})$ and $R_m > R_{m-1} + \sqrt{\varepsilon}$ imply $i, j \in J(R_{m-1})$ and $X_s^i \ne X_s^j$ for all $s \in [R_{m-1}, R_{m-1} + \sqrt{\varepsilon}]$. Therefore, the above is at most

$$\begin{aligned} \sum_{i \ne j \in J(R_{m-1})} P(&X_s^i - X_s^j \ne 0,\ \forall s \in [R_{m-1}, R_{m-1} + \sqrt{\varepsilon}], \\ &|X_s^i - X_s^j| \le \varepsilon^{7/8}\ \exists s \ge R_{m-1} + \sqrt{\varepsilon} | \mathcal{F}_{R_{m-1}}). \end{aligned} \tag{2.48}$$

If Z is as in Lemma 2.6, we may use (b) of that result to bound the above by

$$\begin{aligned} &|J(R_{m-1})|^2 \sup_{z_0 \ne 0} P^{z_0}(|Z_s| \le \varepsilon^{-1/8}\ \exists s \ge \varepsilon^{-3/2}) \\ &\le (M + 1 + (m-1)N_0)^2 \cdot c\varepsilon^{3/8} \end{aligned}$$

Using Lemma 2.4(a), we conclude

$$\begin{aligned} &P(\tau_m < R_m \text{ for some } 1 \le m \le N_T) \\ &\le e^{c_b T}(M+1)(nN_0)^{-1} + \sum_{m=1}^{n} (M + 1 + (m-1)N_0)^2 c\varepsilon^{3/8}. \end{aligned}$$

To bound the sum we note that for $a, b \ge 1$,

$$\sum_{m=1}^{n} (a + (m-1)b)^2 \le \int_0^n (a + xb)^2\, dx = \frac{1}{3b}[(a+nb)^3 - a^3] \le ca^2(nb)^3.$$

Taking $n = \lceil \varepsilon^{-3/32}/N_0 \rceil$ gives the desired bound. A similar calculation (in fact there is some simplification) gives the same upper bound for

$$P(\hat{\tau}_m < \hat{R}_m \text{ for some } 1 \le m \le N_T.)$$

□

Lemma 2.8.

$$P(\min_{1\le m\le N_T} (\tau'_{m+1} - R_m) \le \sqrt{\varepsilon}) \le c_{2.8} e^{c_b T}(M+1)^2 \varepsilon^{1/40}.$$

Proof. Define $S_m \supset G_m$ as G_m (in Lemma 2.1) but without the lower bounds on $\wedge_{i=2}^{m+1} \tau'_i - R_{i-1}$ or $\wedge_{i=1}^{m} \hat{R}_i - \hat{R}_{i-1}$. Note that $S_m \in \mathcal{F}_{R_m}$ and if $\omega \in S_m$, then

$$|X^i_{R_m} - X^j_{R_m}| \ge \varepsilon^{7/8} \text{ for all distinct } i, j \in J(R_m-). \tag{2.49}$$

In addition, since $Y^*_m \le \frac{\varepsilon}{\kappa}\log(1/\varepsilon)$ we have that for all $i \in J(R_m-) - \{\mu_m\}$, $j \in J(R_m) - J(R_m-)$

$$|X^i_{R_m} - X^j_{R_m}| \ge \varepsilon^{7/8} - \frac{\varepsilon}{\kappa}\log(1/\varepsilon) \ge \varepsilon^{7/8}/2 \tag{2.50}$$

since $\varepsilon < \varepsilon_1(\kappa)$ (recall (2.29)).

If T_0 is the return time to zero of the random walk Z in Lemma 2.6, we have (see P 26.2 in [44] for $d = 3$ and project down for $d > 3$)

$$P^{z_0}(T_0 < \infty) \le c|z_0|^{-1}. \tag{2.51}$$

Use (2.49), (2.50), and (2.51) with scaling, and the bound

$$|J(R_m-)| \le M + 1 + (m-1)N_0$$

to see that on $S_m \in \mathcal{F}_{R_m}$,

$$\begin{aligned} &P\left(\tau'_{m+1} - R_m \le \sqrt{\varepsilon} | \mathcal{F}_{R_m}\right) \\ &\quad\le c[(M+1+(m-1)N_0)^2 \varepsilon^{1/8} + (M+1+(m-1)N_0)N_0 \varepsilon^{1/8}] \\ &\quad\le c(M+1)^2 m^2 N_0^2 \varepsilon^{1/8}. \end{aligned} \tag{2.52}$$

Using the bound in Lemma 2.4(a), we conclude

$$\begin{aligned} &P\left(\min_{1\le m\le N_T} \tau'_{m+1} - R_m \le \sqrt{\varepsilon}, S_{N_T}\right) \\ &\quad\le P(N_T > n) + \sum_{k=1}^{n} P(N_T = k, S_k, \min_{1\le m\le k} \tau'_{m+1} - R_m \le \sqrt{\varepsilon}) \\ &\quad\le e^{c_b T}(M+1)(nN_0)^{-1} + \sum_{k=1}^{n}\sum_{m=1}^{k} P(S_m, \tau'_{m+1} - R_m \le \sqrt{\varepsilon}). \end{aligned}$$

Using (2.52) now, the above sum is at most

$$cN_0^2(M+1)^2 n\varepsilon^{1/8} \sum_{m=1}^{n} m^2 \le c(M+1)^2 n^4 \varepsilon^{1/8}.$$

Take $n = \lceil \varepsilon^{-1/40} \rceil$ and use Lemmas 2.4, 2.5, and 2.7 to bound $P(S^c_{N_T})$ to get the desired result. □

2.9 When nothing bad happens, (X, ζ) and $(\hat{X}, \hat{\zeta})$ are close

The next result gives a concrete bound on the difference between X and $\hat{X}$ and deals with the final interval $[R_m, R_m \wedge T]$. Let

$$\bar{G}_m = G_m \cap \{\hat{\tau}_m \geq \hat{R}_m\}, \;\; m \in \mathbb{N},$$

and for $0 < \beta \leq 1/2$, define

$$\begin{aligned}\tilde{G}_T^\beta = \bar{G}_{N_T} \cap \{\sup_{s \leq T} \sup_{j \in \hat{J}(s)} |X_s^j - \hat{X}_s^j| \leq \varepsilon^{1/6}\} \\ \cap \{T \notin \cup_{m=0}^{N_T - 1} [R_m, R_m + 2\varepsilon^\beta]\}. \end{aligned} \tag{2.53}$$

Allowing smaller β values will be useful in Sections 3 and 4, but for now the reader may take $\beta = 1/2$.

Lemma 2.9. *There is a $c_{2.9}$ and $\varepsilon_{2.9}(\kappa) > 0$ so that for any $T \geq 2\varepsilon^\beta$, $0 < \varepsilon < \varepsilon_{2.9}(\kappa)$,*

$$P((\tilde{G}_T^\beta)^c) \leq c_{2.9} e^{c_b T} (M+1)^2 \varepsilon^{\frac{1}{40} \wedge \frac{\beta}{3}}.$$

On $\tilde{G}_T^\beta$ we have $\hat{J}(s) = J(s)$ for all $s \in [T - \varepsilon^\beta, T]$, and $|\hat{X}_T^i - \hat{X}_T^j| \geq \varepsilon^{7/8}$ for all $i \neq j$ in $\hat{J}(T)$.

Proof. Dependence on β will be suppressed. For s as above, Lemma 2.2 implies $\hat{J}(s) = J(s)$ on $\tilde{G}_T$ since $s \in [R_{N_T - 1} + \sqrt{\varepsilon}, R_{N_T})$ on $\tilde{G}_T$. The last assertion of the Lemma holds on $\tilde{G}_T$ because on $\tilde{G}_T$, $\hat{\tau}_{N_T} \geq \hat{R}_{N_T}$ and

$$T \in [R_{N_T - 1} + \sqrt{\varepsilon}, R_{N_T}) = [\hat{R}_{N_T - 1} + \sqrt{\varepsilon}, \hat{R}_{N_T}).$$

Lemmas 2.4, 2.5, 2.7, and 2.8 imply

$$P(\bar{G}_{N_T}^c) \leq c e^{c_b T} (M+1)^2 \varepsilon^{1/40}. \tag{2.54}$$

To deal with the first additional good event in $\tilde{G}_T$, we note that by Lemma 2.3

$$\begin{aligned} &P(G_{N_T}, \sup_{s \leq T} \sup_{j \in \hat{J}(s)} |X_s^j - \hat{X}_s^j| > \varepsilon^{1/6}) \leq P(N_T > n) \\ &\quad + P\Big((n-1)\frac{\varepsilon}{\kappa} \log(1/\varepsilon) + \sum_{i=0}^{n-1} \sup_{j \in \hat{J}(R_i)} \sup_{s \in [R_i, R_i + \sqrt{\varepsilon}]} |\hat{X}_s^j - \hat{X}_{R_i}^j| + |X_s^j - X_{R_i}^j| > \varepsilon^{1/6}\Big) \end{aligned}$$

By (a) in Lemma 2.4 the first term is at most $e^{c_b T}(M+1)(nN_0)^{-1}$. If

$$(n-1)\frac{\varepsilon}{\kappa} \log(1/\varepsilon) < \varepsilon^{1/6}/2, \tag{2.55}$$

then it enough to bound

$$\begin{aligned} &P\left(\sum_{i=0}^{n-1} \sup_{j \in J(R_i)} \sup_{s \in [R_i, R_i + \sqrt{\varepsilon}]} |\hat{X}_s^j - \hat{X}_{R_i}^j| + |X_s^j - X_{R_i}^j| > \frac{\varepsilon^{1/6}}{2}\right) \\ &\quad \leq \sum_{i=0}^{n-1} (M + 1 + iN_0) 2P\Big(\sup_{s \leq \sqrt{\varepsilon}} |B_s^\varepsilon| > \frac{\varepsilon^{1/6}}{4n}\Big) \leq c(M+1)n^2 N_0 \cdot n^2 \varepsilon^{-2/6} \varepsilon^{1/2}, \end{aligned}$$

by the L^2 maximal inequality for martingales. If $n = \lceil \varepsilon^{-1/40} \rceil$ (so that (2.55) holds for $\varepsilon < \varepsilon_{2.9}(\kappa)$) the above gives

$$P\left(\bar{G}_{N_T}, \sup_{s\le T} \sup_{j\in \hat{J}(s)} |X^j_s - \hat{X}^j_s| > \varepsilon^{1/6}\right) \le ce^{c_bT}(M+1)\varepsilon^{1/40}. \tag{2.56}$$

The domination of X by $\bar{X}$ ensures that

$$\cup_{m=0}^{N_T-1}[R_m, R_m + 2\varepsilon^\beta] \subset \cup_{m=0}^{\bar{N}_T-1}[\bar{R}_m, \bar{R}_m + 2\varepsilon^\beta].$$

Therefore (recall $T > 2\varepsilon^\beta$) for any $\ell \in \mathbb{N}$

$$\begin{aligned} &P(T \in \cup_{m=0}^{N_T-1}[R_m, R_m + 2\varepsilon^\beta]) \\ &\quad \le P(\bar{N}_T > \ell) + P(T \in \cup_{m=1}^{\ell-1}[\bar{R}_m, \bar{R}_m + 2\varepsilon^\beta]). \end{aligned}$$

Lemma 2.4(a) shows that the first term is at most $e^{c_bT}(M+1)(\ell N_0)^{-1}$. Conditional on $\mathcal{F}_{\bar{R}_{m-1}}$, $\bar{R}_m - \bar{R}_{m-1}$ is an exponential random variable with rate $(M+1+(m-1)N_0)c^*$, so the second term is at most

$$\begin{aligned} &E\left(\sum_{m=1}^{\ell} P(T - 2\varepsilon^\beta - \bar{R}_{m-1} \le \bar{R}_m - \bar{R}_{m-1} \le T - \bar{R}_{m-1} | \mathcal{F}_{\bar{R}_{m-1}})\right) \\ &\le 2\varepsilon^\beta \sum_{m=1}^{\ell} ((M+1+(m-1)N_0)c^*) \le ce^{c_bT}(M+1)\ell^2\varepsilon^\beta. \end{aligned}$$

Taking $\ell = \lceil \varepsilon^{-\beta/3} \rceil$ then using (2.54) and (2.56) gives the desired bound on $P(\tilde{G}^c_T)$.
□

The next ingredient required for the convergence theorem is:

Lemma 2.10. *Assume $T > 2\varepsilon^\beta$, $t_0 \in [0, \varepsilon^\beta]$, and $\omega \in \tilde{G}^\beta_T$. If $\hat{\zeta}_{t_0}(j) = \zeta_{t_0}(j)$ for all $j \in \hat{J}(T - t_0)$, then $\hat{\zeta}_T(i) = \zeta_T(i)$, $i = 0, \dots, M$. In particular if $\hat{\zeta}_{t_0}(j) = \xi_{t_0}(X^j_{T-t_0})$ for $j \in J(T - t_0)$, then $\hat{\zeta}_T(i) = \xi_T(z_i)$ for $i = 0, \dots, M$.*

Remark 2.1. *By Lemma 2.9, $J(T-t_0) = \hat{J}(T-t_0)$ on $\tilde{G}^\beta_T$, and so all the necessary inputs required for both computations are prescribed in the above result.*

Proof. The last statement is immediate from the first and (2.17) with $r = T$.

By the definition of $G_{N_T} \supset \tilde{G}^\beta_T$ and Lemma 2.2 there is a unique $n < N_T$ so that

$$R_n + \sqrt{\varepsilon} \le T - \varepsilon^\beta \le t - t_0 < T < R_{n+1} \tag{2.57}$$

and

$$\hat{R}_m = R_m \text{ and } \hat{\mu}_m = \mu_m \text{ for } m \le n+1, \ \hat{K}(s) = K(s) \text{ for } s \in [0, T]. \tag{2.58}$$

As was noted in Section 2.6 the inductive definitions of ζ and $\hat{\zeta}$ are identical except the latter has hats on the relevant variables and uses $\approx_t$ in place of $\sim_t$. The above shows that in our current setting the relevant variables are the same with or without hats (recall we are using $\hat{U}_n = U_n$ in our coupled construction of $\hat{\zeta}$) and so it remains to show the equivalence relations are the same and we do this now for the initial extensions. That is, we extended $\hat{\zeta}_{t_0}$ to $\{0, \dots, K(T-t_0)\}$ by $\hat{\zeta}_{t_0}(k) = \hat{\zeta}_{t_0}(j)$ if $k \approx_{T-t_0} j \in \hat{J}(T-t_0) = J(T-t_0)$ (see the above Remark) and extended ζ_{t_0} in the same way but if $k \sim j \in J(T-t_0)$ which means $X^j_{T-t_0} = X^k_{T-t_0}$, and so we now show these equivalencies are the same and hence so are the extensions. Note that in applying (2.24) to extend $\hat{\zeta}_{t_0}$ we are using $\pi_m = \pi_{R_m,\gamma_m}(\sqrt{\varepsilon})$ for $m > 0$ and $\pi_0 = \pi_{0,z}(\sqrt{\varepsilon})$. This means two indices j,k in a family which has branched at time $\hat{R}_m = R_m$, $0 \le m \le n$ (if $m = 0$ this means two initial indices) are equivalent (in the $\approx$ sense) at time $T - t_0$ if their corresponding X paths coalesce by time $R_m + \sqrt{\varepsilon}$. Lemma 2.2 implies that on $\tilde{G}^{\beta}_T$ there are no coalescing events in $[0, T-t_0]$ (in fact on $[0,T]$) except for those in $[R_m, R_m + \sqrt{\varepsilon}]$, involving a common family born at R_m, for $m \le n$. Therefore, the above condition is equivalent to $X^j_{T-t_0} = X^k_{T-t_0}$ and the required result is proved.

The Lemma now follows easily by induction up the tree of X. In place of the above we must show equivalence of the equivalencies used in (2.20) and (2.27) at times R_m. Note here that for the indices of interest in (2.20) and (2.27) this is equivalent to the corresponding equivalencies at times $R_{m-1} + \sqrt{\varepsilon}$ and this follows as above for $m \ge 1$. $\square$

2.10 The branching Brownian motion and computation process

We now define a branching Brownian motion $\hat{X}^0$ starting at $x \in \mathbb{R}^d$ with paths in $\mathcal{D}$. Let $\{\pi^0_n, n \ge 1\}$ be an iid sequence of partitions with law ν_0 (defined in the second paragraph of Section 2.5). Particles in $\hat{X}^0$ branch at rate c^* and at the nth branching time, $|\pi^0_n| - 1$ particles are born at the location of the parent who also remains alive. After birth, particles in $\hat{X}^0$ move as independent Brownian motions in $\mathbb{R}^d$ with variance parameter σ^2. To couple $\hat{X}^0$ with the branching random walk $\hat{X}^\varepsilon$ from Section 2.5 we need two preliminary lemmas which allow us to couple the corresponding particle motions and offspring numbers, respectively, of the two branching processes.

Lemma 2.11. *We may define our scaled random walk B^ε and a d-dimensional Brownian motion B with variance σ^2, starting at 0, on the same space so that for some constant $c_{2.11}$*

$$P\left(\sup_{t \le T} |B^\varepsilon_t - B_t| \ge \sqrt{\varepsilon}\right) \le c_{2.11} T\varepsilon.$$

Proof. Apply Theorem 2.3(i) of Chapter 1 of [10] with $H(x) = x^6$, to see we may define the unscaled random walk B^1 (rate 1, step distribution p) and a Brownian motion as above, B' on the same space so that for all $S > 0$ and $r \ge 1$

$$P\left(\sup_{s \le S} |B^1_s - B'_s| \ge r\right) \le cSr^{-6}. \tag{2.59}$$

Although the above reference applies to discrete time random walks, we apply it to the step distribution $\sum_{i=1}^{N(1)} X_i$, where $\{X_i\}$ are iid $p(\cdot)$ and $N(1)$ is an independent Poisson(1) random variable. We arrive at the above after a short interpolation calculation for B^1.

To get the desired result from (2.59) we set $B_t^\varepsilon = \varepsilon B^1_{\varepsilon^{-2}t}$, $B_t = \varepsilon B'_{\varepsilon^{-2}t}$ and use $r = \varepsilon^{-1/2}$ to conclude that

$$\begin{aligned} P(\sup_{t\leq T} |B_t^\varepsilon - B_t| \geq \sqrt{\varepsilon}) &\leq P(\sup_{t\leq T} |\varepsilon B^1_{\varepsilon^{-2}t} - \varepsilon B'_{\varepsilon^{-2}t}| \geq \sqrt{\varepsilon}) \\ &\leq P(\sup_{t\leq \varepsilon^{-2}T} |B_t^1 - B'_t| \geq \varepsilon^{-1/2}) \\ &\leq c\varepsilon^{-2}T\varepsilon^3 = cT\varepsilon \end{aligned}$$

which proves the desired result. ◻

Lemma 2.12. *For each $\varepsilon > 0$ we may construct the sequence $\{\pi_n^0 : n \geq 1\}$ on the same space as $\{\pi_n^\varepsilon : n \geq 1\}$ so that for all $n \geq 1$,*

$$P(\pi_n^\varepsilon \neq \pi_n^0) \leq c_{2.12}\varepsilon^{3/4}.$$

Proof. The obvious way to couple π_n^ε and π_n^0 is to use the same system of rate one coalescing random walks $\{\hat{B}^{Y^i} : i = 0, \ldots, N_0\}$. If Z is as in Lemma 2.6, then by (2.45) and (2.47)

$$\begin{aligned} P(\pi_n^\varepsilon \neq \pi_n^0) &\leq \sup_x P^x(Z_s = 0 \text{ for some } s \geq \varepsilon^{-3/2}) \\ &\leq c(\varepsilon^{-3/2})^{-1/2} = c\varepsilon^{3/4}. \end{aligned}$$

◻

Let $x_\varepsilon \in \varepsilon\mathbb{Z}^d$ for $\varepsilon > 0$ and assume $x_\varepsilon \to x \in \mathbb{R}^d$. Our goal now is a joint construction of $(\hat{X}^\varepsilon, \hat{X}^0)$ started from (x_ε, x), and associated computation processes $(\hat{\zeta}^\varepsilon, \hat{\zeta}^0)$ with the property that if $\hat{\zeta}^\varepsilon, \hat{\zeta}^0$ have the same inputs then they will have the same outputs with probability close to one.

The branching random walk $\hat{X}^\varepsilon$ starting with a single particle at x_ε, along with the associated index sets $\hat{J}^\varepsilon(\cdot)$, branch times $\{\hat{R}_m^\varepsilon\}$, and parent variables $\{\hat{\mu}_m^\varepsilon\}$, are constructed as in Section 2.5 using the sequence $\{\pi_m^\varepsilon\}$ in Lemma 2.12. There is no initial coalescing step now as we are starting with a single particle. We use the coupled sequence $\{\pi_n^0 : n \geq 1\}$ to define the offspring numbers, branching times $\{R_n^0 : n \geq 1\}$, index sets $J^0(\cdot)$ and parent variables $\{\mu_n^0 : n \geq 1\}$ with the same conditional laws (given $\{\pi_m^0\}$) as in the definition of $\hat{X}^\varepsilon$. We may couple these two constructions so that for all $n \in \mathbb{Z}_+$, on the set

$$G_n^{0,\varepsilon} = \{\pi_m^0 = \pi_m^\varepsilon \text{ for all } 0 \leq m < n\},$$

we have

$$\hat{R}_m^\varepsilon = R_m^0, \hat{\mu}_m^\varepsilon = \mu_m^0, \text{ and } J^0(s) = \hat{J}^\varepsilon(s) \text{ for all } s < R_m^0, \text{ for all } m \leq n. \tag{2.60}$$

Define $N_t^0 = \inf\{m : R_m^0 > t\}$. Using these sequences we follow the prescription in Section 2.5 for constructing $\hat{X}$ but substituting Brownian motion paths for random walk paths. Couple these random walks and Brownian motions as in Lemma 2.11 at least as long as the branching structure of the two are the same. Note that if there are n branching events up to time T there are at most $1 + nN_0$ independent random walk segments and Brownian motions of length at most T to couple (recall our labeling scheme from Section 2.5). In addition to the errors in Lemma 2.11 there will be a small error from the difference in initial positions at time 0, and so we get

$$\begin{aligned} &P(G^{0,\varepsilon}_{N_T^0}, \sup_{s\le T}\sup_{j\in J^0(s)} |\hat{X}_s^{0,j} - \hat{X}_s^{\varepsilon,j}| \ge |x_\varepsilon - x| + \sqrt{\varepsilon}) \qquad (2.61)\\ &\le P(N_T^0 > n) + (1 + nN_0)c_{2.11}T\varepsilon + c_{2.12}n\varepsilon^{3/4}. \end{aligned}$$

The first time $\pi_n^\varepsilon \neq \pi_n^0$ we declare the coupling a failure and complete the definition of $\hat{X}_t^0$ for $t \ge R_n^0$ using random variables independent of $\hat{X}^\varepsilon$.

Fix $T > 0$ and $t_0 \in [0, T)$. Given $\hat{X}_t^0, J^0(t), 0 \le t \le T$, the sequences $\{\pi_n^0\}$, $\{R_n^0\}$, $\{\mu_n^0\}$ an independent sequence of iid uniform $[0, 1]$ random variables $\{U_m^0\}$, and initial inputs $\{\zeta_{t_0}(j) : j \in J^0(T - t_0)\}$, we define a computation process $\hat{\zeta}_t^0, t_0 \le t \le T$. The definition is analogous to that of $\hat{\zeta}_t$ given in Subsection 2.6 for $\hat{X}^\varepsilon$ started at a single point, but we use g_{1-i} in place of g_{1-i}^ε in (2.26). That is, as in (2.25), (2.26), we have

$$V_m^{0,j} = \hat{\zeta}^0_{(T-R_m^0)-}((m-1)N_0 + j),\ j = 1, \dots, N_0,$$

and if $i = \hat{\zeta}^0_{(T-R_m^0)-}(\mu_m^0)$, we have

$$\hat{\zeta}^0_{(T-\hat{R}_m)}(\mu_m^0) = \begin{cases} 1-i & \text{if } U_m^0 \le g_{1-i}(\hat{V}_m^0)/c^* \\ i & \text{otherwise.} \end{cases} \qquad (2.62)$$

We further couple $\hat{\zeta}^0$ and $\hat{\zeta}^\varepsilon$ by using the same sequence of independent uniforms: $\{U_m^0\} = \{U_m\}$ in their inductive definitions. Just as in (2.41) we can show that this sequence is independent of all the other variables used to define $\hat{X}^0$ and $\hat{\zeta}^0$, as required. We let $\hat{\mathcal{F}}_t^0$ denote the right-continuous filtration generated by $\hat{X}^0$, $\hat{X}^\varepsilon$ and $\hat{A}^0(t) = ((R_m^0, \mu_m^0, \pi_m^0, U_m)1(R_m^0 \le t))_{m\in N}$ as well as its counterpart for $\hat{X}^\varepsilon$.

Notation. $\tilde{G}_T^{0,\varepsilon} = G^{0,\varepsilon}_{N_T^0} \cap \{\sup_{s\le T}\sup_{j\in J^0(s)} |\hat{X}_s^{0,j} - \hat{X}_s^{\varepsilon,j}| \le |x_\varepsilon - x| + \sqrt{\varepsilon}\}$,
$\bar{G}_T^{0,\varepsilon} = \tilde{G}_T^{0,\varepsilon} \cap \Big\{U_m \notin \Big[\frac{g_i(\xi)\wedge g_i^\varepsilon(\xi)}{c^*}, \frac{g_i(\xi)\vee g_i^\varepsilon(\xi)}{c^*}\Big] \text{ for all } \xi \in \{0,1\}^{N_0}, m < N_T^0, i = 0, 1\Big\}$.

Lemma 2.13. *(a) On $\tilde{G}_T^{0,\varepsilon}$, we have*

$$\hat{R}_m^\varepsilon = R_m^0, \hat{\mu}_m^\varepsilon = \mu_m^0, \pi_m^\varepsilon = \pi_m^0, \text{ for all } m \le N_T^0, \text{ and } \hat{J}^\varepsilon(s) = J^0(s) \text{ for all } s \le T.$$

(b) $P((\tilde{G}_T^{0,\varepsilon})^c) \le c_{2.13}e^{c_bT}\varepsilon^{3/8}$.
(c) On $\bar{G}_T^{0,\varepsilon}$ we also have for any $t_0 \in [0, T)$, if $\hat{\zeta}^0_{t_0}(j) = \hat{\zeta}^\varepsilon_{t_0}(j)$ for all $j \in J^0(T - t_0)$, then $\hat{\zeta}_T^0(0) = \hat{\zeta}_T^\varepsilon(0)$.
(d) $P((\bar{G}_T^{0,\varepsilon})^c) \le c_{2.13}e^{c_bT}\Big[\varepsilon^{3/8} + \sqrt{\sum_{i=0}^1 \|g_i^\varepsilon - g_i\|_\infty}\Big]$.

Proof. (a) is immediate from (2.60) and the definition of $\tilde{G}_T^{0,\varepsilon}$.
(b) follows from (2.61) and the now familiar bound $P(N_T^0 > n) \leq \frac{e^{c_b T}}{nN_0}$, by setting $n = \lceil \varepsilon^{-3/8} \rceil$.
(c) On $\bar{G}_T^{0,\varepsilon}$, we see from (a) and the inductive definitions of $\hat{\zeta}^0$ and $\hat{\zeta}^\varepsilon$, that all the variables used to define $\hat{\zeta}_T^0(0)$ and $\hat{\zeta}_T^\varepsilon(0)$ coincide. Therefore these outputs can only differ due to the use of g_{i-1} in (2.62) and the use of g_{i-1}^ε in (2.26). By induction we may assume $\hat{V}_m = V_m^0$ and the additional condition defining $\bar{G}_T^{0,\varepsilon}$ now ensures that these two steps produce the same outputs.
(d) The additional condition defining $\bar{G}_T^{0,\varepsilon}$ fails with probability at most (recall $c^* \geq 1$)

$$P(N_T > n) + n2^{N_0}\Big[\sum_{i=0}^{1} \|g_i - g_i^\varepsilon\|_\infty\Big] \leq \frac{e^{c_b T}}{nN_0} + n2^{N_0}\Big[\sum_{i=0}^{1} \|g_i - g_i^\varepsilon\|_\infty\Big].$$

Now let $n = \lceil \sum_{i=0}^1 \|g_i - g_i^\varepsilon\|_\infty \rceil^{-1/2}$ and use (b) to complete the proof. $\square$

3 Proofs of Theorems 1.2 and 1.3

3.1 Proof of Theorem 1.2

We start with a key estimate giving the product structure in Theorem 1.2. This relies on the fact that duals starting at distant points with high probability will not collide. For some results we will need a quantitative estimate. Let $x_\varepsilon^k \in \varepsilon\mathbb{Z}^d$, $y_i \in \mathbb{Z}^d$ and $z_{ik}^\varepsilon = z_{ik} = x_\varepsilon^k + \varepsilon y_i$, for $0 \le i \le L$ and $1 \le k \le K$. Set

$$\Delta_\varepsilon = \min_{0\le i,i'\le L, 1\le k\ne k'\le K} |z_{ik} - z_{i'k'}|\varepsilon^{-1}.$$

The notation is taken to parallel that in the statement of Theorem 1.2 and hypothesis (1.37) implies

$$\lim_\varepsilon \Delta_\varepsilon = \infty. \tag{3.1}$$

Let $X = X^{z,T}$ be the dual process starting at z for the time period $[0,T]$, with associated computation process ζ_t which has initial inputs $\zeta_0(j) = \xi_0^\varepsilon(X_T^j)$, $j \in J(T)$. Let $z_k = (z_{ik}, i = 0,\dots,L)$ and consider the duals $X^{z_k,T}$, $1 \le k \le K$ defined as in Section 2 with their associated uniforms $\{U_m^k\}$ and parent variables $\{\mu_m^k\}$. These duals are naturally embedded in $X^{z,T}$, and although the numbering of the particles may differ, we do have

$$\{X_t^{z,j} : j \in J(t)\} = \cup_{k=1}^K \{X_t^{z_k,j} : j \in J^{z_k}(t)\},\ t \in [0,T]. \tag{3.2}$$

Define

$$\begin{aligned} V_{z,T,\varepsilon} = \inf\{t\in[0,T] :& X_t^{z_k,T,j} = X_t^{z_{k'},T,j'} \\ &\text{for some } 1 \le k \ne k' \le K, j \in J^{z_k}(t), j' \in J^{z_{k'}}(t)\}, \end{aligned} \tag{3.3}$$

where $\inf\emptyset = \infty$.

Lemma 3.1. $P(V_{z,T,\varepsilon} < \infty) \le c_{3.1}(K,L)e^{c_b T}(\Delta_\varepsilon)^{-(d-2)/(d+6)}$.

Proof. We may dominate $X^{z_k,T}$ by the branching random walks $\bar X^{z_k,T}$ from Section 2.8, and define $\bar N_T, \bar N_T^{z_k}$ in the obvious way. Then

$$\begin{aligned} &P(V < \infty) \\ &\le P(\max_{k\le K} \bar N_T^{z_k} > n) \\ &\quad + \sum_{1\le k\ne k'\le K} P(|\bar X_t^{z_k,j} - \bar X_t^{z_{k'},j'}| = 0\ \exists j \in \bar J^{z_k}(t), j' \in \bar J^{z_{k'}}(t), t \le T, \bar N_T^{z_k} \vee \bar N_T^{z_{k'}} \le n). \end{aligned}$$

The first term is bounded by $Ke^{c_b T}n^{-1}$ by Lemma 2.4(a). If $\{Y_m^*\}$ are iid, each equal in law to Y^*, and independent of B^ε, then for $R \ge 1$ the second term is at most

$$\begin{aligned} &\sum_{1\le k\ne k'\le K; 0\le i,i'\le L} (1+nN_0)^2 P_{z_{ik}-z_{i'k'}}(|B_{2t}^\varepsilon| \le \sum_{m=1}^n \varepsilon|Y_m^*|\ \exists t \le T) \\ &\le \sum_{1\le k\ne k'\le K; 0\le i,i'\le L} (1+nN_0)^2 \Big[nCe^{-\kappa R} + P_{z_{ik}-z_{i'k'}}(|B_{2t}^\varepsilon| \le n\varepsilon R\ \exists t \le T)\Big], \end{aligned}$$

where we used (1.6) in the last line. By Lemma 2.6(c), the probability in the last term is at most $c_{2.6}\Delta_\varepsilon^{-(2/3)(d-2)}(nR)^{2(d+1)/3}$, and so if $\delta = \Delta_\varepsilon^{-(2/3)(d-2)}$,

$$P(V<\infty) \le cK^2(L+1)^2 e^{c_b T} N_0^2\Big[n^{-1} + n^3 e^{-\kappa R} + \delta n^2 (nR)^{2(d+1)/3}\Big].$$

Now, optimizing over n and R, set $c_d = \frac{6}{d+6}$, $\kappa R = c_d \log(1/\delta)$ and $n = \lceil e^{\kappa R/4}\rceil$. Here we may assume without loss of generality that $\Delta_\varepsilon \ge M(\kappa)$ so that $R \ge 1$. A bit of arithmetic now shows the above bound becomes

$$P(V<\infty) \le c(K,L)e^{c_b T}\delta^{3/(2d+12)},$$

and the result follows. □

We suppose now that the assumptions of Theorem 1.2 are in force. That is, $T>0$ is fixed, ξ_0^ε has law λ_ε satisfying the local density condition (1.35) for a fixed $r\in(0,1)$, and (1.37) holds. It is intuitively clear that the density hypothesis is weakened by reducing r. To prove this, note that the boundedness of the density and uniformity in x of the convergence in (1.35) shows that the contribution to the density on larger blocks from smaller blocks whose density is not near v is small in L^1. We may therefore approximate the mass in a large block by the mass in smaller sub-blocks of density near v, and use the fact that the contributions close to the boundary of the large block is negligible to derive the density condition (1.35) for the larger blocks. As a result we may assume that $r<1/4$.

By inclusion-exclusion, it suffices to prove for $-1\le L_k\le L$,

$$\begin{aligned}\lim_{\varepsilon\to 0} P(\xi_T^\varepsilon(x_\varepsilon^k+\varepsilon y_{i_j}) = 1,\quad j=0,\dots,L_k,\ k=1,\dots K)&\\ = \prod_{k=1}^K \langle 1\{\xi(y_{i_j})=1,\quad j=0,\dots,L_k\}\rangle_{u(T,x^k)}.&\qquad (3.4)\end{aligned}$$

Allowing k-dependence in L_k and general subsets of the y_i's is needed for the inclusion-exclusion, but to reduce eyestrain we will set $i_j = j$ and $L_k = L$ in what follows. The general case requires only notational changes. By the duality equation (2.17),

$$\begin{aligned}P(\xi_T^\varepsilon(z_{ik}) = 1,\ i=0,\dots,L,k=1,\dots,K)&\qquad (3.5)\\ = P(\zeta_T(i,k)=1, i=0,\dots,L,k=1,\dots,K),&\end{aligned}$$

so (3.4) is then equivalent to

$$\begin{aligned}P(\zeta_T(i,k)=1,i=0,\dots,L,k=1,\dots,K)&\\ \to \prod_{k=1}^K\langle 1\{\xi(y_i)=1,i=0,\dots,L\}\rangle_{u(T,x^k)} \text{ as } \varepsilon\to 0.&\qquad (3.6)\end{aligned}$$

The proof of (3.6) uses the approach of [20], pp. 304-306.

To work with the left-hand side of (3.6) we need the following preliminary result to simplify the initial inputs ζ_0. Define

$$\beta = 1.9r \text{ and } t_\varepsilon = \varepsilon^\beta. \qquad (3.7)$$

Lemma 3.2. *Assume ξ_0^ε is independent of the rescaled random walks $\{B^{\varepsilon,w} : w \in \varepsilon\mathbb{Z}^d\}$ as in (2.8). Then for any $n \in \mathbb{N}$ and $k > 0$,*

$$\lim_{\varepsilon\to 0} \sup_{\substack{|w_1|,\ldots,|w_n|\le k, w_i\in\varepsilon\mathbb{Z}^d \\ w_i\neq w_j \text{ for } i\neq j}} \Big| E\Big(\prod_{i=1}^n \xi_0^\varepsilon(B_{t_\varepsilon}^{\varepsilon,w_i})\Big) - \prod_{i=1}^n v(w_i)\Big| = 0\,.$$

Proof. Recall the definitions of a_ε and Q_ε just before (1.34). For $z_1,\ldots,z_n \in a_\varepsilon\mathbb{Z}^d$ define

$$\Gamma(z_1,\ldots,z_n) = \{B_{t_\varepsilon}^{\varepsilon,w_i} \in z_i + Q_\varepsilon \text{ for } 1 \le i \le n\}\,,$$

and $\gamma(w_i,z_i) = P(B_{t_\varepsilon}^{\varepsilon,w_i} \in z_i + Q_\varepsilon)$, so that $P(\Gamma(z_1,\ldots,z_n)) = \prod_{i=1}^n \gamma(w_i,z_i)$. Let G be the union of the events $\Gamma(z_1,\ldots,z_n)$ over distinct $z_1,\ldots,z_n \in a_\varepsilon\mathbb{Z}^d$ such that $|z_i - w_i| \le k\sqrt{t_\varepsilon}$ for $1 \le i \le n$. We claim that $P(G^c)$ is small for k large enough. To see this, fix $\delta > 0$ and choose k large enough so that

$$P(|B_{t_\varepsilon}^{\varepsilon,w_i} - w_i| \ge (k-1)\sqrt{t_\varepsilon}) = P(|B_{t_\varepsilon}^{\varepsilon,0}| > (k-1)\sqrt{t_\varepsilon}) \le \delta/n\,. \tag{3.8}$$

By a standard estimate (and also since $r < 1/4$), for w_i as above and $i \neq j$,

$$P(|B_{t_\varepsilon}^{\varepsilon,w_i} - B_{t_\varepsilon}^{\varepsilon,w_j}| \le 2a_\varepsilon) \le c|Q_\varepsilon| P(B_{2t_\varepsilon}^{\varepsilon,0} = 0) \le c|Q_\varepsilon|(\varepsilon^{-2}t_\varepsilon)^{-d/2} \le c\varepsilon^{.05rd}\,,$$

which implies

$$P(|B_{t_\varepsilon}^{\varepsilon,w_i} - B_{t_\varepsilon}^{\varepsilon,w_j}| \le 2a_\varepsilon \text{ for some } 1 \le i < j \le n) \le Cn^2\varepsilon^{.05rd}\,. \tag{3.9}$$

By (3.9) and (3.8),

$$P(G^c) \le Cn^2\varepsilon^{.05rd} + \delta\ . \tag{3.10}$$

Now consider the decomposition

$$E\Big(\prod_{i=1}^n \xi_0^\varepsilon(B_{t_\varepsilon}^{\varepsilon,w_i}); G\Big) = \sum_{z_1,\ldots,z_n} E\Big(\prod_{i=1}^n \xi_0^\varepsilon(B_{t_\varepsilon}^{\varepsilon,w_i}); \Gamma(z_1,\ldots,z_n)\Big) \tag{3.11}$$

where the sum is taken over only those $(z_1,\ldots,z_n)$ used in the definition of G. A typical term in this sum takes the form

$$\sum_{e_1,\ldots,e_n\in Q_\varepsilon} E\Big(\prod_{i=1}^n \xi_0^\varepsilon(z_i+e_i) 1(B_{t_\varepsilon}^{\varepsilon,w_i} = z_i + e_i)\Big)$$
$$= \sum_{e_1,\ldots,e_n\in Q_\varepsilon} E\Big(\prod_{i=1}^n \xi_0^\varepsilon(z_i+e_i)\Big) \prod_{i=1}^n P(B_{t_\varepsilon}^{\varepsilon,w_i} = z_i + e_i)\ .$$

Since $\sqrt{t_\varepsilon} \gg a_\varepsilon$, the probabilities $P(B_{t_\varepsilon}^{\varepsilon,w_i} = z_i + e_i) = P(B_{t_\varepsilon}^{\varepsilon,0} = z_i - w_i + e_i)$ are almost constant over $e_i \in Q_\varepsilon$. In fact, a calculation, using the version of the local central limit theorem in the Remark after P7.8 in [44] to expand $P(B_{t_\varepsilon}^{\varepsilon,0} = z_i - w_i + e_i) = P(B_{\varepsilon^{-2}t_\varepsilon}^0 = (z_i - w_i + e_i)/\varepsilon)$, shows that

$$\lim_{\varepsilon\to 0} \sup_{\substack{e,e'\in Q_\varepsilon \\ |z_i-w_i|\le k\sqrt{t_\varepsilon}}} \frac{P(B_{t_\varepsilon}^{\varepsilon,0} = z_i - w_i + e)}{P(B_{t_\varepsilon}^{\varepsilon,0} = z_i - w_i + e')} = 1 \tag{3.12}$$

The continuous time setting is easily accommodated, for example by noting that along multiples of a fixed time it becomes a discrete time random walk.

Consequently, for all sufficiently small $\varepsilon > 0$, we have $k\sqrt{t_\varepsilon} < 1$ and uniformly in $|w_i| \le k$, $|z_i - w_i| \le k\sqrt{t_\varepsilon}$ and $e \in Q_\varepsilon$,

$$1 - \delta \le \frac{|Q_\varepsilon| P(B^{\varepsilon,w_i}_{t_\varepsilon} = z_i + e)}{\gamma(w_i, z_i)} \le 1 + \delta \; . \tag{3.13}$$

Using this bound and the fact that the z_i are distinct we have

$$\begin{aligned}\sum_{e_1,\ldots,e_n \in Q_\varepsilon} & E\Big(\prod_{i=1}^n \xi_0^\varepsilon(z_i + e_i)\Big) \prod_{i=1}^n P(B^{\varepsilon,w_i}_{t_\varepsilon} = z_i + e_i) \\ &\le \sum_{e_1,\ldots,e_n \in Q_\varepsilon} E\Big(\prod_{i=1}^n \xi_0^\varepsilon(z_i + e_i)\Big) \frac{(1+\delta)^n}{|Q_\varepsilon|^n} \prod_{i=1}^n \gamma(w_i, z_i) \\ &= (1+\delta)^n E\Big(\prod_{i=1}^n D(z_i, \xi_0^\varepsilon)\Big) P(\Gamma(z_1, \ldots, z_n)) \,. \end{aligned}$$

The continuity of v implies that for small enough ε, for all $|w| \le k$ and $|z - w| \le k\sqrt{t_\varepsilon}$, $|v(w) - v(z)| < \delta$. Also for sufficiently small ε and $z \in a_\varepsilon \mathbb{Z}^d$, $|z| \le k+1$, we have $P(D(z, \xi_0^\varepsilon) > v(z) + \delta) \le \delta/n$. Thus

$$E\Big(\prod_{i=1}^n D(z_i, \xi_0^\varepsilon)\Big) \le \delta + \prod_{i=1}^n (v(z_i) + \delta) \le \delta + \prod_{i=1}^n (v(w_i) + 2\delta) \; .$$

Returning to the decomposition (3.11), the above bounds imply that for sufficiently small ε,

$$\begin{aligned} E\Big(\prod_{i=1}^n \xi_0^\varepsilon(B^{\varepsilon,w_i}_{t_\varepsilon}); G\Big) &\le (1+\delta)^n \Big[\delta + \prod_{i=1}^n (v(w_i) + 2\delta)\Big] \sum_{z_1,\ldots,z_n} P(\Gamma(z_1, \ldots, z_n)) \\ &\le (1+\delta)^n \Big[\delta + \prod_{i=1}^n (v(w_i) + 2\delta)\Big] \,. \end{aligned}$$

Let $\varepsilon \to 0$ and then $\delta \to 0$ above and in (3.10) to obtain

$$\limsup_{\varepsilon \to 0} \sup_{|w_1|,\ldots,|w_n| \le k} \Big(E\Big(\prod_{i=0}^n \xi_0^\varepsilon(B^{\varepsilon,w_i}_{t_\varepsilon})\Big) - \prod_{i=1}^n v(w_i)\Big) \le 0 \,.$$

A similar argument gives a reverse inequality needed to complete the proof. □

We break the proof of (3.6) into three main steps. Introduce

$$S = T - t_\varepsilon = T - \varepsilon^\beta.$$

Step 1. Reduction to Bernoulli inputs and $K = 1$.
Let $\tilde{X} = \tilde{X}^{z,T}$ be the modification of the dual in which particles ignore reaction and coalescing events on $[S,T]$, and let $\tilde{\zeta}_t$ be the associated computation process with inputs $\tilde{\zeta}_0(j) = \xi_0^\varepsilon(\tilde{X}_T^j)$. That is, $\tilde{X}_t = X_t$ for $t \in [0,S]$, and during the time period $[S,T]$, $\tilde{X}_t^j$, $j \in J(S)$ follows the same path as X_t^j until the first time a reaction or coalescence occurs, at which time all the $\tilde{X}_t^j$ switch to following completely independent B^ε random walks.

On the event $\tilde{G}_T^\beta$ defined in (2.53) there are no reaction or coalescing events during $[S,T]$. Thus, $\tilde{X}_t = X_t$ for all $t \in [0,T]$ on $\tilde{G}_T^\beta$, so it follows from Lemma 2.9 that

$$P(\zeta_t \neq \tilde{\zeta}_t \text{ for some } t \in [0,T]) \leq c_{2.9}[(L+1)K]^2 e^{c_b T} \varepsilon^{\frac{1}{40} \wedge \frac{\beta}{3}}\,. \tag{3.14}$$

Let $\psi_\varepsilon(x) = P_{t_\varepsilon}^\varepsilon \xi_0^\varepsilon(x)$, where

$$P_t^\varepsilon f(x) = E(f(x + B_t^\varepsilon)),\ x \in \varepsilon\mathbb{Z}^d,\ \text{is the semigroup of } B^\varepsilon, \tag{3.15}$$

and let $W_1, W_2, \dots$ be an iid sequence of uniforms on the interval $[0,1]$, independent of ξ_0^ε and the random variables used in Section 2. We will use this sequence throughout the rest of this section and also in Section 4. Define a second computation process $\zeta_t^*, t_\varepsilon \leq t \leq T$, for $\tilde{X}$, with inputs

$$\zeta_{t_\varepsilon}^*(j) = 1\{W_j \leq v(\tilde{X}_S^j)\}, \quad j \in J(S). \tag{3.16}$$

It is clear that conditional on $\sigma(\xi_0^\varepsilon) \vee \mathcal{F}_\infty$, the variables $\zeta_{t_\varepsilon}^*(j)$, $j \in J(S)$, respectively $\tilde{\zeta}_{t_\varepsilon}(j)$, $j \in J(S)$, are independent Bernoulli with means $v(\tilde{X}_S^j)$, respectively $\psi_\varepsilon(\tilde{X}_S^j)$. Let $\bar{X} = \bar{X}^{z,T}$ be the branching random walk dominating X which was introduced in Section 2.8. If we fix $\delta > 0$, then using Lemma 2.4(a) it is not hard to see that there exist n, k such that for all ε sufficiently small,

$$P(|\bar{J}(S)| \leq n \text{ and } |\bar{X}_j^\varepsilon(S)| \leq k \text{ for all } j \in \bar{J}(S)) > 1 - \delta. \tag{3.17}$$

It now follows from (3.17), Lemma 3.2 and the definitions of $\tilde{X}$, $\tilde{\zeta}_0$ and $\zeta_{t_\varepsilon}^*$ that for any $b : \mathbb{Z}^+ \to \{0,1\}$,

$$\begin{aligned}
&|P(\tilde{\zeta}_{t_\varepsilon}(j) = b_j, j \in J(S)) - P(\zeta_{t_\varepsilon}^*(j) = b_j, j \in J(S))| \\
&\quad \leq E(|P(\tilde{\zeta}_{t_\varepsilon}(j) = b_j, j \in J(S)|\mathcal{F}_S \vee \sigma(\xi_0^\varepsilon)) - P(\zeta_{t_\varepsilon}^*(j) = b_j, j \in J(S))|\mathcal{F}_S \vee \sigma(\xi_0^\varepsilon))|) \\
&\quad \to 0 \text{ as } \varepsilon \to 0.
\end{aligned}$$

As a consequence, since both $\tilde{\zeta}_t, \zeta_t^*, t_\varepsilon \leq t \leq T$ are defined relative to $\tilde{X}$ with identical $\{U_m\}$, $\{\mu_m\}$ and $\{R_m\}$, by conditioning on the input values, the above implies

$$\begin{aligned}
&P(\tilde{\zeta}_T(i,k) = 1, i = 0, \dots, L, k = 1, \dots, K) \\
&\qquad - P(\zeta_T^*(i,k) = 1, i = 0, \dots, L, k = 1, \dots, K) \to 0 \text{ as } \varepsilon \to 0.
\end{aligned} \tag{3.18}$$

Let ζ_t^{*,z_k} be the computation process associated with $X^{z_k,T}$, $1 \leq k \leq K$ with inputs as in (3.16). That is, for $j \in J^{z_k}(S)$ there exists a $j' \in J(S)$ with $X_S^{j'} = X_S^{z_k,j}$

(by (3.2)) and we set $\zeta^{*,z_k}_{t_\varepsilon}(j) = 1\{W_{j'} \le v(X^{z_k,j}_S)\}$. Up to time $V = V_{z,T,\varepsilon}$ the duals $X^{z_k,T}$, $k \le K$, use independent random walk steps and branching mechanisms, and on $\{V = \infty\}$ the computation processes ζ^{*,z_k} also use independent uniforms and parent variables as well as independent inputs at time t_ε. It follows that (see below)

$$\begin{aligned}
|P(\zeta^*_T(i,k) = 1, i = 0,\dots,L, k = 1,\dots,K) \\
- \prod_{k=1}^{K} P(\zeta^{*,z_k}_T(i) = 1, 0 \le i \le L)| \\
\le P(V < \infty) \to 0 \text{ as } \varepsilon \to 0.
\end{aligned} \tag{3.19}$$

The last limit follows from Lemma 3.1 and (3.1). Perhaps the easiest way to see the first inequality is to extend $X^{z_k,T}_t$ to $t \in [V,T]$ by using independent graphical representations and define the corresponding computation processes ${\zeta'}^{*,z_k}$ using independent collections of $\{W_j\}$'s for the inputs at time t_ε. The resulting computation processes ${\zeta'}^{*,z_k}$ are then independent, each ${\zeta'}^{*,z_k}$ is equal in law to ζ^{*,z_k}, and the two are identical for all k on $\{V = \infty\}$. On this set we also have

$$\{\zeta^*_T(i,k) : i,k\} = \{\zeta^{*,z_k}_T(i) : i,k\},$$

and so (3.19) follows. It is therefore enough to set $K = 1$ and drop the superscript k. Altering our notation to $z = (z_i)$, $z_i = x_\varepsilon + \varepsilon y_i$ where $x_\varepsilon \to x$, it suffices now to prove

$$P(\zeta^*_T(i) = 1, 0 \le i \le L) \to \langle 1\{\xi(y_i) = 1, 0 \le i \le L\}\rangle_{u(T,x)} \text{ as } \varepsilon \to 0. \tag{3.20}$$

Step 2. Reduction to $L = 0$. Let $\hat{X} = \hat{X}^{z,T}, 0 \le t \le T$ be the branching random walk started at z, with associated computation process $\hat\zeta_t, t_\varepsilon \le t \le T$. We suppose that X and $\hat{X}$ are coupled as in Subsection 2.7, and that $\hat\zeta_t$ has initial inputs

$$\hat\zeta_{t_\varepsilon}(j) = 1\{W_j \le v(\hat{X}^j_S)\}, j \in \hat{J}(S).$$

On the event $\tilde{G}^\beta_T$, $J(S) = \hat{J}(S)$ and all the differences $|X^j_S - \hat{X}^j_S|$, $j \in J(S)$ are small. It therefore follows from (3.17) (if we take $Y^i = 0$ $\bar{X}$ will stochastically dominate $\hat{X}$), the continuity of v, the definitions of $\zeta^*_{t_\varepsilon}$ and $\hat\zeta_{t_\varepsilon}$, and Lemma 2.9 that

$$P(\tilde{G}^\beta_T, \zeta^*_{t_\varepsilon}(j) = \hat\zeta_{t_\varepsilon}(j) \text{ for all } j \in J(S)) \to 1 \text{ as } \varepsilon \to 0. \tag{3.21}$$

By Lemma 2.10, on the event in (3.21), the outputs ζ^*_T and $\hat\zeta_T$ agree, and consequently

$$P(\zeta^*_T(i) = 1, 0 \le i \le L) - P(\hat\zeta_T(i) = 1, 0 \le i \le L) \to 0 \text{ as } \varepsilon \to 0. \tag{3.22}$$

Using the branching structure we can now reduce to the case $L = 0$. To see this, let $\hat{X}^{z_i,T}$ be the branching random walk started from $z_i = x_\varepsilon + \varepsilon y_i$, with associated computation process $\hat\zeta^{z_i}_t, t_\varepsilon \le t \le T$ with initial inputs $\hat\zeta^{z_i}_{t_\varepsilon}(j)$ which, conditional on

$\hat{X}_t^{z_i}, 0 \le t \le S$ are independent with means $v(\hat{X}_S^{z_i,j})$. The branching property and definition of $\hat{X}_0$ in (2.23) imply (recall $\pi_0 = \pi_0(\varepsilon, z)$ from just after (2.21))

$$P(\hat{\zeta}_T(i) = 1, i = 0, \dots, L) = \sum_{\pi \in \Pi_L} P(\pi(0, z) = \pi) \prod_{j \in J(\pi)} P(\hat{\zeta}_T^{z_j}(0) = 1).$$

Since $z_i = x_\varepsilon + \varepsilon y_i$ for $i = 0, \dots, L$, translation invariance implies $\pi_0(\varepsilon, z)$ converges weakly to $\pi(0, y)$, where $i \sim j$ in $\pi(0, y)$ means $\hat{B}_s^{y_i} = \hat{B}_s^{y_j} =$ for some $s \ge 0$. Since $z_j \to x$ as $\varepsilon \to 0$ for $i = 0, \dots, L$, if we can establish, for any $x_\varepsilon \to x$ as $\varepsilon \to 0$,

$$P(\hat{\zeta}_T^{x_\varepsilon}(0) = 1) \to \hat{u}(T, x) \text{ as } \varepsilon \to 0, \tag{3.23}$$

for some $\hat{u} : \mathbb{R}_+ \times \mathbb{R}^d \to [0, 1]$, then the convergence $\pi(\varepsilon, z) \Rightarrow \pi(0, y)$ implies

$$\begin{aligned} P(\hat{\zeta}_T(i) = 1, i = 0, \dots, L) &\to \sum_{\pi \in \Pi_L} P(\pi(0, y) = \pi)(\hat{u}(T, x))^{|\pi|} \\ &= \langle 1\{\xi(y_i) = 1, 0 \le i \le L\} \rangle_{\hat{u}(T,x)} \text{ as } \varepsilon \to 0\,, \end{aligned} \tag{3.24}$$

where (1.40) is used in the last line. Combining this with (3.22) gives the desired result (3.20) but with $\hat{u}$ in place of u, that is, we get

$$P(\zeta_T^*(i) = 1, 0 \le i \le L) \to \langle 1\{\xi(y_i) = 1, 0 \le i \le L\} \rangle_{\hat{u}(T,x)} \text{ as } \varepsilon \to 0. \tag{3.25}$$

We first turn now to the proof of (3.23).

Step 3. Convergence and identification of the limit. Let $\hat{X}^0$ be the branching Brownian motion started at $x \in \mathbb{R}^d$ run over the time period $[0, T]$, with associated computation process $\hat{\zeta}_t^0, t_\varepsilon \le t \le T$ with inputs

$$\hat{\zeta}_{t_\varepsilon}^0(j) = 1\{W_j \le v(\hat{X}_S^{0,j})\}, j \in J^0(S).$$

Using the obvious analogue of (3.17) for $\hat{X}^0$, the continuity of v and the definitions of $\hat{\zeta}_{t_\varepsilon}$ and $\hat{\zeta}_{t_\varepsilon}^0$, Lemma 2.13 (and the uniform convergence of g_i^ε to g_i) implies

$$P(\bar{G}_T^{0,\varepsilon}, \hat{\zeta}_{t_\varepsilon}(j) = \hat{\zeta}_{t_\varepsilon}^0(j) \text{ for all } j \in J^0(S)) \to 1 \text{ as } \varepsilon \to 0. \tag{3.26}$$

By Lemma 2.13(c), on the event in (3.26), $\hat{\zeta}_T(0) = \hat{\zeta}_T^0(0)$, and thus

$$P(\hat{\zeta}_T^{x_\varepsilon}(0) = 1) - P(\hat{\zeta}_T^0(0) = 1) \to 0 \text{ as } \varepsilon \to 0, \tag{3.27}$$

where we note that both quantities in the above depend on ε. If we take the initial inputs for the computation process $\hat{\zeta}^{0,*}$, $0 \le t \le T$, at time 0 to be

$$\hat{\zeta}_0^{0,*}(j) = 1\{W_j \le v(X_T^{0,j})\}, j \in J^0(T), \tag{3.28}$$

it is now routine to see that $P(\hat{\zeta}_{t_\varepsilon}^0 = \hat{\zeta}_{t_\varepsilon}^{0,*}) \to 1$ as $\varepsilon \to 0$, and so by (3.27)

$$\lim_{\varepsilon \to 0} P(\hat{\zeta}_T^{x_\varepsilon}(0) = 1) = P(\hat{\zeta}_T^{0,*}(0) = 1) \equiv \hat{u}(T, x). \tag{3.29}$$

This proves (3.23), hence (3.24) and so to complete the proof of (3.20), and hence Theorem 1.2, we only need show $\hat{u} = u$:

Lemma 3.3. *Let $\hat{X}^0_t, 0 \le t \le T$ be the branching Brownian motion started at $x \in \mathbb{R}^d$, with associated computation process $\hat{\zeta}^{0,*}_t, 0 \le t \le T$ with initial inputs as in (3.28). Then*

$$P(\hat{\zeta}^{0,*}_T(0) = 1) = u(T, x),$$

where u is the solution of the PDE (1.33).

Proof. This is very similar to the proof in Section 2(e) of [20]. Recall P^ε_t is the semigroup of B^ε. Let $x \in \mathbb{R}^d$ and $x_\varepsilon \in \varepsilon\mathbb{Z}^d$ satisfy $|x - x_\varepsilon| \le \varepsilon$ and let ξ^ε be our rescaled particle system where $\{\xi^\varepsilon_0(\varepsilon y) : y \in \mathbb{Z}^d\}$ are independent Bernoulli random variables with means $\{v(\varepsilon y) : y \in \mathbb{Z}^d\}$. If

$$d_\varepsilon(\varepsilon y, \xi_\varepsilon) = -\xi(y)h^\varepsilon_0(y, \xi) + (1 - \xi(y))h^\varepsilon_1(y, \xi),\ y \in \mathbb{Z}^d,\ \xi \in \{0,1\}^{\mathbb{Z}^d}, \tag{3.30}$$

then the martingale problem for ξ^ε shows that (cf. (2.25) of [20])

$$E(\xi^\varepsilon_T(x_\varepsilon)) = E(P^\varepsilon_T \xi^\varepsilon_0(x_\varepsilon)) + \int_0^T E_{x_\varepsilon} \times E(d_\varepsilon(B^\varepsilon_{T-s}, \xi^\varepsilon_s))\, ds, \tag{3.31}$$

where $B^\varepsilon_0 = x_\varepsilon$ under P_{x_ε}. Our hypotheses on ξ^ε_0 imply

$$E(P^\varepsilon_T \xi^\varepsilon_0(x_\varepsilon)) = P^\varepsilon_T v(x_\varepsilon) \to P_T v(x) \text{ as } \varepsilon \to 0,$$

where P_t is the d-dimensional Brownian semigroup with variance σ^2. Recall we have proved ((3.24) and the preceding results) that

$$\lim_{\varepsilon \to 0} P(\xi^\varepsilon_T(x_\varepsilon + \varepsilon y_i) = \eta_i,\ i = 0, \dots, L) = \langle 1\{\xi(y_i) = \eta_i,\ i = 0, \dots, L\}\rangle_{\hat{u}(T,x)}.$$

Now use the above with Fubini's theorem, the uniform convergence of g^ε_i in (1.7) and the coupling of B^ε and B in Lemma 2.11 to conclude that

$$\begin{aligned}
&\lim_{\varepsilon \to 0} E_{x_\varepsilon} \times E(d_\varepsilon(B^\varepsilon_{T-s}, \xi^\varepsilon_s)) \\
&= \lim_{\varepsilon \to 0} E_{x_\varepsilon} \times E \times E_Y\Big(-\xi^\varepsilon_s(B^\varepsilon_{T-s}) g^\varepsilon_0(\xi^\varepsilon_s(B^\varepsilon_{T-s} + \varepsilon Y^1), \dots, \xi_s(B^\varepsilon_{T-s} + \varepsilon Y^{N_0})) \\
&\qquad\qquad + (1 - \xi^\varepsilon_s(B^\varepsilon_{T-s})) g^\varepsilon_1(\xi^\varepsilon_s(B^\varepsilon_{T-s} + \varepsilon Y^1), \dots, \xi^\varepsilon_s(B^\varepsilon_{T-s} + \varepsilon Y^{N_0}))\Big) \\
&= E_x\Big(\langle -\xi(0)h_0(0, \xi) + (1 - \xi(0))h_1(0, \xi)\rangle_{\hat{u}(s, B_{T-s})}\Big) \\
&= E_x(f(\hat{u}(s, B_{T-s}))),
\end{aligned}$$

the last by (1.30). Now use the above to take limits in (3.31) to show that $\hat{u}$ solves the weak form of (1.33). As in Lemma 2.21 of [20] it follows that $\hat{u}$ solves (1.33) and so equals u. □

The following asymptotic independence result follows easily from Step 1 in the above argument.

Proposition 3.4. *If $K \in \mathbb{N}$, there is a $c_{3.4}(K)$ so that if $z_1, \dots, z_K \in \varepsilon\mathbb{Z}^d$ satisfy $\inf_{j \neq k} |z_j - z_k| \geq \varepsilon^{1/4}$ and ξ_0^ε is deterministic, then*

$$|E(\prod_{k=1}^{K} \xi_T^\varepsilon(z_k)) - \prod_{k=1}^{K} E(\xi_T^\varepsilon(z_k))| \leq c_{3.4}(K) e^{c_b T} \varepsilon^{1/12}.$$

Proof. Define $V = V_{z,T,\varepsilon}$ as in (3.3) but now with $z_k \in \varepsilon\mathbb{Z}^d$, that is $L = 0$. Use the dual equation (3.5) and argue just as in the derivation of (3.19) to see that

$$|E(\prod_{i=1}^{K} \xi_T^\varepsilon(z_i)) - \prod_{i=1}^{K} E(\xi_T^\varepsilon(z_i))| \leq P(V < \infty).$$

The fact that ξ_0^ε is deterministic makes the independence argument simpler in this setting. Now use Lemma 3.1 and the separation hypothesis on the z_k's to bound the right-hand side of the above by

$$c_{3.1}(K, 0) e^{c_b T} \varepsilon^{(3/4)(d-2)/(d+6)} \leq c_{3.1}(K, 0) e^{c_b T} \varepsilon^{1/12}.$$

□

3.2 Proof of Theorem 1.3

Proof. Let $t > 0$ and choose $\eta(\varepsilon) \downarrow 0$ so that $\eta(\varepsilon)/\varepsilon \to \infty$ and $\eta(\varepsilon)/\delta(\varepsilon) \to 0$. Recall $I_\delta(x)$ is the semi-open cube containing x defined prior to Theorem 1.3. Write $E((\tilde{u}^\varepsilon(t,x) - u(t,x))^2) =$

$$= \Big(\frac{\varepsilon}{\delta(\varepsilon)}\Big)^{2d} \sum_{x_1, x_2 \in I_\delta(x)} E(\xi_t^\varepsilon(x_1)\xi_t^\varepsilon(x_2) - u(t,x)(\xi_t^\varepsilon(x_1) + \xi_t^\varepsilon(x_2)) + u(t,x)^2).$$

The contribution to the above sum from $|x_1 - x_2| \leq \eta(\varepsilon)$ is trivially asymptotically small, uniformly in x, as $\varepsilon \to 0$. Theorem 1.2 shows that the expectation in the above sum goes to zero uniformly in $x_1, x_2 \in I_\delta(x)$, $|x_1 - x_2| \geq \eta(\varepsilon)$, $x \in [-K, K]^d$ as $\varepsilon \to 0$. The result follows. □

4 Achieving low density

The first step in the proof of Theorem 1.5 is to use the convergence to the partial differential equation in Theorem 1.2, and more particularly the estimates in the proof, to get the particle density in (1.34) low on a linearly growing region.

As we will now use the partial differential equation results in Section 1.3, we begin by giving the short promised proofs of Propositions 1.6 , 1.7 and 1.9.

Proof of Proposition 1.9. Set $\eta = |r|/3$ and let L_δ^0, C_0 and c_0 be the constants in Proposition 1.8, and define

$$L_\delta = L_\delta^0, \quad c_1 = c_0, \quad t_\delta = L_\delta^0 \cdot 3\sqrt{d}/|r|, \quad C_1 = (C_0 \vee 1)e^{c_0 t_\delta}\,.$$

Suppose $t \geq t_\delta$, $L \geq L_\delta$ and $|x| \leq L + (|r|/3)t/\sqrt{d}$. Then we may write $x = x_0 + y$, where

$$|y| \leq \frac{2|r|}{3}\frac{t}{\sqrt{d}} \quad \text{ and } \quad |x_0| \leq L - \frac{|r|t}{3\sqrt{d}} \leq L - \frac{|r|t_\delta}{3\sqrt{d}} = L - L_\delta^0. \tag{4.1}$$

For $t \geq 0$ and $z \in \mathbb{R}^d$ define $\tilde{u}(t,z) = u(t, x_0 + z)$. If $|z| \leq L_\delta^0$, then $|x_0 + z| \leq |x_0| + L_\delta^0 \leq L$, which implies that $\tilde{u}(0,z) \leq \rho - \delta$. Applying Proposition 1.8 to $\tilde{u}$, and recalling the bound on $|y|$ in (4.1), which implies $|y|_2 \leq \frac{2|r|}{3}t$ we have that for $t \geq t_\delta$, and $|x| \leq L + (|r|/3)t/\sqrt{d}$

$$u(t,x) = \tilde{u}(t,y) \leq C_0 e^{-c_0 t} \leq C_1 e^{-c_1 t}\,.$$

Since the right-hand side above is at least 1 if $t \leq t_\delta$, the above bound follows for all $t \geq 0$, and we have proved the result with $w = |r|/6\sqrt{d}$. □

Proof of Proposition 1.7. Extend $f|_{[0,1]}$ to a smooth function $\tilde{f}$ on $[0, 1+\delta_0]$ so that $\tilde{f} > 0$ on $(1, 1+\delta_0)$, $\tilde{f}(1+\delta_0) = 0$, $\tilde{f}'(1+\delta_0) < 0$ and $\int_0^{1+\delta_0} \tilde{f}(u)du < 0$. The situation is now as in Proposition 1.9 with 0, 1 and $1+\delta_0$ playing the roles of 0, ρ and 1. As the solutions take values in $[0,1]$ the extension will not affect the solutions and the result follows from Proposition 1.9. □

Proof of Proposition 1.6. The version of Proposition 1.7 with the roles of 0 and 1 reversed, applied on the interval $(0,\alpha)$ shows there are positive constants L, c, and C so that if $u(0,x) \geq \alpha/2$ for $|x| \leq L$, then

$$u(t,x) \geq \alpha - Ce^{-ct} \text{ for } |x| \leq L + 2wt.$$

It is here that we need $f'(\alpha) < 0$, corresponding to $f'(0) < 0$ in Proposition 1.7. By Theorem 3.1 of Aronson and Weinberger [2] there is a T_0 so that

$$u(T,x) \geq \alpha/2 \text{ for } |x| \leq L \text{ and } T \geq T_0.$$

Therefore we have

$$u(t+T_0, x) \geq \alpha - Ce^{-ct} \text{ for } |x| \leq L + 2w(t+T_0) - 2wT_0,$$

and so for $t \geq 2T_0$,

$$u(t,x) \geq \alpha - Ce^{cT_0}e^{-ct} \text{ for } |x| \leq L + wt.$$

The result follows as we may replace w with $2w$. $\square$

Recall the parameter $r \in (0,1)$, and definitions of a_ε, t_ε, Q_ε, and $D(x,\xi)$ in (1.34). We first show the density $D(x,\xi_T^\varepsilon)$ is close to its mean.

Lemma 4.1. *Let $T > 0$ and assume ξ_0^ε is deterministic.*
(a) If $0 < r < \frac{5}{24}$, then for all $x \in a_\varepsilon\mathbb{Z}^d$,

$$E((D(x,\xi_T^\varepsilon) - E(D(x,\xi_T^\varepsilon)))^2) \leq C_{4.1}e^{c_bT}\varepsilon^{1/12}.$$

(b) If $0 < r \leq 1/(16d)$ and $C = y + [-L,L]^d$ for $y \in \mathbb{R}^d$, then for all $\eta > 0$,

$$P\Big(\sup_{x\in C\cap a_\varepsilon\mathbb{Z}^d} |D(x,\xi_T^\varepsilon) - E(D(x,\xi_T^\varepsilon))| \geq \eta\Big) \leq C_{4.1}\varepsilon^{1/48}L^de^{c_bT}\eta^{-2}.$$

Proof. (a) Note that

$$\begin{aligned} &|\{(z_1,z_2) \in (x+Q_\varepsilon)^2 : |z_1 - z_2| \leq \varepsilon^{1/4}\}| \\ &\quad\leq (2\varepsilon^{-3/4}+1)^d|Q_\varepsilon| \leq c_d|Q_\varepsilon|^2(\varepsilon^{\frac{1}{4}-r})^d. \end{aligned} \tag{4.2}$$

If Σ_z^x denotes the sum over

$$z \in \{(z_1,z_2) \in (x+Q_\varepsilon)^2 : |z_1 - z_2| > \varepsilon^{1/4}\}, \tag{4.3}$$

then by (4.2) and Proposition 3.4 with $K = 2$,

$$\begin{aligned} &E((D(x,\xi_T^\varepsilon) - E(D(x,\xi_T^\varepsilon)))^2) \\ &\quad\leq |Q_\varepsilon|^{-2}\Big[c_d|Q_\varepsilon|^2(\varepsilon^{\frac{1}{4}-r})^d + \textstyle\sum_z^x[E(\displaystyle\prod_{k=1}^2\xi_T^\varepsilon(z_k)) - \prod_{k=1}^2 E(\xi_T^\varepsilon(z_k))]\Big] \\ &\quad\leq c_d\varepsilon^{(\frac{1}{4}-r)d} + 4c_{3.4}(2)e^{c_bT}\varepsilon^{1/12} \\ &\quad\leq C_{4.1}e^{c_bT}\varepsilon^{1/12}, \end{aligned} \tag{4.4}$$

where our condition on r is used in the last line.
(b) Note that

$$|C \cap a_\varepsilon\mathbb{Z}^d| \leq c_dL^da_\varepsilon^{-d} \leq c_dL^d\varepsilon^{-rd} \leq c_dL^d\varepsilon^{-1/16}.$$

The result now follows from (a) and Chebychev's inequality. $\square$

We recall the following hypothesis from Section 1.2:

Assumption 2. There are constants $0 < u_1 < 1$, $c_2, C_2, w > 0$, $L_0 \geq 3$ so that for all $L \geq L_0$, if $u(0,x) \leq u_1$ for $|x| \leq L$ then for all $t \geq 0$

$$u(t,x) \leq C_2e^{-c_2t} \text{ for all } |x| \leq L + 2wt.$$

We also recall the following condition from the same Section: For some $r_0 > 0$,

$$\sum_{i=0}^{1} \|g_i^\varepsilon - g_i\|_\infty \le c_{1.41}\varepsilon^{r_0}. \tag{4.5}$$

We say that $\xi \in \{0,1\}^{\varepsilon\mathbb{Z}^d}$ has density at most κ (respectively, in $[\kappa_1, \kappa_2]$) on $A \subset \mathbb{R}^d$ iff $D(x,\xi) \le \kappa$ (respectively $D(x,\xi) \in [\kappa_1,\kappa_2]$) for all $x \in (a_\varepsilon\mathbb{Z}^d) \cap A$. We set (recall (3.7))

$$r = \frac{1}{16d}, \text{ hence } \beta = \frac{1.9}{16d}, \; t_\varepsilon = \varepsilon^{1.9/(16d)}, T = A_{4.2}\log(1/\varepsilon), \text{ and } S = T - t_\varepsilon, \tag{4.6}$$

where $A_{4.2} = c_b^{-1}\Big(\frac{1}{100d} \wedge \frac{r_0}{4}\Big)$.

Lemma 4.2. *Suppose Assumption 2 and (4.5) hold. Let $u_2 \in (0, u_1)$ and $\gamma_{4.2} = \Big(\frac{c_2}{c_b} \wedge 1\Big)\Big(\frac{1}{120d} \wedge \frac{r_0}{5}\Big)$. There is an $\varepsilon_{4.2} > 0$, depending on (u_1, u_2, w, c_2, C_2) and satisfying*

$$\varepsilon_{4.2}^{\gamma_{4.2}} \le u_2, \tag{4.7}$$

so that if $0 < \varepsilon \le \varepsilon_{4.2}$ and $2 + L_0 \le L \le \varepsilon^{-.001/d}$, then whenever ξ_0^ε has density at most u_2 in $[-L, L]^d$,

$$P(\xi_T^\varepsilon \text{ has density at most } \varepsilon^{\gamma_{4.2}} \text{ in } [-L - wT, L + wT]^d | \xi_0^\varepsilon) \ge 1 - \varepsilon^{.01}.$$

Note that (4.7) allows us to iterate this result and obtain the conclusion on successively larger spatial regions at multiples of T.

The proof of the above Lemma will require some preliminary lemmas.

Lemma 4.3. *If $p_t^\varepsilon(y) = \varepsilon^{-d}P(B_t^\varepsilon = y)$, $y \in \varepsilon\mathbb{Z}^d$, then for $0 < \varepsilon \le 1$,*

$$|p_t^\varepsilon(x) - p_t^\varepsilon(x+y)| \le c_{4.3}|y|t^{-(d+1)/2} \text{ for all } x, y \in \varepsilon\mathbb{Z}^d \text{ and } t > 0.$$

Proof. This is a standard local central limit theorem; for $d = 2$ this is Lemma 2.1 of [5] and the same proof applies in higher dimensions. □

Recall (from (3.15)) that P_t^ε is the semigroup associated with B^ε.

Lemma 4.4. *There is a $c_{4.4}$ such that if $1 > \alpha > \beta/2$, then for $0 < \varepsilon \le 1$,*

$$|P_{t_\varepsilon}^\varepsilon \xi(x) - P_{t_\varepsilon}^\varepsilon \xi(x')| \le c_{4.4}\varepsilon^{(2\alpha-\beta)/(2+d)}$$

for all $x, x' \in \varepsilon\mathbb{Z}^d$ such that $|x - x'| \le 2\varepsilon^\alpha$ and all $\xi \in \{0,1\}^{\varepsilon\mathbb{Z}^d}$.

Proof. Let $-\infty < \delta \le \alpha$, $\Delta = x - x'$ and assume $|\Delta| \le 2\varepsilon^{\alpha}$. Apply Lemma 4.3 to see that

$$\begin{aligned}
&|P^{\varepsilon}_{t_\varepsilon}\xi(x) - P^{\varepsilon}_{t_\varepsilon}\xi(x')| \\
&\le \sum_{z\in\varepsilon\mathbb{Z}^d} |P(B^{\varepsilon}_{t_\varepsilon} = z) - P(B^{\varepsilon}_{t_\varepsilon} = z + \Delta)| \\
&\le \sum_{|z|\le 3\varepsilon^{\delta}} c_{4.3}\varepsilon^{d}|\Delta| t_\varepsilon^{-(d+1)/2} + P(|B^{\varepsilon}_{t_\varepsilon}| > 3\varepsilon^{\delta}) + P(|B^{\varepsilon}_{t_\varepsilon}| \ge 3\varepsilon^{\delta} - \Delta) \\
&\le c\varepsilon^{(\delta-1)d}\varepsilon^{d+\alpha}\varepsilon^{-\beta(d+1)/2} + 2P(|B^{\varepsilon}_{t_\varepsilon}| > \varepsilon^{\delta}).
\end{aligned}$$

If we use Chebychev to bound the last summand by $ct_\varepsilon\varepsilon^{-2\delta} = c\varepsilon^{\beta-2\delta}$ and optimize over δ (setting $\delta = \frac{\beta}{2} - \frac{\alpha-(\beta/2)}{2+d} < \frac{\beta}{2} < \alpha$), we obtain the required upper bound. □

Lemma 4.5. *For any $\eta > 0$ there is an $\varepsilon_{4.5}(\eta) > 0$ so that if $0 < \varepsilon \le \varepsilon_{4.5}$, $u \in [0,1]$, $L > 1$, and $\xi \in \{0,1\}^{\varepsilon\mathbb{Z}^d}$ has density at most u in $[-L, L)^d$, then*

$$P^{\varepsilon}_{t_\varepsilon}\xi(x) \le u + \eta \text{ for all } x \in [-L+1, L-1]^d \cap \varepsilon\mathbb{Z}^d\,. \tag{4.8}$$

Proof. By translation invariance it suffices to prove that for small enough $\varepsilon > 0$ and all $x \in [-a_\varepsilon, a_\varepsilon]^d \cap \varepsilon Z^d$, if ξ has density at most u in $[-1,1)^d$ then $P^{\varepsilon}_{t_\varepsilon}\xi(x) \le u + \eta$. (This addresses the uniformity in L.) Argue as in the upper bound in (3.13) to see that for $\varepsilon < \varepsilon_0(\eta)$,

$$\frac{|Q_\varepsilon| P(B_{t_\varepsilon} = z + e)}{P(B^{\varepsilon,x}_{t_\varepsilon} \in z + Q_\varepsilon)} \le 1 + \frac{\eta}{2} \text{ for all } z \in a_\varepsilon\mathbb{Z}^d,\ |z - x| \le 1 \text{ and } e \in Q_\varepsilon.$$

We therefore have

$$\begin{aligned}
P^{\varepsilon}_{t_\varepsilon}\xi(x) &\le P(|B^{x,\varepsilon}_{t_\varepsilon}| \ge 1/2) \\
&\quad + \sum_{z\in a_\varepsilon\mathbb{Z}^d} 1(|z-x| \le 3/4) \sum_{e\in Q_\varepsilon} \xi(z+e)\frac{1+(\eta/2)}{|Q_\varepsilon|} P(B^{\varepsilon,x}_{t_\varepsilon} \in z + Q_\varepsilon) \\
&\le 4\sigma^2 d t_\varepsilon + \sum_{z\in a_\varepsilon\mathbb{Z}^d} 1(|z-x| \le 3/4) u(1 + \frac{\eta}{2}) P(B^{\varepsilon,x}_{t_\varepsilon} \in z + Q_\varepsilon) \\
&\le 4\sigma^2 d t_\varepsilon + u + \frac{\eta}{2} \le u + \eta,
\end{aligned}$$

for $\varepsilon < \varepsilon_1(\eta)$. □

We are ready for the Proof of Lemma 4.2.

Proof. By the conditioning we may fix a deterministic ξ^{ε}_0 as in the statement of Lemma 4.2. In light of Lemma 4.1 our first and main goal is to bound $E(D(x, \xi^{\varepsilon}_T))$ for a fixed $x \in a_\varepsilon\mathbb{Z}^d \cup [-L - wT, L + wT]^d$. Let $z \in x + Q_\varepsilon$. Let $\tilde{X}^{z,T}$ be the modification of the dual $X^{z,T}$, starting with a single particle at z, in which particles ignore branching and coalescing events on $[S,T]$ by following their own random walk and switching to independent random walk mechanisms when a collision between particles occurs. Hence $X^i_S = \tilde{X}^i_S$ for all $i \in J(S)$, and on $[S,T]$ the particles in $\tilde{X}^{z,T}$ follow independent copies of B^ε. Let $\tilde{\zeta}^\varepsilon$ be the associated computation process, defined just as ζ^ε is for $X^{z,T}$, with initial values $\tilde{\zeta}^\varepsilon_0(j) = \xi^\varepsilon_0(\tilde{X}^j_T), j \in J(S)$ (for $\tilde{X}^{z,T}$ the index set is constant on $[S,T]$). On $\tilde{G}^\beta_T$, $T \notin \cup_{m=0}^{N_T-1}[R_m, R_m + 2\varepsilon^\beta]$, with $\beta < 1/2$, and so

$$[S,T] \cap (\cup_{m=0}^{N_T-1}[R_m, R_m + \sqrt{\varepsilon}]) = \emptyset. \tag{4.9}$$

Therefore on $\tilde{G}^\beta_T$, $X^{z,T}$ has no branching or coalescing events on $[S,T]$, and so $\tilde{X}^{z,T} = X^{z,T}$ on $[0,T]$. This also means (by (2.17)) that, given the common inputs $\zeta^\varepsilon_0(j) = \tilde{\zeta}^\varepsilon_0(j) = \xi^\varepsilon_0(X^j_T)$, $j \in J(T) = J(S)$ we have

$$\tilde{\zeta}^\varepsilon_T(0) = \zeta^\varepsilon_T(0) = \xi^\varepsilon_T(z) \text{ on } \tilde{G}^\beta_T. \tag{4.10}$$

Let $\psi_\varepsilon(x) = P^\varepsilon_{t_\varepsilon}\xi^\varepsilon_0(x)$. Conditional on $\mathcal{F}_S$, $\{\tilde{X}^j_T - \tilde{X}^j_S : j \in \hat{J}(S)\}$ are iid with law $P_0(B^\varepsilon_{t_\varepsilon} \in \cdot)$, and so, conditional on $\mathcal{F}_S$, $\{\tilde{\zeta}^\varepsilon_{T-S}(j) = \xi^\varepsilon_0(\tilde{X}^j_T) : j \in J(S)\}$ are independent Bernoulli rv's with means $\{\psi_\varepsilon(X^j_S) : j \in J(S)\}$. Recall $\{W_j\}$ is an iid sequence of uniform $[0,1]$ rv's independent of $\mathcal{F}_\infty$ (that is of our graphical construction). Let $\{\tilde{\zeta}^{\varepsilon,*}_t(j) : j \in J(T-t), T-S \le t \le T\}$ be the computation process associated with $\tilde{X}^{z,T}$ but with initial inputs $\tilde{\zeta}^{\varepsilon,*}_{T-S}(j) = 1(W_j \le \psi_\varepsilon(X^j_S) : j \in J(S)\}$. Then $\{\tilde{\zeta}^\varepsilon_t : T-S \le t \le T\}$ and $\{\tilde{\zeta}^{\varepsilon,*}_t : T-S \le t \le T\}$ have the same law because the joint law of their Bernoulli inputs and the processes $\tilde{X}^{z,T}_t, t \le S$ and $((\mu_m, U_m)1(R_m \le t), t \le S)$ used to define them are the same. Therefore by (4.10)

$$\begin{aligned}
&|P(\xi^\varepsilon_T(z) = 1) - P(\tilde{\zeta}^{\varepsilon,*}_T(0) = 1)| \\
&\quad = |P(\xi^\varepsilon_T(z) = 1) - P(\tilde{\zeta}^\varepsilon_T(0) = 1)| \le P((\tilde{G}^\beta_T)^c).
\end{aligned} \tag{4.11}$$

Consider now the branching random walk $\hat{X}$ starting with a single particle at z and coupled with $X^{z,T}$ as in Section 2.7, together with its computation process $\{\hat{\zeta}^\varepsilon : t \in [T-S,T]\}$ with initial inputs $\hat{\zeta}_{T-S}(j) = 1(W_j \le \psi_\varepsilon(\hat{X}^j_S)), j \in \hat{J}(S)$. Conditional on $\mathcal{F}_\infty$, these inputs are independent Bernoulli rv's with means $\{\psi(\hat{X}^j_S) : j \in \hat{J}(S)\}$. The computation processes $\tilde{\zeta}^{\varepsilon,*}$ and ζ^ε are identical on $[T-S,T]$ if given the same inputs at time $T-S$. Therefore Lemma 2.10 shows that on $\tilde{G}^\beta_T$ $\hat{\zeta}^\varepsilon_T(0)$ and $\tilde{\zeta}^{\varepsilon,*}_T(0)$ will coincide if given the same inputs at time $T-S$. Therefore

$$\begin{aligned}
&|P(\hat{\zeta}_T(0) = 1) - P(\tilde{\zeta}^{\varepsilon,*}_T(0) = 1)| \\
&\quad \le P((\tilde{G}^\beta_T)^c) + E(P(\hat{\zeta}^\varepsilon_{T-S}(j) \ne \tilde{\zeta}^{\varepsilon,*}_{T-s}(j) \ \exists j \in \hat{J}(S)|\mathcal{F}_\infty)1(\tilde{G}^\beta_T)) \\
&\quad \le P((\tilde{G}^\beta_T)^c) + E\Big(\sum_{j\in\hat{J}(S)} |\psi_\varepsilon(\hat{X}^j_S) - \psi_\varepsilon(X^j_S)|1(\sup_{j\in\hat{J}(S)} |X^j_S - \hat{X}^j_S| \le \varepsilon^{1/6})\Big).
\end{aligned} \tag{4.12}$$

Use Lemma 2.9 to bound the first term above and Lemma 4.4 with $\alpha = 1/6 > \beta/2$ to bound the second, and combine this with (4.11) to conclude that (use $d \ge 3$ and $\beta = 1.9/16d$) for small enough $\varepsilon > 0$,

$$\begin{aligned}
|P(\xi^\varepsilon_T(z) = 1) - P(\hat{\zeta}^\varepsilon_T(0) = 1)| &\le 2c_{2.9}e^{c_bT}\varepsilon^{\beta/3} + E(|\hat{J}(S)|)c_{4.4}\varepsilon^{(1/3-\beta)/(2+d)}) \\
&\le 2c_{2.9}\varepsilon^{1/(40d)} + e^{c_bS}c_{4.4}\varepsilon^{(1/3-\beta)/(2+d)} \\
&\le (2c_{2.9} + c_{4.4})\varepsilon^{1/(40d)}.
\end{aligned} \tag{4.13}$$

To prepare with the coupling with the branching Brownian motion we must extend $\psi_\varepsilon(x) = P^\varepsilon_{t_\varepsilon}\xi^\varepsilon_0(x)$ from $\varepsilon\mathbb{Z}^d$ to $\mathbb{R}^d$ in an appropriate manner. Since ξ^ε_0 has density at most u_2 in $[-L,L]^d$, Lemma 4.5 shows that for $\varepsilon \le \varepsilon_{4.5}(u_1-u_2)$ we may extend ψ_ε in a piecewise linear manner so that

$$\psi_\varepsilon(x) \le u_1 1(|x| \le L-2) + 1(|x| > L-2) \text{ for all } x \in \mathbb{R}^d. \tag{4.14}$$

In addition, using Lemma 4.4 with $\alpha = 1/6$, we may assume the above extension also satisfies

$$|\psi_\varepsilon(x) - \psi_\varepsilon(x')| \le c_{4.4}\varepsilon^{\frac{(1/3)-\beta}{2+d}} \text{ for } x, x' \in \mathbb{R}^d \text{ such that } |x-x'| \le \varepsilon^{1/6}. \tag{4.15}$$

Now consider the branching Brownian motion $\hat{X}^0$ starting with a single particle at z, coupled with $\hat{X}^\varepsilon$ as in Section 3. Consider also its associated computation process $\hat{\zeta}^0$ on $[T-S,T]$ starting with conditionally independent Bernoulli inputs $\{\hat{\zeta}^0_{T-S}(j) = 1(W_j \le \psi_\varepsilon(\hat{X}^{0,j}_S)) : j \in \hat{J}^0(S)\}$. We may argue as in the derivation of (4.13), but now using Lemma 2.13, (4.5), and (4.15) in place of Lemmas 2.9 and 4.4, to conclude after some arithmetic using $d \ge 3$,

$$|P(\hat{\zeta}^\varepsilon_T(0)=1) - P(\hat{\zeta}^0_T(0)=1)| \le e^{c_bT}c_{2.13}(\varepsilon^{3/8} + (c_{1.41}\varepsilon^{r_0})^{1/2}) + c_{4.4}\varepsilon^{1/(40d)}. \tag{4.16}$$

By Lemma 3.3 (we have shifted time by $T-S$), $P(\hat{\zeta}^0_T(0)=1) = u_\varepsilon(S,z)$, where u_ε is the solution of the PDE (1.33) with initial condition $u_\varepsilon(0,\cdot) = \psi_\varepsilon$. Now combine this with (4.13) and (4.16) to see that for small enough ε as above

$$E(\xi^\varepsilon_T(z)) \le c\varepsilon^{(1/(40d))\wedge(r_0/4)} + u_\varepsilon(S,z). \tag{4.17}$$

Now use the bound on the initial condition (4.14) and Assumption 2 in the above to conclude that for $|z| \le L-2+2wS$, and small ε

$$E(\xi^\varepsilon_T(z)) \le c\varepsilon^{(1/(40d))\wedge(r_0/4)} + C_2e^{-c_2S} \le \varepsilon^{\gamma'}, \tag{4.18}$$

where $\gamma' = \left(\frac{c_2}{c_b} \wedge 1\right)\left(\frac{1}{110d} \wedge \frac{9r_0}{40}\right) > \gamma_{4.2}$ and we used (4.6) and some arithmetic. By taking ε smaller if necessary we may assume $2wS - 3 \ge wT$ and so the above holds for $|z| \le L+1+wT$. This shows that

$$E(D(x,\xi^\varepsilon_T)) \le \varepsilon^{\gamma'} \text{ for } x \in a_\varepsilon\mathbb{Z}^d \cap [-L-wT, L+wT]^d. \tag{4.19}$$

Finally apply the above and Lemma 4.1 to conclude that for small enough ε

$$\begin{aligned}
&P\Big(\sup_{x\in[-L-wT,L+wT]^d\cap a_\varepsilon\mathbb{Z}^d} D(x,\xi^\varepsilon_T) \ge \varepsilon^{\gamma_{4.2}}\Big)\\
&\le P(\sup_{x\in[-L-wT,L+wT]^d\cap a_\varepsilon\mathbb{Z}^d} |D(x,\xi^\varepsilon_T) - E(D(x,\xi^\varepsilon_T))| \ge \varepsilon^{\gamma_{4.2}}/2)\\
&\le C_{4.1}(L+wT)^d e^{c_bT}\varepsilon^{1/48}4\varepsilon^{-2\gamma_{4.2}}\\
&\le C\varepsilon^{-.001}\varepsilon^{-1/100d}\varepsilon^{1/48}\varepsilon^{-1/60d} \le \varepsilon^{.01}.
\end{aligned}$$

□

Our next goal is to show that the dual process only expands linearly in time. The first ingredient is a large deviations result. Recall the dominating branching random walk $\{\bar{X}^{\varepsilon,j}(t) : j \in \bar{J}^\varepsilon(t)\}$ introduced at the beginning of Section 2.8 which satisfies

$$\{X_s^{\varepsilon,j} : j \in J^\varepsilon(s)\} \subset \{\bar{X}_s^{\varepsilon,j} : j \in \bar{J}^\varepsilon(s)\}.$$

If $\|X_s^\varepsilon\|_\infty = \sup\{|X^{\varepsilon,j}(s)| : j \in J^\varepsilon(s)\}$ and similarly for $\|\bar{X}_s^\varepsilon\|_\infty$, then the above domination implies

$$\|\bar{X}_s^\varepsilon\|_\infty \geq \|X_s^\varepsilon\|_\infty \text{ for all } s, \varepsilon. \tag{4.20}$$

Recall c^* is as in (2.7).

Lemma 4.6. *Assume $\bar{X}^\varepsilon$ starts from one particle at 0. For each $R > 0$ there is an $\varepsilon_{4.6}(c^*, R) > 0$, nonincreasing in each variable, so that for $0 < \varepsilon \leq \varepsilon_{4.6}$ and $t > 0$,*

$$P(\|\bar{X}_s^\varepsilon\|_\infty \geq 2\rho t \text{ for some } s \leq t) \leq (4d+1)\exp(-t(\gamma(\rho) - c_b)) \quad \text{for all } 0 < \rho \leq R,$$

where $\gamma(\rho) = \min\{\rho/2, \rho^2/3\sigma^2\}$. Moreover, if $\rho \geq \max\{4c_b, 2\sigma^2\}$, then the above bound is at most $(4d+1)\exp(-t\rho/4)$.

Proof. The last assertion is trivial. Let S_t^ε be a random walk that starts at 0, jumps according to p_ε at rate ε^{-2}, and according to q_ε at rate c^*. Since $E|\bar{X}_t^\varepsilon| = \exp(c_b t)$ by summing over the branches of the tree, it suffices to show

$$P(\|S_s^\varepsilon\|_\infty \geq \rho t \text{ for some } s \leq t) \leq (4d+1)\exp\left(-\gamma(\rho)t\right). \tag{4.21}$$

As usual, B_t^ε is the random walk that jumps according to p_ε at rate ε^{-2} and $B_t^{i,\varepsilon}$ is its ith coordinate. By the reflection principle

$$P\left(\sup_{s\leq t} B_s^{\varepsilon,i} \geq \rho t\right) \leq 2P\left(B_t^{\varepsilon,i} \geq \rho t\right) \leq 2e^{-\theta\rho t} E\left(\exp\left(\theta B_t^{\varepsilon,i}\right)\right)$$

for any $\theta > 0$. If $\phi_\varepsilon(\theta) = \sum_x e^{\theta x^i} p_\varepsilon(x)$ then a standard Poisson calculation gives

$$E\left(\exp\left(\theta B_t^{\varepsilon,i}\right)\right) = \exp(t\varepsilon^{-2}(\phi_\varepsilon(\theta) - 1)).$$

By scaling $\phi_\varepsilon(\theta) = \phi_1(\varepsilon\theta)$. Our assumptions imply $\phi_1'(0) = 0$ and $\phi_1''(0) = \sigma^2$ so

$$\varepsilon^{-2}(\phi_1(\varepsilon\theta) - 1) \to \sigma^2\theta^2/2 \text{ as } \varepsilon \to 0.$$

If $0 < \rho \leq R$ and $\theta = \rho/\sigma^2$ in the above, it follows that for $\varepsilon < \varepsilon_0(R)$,

$$e^{-(\rho^2/\sigma^2)t} E\left(\exp\left((\rho/\sigma^2)B_t^{\varepsilon,i}\right)\right) \leq \exp(-\rho^2 t/3\sigma^2),$$

and so,

$$P(\sup_{s\leq t} |B_s^{\varepsilon,i}| \geq \rho t) \leq 4\exp\left(-\rho^2 t/3\sigma^2\right). \tag{4.22}$$

Let J_t^ε be the one dimensional random walk that jumps according to the law of $\varepsilon Y^* = \varepsilon \max_{i\leq N_0} |Y^i|$ at rate c^*, and notice that this will bound the L^∞ norm of the sum of the absolute values of the jumps according to q_ε in S_ε up to time t. If we let $\phi_J(\theta) = E(\exp(\theta Y^*))$, then, arguing as above, we obtain

$$P(J_t^\varepsilon \geq \rho t) \leq \exp(-\rho\theta t + c^* t(\phi_J(\varepsilon\theta) - 1)).$$

The exponential tail of Y^* (from (1.6)) shows that $(\phi_J(\varepsilon\theta)-1)/\varepsilon\theta \to EY^*$ as $\varepsilon \to 0$, and so, if we set $\theta = 1$, then for small ε, $c^*(\phi_J(\varepsilon) - 1) \leq \rho/2$. (The choice of ε here works for all ρ because we may assume without loss of generality that $\rho \geq \rho_0 > 0$ as the Lemma is trivial for small ρ.) Therefore

$$P(J_t^\varepsilon \geq \rho t) \leq \exp(-\rho t/2). \tag{4.23}$$

To derive (4.21), write

$$P(\sup_{s\leq t} \|S_s^\varepsilon\|_\infty \geq 2\rho t) \leq \sum_{i=1}^d P(\sup_{s\leq t} |B_s^{\varepsilon,i}| \geq \rho t) + P(J_t^\varepsilon \geq \rho t),$$

and use (4.22) and (4.23). □

Our next result uses the large deviation bound in Lemma 4.6 to control the movement of all the duals that start in a region. Recall that $X^{x,U}$ is the dual for ξ^ε starting with one particle at x from time U. For $x \in \mathbb{R}^d$ and $r > 0$ let $Q(x,r) = [x-r, x+r]^d$ and $Q^\varepsilon(x,r) = Q(x,r) \cap \varepsilon\mathbb{Z}^d$. Write $Q(r)$ for $Q(0,r)$ and $Q^\varepsilon(r)$ for $Q^\varepsilon(0,r)$.

Lemma 4.7. *For $c > 0$, $b \geq 4c_b \vee 2\sigma^2$, $L \geq 1$ and $U \geq T' = c\log(1/\varepsilon)$, let*

$$\bar{p}_\varepsilon(b,c,L,U) = P(X_t^{x,u} \text{ is not contained in } Q(L+2bT') \\ \text{for some } u \in [U-T', U], t \leq T' \text{ and some } x \in Q^\varepsilon(L)).$$

Let $c'_d = 12(4d+1)3^d$. There exists $\varepsilon_{4.6}(c^*, b) > 0$ such that if $0 < \varepsilon \leq \varepsilon_{4.6}$

$$\bar{p}_\varepsilon(b,c,L,U) \leq c'_d L^d (c\log(1/\varepsilon) + 1)\varepsilon^{q-d}$$

where $q = (\frac{bc}{4} - 2) \wedge c\varepsilon^{-2}$.

Proof. By translation invariance it suffices to take $U = T'$. For $x \in \varepsilon\mathbb{Z}^d$ let $\{T_i(x) : i \geq 0\}$ be the successive jump times of the reversed Poisson process, starting at time T', determined by the $T_n^x, T_n^{*,x}$. Also let N_x be the number of such jumps up to time T', so that N_x is Poisson with mean $(c^* + \varepsilon^{-2})T'$. The process $\xi_t^\varepsilon(x)$ is constant for $t \in (T' - T_{i+1}(x), T' - T_i(x)]$ and for such t the dual $X^{x,t}(v)$ is

$$X^{x,T'-T_i(x)}(v + (T' - T_i(x) - t)),$$

that is, one is a simple translation of the other. This means for t as above

$$\cup_{v\leq t} X^{x,t}(v) \subset \cup_{v \leq T'-T_i(x)} X^{x,T'-T_i(x)}(v), \tag{4.24}$$

(in fact equality clearly holds). As a result, in $\bar{p}_\varepsilon(b,c,L,T')$ we only need consider t to be one of the times $T'-T_i(x)$ for $0\le i\le N_x$ and we may bound $1-\bar{p}_\varepsilon(b,c,L,T')$ by

$$\begin{aligned}
&P(\exists x\in Q^\varepsilon(L)\text{ s.t. } N_x\ge 3T'(\varepsilon^{-2}+c^*))\\
&\qquad + P(\exists x\in Q^\varepsilon(L),\ 0\le T_i(x)\le 3T'(\varepsilon^{-2}+c^*)\text{ s.t.}\\
&\qquad\qquad \sup_{v\le T'-T_i(x)}\|X^{x,T'-T_i(x)}(v)\|_\infty>2bT')\\
&\le (2L\varepsilon^{-1}+1)^d\exp\{-3T'(\varepsilon^{-2}+c^*)\}E(e^{N_x})\\
&\qquad+(2L\varepsilon^{-1}+1)^d(3T'(\varepsilon^{-2}+c^*)+1)(4d+1)\exp(-T'b/4).
\end{aligned}$$

Here we are using Lemma 4.6 and the strong Markov property at $T_i(x)$ for the filtration generated by the reversed Poisson processes $\mathcal{F}_t$. Some arithmetic shows the above is at most

$$\begin{aligned}
&3^d(L\vee\varepsilon)^d\varepsilon^{-d}\Big[\exp(-3T'(\varepsilon^{-2}+c^*))\exp((\varepsilon^{-2}+c^*)T'(e-1))\\
&\qquad\qquad+(4d+1)(3T'(\varepsilon^{-2}+c^*)+1)\varepsilon^{bc/4}\Big]\\
&\le 3^d(L\vee\varepsilon)^d\varepsilon^{-d}\Big[\exp(-T'(\varepsilon^{-2}+c^*))+6(4d+1)(c\log(1/\varepsilon)\varepsilon^{-2}+1)\varepsilon^{bc/4}\Big]\\
&\le 3^d(L\vee\varepsilon)^d\varepsilon^{-d}\Big[\varepsilon^{(c\varepsilon^{-2})}+6(4d+1)(c\log(1/\varepsilon)+1)\varepsilon^{bc/4-2}\Big]\\
&\le c_d'(L\vee\varepsilon)^d(c\log(1/\varepsilon)+1)\varepsilon^{-d}\varepsilon^{(bc/4-2)\wedge c\varepsilon^{-2}}
\end{aligned}$$

$\square$

5 Percolation results

To prove Theorems 1.4 and (especially) 1.5 we will use block arguments that involve comparison with oriented percolation. Let $D = d+1$, where for now we allow $d \geq 1$, and let $\mathcal{A}$ be any $D \times D$ matrix so that (i) if x has $x_1 + \cdots + x_D = 1$ then $(\mathcal{A}x)_D = 1$, and (ii) if x and y are orthogonal then so are $\mathcal{A}x$ and $\mathcal{A}y$. Geometrically, we first rotate space to take $(1/D, \ldots 1/D)$ to $(0, \ldots, 0, 1/\sqrt{D})$ and then scale $x \to x\sqrt{D}$. Let $\mathcal{L}_D = \{\mathcal{A}x : x \in \mathbb{Z}^D\}$. The reason for this choice of lattice is that if we let $\mathcal{Q} = \{\mathcal{A}x : x \in [-1/2, 1/2]^D\}$, then the collection $\{z + \mathcal{Q}, z \in \mathcal{L}_D\}$ is a tiling of space by rotated cubes. When $d = 1$, $\mathcal{L}_2 = \{(m,n) : m+n \text{ is even}\}$ is the usual lattice for block constructions (see Chapter 4 of [14]).

Let $\mathcal{H}_k = \{z \in \mathcal{L}_D : z_D = k\} = \{\mathcal{A}x : x \in \mathbb{Z}^D, \sum_i x_i = k\}$ be the points on "level" k. We will often write the elements of $\mathcal{H}_k$ in the form (z,k) where $z \in \mathbb{R}^d$. Let $\mathcal{H}'_k = \{z \in \mathbb{R}^d : (z,k) \in \mathcal{H}_k\}$. When $d = 2$, the points in $\mathcal{H}'_0$ are the vertices of a triangulation of the plane using equilateral triangles, and the points in $\mathcal{H}'_1$ are obtained by translation. One choice of $\mathcal{A}$ leads to Figure 7, where $\mathcal{H}'_1$ and $\mathcal{H}'_2$ are obtained by translating $\mathcal{H}'_0$ upward by $\sqrt{2}$ and $2\sqrt{2}$, respectively, and $\mathcal{H}'_3 = \mathcal{H}'_0$.

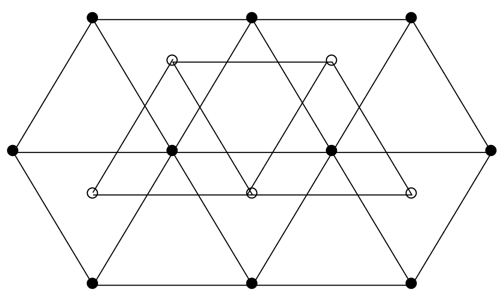


Figure 7: $\mathcal{H}'_0$ (black dots) and $\mathcal{H}'_1$ (white dots) in $\mathcal{L}_3$

In $d \geq 3$ dimensions (the case we will need for our applications in this work) the lattice is hard to visualize so we will rely on arithmetic. Let $\{e_1, \ldots, e_D\}$ be the standard basis in $\mathbb{R}^D$, and put $v_i = \mathcal{A}e_i$, $i = 1, \ldots, D$. By the geometric description of $\mathcal{A}$ given above, $v_i \in \mathcal{H}_1$ has length $\sqrt{D}$, and writing $v_i = (v'_i, 1)$, $v'_i \in \mathbb{R}^d$ has length $\sqrt{D-1}$. For $i \neq j$, $\|v'_i - v'_j\|_2 = \|v_i - v_j\|_2 = \sqrt{2D}$, the last by orthogonality of v_i and v_j. The definitions easily imply that $\mathcal{H}'_{k+1} = v'_i + \mathcal{H}'_k \equiv \{v'_i + x : x \in \mathcal{H}'_k\}$ for each i and k. Note that $Dv'_i \in \mathcal{H}'_0$ because $Dv_i - (0, \ldots, 0, D) \in \mathcal{H}_0$. This implies that $\mathcal{H}'_{k+D} = Dv'_i + \mathcal{H}'_k = \mathcal{H}'_k$.

For $x \in \mathcal{H}'_k$ let $\mathcal{V}_x \subset \mathbb{R}^d$ be the Voronoi region for x associated with the points in $\mathcal{H}'_k$, i.e., the closed set of points in $\mathbb{R}^d$ that are closer to x in Euclidean norm than to all the other points of $\mathcal{H}'_k$ (including ties). If $\mathcal{V} = \mathcal{V}_0$ (in $d = 2$, $\mathcal{V}_0$ is the hexagon in Figure 8 inside the connected six white dots), then the translation invariance of $\mathcal{H}'_0$ and fact that $\mathcal{H}'_k = kv'_i + \mathcal{H}'_0$ show that $\mathcal{V}_x = x + \mathcal{V}$ for all $x \in \cup\mathcal{H}'_k$. It is immediate

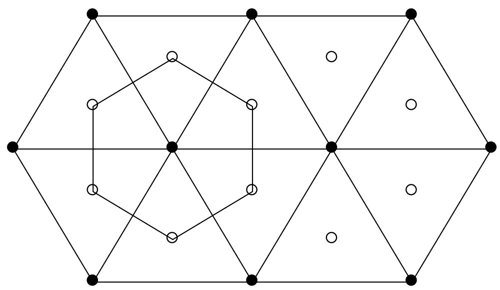


Figure 8: $\mathcal{H}'_0$ (black dots) and Voronoi region about 0 (inside white dots) in $\mathcal{L}_3$

from the definition of Voronoi region that for each k,

$$\cup_{x\in\mathcal{H}'_k}\mathcal{V}_x = \mathbb{R}^d. \tag{5.1}$$

Furthermore, $\mathcal{V}_x$ is contained in the closed ball of radius D centered at x. (To see this we may set $x = k = 0$ and transfer the problem to $\mathbb{Z}^D$ via $\mathcal{A}^{-1}$. It then amounts to noting that if $x \in \mathbb{R}^D$ satisfies $\sum x_i = 0$ and $\|x\|_2 > \sqrt{D}$, then there are $i \neq j$ s.t. $x_i > 1, x_j < 0$ or $x_i < -1, x_j > 0$, and so $\|x \pm (e_i - e_j)\|_2 < \|x\|_2$.) From this inclusion we see that for any $L > 0$,

$$\begin{gathered}\text{if } c_L = L/(2D) \text{ then } c_L\mathcal{V}_x \subset c_L x + [-L, L]^d, \\ \text{and so } \cup_{x\in\mathcal{H}'_k} c_L x + [-L, L]^d = \mathbb{R}^d.\end{gathered} \tag{5.2}$$

The above also holds with $2c_L$ in place of c_L but the above ensures a certain overlap in the union which makes it more robust. Finally, one can check that for some positive $c_{5.3}(D)$,

$$\text{if } a \in \mathcal{V}_x, b \notin \mathcal{V}_x \text{ and } |a - b| < c_{5.3} \text{ then } b \in \cup_{i\neq j}\mathcal{V}_{x+v'_i-v'_j}. \tag{5.3}$$

For this, note that $x + v'_i - v'_j$, $1 \le i \neq j \le D$ are the $D(D-1)$ "neighboring points to x" in $\mathcal{H}'_k$, corresponding to the 6 black vertices of the hexagonal around 0 in Figure 8 for $x = 0$ and $D = 3$. The above states that the $D(D-1)$ corresponding Voronoi regions provide a solid annulus about $\mathcal{V}_x$, as is obvious from Figure 8 for $D = 3$.

Our oriented percolation process will be constructed from a family of random variables $\{\eta(z), z \in \mathcal{L}_D\}$ taking values 0 or 1, where 0 means closed and 1 means open. In the block construction, one usually assumes that the collection of $\eta(z)$ is "M dependent with density at least $1-\theta$" which means that for any k,

$$\begin{gathered}P(\eta(z_i) = 1|\eta(z_j), j \neq i) \ge (1-\theta), \\ \text{whenever } z_i \in \mathcal{L}_D, 1 \le i \le k \text{ satisfy } |z_i - z_j| > M \text{ for all } i \neq j.\end{gathered} \tag{5.4}$$

Our process will satisfy the modified condition

$$P(\eta(z_k)=1|\eta(z_j),j<k)\geq(1-\theta) \text{ whenever } z_j=(z'_j,n_j)\in\mathcal{L}_D,\ 1\leq j\leq k \quad (5.5)$$
$$\text{satisfy } n_j<n_k \text{ or } (n_j=n_k \text{ and } |z'_j-z'_k|>M) \text{ for all } j<k.$$

It is typically not difficult to prove results for M-dependent percolation processes with θ small (see Chapter 4 of [14]), but in Section 7 we will simplify things by applying Theorem 1.3 of [37] to reduce to the case of independent percolation. By that result, under (5.4), there is a constant Δ depending on D and M such that if

$$1-\theta'=\Big(1-\frac{\theta^{1/\Delta}}{(\Delta-1)^{(\Delta-1)/\Delta}}\Big)\big(1-(\theta(\Delta-1))^{1/\Delta}\big), \tag{5.6}$$

we may couple $\{\eta(z),z\in\mathcal{L}_D\}$ with a family $\{\zeta(z),z\in\mathcal{L}_D\}$ of iid Bernoulli random variables with $P(\zeta(z)=1)=1-\theta'$ such that $\zeta(z)\leq\eta(z)$ for all $z\in\mathcal{L}_D$.

We will now show the above conclusion remains valid in our setting with the modified condition (5.5) in place of (5.4).

Lemma 5.1. *Assume $\{\eta(z):x\in\mathcal{L}_D\}$ is a collection of $\{0,1\}$-valued random variables satisfying (5.5). There is a constant Δ depending on D and M such that if θ' is given by (5.6), then we may couple $\{\eta(z),z\in\mathcal{L}_D\}$ with a family $\{\zeta(z),z\in\mathcal{L}_D\}$ of iid Bernoulli random variables with $P(\zeta(z)=1)=1-\theta'$ such that $\zeta(z)\leq\eta(z)$ for all $z\in\mathcal{L}_D$.*

Proof. The result is proved by modifying the proof of Theorem 1.3 of [37]. In that argument we will set $\Delta=1+\#\{z'\in\mathcal{H}'_0;|z|\leq M\}$ and G will be the graph with vertex set $\mathcal{L}_D$ and edges drawn between z,z' in some $\mathcal{H}'_k$ iff $|z-z'|\leq M$. There are no edges between distinct "levels" in $\mathcal{L}_D$. In the proof of Theorem 1.3 of [37] a finite subset $S=\{z_1,\dots,z_k\}$ is totally ordered ($z_1\leq z_2\leq\dots z_k$) so that each z_j is adjacent in G to at most $\Delta-1$ of its predecessors. With our choice of Δ we may also order S so that $z_i=(z'_i,n_i)$ where $n_1\leq n_2\leq\dots\leq n_k$. The proof of Theorem 1.3 in [37] then goes through providing the conclusion of Proposition 1.2 holds where $\{\eta(z_{s_0})=1\}$ is now conditioned on the corresponding values at vertices $z_{s_1}=(z'_{s_1},n_1),\dots,z_{s_j})=(z'_{s_j},n_j)$ where $\max(n_1,\dots,n_j)\leq n_0$ and $z_{s_0}=(z'_{s_0},n_0)$. It is now easy to modify the inductive proof of Proposiiton 1.2 so that one is always conditioning on values at lower or equal levels n_i. In particular (5.5) suffices to derive the lower bound on the numerator of (1.6) in [37]. ◻

Remark The above argument clearly applies to more general "oriented" graphs than $\mathcal{L}_D$.

In view of the comparison, and the fact that $\theta'\to0$ as $\theta\to0$, we can for the rest of the section suppose:

$$\eta(z) \text{ are i.i.d. with } P(\eta(z)=1)=1-\theta. \tag{5.7}$$

We now define the edge set $\mathcal{E}_\uparrow$ for $\mathcal{L}_D$ to be the set of all oriented edges from z to $z+v_i$, $z\in\mathcal{L}_D$, $1\leq i\leq D$. A sequence of points $z_0,\dots,z_{n-1},z_n$ in $\mathcal{L}_D$ is called an open path from z_0 to z_n, and we write $z_0 \to z_n$, if there is an edge in $\mathcal{E}_\uparrow$ from z_i to z_{i+1} and z_i is open for $i = 0, \dots, n-1$. Note that z_n does not have to be open if $n \ge 1$ but z_0 does. In Sections 6 and 7 we will employ a block construction and determine suitable parameters so that $(x, n) \in \mathcal{H}_n$ being open will correspond to a certain "good event" occurring for our Poisson processes in the space-time block $(c_L x + [-K_1 T, K_1 T]^d) \times [nJ_1 T, (n+1)J_1 T]$ for appropriate L, K_1 and J_1.

Given an initial set of "wet" sites $W_0 \subset \mathcal{H}_0$, we say $z \in \mathcal{H}_n$ is wet if $z_0 \to z$ for some initial wet site z_0. Let $\bar{W}_n$ be the set of wet sites in $\mathcal{H}_n$ when all the sites in $\mathcal{H}_0$ are wet, and let W_n^0 be the set of wet sites in $\mathcal{H}_n$ when only $0 \in \mathcal{H}_0$ is wet. Let $\Omega_\infty^0 = \{W_n^0 \neq \emptyset \text{ for all } n \ge 0\}$.

Lemma 5.2. *(i)* $\inf_{x \in \mathcal{H}_n} P(x \in \bar{W}_n) \ge P(\Omega_\infty^0) \to 1$ *as* $\theta \to 0$.
(ii) Let $\mathcal{H}_n^r = \{(z, n) \in \mathcal{L}_D : z \in [-r, r]^d\}$. *There are* $\theta_{5.2} > 0$ *and* $r_{5.2} > 0$ *such that if* $\theta < \theta_{5.2}$ *and* $r \le r_{5.2}$, *then as* $N \to \infty$,

$$P(\Omega_\infty^0 \text{ and } W_n^0 \cap \mathcal{H}_n^{rn} \neq \bar{W}_n \cap \mathcal{H}_n^{rn} \text{ for some } n \ge N) \to 0. \tag{5.8}$$

Proof. The first result follows from well-known $d = 1$ results, e.g., see Theorem 4.1 of [12]. The second result is weaker than a "shape theorem" for W_n^0, which would say the following, using the notation $A' = \{x' : (x', n) \in A\}$ for $A \subset \mathcal{H}_n$. For $\theta < \theta_c$ there is a convex set $\mathcal{D} \subset \mathbb{R}^d$, containing the origin in its interior, so that on Ω_∞^0,

$$(W_n^0)' \approx n\mathcal{D} \cap (\bar{W}_n)'$$

for all large n. More precisely with probability 1, if $\delta > 0$ there is a random n_δ such that $(W_n^0)' \subset n(1+\delta)\mathcal{D}$ and $(W_n^0)' \supset (1-\delta)n\mathcal{D} \cap (\bar{W}_n)'$ for all $n \ge n_\delta$. The technology exists to prove such a result for oriented percolation on $\mathcal{L}_D$, but unfortunately no one has written down the details. The argument is routine but messy, so we content ourselves to remark that (ii) can be established by the methods used in Durrett and Griffeath [17] to prove the shape theorem for the d-dimensional contact process with large birth rates: one uses percolation in two dimensional subspaces $A(me_i + ne_j)$, $1 \le i < j \le n$ and self-duality. □

Call sites in $\bar{V}_n = \mathcal{H}_n \setminus \bar{W}_n$ dry. In Section 7, when we are trying to show that ξ_t^ε dies out, the block construction will imply for appropriate L and J_1,

$$\text{if } (z, n) \in W_n^0, \text{ then } \Big(c_L z + [-L, L]^d\Big) \times [(n-1)J_1 T, nJ_1 T] \text{ is } \varepsilon\text{-empty}, \tag{5.9}$$

where a region is ε-empty if $\xi_t^\varepsilon(x) = 0$ for all (x, t) in the region. This will not be good enough for our purposes because the space-time regions associated with points in $V_n^0 = \mathcal{H}_n \setminus W_n^0$ might be occupied by particles. To identify the locations where there might be 1's in ξ_t we will work backwards in time. However in our coarser grid $\mathcal{L}_D$, 1's may spread sideways through several dry regions and so we need to introduce an additional set of edges for $\mathcal{L}_D$. Let $\mathcal{E}_\downarrow$ consist of the set of oriented edges from z to $z - v_i$ for $1 \le i \le D$, and from z to $z + v_i - v_j$ for $1 \le i \ne j \le D$, $z \in \mathcal{L}_D$.

We assume for the rest of this section that

$$d \geq 2,$$

since we will in fact applying these results only for $d \geq 3$. Our next goal is to prove an exponential bound on the size of clusters of dry sites. Up to this point the definitions are almost the same as the ones in Durrett [13]. However, we must now change the details of the contour argument there, so that it is done on the correct graph. Let $y \in \mathcal{L}_D$ with $y_D = n \geq 0$ (write $y \in \mathcal{L}_D^+$). In addition to P as in (5.7), for $M > 0$ we also work with a probability $\bar{P} = \bar{P}_{n,M}$ under which $\eta(z) = 1$ for $z = (z', m) \in \mathcal{L}_D^+$ satisfying $m \leq n$ and $|z'| \geq M$, and the remaining $\eta(z)$'s are as in (5.7). Therefore under $\bar{P}$ the sets of wet sites $\{\bar{W}_n\}$ will be larger, although we will use the same notation since their definition is the same under either probability law. If y is wet put $D_y = \emptyset$, and otherwise let D_y be the connected component in $(\mathcal{L}_D, \mathcal{E}_\downarrow)$ of dry sites containing y. That is, $z \in D_y$ iff there are $z_1 = y, z_2, \ldots, z_K = z$ all in $\mathcal{L}_D$ so that the edge from z_i to z_{i+1} is in $\mathcal{E}_\downarrow$ and each z_i is dry. Since all sites in $\mathcal{H}_0$ are wet, $D_y \subset \{z \in \mathcal{L}_D : n \geq z_D > 0\}$, and under $\bar{P}_{n,M}$, $D_y \subset \{z \in \mathcal{L}_D : n \geq z_D > 0, |(z_1, \ldots, z_{D-1})| < M\}$. We assume that ω satisfies

$$D_y(\omega) \text{ is finite.} \tag{5.10}$$

The fact that (5.10) holds a.s. under $\bar{P}_{n,M}$ is the reason this law was introduced. To make D_y into a solid object we consider the compact solid

$$R_y = \cup_{z \in D_y}(z + \mathcal{Q}) \subset \mathbb{R}^d \times \mathbb{R}_+ \, .$$

If R_y^c is the complement of R_y in $\mathbb{R}^d \times \mathbb{R}_+$, we claim that both R_y and R_y^c are path-connected. For R_y, suppose for concreteness that $D = 3$ and note that for the diagonally adjacent points $y(0) = \mathcal{A}(0, 0, 1)$ and $y(1) = \mathcal{A}(1, -1, 1)$, $R_{y(0)} \cap R_{y(1)}$ contains the edge $\mathcal{A}(\{1/2\} \times \{-1/2\} \times [1/2, 3/2])$. For R_y^c, if $x \in R_y^c$ then there exists $[x] \in \mathcal{L}_D^+ \setminus D_y$ such that $x \in [x] + \mathcal{Q}$ and the line segment from x to $[x]$ is contained in R_y^c. We first assume $[x] \in \mathcal{H}_k$ for some $k \in \{1, 2 \ldots, n\}$. If $[x]$ is wet then there must be a path in R_y^c connecting $[x]$ to $\mathcal{H}_0$. Suppose $[x]$ is dry, and let $z_0, z_1, \ldots, z_K$ be a path in $\mathcal{E}_\downarrow$ connecting $z_0 = y$ to $z_K = [x]$. At least one site on this path must be wet (else $[x] \in D_y$), so let z_j be the first wet site encountered starting at z_K. Then for each $i > j$, z_i is dry and $z_i \notin D_y$ (or else $[x]$ would in D_y). Thus $\cup_{i=j}^K (z_i + \mathcal{Q})$ is path-connected, contained in R_y^c, and z_j is connected to $\mathcal{H}_0$ by a path in R_y^c. Note that $\mathcal{H}_0 \subset (\mathbb{R}^d \times \{0\}) \cap R_y^c \equiv \tilde{\mathcal{H}}_0$ which is path-connected because the rotated cubes making up R_y can only intersect $\mathbb{R}^d \times \{0\}$ in a discrete set of points (since $D_y \subset \{z_D > 0\}$). It is here that we use $d \geq 2$. Now suppose $[x] \in \mathcal{H}_k$ for some $k > n$. $\tilde{\mathcal{H}}_0$ is also connected to $\mathcal{H}_{n+1}$ by a path in R_y^c (assuming $\theta < 1$). This allows us to connect $[x]$ to $\tilde{\mathcal{H}}_0$ and so conclude that R_y^c is path-connected.

Let Γ_y be the boundary R_y. To study Γ_y we need some notation. We define the plus faces of $[-1/2, 1/2]^D$ to be $[-1/2, 1/2]^m \times \{1/2\} \times [-1/2, 1/2]^{D-m-1}$, and define the minus faces to be $[-1/2, 1/2]^m \times \{-1/2\} \times [-1/2, 1/2]^{D-m-1}$, $m = 0, \ldots, D-1$. The images of the plus and minus faces of $[-1/2, 1/2]^d$ under A constitute the plus and minus faces of $Q = A([-1/2, 1/2]^d)$, which are used to defined the plus and minus faces of Γ_y in the obvious way. Note that the plus faces of Γ_y will have outward normal v_i for some i while the minus faces will have outward normal $-v_i$ for some i.

Lemma 5.3. *If (5.10) holds, then Γ_y is connected and bounded.*

Proof. For $\varepsilon > 0$ let $R_y^\varepsilon = \{x \in \mathbb{R}^d : |x - w|_\infty < \varepsilon \text{ for some } w \in R_y\}$. Since R_y is connected, so is R_y^ε. If $\kappa(U)$ denotes the number of path-connected components of a set U, it is a consequence of the Mayer-Vietoris exact sequence with $n = 0$ that for open sets $U, V \subset \mathbb{R}^D$ with $U \cup V = \mathbb{R}^D$,

$$\kappa(U \cap V) = \kappa(U) + \kappa(V) - 1.$$

See page 149 of [31] and also Proposition 2.7 of that reference. Applying this to the open connected (hence path-connected) sets R_y^ε and R_y^c whose union is $\mathbb{R}^D$, we find that $R_y^\varepsilon \cap R_y^c$ is path-connected.

Finally, $R_y^\varepsilon \cap R_y^c$ is homotopic to Γ_y, and therefore Γ_y is also path-connected. Boundedness is immediate from (5.10). $\square$

For the next result we follow the proof of Lemma 6 from [13]. A *contour* will be a finite union of faces in $\mathcal{L}_D$ which is connected.

Lemma 5.4. *There are constants $C_{5.4}$ and $\mu_{5.4}$ which only depend on the dimension D so that the number of possible contours with N faces, containing a fixed face, is at most $C_{5.4}(\mu_{5.4})^N$.*

Proof. Make the set of faces of $\mathcal{L}_D$ into a graph by connecting two if they share a point in common. Note that by the above definition a contour corresponds to a finite connected subset of this graph. Each point in the graph has a constant degree $\nu = \nu(D)$. An induction argument shows that any connected set of N vertices has at most $N(\nu - 2) + 2$ boundary points. (Adding a new point removes 1 boundary point and adds at most $\nu - 1$ new ones.) Consider percolation on this graph in which sites are open with probability a and closed with probability $1 - a$. Let 0 be a fixed point of the graph corresponding to our fixed face, and $\mathcal{C}_0$ be the component containing 0. If B_N is the number of components of size N containing 0, then

$$1 \geq P(|\mathcal{C}_0| = N) \geq B_N a^N (1 - a)^{N(\nu-2)+2}.$$

Rearranging, we get $B_N \leq C\mu^N$ with $C = (1 - a)^{-2}$ and $\mu = a^{-1}(1 - a)^{-(\nu-2)}$. Taking the derivative of $-\log a - (\nu - 2)\log(1 - a)$ and setting it equal to 0, we see that $a = 1/(\nu - 1)$ optimizes the bound, and gives constants that only depend on the degree ν. $\square$

Lemma 5.5. *If $\theta_{5.5} = (2\mu_{5.4})^{-2D}$, then $\theta \leq \theta_{5.5}$ implies that for all $y = (y', n) \in \mathcal{L}_D^+$ and all $M > |y'|$, $\bar{P}_{n,M}(|\Gamma_y| \geq N) \leq 2C_{5.4}2^{-N}$ for all $N \in \mathbb{N}$.*

Proof. By Lemma 5.3 if $D_y \neq \emptyset$ we see that under $\bar{P}_{n,M}$, Γ_y is a contour which by definition contains the plus faces of $y+Q$. Given a plus face in Γ_y if we travel the line perpendicular to $\mathbb{R}^d \times \{0\}$ and through the center of the face, then we enter and leave the set an equal number of times, so the number of plus faces of Γ_y is equal to the number of minus faces. Thus, if the contour Γ_y has size N there are $N/2$ minus faces. It is easy to see that a point of $\bar{W}_j$ adjacent to a minus face associated with a point in $\bar{V}_{j+1}$ must be closed for otherwise it would wet the point in $\bar{V}_{j+1}$ (recall the outward normal of a minus face is $-v_i$ for some i). The point of $\bar{W}_j$ that we have identified might be associated with as many as D minus faces, but in any case for a contour of size N there must be at least $N/2D$ associated closed sites. Taking $\theta \leq (2\mu_{5.4})^{-2D}$, using Lemma 5.4 to bound the number of possible contours containing a fixed plus face of $y+Q$, and summing the resulting geometric series now gives the result. $\square$

It follows from the above and an elementary isoperimetric inequality that there are finite positive constants C, c such that for all $y=(y',n) \in \mathcal{L}_D^+$ and $M > |y'|$,

$$\text{if } \theta \leq \theta_{5.5} \text{ then } \bar{P}_{n,M}(|D_y| \geq N) \leq C\exp(-cN^{(D-1)/D}) \text{ for all } N \in \mathbb{N}. \tag{5.11}$$

Now fix $r>0$ and let $\mathcal{B}_n$ be the dry sites in $\mathcal{H}_n^{rn/4}$ connected to the complement of $\cup_{m=n/2}^n \mathcal{H}_m^{rm/2}$ by a path of dry sites on the graph with edges $\mathcal{E}_\downarrow$, where as for open sites the last site in such a path need not be dry.

Lemma 5.6. *If $\theta \leq \theta_{5.5}$ then*

$$P(\mathcal{B}_n \neq \emptyset \textit{ infinitely often}) = 0.$$

Proof. Let $M > n(r+\sqrt{2D})$. We couple the iid Bernoulli random variables $\{\eta(z) : z \in \mathcal{L}_D\}$ (under P) with the corresponding random field $\bar{\eta}$ (under $\bar{P} = \bar{P}_{n,M}$) so that

$$\eta(z) = \bar{\eta}(z) \quad \forall z=(z',m) \;\text{ where } |z'|<M \text{ or } m>n.$$

We claim that $z \in \cup_{m=n/2}^n \mathcal{H}_m^{rm} \equiv \hat{\mathcal{H}}_n$ is wet for η iff it is wet for $\bar{\eta}$. It clearly suffices to fix $z=(z',m) \in \hat{\mathcal{H}}_n$ which is wet for $\bar{\eta}$ and show it is wet for η. A path of sites $z_i = (z_i', i)$, $i=0,\dots,m$ with edges in $\mathcal{E}_\uparrow$ from $\mathcal{H}_0$ to z satisfies $\max_{i\leq m}|z_i'| \leq rn + \sqrt{2D}n < M$. This is because the edges in $\mathcal{E}_\uparrow$ have length at most $\sqrt{2D}$. Therefore if the sites in the path are open in $\bar{\eta}$, then they will also be open in η. This proves the claim.

Next note that if $y \in \mathcal{H}_n^{rn/4}$, then $y \in \mathcal{B}_n$ for η iff $y \in \mathcal{B}_n$ for $\bar{\eta}$. This is because the path of dry sites connecting y to the complement of $\cup_{m=n/2}^n \mathcal{H}_m^{rm/2}$ can be taken to be inside $\hat{\mathcal{H}}_n$ and so we may apply the claim in the last paragraph. It now follows from the above bound on the length of the edges in $\mathcal{E}_\downarrow$ that

$$P(y \in \mathcal{B}_n) = \bar{P}_{n,M}(y \in \mathcal{B}_n) \leq \bar{P}_{n,M}\Big(|D_y| \geq \frac{c(r)n}{\sqrt{2D}}\Big).$$

The number of sites in $\mathcal{H}_n^{rn/4}$ is at most Cn^d, and the bound in (5.11) shows that $P(\mathcal{B}_n \neq \emptyset) \leq \sum_{y \in \mathcal{H}_n^{rn/4}} P(y \in \mathcal{B}_n)$ is summable over n. $\square$

Remark 5.1. *We will prove in Section 7 that if wet sites have the property in* (5.9)*, and the kernels* $p(\cdot)$ *and* $q(\cdot)$ *are finite range, then for an appropriate* $r > 0$*,* $\mathcal{B}_n = \emptyset$ *will imply that on* Ω^0_∞ *all sites in* $[-c_{L,d}rn, c_{L,d}rn]^d$ *will be vacant at times* $t \in [(n-1)J_1T, nJ_1T]$*. This linearly growing dead zone will guarantee extinction of the 1's.*

6 Existence of stationary distributions

With the convergence of the particle system to the PDE established and the percolation result introduced, we can infer the existence of stationary distributions by using a "block construction". Recall that our voter model perturbations take values in $\{0,1\}^{\varepsilon\mathbb{Z}^d}$ and so our stationary distributions will be probabilities on this space of rescaled configurations. We begin with a simple result showing that for stationary distributions, having some 1's a.s. or infinitely many 1's a.s. are equivalent. Let $|\xi| = \sum_x \xi(x)$.

Lemma 6.1. *If ν is a stationary distribution for a voter perturbation, then*

$$|\xi| = \infty \quad \nu - a.s. \quad \textit{iff} \quad |\xi| > 0 \quad \nu - a.s.$$

Proof. It suffices to prove

$$\nu(|\xi| < \infty) > 0 \text{ implies } \nu(|\xi| = 0) > 0. \tag{6.1}$$

Assume first that the 0 configuration is a trap. Then if $|\xi_0| = K < \infty$, (1.27) shows the sum of the flip rates is finite and so it is easy to prescribe a sequence of K flips which occur with positive probability and concludes with the 0 state. By stationarity we get the implication in (6.1).

Assume next that 0 is not a trap, which means $g_1^\varepsilon(0,\ldots,0) > 0$. We claim that $\nu(|\xi| < \infty) = 0$, which implies the required result. Intuitively this is true because configurations with finitely many 1's have an infinite rate of production of 1's. One way to prove this formally is through generators. Let Ω^ε be the generator of our voter perturbation, Ω_v be the generator of the voter model in (1.2) and for $i = 0, 1$

$$\Omega_i \psi(\xi) = \sum_{x\in\mathbb{Z}^d} 1(\xi(x) = 1-i) E(g_i^\varepsilon(\xi(x+Y^1),\ldots,\xi(x+Y^{N_0})))(\psi(\xi^x) - \psi(\xi)).$$

Here ψ will be a bounded function on $\{0,1\}^{\mathbb{Z}^d}$ depending on finitely many coordinates, and we recall that ξ^x is ξ with the coordinate at x flipped to $1-\xi(x)$. Recall that $\xi_\varepsilon(\varepsilon x) = \xi(x)$ for $\xi \in \{0,1\}^{\mathbb{Z}^d}, x \in \mathbb{Z}^d$. For ψ as above define ψ_ε on $\{0,1\}^{\varepsilon\mathbb{Z}^d}$ by $\psi_\varepsilon(\xi_\varepsilon) = \psi(\xi)$. Then by (1.27) and (1.28),

$$\Omega^\varepsilon \psi_\varepsilon(\xi_\varepsilon) = (\varepsilon^{-2} - \varepsilon_1^{-2})\Omega_v\psi(\xi) + \Omega_0\psi(\xi) + \Omega_1\psi(\xi). \tag{6.2}$$

For $0 < r < R$, let $A(r,R) = \{x \in \mathbb{Z}^d : r \le |x| \le R\}$ and

$$\psi_{r,R}(\xi) = 1(\xi|_{A(r,R)} \equiv 0), \quad \xi \in \{0,1\}^{\mathbb{Z}^d}.$$

Considering two cases $x \in A(r,R)$ and $x \notin A(r,R)$ we have

$$\text{if } \xi(x) = 0 \text{ then } \psi_{r,R}(\xi^x) - \psi_{r,R}(\xi) \le 0. \tag{6.3}$$

Since $\psi_{r,R}(\xi^x) - \psi_{r,R}(\xi) = 1$ only if x is the only site in $A(r,R)$ where $\xi(x) = 1$, we have

$$\Omega_v\psi_{r,R}(\xi) \le 1, \quad \Omega_0\psi_{r,R}(\xi) \le \|g_0^\varepsilon\|_\infty. \tag{6.4}$$

Choose λ so that $P(Y^* \leq \lambda) \geq 1/2$, where Y^* is as in (1.6). Flipping a site from 0 to 1 cannot increase $\psi_{r,R}$, and $\psi_{r,R}(\xi) = 1$ implies $\xi(x) = 0$ for all $x \in A(r, R)$, so we have

$$\begin{aligned}\Omega_1 \psi_{r,R}(\xi) &\leq - \sum_{x \in A(r,R)} (1 - \xi(x)) g_1^\varepsilon(0, \ldots, 0) P(\xi(x + Y^i) = 0 \text{ for } 1 \leq i \leq N_0) \psi_{r,R}(\xi) \\ &\leq - \frac{g_1^\varepsilon(0, \ldots, 0)}{2} \psi_{r,R}(\xi) |A(r + \lambda, R - \lambda)|. \qquad (6.5)\end{aligned}$$

The stationarity of ν implies, see Theorem B.7 of Liggett [36], that if $\psi = \psi_{r,R}$ then $\int \Omega^\varepsilon \psi_\varepsilon d\nu = 0$. Using (6.2), (6.4) and (6.5), and noting that

$$\int \psi_\varepsilon \, d\nu = \nu(\xi \equiv 0 \text{ on } A(\varepsilon r, \varepsilon R)),$$

we have

$$0 \leq (\varepsilon^{-2} - \varepsilon_1^{-2}) + \|g_0^\varepsilon\|_\infty - \frac{g_1^\varepsilon(0, \ldots, 0)}{2} |A(r + \lambda, R - \lambda)| \nu(\xi \equiv 0 \text{ on } A(\varepsilon r, \varepsilon R)).$$

Rearranging this inequality we get

$$\nu(\xi \equiv 0 \text{ on } A(\varepsilon r, \varepsilon R)) \leq \frac{2((\varepsilon^{-2} - \varepsilon_1^{-2}) + \|g_0^\varepsilon\|_\infty)}{g_1^\varepsilon(0, \ldots, 0)|A(r + \lambda, R - \lambda)|}$$

(recall $g_1^\varepsilon(0, \ldots, 0) > 0$). Letting $R \to \infty$ we conclude that $\nu(\xi \equiv 0 \text{ on } A(\varepsilon r, \infty)) = 0$. In words, for ν-a.a. configurations there is a 1 outside the ball of radius εr. As this holds for all $r < \infty$, there are infinitely many ones with probability 1 under ν.
□

Assumption 1 and (1.41) are in force throughout the rest of this section and we drop dependence on the parameters w, v_i, L_0, L_1, r_0, etc. arising in those hypotheses in our notation. We continue to work with the particle densities $D(x, \xi)$ using the choice of r in (4.6). We start with a version of Lemma 4.2 which is adapted for proving coexistence. We let

$$L_2 = 3 + L_0 \vee L_1.$$

Lemma 6.2. *There is a $C_{6.2} > 0$ and for every $\eta > 0$, there are $T_\eta \geq 1$ and $\varepsilon_{6.2}(\eta) > 0$ so that for $t \in [T_\eta, C_{6.2} \log(1/\varepsilon)]$ and $0 < \varepsilon < \varepsilon_{6.2}$, if*

$$\xi_0^\varepsilon \textit{ has density in } [v_0 + \eta, v_1 - \eta] \textit{ on } [-L_2, L_2]^d,$$

then

$$P(\xi_t^\varepsilon \textit{ has density in } [u_* - \eta, u^* + \eta] \textit{ on } [-wt, wt]^d | \xi_0^\varepsilon) \geq 1 - \varepsilon^{.01}.$$

The proof is derived by making minor modifications to that of Lemma 4.2 and so is omitted. We will always assume $\eta > 0$ is small enough so that

$$0 < v_0 + \eta \leq u_* - \eta < u^* + \eta \leq v_1 - \eta < 1.$$

The one-sided versions of the above Lemma also hold (recall Lemma 4.2 on which the proof is based is a one-sided result), that is, with only one-sided bounds on the densities in the hypothesis and conclusion.

Theorem 1.4. *Consider a voter model perturbation on $\mathbb{Z}^d$ satisfying (1.41). Suppose Assumption 1. If $\varepsilon > 0$ is small enough, then coexistence holds and the non-trivial stationary distribution ν may be taken to be translation invariant.*

If $\eta > 0$ and $\varepsilon > 0$ is small enough, depending on η, then any stationary distribution ν such that

$$\nu\Big(\sum_x \xi(x) = 0 \text{ or } \sum_x (1-\xi(x)) = 0\Big) = 0 \tag{6.6}$$

satisfies $\nu(\xi(x) = 1) \in (u_* - \eta, u^* + \eta)$ for all x.

Proof. We use the block construction in the form of Theorem 4.3 of [14]. This result is formulated for $D = 2$ but it is easy to extend the proof to $D \geq 3$, and we use this extension without further comment. Recall $Q(r) = [-r, r]^d$ and $Q^\varepsilon(r) = Q(r) \cap \varepsilon\mathbb{Z}^d$. Let $U = (C_{6.2}/2)\log(1/\varepsilon)$, $L = wU/(\alpha_0 D + 1)$, where $\alpha_0 > 0$ is a parameter to be chosen below, and $I^*_\eta = [u_* - \eta/4, u^* + \eta/4]$. Next we define the sets H and G_ξ which appear in the above Theorem. Let

$$H = \{\xi \in \{0,1\}^{\varepsilon\mathbb{Z}^d} : \xi \text{ has density in } I^*_\eta \text{ on } Q(L)\},$$

that is, if $Q_\varepsilon = [0, a_\varepsilon)^d \cap \varepsilon\mathbb{Z}^d$ then the fraction of occupied sites in $x + Q_\varepsilon$ is in $I^*_\eta = [u^* - \eta/4, u^* + \eta/4]$ whenever $x \in a_\varepsilon\mathbb{Z}^d \cap [-L, L]^d$. If $L' = L + 1$, then $\{\xi \in H\}$ depends on $\xi|_{[-L',L']^d}$. Here we need to add 1 as the cubes of side a_ε with "lower left-hand corner" at $x \in [-L, L]^d$ will be contained in $[-L', L']^d$. This verifies the measurability condition in Theorem 4.3 of [14] with $L' = L + 1$ in place of L which will affect nothing in the proof of Theorem 4.3.

Let G_ξ be the event on which (a) if $\xi^\varepsilon_0 = \xi$, then ξ^ε_U has density in I^*_η on $Q(wU)$ and (b) for all $z \in Q^\varepsilon(wU + 1)$ and all $t \leq U$, $X^{z,U}_t \subset Q((w + b_0)U + 1)$, where $b_0 = 16(3 + d)/C_{6.2}$. Note that

$$\begin{aligned} G_\xi \in &\sigma\Big(\Lambda^y_r|_{[0,U]\times\varepsilon\mathbb{Z}^{dN_0}\times[0,1]}, \Lambda^y_w|_{[0,U]\times\varepsilon\mathbb{Z}^d} : y \in Q^\varepsilon((w + b_0)U) + 1\Big) \\ \equiv &\mathcal{G}(Q((b_0 + w)U + 1)) \times [0, U]) \end{aligned} \tag{6.7}$$

Informally, $\mathcal{G}(R)$ is the σ-field of generated by the points in the graphical representation that lie in R. The above measurability is easy to verify using the duality relation (2.17).

Consider now the Comparison Assumptions prior to Theorem 4.3 of [14]. In our context we need to show

Lemma 6.3. *For $0 < \varepsilon < \varepsilon_{6.3}(\eta)$:*

(i) if $\xi^\varepsilon_0 \in H$, then on $G_{\xi^\varepsilon_0}$, ξ^ε_U has density in I^*_η on $\alpha_0 L v'_i + [-L, L]^d$, $1 \leq i \leq D$,

(ii) if $\xi \in H$, then $P(G_\xi) \geq 1 - \varepsilon^{0.009}$.

Proof. By assuming $\varepsilon < \varepsilon_1(\eta)$ we have $U \geq T_{\eta/4}$ and $L \geq L_2$. Using the definition of L and the fact that $|v_i'| \leq \|v_i'\|_2 = \sqrt{D-1}$ one easily checks that

$$\alpha_0 L v_i' + [-L, L]^d \subset [-wU, wU]^d \text{ for } i = 1, \dots, D. \tag{6.8}$$

Part (a) of the definition of G_ξ now gives (i). By Lemma 4.7 with parameters $L = wU + 1$, $2b = b_0$, $c = C_{6.2}/2$ and $T' = U$, and Lemma 6.2, for $\xi \in H$ we have for $\varepsilon < \varepsilon_{6.2}(\eta)$,

$$\begin{aligned} P(G_\xi^c) &\leq \varepsilon^{.01} + c_d'(wU+1)^d(U+1)\varepsilon^{((b_0 C_{6.2})/16)-2-d} \\ &\leq \varepsilon^{.01} + c(\log(1/\varepsilon))^{d+1}\varepsilon \leq \varepsilon^{.009}, \end{aligned}$$

where the last two inequalities hold for small ε. We may reduce $C_{6.2}$ to ensure that $b = b_0/2$ satisfies the lower bound in Lemma 4.7. This proves (ii). $\square$

Continue now with the proof of Theorem 1.4. Let $\varepsilon < \varepsilon_{6.3}$ and define

$$V_n = \{(x, n) \in \mathcal{H}_n : \xi_{nU}^\varepsilon \text{ has density in } I_\eta^* \text{ on } \alpha_0 Lx + [-L, L]^d\}.$$

(To be completely precise in the above we should shift $\alpha_0 Lx$ and $\alpha_0 L v_i'$ to the point in $\varepsilon\mathbb{Z}^d$ "below and to the left of it" but the adjustments become both cumbersome and trivial so we suppress such adjustments in what follows.) If we let

$$R_{y,n} = (y\alpha_0 L, nU) + Q((b_0 + w)U + 1) \times [0, U], \text{ for } (y, n) \in \mathcal{L}_D$$

and

$$M = \left\lceil \frac{2(b_0 + w)(\alpha_0 D + 1)}{\alpha_0 w} \right\rceil,$$

then $R_{y_1,m} \cap R_{y_2,n} = \emptyset$ if $|(y_1, m) - (y_2, n)| > M$. Since $\mathcal{G}(R_i)$, $1 \leq i \leq k$ are independent for disjoint R_i's, Lemma 6.3 allows us to apply the proof of Theorem 4.3 of [14]. This shows there is an M-dependent (in the sense of (5.5)) oriented percolation process $\{W_n\}$ on $\mathcal{L}_D$ with density at least $1 - \varepsilon^{.009}$ such that $W_0 = V_0$ and $W_n \subset V_n$ for all $n \geq 0$. We note that although a weaker definition of M-dependence is used in [14] (see (4.1) of that reference), the proof produces $\{W_n\}$ as in (5.5). By Lemma 5.2 with $r = r_{5.2}$ and $\theta = \varepsilon^{.009}$, if $\varepsilon < \varepsilon_1(\eta)$, then

$$\begin{aligned} \lim_{n\to\infty} \inf_{(x,n)\in\mathcal{H}_n^{rn}} & P(\xi_{nU}^\varepsilon \text{ has density in } I_\eta^* \text{ on } \alpha_0 Lx + [-L, L]^d) \\ &\geq \left(1 - \frac{\eta}{4}\right) P(0 \in V_0). \end{aligned} \tag{6.9}$$

We will choose different values of α_0 to first prove the existence of a stationary law, and then to establish the density bound for any stationary distribution. For the first part, set $\alpha_0 = 3$ and take $\{\xi_0^\varepsilon(x) : x \in \varepsilon\mathbb{Z}^d\}$ to be iid Bernoulli variables with mean $u = (u_* + u^*)/2$. The weak law of large numbers implies that if ε is small enough

$$P(\xi_0^\varepsilon \text{ has density in } I_\eta^* \text{ on } [-L, L]^d) \geq \frac{1}{2}. \tag{6.10}$$

Since $\alpha_0 = 3$, $L \geq 3$ and $|x-y| \geq \|x-y\|_2/\sqrt{D} \geq 1$ for all $x \neq y \in \mathcal{H}'_n$, $\{\alpha_0 Lx + [-L', L']^d : x \in \mathcal{H}'_n\}$ is a collection of disjoint subsets of $\mathbb{R}^d$ for each n. This and the measurability property of $\{\xi \in H\}$ noted above shows that if $0 < \varepsilon < \varepsilon_0(\eta)$ then $\{V_n\}$ is bounded below by an M-dependent (as in (5.5)) oriented percolation process, $\{W_n^{1/2}\}$, with density at least $\geq 1 - \varepsilon^{.009}$ starting with an iid Bernoulli (1/2) field. Having established that our process dominates oriented percolation, it is now routine to show the existence of a nontrivial stationary distribution. We will spell out the details for completeness. Recall the notation introduced just prior to Lemma 4.7: $Q(x,r) = [x-r, x+r]^d$ and $Q^\varepsilon(x,r) = Q(x,r) \cap \varepsilon\mathbb{Z}^d$.

Lemma 6.4. *Assume $\alpha_0 = 3$ and $\{\xi_0^\varepsilon(x) : x \in \varepsilon\mathbb{Z}^d\}$ are as above. There is an $\varepsilon_{6.4}(\eta) > 0$ so that for any $\varepsilon \in (0, \varepsilon_{6.4}(\eta))$ and any $k \in \mathbb{N}$ there are $t_1(k,\varepsilon)$, $M_1(k,\varepsilon) > 0$ so that for $t \geq t_1$,*

$$P\Big(\sum_{|x| \leq M_1} \xi_t^\varepsilon(x) \geq k \text{ and } \sum_{|x|\leq M_1} 1 - \xi_t^\varepsilon(x) \geq k\Big) \geq 1 - \frac{2}{k}.$$

Proof. As in Theorem A.3 of [14] for $k \in \mathbb{N}$ there are $n_0, \ell_0, M_0 \in \mathbb{N}$ and $z_1, \dots, z_{4k} \in Q(M_0)$ satisfying $|z_i - z_j| > 3M + 2\ell_0 + 1$ for $i \neq j$, such that for $n \geq n_0$ with probability at least $1 - k^{-1}$

$$W_n^{1/2} \cap Q(z_j, \ell_0) \neq \emptyset \text{ for } j = 1, \dots, 4k. \tag{6.11}$$

where $Q(z_j, \ell_0) = z_j + Q(\ell_0)$. The above implies there are $\sigma(\xi_{nU}^\varepsilon)$-measurable $y_j \in Q^\varepsilon(z_j, \ell_0)$ such that

$$\xi_{nU}^\varepsilon \text{ has density in } I_\eta^* \text{ on } 3Ly_j + [-L,L]^d, \quad j = 1, \dots, 4k. \tag{6.12}$$

This proves the result for $t = nU$. Intermediate times can be easily handled using Lemma 6.2 and the finite speed of the dual (Lemma 4.7). Those results show that for a fixed $\varepsilon < \varepsilon_{6.2}$ and $t \geq (n_0+1)U$, if we choose $n \geq n_0$ so that $t \in [(n+1)U, (n+2)U]$ (use $T_{\eta/4} \leq 2U = C_{6.2} \log(1/\varepsilon)$ in applying Lemma 6.2), then on the event in (6.12) we have

$$\begin{aligned}
&P(\xi_t^\varepsilon \text{ has density in } I_\eta^* \text{ on } 3Ly_j + [-L,L]^d, \\
&\quad \text{and } X_s^{x,t} \in Q(3Ly_j, L' + b_0 U) \text{ for all } x \in 3Ly_j + [-L', L']^d \text{ and } s \in [0,t] | \xi_{nU}^\varepsilon) \\
&\quad \geq 1 - \varepsilon^{.01} - c_1(\log(1/\varepsilon))^{d+1}\varepsilon \geq \frac{1}{2}.
\end{aligned}$$

where in the last inequality we may have needed to make ε smaller.

Our separation condition on the $\{z_j\}$ and $L \geq 3$ implies that $Q(3Ly_j, L' + b_0 U)$, $j = 1, \dots, 4k$ are disjoint and so the events on the left-hand side are conditionally independent as j varies. Therefore a simple binomial calculation shows that

$$\begin{aligned}
&P(|\{j \leq 4k : \xi_t^\varepsilon \text{ has density in } I_\eta^* \text{ on } 3Ly_j + [-L,L]^d\}| \geq k) \\
&\geq \Big(1 - \frac{1}{k}\Big)\Big(1 - \frac{1}{k}\Big) \geq 1 - \frac{2}{k}.
\end{aligned}$$

Here the first $1 - \frac{1}{k}$ comes from establishing (6.12) and the second $1 - \frac{1}{k}$ comes from the binomial error in getting fewer than k points with appropriate density at time t. Since the above event implies the required event with $M_1 = 3L(M_0 + \ell_0) + L$ we are done. $\square$

Fix $\varepsilon < \varepsilon_{6.4}$. By Theorem I.1.8 of [35] there is a sequence $t_n \to \infty$ s.t. $t_n^{-1} \int_0^{t_n} 1(\xi_s \in \cdot)\, ds \to \nu$ in law where ν is a translation invariant stationary distribution for our voter perturbation. Lemma 6.4 easily shows that there are infinitely many 0's and 1's ν-a.s., proving the first part of Theorem 1.4.

Turning to the second assertion, by Lemma 6.1 and symmetry it suffices to show that for $\varepsilon < \varepsilon_2(\eta)$ and any given stationary ν with infinitely many 0's and 1's a.s. then

$$\sup_x \mu(\xi(x) = 1) \le u^* + \eta.$$

Start the system with law ν. We claim that

Lemma 6.5. *There is a $\sigma(\xi_0^\varepsilon)$-measurable r.v. $x_0 \in \varepsilon\mathbb{Z}^d$ such that $\xi_0^\varepsilon \equiv 0$ on $Q^\varepsilon(x_0, L)$ a.s. More generally w.p. 1 there is an infinite sequence $\{x_i : i \in \mathbb{Z}_+\}$ of such random variables satisfying $|x_i - x_j| \ge 4L + 1$ for all $i \ne j$.*

Proof. To see this condition on ξ_0^ε, choose x_0 so that $\xi_0^\varepsilon(x_0) = 0$ and note that if R_1^x is the first reaction time of the dual $X^{x,\varepsilon}$, the event "$\xi_0^\varepsilon \equiv 0$ on $Q^\varepsilon(x_0, L)$" occurs if for all $x \in x_0 + [-L, L]^d$, $R_1^x > 1$, $X_1^{x,\varepsilon} = x_0$, and $\sup_{s \le 1} |X_s^{x,\varepsilon} - x| \le 1$. Call the last event $A(x_0)$. The last condition has been imposed so that if $|x_0 - x_1| \ge 4L + 1$ then the events $A(x_0)$ and $A(x_1)$ are (conditionally) independent. Clearly they have positive probability. Given our initial configuration with $|\{y : \xi_0^\varepsilon(y) = 0\}| = \infty$ a.s., we can pick an infinite sequence x_i, $i \in \mathbb{N}$, with $\xi_0^\varepsilon(x_i) = 0$ and $|x_j - x_i| \ge 4L + 1$ when $j > i$, so the strong law of large numbers implies that at time 1 there will be infinitely many x_i with $\xi_1^\varepsilon(x) = 0$ for all $x \in Q^\varepsilon(x_i, L)$. By stationarity this also holds at time 0. $\square$

Now condition on ξ_0^ε, shift our percolation construction in space by x_0, set $\alpha_0 = (2D)^{-1}$ and only require the density to be at most $u^* + \eta/4$ in our definition of V_n which now becomes

$$V_n = \{(x, n) \in \mathcal{H}_n : \xi_{nU}^\varepsilon \text{ has density at most } u^* + \eta/4 \text{ on } x_0 + c_L x + [-L, L]^d\},$$

where we recall from (5.2) that $c_L = L/(2D)$. (Here we are using the one-sided version of Lemma 6.2 mentioned above, after its statement.) Then $0 \in V_0$ and the one-sided analogue of (6.9) shows that if $\varepsilon < \varepsilon_3(\eta)$, then

$$\lim_{n\to\infty} \inf_{(x,n)\in\mathcal{H}_n^{rn}} P(x \in V_n) \ge 1 - \frac{\eta}{4}.$$

Recall from (5.2) that $\cup_{x\in\mathcal{H}'_n} x_0 + c_L x + [-L, L]^d = \mathbb{R}^d$, so this implies for any $x \in \mathbb{R}^d$ and n large enough,

$$P(\xi_{nU}^\varepsilon \text{ has density at most } u^* + \frac{\eta}{4} \text{ on } x + [-L, L]^d) \ge 1 - \frac{\eta}{3},$$

and so by stationarity

$$\nu(\xi^\varepsilon \text{ has density at most } u^* + \frac{\eta}{4} \text{ on } x + [-L, L]^d) \geq 1 - \frac{\eta}{3} \text{ for all } x \in \mathbb{R}^d.$$

To complete the proof, run the dual for time t_ε (t_ε as in (4.6)) and apply Lemma 4.5 with $u = u^* + \frac{\eta}{4}$ to see that for $x \in \varepsilon\mathbb{Z}^d$ and $\varepsilon < \varepsilon_3(\eta) \wedge \varepsilon_{4.5}(\eta/3)$,

$$\begin{aligned}
\nu(\xi(x) = 1) &= P(\xi^\varepsilon_{t_\varepsilon}(x) = 1) \\
&\leq P(R_1 \leq t_\varepsilon) + E(P(R_1 > t_\varepsilon, \xi^\varepsilon_{t_\varepsilon}(x) = 1 | \xi^\varepsilon_0)) \\
&\leq (1 - e^{-c^* t_\varepsilon}) + E(P(\xi^\varepsilon_0(B^{\varepsilon,x}_{t_\varepsilon}) = 1 | \xi^\varepsilon_0)) \\
&\leq c^* t_\varepsilon + \frac{\eta}{3} + u^* + \frac{\eta}{4} + \frac{\eta}{3} \leq u^* + \eta,
\end{aligned}$$

where ε is further reduced, if necessary, for the last inequality. $\square$

7 Extinction of the process

7.1 Dying out

Our goal in this section is to show that if $f'(0) < 0$ and $|\xi_0^\varepsilon|$ is $o(\varepsilon^{-d})$, then with high probability ξ_t^ε will be extinct by time $O(\log(1/\varepsilon))$. Throughout this Section we assume that $0 < \varepsilon \le \varepsilon_0$ and that (1.43) holds, i.e., $g_1^\varepsilon(0,\dots,0) = 0$ for $0 < \varepsilon \le \varepsilon_0$

Recall from (3.30) the drift at εx in the rescaled state $\xi_\varepsilon \in \{0,1\}^{\varepsilon\mathbb{Z}^d}$ (recall the notation prior to (1.25)) is

$$d_\varepsilon(\varepsilon x, \xi_\varepsilon) = (1-\xi(x))h_1^\varepsilon(x,\xi) - \xi(x)h_0^\varepsilon(x,\xi),$$

and define the total drift for $|\xi_\varepsilon| < \infty$ by

$$\psi_\varepsilon(\xi_\varepsilon) = \sum_x d_\varepsilon(\varepsilon x, \xi_\varepsilon). \tag{7.1}$$

Recall from (1.5) and (2.6) that

$$h_i^\varepsilon(x,\xi) = E_Y(g_i^\varepsilon(\xi(x+Y^1),\dots\xi(x+Y^{N_0}))), \tag{7.2}$$

where E_Y denotes the expected value over the distribution of $(Y^1,\dots Y^{N_0})$, and also that

$$c^* = c^*(g) = \sup_{0<\varepsilon\le\varepsilon_0/2} \|g_1^\varepsilon\|_\infty + \|g_0^\varepsilon\|_\infty + 1, \quad c_b = c^* N_0. \tag{7.3}$$

It will be convenient to write

$$\xi_\varepsilon(\varepsilon x + \varepsilon\bar{Y}) = (\xi(x+\varepsilon Y^1),\dots\xi(x+\varepsilon Y^{N_0})).$$

If $\mathcal{H}_t$ is the right-continuous filtration generated by the graphical representation, then

$$|\xi_t^\varepsilon| = |\xi_0^\varepsilon| + M_t^\varepsilon + \int_0^t \psi_\varepsilon(\xi_s^\varepsilon)\,ds, \tag{7.4}$$

where M^ε is a zero mean L^2-martingale. This is easily seen by writing $\xi_t^\varepsilon(x)$ as a solution of a stochastic differential equation driven by the Poisson point processes in the graphical representation and summing over x. The integrability required to show M^ε is a square integrable martingale is readily obtained by dominating $|\xi^\varepsilon|$ by a pure birth process (the rates c_ε are uniformly bounded for each ε) and a square function calculation.

Lemma 7.1. *For any finite stopping time S*

$$e^{-c_b t}|\xi_S^\varepsilon| \le E(|\xi_{S+t}^\varepsilon||\mathcal{H}_S) \le e^{c_b t}|\xi_S^\varepsilon|.$$

Proof. By the strong Markov property it suffices to prove the result when $S = 0$. The fact that $d_\varepsilon(\varepsilon x, \xi_\varepsilon) \ge -\|g_0^\varepsilon\|_\infty$ implies $\psi_\varepsilon(\xi_s^\varepsilon) \ge -\|g_0^\varepsilon\|_\infty|\xi_s^\varepsilon|$. It follows from (1.43) and (7.2) that

$$d_\varepsilon(\varepsilon x, \xi_\varepsilon) \le \|g_1^\varepsilon\|_\infty \sum_{y:\xi(y)=1}\sum_{i=1}^{N_0} P(Y^i = y - x).$$

Summing over x and then y, we get $\psi_\varepsilon(\xi^\varepsilon_s) \le N_0\|g^\varepsilon_1\|_\infty|\xi^\varepsilon_s|$ and (recalling (7.3)) the desired result follows by taking means in (7.4) and using Gronwall's Lemma. □

Let $\xi^{\varepsilon,0}$ be the voter model constructed from the same graphical representation as ξ^ε by only considering the voter flips. We always assume $\xi^{\varepsilon,0}_0 = \xi^\varepsilon_0$.

Lemma 7.2. *If $c_{7.2} = 4(2N_0+1)c^*$ then*

$$E(|\psi_\varepsilon(\xi^\varepsilon_s) - \psi_\varepsilon(\xi^{\varepsilon,0}_s)|) \le c_{7.2}[e^{c^*(N_0+1)s} - 1]|\xi^\varepsilon_0|\,.$$

Proof. Let $\xi^\varepsilon_s(\varepsilon x + \varepsilon\tilde{Y}) = (\xi^\varepsilon_s(\varepsilon x + \varepsilon Y^0), \dots, \xi^\varepsilon_s(\varepsilon x + \varepsilon Y^{N_0}))$, where $Y^0 = 0$, $\tilde{Y}$ is independent of ξ^ε, and note that in contrast to $\bar{Y}$, $\tilde{Y}$ contains 0. Let

$$D_\varepsilon(\eta_0, \eta_1, \dots \eta_{N_0}) = -\eta_0 g^\varepsilon_0(\eta_1, \dots, \eta_{N_0}) + (1-\eta_0)g^\varepsilon_1(\eta_1, \dots, \eta_{N_0}),$$

and note that

$$\begin{aligned} E(|\psi_\varepsilon(\xi^\varepsilon_s) - \psi_\varepsilon(\xi^{\varepsilon,0}_s)|) &\le E\Big(\sum_x |D_\varepsilon(\xi^\varepsilon_s(\varepsilon x + \varepsilon\tilde{Y}) - D_\varepsilon(\xi^{\varepsilon,0}_s(\varepsilon x + \varepsilon\tilde{Y}))|\Big) \\ &\le 2\|D_\varepsilon\|_\infty E\Big(\sum_x [\max_{0\le i\le N_0} \xi^\varepsilon_s(\varepsilon x + \varepsilon Y_i) \vee \xi^{\varepsilon,0}_s(\varepsilon x + \varepsilon Y_i)] \qquad (7.5) \\ &\qquad \times 1\{\xi^\varepsilon_s(\varepsilon x + \varepsilon\tilde{Y}) \ne \xi^{\varepsilon,0}_s(\varepsilon x + \varepsilon\tilde{Y})\}\Big), \end{aligned}$$

because for fixed x if the latter summand is zero, so is the former, and if the latter summand is 1, the former is at most $2\|D_\varepsilon\|_\infty$.

Let $X_t = X^{z,s}_t$, $t \in [0,s]$ be the dual of ξ^ε starting at $(z_0, \dots, z_{N_0}) = \varepsilon x + \varepsilon\tilde{Y}$ at time s and let R_m, $m \ge 1$ be the associated branching times. We claim that

$$\begin{aligned} E\Big(\Big[\max_{0\le i\le N_0} \xi^\varepsilon_s(\varepsilon x + \varepsilon Y^i)\Big]1\{\xi^\varepsilon_s(\varepsilon x + \varepsilon\tilde{Y}) \ne \xi^{\varepsilon,0}_s(\varepsilon x + \varepsilon\tilde{Y})\}\Big) \qquad (7.6) \\ \le E\Big(\sum_{\ell\in J(s)} \xi^\varepsilon_0(X^\ell_s)1\{R_1 \le s\}\Big). \end{aligned}$$

To see this, note that:

(i) if $R_1 > s$, then there are no branching events and so $(X_t, t \le s)$ is precisely the coalescing dual used to compute the rescaled voter model values $\xi^{\varepsilon,0}_s(\varepsilon x + \varepsilon\tilde{Y})$.

(ii) In the case $R_1 \le s$, if $\xi^\varepsilon_0(X^\ell_s) = 0$ for all $\ell \in J(s)$ then $\xi^\varepsilon_s(\varepsilon x + \varepsilon Y^i) = 0$ for $0 \le i \le N_0$ because working backwards from time 0 to time s, we see thanks to (1.43) and (2.19) that no site can flip due to a reaction, and again we have $\xi^\varepsilon(\varepsilon x + \varepsilon\tilde{Y}) = \xi^{\varepsilon,0}(\varepsilon x + \varepsilon\tilde{Y})$.

Similar reasoning and the fact that the dual $(X^{0,j}_t, j \in J^0(t))$ of the voter model $\xi^{\varepsilon,0}$ with the same initial condition z satisfies $J^0(t) \subset J(t)$ for all $t \le s$ a.s., shows that

$$\begin{aligned} E\Big(\Big[\max_{0\le i\le N_0} \xi^{\varepsilon,0}_s(\varepsilon x + \varepsilon Y^i)\Big]1\{\xi^\varepsilon_s(\varepsilon x + \varepsilon\tilde{Y}) \ne \xi^{\varepsilon,0}_s(\varepsilon x + \varepsilon\tilde{Y})\}\Big) \qquad (7.7) \\ \le E\Big(\sum_{\ell\in J(s)} \xi^\varepsilon_0(X^\ell_s)1\{R_1 \le s\}\Big). \end{aligned}$$

If E_0 denotes expectation with respect to the law of $X_t^{z,s}$ when $x = 0$ then, using (7.6) and (7.7), we may bound (7.5) by

$$4c^* E_0\Big(\sum_{\ell \in J(s)} \sum_x \xi_0^\varepsilon(\varepsilon x + X_s^\ell) 1\{R_1 \le s\}\Big)$$

Bounding by the dominating branching random walk $\bar{X}$, and using $|\bar{J}(\bar{R}_1)| = 2N_0+1$ and $P(\bar{R}_1 \le s) = 1 - e^{-c^*(N_0+1)s}$, we see the expected value in the last formula is at most

$$\begin{aligned}&|\xi_0^\varepsilon| E(|\bar{J}(s)| 1\{\bar{R}_1 \le s\}) \le |\xi_0^\varepsilon| e^{c^* N_0 s} E(|\bar{J}(\bar{R}_1)| 1\{\bar{R}_1 \le s\}) \\ &\le (2N_0+1)|\xi_0^\varepsilon| e^{c^* N_0 s}(1 - e^{-c^*(N_0+1)s}) \le (2N_0+1)|\xi_0^\varepsilon|(e^{c^*(N_0+1)s} - 1),\end{aligned}$$

which proves the desired result. $\square$

For the next step in the proof we recall the notation from Section 1.8. We assume Y is independent from the coalescing random walk system $\{\hat{B}^x : x \in \mathbb{Z}^d\}$ used to define $\tau(A)$ and $\tau(A,B)$. Recall from (1.92) and (1.89) that under (1.43)

$$\theta \equiv f'(0) = \sum_{S \in \hat{\mathcal{P}}_{N_0}} \hat{\beta}(S) P(\tau(Y^S) < \infty, \tau(Y^S, \{0\}) = \infty) - \hat{\delta}(S) P(\tau(Y^S \cup \{0\}) < \infty).$$

For $M > 0$ define

$$\theta_M^\varepsilon = \sum_{S \in \hat{\mathcal{P}}_{N_0}} \hat{\beta}_\varepsilon(S) P(\tau(Y^S) \le M < \tau(Y^S, \{0\})) - \hat{\delta}_\varepsilon(S) P(\tau(Y^S \cup \{0\}) \le M)\ .$$

It follows from (1.93) that (with or without the ε's)

$$\sum_{S \in \hat{\mathcal{P}}_{N_0}} |\hat{\beta}_\varepsilon(S)| + |\hat{\delta}_\varepsilon(S)| \le 2^{2N_0}(\|g_1^\varepsilon\|_\infty + \|g_0^\varepsilon\|_\infty) \le 2^{2N_0} c^*(g) \tag{7.8}$$

(recall here that $\tilde{g}_i^\varepsilon = g_i^\varepsilon$ by our $\varepsilon_1 = \infty$ convention). It is clear that $\lim_{M\to\infty, \varepsilon\to 0} \theta_M^\varepsilon = \theta$, but we need information about the rate.

Lemma 7.3. *There is a $\varepsilon_{7.3}(M) \downarrow 0$ (independent of the g_i^ε) so that*

$$|\theta_M^\varepsilon - \theta| \le 2^{2N_0}\Big[\|g_1^\varepsilon - g_1\|_\infty + \|g_0^\varepsilon - g_0\|_\infty + c^*(g)\varepsilon_{7.3}(M)\Big].$$

Proof. Define

$$\varepsilon_{7.3}(M) = \frac{1}{2} \sup_{S \in \hat{\mathcal{P}}_{N_0}} \{P(M < \tau < \infty) : \tau = \tau(Y^S), \tau(Y^S \cup \{0\}), \tau(Y^S, \{0\})\},$$

and note that $\varepsilon_{7.3}(M) \downarrow 0$ as $M \uparrow \infty$. Using (1.94) and (7.8) (the latter without the ε's) we have

$$\begin{aligned}|\theta^\varepsilon_M - \theta| \le &\sum_{S\in\hat{\mathcal{P}}_{N_0}} |\hat\beta_\varepsilon(S) - \hat\beta(S)| + |\hat\delta_\varepsilon(S) - \hat\delta(S)| \\ &+ \sum_{S\in\hat{\mathcal{P}}_{N_0}} \Big[|\hat\beta(S)||P(\tau(Y^S) \le M < \tau(Y^S,\{0\})) - P(\tau(Y^S) < \infty = \tau(Y^S,\{0\}))| \\ &\qquad + |\hat\delta(S)|P(M < \tau(Y^S\cup\{0\}) < \infty) \\ &\le 2^{2N_0}[\|g_1^\varepsilon - g_1\|_\infty + \|g_0^\varepsilon - g_0\|_\infty] + 2^{2N_0}c^*(g)\varepsilon_{7.3}(M).\end{aligned}$$

The result follows. □

To exploit the inequality in Lemma 7.2 we need a good estimate of $E(\psi_\varepsilon(\xi_s^{\varepsilon,0}))$ for small s.

Lemma 7.4. *There is a constant $c_{7.4}$ (independent of g_i^ε) such that for $\varepsilon, \delta > 0$,*

$$E(\psi_\varepsilon(\xi_\delta^{\varepsilon,0})) = \theta^\varepsilon_{\delta\varepsilon^{-2}}|\xi_0^\varepsilon| + \eta_{7.4}(\varepsilon,\delta), \tag{7.9}$$

where $|\eta_{7.4}(\varepsilon,\delta)| \le c_{7.4}c^*(g)\delta^{-d/2}|\xi_0^\varepsilon|^2\varepsilon^d$.

Proof. As usual we assume Y is independent of $\xi^{\varepsilon,0}$. Summability issues in what follows are handled by Lemma 7.1 (and its proof) with $g_i^\varepsilon \equiv 0$. The representation (1.87) and (7.1) imply that

$$\begin{aligned}E(\psi_\varepsilon(\xi_\delta^{\varepsilon,0})) &= \sum_S \hat\beta_\varepsilon(S)E_0^\varepsilon(S) - \hat\delta_\varepsilon(S)E_1^\varepsilon(S) \\ E_0^\varepsilon(S) &= \sum_{x\in\mathbb{Z}^d} E\Big((1-\xi_\delta^{\varepsilon,0}(\varepsilon x))\prod_{i\in S}\xi_\delta^{\varepsilon,0}(\varepsilon x + \varepsilon Y^i)\Big) \\ E_1^\varepsilon(S) &= \sum_{x\in\mathbb{Z}^d} E\Big(\xi_\delta^{\varepsilon,0}(\varepsilon x)\prod_{i\in S}\xi_\delta^{\varepsilon,0}(\varepsilon x + \varepsilon Y^i)\Big).\end{aligned} \tag{7.10}$$

We will use duality between $\xi^{\varepsilon,0}$ and $\{\hat B^x\}$ (see (V.1.7) of [35]) to argue that

$$\begin{aligned}E_0^\varepsilon(S) &\approx |\xi_0^\varepsilon|P(\tau(Y^S) \le \delta\varepsilon^{-2} < \tau(Y^S,\{0\})),\ \emptyset \ne S \subset \{1,\dots,N_0\} \\ E_1^\varepsilon(S) &\approx |\xi_0^\varepsilon|P(\tau(Y^S,\{0\}) \le \delta\varepsilon^{-2}) \text{ all } S \subset \{1,\dots,N_0\}.\end{aligned}$$

Beginning with the first of these, note that duality implies (recall $Y^0 \equiv 0$)

$$\begin{aligned}E_0^\varepsilon(S) &= \sum_{x\in\mathbb{Z}^d} E\Big((1-\xi_0^\varepsilon(\varepsilon\hat B^x_{\delta\varepsilon^{-2}}))\prod_{i\in S}\xi_0^\varepsilon(\varepsilon\hat B^{x+Y^i}_{\delta\varepsilon^{-2}})1\{\tau(x+Y^S,\{x\}) > \delta\varepsilon^{-2}\}\Big) \\ &= \sum_{x\in\mathbb{Z}^d} E\Big(\prod_{i\in S}\xi_0^\varepsilon(\varepsilon\hat B^{x+Y^i}_{\delta\varepsilon^{-2}})1\{\tau(x+Y^S,\{x\}) > \delta\varepsilon^{-2}\}\Big) \\ &\quad - \sum_{x\in\mathbb{Z}^d} E\Big(\prod_{i\in S\cup\{0\}}\xi_0^\varepsilon(\varepsilon\hat B^{x+Y^i}_{\delta\varepsilon^{-2}})1\{\hat\tau(x+Y^S,\{x\}) > \delta\varepsilon^{-2}\}\Big) \\ &\equiv \Sigma_1 - \Sigma_2.\end{aligned} \tag{7.11}$$

If $\tau(x+Y^S) > \delta\varepsilon^{-2}$ there are $i \neq j \in S$ so that $\tau(\{x+Y^i\},\{x+Y^j\}) > \delta\varepsilon^{-2}$. If we condition on the values of the Y^i, Y^j in the next to last line below,

$$\begin{aligned}
&\sum_{x\in\mathbb{Z}^d} E\Big(\prod_{i\in S}\xi_0^\varepsilon(\varepsilon\hat{B}^{x+Y^i}_{\delta\varepsilon^{-2}})1\{\tau(x+Y^S)>\delta\varepsilon^{-2}\}\Big)\\
&\le \sum_{x\in\mathbb{Z}^d}\sum_{1\le i<j\le N_0} E(\xi_0^\varepsilon(\varepsilon\hat{B}^{x+Y^i}_{\delta\varepsilon^{-2}})\xi_0^\varepsilon(\varepsilon\hat{B}^{x+Y^j}_{\delta\varepsilon^{-2}})1\{\tau(\{x+Y^i\},\{x+Y^j\})>\delta\varepsilon^{-2}\})\\
&\le \sum_{w\in\mathbb{Z}^d}\sum_{z\in\mathbb{Z}^d}\xi_0^\varepsilon(\varepsilon w)\xi_0^\varepsilon(\varepsilon z)\sum_{1\le i<j\le N_0}\sum_{x\in\mathbb{Z}^d} P(\hat{B}^{x+Y^i}_{\delta\varepsilon^{-2}}=w,\,\hat{B}^{x+Y^j}_{\delta\varepsilon^{-2}}=z,\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \tau(\{x+Y^i\},\{x+Y^j\})>\delta\varepsilon^{-2}\})\\
&\le \sum_{w\in\mathbb{Z}^d}\sum_{z\in\mathbb{Z}^d}\xi_0^\varepsilon(\varepsilon w)\xi_0^\varepsilon(\varepsilon z)\sum_{1\le i<j\le N_0} P(\hat{B}^0_{2\delta\varepsilon^{-2}}=w-z-Y_i+Y_j)\\
&\le N_0(N_0-1)|\xi_0^\varepsilon|^2 c(1+2\delta\varepsilon^{-2})^{-d/2},
\end{aligned}\tag{7.12}$$

where the local central limit theorem (e.g. (A.7) in [6]) is used in the last line. A similar calculation shows that

$$\Sigma_2 \le |\xi_0^\varepsilon|^2 c(1+2\delta\varepsilon^{-2})^{-d/2}. \tag{7.13}$$

To see this, note that $\tau(x+Y^S,\{0\}) > \delta\varepsilon^{-2}$ implies that for $i_0 \in S$ (this is where we require S non-empty) $\tau(\{x+Y^{i_0}\},\{x\}) > \delta\varepsilon^{-2})$ and we may repeat the above with $i = i_0$ and $j = 0$. Returning to the study of Σ_1, taking any $i_0 \in S$ we have

$$\begin{aligned}
&\sum_{x\in\mathbb{Z}^d} E\Big(\prod_{i\in S}\xi_0^\varepsilon(\varepsilon B^{x+Y^i}_{\delta\varepsilon^{-2}})1\{\tau(x+Y^S)\le\delta\varepsilon^{-2}<\tau(x+Y^S,\{x\})\}\Big)\\
&=\sum_{x\in\mathbb{Z}^d} E\Big(\xi_0^\varepsilon(\varepsilon x+\varepsilon B^{Y^{i_0}}_{\delta\varepsilon^{-2}})1\{\tau(Y^S)\le\delta\varepsilon^{-2}<\tau(Y^S,\{0\})\}\Big)\\
&=|\xi_0^\varepsilon|P(\tau(Y^S)\le\delta\varepsilon^{-2}<\tau(Y^S,\{0\})).
\end{aligned}\tag{7.14}$$

Together (7.12) and (7.14) bound Σ_1. Using this with (7.13) in (7.11), we conclude that

$$E_0^\varepsilon(S) = |\xi_0^\varepsilon|P(\tau(Y^S)\le\delta\varepsilon^{-2}<\tau(Y^S,\{0\})) + \eta_1(\varepsilon,\delta,S), \tag{7.15}$$

where $|\eta_1(\varepsilon,\delta,S)| \le cN_0^2|\xi_0^\varepsilon|^2\delta^{-d/2}\varepsilon^d$. A similar, and simpler, argument shows that for $S \subset \{1,\dots,N_0\}$,

$$E_1^\varepsilon(S) = |\xi_0^\varepsilon|P(\tau(Y^S\cup\{0\})\le\delta\varepsilon^{-2}) + \eta_2(\varepsilon,\delta,S), \tag{7.16}$$

where $|\eta_2(\varepsilon,\delta,S)| \le cN_0(N_0+1)|\xi_0^\varepsilon|^2\delta^{-d/2}\varepsilon^d$.

Now use (7.15), (7.16) and the fact that $\hat{\beta}_\varepsilon(\emptyset) = 0$ (by (1.43)), to obtain (7.9) with

$$|\eta_{7.4}(\varepsilon,\delta)| \le \sum_S(|\hat{\beta}_\varepsilon(S)| + |\hat{\delta}_\varepsilon(S)|)cN_0(N_0+1)\delta^{-d/2}|\xi_0^\varepsilon|^2\varepsilon^d.$$

Finally use (7.8) to complete the proof. □

For $0 < \eta_1 < 1$, let $T(\eta_1) = T_\varepsilon(\eta_1) = \inf\{t \geq \varepsilon^{\eta_1} : |\xi^\varepsilon_{t-\varepsilon^{\eta_1}}| \geq (\varepsilon^{-1+\frac{\eta_1}{2}})^d \varepsilon^{\eta_1}\}$ and note that $T(\eta_1) - \varepsilon^{\eta_1}$ is an $(\mathcal{H}_t)$-stopping time.

Lemma 7.5. *There is a $c_{7.5}$ so that if $\eta_1 \in (0,1)$, then for all $s \geq \varepsilon^{\eta_1}$*

$$E(\psi(\xi^\varepsilon_s)|\mathcal{H}_{s-\varepsilon^{\eta_1}}) \leq [\theta^\varepsilon_{\varepsilon^{\eta_1}\varepsilon^{-2}} + c_{7.5}\varepsilon^{\eta_1}]\,|\xi^\varepsilon_{s-\varepsilon^{\eta_1}}| \quad \text{a.s. on } \{T(\eta_1) > s\}.$$

Proof. Let $\delta = \varepsilon^{\eta_1}$. If $|\xi^\varepsilon_0| \leq (\varepsilon^{-1+\frac{\eta_1}{2}})^d\varepsilon^{\eta_1}$, then Lemmas 7.2 and 7.4 imply $E(\psi(\xi^\varepsilon_\delta)) = |\xi^\varepsilon_0|\theta^\varepsilon_{\delta\varepsilon^{-2}} + \eta'(\varepsilon)$ with

$$\begin{aligned}|\eta'(\varepsilon)| &\leq c_{7.2}[e^{(N_0+1)c^*\varepsilon^{\eta_1}} - 1]|\xi^\varepsilon_0| + c_{7.4}c^*(g)\varepsilon^{-\eta_1 d/2+d}|\xi^\varepsilon_0|^2 \\ &\leq c_{7.2}c^*(N_0+1)e^{(N_0+1)c^*}\varepsilon^{\eta_1}|\xi^\varepsilon_0| + c_{7.4}c^*(g)|\xi^\varepsilon_0|\varepsilon^{\eta_1}.\end{aligned}$$

For the second term we used the bound on $|\xi^\varepsilon_0|$. The result now follows from the above by the Markov property and the definition of $T(\eta_1)$. □

Lemma 7.6. *Let $\beta, \eta_2 \in (0,1]$. There is an $\varepsilon_{7.6}(\beta,\eta_2) \in (0,1)$, so that if $0 < \varepsilon \leq \varepsilon_{7.6}$ and $\theta = f'(0) \leq -\eta_2$, then $|\xi^\varepsilon_0| \leq \varepsilon^{-d+\beta}$ implies*

$$P(|\xi^\varepsilon_t| > 0) \leq 6e^{2c_b}\varepsilon^{\beta/2} \quad \text{for all } t \geq \tfrac{2d}{\eta_2}\log(1/\varepsilon).$$

Proof. Let $\lambda \leq \eta_2/2$, $\eta_1 = \beta(2+d)^{-1}$, $T = T_\varepsilon(\eta_1)$ and $\delta = \varepsilon^{\eta_1}$. An integration by parts using (7.4) shows that for $t \geq \delta$,

$$e^{\lambda(t\wedge T)}|\xi^\varepsilon_{t\wedge T}| = e^{\lambda\delta}|\xi^\varepsilon_\delta| + \int_\delta^t 1\{r < T\}[\lambda e^{\lambda r}|\xi^\varepsilon_r| + e^{\lambda r}\psi_\varepsilon(\xi^\varepsilon_r)]dr + N^\varepsilon_t,$$

where N^ε is a mean 0 martingale. Since $\{r < T\} \in \mathcal{H}_{r-\delta}$, we have for $\delta \leq s \leq t$

$$\begin{aligned}&E(e^{\lambda(t\wedge T)}|\xi^\varepsilon_{t\wedge T}| - e^{\lambda(s\wedge T)}|\xi^\varepsilon_{s\wedge T}||\mathcal{H}_{s-\delta}) \\ &\quad = \int_s^t E(1\{r<T\}e^{\lambda r}E(\lambda|\xi^\varepsilon_r| + \psi_\varepsilon(\xi^\varepsilon_r)|\mathcal{H}_{r-\delta})|\mathcal{H}_{s-\delta})dr.\end{aligned}$$

Using Lemmas 7.1 and 7.5 the above is at most

$$E\Big(\int_s^t 1\{r<T\}e^{\lambda r}\gamma(\varepsilon)|\xi^\varepsilon_{r-\delta}|dr|\mathcal{H}_{s-\delta}\Big), \tag{7.17}$$

where $\gamma(\varepsilon) = \lambda e^{c_b\delta} + \theta^\varepsilon_{\delta\varepsilon^{-2}} + c_{7.5}\delta$. Recall $\delta = \varepsilon^{\eta_1}$ and $\theta = f'(0) \leq -\eta_2$. By Lemma 7.3 and the uniform convergence of the g^ε_i to g_i there is a $\varepsilon_1(\beta,\eta_2) > 0$ so that if $0 < \varepsilon \leq \varepsilon_1$, then

$$\begin{aligned}\gamma(\varepsilon) &\leq \frac{\eta_2}{2}e^{c_b\varepsilon^{\eta_1}} - \eta_2 + 2^{2N_0}\Big[\sum_{i=0}^1 \|g^\varepsilon_i - g_i\|_\infty\Big] + 2^{2N_0}c^*\varepsilon_{7.3}(\varepsilon^{\eta_1}\varepsilon^{-2}) + c_{7.5}\varepsilon^{\eta_1} \\ &\leq -\eta_2/4 < 0.\end{aligned}$$

We assume $0 < \varepsilon \le \varepsilon_1$ in what follows. Since the bound in (7.17) is therefore non-positive and our assumption on $|\xi_0^\varepsilon|$ implies $T > \delta$, we may use Lemma 7.1 and the fact that $\delta \le 1$ to see that for $t \ge \delta$,

$$E(|\xi^\varepsilon_{t\wedge T}|e^{\lambda(t\wedge T)}) \le e^{\lambda\delta}E(|\xi^\varepsilon_\delta|) \le e^{(\lambda+c_b)\delta}|\xi^\varepsilon_0| \le e^{\eta_2+c_b}\varepsilon^{\beta-d}. \tag{7.18}$$

Now $|\xi^\varepsilon_t| \ge 1$ if it is positive so

$$P(|\xi^\varepsilon_t| > 0) \le E(|\xi^\varepsilon_{t\wedge T}|e^{\lambda(T\wedge t)}1\{T \ge t\})e^{-\lambda t} + P(T < t). \tag{7.19}$$

Let $t \ge (2d/\eta_2)\log(\varepsilon^{-1})$ and use (7.18) with $\lambda = \eta_2/2$ to see that the first term is at most

$$e^{\eta_2+c_b}\varepsilon^{\beta-d}\varepsilon^d = e^{\eta_2+c_b}\varepsilon^\beta. \tag{7.20}$$

To bound $P(T < t)$, we note that $|\xi^\varepsilon_{T-\delta}| \ge (\varepsilon^{-1+\frac{\eta_1}{2}})^d\varepsilon^{\eta_1}$ if $T < \infty$, so

$$P(T < t) \le E(|\xi^\varepsilon_{T-\delta}|1\{T < t\})(\varepsilon^{-1+\frac{\eta_1}{2}})^{-d}\varepsilon^{-\eta_1}.$$

By making ε_1 smaller, depending on β, we can assume that $(2d/\eta_2)\log(\varepsilon^{-1}) \ge \varepsilon^{\eta_1} = \delta$. Let $S = (T-\delta)\wedge(t-\delta)$, note $\{T < t\} \in \mathcal{H}_S$, and use the lower bound in Lemma 7.1 to conclude the first inequality in

$$E(|\xi^\varepsilon_{T-\delta}|1\{T < t\}) \le e^{c_b\delta}E(|\xi^\varepsilon_{T\wedge t}|) \le e^{2c_b+\eta_2}\varepsilon^{\beta-d}.$$

The second inequality comes from (7.18) with $\lambda = 0$ (recall that $t \ge 2d/\eta_2)\log(\varepsilon^{-1}) \ge \varepsilon^{\eta_1} = \delta$) and $\delta \le 1$. Using the last two equations with (7.20) in (7.19), we conclude that

$$\begin{aligned}P(|\xi^\varepsilon_t| > 0) &\le e^{\eta_2+c_b}\varepsilon^\beta + e^{2c_b+\eta_2}\varepsilon^{\beta-d}(\varepsilon^{-1+\frac{\eta_1}{2}})^{-d}\varepsilon^{-\eta_1}\\ &\le e^{\eta_2+2c_b}[\varepsilon^\beta + \varepsilon^{\beta-\eta_1(1+\frac{d}{2})}] \le 2e^{1+2c_b}\varepsilon^{\beta/2},\end{aligned}$$

where the definition of η_1 is used in the last line. The result follows. $\square$

7.2 The Dead Zone

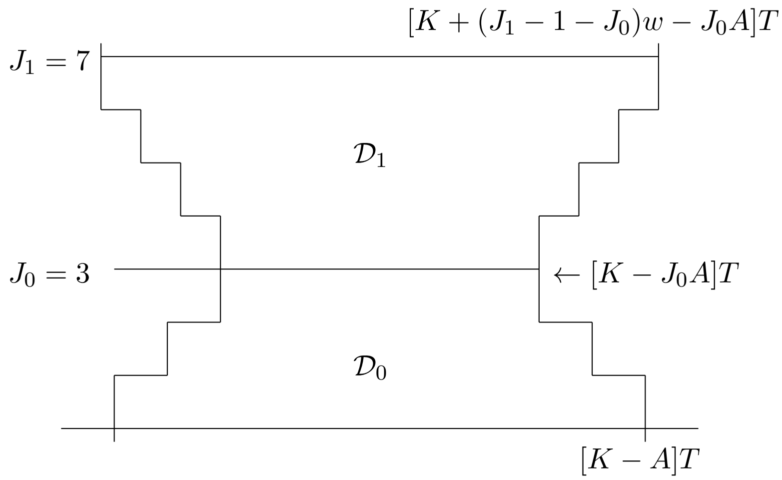
For the remainder of this Section we suppose (1.41), (1.43) and Assumption 2 are in force and $-f'(0) \ge \eta_2 \in (0,1]$. We also assume that $p(\cdot)$ and $q(\cdot)$ have finite supports. More specifically, $R_0 \in \mathbb{N}$ satisfies

$$\{x \in \mathbb{Z}^d : p(x) > 0\} \subset [-R_0, R_0]^d \text{ and } \{x \in \mathbb{Z}^{dN_0} : q(x) > 0\} \subset [-R_0, R_0]^{dN_0}. \tag{7.21}$$

In order to connect with the percolation results from Section 5 we need certain space-time regions suitable for applying Lemma 4.2 to decrease particle density, Lemma 4.7 to control the spread of duals, and Lemma 7.6 to actually kill off particles. Recall that $Q^\varepsilon(r) = [-r,r]^d \cap (\varepsilon\mathbb{Z}^d)$. For $J_0 < J_1 \in \mathbb{N}, 0 < w < 1, A, K > 1$, and $T > 0$ define regions $\mathcal{D}(J_0, J_1, w, A, K) = \mathcal{D}_0 \cup \mathcal{D}_1$, where

$$\begin{aligned}\mathcal{D}_0 &= \cup_{j=1}^{J_0}(Q^\varepsilon((K-jA)T) \times [(j-1)T, jT]),\\ \mathcal{D}_1 &= \cup_{j=J_0}^{J_1-1}(Q^\varepsilon((K+jw-(w+A)J_0)T) \times [jT,(j+1)T]).\end{aligned}$$

For help with the definition consult the following picture:



The speed $w > 0$ is as in Assumption 2 (and may be assumed to be < 1), and $T = A_{4.2} \log(1/\varepsilon)$ is the same as in (4.6). For the regions $\mathcal{D}_0, \mathcal{D}_1, \mathcal{E}$ we take

$$J_0 = \left\lceil \frac{2d}{\eta_2 A_{4.2}} \right\rceil + 1, \quad A = \frac{8(2d+3))}{A_{4.2}} \vee (2c_b \vee 2\sigma^2)$$
$$K = 2 + AJ_0, \text{and } J_1 = J_0 + 1 + \left\lceil \frac{K + AJ_0}{w} \right\rceil. \tag{7.22}$$

The choice $K = 2 + AJ_0$ implies that $Q^\varepsilon(2T) \times \{J_0 T\}$ is the "top" of $\mathcal{D}_0$ and the "bottom" of $\mathcal{D}_1$. The choice of J_1 implies

the top of $\mathcal{D}$ contains $Q^\varepsilon(2KT) \times \{J_1 T\}$ and is contained in $Q^\varepsilon((2K+1)T) \times \{J_1 T\}$, and $\mathcal{D}$ contains the region $Q^\varepsilon(2T) \times [0, J_1 T]$. (7.23)

Recall from Section 5 that a region $\mathcal{C}$ in $\mathbb{R}^d \times \mathbb{R}_+$ is ε-empty iff $\xi_t^\varepsilon(x) = 0$ for all $(t, x) \in \mathcal{C}$, where ξ^ε is our voter model perturbation as usual. If $A \subset \mathbb{R}^d$ let $\xi_0^\varepsilon(A) = \sum_{x \in A \cap \varepsilon\mathbb{Z}^d} \xi_0^\varepsilon(x)$.

Lemma 7.7. *With T, w, J_0, J_1, A, K as above there exist $\varepsilon_{7.7}, c_{7.7} > 0$ depending on u_1, u_2, w, c_2, C_2 (from Assumption 2) and $r_0, \gamma_{4.2}, \eta_2$ such that such that if $0 < \varepsilon \le \varepsilon_{7.7}$ and*

$$\xi_0^\varepsilon(Q(KT)) = 0, \tag{7.24}$$

then

$$P\Big(\mathcal{D}(J_0, J_1, w, A, K) \text{ is } \varepsilon\text{-empty}\Big) \ge 1 - c_{7.7}\varepsilon^{.01 \wedge \frac{\gamma_{4.2}}{4}}, \tag{7.25}$$

and with probability at least $1 - c_{7.7}\varepsilon^d$,

$$\text{for all } j = 1, \ldots, J_1, (x, u) \in (Q^\varepsilon((K + J_1 w + A(J_1 - j))T) \times [(j-1)T, jT])$$
$$\text{and } t \in [0, u - (j-1)T], X_t^{x,u} \subset Q^\varepsilon((K + J_1 w + A(J_1 - j + 1))T). \tag{7.26}$$

Proof. We begin with some notation for describing events in which the dual process is confined to certain space-time regions. For $j \ge 1$ and $0 < r < s$ let $\Gamma_T(j,r,s)$ be the event

$$X_t^{x,u} \subset Q^\varepsilon(sT)\ \forall\ x \in Q^\varepsilon(rT),\ \ u \in [(j-1)T, jT],\ \text{and } t \in [0, u-(j-1)T].$$

On $\Gamma_T(j,r,s)$, duality and (1.43) imply

$$\xi^\varepsilon_{(j-1)T}(Q^\varepsilon(sT)) = 0 \text{ implies } Q^\varepsilon(rT) \times [(j-1)T, jT] \text{ is } \varepsilon\text{-empty}. \tag{7.27}$$

Step 1. We first check that $\mathcal{D}_0$ is empty with high probability. For $j \in \{1, \dots J_0\}$ we bound the probability of $\Gamma_T(j, K-jA, K-(j-1)A)$ by using Lemma 4.7. If we set $c = A_{4.2}$, $U = jT$, $L = (K-jA)T$ and $2b = A$, then evaluating q in the lemma we obtain

$$q = (\frac{AA_{4.2}}{8} - 2) \wedge A_{4.2}\varepsilon^{-2} \ge 2d+1$$

if $A_{4.2}\varepsilon^{-2} \ge (2d+1)$. Hence the bound on $\bar{p}_\varepsilon$ in Lemma 4.7 gives us

$$P(\Gamma_T(j, K-jA, K-(j-1)A)) \ge 1 - c_d'(K-jA)^d\varepsilon^d \tag{7.28}$$

for $\varepsilon < \varepsilon_{4.6}(A/2)$ such that $A_{4.2}\varepsilon^{-2} \ge 2d+1$ and $\varepsilon(A_{4.2}\log(1/\varepsilon)+1)^d \le 1$.

By (7.27), on the intersection

$$\cap_{j=1}^{J_0} \Gamma_T(j, K-jA, K-(j-1)A),$$

for each $j \in \{1, \dots J_0\}$, if $\xi^\varepsilon_{(j-1)T}(Q^\varepsilon((K-(j-1)A)T)) = 0$ then $Q^\varepsilon((K-jA)T) \times [(j-1)T, jT]$ is ε-empty. Iterating this, (7.24) and (7.28) imply that for some positive ε_0,

$$P\left(\mathcal{D}_0 \text{ is } \varepsilon\text{-empty}\right) \ge 1 - c_d' J_0 K^d \varepsilon^d \text{ if } \varepsilon < \varepsilon_0. \tag{7.29}$$

Here, and throughout the proof, ε_0 will denote a positive constant depending only on our fixed parameters including r_0.

Step 2. By taking ε small enough we may assume that (recall L_0 is as in Assumption 2)

$$2 + L_0 \le KT \le (K + wJ_1)T \le \varepsilon^{-.001/d}. \tag{7.30}$$

For $j \in \{1, \dots, J_1 - J_0\}$, on account of (7.24), we may apply Lemma 4.2 and the Markov property $J_1 - J_0$ times and conclude that for $\varepsilon < \varepsilon_{4.2}$,

$$\begin{aligned} &P(\xi^\varepsilon_{jT} \text{ has density at most } \varepsilon^{\gamma_{4.2}} \text{ in } Q^\varepsilon((K+wj)T) \text{ for } j = 1, \dots, J_1 - J_0)) \\ &\quad \ge 1 - (J_1 - J_0)\varepsilon^{.01}. \end{aligned}$$

When the above event occurs, for any $j \in \{1, 2, \ldots, J_1 - J_0\}$, (recall that $Q_\varepsilon = [0, a_\varepsilon)^d \cap (\varepsilon\mathbb{Z}^d)$)

$$\begin{aligned}
&\xi^\varepsilon_{jT}(Q^\varepsilon((K + wj)T)) \\
&\quad \leq \sum_{x \in a_\varepsilon \mathbb{Z}^d \cap Q^\varepsilon((K+wj)T)} \xi^\varepsilon_{jT}(x + Q_\varepsilon) + \sum_{x \in Q^\varepsilon((K+wj)T) - Q((K+wj)T - a_\varepsilon)} \xi^\varepsilon_{jT}(x) \qquad (7.31)\\
&\quad \leq |Q_\varepsilon| \varepsilon^{\gamma_{4.2}} \mathrm{card}(a_\varepsilon \mathbb{Z}^d \cap Q^\varepsilon((K + wj)T) + c_d \varepsilon^{-d}[(K + wj)T]^{d-1} a_\varepsilon \\
&\quad \leq c_d |Q_\varepsilon| \varepsilon^{\gamma_{4.2}} a_\varepsilon^{-d}((K + wj)T)^d + c_d \varepsilon^{-d}[(K + wj)T]^{d-1} a_\varepsilon \\
&\quad \leq c_d (K + wJ_1)^d A^d_{4.2} (\log(\varepsilon^{-1}))^d [\varepsilon^{\gamma_{4.2} - d} + \varepsilon^{(1/16d) - d}] \\
&\quad \leq \varepsilon^{\gamma_{4.2}/2 - d}, \qquad (7.32)
\end{aligned}$$

for small enough ε, where we have used $\gamma_{4.2} \leq (16d)^{-1}$ in the last line. We have shown that for all ε smaller than some positive ε_0,

$$P(\xi^\varepsilon_{jT}(Q^\varepsilon((K+wj)T)) \leq \varepsilon^{\gamma_{4.2}/2-d} \text{ for } j = 1, \ldots, J_1 - J_0) \geq 1 - (J_1 - J_0)\varepsilon^{.01}. \quad (7.33)$$

Step 3. Fix $j \in \{1, \ldots, J_1 - J_0\}$, and define $(\hat{\xi}^{j,\varepsilon}_t, t \geq jT)$ by setting

$$\hat{\xi}^{j,\varepsilon}_{jT}(x) = \begin{cases} \xi^\varepsilon_{jT}(x) & \text{if } x \in Q^\varepsilon((K + wj)T), \\ 0 & \text{otherwise,} \end{cases}$$

and then using our Poisson processes $\{\Lambda^x_r, \Lambda^x_w : x \in \varepsilon\mathbb{Z}^d\}$ to continue constructing $\hat{\xi}^{j,\varepsilon}_t$ in the same way as ξ^ε_t is constructed. By Lemma 7.6, if $\varepsilon < \varepsilon_{7.6}(\gamma_{4.2}/2, \eta_2)$,

$$\begin{aligned}
&\xi^\varepsilon_{jT}(Q^\varepsilon((K + jw)T) \leq \varepsilon^{(\gamma_{4.2}/2) - d} \text{ implies} \\
&\qquad\qquad P(|\hat{\xi}^{j,\varepsilon}_t| > 0 | \xi^\varepsilon_{jT}) \leq 6e^{2c_b} \varepsilon^{\gamma_{4.2}/4} \text{ for all } t \geq (j + J_0 - 1)T. \qquad (7.34)
\end{aligned}$$

Using $\hat{\xi}^{j,\varepsilon}_t$, we will show that with high probability,

$$\begin{aligned}
&\xi^\varepsilon_{jT}(Q^\varepsilon((K + jw)T) \leq \varepsilon^{(\gamma_{4.2}/2) - d} \text{ implies} \\
&\qquad\qquad Q^\varepsilon((K + jw - J_0 A)T) \times [(j - 1 + J_0)T, (j + J_0)T] \text{ is } \varepsilon\text{-empty.} \qquad (7.35)
\end{aligned}$$

To do this, define the event

$$\Gamma_T(j) = \cap_{i=1}^{J_0} \Gamma_T(j + i, K + wj - iA, K + wj - (i - 1)A).$$

Using Lemma 4.7 as in Step 1 we have for small enough ε

$$P(\Gamma_T(j + i, K + wj - iA, K + wj - (i - 1)A)) \geq 1 - c'_d(K + wj - iA)^d \varepsilon^d$$

and thus

$$P(\Gamma_T(j)) \geq 1 - c'_d(3K)^d J_0 \varepsilon^d \qquad (7.36)$$

for small enough ε.

Observe that on the event $\Gamma_T(j)$ we have

$$X_t^{x,u} \subset Q^\varepsilon((K+jw)T) \; \forall \; x \in Q^\varepsilon((K+jw-J_0A)T),$$
$$u \in [(j-1+J_0)T,(j+J_0)T], \text{ and } t \in [0,u-jT]. \quad (7.37)$$

Therefore, by duality, on $\Gamma_T(j)$,

$$\xi_t^\varepsilon(x) = \hat{\xi}_t^{j,\varepsilon}(x) \text{ for all}$$
$$(x,t) \in Q^\varepsilon((K+jw-J_0A)T) \times [(j-1+J_0)T,(j+J_0)T].$$

Combining this observation with (7.34) and (7.36) we see that the event in (7.35) has probability at least

$$1 - 6e^{2c_b}\varepsilon^{\gamma_{4.2}/4} - c_d'(3K)^d J_0 \varepsilon^d$$

for ε smaller than some ε_0.

Step 4 We can now sum the last estimate over $j = 1,\dots,J_1-J_0$ and use (7.33) to obtain

$$P(\mathcal{D}_1 \text{ is } \varepsilon\text{-empty}) \geq 1 - J_1(\varepsilon^{.01} + 6e^{2c_b}\varepsilon^{\gamma_{4.2}/4} + c_d'(3K)^d J_0\varepsilon^d) \quad (7.38)$$

for small enough ε. (Actually we get a slightly larger set than $\mathcal{D}_1$.) Combine (7.29) and (7.38) to obtain (7.25).

Step 5 Finally, using the notation from Step 1, the event in (7.26) is just

$$\cap_{j=1}^{J_1}\Gamma_T(j, K+wJ_1+(J_1-j)A, K+wJ_1+(J_1-j+1)A).$$

As in Step 1, we can use Lemma 4.7 to bound the probability of this intersection by $1 - c_d' J_1(K+J_1(w+A))^d\varepsilon^d$ for small enough ε, so we are done. $\square$

Let $K_1 = K + J_1(w+A)$. For $\xi \in \{0,1\}^{\varepsilon\mathbb{Z}^d}$, let G_ξ^ε be the event, depending on our graphical representation, on which $\mathcal{D} = \mathcal{D}(J_0,J_1,w,A,K)$ is ε-empty if $\xi_0^\varepsilon = \xi$, and on which (7.26) holds. Note that (7.26) implies all the duals starting at $(x,u) \in \mathcal{D}$ and run up until time u remain in $Q(K_1T)$. Hence duality implies that G_ξ^ε is $\mathcal{G}(Q(K_1T)\times[0,J_1T])$-measurable, where we recall from (6.7) that $\mathcal{G}(R)$ is the σ-field generated by the Poisson points in the graphical representation in the region R. By the inclusion (7.23) we have

$$\text{on } G^\varepsilon_{\xi_0^\varepsilon},\; Q^\varepsilon(2T)\times[0,J_1T] \text{ is } \varepsilon - \text{ empty, and } \xi^\varepsilon_{J_1T}(Q^\varepsilon(2KT)) = 0, \quad (7.39)$$

providing that $\xi_0^\varepsilon \in H = \{\xi \in \{0,1\}^{\varepsilon\mathbb{Z}^d} : \xi(Q^\varepsilon(KT)) = 0\}$. Adding the bounds in Lemma 7.7 we see that

$$\text{if } \xi \in H, \text{ then } P(G_\xi^\varepsilon) \geq 1 - 2c_{7.7}\varepsilon^{.01\wedge\frac{\gamma_{4.2}}{2}} \text{ if } \varepsilon < \varepsilon_{7.7}. \quad (7.40)$$

7.3 Proof of Theorem 1.5

Proof of Theorem 1.5. We continue to take $T = A_{4.2}\log(1/\varepsilon)$, and with K, J_1 from (7.22) we define

$$L = T, \quad T' = J_1 T,$$

and set $c_L = L/(2D)$ as before. We set $\bar\xi^\varepsilon(y) = 1(|y| > L)$, $y \in \varepsilon\mathbb{Z}^d$, and σ_z, $z \in \varepsilon\mathbb{Z}^d$ denote the translation operators on $\{0,1\}^{\varepsilon\mathbb{Z}^d}$. For $(x, n) \in \mathcal{L}_D$ let

$$\xi^{x,n} = \begin{cases} \sigma_{-c_L x}(\xi^\varepsilon_{nT'}) & \text{if } \sigma_{-c_L x}(\xi^\varepsilon_{nT'}) \in H \\ \bar\xi^\varepsilon & \text{otherwise,} \end{cases}$$

and define the percolation variables

$$\eta(x,n) = 1(G^\varepsilon_{\xi^{x,n}} \text{ occurs in the graphical representation in which the Poisson processes are translated by } -c_L x \text{ in space and } -nT' \text{ in time}). \tag{7.41}$$

In the percolation argument which follows it is the first part of the definition of $\xi^{x,n}$ that will matter; the $\bar\xi^\varepsilon$ is really only a place-holder which allows us to define η when the translated configuration is not in H. As in the proof of Theorem 1.4 in Section 6, we are actually translating in space by the "lower left hand corner" in $\varepsilon\mathbb{Z}^d$ associated with $-c_L x$ and as before suppress this in our notation. In Section 6 we used Theorem 4.3 of [14]; here we copy the key definition in its proof. Using the measurability of G^ε_ξ, the independence of $\mathcal{G}(R)$ for disjoint regions R, and (7.40) one can check that for any any $M > 4DK_1$, the family $\{\eta(z), z \in \mathcal{L}_D\}$ satisfies the modified M-dependent condition (5.5) with $\theta = 2c_{7.7}\varepsilon^{.01\wedge\frac{\gamma_{4.2}}{4}}$. To see this argue exactly as in the proof of Theorem A.4 of [14].

Using the percolation results from Section 5, we will show

Lemma 7.8. *There exists $\bar r > 0$ such that for ε small enough $\xi^\varepsilon_0(Q^\varepsilon(KT)) = 0$ implies*

$$P(\Lambda) \equiv P(\xi^\varepsilon_t(Q^\varepsilon(\bar r t)) = 0 \text{ for all large } t) > 1/2.$$

Intuitively this is an immediate consequence of Lemmas 5.2 and 5.6. The first result implies that on Ω^0_∞ then for large n, the wet sites satisfy $W^0_n \cap \mathcal{H}^{rn}_n = \bar W_n \cap \mathcal{H}^{rn}_n$. The second result shows that if $\mathcal{B}_n$ is the collection of dry sites in $\mathcal{H}^{rn/4}_n$ connected to the complement of $\cup^n_{m=n/2}\mathcal{H}^{rm/2}_m$ by a path of dry sites on the graph with edge set $\mathcal{E}_\downarrow$ then $\mathcal{B}_n = \emptyset$ eventually. Wet sites in $\mathcal{H}^{rn}_n$ will correspond to space-time blocks that are empty of 1's while dry sites (i.e. not wet sites) in $\mathcal{H}^{rn}_n$ correspond to space-time blocks which *may* contain a 1. If a dry site in $\mathcal{H}^{rn/4}_n$ corresponds to a block containing a 1 there must be a dual path of 1's leading from this 1 to a site outside of $\cup^n_{m=n/2}\mathcal{H}^{rm/2}_m$. This corresponds to a path of dry sites in $\mathcal{E}_\downarrow$ and so cannot happen for large n since $\mathcal{B}_n = \emptyset$ for large n. Thinking of the corresponding space time regions are being filled with concrete, and the dry sites as air spaces, we see that there cannot be a 1 in $\mathcal{H}^{rn/4}_n$ unless some air space reaches outside of $\cup^n_{m=n/2}\mathcal{H}^{rm/2}_m$. We now give a formal proof.

Proof. Fix ξ_0^ε as above and recall $W_n^0, \bar{W}_n, \mathcal{B}_n$ and Ω_∞^0 from Section 5. In particular $W^0, \bar{W}$ are constructed from an iid Bernoulli field which is bounded above by $\eta(z,n),\ z \in \mathcal{H}_n, n \geq 1$. By (7.40) and our condition on ξ_0^ε (which implies $\xi^{0,0} = \xi_0^\varepsilon$ in the definition of $\eta(0,0) = 1$) we see that

$$P(\eta(0,0) = 1) \geq 1 - 2c_{7.7}\varepsilon^{.01\wedge(\gamma_{4.2}/4)} \geq 3/4,$$

for ε small enough. By working with $P(\cdot|\eta(0,0) = 1)$ in place of P we may assume $\eta(0,0) \equiv 1$ at a cost of proving (under our new P) that

$$P(\Lambda) > 3/4. \tag{7.42}$$

Assume $n \geq 1$ and $(y,n) \in W_n^0$. Then for some i, letting $y' = y - v_i'$, $(y', n-1) \in W_{n-1}^0$ with $\eta(y', n-1) = 1$ (if $n = 1$ we use $\eta(0,0) \equiv 1$ here). Continue to trace back the set of open sites $y' = y_{n-1}', \dots, y_0' = 0$. Proceeding through the y_i' values, using the second part of (7.39) and $c_L v_i' + [-L, L]^d \subset [-2L, 2L]^d$ for $i = 1, \dots, D$, we see that $\xi^{y_i', i} = \sigma_{-c_L y_i'}(\xi_{iT'}^\varepsilon)$ in the definition of $\eta(y_i', i) = 1$. Therefore (7.39) and translation invariance, show that $\eta(y', n-1) = 1$ implies

$$\xi_t^\varepsilon(c_L y' + Q^\varepsilon(2L)) = 0 \text{ for all } t \in [(n-1)T', nT'].$$

Since $c_L y + Q^\varepsilon(L) \subset c_L y' + Q^\varepsilon(2L)$ we obtain

$$(y,n) \in W_n^0 \text{ implies } \xi_t^\varepsilon(c_L y + Q^\varepsilon(L)) = 0 \text{ for all } t \in [(n-1)T', nT']. \tag{7.43}$$

This confirms (5.9) in Section 5.

Next by Lemma 5.2 we may assume ε is small enough (independent of the choice of ξ_0^ε) so that $P(\Omega_\infty^0) > 3/4$ and $\theta < \theta_{5.2} \wedge \theta_{5.5}$. Let $r = r_{5.2}$ and assume $\omega \in \Omega_\infty^0$. By Lemma 5.2 there is an $n_0 \in \mathbb{N}$ so that

$$W_k^0 \cap \mathcal{H}_k^{rk} = \bar{W}_k \cap \mathcal{H}_k^{rk} \ \forall\ k \geq n_0. \tag{7.44}$$

Let $\bar{r} = \frac{r}{16DJ_1}$ and assume $\omega \notin \Lambda$. The latter implies that for infinitely many $n > 2n_0$ there are $t \in [(n-1)T', nT']$ and $x \in Q^\varepsilon(\bar{r}t)$ with $\xi_t(x) = 1$. We claim that this implies

$$\mathcal{B}_n \neq \emptyset \text{ for } n \text{ as above}. \tag{7.45}$$

Lemma 5.6 implies the above is a null set, so it follows that $P(\Omega_\infty^0 \setminus \Lambda) = 0$ and so (7.42) would be proved (recall $P(\Omega_\infty^0) > 3/4$).

To prove (7.45), fix such an n and x, t and consider the dual starting at (x,t). We are going to make use of (1.43), which ensures that not all of the inputs to the dual can be 0. By (5.2) there must exist some $(y,n) \in \mathcal{H}_n$ such that $x \in c_L \mathcal{V}_y \subset c_L y + Q^\varepsilon(L)$ and a bit of arithmetic using the definition of $\bar{r}$ gives

$$|y| \leq \frac{|x|}{c_L} + \frac{L}{c_L} \leq \frac{\bar{r}nT'}{c_L} + 2D \leq \frac{rn}{4},$$

and we have taken n_0 big enough for the last inequality. Hence $(y,n) \in \mathcal{H}_n^{rn/4}$ and so the fact that $\xi_t(x) = 1$, (7.43) and (7.44) imply $(y,n) \notin \bar{W}_n$, i.e., (y,n) is dry. By duality and the finite range assumption (recall (7.21)), there must exist $x' \in \varepsilon\mathbb{Z}^d$ and $t' \in [(n-1)T', t)$ such that $|x-x'| \le R_0\varepsilon$ and $\xi^\varepsilon_{t'}(x') = 1$. That is, $t' \ge (n-1)T'$ is the first time below t that the dual jumps or $t' = (n-1)T'$ if there is no such time in which case $x' = x$. We may assume ε is small enough so that $R_0\varepsilon/c_L \le c_{5.3}$, in which case by (5.3) $x' \in c_L\mathcal{V}_{y'}$ for some y' of the form $y + v'_i - v'_j$ ($y = y'$ is included). If $(y', n) \in \mathcal{H}_n^{rn/2} \subset \mathcal{H}_n^{rn}$, it follows from (7.43) that (y', n) must be dry, and thus $(y', n) \in D_{(y,n)}$.

Continue the above construction until either we reach a point $(y'', n) \in (\mathcal{H}_n^{rn/2})^c$ with all earlier points in our path from (y, n) being dry, or we obtain x'', y'' such that $\xi^\varepsilon_{(n-1)T'}(x'') = 1$, $(y'', n) \in D_{(y,n)} \cap \mathcal{H}_n^{rn/2}$ and $x'' \in c_L\mathcal{V}_{y''}$. In the former case $\mathcal{B}_n \neq \emptyset$ (recall the precise definition prior to Lemma 5.6). In the latter case if $(y'' - v'_i, n-1) \notin \mathcal{H}_{n-1}^{r(n-1)/2}$ for some i, then (7.45) holds. If not, then as one easily checks $|c_L(y'' - v'_i) - x''| < L$, and so arguing as above, we see that $(y'' - v'_i, n-1)$ is dry. Therefore the iteration can be continued until it stops as above or continues down to time $(\frac{n}{2} - 1)T'$, again forcing (7.45) in either case. □

Having established Lemma 7.8 the rest of the proof of Theorem 1.5 is routine. The proof of Lemma 6.5 shows that if we start from an initial configuration with infinitely many 0's then at time 1 there will be infinitely many cubes of the form $c_Lx + Q^\varepsilon(L)$ with $x \in \mathcal{H}_0$ that are ε-empty. By the Markov property this will hold at all times $N \in \mathbb{N}$ a.s. The above shows that if x_0 is chosen so that $\xi^\varepsilon_1(x_0 + Q(L)) = 1$, then w. p. at least $1/2$, $\xi^\varepsilon_{1+t} \equiv 0$ on $c_Lx_0 + Q(\bar{r}t)$ for all large t. If this fails at some time we can try again at a later time N by the above and after a geometric (1/2) number of trials we will succeed and produce a linearly growing set of 0's starting at some space-time location. Therefore the 0's take over. □

Acknowledgement. We thank an anonymous referee for an extremely careful and detailed report which caught a number of misprints and suggested a number of improvements.

References

[1] Aronson, D.G. and Weinberger, H.F. (1975) Nonlinear diffusion in population genetics, combustion, and nerve pulse propagation. Pages 5–49 in *Partial Differential Equations and Related Topics*. Springer Lecture Notes in Math 446, Springer, New York.

[2] Aronson, D.G. and Weinberger, J.F. (1978) Multidimensional nonlinear diffusion arising in population genetics. *Adv. in Math.* **30** 33-76.

[3] Bramson, M. (1983) The convergence of solutions of the Kolmogorov nonlinear diffusion equation to traveling waves. *Memoirs Amer. Math. Soc.* **85** 190 pages.

[4] Cox, J.T. and Durrett, R. (1991) Nonlinear voter models, in *Random Walks, Brownian Motion and Interacting Particle Systems*, R. Durrett and H. Kesten, editors, pages 203-213, Birkhauser.

[5] Cox, J.T., Merle, M. and Perkins, E.A. (2009) Co-existence in a two-dimensional Lotka-Volterra model. *Elect. J. Probab.* **15** 1190-1266.

[6] Cox, J.T., Durrett, R. and Perkins, E.A. (2000) Rescaled voter models converge to super- Brownian motion. *Ann. Probab.* **28** 185-234.

[7] Cox, J.T. and Perkins, E.A. (2005) Rescaled Lotka-Volterra models converge to super-Brownian motion. *Ann. Probab.* **33** 904-947.

[8] Cox, J.T. and Perkins, E.A. (2007) Survival and coexistence in stochastic spatial Lotka-Volterra models. *Prob. Theory Rel. Fields* **139** 89-142.

[9] Cox, J.T. and Perkins, E.A. (2008) Renormalization of the Two-dimensional Lotka-Volterra Model. *Ann. Appl. Prob.* **18** 747-812.

[10] Csörgo, M. and Horvath, L. (1993) *Weighted Approximations in Probability and Statistics.* John Wiley and Sons, New York.

[11] De Masi, A., Ferrari, P. and Lebowitz, J. (1986) Reaction-diffusion equations for interacting particle systems. *J. Statist. Phys.* **44** 589-644.

[12] Durrett, R., (1984) Oriented percolation in two dimensions. *Ann. Probab.* **12**, 999–1040.

[13] Durrett, R. (1992) Multicolor particle systems with large threshold and range. *J. Theoretical Prob.*, **5** 127-152.

[14] Durrett, R. (1995) Ten Lectures on Particle Systems, in *École d'Été de Probabilités de Saint Flour*, Lecture Notes Math.,**1608** pages 97-201, Springer-Verlag, Berlin.

[15] Durrett, R. (2002) Mutual invadability implies coexistence in spatial models. *Memoirs of the AMS.* **156** Number 740.

[16] Durrett, R. (2009) Coexisence in stochastic spatial models. *Ann. Appl. Probab.* **19**, 477–496

[17] Durrett, R., and Griffeath, D., (1982) Contact processes in several dimensions, *Z. fur Wahr.*, **53**, 535-552.

[18] Durrett, R. and Lanchier, N. Coexistence in host pathogen systems. *Stoch. Proc. Appl.*, to appear.

[19] Durrett, R. and Levin, S. (1994) The importance of being discrete (and spatial). *Theoret. Pop. Biol.* **46** 363-394.

[20] Durrett, R. and Neuhauser, C. (1994) Particle systems and reaction-diffusion equations. *Ann. Prob.* **22** 289-333.

[21] Durrett, R. and Neuhauser, C. (1997) Coexistence results for some competition models. *Ann. Appl. Probab.*, **7** 10-45.

[22] Durrett, R. and Reminik, D. (2012) Evolution of dispersal distance. *J. Math. Biol.*, **64** 657-666.

[23] Durrett, R. and Schinazi, R. (1993) Asymptotic critical value for a competition model. *Ann. Applied. Prob.* **3** 1047-1066,

[24] Durrett, R. and Zähle, I. On the width of hybrid zones. *Stoch. Proc. Appl.* **117** (2007), 1751–1763

[25] Ethier, S.N. and Kurtz, T.G. (1986) *Markov Processes, Characterization and Convergence.* John Wiley and Sons, New York.

[26] Fife, P.C. and McLeod, J.B. (1977) The approach of solutions of nonlinear diffusion equations to travelling front solutions. *Arch. Rat. mech. Anal.* **65** 335-361.

[27] Fife, P.C. and McLeod, J.B. (1981) A phase plane discussion of convergence to traveling fronts for nonlinear diffusion. *Arch. Rat. mech. Anal.* **75** 281-314.

[28] Griffeath, D. (1978) Additive and cancellative interacting particle systems. *Springer Lecture Notes in Math.* **714**.

[29] Griffeath, D., (1981) The basic contact process. *Stoch. Proc. Appl.* **11**, 151-185.

[30] Harris, T.E. (1972) Nearest neighbor Markov interaction processes on multi-dimensional lattices. *Adv. Math.* **9** 66-89.

[31] Hatcher, A.(2002). *Algebraic Topology*, Cambridge University Press, Cambridge.

[32] Kolmogorov, A.N., Petrovsky, I.G., and Piscounov, N.S. (1937) Étude de l'équation de la diffusion avec croissance de la quantité de matière et son application à un problème biologique. *Moscou Universitet. Bull. Math.***1** 1-25.

[33] Lambiotte, R., Saramäki, J., and Blondel, V.D. (2009) Dyanmics of latent voters. *Physical Review E* 79, paper 046107

[34] Lambiotte, R., and Redner, S. (2007) Vacillating voter models. *J. Stat. Mech.: Theory and Experiment* L10001

[35] Liggett, T.M (1985). *Interacting Particle Systems*, Springer-Verlag, New York.

[36] Liggett, T.M (1999). *Interacting Particle Systems PART 2*, Springer-Verlag, New York.

[37] Liggett, T.M., Schonmann, R.H., Stacey, A.M. (1997) Domination by product measures. *Ann. Probab.* **25** 71-95.

[38] McKean, H.P. (1975) Application of Brownian motion to the equation of Kolmogorov-Petrovskii-Piscounov. *Comm. Pure Appl. Math.* **28** 323-331.

[39] Molofsky, J., Durrett, R., Dushoff, J., Griffeath, D., and Levin, S. (1999) Local frequency dependence and global coexistence. *Theor. Pop. Biol.* **55**, 270–282

[40] Neuhauser, C. and Pacala, S.W. (1999) An explicitly spatial version of the Lotka-Volterra model with interspecific competition. *Ann. Appl. Probab.* **9** 1226-1259.

[41] Ohtsuki, H., Hauert, C., Lieberman, E., and Nowak, M.A. (2006) A simple rule for the evolution of cooperation on graphs and social networks. *Nature.* **441**, 502–505

[42] Perkins, E. (2002) Measure-valued processes and interactions, in *École d'Été de Probabilités de Saint Flour XXIX-1999*, Lecture Notes Math. **1781**, pages 125-329, Springer-Verlag, Berlin.

[43] Presutti, E. and Spohn, H. (1983) Hydrodynamics of the voter model. *Ann. Probab.* **11** 867-875.

[44] Spitzer, F.L. (1976) *Principles of Random Walk*, 2nd ed. Springer-Verlag, New York.

[45] Weinberger, H. F. (1982) Long-time behavior of a class of biological models. *Siam J. Math. Anal.* **13** 353-396.

Mathematics Department, Syracuse University,
215 Carnegie Hall, Syracuse, NY 13244-1150, USA
E-mail address: jtcox@syr.edu

Mathematics Department, Duke University,
Box 90320, Durham, NC 27708-0320, USA
E-mail address: rtd@math.duke.edu

Department of Mathematics, The University of British Columbia,
1984 Mathematics Road, Vancouver, B.C., Canada V6T 1Z2
E-mail address: perkins@math.ubc.ca