

# A PHASE TRANSITION FOR MEASURE-VALUED SIR EPIDEMIC PROCESSES

BY STEVEN P. LALLEY\* EDWIN A. PERKINS† AND XINGHUA ZHENG‡

*University of Chicago, University of British Columbia, and Hong Kong University of Science and Technology*

We consider measure-valued processes  $X_t$  that solve the following martingale problem: For a given initial measure  $X_0$ , and for all smooth, compactly supported test functions  $\varphi$ ,

$$X_t(\varphi) = X_0(\varphi) + \frac{1}{2} \int_0^t X_s(\Delta\varphi) ds + \theta \int_0^t X_s(\varphi) ds - \int_0^t X_s(L_s\varphi) ds + M_t(\varphi).$$

Here  $L_s(x)$  is the local time density process associated with  $X_t$ , and  $M_t(\varphi)$  is a martingale with quadratic variation  $[M(\varphi)]_t = \int_0^t X_s(\varphi^2) ds$ . Such processes arise as scaling limits of SIR epidemic models. We show that there exist critical values  $\theta_c(d) \in (0, \infty)$  for dimensions  $d = 2, 3$  such that if  $\theta > \theta_c(d)$ , then the solution survives forever with positive probability, but if  $\theta < \theta_c(d)$ , then the solution dies out in finite time with probability 1. For  $d = 1$  we prove that the solution dies out almost surely for all values of  $\theta$ . We also show that in dimensions  $d = 2, 3$  the process dies out locally almost surely for any value of  $\theta$ , that is, for any compact set  $K$ , the process  $X_t(K) = 0$  eventually.

## CONTENTS

1	Introduction . . . . .	2
1.1	Epidemic models and their continuum limits . . . . .	2
1.2	Main results: survival . . . . .	4
1.3	Proof strategy and heuristics . . . . .	5
1.4	Epidemics with suppression . . . . .	7
1.5	Relations with scaling laws for contact processes . . . . .	7
1.6	Plan of the paper . . . . .	8
2	Preliminaries on the epidemic processes . . . . .	9
2.1	Dawson’s Girsanov Theorem; existence and uniqueness . . . . .	9
2.2	Discrete epidemic models . . . . .	13
2.3	Comparison lemmas . . . . .	18
2.4	The Sandwich Lemma . . . . .	23

---

\*Research partially supported by NSF Grant DMS-1106669

†Research partially supported by an NSERC Discovery Grant

‡Research partially supported by NSERC (Canada) and GRF 606010 of the HKSAR

*AMS 2000 subject classifications:* Primary 60H30, 60K35; secondary 60H15

*Keywords and phrases:* Spatial epidemic, Dawson-Watanabe process, Phase transition, Local extinction

2.5	Scaling . . . . .	26
3	Preliminaries on (supercritical) super-Brownian motions . . . . .	26
3.1	Uniform regularity of super-Brownian motions . . . . .	27
3.2	Local time densities of super-Brownian motions . . . . .	28
3.3	Local time densities of supercritical super-Brownian motions . . . . .	32
3.4	Propagation of supercritical super-Brownian motions . . . . .	34
4	A weak form of local extinction and its consequences . . . . .	34
4.1	A weak form of local extinction . . . . .	35
4.2	Universality of the critical values $\theta_c$ . . . . .	36
4.3	Extinction in dimension one . . . . .	37
5	Proof of survival when $d = 2$ or $3$ . . . . .	38
5.1	Scaled process . . . . .	38
5.2	Sandwich Lemma . . . . .	39
5.3	Percolation probability estimates . . . . .	39
5.4	Proof of survival . . . . .	42
6	Proof of extinction when $d = 2$ or $3$ . . . . .	45
6.1	Scaled process . . . . .	45
6.2	Proof of Lemma 6.3 and Proposition 6.4 . . . . .	48
6.3	Proof of extinction for the original equation . . . . .	53
7	A strong form of local extinction . . . . .	54
7.1	Re-infection at large times . . . . .	54
7.2	Finite mass flux . . . . .	56
7.3	Local extinction . . . . .	60
A	Appendix: Proof of (2.17) for $d = 2, 3$ . . . . .	63
	Acknowledgments . . . . .	66
	REFERENCES . . . . .	66

## 1. Introduction.

1.1. *Epidemic models and their continuum limits.* The use of stochastic processes to model epidemics can be traced to McKendrick (1926) and Kermack and McKendrick (1927), who proposed a simple continuous-time, mean-field model of an *SIR* (for *susceptible-infected-removed* or *susceptible-infected-recovered*) epidemic. The corresponding discrete-time model (known variously as the *Reed-Frost* or the *chain-binomial* model – see Daley and Gani (1999) for background) was proposed several years later, in 1928, by Reed and Frost in lectures at Johns Hopkins University. In these models, an *infected* individual remains infected for a certain period of time, during which he/she can transmit the disease to *susceptible* individuals, and then *recovers*, after which he/she is immune to further infection. Both models are *mean-field* models: the rate of infection transmission is the same for all pairs of

infected and susceptible members of the population. The Reed-Frost model is of particular interest not only because of its use in modeling epidemics and epidemic-like processes but because of its close relation to the *Erdős-Renyi random graph* model. In particular, given a realization of an Erdős-Renyi graph, whose vertices are marked either  $S$  or  $I$ , a realization of the Reed-Frost process can be obtained by defining  $I_n$ , the infected set at time  $n$ , to be the set of all vertices at (graph) distance  $n$  from the set  $I$ . The union of the connected components of the Erdős-Renyi graph that contain vertices in the set  $I = I_0$  consists of all individuals ever infected during the course of the epidemic.

Spatial versions of the above models have a rich history in both the mathematical and biological literature. Bailey (1967) considered a spatial version of the Reed-Frost model and Mollison (1977) is a good source of information about a range of related stochastic spatial models. Cox and Durrett (1988) prove a shape theorem for a related continuous time/discrete space model in two dimensions which is clearly similar in spirit to our main theorems below on survival and local extinction for a continuum model in two and three dimensions.

The *SIR* models differ qualitatively from *SIS* and *SIRS* models, such as the *stochastic logistic* model, in that the progress of the epidemic depends on an exhaustible resource which is gradually consumed. This leads to interesting critical behavior, as was discovered by Martin-Löf (1998) and Aldous (1999). Martin-Löf proved, in particular, that at criticality (when the probability of transmission from an infected to a susceptible individual is  $p = p_c = 1/N$ , where  $N$  is the size of the population) then as  $N \rightarrow \infty$ , after suitable scaling, the total number of individuals ever infected converges in law to the first passage time of a Wiener process to a parabolic boundary. Dolgoarshinnykh and Lalley (2006) subsequently showed that for suitable initial conditions the Kermack-McKendrick epidemic process, after rescaling, converges weakly as  $N \rightarrow \infty$  to a continuous-time process  $I_t$  that satisfies the stochastic differential equation

$$(1.1) \quad \begin{aligned} dI_t &= (\lambda I_t - I_t R_t) dt + \sqrt{I_t} dW_t \quad \text{where} \\ dR_t &= I_t dt. \end{aligned}$$

The proof can easily be adapted to show that the Reed-Frost process has the same limit. The parameter  $\lambda \in \mathbb{R}$  represents the *transmission rate* of the disease: it is related to the infected-susceptible transmission probability  $p$  in the Reed-Frost model by  $p = 1/N + \lambda/N^{4/3}$ . It is not difficult to see (using well known facts about Feller's diffusion) that for any value of  $\lambda$  the process  $I_t$  defined by (1.1) is eventually absorbed at 0.

The subject of this paper is a stochastic partial differential analogue of the system (1.1) that arises as a scaling limit of a spatial version of the Reed-Frost process proposed by Lalley (2009) as a crude model for an epidemic in a geographically stratified population. In this model, populations of size  $N$  ("villages") are located at each lattice point of  $\mathbb{Z}^d$ ; the rules of transmission are the same as in the Reed-Frost model, *except* that infectious contacts are permitted only for infected-susceptible pairs in the same or neighboring villages. (The model is described in more detail in Section 2.2 below.) Large-population ( $N \rightarrow \infty$ ) limit theorems for near-critical versions of these spatial *SIR* processes were proved for  $d = 1$  in Lalley (2009) and for  $d = 2, 3$  in Lalley and Zheng (2010). The limit processes are now continuous finite measure-valued processes  $X = (X_t)_{t \geq 0}$ ; for each time  $t$ , the random measure  $X_t$  represents

the infected set (more precisely, its distribution in space), and  $R_t = \int_0^t X_s ds$  the recovered set. The dynamics of the model are specified by the following martingale problem. For any Radon measure  $\mu$  on  $\mathbb{R}^d$  and any integrable or non-negative measurable function  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ , write  $\mu(\varphi)$  or  $\langle \mu, \varphi \rangle$  for the integral  $\int \varphi d\mu$ . Then for any initial mass distribution  $X_0 = \mu$  and any test function  $\varphi \in C_c^2(\mathbb{R}^d)$ ,

$$(1.2) \quad X_t(\varphi) = \mu(\varphi) + \frac{1}{2} \int_0^t X_s(\Delta\varphi) ds + \theta \int_0^t X_s(\varphi) ds - \int_0^t X_s(L_s\varphi) ds + M_t(\varphi).$$

Here  $C_c^2(\mathbb{R}^d)$  stands for the space of compactly supported twice differentiable with continuous second derivative functions on  $\mathbb{R}^d$ ,  $M_t(\varphi)$  is a continuous martingale with quadratic variation  $[M(\varphi)]_t = \int_0^t X_s(\varphi^2) ds$ , and  $L_t(x)$  is the Sugitani *local time density process* of  $X$ , that is, for each  $t \geq 0$  the function  $L_t(x)$  is the density of the occupation measure  $R_t$ . (Throughout this article, unless otherwise specified, the martingale  $M_t(\varphi)$  in a martingale problem such as (1.2) will be a martingale relative to the minimal right continuous filtration of the process  $X$ , i.e.,  $\mathcal{F}_t^X := \cap_{u>t} \sigma(X_s, s \leq u)$ ). Dawson's Girsanov formula (Section 2.1 below) implies that on a suitable probability space there exists a solution to (1.2), that solutions are unique in law, and that the law is absolutely continuous on finite time intervals with respect to the law of super-Brownian motion (see the definition below in Subsection 1.4). However, because the Sugitani local time process  $L_t$  depends on the entire past of the spatial epidemic  $X_t$ , solutions  $X_t$  will not generally be Markov (although the vector-valued process  $(X_t, L_t)$  will be). Henceforth, we shall call a measure-valued processes  $X$  satisfying (1.2) a *spatial epidemic process with transmission rate  $\theta$  and initial mass distribution  $\mu$* .

The martingale problem (1.2) is a natural spatial analogue of the stochastic differential equation (1.1). In both problems, the key qualitative feature is a ‘‘resource depletion’’ term: in (1.1), it is the integral  $\int_0^t I_s R_s ds$ , whereas in the martingale problem (1.2) it is the integral  $\int_0^t X_s(L_s\varphi) ds$ . It seems likely that processes  $X_t$  governed by (1.2) – or similar equations incorporating depletion terms – should also arise as continuum limits of models for various other physical (combustion), chemical (reaction-diffusion), and biological processes (foraging) in which there is an exhaustible resource upon which the process depends. In fact, Mueller and Tribe (2011) have suggested (see their Remark at the end of Section 6) that they should also occur as scaling limits of certain stochastic reaction-diffusion systems.

**1.2. Main results: survival.** A measure-valued process  $X$  *survives* if  $X_t(1) > 0$  for all  $t > 0$ ; it *dies out*, or becomes *extinct*, if  $X_t = 0$  for large enough  $t$ . For processes governed by equation (1.2), the question of whether or not there is survival or extinction is of fundamental importance. Mueller and Tribe (2011) (see again the Remark at the end of Section 6) have conjectured that there is a critical value  $\theta_c = \theta_c(d) \in (0, \infty)$  for the transmission rate below which extinction is certain and above which survival has positive probability. Our main result states that under a mild restriction on the initial measure  $\mu$  this is true in dimensions  $d = 2$  and  $d = 3$ , but that in  $d = 1$  extinction is certain at all values of the parameter  $\theta$ . The restriction on the initial measure is as follows:

ASSUMPTION 1.1. *The measure  $\mu$  has compact support and finite total mass, and when*

$d = 2$  or  $3$ , its convolution  $\mu * q_t$  with the integrated Gauss kernel

$$(1.3) \quad q_t(x) = \int_0^t p_s(x) ds, \quad \text{where} \quad p_t(x) = \frac{e^{-|x|^2/2t}}{(2\pi t)^{d/2}},$$

is jointly continuous in  $(t, x) \in [0, \infty) \times \mathbb{R}^d$ .

Theorem 2 of Sugitani (1989) asserts that in dimensions  $d = 2$  and  $d = 3$  a super-Brownian motion with initial mass distribution  $\mu$  satisfying Assumption 1.1 has a local time density process  $L_t(x)$  that is jointly continuous in  $t \geq 0$  and  $x \in \mathbb{R}^d$ . Since the law of a spatial epidemic  $X$  is absolutely continuous relative to that of super-Brownian motion, spatial epidemics must also have jointly continuous local time processes in  $d = 2, 3$ . In dimension  $d = 1$ , the existence and continuity of the local time process follows from the fact that the state of a super-Brownian motion at any time  $t$  is absolutely continuous with respect to Lebesgue measure, with a jointly continuous density. Thus, the equation (1.2) makes sense in all dimensions  $d \leq 3$ , and so henceforth we assume that  $d \leq 3$ .

**THEOREM 1.2.** *There exist critical values  $\theta_c = \theta_c(2) > 0$  and  $\theta_c = \theta_c(3) > 0$  such that the following is true: If  $d = 2$  or  $d = 3$  and  $X$  is a spatial epidemic process in  $\mathbb{R}^d$  with transmission rate  $\theta$  and initial mass distribution  $\mu$  satisfying Assumption 1.1, then*

- (a) *if  $\theta < \theta_c$  then  $X$  dies out almost surely, but*
- (b) *if  $\theta > \theta_c$  then  $X$  survives with positive probability.*

*If  $X$  is a spatial epidemic in  $\mathbb{R}^1$  with any transmission rate  $\theta$  and any finite initial mass distribution  $\mu$  then  $X$  dies out almost surely.*

Thus, in dimensions 2 and 3 a spatial epidemic can survive if the transmission rate is sufficiently high. However, since the process feeds on a substrate which is gradually consumed in infected areas, it is natural to conjecture that the epidemic should survive in a transient wave which sweeps through space. The following result partly establishes the validity of this picture.

**THEOREM 1.3.** *Let  $X$  be a spatial epidemic with arbitrary transmission rate  $\theta \in \mathbb{R}$  and initial mass distribution satisfying Assumption 1.1. For any compact set  $K \subset \mathbb{R}^d$ , with probability one,*

$$(1.4) \quad X_t(K) = 0 \quad \text{eventually.}$$

Consequently, with probability one the local time  $L_t(x)$  at any point  $x$  is eventually constant. Since the local time  $L_t(x)$  is jointly continuous in its arguments, it follows that  $L_\infty(x) := \lim_{t \rightarrow \infty} L_t(x)$  is finite and continuous in  $x$  almost surely.

**1.3. Proof strategy and heuristics.** The proofs of Theorems 1.2–1.3 are rather technical, largely because of difficulties that will arise in carrying out comparison arguments for measure-valued processes defined by stochastic partial differential equations in which the entire histories of the solutions (e.g., local time density) influence the coefficients. However,

the ideas behind the results can be explained, at least roughly, in simple terms. Consider first the assertion of global extinction in one dimension. If the epidemic process  $X$  were to survive with positive probability, then on this event its total mass  $X_t(1)$  would diverge to  $\infty$ , since otherwise the process would be presented with infinitely many opportunities to become extinct (see Lemma 2.15). In addition, by a large deviations calculation on a dominating super-Brownian motion with drift  $\theta$  (see Pinsky (1995)), there exists  $c < \infty$  such that  $\text{Supp}(X_t) \subset [-ct, ct]^d$  for all large  $t$ . Therefore, for  $d = 1$ , on the event of survival and for large  $t$ , the average value of  $L_t(\cdot)$  must satisfy

$$(2ct)^{-1} \int_{-ct}^{ct} L_t(x) dx = (2ct)^{-1} \|L_t\|_1 = (2ct)^{-1} \int_0^t X_s(1) ds \rightarrow \infty.$$

If (1.2) were valid for the function  $\varphi \equiv 1$  (it is only assumed for compactly supported functions), then it would follow that for large  $t$  the drift term in (1.2) for the total mass  $X_t(1)$  would eventually turn (very) negative, making it impossible for  $X_t(1)$  to remain positive. The formal proof in Section 4.3 makes this heuristic argument precise.

A local variation of this argument (which is harder to justify rigorously – see Section 7) explains the strong local extinction asserted in Theorem 1.3. We will show that in order for  $X_t(K) > 0$  to occur at indefinitely large times, for some ball  $B$  centered at the origin, it must be the case that  $X_t(3B)$  integrates to  $\infty$ . This, however, would imply that the local time in  $2B \setminus B$  would grow indefinitely, eventually making the drift in the equation (1.2) for  $X_t(B)$  negative.

A different line of argument makes it at least plausible that in dimensions  $d \geq 2$  the epidemic  $X$  might survive with positive probability when the transmission rate  $\theta$  is sufficiently large. If  $\theta$  is large, then equation (1.2) implies that when the infection first enters a region  $K$  of space it will, at least for a while, grow at least as fast as a super-Brownian motion with a large constant drift. Thus, with high probability, the total mass  $X_t(K)$  will become large long before the local time  $L_t$  becomes appreciable in  $K$ . In particular, for a cube  $K$ , if the size  $X_t(K)$  of the infected set reaches a certain threshold before the local time exceeds a fraction of this level, then the epidemic will have high probability of spreading to neighboring cubes quickly, and the infection in these cubes will have similarly high probability of spreading to neighboring cubes, and so on. Since high-density oriented site percolation in dimensions  $d \geq 2$  has infinite clusters, with positive probability, it should then follow that the epidemic will reach infinitely many cubes with positive probability. It will take some work to implement this plan. This is done in Section 5 after some important groundwork is laid in Sections 2 and 3.

For the extinction assertion of Theorem 1.2 we will adapt the corresponding argument of Mueller and Tribe (1994). For small  $\theta > 0$  it is possible to rescale  $X$  so that the total mass process can be dominated by a subcritical branching process which dies out. The actual implementation of this idea is carried out in a slightly different manner in Section 6 (see Proposition 6.1). A key observation, used here and elsewhere in this work, is that if the initial state is split up into pieces, then one can couple the epidemics so that the survival probability is dominated by the sum of the survival probabilities corresponding to the pieces (see Lemma 2.19).

1.4. *Epidemics with suppression.* Our results extend to a somewhat larger class of measure-valued processes that incorporate location-dependent *local suppression*. Let  $K : \mathbb{R}^d \rightarrow \mathbb{R}_+$  be a bounded, continuous (or, more generally, piecewise continuous), nonnegative function; call this the *suppression rate*. A *spatial epidemic with local suppression rate  $K$ , transmission rate  $\theta$ , branching rate  $\gamma > 0$ , and inhibition parameter  $\beta \geq 0$*  is a solution to the martingale problem  $(MP)_{\mu, K}^{\theta, \beta, \gamma}$

$$(1.5) \quad X_t(\varphi) = \mu(\varphi) + \int_0^t X_s(\Delta\varphi/2 + \theta\varphi - K\varphi - \beta L_s(X)\varphi) ds + \sqrt{\gamma}M_t(\varphi), \quad \varphi \in C_c^2(\mathbb{R}^d),$$

where  $X_t$  is a continuous finite measure-valued process,  $M_t(\varphi)$  is a continuous martingale with quadratic variation  $[M(\varphi)]_t = \int_0^t X_s(\varphi^2) ds$  and  $L_t(x)$  is the local time density of  $X_t$ . When  $K \equiv 0$  and  $\beta = \theta = 0$  and  $\gamma = 1$ , a process  $X$  satisfying (1.5) is a *super-Brownian motion*; more generally, when  $K \equiv 0$  and  $\beta = 0$ , it is a *super-Brownian motion with drift  $\theta$*  and branching rate  $\gamma$ ; and when  $\beta = 0$  it is a *super-Brownian motion with killing*, with local killing rate  $K$ , drift  $\theta$  and branching rate  $\gamma$ . Theorems 1.2 and 1.3 extend to all processes governed by (1.5) with  $\beta > 0$ ; the critical values  $\theta_c$  will depend on the parameters, but not on the suppression rate function  $K$  if we restrict  $K$  to be compactly supported. In the interests of simplicity we shall prove our main results only in the case  $K \equiv 0$ . However, solutions of the martingale problem (1.5) will be needed in the proofs of the main results even in the special case  $K \equiv 0$  as they will arise naturally in the Markov property of solutions to (1.2).

1.5. *Relations with scaling laws for contact processes.* As noted above, in *SIR* epidemics, unlike *SIS* and *SIRS* epidemics, the population of susceptible individuals is gradually depleted during the course of the epidemic. It is this that accounts for the depletion term  $-\int_0^t X_s(L_s\varphi) ds$  in the martingale problem (1.2), which in turn is responsible for the local extinction asserted in Theorem 1.3. Spatial models of *SIS* and *SIRS* lead to measure-valued processes with different qualitative behavior. One such model that has been studied in some detail is the *long-range contact process*, cf. Bramson et al. (1989), Mueller and Tribe (1994), Durrett and Perkins (1999). In this model, only one individual, who can be either infected or susceptible, inhabits each lattice point, but infectious contact is allowed at distances up to  $L$  (usually the  $\ell_\infty$  metric is used). Scaling limits obtain for the limiting regime  $L \rightarrow \infty$  (see Mueller and Tribe (1995) and Durrett and Perkins (1999) for details). Bramson et al. (1989) determined the long-range functional dependence of the critical value  $\lambda_c(L)$  on  $L$  (but not the precise constants): in dimension  $d = 1$ , they showed that for large  $L$ ,

$$(1.6) \quad 0 < cL^{-2/3} \leq \lambda_c(L) - 1 \leq CL^{-2/3}.$$

The long-conjectured (but still unresolved) link between the discrete and continuum settings in  $d = 1$  is that

$$(1.7) \quad \lambda_c(L) - 1 \sim \theta_c L^{-2/3}.$$

Durrett and Perkins (1999) established weak convergence of long-range contact processes to a super-Brownian motion in dimensions  $d \geq 2$ , while for  $d = 1$  Mueller and Tribe (1995)

showed that the scaling limit of the long-range contact process is governed by the stochastic PDE

$$(1.8) \quad \frac{\partial u}{\partial t} = \frac{u''}{6} + \theta u - u^2 + \sqrt{2u}\dot{W}.$$

In this equation the local time density in the third integral of (1.2) is replaced by the density,  $u_t$ , of  $X_t$ . This reflects the fact that in the contact (and other *SIS*) processes, the susceptible population is depleted locally by the current size of the infected set. The results of Durrett and Perkins (1999) show this effect induces a killing term with a known constant rate. Mueller and Tribe (1994) showed that there is a phase transition in equation (1.8) in that there is positive probability of survival for  $\theta$  above a critical  $\theta_c > 0$  and a.s. extinction below it. By contrast, the martingale problems (1.2) have solutions in up to 3 dimensions, whereas (1.8) only makes sense in one spatial dimension (since super-Brownian motion has the property that the mass distributions  $X_t$  at positive times are absolutely continuous only in dimension 1).

The discrete *SIR* models underlying our continuous models are described in Section 2.2 below. The analogue to (1.6) in this discrete setting is also described in Section 2.2.

1.6. *Plan of the paper.* The remainder of the paper is devoted to the proofs of Theorems 1.2–1.3. The plan is as follows. In Section 2 we discuss existence and uniqueness of solutions to a class of martingale problems including (1.2), weak convergence of certain discrete processes to spatial epidemics, and basic comparison principles. In Section 3 we discuss some regularity properties of (supercritical) super-Brownian motions and their local time densities. In Sections 4.2 and 4.3 we prove that the critical values  $\theta_c$  in dimensions 2 and 3 do not depend on the initial mass distributions, and we prove that spatial epidemics in  $\mathbb{R}^1$  die out almost surely at all values of the transmission rate  $\theta$ . In Section 5 we prove that spatial epidemics in dimensions 2 and 3 can survive if the transmission rate  $\theta$  is sufficiently high; and in Section 6 we prove that at low values of  $\theta$  extinction is certain. We prove a weak form of local extinction in Section 4.1, and finally, in Section 7, we prove Theorem 1.3.

**Standing Notation.** For any  $a \geq 0$ ,  $[a]$  stands for its integer part. For any Borel subset  $D \subseteq \mathbb{R}^d$ , let  $\mathcal{M}(D)$  be the space of finite Borel measures on  $D$ , equipped with the weak topology, and let  $\mathcal{M}_c(D)$  be the subset consisting of all measures with compact support in  $D$ . These spaces are partially ordered in a natural way: we write  $\mu \leq \nu$  to mean that for all nonnegative, bounded functions  $\varphi$ ,

$$\int \varphi d\mu \leq \int \varphi d\nu.$$

For a measure  $\mu \in \mathcal{M}(D)$  and a nonnegative measurable function  $f : D \rightarrow \mathbb{R}_+$ , we shall continue to use the shorthand notations  $\mu(f)$  or  $\langle \mu, f \rangle$  to denote the integral of  $f$  against  $\mu$  and also write  $|\mu|$  for  $\mu(1)$ , the total mass of  $\mu$ . Let  $C_b(\mathbb{R}^d)$  be the space of bounded and continuous functions on  $\mathbb{R}^d$ , endowed with the sup-norm topology, and let  $C_c(\mathbb{R}^d)$  be the space of compactly supported continuous functions on  $\mathbb{R}^d$ . Furthermore, for any  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  and any  $r > 0$ , let  $Q_r(x) = [x_1 - r/2, x_1 + r/2) \times \dots \times [x_d - r/2, x_d + r/2)$  be the (half-closed,

half-open) cube of side length  $r$  centered at  $x$ , and, for notational ease,  $Q(x) := Q_1(x)$ . Finally, let  $C_p(\mathbb{R}^d, \mathbb{R}_+)$  be the space of non-negative piecewise constant functions on  $\mathbb{R}^d$  satisfying the following conditions: each such function is supported by  $\cup_i Q(x_i)$  for finitely many  $x_i \in \mathbb{Z}^d$  and is constant on each cube.

**Conventions.** Throughout the paper,  $C, c, C_1$ , etc. denote generic constants whose values may change from line to line. The notation  $Y_n = o_P(f(n))$  means that  $Y_n/f(n) \rightarrow 0$  in probability; and  $Y_n = O_P(f(n))$  means that the sequence  $|Y_n|/f(n)$  is tight. Also, for any  $a, b \in \mathbb{R}$ ,  $a \wedge b := \min(a, b)$  and  $a \vee b := \max(a, b)$ . Finally, we use a “local scoping rule” for notation: Any notation introduced in a proof is local to the proof, unless otherwise indicated.

## 2. Preliminaries on the epidemic processes.

2.1. *Dawson’s Girsanov Theorem; existence and uniqueness.* Existence and uniqueness of solutions (in the weak sense) to a class of martingale problems similar to (1.5) was established in Mueller and Tribe (2011) using Dawson’s Girsanov theorem. Existence in the special case  $K \equiv 0$ ,  $\theta = 0$  was also proved in Lalley (2009) and Lalley and Zheng (2010), by weak convergence methods, which extend trivially to the general case. Nevertheless, since Dawson’s Girsanov formula will be of crucial importance in many of the arguments to follow, we begin by reviewing the essential facts. We first state a variant of Dawson’s Girsanov Theorem (Theorem IV.1.6 in Perkins (2002)) tailored to our needs.

Let  $\Omega = D([0, \infty); \mathcal{M}_c(\mathbb{R}^d))$  be the canonical path space for compactly supported measure-valued processes, with coordinate maps  $X_t : \Omega \rightarrow \mathcal{M}_c(\mathbb{R}^d)$  and associated filtration  $\mathbb{F} = (\mathcal{F}_t^X)_{t \geq 0}$ . Fix a probability measure  $P$  on  $(\Omega, \mathcal{F}_\infty^X)$ , and suppose that there is a linear mapping  $\psi \mapsto (M_t(\psi))_{t \geq 0}$  from the space  $C_c^2(\mathbb{R}^d)$  to the space of  $\mathbb{F}$ -adapted, continuous martingales such that  $M_0(\psi) = 0$  and such that  $M(\psi)$  has quadratic variation  $[M(\psi)]_t = \int_0^t \langle X_s, \psi^2 \rangle ds$ . This mapping extends to an *orthogonal martingale measure*  $dM(s, x)$  (see Walsh (1986)). For any previsible  $\times$  Borel process  $B : \mathbb{R}_+ \times \Omega \times \mathbb{R}^d \rightarrow \mathbb{R}$ , we say that  $B$  is  $L^2$ -admissible if

$$(2.1) \quad \int_0^t \langle X_s, B_s^2 \rangle ds < \infty \quad \text{for all } t \geq 0 \quad P\text{-almost surely.}$$

If  $B$  is  $L^2$ -admissible, then the stochastic integrals

$$(2.2) \quad \int_0^t \int_{\mathbb{R}^d} B_s(x) dM(s, x)$$

exist and constitute a continuous,  $\mathbb{F}$ -adapted local martingale with quadratic variation process  $\int_0^t \langle X_s, B_s^2 \rangle ds$ . Consequently, for each  $\gamma > 0$ , the process

$$(2.3) \quad \mathcal{E}_t^B = \exp \left( \frac{1}{\sqrt{\gamma}} \int_0^t \int_{\mathbb{R}^d} B_s(x) dM(s, x) - \frac{1}{2\gamma} \int_0^t \langle X_s, B_s^2 \rangle ds \right),$$

is a continuous local martingale.

LEMMA 2.1. (*Dawson's Girsanov Theorem*) Let  $P$  be a probability measure on  $(\Omega, \mathcal{F}_\infty)$  such that under  $P$  the coordinate process  $(X_t)_{t \geq 0}$  satisfies the following martingale problem: For some uniformly bounded  $L^2$ -admissible integrand  $A$ , for all  $\psi \in C_c^2(\mathbb{R}^d)$ ,

$$(2.4) \quad X_t(\psi) = X_0(\psi) + \frac{\alpha}{2} \int_0^t \langle X_s, \Delta \psi \rangle ds + \int_0^t \langle X_s, A_s \psi \rangle ds + \sqrt{\gamma} M_t^P(\psi),$$

where  $M_t^P(\psi)$  is a continuous  $\mathcal{F}_t^X$ -martingale with quadratic variation  $[M^P(\psi)]_t = \int_0^t X_s(\psi^2) ds$ .

(a) Suppose that  $Q$  is another probability measure on  $(\Omega, \mathcal{F}_\infty^X)$  such that under  $Q$  the coordinate process  $(X_t)_{t \geq 0}$  satisfies the martingale problem

$$(2.5) \quad X_t(\psi) = X_0(\psi) + \frac{\alpha}{2} \int_0^t \langle X_s, \Delta \psi \rangle ds + \int_0^t \langle X_s, (A_s + B_s) \psi \rangle ds + \sqrt{\gamma} M_t^Q(\psi),$$

for all  $\psi \in C_c^2(\mathbb{R}^d)$ , where  $B$  is a uniformly bounded  $L^2$ -admissible integrand and  $M_t^Q(\psi)$  is a continuous martingale (under  $Q$ ) with quadratic variation  $[M^Q(\psi)]_t = \int_0^t X_s(\psi^2) ds$ . Suppose also that the restrictions of  $P$  and  $Q$  to the  $\sigma$ -algebra  $\mathcal{F}_0^X$  are equal. Then for each  $t < \infty$  the measures  $P$  and  $Q$  on  $\mathcal{F}_t^X$  are mutually absolutely continuous, with likelihood ratio

$$(2.6) \quad \left. \frac{dQ}{dP} \right|_{\mathcal{F}_t^X} = \mathcal{E}_t^B.$$

In particular,  $Q$  is uniquely determined on  $\mathcal{F}_\infty^X$  by the martingale problem (2.5).

(b) Conversely, if  $Q$  is the probability measure determined by the likelihood ratios (2.6), then under  $Q$  the process  $X$  satisfies the martingale problem (2.5).

We next apply the above to prove that the martingale problem (1.5) is well-posed. Recall that for each  $x \in \mathbb{R}^d$ ,  $Q_r(x)$  stands for the cube of side length  $r$  centered at  $x$ , and  $Q(x) = Q_1(x)$ . For any continuous path  $X_t$  valued in  $\mathcal{M}_c(\mathbb{R}^d)$ , define

$$L(t, X, x) = L_t(X, x) = L_t^X(x) = \limsup_{\varepsilon \downarrow 0} \frac{\int_0^t X_s(Q_\varepsilon(x)) ds}{\varepsilon^d}.$$

When there is no confusion, we shall suppress the dependence on  $X$  and abbreviate  $L_t(X, x)$  as  $L_t(x)$ . If  $X_t$  is an adapted process on the filtered space  $(\Omega, \mathbb{F})$  then  $L(t, X, x)$  is nonnegative, nondecreasing in  $t$ , and  $\mathcal{P} \times \mathcal{B}^d$ -measurable, where  $\mathcal{B}^d$  is the Borel  $\sigma$ -field on  $\mathbb{R}^d$  and  $\mathcal{P}$  is the previsible  $\sigma$ -field. If  $X$  has a local time density,  $L(t, X, x)$  will be a jointly measurable version of it.

THEOREM 2.2. Let  $\mu \in \mathcal{M}_c(\mathbb{R}^d)$  satisfy Assumption 1.1 and let  $K \in C_p(\mathbb{R}^d, \mathbb{R}_+)$ . For any fixed  $\theta \in \mathbb{R}$  and  $\gamma > 0$ , denote by  $P_\mu = P_{\mu,0}^{\theta,0,\gamma}$  the law of a super-Brownian motion with initial mass distribution  $\mu$ , drift  $\theta$ , and branching rate  $\gamma$ .

(a) If  $X$  solves the martingale problem (1.5) with initial value  $X_0 = \mu$  then the law  $P_{\mu,K} := P_{\mu,K}^{\theta,\beta,\gamma}$  of  $X$  on the canonical path space is unique and given by

$$(2.7) \quad \frac{dP_{\mu,K}}{dP_{\mu}} \Big|_{\mathcal{F}_t^X} = \mathcal{E}_t^B \quad \text{where } B(s, \omega, x) = -(K(x) + \beta L(s, X, x))$$

and  $dM(s, x)$  is the orthogonal martingale measure under  $P_{\mu}$ . Conversely, if  $P_{\mu,K}$  is the probability measure specified by (2.7) then under  $P_{\mu,K}$  the coordinate process  $X_t$  satisfies the martingale problem (1.5).

(b) The mapping  $(\mu, K) \mapsto P_{\mu,K}$  is jointly measurable with respect to the appropriate Borel fields.

(c) Under  $P_{\mu,K}$  the local time process  $L_t(x)$  is jointly continuous in  $(t, x)$  and almost surely is the density of the occupation measure  $R_t = \int_0^t X_s ds$ .

(d) Under the measure  $P_{\mu,K}$  the process  $(X, L)$  is strong Markov, that is, for any  $\mathcal{F}_t^X$ -stopping time  $\tau$ ,

$$P_{\mu,K}(X_{\tau+} \in A | \mathcal{F}_{\tau}) = P_{X_{\tau}, K + \beta L_{\tau}}(A) \quad \text{almost surely on } \{\tau < \infty\} \text{ for all } A \in \mathcal{F}_{\infty}^X.$$

(e) For any pair  $K, K'$  of suppression rate functions, the probability measures  $P_{\mu,K}$  and  $P_{\mu,K'}$  are mutually absolutely continuous on  $\mathcal{F}_t^X$ , with likelihood ratio

$$(2.8) \quad \frac{dP_{\mu,K'}}{dP_{\mu,K}} \Big|_{\mathcal{F}_t^X} = \exp \left\{ \frac{1}{\sqrt{\gamma}} \int_0^t \int (K(x) - K'(x)) dM_K(s, x) - \frac{1}{2\gamma} \int_0^t \int (K(x) - K'(x))^2 X_s(dx) ds \right\},$$

where  $dM_K(s, x)$  is the orthogonal martingale measure under  $P_{\mu,K}$ .

REMARK 2.3. Assertion (b) guarantees that if  $X_0$  and  $K_0$  are random and  $\mathcal{F}_0^X$ -measurable then the random probability measure  $P_{X_0, K_0}$  is  $\mathcal{F}_0^X$ -measurable. Similarly, if  $X_{\tau}$  and  $K_{\tau}$  are  $\mathcal{F}_{\tau}^X$ -measurable then  $P_{X_{\tau}, K_{\tau}}$  is  $\mathcal{F}_{\tau}^X$ -measurable. Moreover, since  $P_{X_0, K_0}$  is a regular conditional distribution on the canonical path space given  $\mathcal{F}_0^X$ , it follows from (d) that the strong Markov property holds when the initial condition  $X_0$  and the suppression rate function  $K_0$  are random.

REMARK 2.4. Since the local time density  $L_t$  is not uniformly bounded on finite time intervals, the exponential process  $\mathcal{E}_t^B$  is not *a priori* a martingale. Part of the assertion of the theorem is that in fact  $\mathcal{E}_t^B$  is a martingale, and hence that (2.7) defines a probability measure on  $\mathcal{F}_t^X$ .

PROOF OF THEOREM 2.2. (a) First we claim that any solution  $X_t$  to martingale problem (1.5) has the property that its local time density  $L_t^X(x, \omega)$  is bounded in  $(t, x)$  for  $t$  in finite intervals and for every  $t < \infty$  has compact support in  $x$  for almost every  $\omega$ . This follows because on some probability space a version of  $X_t$  can be coupled with a super-Brownian motion  $\bar{X}_t$  with drift  $\theta$  and branching rate  $\gamma$  such that  $\bar{X}_t \geq X_t$  for all  $t \geq 0$  almost surely. See Proposition IV.1.4 in Perkins (2002) which we apply with  $D = 0$ ,  $C_t(\varphi) = \int_0^t X_s(L_s^X \varphi) ds$ , and only to the first coordinate of the pair of processes considered there. To apply the above result we need to show that  $t \rightarrow C_t$  is a continuous  $\mathcal{M}(\mathbb{R}^d)$ -valued process. For  $\varphi \in C_c^2(\mathbb{R}^d)$ ,  $C_t(\varphi)$  is continuous by the martingale problem. It is easy to extend the martingale problem to  $\varphi = 1$  by taking limits and the continuity of  $t \rightarrow C_t(1)$  follows. This establishes the required continuity. Since super-Brownian motion has a continuous local time process with compact support in any finite time interval, by Sugitani's theorem, it follows that the process  $X$  also has a local time density  $L_t^X(x)$  with the advertised properties.

Unfortunately, we cannot directly apply the previous lemma to conclude (2.7), because  $L_t^X$  is not *uniformly* bounded in  $\omega$ . To circumvent this problem we use a localization argument. Fix  $0 < b < \infty$ , and consider the exponential process  $\mathcal{E}_{t \wedge \tau(b)}^B$ , where

$$\tau(b) = \inf\{t : \max_x |B_t(x)| \geq b\}.$$

By Lemma 2.1, the process  $\mathcal{E}_{t \wedge \tau(b)}^B$  is a martingale, and so under the probability measure  $Q^b$  specified by equation (2.7) (with the stopped exponential martingale as the likelihood ratio) the process  $X$  satisfies the martingale problem (1.5) with  $K(x) + \beta L(s, X, x)$  replaced by its stopped value. But the preceding paragraph implies that for each  $t$ ,  $Q^b(\tau(b) \leq t) \rightarrow 0$  as  $b \rightarrow \infty$ , i.e.,  $\lim_{b \rightarrow \infty} E_\mu(\mathcal{E}_{t \wedge \tau(b)}^B \mathbf{1}_{\tau(b) \leq t}) = 0$ . Therefore

$$\begin{aligned} E_\mu(\mathcal{E}_t^B) &\geq E_\mu(\mathcal{E}_t^B \mathbf{1}_{\tau(b) > t}) = E_\mu(\mathcal{E}_{t \wedge \tau(b)}^B \mathbf{1}_{\tau(b) > t}) \\ &= E_\mu(\mathcal{E}_{t \wedge \tau(b)}^B) - E_\mu(\mathcal{E}_{t \wedge \tau(b)}^B \mathbf{1}_{\tau(b) \leq t}) \rightarrow 1. \end{aligned}$$

On the other hand, by Fatou's lemma,  $E_\mu(\mathcal{E}_t^B) \leq 1$ , and so  $E_\mu(\mathcal{E}_t^B) = 1$ . It follows that  $\mathcal{E}_t^B$  is a martingale under  $P_\mu$  and that under the probability measure defined by (2.7) the process  $X$  satisfies the martingale problem (1.5).

(b, c) These are easy consequences of (a), the continuity of  $\mu \rightarrow P_\mu$ , and Sugitani's theorem.

(d) It suffices to consider a finite-valued  $\tau$ . By (c), the local time  $L_t$  is the occupation density of  $X_t$  under  $P_{\mu, K}$ , so it follows that

$$L_{\tau+t}(X, x) = L_\tau(X, x) + L_t(X_{\tau+}, x) \text{ for all } (t, x) \text{ almost surely.}$$

If  $Q(\omega)$  is a regular conditional probability for  $X_{\tau+}$  given  $\mathcal{F}_\tau$ , then it follows easily from this that almost surely under  $Q(\omega)$  the coordinate process satisfies the martingale problem (1.5) with  $K$  replaced by  $K + L_\tau$ . Therefore, by the uniqueness in law of solutions,  $Q(\omega) = P_{X_\tau(\omega), K(\omega) + \beta L_\tau(\omega)}$  almost surely. The strong Markov property now follows.

(e) This follows immediately from (a).  $\square$

In the course of proving (a) we have also established the following:

PROPOSITION 2.5. *Let  $X_t$  be a solution of the martingale problem (1.5) where  $\mu$  and  $K$  are as in Theorem 2.2. Then on some probability space a version of  $X$  can be coupled with a dominating super-Brownian motion  $\bar{X}$ , with the same initial mass distribution  $\mu$ , and drift  $\theta$ , so that  $\bar{X}_t \geq X_t$  for all  $t \geq 0$  a.s. We will call  $\bar{X}$  the super-Brownian motion envelope.*

REMARK 2.6. Lemma 2.1 holds equally well on the larger space of continuous  $\mathcal{M}(\mathbb{R}^d)$ -valued paths (as in Perkins (2002)). The proof of Theorem 2.2 also holds on this larger space if one starts with compactly supported initial conditions. That is, the solutions necessarily have compact supports for all  $t$  by the domination in (a). This slightly strengthens the uniqueness part and may be used implicitly below without further comment. The main reason for restricting to compactly supported measures is the use of Proposition 2.9 below in the proof of our main result Theorem 1.2.

2.2. *Discrete epidemic models.* Measure-valued processes that satisfy the martingale problem (1.2) arise naturally as weak limits of discrete, finite-population stochastic models of spatial epidemics. Here we describe one such class of models, following Lalley (2009) and Lalley and Zheng (2010). Several of the couplings we shall develop later in the paper involving measure-valued epidemics will be constructed by first building corresponding couplings for discrete epidemics, then using the weak convergence of the discrete to the measure-valued processes to prove that they extend to the measure-valued setting.

The discrete *SIR-d* epidemic models take place in populations of size  $N$  located at each of the sites of the integer lattice  $\mathbb{Z}^d$ . We shall call  $N$  the *village size*. Each of the  $N$  individuals (or *particles*) at a site  $x \in \mathbb{Z}^d$  may at any time be either *susceptible*, *infected*, *recovered*, or *removed*. Infected individuals remain infected for one unit of time, and then recover, after which they are immune to further infection. The rules governing the transmission of infection are as follows: at each time  $i = 1, 2, \dots$ , for each pair  $(i_x, s_y)$  of an infected individual located at  $x$  and a susceptible individual at  $y$ ,  $i_x$  infects  $s_y$  with probability  $p_N(x, y)$ , where

$$(2.9) \quad p_N(x; y) = p_N^\theta(x; y) = \frac{1 + \theta/N^\alpha}{(2d + 1)N} \quad \text{if } |y - x| \leq 1 \quad \text{and } = 0 \text{ otherwise,}$$

where  $|z|$  is the Euclidean norm of  $z$  and

$$\alpha = \alpha(d) = 2/(6 - d)$$

is the critical exponent, see Theorem 1 in Lalley (2009) and Theorem 2 in Lalley and Zheng (2010). For the *SIR-d* model with village size  $N$ , define

$$\begin{aligned} \mathbb{X}_i^N(x) &:= \text{set of infected particles at } x \text{ at time } i; & X_i^N(x) &:= |\mathbb{X}_i^N(x)|; \\ \mathbb{K}^N(x) &:= \text{set of removed particles at } x \text{ (at time 0);} & K^N(x) &:= |\mathbb{K}^N(x)|; \\ \mathbb{R}_n^N(x) &:= \text{set of recovered particles at } x \text{ at time } n; & R_n^N(x) &:= |\mathbb{R}_n^N(x)|; \quad \text{and} \\ \mathbb{X}_i^N &:= \cup_x \mathbb{X}_i^N(x), & \mathbb{K}^N &:= \cup_x \mathbb{K}^N(x), \quad \text{and} & \mathbb{R}_n^N &:= \cup_x \mathbb{R}_n^N(x). \end{aligned}$$

Theorem 1.2 and Proposition 2.9 below suggest, after an interchange of limits, that the critical infection probability  $p_c(N)$  for the SIR model satisfy

$$(2.10) \quad 0 < cN^{-\alpha} \leq (2d + 1)N \cdot p_c(N) - 1 \leq CN^{-\alpha} \text{ for large } N \text{ and } d = 2, 3.$$

This would be consistent with the result (1.6) for the long-range contact process. Whether or not there is a stronger relation (as in (1.7)) involving the exact constants  $\theta_c$  in Theorem 1.2 is another interesting open question.

**The Standard Construction.** We now describe a way to construct this process using a *percolation structure*. Connections between SIR epidemics and bond percolation go back at least to Mollison (1977) (see p. 322) in the continuous setting and were used extensively by Cox and Durrett (1988), again in the continuous time setting. The construction we use is a modification of the constructions in Lalley (2009) and Lalley and Zheng (2010). We shall call this the *Standard Construction*. The percolation structure is a random graph with vertex set  $\mathbb{Z}^d \times \{1, 2, \dots, N\}$ ; the vertex  $(x, i)$  represents the  $i$ th individual (or particle) in the “village”  $\mathcal{V}_x$  situated at location  $x \in \mathbb{Z}^d$ . For each pair  $(x, i)$  and  $(y, j)$  of vertices whose spatial locations differ by at most 1 (that is,  $|x - y| \leq 1$ ) a  $p_N$ -coin toss determines whether or not there is an edge between  $(x, i)$  and  $(y, j)$ . (As is often the case in such constructions, it is useful, for comparison purposes, to assume that these coin tosses are realized using independent Uniform-[0, 1] random variables.) Thus, the resulting random graph  $\mathcal{G} = \mathcal{G}^N$  has edges only between vertices in the same or neighboring villages.

The spatial epidemic is defined by a deterministic algorithm on the random graph  $\mathcal{G}$ . Since the village size  $N$  is fixed in this algorithm, we shall omit all superscripts  $N$  in the specification of the algorithm. The colors *green*, *blue*, *red*, and *black* will be used to denote susceptible, infected, recovered, and removed vertices in each generation. For the 0th generation, designate  $K(x)$  vertices at location  $x$  as *black*; the set of black vertices will not change during the course of the epidemic. Next, color  $X_0(x)$  vertices in  $\mathcal{V}_x$  *blue*, and all remaining vertices *green*. (Thus, in generation 0 there are no red vertices.) Now define a time evolution as follows. In generation  $n + 1$ , the set  $\mathbb{X}_{n+1}(x)$  of blue vertices will consist of all vertices that were green in generation  $n$  and were connected by edges of the random graph to blue vertices (that is, vertices in  $\mathbb{X}_n = \cup_y \mathbb{X}_n(y)$ ). Finally, all vertices that were blue in generation  $n$  become red in generation  $n + 1$ , and remain red in all subsequent generations (that is,  $\mathbb{R}_{n+1}(x) = \mathbb{R}_n(x) \cup \mathbb{X}_n(x)$ ).

The virtue of this construction is that all quantities of interest can easily be described in terms of the geometry of the random graph  $\mathcal{G}' = \mathcal{G} \setminus \mathbb{K}$  obtained by deleting all black vertices from  $\mathcal{G}$ . The set  $\mathbb{X}_n(x)$  consists of all vertices at site  $x$  at distance  $n$  in the graph  $\mathcal{G}'$  from the set of vertices that were colored blue in generation 0. Similarly, the set  $\mathbb{R}_n(x)$  consists of all vertices at site  $x$  at distance  $< n$  from the blue vertices in generation 0. The set  $\mathbb{R}_\infty$  of vertices that are *ever* infected during the course of the epidemic is the union of the connected clusters of the blue vertices of generation 0 in  $\mathcal{G}'$ . It is immediately obvious from this that the recovered sets  $\mathbb{R}_n(x)$  are non-increasing in the initial condition  $\mathbb{K}(\cdot)$ , and non-decreasing in  $\mathbb{X}_0(\cdot)$  and the transmission parameter  $\theta$ . □

Denote by

$$(2.11) \quad \mathbb{P}^n = (P_n(x, y))_{x, y \in \mathbb{Z}^d} = (P_n(y - x))_{x, y \in \mathbb{Z}^d}$$

the transition probability kernel of the simple random walk on  $\mathbb{Z}^d$ , that is,  $\mathbb{P}^n = \mathbb{P} * \mathbb{P}^{n-1}$  is the  $n$ th convolution power of the one-step transition probability kernel given by

$$(2.12) \quad P_1(x, y) = 1/(2d + 1) \quad \text{for } |y - x| \leq 1,$$

and let  $\sigma^2 = 2/(2d + 1)$  be the variance of the distribution  $P_1(0, \cdot)$ . Let  $G_n(x, y)$  be the associated Green's function

$$G_n(x, y) := \sum_{1 \leq i < n} P_i(x, y), \quad G_n(x) := G_n(0, x),$$

and for any finite measure  $\mu$  on  $\mathbb{Z}^d$  denote by  $\mu G_n(x) = (\mu * G_n)(x)$  the convolution of  $\mu$  with  $G_n$ .

Next, we explain the re-scaling of the discrete epidemics that gives weak convergence to the measure-valued epidemics determined by the martingale problem (1.5).

DEFINITION 2.7. The *Feller-Watanabe scaling operator*  $\mathcal{F}_N$  scales mass by  $1/N^\alpha$  and space by  $1/\sqrt{N^\alpha \sigma^2}$ , that is, for any finite Borel measure  $\mu$  on  $\mathbb{R}^d$  and any test function  $\varphi$ ,

$$(2.13) \quad \langle \varphi, \mathcal{F}_N \mu \rangle = N^{-\alpha} \int \varphi(x/\sqrt{N^\alpha \sigma^2}) \mu(dx).$$

DEFINITION 2.8. The *Sugitani scaling operator*  $\mathcal{S}_N$  scales mass by  $1/N^{\alpha(2-d/2)}$  and space by  $1/\sqrt{N^\alpha \sigma^2}$ , that is, for any function  $f$ ,

$$(2.14) \quad (\mathcal{S}_N f)(x) = \frac{f(\sqrt{N^\alpha \sigma^2} x)}{N^{\alpha(2-d/2)}}.$$

When the function  $f$  is only defined for  $x \in \mathbb{Z}^d$ , define  $(\mathcal{S}_N f)(x)$  for  $x \in \mathbb{Z}^d/[\sqrt{N^\alpha \sigma^2}]$  as above and extend it to a continuous function on  $\mathbb{R}^d$  by a suitable piecewise linear interpolation.

The following weak convergence theorem is a slight variant of the main results in Lalley (2009) and Lalley and Zheng (2010).

PROPOSITION 2.9. Assume that  $d \leq 3$ , and suppose that the initial configurations  $\mu^N := X_0^N$  and  $K^N$  are both supported by finitely many integer sites and are such that for some measure  $\mu$  satisfying Assumption 2.10 below and some  $K \in C_p(\mathbb{R}^d, \mathbb{R}_+)$ , the following conditions are satisfied, where  $\Longrightarrow$  denotes weak convergence on the respective spaces:

(a) if  $d = 1$ , then  $\mu^N(\sqrt{N^\alpha \sigma^2} \cdot)/\sqrt{N^\alpha}$  are supported by a common compact interval, and (after linear interpolation to be continuous functions on  $\mathbb{R}$ )

$$(2.15) \quad \frac{\mu^N(\sqrt{N^\alpha \sigma^2} x)}{\sqrt{N^\alpha}} \Longrightarrow X_0(x) \in C_c(\mathbb{R});$$

(b) if  $d = 2$  or  $3$ , then

$$(2.16) \quad \mathcal{F}_N \mu^N \Longrightarrow \mu,$$

$$(2.17) \quad \mathcal{S}_N(\mu^N * G_{[N^\alpha t]}) \Longrightarrow \mu * q_t \in C_b([0, \infty) \times \mathbb{R}^d),$$

where the second convergence is in  $D([0, \infty); C_b(\mathbb{R}^d))$ ; and

(c) in all dimensions,

$$(2.18) \quad K^N(x) = [N^{\alpha(2-d/2)} \cdot K(x/[\sqrt{N^\alpha \sigma^2}])], \quad \text{for all } x \in \mathbb{Z}^d.$$

Then we have the following weak convergence

$$(2.19) \quad (\mathcal{F}_N X_{[N^\alpha t]}^N, \mathcal{S}_N R_{[N^\alpha t]}^N) \implies (X_t, L_t(x))$$

in  $D([0, \infty); \mathcal{M}_c(\mathbb{R}^d)) \times D([0, \infty); C_b(\mathbb{R}^d))$ , where the limit process  $X$  has initial configuration  $X_0 = \mu$ , solves (1.5) with  $\gamma = 1$ ,  $\beta = 1$ ,  $\theta$  as in (2.9), and suppression rate  $K$ , and  $L_t(x)$  is its local time density process.

ASSUMPTION 2.10. *The finite measure  $\mu$  has compact support. When  $d = 1$ ,  $\mu$  has a density  $X_0(x) \in C_c(\mathbb{R})$ , and for  $d = 2, 3$  for some  $C_\mu > 0$ ,  $\mu$  satisfies*

$$(2.20) \quad \sup_{x \in \mathbb{R}^d} \mu(B(x, r)) \leq \begin{cases} C_\mu (\log 1/r)^{-3} & \text{if } d = 2, \\ C_\mu r (\log 1/r)^{-2} & \text{if } d = 3, \end{cases} \quad \text{for all } r \in (0, 1].$$

REMARK 2.11. It is easy to see that Assumption 2.10 implies Assumption 1.1. Take the case  $d = 3$  for example. For any  $(t_n, x_n) \rightarrow (t, x)$ , we want to show that  $\int_y q_{t_n}(y - x_n) d\mu(y) \rightarrow \int_y q_t(y - x) d\mu(y)$ . Since  $\mu(\{x\}) = 0$ , we have

$$q_{t_n}(y - x_n) \rightarrow q_t(y - x) \text{ for } \mu\text{-a.a. } y,$$

and hence it suffices to show that  $\{q_{t_n}(y - x_n)\}$  is uniformly integrable with respect to  $\mu$ , which in turn reduces to show

$$\limsup_{\delta \rightarrow 0} \sup_n \int_{|y - x_n| \leq \delta} q_{t_n}(y - x_n) d\mu(y) = 0.$$

To see this, let  $M(r) = \mu(B(x_n, r))$  for  $r \geq 0$ . The elementary bound  $q_t(z) \leq C|z|^{-1}$  and an integration by parts lead to

$$(2.21) \quad \begin{aligned} \int_{|y - x_n| \leq \delta} q_{t_n}(y - x_n) d\mu(y) &\leq C \int_{|y - x_n| \leq \delta} |y - x_n|^{-1} d\mu(y) \\ &= C \int_0^\delta r^{-1} dM(r) \\ &= Cr^{-1} M(r)|_0^\delta + C \int_0^\delta r^{-2} M(r) dr \\ &\leq C(\log(1/\delta))^{-2} + C \int_0^\delta r^{-1} (\log(1/r))^{-2} dr, \end{aligned}$$

which goes to 0 as  $\delta \rightarrow 0$ . A similar argument applies for  $d = 2$ .

REMARK 2.12. For any  $\mu \in \mathcal{M}_c(\mathbb{R}^d)$  and any fixed  $\theta \in \mathbb{R}, \gamma > 0$ , by Theorems III.4.2 and III.3.4. in Perkins (2002), if  $X$  is a super-Brownian motion with initial mass distribution  $\mu$ , drift  $\theta$ , and branching rate  $\gamma$ , then Assumption 2.10 is satisfied by  $X_t$  for all  $t > 0$

almost surely. Furthermore, for any  $K \in C_p(\mathbb{R}^d, \mathbb{R}_+)$  and  $\beta > 0$ , by the absolute continuity between the laws  $P_{\mu,0}^{\theta,0,\gamma}$  and  $P_{\mu,K}^{\theta,\beta,\gamma}$ , the same is true for a spatial epidemic with initial mass distribution  $\mu$ , local suppression rate  $K$ , transmission rate  $\theta$ , branching rate  $\gamma$  and inhibition parameter  $\beta$ .

REMARK 2.13. For  $\mu$  satisfying Assumption 2.10 there are rescaled counting measure  $\mu^N$  satisfying the hypotheses of the above theorem and hence Proposition 2.9 implies, among other things, that for any suppression rate function  $K \in C_p(\mathbb{R}^d, \mathbb{R}_+)$ , the measure-valued epidemic process  $X$  satisfying (1.5) is a weak limit of appropriately scaled discrete *SIR* epidemics. When  $d = 1$ , for each  $x \in \mathbb{Z}/\sqrt{N^\alpha\sigma^2}$ , let  $\mu^N(x\sqrt{N^\alpha\sigma^2}) = [\sqrt{N^\alpha} \cdot X_0(x)]$ . Then (2.15) is obvious. When  $d = 2$  or  $3$ , the required sequence  $\{\mu^N\}$  satisfying (2.16) and (2.17) can be built as follows. Recall that for each  $x \in \mathbb{Z}^d$ ,  $Q(x)$  stands for the (half-closed, half-open) unit cube centered at  $x$ .  $\mathbb{R}^d$  can hence be decomposed as a non-overlapping union of  $Q(x)$ 's for  $x \in \mathbb{Z}^d$ , and so for any  $y \in \mathbb{R}^d$ , we can find a unique  $x \in \mathbb{Z}^d$  such that  $y \in Q(x)$ , and with a slight abuse of notation, denote such an  $x$  by  $[y]$ . Next, let  $\{X_i\}$  be a sequence of i.i.d. random variables with probability distribution  $\mu/|\mu|$ , and let

$$\mu^N = \sum_{i=1}^{\lfloor N^\alpha \cdot |\mu| \rfloor} \delta_{[X_i \sqrt{N^\alpha \sigma^2}]}.$$

(Note that  $\alpha < 1$  and on each integer site there are  $N$  vertices, so for all  $N$  large enough,  $\mu^N$  can be realized as a counting measure on the graph  $\mathbb{Z}^d \times \{1, 2, \dots, N\}$ .) One can then show that  $\{\mu^N\}$  satisfies (2.16) and (2.17) almost surely. In fact, (2.16) holds trivially by the strong law of large numbers (SLLN), the uniform continuity of test functions, and the simple bound

$$(2.22) \quad \left| \frac{[X_i \sqrt{N^\alpha \sigma^2}]}{\sqrt{N^\alpha \sigma^2}} - X_i \right| \leq \frac{1}{\sqrt{N^\alpha \sigma^2}}, \quad \text{for all } i.$$

The verification of (2.17) is given in the Appendix.

REMARK 2.14. The arguments of Lalley (2009) and Lalley and Zheng (2010) are based on the fact that each of the discrete *SIR* epidemics has law absolutely continuous with respect to the law of a critical branching random walk with the same initial condition. The Radon-Nikodym derivatives can be written explicitly as products, and these can be shown to converge to exponentials of the form  $\mathcal{E}_t^B$  appearing in (2.7). Since branching random walks, after rescaling, converge to super-Brownian motions, it follows that the rescaled discrete *SIR* epidemics converge to processes related to super-Brownian motion by (2.7), that is, processes that solve the martingale problem (1.2).

Routine modifications of these arguments can be used to establish weak convergence for a variety of discrete processes similar to or related to the discrete *SIR* epidemics constructed above. In particular, the convergence (2.19) can be extended to *joint* weak convergence for *coupled* *SIR* epidemics with suitable initial conditions. For example, let  $\mu^{N,A}, \mu^{N,B}$  be initial conditions satisfying the hypotheses (2.15)–(2.17), and let  $\mathbb{X}^{N,A}, \mathbb{X}^{N,B}, \mathbb{X}^N$  be discrete *SIR* epidemics all constructed using the same percolation structure  $\mathcal{G}^N$ , with the same initially

removed sets  $\mathbb{K}^N(x)$ , in such a way that the sets  $\mathbb{X}_0^{N,A}(x)$  and  $\mathbb{X}_0^{N,B}(x)$  are non-overlapping, with cardinalities  $\mu^{N,A}(x)$  and  $\mu^{N,B}(x)$ , and such that

$$(2.23) \quad \mathbb{X}_0^N(x) = \mathbb{X}_0^{N,A}(x) \cup \mathbb{X}_0^{N,B}(x).$$

Then after rescaling, the processes  $X^{N,A}$ ,  $X^{N,B}$ , and  $X^N$  converge *jointly* in law to (dependent) measure-valued epidemics  $X_t^A$ ,  $X_t^B$ , and  $X_t$ , with initial mass distributions  $\mu^A$ ,  $\mu^B$ , and  $\mu^A + \mu^B$ , respectively, whose local time densities satisfy

$$(2.24) \quad L_t^A \vee L_t^B \leq L_t \leq L_t^A + L_t^B.$$

(The arguments that follow will not rely in an essential way on this joint convergence, however. All that is needed is that *subsequences* converge jointly, as this is enough to guarantee the existence of coupled measure-valued processes satisfying the same monotonicity properties (such as (2.24)). Joint convergence along subsequences follows trivially from the weak convergence of marginals, since this implies joint tightness.)

□

**2.3. Comparison lemmas.** The construction of the measure-valued spatial epidemic process as the weak limit of discrete epidemics and the Girsanov formulas (2.7)–(2.8) lead to a number of basic comparison principles that will be used in the proof of Theorem 1.2. We formulate these as *couplings*, in which two epidemic processes (or super-Brownian motions) are constructed on a common probability space in such a way that various functionals of the processes (e.g., the limiting local time densities  $L_\infty(x)$ ) are ordered.

**LEMMA 2.15.** *Suppose that  $X_t$  has law  $P_{\mu,K}^{\theta,1,\gamma}$  for some  $K \in C_p(\mathbb{R}^d, \mathbb{R}_+)$  and some initial condition  $\mu$  that satisfies Assumption 1.1. Then*

$$P(X \text{ survives}) = P\left(\lim_{t \rightarrow \infty} \int_{\mathbb{R}^d} L_t(x) dx = \infty\right) = P(\lim_{t \rightarrow \infty} |X_t| = \infty).$$

**PROOF.** This uses the existence of a coupling between the measure-valued epidemic  $X_t$  and its super-Brownian motion envelope (Proposition 2.5). If  $Z_s := |X_s|$  is the total mass at time  $s$ , then  $\lim_{t \rightarrow \infty} \int_{\mathbb{R}^d} L_t(x) dx = \int_0^\infty Z_s ds$ . Because  $Z_s$  is continuous and 0 is an absorbing state (e.g. by the strong Markov property in Theorem 2.2), if  $\int_0^\infty Z_s ds = \infty$  then  $X_t$  must survive. On the other hand, if  $\liminf_{t \rightarrow \infty} Z_t < \infty$ , then there exists  $M \in \mathbb{N}$  and an infinite sequence of stopping times  $\tau_n \rightarrow \infty$  such that

$$(2.25) \quad \tau_{n+1} \geq \tau_n + 1, \quad \text{and} \quad Z_{\tau_n} \leq M.$$

Consider the time period  $[\tau_n, \tau_n + 1]$ . By the strong Markov property and the existence of a monotone coupling between a spatial epidemic and its super-Brownian motion envelope, the process  $Z_{t+\tau_n}$  is dominated by Feller diffusion with drift  $\theta$  and initial total mass less than  $M$ . This dies out in the next one unit of time with positive probability, independent of  $n$ , hence so does  $X_{t+\tau_n}$  for  $t \leq 1$ . It follows that with probability 1, if  $X$  survives then  $\lim_{t \rightarrow \infty} Z_t = \infty$ . As the latter trivially implies  $\int_0^\infty Z_s ds = \infty$ , the proof is complete. □

LEMMA 2.16. *Fix  $\theta < \theta^*$  and  $\gamma > 0$ . For any initial mass distribution  $\mu$  that satisfies Assumption 2.10 and any  $K \in C_p(\mathbb{R}^d, \mathbb{R}_+)$ , there exist on some probability space epidemic processes  $(X, X^*) \in D([0, \infty); \mathcal{M}_c(\mathbb{R}^d))^2$  with laws  $P_{\mu, K}^{\theta, 1, \gamma}$  and  $P_{\mu, K}^{\theta^*, 1, \gamma}$  and local time densities  $L_t, L_t^*$ , respectively, such that almost surely, for every  $t \geq 0$ ,*

$$(2.26) \quad L_t \leq L_t^*.$$

PROOF. This follows from the weak convergence result (2.19) and the Standard Construction. Recall that in the standard construction of the discrete *SIR* epidemics, the evolution is determined by the random graph  $\mathcal{G}'$  in which edges are present with probabilities  $p_N(x; y) = p_N^\theta(x; y)$  given by (2.9). These probabilities are increasing in  $\theta$ . Consequently, it is possible (using auxiliary Uniform-[0,1] random variables) to simultaneously construct random graphs  $\mathcal{G}'$  and  $\mathcal{G}'_*$  with percolation probabilities  $p_N^\theta(x; y)$  and  $p_N^{\theta^*}(x; y)$ , respectively, in such a way that the edge set of  $\mathcal{G}'$  is contained in that of  $\mathcal{G}'_*$ . This forces

$$\mathbb{R}_n(x) \subseteq \mathbb{R}_n^*(x) \quad \text{for all } n \geq 0 \quad \text{and } x \in \mathbb{Z}^d,$$

and hence also  $R_n(x) \leq R_n^*(x)$ . This inequality will be preserved upon taking weak limits, so we obtain (2.26).  $\square$

REMARK 2.17. It follows immediately from (2.26) and Lemma 2.15 that

$$(2.27) \quad P(X \text{ survives}) \leq P(X^* \text{ survives}).$$

LEMMA 2.18. *Let  $K, K^* \in C_p(\mathbb{R}^d; \mathbb{R}_+)$  be suppression rate functions such that  $K \leq K^*$ . Then for any  $\mu$  satisfying Assumption 2.10, there exist spatial epidemics  $(X, X^*) \in D([0, \infty); \mathcal{M}_c(\mathbb{R}^d))^2$  with marginal laws  $P_{\mu, K}^{\theta, 1, \gamma}$  and  $P_{\mu, K^*}^{\theta, 1, \gamma}$  and local time densities  $L_t, L_t^*$ , respectively, such that almost surely,*

$$(2.28) \quad L_t \geq L_t^* \quad \text{for all } t \geq 0,$$

and  $P(X \text{ survives}) \geq P(X^* \text{ survives})$ .

PROOF. The existence of the coupling follows directly from the weak convergence (2.19) and the Standard Construction, because in this construction, increasing the removed sets  $\mathbb{K}$  decreases the sizes of the connected components. The assertion about survival probabilities follows from (2.28), by Lemma 2.15.  $\square$

LEMMA 2.19. *Let  $\mu_0, \nu_0$  be initial mass distributions satisfying Assumption 2.10, and  $\mu = \mu_0 + \nu_0$ . Then on some probability space there exist epidemic processes  $(X, X^1, X^2) \in D([0, \infty); \mathcal{M}_c(\mathbb{R}^d))^3$ , with initial conditions  $\mu, \mu_0$ , and  $\nu_0$  and marginal laws  $P_{\mu, 0}^{\theta, 1, \gamma}$ ,  $P_{\mu_0, 0}^{\theta, 1, \gamma}$ , and  $P_{\nu_0, 0}^{\theta, 1, \gamma}$  respectively, such that*

$$(2.29) \quad \max(L_t^1(x), L_t^2(x)) \leq L_t(x) \leq L_t^1(x) + L_t^2(x), \quad \text{for all } t \geq 0 \text{ and } x \in \mathbb{R}^d.$$

Consequently,

$$(2.30) \quad P(X \text{ survives}) \leq P(X^1 \text{ survives}) + P(X^2 \text{ survives}) \quad \text{and}$$

$$(2.31) \quad P(X \text{ survives}) \geq P(X^i \text{ survives}), \quad \text{for each } i = 1, 2.$$

PROOF. This follows by the same argument as the preceding lemma (see Remark 2.14).  $\square$

Lemma 2.19 describes the effect of adding infected mass at time 0. The next lemma concerns the effect of introducing additional infected mass at a time  $t_* > 0$  after the epidemic has already begun. Let  $\mu, \mu_0, \nu_0$  be initial mass distributions satisfying the hypotheses of Lemma 2.19. Say that  $X_t^*$  is a *measure-valued epidemic with immigration* at time  $t_*$  if it satisfies the following martingale problem: For all  $\varphi \in C_c^2(\mathbb{R}^d)$ ,

$$(2.32) \quad X_t^*(\varphi) = \mu_0(\varphi) + \mathbf{1}_{[t_*, \infty)}(t)\nu_0(\varphi) + \int_0^t X_s^*(\Delta\varphi/2 + \theta\varphi - K\varphi - \beta L_s^*\varphi) ds + \sqrt{\gamma}M_t^*(\varphi),$$

where  $M_t^*$  is a continuous martingale with quadratic variation  $[M^*(\varphi)]_t = \int_0^t X_s^*(\varphi^2) ds$  and  $L^* = L^{X^*}$  is the local time density of  $X^*$ . Existence and uniqueness of solutions to (2.32) follows from Theorem 2.2 and the Markov property.

LEMMA 2.20. *Let  $\mu_0, \nu_0$  be initial mass distributions satisfying Assumption 2.10, and  $\mu = \mu_0 + \nu_0$ . On some probability space there exist a solution  $X_t$  to (1.5) with initial mass distribution  $\mu$  and a solution  $X_t^*$  to (2.32) such that*

$$(2.33) \quad L_t^X \geq L_t^{X^*} \quad \forall t \geq 0 \quad \text{and} \quad L_\infty^X = L_\infty^{X^*}.$$

PROOF. This is by discrete approximation, using the Standard Construction of the discrete *SIR* epidemics. On each percolation structure  $\mathcal{G} = \mathcal{G}^N$ , we construct a pair of epidemics. The first, denoted by  $\mathbb{X} = \mathbb{X}^N$ , is constructed using initially infected sets  $\mathbb{X}_0(x) = \mathbb{X}_0^N(x)$  such that both (2.16) and (2.17) hold. The second, denoted by  $\mathbb{Y} = \mathbb{Y}^N$ , has initially infected sets  $\mathbb{Y}_0(x) \subseteq \mathbb{X}_0(x)$  such that after Feller-Watanabe rescaling the initial mass distributions converge to  $\mu_0$  (see Remark 2.13 in Section 2.2). This second epidemic  $\mathbb{Y}$  has spontaneous new infections at time  $[N^\alpha t_*]$ : in particular, all individuals in the sets

$$\mathbb{X}_0 \setminus \mathbb{Y}_0 := \cup_x (\mathbb{X}_0(x) \setminus \mathbb{Y}_0(x))$$

who are not yet recovered become infected. Thus, the time evolution of the epidemic  $\mathbb{Y}_n$  is determined by the random graph  $\mathcal{G}' := \mathcal{G} \setminus \cup_x \mathbb{K}(x)$  as follows: (1) For  $n < [N^\alpha t_*]$ , the recovered set  $\mathbb{R}_n^Y$  consists of all vertices at graph distance  $< n$  from the initially infected set  $\cup_x \mathbb{Y}_0(x)$ . (2) For  $n \geq [N^\alpha t_*]$ , the set  $\mathbb{R}_n^Y$  consists of all vertices  $v$  such that either the graph distance of  $v$  from  $\cup_x \mathbb{Y}_0(x)$  is  $< n$ , or the graph distance of  $v$  from  $\mathbb{X}_0 \setminus \mathbb{Y}_0$  is  $< n - [N^\alpha t_*]$ .

From the construction above and the Standard Construction described earlier, it is clear that

$$(2.34) \quad \text{for all } n \geq 0, \mathbb{R}_n^X \supseteq \mathbb{R}_n^Y, \text{ and } \mathbb{R}_n^X \subseteq \mathbb{R}_{n+[N^\alpha t_*]}^Y.$$

Set

$$Y_t^N = \mathcal{F}_N | \mathbb{Y}_{[N^\alpha t]} | \quad \text{and} \quad R_t^{Y,N} = \mathcal{S}_N | \mathbb{R}_{[N^\alpha t]}^Y |,$$

where  $\mathcal{F}_N$  and  $\mathcal{S}_N$  are the Feller-Watanabe and Sugitani rescaling operators.

CLAIM 2.21. *The vector-valued process  $(Y^N, R^{Y,N})$  converges weakly to a process  $(X^*, L^*)$  such that  $X^*$  solves the martingale problem (2.32), and  $L^* = L^{X^*}$  is the local time density of  $X^*$ .*

By passing to a subsequence, if necessary, it follows from Proposition 2.9, the above Claim, and (2.34) that

$$L_t^{X^*} \leq L_t^Y \leq L_{t+t_*}^{X^*} \text{ for all } t \geq 0 \text{ a.s.}$$

This clearly implies (2.33).

PROOF OF THE CLAIM (SKETCH). This is done by following the likelihood ratio strategy described in Remark 2.14. As used in Lalley (2009) and Lalley and Zheng (2010), this strategy was based on the fact that each discrete *SIR* epidemic considered had law absolutely continuous with respect to the law of a critical branching random walk with the same initial condition. The bulk of the proof consisted of showing that the likelihood ratios converged in law, *under the branching random walk measure*, to the Radon-Nikodym derivative of a measure-valued epidemic relative to the law of super-Brownian motion. For the processes considered in this claim, the appropriate comparison processes are not standard branching random walks, but rather *branching random walks with immigration* in which new particles are introduced at times  $[N^\alpha t_*]$  in such a way that after Feller-Watanabe rescaling the mass distributions of these new particles converge to  $\nu_0$ . The laws of these processes converge, after rescaling, to the law of super-Brownian motion with immigration at time  $t_*$ , that is, a process  $Y^*$  satisfying the martingale problem (2.32) with  $\beta = 0$  and  $K = 0$ . (This follows easily from the standard convergence theorem for critical branching random walks, because the effect of the immigration is simply to superimpose an independent branching random walk shifted in time by  $[N^\alpha t_*]$ .)

Consider the likelihood ratios for the law of the epidemic process  $Y^N$  relative to that of the corresponding branching random walk with immigration. These are products of factors indexed by (discrete) times  $t$  and lattice sites  $x \in \mathbb{Z}^d$  (see Lalley (2009), eqn. (53)). For  $t \leq [N^\alpha t_*]$  the factors are exactly the same as in the case where there is no immigration. Beginning with time  $t = [N^\alpha t_*]$ , new factors are introduced; these indicate the relative likelihood ratios for the newly introduced immigrants and their offspring. Under the law of the branching random walks with immigration the immigrants and their offspring evolve *independently* of the progeny of the original (time 0) particles. Using this fact, one can show, in much the same manner as in Lalley (2009) and Lalley and Zheng (2010), that the likelihood ratios converge weakly (under the branching random walk with immigration laws) to the Radon-Nikodym derivative of the process  $X^*$  relative to super-Brownian motion with immigration. In carrying out this final step, the main hurdle is showing that in the epidemics with immigration, the numbers of individuals in the sets  $\mathbb{X}_0 \setminus \mathbb{Y}_0$  who are infected prior to time  $[N^\alpha t_*]$  is of order  $O_P(1)$ . Here is a brief synopsis of the argument: Since the local time densities, after rescaling, converge, the maximum number of recovered individuals at time  $[N^\alpha t_*]$  at any site is of order  $O_P(N^{\alpha(2-d/2)})$ . Consequently, because  $\mathbb{X}_0 \setminus \mathbb{Y}_0$  has cardinality on the order  $N^\alpha$ , the number of individuals in  $\mathbb{X}_0 \setminus \mathbb{Y}_0$  infected prior to time  $[N^\alpha t_*]$  is of order

$$(2.35) \quad O(N^\alpha) \times \frac{O_P(N^{\alpha(2-d/2)})}{N} = O_P(1),$$

since  $\alpha = 2/(6 - d)$ . □

This completes the proof of Lemma 2.20. □

In the proof of Theorem 1.2 it will be necessary to compare the evolution of a measure-valued epidemic  $X$  with a coupled process in which additional infected mass is introduced at a *random* time. Say that  $X_t^*$  is a *measure-valued epidemic with immigration* at time  $\tau$  if it satisfies the following martingale problem : For all  $\varphi \in C_c^2(\mathbb{R}^d)$ ,

$$(2.36) \quad X_t^*(\varphi) = \mu_0(\varphi) + \mathbf{1}_{[\tau, \infty)}(t)\nu_0(\varphi) + \int_0^t X_s^* (\Delta\varphi/2 + \theta\varphi - K\varphi - \beta L_s^* \varphi) ds + \sqrt{\gamma} M_t^*(\varphi),$$

where  $\tau$  is a finite stopping time relative to the filtration  $\mathcal{F}^{X^*}$ ,  $M_t^*$  is a continuous martingale with quadratic variation  $[M^*(\varphi)]_t = \int_0^t X_s^*(\varphi^2) ds$ , and  $L^* = L^{X^*}$  is the local time density.

LEMMA 2.22. *Let  $\mu_0, \nu_0$  be initial mass distributions satisfying Assumption 2.10, and  $\mu = \mu_0 + \nu_0$ . Then on some probability space there exist epidemics  $(X_t, X_t^*) \in D([0, \infty); \mathcal{M}_c(\mathbb{R}^d))^2$  such that (1)  $X_t$  solves the martingale problem (1.5) with initial value  $X_0 = \mu$ ; (2)  $X_t^*$  solves the martingale problem (2.36); and (3)*

$$(2.37) \quad L_t^X \geq L_t^{X^*}, \text{ for all } t \geq 0 \quad \text{and} \quad L_\infty^X = L_\infty^{X^*}.$$

PROOF. By the usual continuity (weak convergence) arguments, it suffices to prove this for stopping times  $\tau$  that take values in a finite set. By a routine induction on the cardinality of this finite set, it suffices to consider stopping times that take values in a two-element set  $\{0, t_*\}$ . For such stopping times, the result follows from Lemma 2.20, since this can be applied conditionally on  $\mathcal{F}_0$ . □

Now the results of Lemmas 2.18, 2.19, and 2.22 can be combined, allowing us to couple the measure-valued epidemic  $X$  with a measure-valued process  $X^*$  in which the infected mass is decreased and the suppression rate increased at a random time  $\tau$ . Here we will use the strong Markov property (Theorem 2.2(c)) and Remark 2.3. The process  $X^*$  will satisfy the following martingale problem: for every  $\varphi \in C_c^2(\mathbb{R}^d)$ ,

$$(2.38) \quad X_t^*(\varphi) = \begin{cases} \mu_0(\varphi) + \int_0^t X_s^* (\Delta\varphi/2 + \theta\varphi - K\varphi - \beta L_s^{X^*} \varphi) ds + \sqrt{\gamma} M_t^*(\varphi), & \text{for all } t < \tau, \\ Y_\tau(\varphi) + \int_\tau^t X_s^* (\Delta\varphi/2 + \theta\varphi - K_\tau^* \varphi - \beta L_s^{X^*} \varphi) ds + \sqrt{\gamma} (M_t^*(\varphi) - M_\tau^*(\varphi)), & \text{for all } t \geq \tau, \end{cases}$$

where  $L_s^{X^*}$  is the local time density and

- (i)  $\tau$  is a finite  $\mathcal{F}^{X^*}$ -stopping time;

- (ii)  $Y_\tau$  is an  $\mathcal{F}_{\tau-}^{X^*}$ -measurable random measure satisfying  $Y_\tau \leq X_{\tau-}^* + \nu_0$ ;
- (iii)  $K_\tau^*$  is an  $\mathcal{F}_\tau^{X^*}$ -measurable random element of  $C_p(\mathbb{R}^d, \mathbb{R}_+)$  satisfying  $K_\tau^* \geq K$ ; and
- (iv)  $M_t^*(\varphi)$  is an  $\mathcal{F}^{X^*}$ -continuous martingale with quadratic variation  $[M^*(\varphi)]_t = \int_0^t X_s^*(\varphi^2) ds$ .

PROPOSITION 2.23. *Let  $\mu_0, \nu_0$  be initial mass distributions satisfying Assumption 2.10, and  $\mu = \mu_0 + \nu_0$ . Then there exist epidemics  $(X_t, X_t^*) \in D([0, \infty); \mathcal{M}_c(\mathbb{R}^d))^2$  such that*

- (i)  $X_t$  solves the martingale problem (1.5) with initial value  $X_0 = \mu$ ;
- (ii)  $X_t^*$  satisfies the martingale problem (2.38); and
- (iii) the local time densities of  $X$  and  $X^*$  satisfy

$$L_t^X \geq L_t^{X^*} \text{ for all } t \geq 0.$$

This proposition can be used iteratively, using the strong Markov property and standard arguments, so as to allow immigration and increases in the suppression rate  $K^*$  at increasing stopping times  $0 = \tau_0 \leq \tau_1 \leq \tau_2 \leq \dots < \infty$ . The associated martingale problem is as follows: for every  $\varphi \in C_c^2(\mathbb{R}^d)$ , and for all  $\tau_i \leq t < \tau_{i+1}$ ,

$$(2.39) \quad \begin{aligned} X_t^*(\varphi) &= \mu_i(\varphi) \\ &+ \int_{\tau_i}^t X_s^* (\Delta\varphi/2 + \theta\varphi - K_i^*\varphi - \beta(L_s(X^*) - L_{\tau_i}(X^*))\varphi) ds + \sqrt{\gamma}(M_t^*(\varphi) - M_{\tau_i}^*(\varphi)), \end{aligned}$$

where

- (i) for each  $i = 1, 2, \dots$ ,  $\mu_i$  and  $\nu_i$  are  $\mathcal{F}_{\tau_i-}^{X^*}$ -measurable random measures satisfying Assumption 2.10 and such that

$$\mu_i + \nu_i \leq X_{\tau_i-}^* + \nu_{i-1};$$

- (ii)  $K_0^* \equiv 0$ , and  $K_i^* \in C_p(\mathbb{R}^d, \mathbb{R}_+)$  is, for each  $i$ , an  $\mathcal{F}_{\tau_i}^{X^*}$ -measurable random function such that

$$K_i^* \geq K_{i-1}^* + \beta(L_{\tau_i}(X^*) - L_{\tau_{i-1}}(X^*)); \quad \text{and}$$

- (iii)  $M_t^*$  is a continuous  $\mathcal{F}^{X^*}$ -martingale with quadratic variation  $[M^*(\varphi)]_t = \int_0^t X_s^*(\varphi^2) ds$ .
- The existence of a solution to this martingale problem follows from the strong Markov property (Theorem 2.2 (d)). Proposition 2.23 and a standard induction argument now yield the following comparison result.

PROPOSITION 2.24. *Let  $\mu_0, \nu_0$  be initial mass distributions satisfying Assumption 2.10, and  $\mu = \mu_0 + \nu_0$ . Then on some probability space there exist measure-valued processes  $X$  and  $X^*$  such that (i)  $X$  solves the martingale problem  $(MP)_{\mu,0}^{\theta,\beta,\gamma}$  specified in (1.5); (ii)  $X^*$  solves the martingale problem (2.39); and (iii) the corresponding local time processes satisfy*

$$L_t^X \geq L_t^{X^*} \text{ for all } t \geq 0.$$

2.4. *The Sandwich Lemma.* The discrete SIR-d process  $X^N$  is naturally associated with a *branching envelope*. This is a nearest-neighbor branching random walk  $\overline{X}_n^N$  with initial

condition  $\bar{X}_0^N = X_0^N$  and offspring distribution  $\text{Bin}((2d+1)N, p_N^\theta(0,0))$  that dominates  $X_n^N$ , that is, such that for each  $n \geq 0$  and  $x \in \mathbb{Z}^d$ ,

$$X_n^N(x) \leq \bar{X}_n^N(x).$$

See Section 1.6 of Lalley (2009) for details concerning the construction. Since branching random walks, after Feller-Watanabe rescaling, converge weakly to super-Brownian motions, the vector-valued processes  $(X^N, \bar{X}^N)$ , similarly rescaled, have marginals that converge weakly. It follows that after rescaling the laws of the vector-valued processes  $(X^N, \bar{X}^N)$  are tight. Hence, any subsequence has a weakly convergent subsequence, and the limit process  $(X, \bar{X})$  must satisfy  $X_t \leq \bar{X}_t$ . The component processes  $X$  and  $\bar{X}$  of any such weak limit must be a measure-valued epidemic (that is, a solution of the martingale problem (1.5) with  $\gamma = 1$ ) and a super-Brownian motion with drift  $\theta$ , respectively. This gives another proof of Proposition 2.5. Next is a result which also gives a *lower bound* on the epidemic process.

LEMMA 2.25. *For any measure  $\mu \in \mathcal{M}_c(\mathbb{R}^d)$  satisfying Assumption 2.10, any  $\kappa > 0$ ,  $\theta \in \mathbb{R}$ , and any function  $K \in C_p(\mathbb{R}^d, \mathbb{R}_+)$  there exist, on some probability space, measure-valued processes  $X, \bar{X}, \underline{X}$  with common initial state  $X_0 = \bar{X} = \underline{X} = \mu$  such that*

$$\underline{X}_t \leq X_t \leq \bar{X}_t \quad \text{for all } t \leq \tau,$$

where

$$\tau = \inf \left\{ t \geq 0 : \left( \max_x K(x) \right) + \left( \max_x L_t^X(x) \right) \geq \kappa \right\},$$

with the following laws:

- (i)  $X_t$  is a spatial epidemic with local suppression rate  $K$ , transmission rate  $\theta$ , branching rate  $\gamma = 1$  and inhibition parameter  $\beta = 1$  ;
- (ii)  $\bar{X}_t$  is a super-Brownian motion with drift  $\theta$ ; and
- (iii)  $\underline{X}_t$  is a super-Brownian motion with drift  $\theta - \kappa$ .

The proof will once again be based on discrete approximations. We shall build approximating discrete epidemic processes that satisfy the analogous sandwich relationship. The construction makes use of the following lemma. First, observe that in a discrete SIR epidemic, when two infected individuals simultaneously attempt to infect the same susceptible individual, all but one of the attempts fail; call such an occurrence a *collision*. (Hence, for example, when three infected individuals simultaneously attempt to infect the same susceptible individual, then the number of collisions would be  $\binom{3}{2} = 3$ .)

By slightly modifying the proof of Lemma 9 in Lalley and Zheng (2010), in particular, by noticing that the statement right above equation (62) therein also holds for the way that we count the number of collisions here, we get the following

LEMMA 2.26 (A slight variant of Lemma 9 and equations (62) – (64) in Lalley and Zheng (2010)). *For each pair  $(n, x) \in \mathbb{N} \times \mathbb{Z}^d$ , let  $\Gamma_n^N(x)$  be the number of collisions at site  $x$  and time  $n$  in the SIR- $d$  epidemic with village size  $N$ . Assume that the hypotheses (2.16)-(2.17) of Proposition 2.9 are satisfied. Then for any fixed  $T \geq 0$ ,*

$$(2.40) \quad E \sum_{n \leq N^\alpha T} \sum_x \Gamma_n^N(x) = o(N^\alpha).$$

A direct consequence of the previous lemma is that if we define a **Modified SIR** process in the following way:

**Modified SIR process:** At any site/time  $(x, t)$ , each particle produces  $\text{Bin}((N - K^N(y) - R_t^N(y)), p_N^\theta(x, y))$  number of offspring at neighboring sites  $y$ , where  $R_k^N(y) = \sum_{i < k} X_i^N(y)$ ,

then the Modified SIR process can be constructed together with the original SIR process, in much the same way as for the branching envelope with the original SIR process, such that: (1) the Modified SIR process always dominates the original SIR process; and (2) the discrepancy  $D_t(x) \geq 0$  between them at site  $x$  and time  $t$  satisfies

$$(2.41) \quad \max_{t \leq N^{\alpha T}} \sum_x D_t(x) = o_P(N^\alpha).$$

Therefore after the Feller-Watanabe scaling as in Proposition 2.9, the modified SIR process will converge to the same limit as in Proposition 2.9.

(To show (2.41), observe that if we let  $D_n := \sum_x D_n(x)$ , and  $\tilde{D}_n = D_n / (1 + \theta/N^\alpha)^n$ , then  $\tilde{D}_n$  is a sub-martingale:  $E(\tilde{D}_{n+1} | \mathcal{F}_n) \geq \tilde{D}_n$ , with the inequality due to collisions at generation  $n + 1$ . Further note that for any  $T > 0$ ,

$$\begin{aligned} E(\tilde{D}_{\lfloor N^{\alpha T} \rfloor}) &\leq \frac{E \sum_{n \leq N^{\alpha T}} ((1 + \theta/N^\alpha)^{[N^{\alpha T}] - n} \cdot \sum_x \Gamma_n^N(x))}{(1 + \theta/N^\alpha)^{[N^{\alpha T}]}} \\ &= O \left( E \left( \sum_{n \leq N^{\alpha T}} \sum_x \Gamma_n^N(x) \right) \right) = o(N^\alpha). \end{aligned}$$

(2.41) then follows from the Doob's martingale inequality.)

We now prove Lemma 2.25.

**PROOF OF LEMMA 2.25.** We shall build approximating particle systems that satisfy the analogous sandwich relationship. Choose  $X_0^N$  and  $K^N$  such that (2.16)~(2.18) are satisfied. The super-solution  $\bar{X}_t^N(x)$  is a nearest-neighbor branching random walk with initial configuration  $X_0^N$  and such that at any site/time  $(x, t)$ , each particle at site  $x$  produces  $\text{Bin}(N, p_N^\theta(x, y))$  number of offspring at neighboring sites  $y$ . By Watanabe's Theorem,  $\bar{X}^N$  converges to the desired  $\bar{X}$ . As noted above, the Modified SIR  $X_t^N$  will approximate  $X_t$ . Define the stopping time

$$\tau^N = \min \left\{ t \geq 0 : \left( \max_x K^N(x) \right) + \left( \max_x R_t^N(x) \right) \geq \kappa N^{\alpha(2-d/2)} \right\}.$$

We may assume that  $\kappa > \sup_x K(x)$  (or the result is trivial). The sub-solution  $\underline{X}_t^N(x)$  is a nearest-neighbor branching random walk with initial configuration  $X_0^N$  and such that at any site/time  $(x, t)$ , each particle at  $x$  produces  $\text{Bin}([N - \kappa N^{\alpha(2-d/2)}], p_N^\theta(x, y))$  number of offspring at neighboring sites  $y$ . By Watanabe's Theorem  $\underline{X}^N$  converges weakly to the super-Brownian motion  $\underline{X}$ . It is clear that before time  $\tau^N$ , the three processes  $\bar{X}^N(x)$ ,  $X^N(x)$  and  $\underline{X}^N(x)$  can be built on a common probability space such that

$$\underline{X}_t^N \leq X_t^N \leq \bar{X}_t^N \quad \text{for all } t \leq \tau^N.$$

By Skorokhod's representation theorem and Proposition 2.9 we may assume  $\liminf_N \tau^N \geq \tau$  a.s. Here we use  $\max_x K^N(x)/N^{(\alpha(2-d/2))} \rightarrow \max_x K(x)$  and the fact that there is a greater than or equal to sign in the definition of  $\tau$ . By taking limits in the above, along a subsequence if necessary to get joint convergence, we complete the proof.  $\square$

2.5. *Scaling.* It will be necessary, in some of the arguments to follow, to rescale time and/or space. When a super-Brownian motion, or more generally a solution to the martingale problem (1.5) is rescaled, its diffusion rate may change, that is, the Laplacian in (1.5) may be multiplied by a constant  $\alpha$ . The resulting martingale problem is as follows:

$$(2.42) \quad X_t(\varphi) = X_0(\varphi) + \frac{\alpha}{2} \int_0^t X_s(\Delta\varphi) ds + \theta \int_0^t X_s(\varphi) ds - \int_0^t (X_s(K\varphi) + \beta X_s(L_s\varphi)) ds + \sqrt{\gamma} M_t(\varphi),$$

where  $\alpha, \beta, \gamma > 0$  and  $\theta \in \mathbb{R}$  are constants,  $K \in C_p(\mathbb{R}^d, \mathbb{R}^+)$ , and  $M_t(\varphi)$  is a continuous martingale with quadratic variation  $[M(\varphi)]_t = \int_0^t X_s(\varphi^2) ds$ . As usual,  $L_t$  is the local time density of the process  $X$ . We shall refer to this martingale problem as  $(MP)_{\mu, K}^{\theta, \beta, \gamma, \alpha}$  and continue to write  $(MP)_{\mu, K}^{\theta, \beta, \gamma}$  if  $\alpha = 1$ .

LEMMA 2.27. *Let  $X$  solve the martingale problem  $(MP)_{\mu, K}^{\theta, \beta, \gamma, \alpha}$ . For any constants  $a, b, c > 0$ , define a new measure-valued process  $U$  by*

$$(2.43) \quad U_t(\psi) = c \int_x \psi(bx) X_{at}(dx), \quad \text{for all bounded measurable } \psi \text{ on } \mathbb{R}^d.$$

*Then  $U_t$  solves the martingale problem  $(MP)_{\mu', K'}^{\theta', \beta', \gamma', \alpha'}$  with parameters*

$$\theta' = a\theta, \quad \beta' = \frac{a^2 b^d \beta}{c}, \quad \gamma' = ac\gamma, \quad \alpha' = ab^2\alpha, \quad K'(x) = aK(x/b),$$

*and initial measure defined by  $\int \psi(x) \mu'(dx) = c \int \psi(bx) \mu(dx)$ . The local time densities  $L = L^X$  and  $L^U$  are related by*

$$(2.44) \quad L_s^U(x) = \frac{c}{ab^d} L_{as} \left( \frac{x}{b} \right), \quad \text{for all } x, t.$$

PROOF. This is by routine calculations.  $\square$

REMARK 2.28. Based on the above result, one can show that by choosing  $a, b$  and  $c$  appropriately, the scaling as in (2.43) would transform the martingale problem  $(MP)_{\mu, 0}^{\theta, \beta, \gamma, \alpha}$  into  $(MP)_{\mu', 0}^{\theta', 1, 1, 1}$ ; in other words, the model is a one parameter model.

**3. Preliminaries on (supercritical) super-Brownian motions.** In this section we present some regularity results for super-Brownian motions and their local times. The results are only of interest, and only will be used, for  $d > 1$  and so we assume  $d = 2$  or  $3$  throughout this Section.

3.1. *Uniform regularity of super-Brownian motions.* In this and the following subsection, let  $Y_t = Y_t^\mu$  be a (driftless) super-Brownian motion with initial state  $Y_0 = \mu$ , and let  $P_\mu = P_{\mu,0}^{0,0,\dagger}$  be its law. Denote by  $B_r(y)$  the open Euclidean ball in  $\mathbb{R}^d$  centered at  $y$  of radius  $r$ , and for any measure  $\mu \in \mathcal{M}(\mathbb{R}^d)$ , define

$$(3.1) \quad D(\mu, r) = \sup\{\mu(B_r(y)) : y \in \mathbb{R}^d\}.$$

For any function  $\varphi$  and any measure  $\mu \in \mathcal{M}(\mathbb{R}^d)$ , set  $\mu\varphi = \mu * \varphi$ , where  $*$  denotes convolution. In particular, for any  $t \geq 0$ ,

$$(\mu p_t)(x) = \int p_t(x-y)\mu(dy), \quad \text{and} \quad (\mu q_t)(x) = \int_0^t \int p_s(x-y)\mu(dy) ds,$$

where  $p_t$  and  $q_t$  are the Gauss kernel and the integrated Gauss Kernel in (1.3) respectively. For  $r \in (0, 1]$  let

$$h(r) = \sqrt{r \log \frac{1}{r}}, \quad \text{and} \quad \varphi(r) = r^2 \left(1 + \log \frac{1}{r}\right)^2.$$

Finally, for  $T, r_0, C_i > 0$  introduce the event

$$G_T(r_0; C_1, C_2) = \{D(Y_t, r) \leq C_1 (D(\mu p_t, C_2 r) + \varphi(r)) \quad \text{for all } r \leq r_0 \text{ and } t \leq T\}.$$

The following lemma is an easy consequence of the proof of Theorem 4.7 in Barlow et al. (1991).

LEMMA 3.1. *If  $K \geq 1$  there are constants  $C_1, C_2 > 0$  (depending on  $K$ ), and for any  $T > 0$  there is an  $r_0(K, T) \in (0, 1]$  such that for all  $\lambda \geq 1$  and any  $\mu$  with  $|\mu| = \lambda$ ,*

$$(3.2) \quad P_\mu \left( G_T \left( r_0 e^{-\lambda}; C_1, C_2 \right) \right) \geq 1 - e^{-K\lambda}.$$

PROOF. This is a quantitative version of Theorem 4.7 of Barlow et al. (1991). The proof of that result shows for  $K \geq 1$  there are constants  $C_1, C_2, C_3 \geq 1$  such that for all  $\lambda \geq 1$ ,  $T > 0$ ,  $n \in \mathbb{N}$  and  $\mu$  with  $|\mu| = \lambda$ ,

$$(3.3) \quad P_\mu(G_T(h(2^{-n}); C_1, C_2)^c) \leq C_3(T+1)(\lambda+1)2^{-Kn} + C_3\lambda 2^{-Kn} \leq C_3(2T+3)\lambda 2^{-Kn}.$$

Here one has to chase constants a bit to check that the constant  $c_{2,2}$  in the proof of the above Theorem in Barlow et al. (1991) may be taken to be as large as you like at the cost of our  $C_2$  and their  $c_{4,2}$  being large. The latter can then be handled in the key bound in the proof of Theorem 4.7 in Barlow et al. (1991) by taking our  $C_1$  large enough. Now choose  $n_0 \geq 2$  in  $\mathbb{N}$  so that

$$(3.4) \quad C_3(2T+3)2^{-Kn_0} \leq e^{-K\lambda}\lambda^{-1} < C_3(2T+3)2^{-Kn_0+K}.$$

The above definition implies

$$h(2^{-n_0}) = 2^{-n_0/2}(n_0 \log 2)^{1/2} \geq \left( \frac{e^{-K\lambda}\lambda^{-1}}{2^K C_3(2T+3)} \right)^{1/(2K)} [2 \log 2]^{1/2} \geq r_0(K, T)e^{-\lambda},$$

where in the last inequality we used the simple fact that for all  $\lambda, K \geq 1$ ,  $\lambda^{-1/(2K)} \geq \lambda^{-1} \geq \exp(-\lambda/2)/2$ . Therefore (3.3) and (3.4) imply that

$$P_\mu(G_T(r_0 e^{-\lambda}; C_1, C_2)^c) \leq e^{-K\lambda}.$$

□

To formulate the next result we introduce the following

DEFINITION 3.2. For any positive constants  $A, \lambda$ , and  $r_0$ , with  $r_0 \leq 1$ , and any measure  $\mu \in \mathcal{M}(\mathbb{R}^d)$ , we say that  $\mu$  is  $(A, \lambda, r_0)$ -admissible if

$$D(\mu, r) \leq Ar^2 \left( \lambda r^{d-2} + \left(1 + \log \frac{1}{r}\right)^2 \right) \equiv \psi(r) \quad \text{for all } r \leq r_0 e^{-\lambda}.$$

COROLLARY 3.3. For any fixed  $K \geq 1$  and  $T > 0$ , there exist positive constants  $A = A(K, T)$  and  $r_0 = r_0(K, T) \leq 1$ , such that for all  $\lambda \geq 1$  and  $\mu$  with  $|\mu| = \lambda$ ,

$$(3.5) \quad P_\mu(Y_T \text{ is } (A, \lambda, r_0)\text{-admissible}) \geq 1 - e^{-K\lambda}.$$

PROOF. This follows from Lemma 3.1 by noticing that

$$D(\mu p_T, C_2 r) \leq C_4(T) \lambda r^d.$$

□

3.2. *Local time densities of super-Brownian motions.* Recall (equation (1.3)) that  $p_t(x)$  and  $q_t(x)$  are the Gauss kernel and the integrated Gauss kernel, respectively.

LEMMA 3.4. Suppose that  $\mu \in \mathcal{M}(\mathbb{R}^d)$  satisfies  $|\mu| = \lambda$  and is  $(A, \lambda, r_0)$ -admissible for some constants  $A$  and  $r_0$ . For any  $0 \leq \beta < 2 - d/2$  and any fixed  $T > 0$ , define

$$(3.6) \quad \Xi_1(T) := \max_x (\mu q_T)(x), \quad \text{and} \quad \Xi_2(T) := \max_x \int_0^T \int s^{-\beta/2} p_s(x-y) \mu(dy) ds.$$

Then there exists a constant  $A' = A'(A, r_0, T, \beta) > 0$  such that for both  $i = 1, 2$  and for all  $\lambda \geq 1$ ,

$$(3.7) \quad \Xi_i(T) \leq \kappa_d(\lambda) := \begin{cases} A' \lambda^2 & \text{when } d = 2 \\ A' \lambda^2 e^\lambda & \text{when } d = 3. \end{cases}$$

PROOF. We shall only prove the result for  $\Xi_1(T)$ ; the proof for  $\Xi_2(T)$  is similar. Let  $\tilde{r}_0 = r_0 e^{-\lambda}$ . We first deal with the integral for  $t \in [\tilde{r}_0^{8/3}, T]$ : to do so, for any fixed  $x \in \mathbb{R}^d$  we cover  $\mathbb{R}^d$  with balls  $B_i$  of radius  $\tilde{r}_0$  and with center of distance  $d_i = k \tilde{r}_0$  to  $x$  for some  $k \in \mathbb{Z}_{\geq 0}$ . Then for any  $t > 0$ ,

$$\begin{aligned} (2\pi)^{d/2} (\mu p_t)(x) &= \int t^{-d/2} \exp\left(-\frac{|x-y|^2}{2t}\right) \mu(dy) \\ &\leq \sum_i \int_{y \in B_i} t^{-d/2} \exp\left(-\frac{\min(0, d_i - \tilde{r}_0)^2}{2t}\right) \mu(dy). \end{aligned}$$

The balls can be chosen in such a way that for any  $k \geq 3$  there are at most  $C(k-2)^{d-1}$  balls with center of distance  $k\tilde{r}_0$  to  $x$ . It is then easy to see that there exist constants  $C_i$  such that

$$\begin{aligned}
& \int t^{-d/2} \exp\left(-\frac{|x-y|^2}{2t}\right) \mu(dy) \\
& \leq C_1 t^{-d/2} \psi(\tilde{r}_0) + C_2 \sum_{k=3}^{\infty} t^{-d/2} \exp\left(-\frac{(k-1)^2 \tilde{r}_0^2}{2t}\right) \cdot (k-2)^{d-1} \psi(\tilde{r}_0) \\
& \leq C_1 t^{-d/2} \psi(\tilde{r}_0) + C_2 \psi(\tilde{r}_0) \int_1^{\infty} t^{-d/2} a^{d-1} \exp\left(-\frac{a^2 \tilde{r}_0^2}{2t}\right) da \\
& \leq C_1 \psi(\tilde{r}_0) t^{-d/2} + C_3 \psi(\tilde{r}_0) \tilde{r}_0^{-d} \\
& \leq \begin{cases} C_1 \psi(\tilde{r}_0) t^{-1} + C_3 (\lambda + (1 + \log(1/\tilde{r}_0))^2) & \text{when } d = 2 \\ C_1 \psi(\tilde{r}_0) t^{-3/2} + C_3 (\lambda + (1 + \log(1/\tilde{r}_0))^2 / \tilde{r}_0) & \text{when } d = 3. \end{cases}
\end{aligned}$$

Therefore

$$\begin{aligned}
& \int_{\tilde{r}_0^{8/3}}^T \int t^{-d/2} \exp\left(-\frac{|x-y|^2}{2t}\right) \mu(dy) dt \\
& \leq \begin{cases} C_4 \psi(\tilde{r}_0) (\log T + \log(1/\tilde{r}_0)) + C_3 T (\lambda + (1 + \log(1/\tilde{r}_0))^2) & \text{when } d = 2 \\ C_4 \psi(\tilde{r}_0) \tilde{r}_0^{(-4/3)} + C_3 T (\lambda + (1 + \log(1/\tilde{r}_0))^2 / \tilde{r}_0) & \text{when } d = 3. \end{cases}
\end{aligned}$$

which can be bounded by  $\kappa_d(\lambda)$  for all  $\lambda \geq 1$  for an appropriate choice of  $A'$ . Now we deal with the integral for  $t \in [0, \tilde{r}_0^{8/3}]$ :

$$\begin{aligned}
& \int_0^{\tilde{r}_0^{8/3}} \int t^{-d/2} \exp\left(-\frac{|x-y|^2}{2t}\right) \mu(dy) dt \\
& = \int_0^{\tilde{r}_0^{8/3}} \left( \int_{|x-y| \leq t^{3/8}} + \int_{|x-y| > t^{3/8}} \right) t^{-d/2} \exp\left(-\frac{|x-y|^2}{2t}\right) \mu(dy) dt \\
& \leq \int_0^{\tilde{r}_0^{8/3}} t^{-d/2} \psi(t^{3/8}) dt + \lambda \int_0^{\tilde{r}_0^{8/3}} t^{-d/2} \exp\left(-\frac{1}{2t^{1/4}}\right) dt \\
& \leq C_5 + C_6 \lambda.
\end{aligned}$$

□

The following lemma is implicit in Sugitani (1989).

LEMMA 3.5. *Suppose that  $\mu \in \mathcal{M}(\mathbb{R}^d)$  and for all  $t > 0$ ,  $\Xi_1(t) < \infty$ , and for some  $0 < \beta < 2 - d/2$  and all  $t > 0$ ,  $\Xi_2(t) < \infty$ , where  $\Xi_1$  and  $\Xi_2$  are defined in (3.6). Define*

$$Z_t(x) = L_t(x) - (\mu q_t)(x).$$

*Then for any  $T > 0$ , there exist constants  $\eta_0 = \eta_0(\beta, T) > 0$ ,  $C_i = C_i(\beta, T) > 0$  such that for all  $0 \leq \eta < \eta_0$  and  $t \leq T$ ,*

$$(3.8) \quad E_\mu \exp\left(\frac{\eta(Z_t(a) - Z_t(b))}{|a-b|^\beta}\right) \leq \exp(C_1 \Xi_2(2t) \cdot \eta), \quad \text{for all } 0 < |a-b| \leq 2;$$

and

$$(3.9) \quad E_\mu \exp(\eta Z_t(a)) \leq \exp(C_2 \Xi_1(2t) \cdot \eta), \quad \text{for all } a \in \mathbb{R}^d.$$

PROOF. The second claim (3.9) follows from Lemma 3.4 in Sugitani (1989).

To prove (3.8), following (3.34) in Sugitani (1989), for a random variable  $X$  we say that

$$E \exp(\eta X) = \exp\left(\sum_{n=1}^{\infty} c_n \eta^n\right)$$

holds formally if for all  $k \geq 1$ ,  $E|X|^k < \infty$  and

$$EX^k = \left( \frac{d^k \left( \exp\left(\sum_{n=1}^k c_n \eta^n\right)\right)}{d\eta^k} \right) \Big|_{\eta=0}.$$

By (3.38), (3.45) and (3.48) in Sugitani (1989), we have formally

$$(3.10) \quad E_\mu \exp(\eta(Z_t(a) - Z_t(b))) = \exp\left(2 \sum_{n=2}^{\infty} \left(\frac{\eta}{2}\right)^n \langle \mu, \nu_n(t, \cdot) \rangle\right),$$

where for  $n \geq 2$ , and  $x \in \mathbb{R}^d$ ,

$$(3.11) \quad |\nu_n(t, x)| \leq b_n \cdot |a - b|^{n\beta} t^{2-(d+\beta)/2} \int_0^{2t} s^{-\beta/2} (p_s(a-x) + p_s(b-x)) ds,$$

and  $\{b_n\}$  are defined inductively as follows:

$$b_1 = C_4 > 0; \quad b_n = C_5 \sum_{k=1}^{n-1} b_k b_{n-k}.$$

Using the proof of Lemma 3.4 in Sugitani (1989), if we let  $f(\eta) = \sum_{n=1}^{\infty} b_n \eta^n$ , then for some  $\delta > 0$ ,

$$(3.12) \quad f(\eta) - C_4 \eta = C_5 f(\eta)^2; \quad f(\eta) = \frac{1 - \sqrt{1 - 4C_4 C_5 \eta}}{2C_5} \leq C\eta \quad \text{for } 0 \leq \eta \leq \delta.$$

This shows that  $\sum_n b_n \eta^n$  has a positive radius of convergence, and the formal equation (3.10) is indeed an equation when  $\eta$  is sufficiently close to 0 because the Taylor series for the analytic function on the right-hand side is given by the left-hand side. Relation (3.8) then follows easily from the upper bounds (3.11) and (3.12). □

**COROLLARY 3.6.** *Under the assumptions of the previous lemma, for any fixed  $T > 0$ , there exist constants  $\eta_0 = \eta_0(\beta, T) > 0$ ,  $C_i = C_i(\beta, T) > 0$  such that for all  $0 < \eta < \eta_0$ ,*

$$(3.13) \quad E_\mu \exp\left(\frac{\eta |L_T(a) - L_T(b)|}{|a - b|^\beta}\right) \leq 2 \exp(C_1 \Xi_2(2T) \eta), \quad \text{for all } 0 < |a - b| \leq 2;$$

and

$$(3.14) \quad E_\mu \exp(\eta L_T(a)) \leq \exp(C_2 \Xi_1(2T) \eta), \quad \text{for all } a \in \mathbb{R}^d.$$

PROOF. Relation (3.14) follows easily from (3.9). As for (3.13), by (3.8) and the elementary inequality  $e^{|x|} \leq e^x + e^{-x}$ ,

$$E_\mu \exp \left( \frac{\eta |L_T(a) - L_T(b)|}{|a - b|^\beta} \right) \leq 2 \exp(C_1 \Xi_2(2T)\eta) \cdot \exp \left( \frac{\eta |(\mu q_T)(a) - (\mu q_T)(b)|}{|a - b|^\beta} \right)$$

By (3.44) in Sugitani (1989) we have for all  $x, y$  and  $t > 0$ ,

$$|p_t(x) - p_t(y)| \leq c(\beta) t^{-\beta/2} |x - y|^\beta (p_{2t}(x) + p_{2t}(y)),$$

and so

$$\begin{aligned} |(\mu q_T)(a) - (\mu q_T)(b)| &\leq C |a - b|^\beta \int_0^T \int_x t^{-\beta/2} (p_{2t}(a - x) + p_{2t}(b - x)) \mu(dx) ds \\ &\leq C |a - b|^\beta \Xi_2(2T). \end{aligned}$$

□

LEMMA 3.7. *Suppose that  $\Upsilon(x)$  is an almost surely continuous random field on  $\mathbb{R}^d$  such that for some  $\eta > 0$  and  $\beta > 0$ ,*

$$(3.15) \quad \begin{cases} E \exp \left( \frac{\eta |\Upsilon(a) - \Upsilon(b)|}{|a - b|^\beta} \right) \leq C_1, & \text{for all } 0 < |a - b| \leq \sqrt{d}; \quad \text{and} \\ E \exp(\eta \Upsilon(a)) \leq C_2, & \text{for all } a \in \mathbb{R}^d. \end{cases}$$

Let  $M_L = \max_{a \in Q_L(0)} \Upsilon(a)$  ( $Q_L(0)$  is the cube of side length  $L$  centered at 0). Then for all  $L \in \mathbb{N}$  and  $m \geq 0$ ,

$$(3.16) \quad P(M_L \geq m) \leq (C_1 e^{2d/\beta} + C_2) L^d \exp \left( -\frac{\eta m}{1 + \gamma} \right),$$

where  $\gamma = 8d^{\beta/2}$ . In particular, for any  $K > 0$ , there exists  $C > 0$ , depending only on  $K, C_1, C_2$  and  $\beta$ , such that for all  $L \geq 1$  and  $\lambda \geq 1$ ,

$$(3.17) \quad P \left( M_L \geq C \frac{\lambda + \log L}{\eta} \right) \leq e^{-K\lambda}.$$

PROOF. The inequality (3.17) follows by plugging  $C \frac{\lambda + \log L}{\eta}$  as  $m$  into (3.16) and noticing that when  $C$  is large enough, the factor  $\exp \left( -\frac{C\lambda}{2(1+\gamma)} - \frac{C \log L}{1+\gamma} \right)$  would be smaller than  $[(C_1 e^{2d/\beta} + C_2) L^d]^{-1}$  for all  $\lambda, L \geq 1$ . Furthermore, it suffices to prove (3.16) for  $L = 1$  as the results then follow trivially by dividing  $Q_L(0)$  into unit cubes. We apply Lemma 1 of Garsia (1972) with  $p(u) = u^\beta$ ,  $\Psi(u) = \exp \left( \frac{\eta |u|}{d^{\beta/2}} \right)$ ,  $Q_1 = Q_1(0)$  and

$$B = \int_{Q_1} \int_{Q_1} \exp \left( \frac{\eta |\Upsilon(x) - \Upsilon(y)|}{|x - y|^\beta} \right) dx dy.$$

It is easy to check that this  $B$  satisfies the hypothesis of Lemma 1 in the above reference. That result, or more precisely (10) in the proof, implies

$$\begin{aligned} \overline{M_1} &\leq \Upsilon(0) + 8 \int_0^1 \Psi^{-1} \left( \frac{B}{u^{2d}} \right) d(u^\beta) \\ &\leq \Upsilon(0) + \frac{\gamma}{\eta} \left[ \log(B) + 2d \int_0^1 \log(1/u) d(u^\beta) \right] \\ &= \Upsilon(0) + \frac{\gamma}{\eta} [\log(B) + (2d/\beta)]. \end{aligned}$$

Therefore for  $x \geq 0$ ,

$$\begin{aligned} P(M_1 \geq (1 + \gamma)x) &\leq C_2 \exp(-\eta x) + P(\log(B) \geq \eta x - (2d/\beta)) \\ &\leq C_2 \exp(-\eta x) + E(B) \exp((2d/\beta) - \eta x) \\ &= (C_2 + C_1 e^{2d/\beta}) e^{-\eta x}, \end{aligned}$$

which is (3.16) for  $L = 1$ , and where (3.15) is used to see that  $E(B) \leq C_1$ .  $\square$

Combining Lemma 3.4, Corollary 3.6 and Lemma 3.7 we obtain the following for the local time  $L$  of the super-Brownian motion  $Y_t$ .

**PROPOSITION 3.8.** *For any  $T > 0$ ,  $M > 0$ ,  $K > 0$ ,  $A > 0$  and  $r_0 > 0$  there exists a constant  $A''$ , depending only on  $(T, M, K, A, r_0)$ , so that for all  $\lambda \geq 1$  and all  $(A, \lambda, r_0)$ -admissible  $\mu \in \mathcal{M}(\mathbb{R}^d)$  satisfying  $|\mu| = \lambda$ , the local time,  $L_T(x)$ , of the super-Brownian motion  $Y_t^\mu$  satisfies*

$$P_\mu \left( \max_{|x| \leq M e^\lambda} L_T(x) \geq A'' \lambda \kappa_d(\lambda) \right) \leq e^{-K\lambda},$$

where  $\kappa_d(\lambda)$  is defined in (3.7).

**PROOF.** By Lemma 3.4 and Corollary 3.6, for any fixed  $0 \leq \eta < \eta_0$ , if we let  $\eta(\lambda) = \eta/\kappa_d(\lambda)$ , then for some fixed  $C_1$  and  $C_2$ , for all  $\lambda \geq 1$ , the assumptions (3.15) hold for the random field  $L_T(x)$  and  $0 < \beta < 2 - d/2$  by replacing  $\eta$  with  $\eta(\lambda)$ . The conclusion then follows from (3.17).  $\square$

**3.3. Local time densities of supercritical super-Brownian motions.** In this and the following subsection  $Y_t = Y_t^\mu$  is a super-Brownian motion with drift one starting at an initial state  $\mu$ , let  $P_\mu^1 = P_{\mu,0}^{1,0,1}$  be its law. Further denote by  $P_\mu^0 = P_{\mu,0}^{0,0,1}$  the law of a (driftless) super-Brownian motion starting at  $\mu$ . By Lemma 2.1 we have

$$\frac{dP_\mu^1}{dP_\mu^0} \Big|_{\mathcal{F}_t^Y} := \Phi_t = \exp \left( M_t^0(1) - \frac{1}{2} \int_0^t |Y_s| ds \right),$$

where  $M^i$  denote the martingale measure under  $P_\mu^i$ ,  $i = 0, 1$ , and therefore

$$(3.18) \quad M_t^0(1) = M_t^1(1) + \int_0^t |Y_s| ds.$$

LEMMA 3.9. For any  $K \geq 1$ ,  $T > 0$ ,  $\lambda \geq 1$  and  $\mu \in \mathcal{M}(\mathbb{R}^d)$  with  $|\mu| = \lambda$ ,

$$(3.19) \quad E_\mu^0 \left( \Phi_T \cdot \mathbf{1}_{(\Phi_T \geq e^{K\lambda})} \right) \leq \frac{5e^T}{K}.$$

PROOF. Using (3.18) we see the above expectation equals

$$\begin{aligned} E_\mu^0 \left( \Phi_T \cdot \mathbf{1}_{\left( M_T^0 - \frac{1}{2} \int_0^t |Y_s| ds \geq K\lambda \right)} \right) &= P_\mu^1 \left( M_T^1(1) + \frac{1}{2} \int_0^t |Y_s| ds \geq K\lambda \right) \\ &\leq \frac{4E_\mu^1 \left( (M_T^1(1))^2 \right)}{K^2 \lambda^2} + \frac{E_\mu^1 \left( \int_0^T |Y_s| ds \right)}{K\lambda} \\ &= \frac{4\lambda(e^T - 1)}{K^2 \lambda^2} + \frac{\lambda(e^T - 1)}{K\lambda} \leq \frac{5e^T}{K}. \end{aligned}$$

□

PROPOSITION 3.10. For any fixed  $T > 0$  and any  $\varepsilon > 0$ , there exist constants  $A = A(T, \varepsilon) > 0$  and  $r_0 = r_0(T, \varepsilon) \in (0, 1]$  such that for all  $\lambda \geq 1$  and all  $\mu$  with  $|\mu| = \lambda$ ,

$$(3.20) \quad P_\mu^1 (Y_T \text{ is } (A, \lambda, r_0)\text{-admissible}) \geq 1 - \varepsilon.$$

PROOF. Let  $G_{T,\lambda}$  denote the event in (3.20). For  $T, \varepsilon$  as above choose  $K = K(\varepsilon, T) \geq 2$  so that

$$\frac{10e^T}{K} + e^{-K/2} < \varepsilon,$$

and then choose  $A$  and  $r_0$  as in Corollary 3.3 for this choice of  $K$  and  $T$ , so that they depend ultimately on  $T$  and  $\varepsilon$ . Then the previous Lemma and Corollary 3.3 imply that

$$\begin{aligned} P_\mu^1(G_{T,\lambda}^c) &= E_\mu^0(\Phi_T \cdot \mathbf{1}_{G_{T,\lambda}^c}) \\ &\leq E_\mu^0(\Phi_T \cdot \mathbf{1}_{(\Phi_T \geq e^{\lambda K/2})}) + e^{\lambda K/2} P_\mu^0(G_{T,\lambda}^c) \\ &\leq 10e^T/K + e^{-\lambda K/2} < \varepsilon, \end{aligned}$$

where the choice of  $K$  is used in the last inequality. □

The same reasoning, but now using Proposition 3.8 in place of Corollary 3.3, gives the following:

PROPOSITION 3.11. For any positive constants  $T, M$  and  $\varepsilon, A > 0$  and  $r_0 > 0$  there exists a constant  $A''$ , depending only on  $(T, M, \varepsilon, A, r_0)$ , so that for all  $\lambda \geq 1$  and all  $(A, \lambda, r_0)$ -admissible  $\mu \in \mathcal{M}(\mathbb{R}^d)$  satisfying  $|\mu| = \lambda$ , the local time,  $L_T(x)$ , of  $Y_t^\mu$  satisfies

$$P_\mu^1 \left( \max_{|x| \leq M e^\lambda} L_T(x) \leq A'' \lambda \kappa_d(\lambda) \right) \geq 1 - \varepsilon.$$

3.4. *Propagation of supercritical super-Brownian motions.* We continue to let  $Y_t = Y_t^\mu$  be a super-Brownian motion with drift one starting at  $\mu$ , and let  $P_\mu^1$  denote its law. Recall that for  $x \in \mathbb{Z}^d$ ,  $Q_r(x)$  denotes the cube of side length  $r$  centered at  $x$ , and  $Q(x) := Q_1(x)$ .

LEMMA 3.12. *For any  $T \geq 1$  and  $\varepsilon > 0$ , there exists a constant  $M = M(T, \varepsilon) > 0$  such that for any  $\lambda \geq e$  and any  $\mu$  satisfying  $|\mu| = \lambda$  and  $\text{Supp}(\mu) \subseteq Q(0)$ , we have*

$$P_\mu^1 \left( \text{Supp}(Y[0, T]) \subseteq Q_{M\sqrt{\log \lambda}}(0) \right) \geq 1 - \varepsilon.$$

PROOF. This is a direct consequence of Theorem A in Pinsky (1995).  $\square$

Write  $\mathcal{N}(0) = \{x \in \mathbb{Z}^d : \|x\|_1 = 1\}$  for the nearest neighbors of the origin in  $\mathbb{Z}^d$ . Fix a  $T$  sufficiently large such that

$$(3.21) \quad \min_{x \in \mathcal{N}(0)} \min_{y \in Q(0)} e^T (\mathbf{1}_{Q(x)} * p_T)(y) \geq 2,$$

where  $p_t(x)$  is the Gauss kernel in (1.3), and  $*$  denotes convolution.

LEMMA 3.13. *For any  $\varepsilon > 0$  and  $T$  as above, there exists  $\lambda_0 = \lambda_0(T, \varepsilon) > 0$  such that for any  $\mu$  satisfying  $\text{Supp}(\mu) \subseteq Q(0)$  and  $|\mu| = \lambda \geq \lambda_0$ ,*

$$P_\mu^1 (Y_T(Q(x)) \geq \lambda \text{ for all } x \in \mathcal{N}(0)) \geq 1 - \varepsilon.$$

PROOF. By a well-known moment formula (see, e.g., Exercise II.5.2 in Perkins (2002)), together with the assumption that  $\text{Supp}(\mu) \subseteq Q(0)$  and (3.21), for any  $x \in \mathcal{N}(0)$ ,

$$EY_T(Q(x)) = e^T \langle \mu, \mathbf{1}_{Q(x)} * p_T \rangle \geq 2|\mu| = 2\lambda,$$

and

$$\text{Var}(Y_T(Q(x))) \leq e^{2T} \langle \mu, \int_0^T (\mathbf{1}_{Q(x)} * p_{(T-s)})^2 * p_s \, ds \rangle.$$

Consequently, by the Chebyshev inequality,

$$\begin{aligned} P(Y_T(Q(x)) \leq \lambda) &\leq P \left( |Y_T(Q(x)) - EY_T(Q(x))| \geq \frac{1}{2} EY_T(Q(x)) \right) \\ &\leq \frac{4e^{2T} \langle \mu, \int_0^T (\mathbf{1}_{Q(x)} * p_{(T-s)})^2 * p_s \, ds \rangle}{(2\lambda)^2} \\ &\leq C_T \lambda^{-1}. \end{aligned}$$

The conclusion follows.  $\square$

#### 4. A weak form of local extinction and its consequences.

4.1. *A weak form of local extinction.* Let  $V_d = \pi^{d/2}/\Gamma(1 + d/2)$  be the volume of a unit  $d$ -dimensional ball.

PROPOSITION 4.1. *There exists  $\kappa < \infty$  such that for any  $\theta \in \mathbb{R}, \gamma > 0$ , and  $K \in C_p(\mathbb{R}^d, \mathbb{R}_+)$ , if  $X$  solves  $(MP)_{\mu, K}^{\theta, 1, \gamma}$  and  $\mu$  satisfies Assumption 1.1, then for any  $N \geq 1$ ,*

$$(4.1) \quad E\langle L_\infty, \mathbf{1}_{B_N(0)} \rangle \leq \frac{2|\mu|}{\kappa + 2\theta^+} + V_d(\kappa + 2\theta^+)(N + 1)^d,$$

and

$$(4.2) \quad E\langle L_\infty^2, \mathbf{1}_{B_N(0)} \rangle \leq 4|\mu| + V_d(\kappa + 2\theta^+)^2(N + 1)^d.$$

PROOF. First, observe that there exists  $\kappa > 0$  such that for any  $N \geq 1$  there exists a function  $\varphi = \varphi_N \in C^2$  such that

$$(4.3) \quad |\Delta\varphi| \leq \kappa\sqrt{\varphi} \quad \text{and} \quad \mathbf{1}_{B_N(0)} \leq \varphi \leq \mathbf{1}_{B_{N+1}(0)}.$$

For example, set  $\varphi(x) = \psi(|x|)$  where  $\psi = \psi_N : \mathbb{R} \rightarrow [0, 1]$  is a smooth, even function bounded above and below by the indicators of  $[-N - 1, N + 1]$  and  $[-N, N]$ , monotone on  $[-N - 1, -N]$  (and therefore also on  $[N, N + 1]$ ), and such that (for instance)  $\psi(x) = (x + N + 1)^4$  for  $x \in [-N - 1, -N - 1/2]$ . Because  $\varphi \in C_c^2(\mathbb{R}^d)$ , the martingale identity (1.5) applies (with  $\beta = 1$ ), so after taking expectations we obtain

$$E\langle X_t, \varphi \rangle = \langle \mu, \varphi \rangle + \frac{1}{2}E\langle L_t, \Delta\varphi \rangle + \theta E\langle L_t, \varphi \rangle - E\langle L_t, K\varphi \rangle - E \int_0^t \langle X_s, L_s\varphi \rangle ds.$$

A routine integration by parts shows that

$$\int_0^t \langle X_s, L_s\varphi \rangle ds = \frac{1}{2}\langle L_t^2, \varphi \rangle.$$

Since  $|\Delta\varphi| \leq \kappa\sqrt{\varphi}$  and  $\langle X_t, \varphi \rangle \geq 0$ , it follows that

$$(4.4) \quad \begin{aligned} -2\langle \mu, \varphi \rangle &\leq \kappa E\langle L_t, \sqrt{\varphi} \rangle + 2\theta^+ E\langle L_t, \varphi \rangle - E\langle L_t^2, \varphi \rangle \\ &\leq (\kappa + 2\theta^+) E\langle L_t, \sqrt{\varphi} \rangle - E\langle L_t^2, \varphi \rangle \\ &\leq (\kappa + 2\theta^+) E\langle L_t, \sqrt{\varphi} \rangle - (V_d^{-1}(N + 1)^{-d}) (E\langle L_t, \sqrt{\varphi} \rangle)^2, \end{aligned}$$

the last by Cauchy-Schwartz and the fact that  $\varphi$  has support contained in  $B_{N+1}(0)$ . This clearly gives an upper bound on  $E\langle L_t, \sqrt{\varphi} \rangle$  that is independent of  $t$ . In fact, (4.4) implies that

$$\begin{aligned} \left( E\langle L_t, \sqrt{\varphi} \rangle - \frac{1}{2}V_d(N + 1)^d(\kappa + 2\theta^+) \right)^2 &\leq \frac{1}{4}(V_d(N + 1)^d)^2(\kappa + 2\theta^+)^2 + 2\langle \mu, \varphi \rangle V_d(N + 1)^d \\ &\leq \left( \frac{1}{2}V_d(N + 1)^d(\kappa + 2\theta^+) + \frac{2\langle \mu, \varphi \rangle}{(\kappa + 2\theta^+)} \right)^2, \end{aligned}$$

and hence

$$\begin{aligned} E\langle L_t, \sqrt{\varphi} \rangle &\leq \frac{1}{2}V_d(N+1)^d(\kappa+2\theta^+) + \left( \frac{1}{2}V_d(N+1)^d(\kappa+2\theta^+) + \frac{2\langle \mu, \varphi \rangle}{(\kappa+2\theta^+)} \right) \\ &= V_d(\kappa+2\theta^+)(N+1)^d + \frac{2\mu(\varphi)}{\kappa+2\theta^+}. \end{aligned}$$

Letting  $t \rightarrow \infty$  yields

$$(4.5) \quad E\langle L_\infty, \sqrt{\varphi} \rangle \leq V_d(\kappa+2\theta^+)(N+1)^d + \frac{2\mu(\varphi)}{\kappa+2\theta^+}.$$

Relation (4.1) follows, since  $\sqrt{\varphi}$  bounds the indicator function of  $B_N(0)$ . Finally, by the second inequality in (4.4),

$$E\langle L_\infty^2, \varphi \rangle \leq 2|\mu| + (\kappa+2\theta^+)E\langle L_\infty, \sqrt{\varphi} \rangle.$$

Relation (4.2) follows from (4.5). □

REMARK 4.2. The above Proposition easily shows that each of the terms on the right-hand side of (1.2) converges a.s. as  $t \rightarrow \infty$ . Therefore  $X_t(\varphi)$  converges a.s. as  $t \rightarrow \infty$  and clearly the limit must be 0 by the above. This shows that  $X_t(K)$  approaches 0 as  $t \rightarrow \infty$  for all compact sets  $K$  a.s. Our Theorem 1.3 asserts a much stronger result, namely that  $X_t(K) = 0$  for large enough  $t$  a.s.

4.2. *Universality of the critical values  $\theta_c$ .* For any  $\mu$  satisfying Assumption 2.10, if  $X$  solves (1.2) with  $\theta \leq 0$ , then  $P(X \text{ survives}) = 0$  because  $X$  is dominated by a critical super-Brownian motion (by Proposition 2.5), which goes extinct almost surely (see, e.g., Eqn (5.7) in Feller (1951) or (II.5.12) in Perkins (2002)). Lemma 2.16 and Remark 2.17 therefore imply that for any such  $\mu$  and any function  $K \in C_p(\mathbb{R}^d; \mathbb{R}_+)$  there is a critical value  $\theta_c(\mu, K) \in [0, \infty]$  so that a spatial epidemic  $X$  with suppression rate  $K$  and transmission parameter  $\theta$  (see (1.5)) survives with positive probability if  $\theta > \theta_c(\mu, K)$  and with zero probability if  $\theta < \theta_c(\mu, K)$ .

PROPOSITION 4.3. *The critical value  $\theta_c(\mu, K)$  depends only on the dimension  $d$  and not on the choice of  $0 \neq \mu$  satisfying Assumption 1.1 or  $K \in C_p(\mathbb{R}^d; \mathbb{R}_+)$ .*

PROOF. In this argument  $\theta$  will be fixed and  $\gamma = 1$ , so we suppress the dependence of the laws  $P_{\mu, K}^{\theta, 1, \gamma}$  on  $\theta$  and  $\gamma$ . By Theorem 2.2, for any measure  $\mu \in \mathcal{M}_c(\mathbb{R}^d)$  satisfying Assumption 1.1 and any two suppression rate functions  $K, K' \in C_p(\mathbb{R}^d; \mathbb{R}_+)$ , the laws  $P_{\mu, K}$  and  $P_{\mu, K'}$  are mutually absolutely continuous on  $\mathcal{F}_t^X$ , with Radon-Nikodym derivative (2.8). Since  $K$  and  $K'$  both have compact support, inequality (4.2) of Proposition 4.1 implies that the integrals in the likelihood ratio converge, so that

$$\lim_{t \rightarrow \infty} \left( \frac{dP_{\mu, K'}}{dP_{\mu, K}} \right)_{\mathcal{F}_t} := Y$$

exists and is positive  $P_{\mu,K}$ -almost surely. Hence, by Fatou's lemma,

$$P_{\mu,K'}(X \text{ survives}) = \lim_{t \rightarrow \infty} P_{\mu,K'}(|X_t| > 0) \geq E_{\mu,K} \left( Y \mathbf{1}_{\{X \text{ survives}\}} \right).$$

It follows that if  $X$  survives with positive  $P_{\mu,K}$  probability then it also survives with  $P_{\mu,K'}$  probability. Reversing the roles of  $K$  and  $K'$  shows that the reverse is also true. Therefore, the critical value  $\theta(\mu, K)$  does not depend on  $K$ .

To complete the proof, it suffices, in view of the preceding paragraph, to prove that if  $X$  survives with positive probability under  $P_{\mu,0}$  then it survives with positive probability under  $P_{\nu,0}$  for  $\nu \neq 0$ . By the Markov property (Theorem 2.2 (d)),

$$P_{\mu,0}(X \text{ survives}) = E_{\mu,0}(P_{X_1,L_1}(X \text{ survives})),$$

and similarly for  $P_{\nu,0}$ . By the argument of the preceding paragraph,

$$P_{X_1,L_1}(X \text{ survives}) > 0 \iff P_{X_1,0}(X \text{ survives}) > 0,$$

so for both  $\omega = \mu$  and  $\omega = \nu$ ,

$$P_{\omega,0}(X \text{ survives}) > 0 \iff E_{\omega,0}(P_{X_1,0}(X \text{ survives})) > 0.$$

But the laws of  $X_1$  under  $P_{\mu,0}$  and  $P_{\nu,0}$  are mutually absolutely continuous. (This can be seen as follows. First, by the absolute continuity results in Evans and Perkins (1991) (see, for example Theorem III.2.2 in Perkins (2002)), if  $P_\mu$  and  $P_\nu$  are the laws of super-Brownian motions with initial conditions  $\mu$  and  $\nu$ , then the distributions of  $X_1$  under  $P_\mu$  and  $P_\nu$  are mutually absolutely continuous. Second, by Theorem 2.2 (a), for any initial measure  $\omega$  the measures  $P_\omega$  and  $P_{\omega,0}$  are mutually absolutely continuous.) Therefore,

$$P_{\mu,0}(X \text{ survives}) > 0 \iff P_{\nu,0}(X \text{ survives}) > 0.$$

□

Note that the above arguments do not require  $\mu$  to satisfy the stronger Assumption 2.10, instead just the original Assumption 1.1.

#### 4.3. Extinction in dimension one.

**PROPOSITION 4.4.** *If  $d = 1$  then for every  $\theta \in \mathbb{R}$  and every initial measure  $\mu$  that satisfies Assumption 1.1, the solution  $X_t$  of the martingale problem (1.2) dies out almost surely.*

**PROOF.** First, by Proposition 2.5, on some probability space there is a version of the process  $X$  and a super-Brownian motion  $\bar{X}$  with drift  $\theta$  such that  $X_0 = \bar{X}_0 = \mu$  and  $X_t \leq \bar{X}_t$  for all  $t \geq 0$ .

By a result of Pinsky (1995), there is a positive constant  $C = C_\theta < \infty$  such that almost surely the support of the random measure  $\bar{X}$  is eventually contained in the interval  $[-Ct, Ct]$ . Since  $\bar{X}$  dominates  $X$ , the same is true for  $X$ . Now by Lemma 2.15, on the event that  $X$

survives, the total mass of the measure  $X_t$  must diverge. Because this mass is (eventually) contained in  $[-Ct, Ct]$ , it follows from L'Hospital's rule that on the event of survival, the occupation density process  $L_t(x)$  must satisfy

$$\frac{1}{t} \int_{-Ct}^{Ct} L_t(x) dx = \frac{1}{t} \int_0^t |X_u| du \longrightarrow \infty.$$

Hence, if there is positive probability of survival then

$$\frac{1}{t} E \int_{-Ct}^{Ct} L_t(x) dx \longrightarrow \infty.$$

But this contradicts (4.1) in Proposition 4.1.  $\square$

**5. Proof of survival when  $d = 2$  or  $3$ .** In this section we prove that in dimensions 2 and 3, for all sufficiently large values of the transmission rate  $\theta$ , spatial epidemics – that is, solutions of the martingale problem (1.2) – survive with positive probability. By Proposition 4.3, the critical value  $\theta_c$  for survival in dimensions  $d = 2, 3$  does not depend on the initial mass distribution  $\mu$ , hence it suffices to prove that for *some* finite measure  $\mu$  there is positive probability of survival. The proof will make use of an auxiliary 3-dependent *site percolation process*: this will be constructed in such a way that if percolation occurs with positive probability then the epidemic must survive with positive probability. We will show that by taking  $\theta$  sufficiently large, we can make the density of the site percolation arbitrarily close to 1. Since percolation occurs with positive probability in a site percolation process when the density is near 1 (see, e.g., Theorem 4.1 of Durrett (1995)), it will follow that for large values of  $\theta$  the epidemic process will survive with positive probability. We refer the reader to Chapter 4 of Durrett (1995) for terminology and a general framework for such comparison arguments.

5.1. *Scaled process.* We assume  $d = 2$  or  $3$  throughout this section. Let  $X_t$  be a spatial epidemic process with transmission rate  $\theta$  and initial mass distribution  $\mu$ , that is, a solution to the martingale problem (1.2). It will be convenient to work with a rescaled version of the spatial epidemic defined as follows: for any  $\theta > 0$ ,

$$U_t(\psi) = \theta X_{t/\theta}(\psi(\sqrt{\theta} \cdot)), \quad \text{for all } \psi \in C_c^2(\mathbb{R}^d).$$

The effect of this rescaling is described by Lemma 2.27: in particular,  $U_t$  satisfies the martingale problem  $(MP)_{\tilde{\mu}, 0}^{1, \beta, 1}$  with  $\beta = \theta^{(d-6)/2}$  and  $\tilde{\mu}$  defined by  $\int \psi(x) d\tilde{\mu}(x) = \theta \int \psi(\sqrt{\theta}x) d\mu(x)$ . For notational ease, we will use the notation

$$(5.1) \quad \beta = \beta(\theta) = \theta^{(d-6)/2}$$

in this section, and we will drop the tilde on the initial measure  $\mu$ . We will show that when  $\theta$  is sufficiently large, for a suitable initial condition  $\mu$ , the process  $U$  survives with positive probability.

5.2. *Sandwich Lemma.* By Lemma 2.25, a spatial epidemic process can be bounded below and above by super-Brownian motions with different drift terms up to the time that its local time density exceeds some threshold. We now explain how the result of Lemma 2.25 translates to the rescaled processes.

For any  $\mu \in \mathcal{M}_c(\mathbb{R}^d)$  satisfying Assumption 2.10, and any function  $K \in C_p(\mathbb{R}^d, \mathbb{R}_+)$ , let  $U$  be a solution of the martingale problem  $(MP)_{\mu, K}^{1, \beta, 1}$ , i.e., the spatial epidemic with transmission rate 1, branching rate 1, inhibition parameter  $\beta$ , local suppression rate  $K$ , and initial mass distribution  $\mu$ . In addition, for any fixed constant  $\kappa > 0$ , let  $\bar{U}_t$  and  $\underline{U}_t$  be super-Brownian motions with drift 1 and drift  $1 - \beta \cdot \kappa$  respectively. Denote by  $\bar{M}_t$  and  $\underline{M}_t$  the orthogonal martingale measures associated with  $\bar{U}_t$  and  $\underline{U}_t$ , respectively.

LEMMA 5.1. *Versions of the processes  $U$ ,  $\bar{U}$  and  $\underline{U}$ , all with the same initial condition  $\mu$ , can be built on a common probability space in such a way that*

$$\underline{U}_t \leq U_t \leq \bar{U}_t \quad \text{for all } t \leq \tau,$$

where

$$\tau = \inf \left\{ t \geq 0 : \left( \max_x K(x) \right) + \left( \max_x \beta L_t(U, x) \right) \geq \beta \kappa \right\}.$$

PROOF. This follows by first rescaling  $U$ ,  $\bar{U}$  and  $\underline{U}$  as in Lemma 2.27 so that the  $\beta$  parameter becomes 1 and the drift parameters (or transmission rates) of  $\underline{U}$  and  $\bar{U}$  become  $\theta(1 - \beta\kappa)$  and  $\theta$ , respectively. Then one may apply Lemma 2.25 to the rescaled process. Finally undoing the scaling leads to the required conclusion.  $\square$

By Lemma 2.1(a), the law of  $\underline{U}$  is absolutely continuous with respect to that of  $\bar{U}$ , and the likelihood ratio on  $\mathcal{F}_t$  is

$$(5.2) \quad \text{LR}_t^\kappa = \exp \left\{ -\beta \kappa \bar{M}_t(1) - \frac{\beta^2 \kappa^2}{2} \int_0^t |\bar{U}_s| ds \right\}.$$

5.3. *Percolation probability estimates.* Recall that  $Q_r(x)$  denotes the cube of side length  $r$  centered at  $x$ , and, as before, we abbreviate  $Q(x) := Q_1(x)$ . The auxiliary site percolation processes will be constructed by partitioning the space  $\mathbb{R}^d$  into cubes  $Q(x)$  of side length 1 centered at lattice points  $x \in \mathbb{Z}^d$ , and then using the behavior of the superprocesses in the cube  $Q(x)$  to determine whether the site  $x$  will be occupied or not in the auxiliary percolation process. Roughly, a site  $x$  will be occupied if, within a certain fixed amount of time  $T < \infty$ , the measure-valued process  $U$  started from a certain initial mass distribution supported by  $Q(x)$  manages to generate a sufficiently large total mass in each of the adjacent cubes  $Q(y)$  while simultaneously not accumulating too much local time. The objective of this section is to develop estimates that will allow us to conclude that if  $\theta$  is large (hence  $\beta$  is small), then  $x$  is occupied with high probability.

Define the grid  $\Gamma$  to be  $\mathbb{Z}_+^2$  when  $d = 2$  and  $\mathbb{Z}_+^2 \times \{0\}$  when  $d = 3$ , where  $\mathbb{Z}_+^2 := \{x = (x_1, x_2) \in \mathbb{Z}^2 : x_i \geq 0, i = 1, 2\}$ . For  $x, y \in \Gamma$ , we say that

$$(5.3) \quad x \prec y \begin{cases} \text{if } \|x\|_1 < \|y\|_1 & \text{or} \\ \|x\|_1 = \|y\|_1 \text{ and } x_1 < y_1. \end{cases}$$

Here  $\|x\|_1 := \sum_i |x_i|$  is the  $\ell_1$ -norm. This defines a *total* order on  $\Gamma$ , and so the points of the lattice can be enumerated as  $0 = x(1) \prec x(2) \prec \dots$ . The notation  $x \preceq y$  is understood as  $x \prec y$  or  $x = y$ . Define  $\mathcal{A}(x)$  to be the set of  $y \in \Gamma$  such that  $x \prec y$  and  $\|x - y\|_1 = 1$ . In other words, when  $d = 2$ , for any  $(x_1, y_1) \in \mathbb{Z}_+^2$ ,  $\mathcal{A}((x_1, y_1)) = \{(x_1, y_1 + 1), (x_1 + 1, y_1)\}$ ; and similarly for  $d = 3$ . We shall call any  $y \in \mathcal{A}(x)$  an “immediate offspring” of  $x$ , and  $x$  an “immediate predecessor” of  $y$ .

Fix  $T$  so large that (3.21) holds. For any  $\varepsilon > 0$ , let  $A(\varepsilon) = A(T, \varepsilon)$  and  $r_0(\varepsilon) = r_0(T, \varepsilon)$  be the constants specified in Proposition 3.10. For any measure-valued process  $X$  with local time density  $L_t(X)$ , and for any  $x \in \Gamma$ ,  $M > 0$ ,  $\chi > 0$ ,  $\lambda > 0$  and  $\varepsilon > 0$ , define the following events:

$$(5.4) \quad \begin{cases} F^1(M; X, x) = \{\text{Supp}(L_T(X)) \subseteq Q_M(x)\}; \\ F^2(\chi; X) = \{\max_y L_T(X, y) \leq \chi\}; \\ F^3(X, x) = \{X_T(Q(y)) \geq |X_0|, \text{ for all } y \in \mathcal{A}(x)\}; \\ F^4(\varepsilon; X) = \{X_T \text{ is } (A(\varepsilon/4), |X_0|, r_0(\varepsilon/4))\text{-admissible}\}. \end{cases}$$

Observe that these events depend on the choice of  $T$ . For brevity we will write

$$\bar{F}^1(M) = F^1(M; \bar{U}, 0), \quad \bar{F}^2(\chi) = F^2(\chi; \bar{U}), \quad \bar{F}^3 = F^3(\bar{U}, 0), \quad \text{and} \quad \bar{F}^4(\varepsilon) = F^4(\varepsilon; \bar{U}).$$

Define functions

$$(5.5) \quad f_d(\theta) = \begin{cases} \theta^{1/2} & \text{when } d = 2 \\ \log \theta & \text{when } d = 3, \end{cases}$$

and

$$(5.6) \quad \widetilde{M} = \widetilde{M}(M, \theta) = [M\sqrt{\log f_d(\theta)} + 1], \quad \text{and} \quad \chi = \chi(A'', \theta) = A'' f_d(\theta) \kappa_d(f_d(\theta)),$$

where  $\kappa_d(\cdot)$  is the function defined in (3.7).

LEMMA 5.2. *For any  $\varepsilon_0 > 0$ , there exist positive constants  $\theta_0, M$  and  $A''$ , depending only on  $T$  and  $\varepsilon_0$ , such that if  $\theta > \theta_0$  then for any initial measure  $\mu$  supported by  $Q(0)$ , of total mass  $|\mu| = f_d(\theta)$  and  $(A(\varepsilon_0/4), f_d(\theta), r_0(\varepsilon_0/4))$ -admissible, the super-Brownian motion  $\bar{U}_t$  with drift 1 and initial mass distribution  $\mu$  satisfies*

$$(5.7) \quad P\left(\bar{F}^1(\widetilde{M}) \cap \bar{F}^2(\chi) \cap \bar{F}^3 \cap \bar{F}^4(\varepsilon_0)\right) \geq 1 - \varepsilon_0.$$

PROOF. This is a direct consequence of Lemma 3.12, Proposition 3.11, Lemma 3.13, and Proposition 3.10. More specifically, by Lemma 3.12, there exists constant  $M = M(T, \varepsilon_0)$  such that  $P(\bar{F}^1(\widetilde{M})) \geq 1 - \varepsilon_0/4$ . Moreover, by Proposition 3.11 with  $\varepsilon = \varepsilon_0/4$ , there exists  $A''$  such that  $P(\bar{F}^2(\chi)) \geq 1 - \varepsilon_0/4$ . (Note that by Proposition 3.11,  $A''$  only depends on  $(T, M, \varepsilon_0, A, r_0)$ , and in our case the  $M, A$  and  $r_0$  all only depend on  $(T, \varepsilon_0)$ , so ultimately  $A''$  only depends on  $(T, \varepsilon_0)$ .) Next, by Lemma 3.13, there exists  $\theta_0 > 0$  such that for any  $\theta > \theta_0$ ,

$$(5.8) \quad P(\bar{F}^3) \geq 1 - \varepsilon_0/4.$$

Finally, by Proposition 3.10, as long as  $\theta$  is such that  $f_d(\theta) \geq 1$ ,  $P(\overline{F}^4(\varepsilon_0)) \geq 1 - \varepsilon_0/4$ .  $\square$

The next result explains the choice of  $f_d(\theta)$ .

COROLLARY 5.3. *For any positive constants  $\theta, M$  and  $A''$ , let  $\underline{U}_t = \underline{U}_t^{\theta, \kappa}$  be the super-Brownian motion with drift  $1 - \beta\kappa$  and initial mass distribution  $\mu$ , where*

$$(5.9) \quad \kappa = \widetilde{M}^2 \chi = [M\sqrt{\log f_d(\theta)} + 1]^2 \cdot A'' f_d(\theta) \kappa_d(f_d(\theta)).$$

*Then for any  $\varepsilon_0 \in (0, 1)$ , there exists  $\theta_0 > 0$  such that if  $\theta > \theta_0$  and if the initial condition  $\mu$  is supported by  $Q(0)$  and of total mass  $|\mu| = f_d(\theta)$ , then*

$$(5.10) \quad P(F^3(\underline{U}, 0)) \geq 1 - 3\varepsilon_0/2.$$

PROOF. Let  $F^5(\varepsilon_0) = \{LR_T^\kappa \leq 1 + \varepsilon_0\}$  for the likelihood ratio  $LR_t^\kappa$  defined in (5.2). Using the fact that  $\overline{M}_t(1) = \underline{M}_t(1) - \beta\kappa \int_0^t |\underline{U}_s| ds$ , we have

$$(5.11) \quad \begin{aligned} & E^{\overline{U}}(LR_T^\kappa \cdot \mathbf{1}_{(F^5(\varepsilon_0))^c}) \\ &= P^{\underline{U}} \left( -\beta\kappa \overline{M}_T(1) - (\beta^2 \kappa^2 / 2) \int_0^T |\underline{U}_s| ds \geq \log(1 + \varepsilon_0) \right) \\ &= P^{\underline{U}} \left( -\beta\kappa \underline{M}_T(1) + (\beta^2 \kappa^2 / 2) \int_0^T |\underline{U}_s| ds \geq \log(1 + \varepsilon_0) \right). \end{aligned}$$

Here and below we use  $E^{\overline{U}}$  and  $P^{\overline{U}}$  ( $E^{\underline{U}}$  and  $P^{\underline{U}}$ , resp.) to indicate that the expectation and probability are taken with respect to the law of  $\overline{U}$  ( $\underline{U}$ , resp.). Since for any  $s \geq 0$ ,  $E^{\underline{U}}(|\underline{U}_s|) = |\underline{U}_0| e^{(1-\beta\kappa)s}$  (see, e.g., Eqn (5.4) in Feller (1951)), we have that

$$E^{\underline{U}} \int_0^T |\underline{U}_s| ds \leq f_d(\theta) \int_0^T e^s ds \leq C_T f_d(\theta).$$

This and the definitions of  $\beta$  (in (5.1)),  $\kappa$  and  $f_d$  imply that  $E^{\underline{U}}(\beta^2 \kappa^2 \int_0^T |\underline{U}_s| ds) = o(1)$  as  $\theta$  goes to infinity. Since  $\beta\kappa \underline{M}_T(1)$  has square function  $\beta^2 \kappa^2 \int_0^T |\underline{U}_s| ds$ , we see from the above that both terms inside the  $P^{\underline{U}}$ -probability in (5.11) approach 0 in probability as  $\theta \rightarrow \infty$ . It follows that there exists  $\theta_0 > 0$  such that for any  $\theta > \theta_0$ ,  $E^{\overline{U}}(LR_T^\kappa \cdot \mathbf{1}_{(F^5(\varepsilon_0))^c}) < \varepsilon_0$ . Therefore by (5.8), the complement of the event  $F^3(\underline{U}, 0)$  has probability bounded above by

$$\varepsilon_0 + E^{\overline{U}} \left( LR_T^\kappa \cdot \mathbf{1}_{\{LR_T^\kappa \leq 1 + \varepsilon_0\}} \cdot \mathbf{1}_{(F^3)^c} \right) \leq \varepsilon_0 + (1 + \varepsilon_0) \cdot \frac{\varepsilon_0}{4} \leq \frac{3\varepsilon_0}{2}.$$

$\square$

Combining the sandwich lemma (Lemma 5.1) and the previous two results we obtain

PROPOSITION 5.4. *For any  $\varepsilon_0 > 0$  there exist positive constants  $\theta_0, M, A'$  such that, for any  $\theta > \theta_0$ , any initial condition  $\mu$  satisfying the hypotheses of Lemma 5.2, and any  $K \in C_p(\mathbb{R}^d, \mathbb{R}_+)$  such that*

$$K(x) + \beta\chi \cdot \mathbf{1}_{Q_{\widetilde{M}}(0)}(x) \leq \beta\kappa, \quad \text{for all } x \in \mathbb{R}^d,$$

the process  $U$  solving  $(MP)_{\mu, K}^{1, \beta, 1}$  satisfies

$$(5.12) \quad P\left(F^1(\widetilde{M}; U, 0) \cap F^2(\chi; U) \cap F^3(U, 0) \cap F^4(\varepsilon_0; U)\right) \geq 1 - 3\varepsilon_0,$$

where the events  $F^i$  ( $i = 1, 2, 3, 4$ ) are defined as in (5.4) by replacing  $X$  with  $U$  and  $x$  with 0.

PROOF. On the event  $\overline{F}^1(\widetilde{M}) \cap \overline{F}^2(\chi)$ ,

$$L_T(\overline{U}, x) \leq \chi \cdot \mathbf{1}_{Q_{\widetilde{M}}(0)}(x).$$

Therefore by the assumption on  $K$  and Lemma 5.1,

$$\underline{U}_t \leq U_t \leq \overline{U}_t, \quad \text{for all } t \leq T \quad \text{on } \overline{F}^1(\widetilde{M}) \cap \overline{F}^2(\chi).$$

The required bound now follows from Lemma 5.2, Corollary 5.3 and an elementary argument.  $\square$

#### 5.4. Proof of survival.

PROPOSITION 5.5. *For some finite measure  $\mu$  and some  $\theta < \infty$ , if  $U_t$  solves  $(MP)_{\mu, 0}^{1, \beta, 1}$  with  $\beta = \theta^{(6-d)/2}$ , then*

$$P(U \text{ survives}) > 0.$$

PROOF. Fix a  $T$  so that (3.21) holds. Fix  $\varepsilon_0 > 0$  small enough such that any 3-dependent oriented site percolation process on  $\mathbb{Z}_+^2$  with density at least  $(1 - 6\varepsilon_0)$  has positive probability of percolation. For this  $\varepsilon_0$ , let  $\theta > \theta_0$ , where  $\theta_0$  is as in Proposition 5.4. Then choose a measure  $\mu$  so that it satisfies the hypotheses of Lemma 5.2 with  $\varepsilon_0$  specified as above. Let  $L_t(x)$  denote the local time density of  $U$ , and let  $L_\infty(x) = \lim_{t \rightarrow \infty} L_t(x)$  for all  $x \in \mathbb{R}^d$ . By Lemma 3.12 and (a scaled version of) Proposition 2.5, almost surely,

$$L_\infty \text{ is not compactly supported} \implies U \text{ survives.}$$

It therefore suffices to show that  $L_\infty$  is not compactly supported with positive probability. To do so, we will specify an algorithm that produces a (random) set  $\Omega$  consisting of integer sites such that

- (i)  $L_\infty(Q(x)) > 0$  for all  $x \in \Omega$ ; and
- (ii)  $\Omega$  is infinite with positive probability.

The set  $\Omega$  will be the connected cluster containing the origin in a 3–dependent site percolation process with density  $\geq 1 - 6\varepsilon_0$ .

Let us first give an overview of the algorithm. Recall that the grid  $\Gamma$  is defined to be  $\mathbb{Z}_+^2$  when  $d = 2$  and  $\mathbb{Z}_+^2 \times \{0\}$  when  $d = 3$ . Initially all sites  $x \in \Gamma$  are designated *vacant* (that is,  $\Omega = \emptyset$ ). Our algorithm relies on the comparison in Proposition 2.24. Starting from the origin, following the total order  $0 = x(1) \prec x(2) \prec \dots$  on  $\Gamma$  introduced in (5.3), we shall define stopping times  $\tau_i$ , random measures  $\mu_i, \nu_i$  and suppression rates  $K_i^*$ . Proposition 2.24 allows us to couple  $U$  with another process  $U^*$ , which, on any time interval between two successive stopping times, is a usual spatial epidemic process. The set  $\Omega$  will be determined by  $U^*$ . Proposition 2.24 ensures that  $L_t^U \geq L_t^{U^*}$ , which will be used to ensure property (i). Depending on how  $U^*$  behaves for  $t \in [\tau_{i-1}, \tau_i]$ , we may change the status of site  $x = x(i)$  from vacant to *occupied*, and add  $x$  to the set  $\Omega$ . Roughly speaking, this will be done if and only if the spatial epidemic  $U_t^*$  for  $t \in [\tau_{i-1}, \tau_i]$  succeeds in (1) putting enough mass in adjacent cubes at time  $\tau_i$ ; and (2) accumulating only a small amount of local time. On the event that the status of site  $x$  is changed to occupied, for each successor  $y \in \mathcal{A}(x)$  we will be able to extract a “nice” mass distribution  $\mu^y$  in such a way that if a spatial epidemic is initiated by  $\mu^y$  then it will have high probability of making the two events (1) and (2) occur; in other words, so that site  $y$  will also be added to  $\Omega$  with high probability. By keeping this probability above the percolation threshold we will ensure that the random set  $\Omega$  consisting of all the *occupied* sites will be infinite with positive probability.

We now introduce some notation. In addition to  $\theta_0$ , assume  $M, A''$  are as in Proposition 5.4, so that (5.12) holds. In the algorithm, we will repeatedly use stopping rules  $\tau = \tau(Y; \ell; R)$  defined as follows: for a measure-valued process  $Y \in C([0, \infty); \mathcal{M}_c(\mathbb{R}^d))$  with local time  $L_t^Y$ , a threshold  $\ell > 0$ , and a region  $R \subseteq \mathbb{R}^d$ ,

$$(5.13) \quad \tau(Y; \ell; R) := \inf\{t : \max_x L_t^Y(x) \geq \ell \text{ or } \text{Supp}(L_t^Y) \not\subseteq R\} \wedge T.$$

We will also repeatedly use the notation  $F^1, \dots, F^4$  as introduced in (5.4) to define the so-called “good events”. For notational ease, for each  $i = 1, 2, \dots$ , associated with site  $x(i)$  in the above overview, we write  $U_t^i = U_{t+\tau_{i-1}}^*$  for  $t \geq 0$  (i.e., the process  $U^*$  shifted and restricted to  $t \geq \tau_{i-1}$ ), and

$$(5.14) \quad G^i = F^1(\widetilde{M}; U^i, x(i)) \cap F^2(\chi; U^i) \cap F^3(U^i, x(i)) \cap F^4(\varepsilon_0; U^i).$$

The event  $G^i$  will be called a “good” event. In plain language, ignoring the technical restriction  $F^4$ , on such a good event, before time  $T$ , the spatial epidemic  $U^i$  has not accumulated local time density more than  $\chi \cdot \mathbf{1}_{Q_{\widetilde{M}}(x(i))}$ , and in the meanwhile, at time  $T$ , it spreads at least  $|U_0^i|$  amount of mass in all the cubes  $Q(y)$  for  $y \in \mathcal{A}(x(i))$ .

Now we describe our algorithm in detail. In order to apply Proposition 2.24, we need to define four sequences: random measures  $\mu_i, \nu_i$ , suppression rate functions  $K_i^*$  and stopping times  $\tau_i$ . The random measures  $\mu_i$  and  $\nu_i$  will be defined through an auxiliary random measure sequence  $w_i$ . The suppression rate functions  $K_i^*$  will be deterministic functions as follows:  $K_0^* \equiv 0$ , and for  $i \geq 1$ ,

$$(5.15) \quad K_i^* = \beta\chi \cdot \sum_{j=1}^i \mathbf{1}_{Q_{\widetilde{M}}(x(j))}.$$

Observe that for each  $i$ ,  $K_i^*$  is a summation of *moving windows*, and is bounded by  $\beta\kappa$  everywhere (recall that  $\kappa$  is defined in (5.9) and note each point is covered by at most  $\widetilde{M}^2$  cubes of the form  $Q_{\widetilde{M}}(x_j)$  for centers in our 2-dimensional grid).

We start with site  $x(1) = 0$ . The  $\tau_0$ ,  $\mu_0$  and  $\nu_0$  are all deterministic:  $\tau_0 = 0$ ,  $\mu_0 = \mu$  and  $\nu_0 = 0$ . Let  $\tau_1 = \tau(U^1; \chi, Q_{\widetilde{M}}(x(1)))$ . By Proposition 5.4, the good event  $G^1$  occurs with probability  $\geq 1 - 3\varepsilon_0$ . Observe also that  $\tau_1 > 0$  almost surely and  $\tau_1 = T$  on  $G^1$ . If the good event  $G^1$  occurs, then we change the status of site 0 to be occupied. Further define

$$(5.16) \quad w_1 = \begin{cases} \sum_{z \in \mathcal{A}(x(1))} \frac{|\mu|}{U_{\tau_1}^*(Q(z))} \cdot U_{\tau_1}^*(\cdot \cap Q(z)), & \text{if } G^1 \text{ occurs;} \\ 0, & \text{otherwise.} \end{cases}$$

We now work with site  $y = x(i)$  for  $i \geq 2$ . We proceed according to whether the site  $y$  is an immediate offspring of some occupied site or not.

Case I. Site  $y$  is an immediate offspring of some occupied site. Define

$$(5.17) \quad (\mu_{i-1}, \nu_{i-1}) = (w_{i-1}(\cdot \cap Q(y)), w_{i-1}(\cdot \cap Q(y)^c)).$$

Then  $\mu_{i-1}$  is a measure supported by  $Q(y)$ , of total mass  $|\mu| = f_d(\theta)$ , and  $(A(\varepsilon_0/4), f_d(\theta), r_0(\varepsilon_0/4))$ -admissible. Let  $\tau_i = \tau_{i-1} + \tau(U^i; \chi, Q_{\widetilde{M}}(y))$ . By Proposition 5.4 (with an apparent spatial translation), the good event  $G^i$  occurs with probability  $\geq 1 - 3\varepsilon_0$ . Observe also that  $\tau_i - \tau_{i-1} = T$  on  $G^i$ . If the good event  $G^i$  occurs, then we change the status of site  $y$  to occupied. Moreover, according to whether  $G^i$  occurs or not, we define  $w_i$  as follows:

$$(5.18) \quad w_i = \begin{cases} \nu_{i-1} + \sum_{z \in \widetilde{\mathcal{A}}(y)} \frac{|\mu|}{U_{\tau_i}^*(Q(z))} \cdot U_{\tau_i}^*(\cdot \cap Q(z)), & \text{if } G^i \text{ occurs;} \\ \nu_{i-1}, & \text{otherwise.} \end{cases}$$

where

$$\widetilde{\mathcal{A}}(y) = \{z \in \mathcal{A}(y) : z \notin \mathcal{A}(u) \text{ for } u \text{ which is occupied and } \prec y\}.$$

Case II. Site  $y$  is not an immediate offspring of any occupied site. Then we set  $(\mu_{i-1}, \nu_{i-1}) = (0, w_{i-1})$ ,  $\tau_i = \tau_{i-1}$ , and  $w_i = w_{i-1}$ .

In either case at time  $\tau_i$  we proceed to site  $x(i+1)$ .

It is easy to see that such defined  $\mu_i, \nu_i, K_i^*$  and  $\tau_i$  satisfy the conditions of Proposition 2.24, and therefore the processes  $U$  and  $U^*$  can be coupled such that

$$L_t^U \geq L_t^{U^*}, \quad \text{for all } t \geq 0.$$

Now if we let  $\Omega$  be the set of all occupied sites, then by the algorithm above, for any  $x = x(i) \in \Omega$ ,

$$L_\infty^{U^*}(Q(x)) \geq L_{\tau_i}^{U^*}(Q(x)) - L_{\tau_{i-1}}^{U^*}(Q(x)) > 0,$$

and hence  $\Omega$  satisfies the condition (i).

We now show that  $\Omega$  is infinite with positive probability. Define a site percolation on  $\Gamma$  as follows: for each  $x \in \Gamma$ , if  $x$  is occupied, then we let  $\xi(x) = 1$  if both  $y \in \mathcal{A}(x)$  are occupied,

and  $= 0$  otherwise; if  $x$  is vacant, then we let  $\xi(x)$  be a Bernoulli( $1 - 6\varepsilon_0$ ) random variable that is independent of everything else.

We know that the origin is occupied with positive probability. We claim that on the event that the origin is occupied,  $\Omega$  contains the collection of sites reachable from the origin. We may assume that  $\xi(0) = 1$  since otherwise we are done. But when  $0$  is occupied,  $\xi(0) = 1$  implies that both  $y \in \mathcal{A}(0)$  are occupied. By induction the conclusion follows.

It remains to show that the above defined site percolation is a 3-dependent site percolation with density at least  $(1 - 6\varepsilon_0)$ , i.e., we need to show that for any  $n \geq 1$  and any  $1 \leq i_1 < \dots < i_n$  such that  $\|x(i_j) - x(i_k)\|_1 \geq 3$ ,

$$P(\xi(x(i_j)) = 0 \text{ for all } j = 1, \dots, n) \leq (6\varepsilon_0)^n.$$

Since when a site  $x$  is vacant,  $\xi(x)$  is a Bernoulli( $1 - 6\varepsilon_0$ ) random variable independent of everything else, we need only to show

$$(5.19) \quad P(\xi(x(i_j)) = 0 \text{ for all } j = 1, \dots, n \mid \text{all } x(i_j)\text{'s are occupied}) \leq (6\varepsilon_0)^n.$$

Let us first consider the  $n = 1$  case. When  $x := x(i_1)$  is occupied, by construction, each  $y \in \mathcal{A}(x)$  is occupied with probability at least  $1 - 3\varepsilon_0$ , hence the probability that both  $y \in \mathcal{A}(x)$  are occupied is at least  $1 - 6\varepsilon_0$ . (5.19) follows.

In general, for each  $m \geq 0$ , we define  $\mathcal{G}_m$  to be the  $\sigma$ -algebra generated by  $\{U_t^* : 0 \leq t \leq \tau_m\}$ . Then for each  $i \geq 1$ , the good event  $G^i$  is measurable with respect to  $\mathcal{G}_i$ , and hence the Bernoulli random variable  $\xi(x(i))$  is measurable with respect to  $\mathcal{G}_\ell$  where  $\ell$  is the index of the second  $y \in \mathcal{A}(x(i))$ . Now since  $x(i_j)$ 's are at least distance 3 from each other, if we let  $\ell_j$  be the index of the second  $y \in \mathcal{A}(x(i_j))$ , then

$$\ell_j < \ell_n - 2, \quad \text{for all } j < n.$$

Hence by further conditioning on  $\mathcal{G}_{\ell_n - 2}$ , (5.19) reduces to the  $n = 1$  case and hence holds.  $\square$

**6. Proof of extinction when  $d = 2$  or  $3$ .** As the title suggests we shall assume  $d = 2$  or  $3$  throughout this Section.

### 6.1. Scaled process.

PROPOSITION 6.1. *Suppose  $U_t$  is such that for each  $\psi \in C_c^2(\mathbb{R}^d)$ ,*

$$(6.1) \quad U_t(\psi) = U_0(\psi) + \frac{\alpha}{2} \int_0^t U_s(\Delta\psi) ds + \varepsilon \int_0^t U_s(\psi) ds - \beta \int_0^t U_s(L_t^U \cdot \psi) ds + \sqrt{\gamma} M_t(\psi),$$

where  $M_t(\psi)$  is a martingale with quadratic variation  $[M(\psi)]_t = \int_0^t U_s(\psi^2) ds$ . There exist positive constants  $\varepsilon_0$  and  $\zeta$  such that if the initial condition  $U_0$  belongs to the class

$$(6.2) \quad \mathcal{C} := \{\mu \text{ satisfying Assumption 2.10, } \text{Supp}(\mu) \subseteq Q(0), \text{ and } |\mu| = 2\},$$

and the positive parameters  $\alpha, \varepsilon, \beta$  and  $\gamma$  satisfy Assumption 6.2 below, then

$$P(U \text{ dies out}) = 1.$$

ASSUMPTION 6.2.

$$\varepsilon \leq \frac{\beta}{2 \cdot 3^d}, \quad \max\left(\varepsilon, \frac{\alpha}{\gamma}, \frac{\sqrt{\varepsilon}}{\gamma}\right) \leq \varepsilon_0, \quad \text{and} \quad \min\left(\frac{\beta}{\varepsilon^2}, \frac{\beta^2}{\varepsilon^3 \gamma}, \frac{1}{\gamma}, \frac{\beta}{\gamma}\right) \geq \zeta.$$

We denote by  $P_\mu^{\alpha, \varepsilon, \beta, \gamma}$  the law of  $U_t$  satisfying (6.1) with  $U_0 = \mu \in \mathcal{C}$ . Then we can rephrase the conclusion of Proposition 6.1 as

$$p^{\alpha, \varepsilon, \beta, \gamma} = 0,$$

where

$$(6.3) \quad p^{\alpha, \varepsilon, \beta, \gamma} := \sup_{\mu \in \mathcal{C}} P_\mu^{\alpha, \varepsilon, \beta, \gamma}(U \text{ survives}).$$

When there is no confusion about the initial configuration  $\mu$ , we omit  $\mu$  and write  $P^{\alpha, \varepsilon, \beta, \gamma}$  and sometimes just write  $P$ . Note that  $P^{\alpha, \varepsilon, 0, \gamma}$  denotes the law of a Dawson-Watanabe process without any local time killing, and  $P^{\alpha, 0, 0, \gamma}$  the law of driftless Dawson-Watanabe process. By (a scaled version of) Proposition 2.5 we see that when  $\beta > 0$ ,

$$(6.4) \quad U^{\alpha, \varepsilon, \beta, \gamma} \lesssim U^{\alpha, \varepsilon, 0, \gamma},$$

where  $U^{\alpha, \varepsilon, \beta, \gamma}$  has law  $P_\mu^{\alpha, \varepsilon, \beta, \gamma}$ ,  $U^{\alpha, \varepsilon, 0, \gamma}$  has law  $P_\mu^{\alpha, \varepsilon, 0, \gamma}$  and the above notation means we can define versions of these processes on the same space with  $U_t^{\alpha, \varepsilon, \beta, \gamma} \leq U_t^{\alpha, \varepsilon, 0, \gamma}$  for all  $t \geq 0$  almost surely. Furthermore, by Lemma 2.1, the laws  $P_\mu^{\alpha, \varepsilon, 0, \gamma}$  and  $P_\mu^{\alpha, 0, 0, \gamma}$  are related to each other via the likelihood ratio

$$(6.5) \quad \left. \frac{dP_\mu^{\alpha, \varepsilon, 0, \gamma}}{dP_\mu^{\alpha, 0, 0, \gamma}}(U) \right|_{\mathcal{F}_t} = \exp\left(\frac{\varepsilon}{\sqrt{\gamma}} M_t(1) - \frac{\varepsilon^2}{2\gamma} \int_0^t |U_s| ds\right).$$

We introduce the following notation:

$$V_t = |U_t|; \quad \tau_3 = \inf\{t : \text{Supp}(L_t^U) \not\subseteq Q_3(0)\},$$

and for any continuous real valued process  $X$  and any  $c \in \mathbb{R}$ , we let  $T_c(X)$  be the hitting time:

$$T_c(X) = \inf\{t : X_t = c\}.$$

Finally, define  $\tau$  to be the first time that  $V_t$  hits 0 or 4 or that  $U_t$  exits  $Q_3(0)$ , i.e.,

$$(6.6) \quad \tau = T_0(V) \wedge T_4(V) \wedge \tau_3.$$

LEMMA 6.3.  $\tau < \infty$  almost surely.

PROPOSITION 6.4. *There exist constants  $\varepsilon_0$  and  $\zeta$  such that if the parameters  $\alpha, \varepsilon, \beta$  and  $\gamma$  satisfy Assumption 6.2, then*

$$(6.7) \quad \sup_{\mu \in \mathcal{C}} P_\mu^{\alpha, \varepsilon, \beta, \gamma}(V_\tau > 0) < \frac{1}{2 \cdot 3^d} := p_c.$$

We will prove these results in the next sub-Section. Proposition 6.4 is analogous to Lemma 2.3.1 in Mueller and Tribe (1994). Once we have the Proposition, we can prove Proposition 6.1 by constructing a sub-critical branching process as in Mueller and Tribe (1994), or more directly as follows.

PROOF OF PROPOSITION 6.1. Suppose that the positive parameters  $\alpha, \varepsilon, \beta$  and  $\gamma$  satisfy the assumption of Proposition 6.4. Let

$$r_1 := \frac{\sup_{\mu \in \mathcal{C}} P_\mu^{\alpha, \varepsilon, \beta, \gamma}(V_\tau > 0)}{p_c} < 1.$$

By the definition (6.3) of  $p^{\alpha, \varepsilon, \beta, \gamma}$ , we can find a  $\mu \in \mathcal{C}$  such that

$$(6.8) \quad P_\mu^{\alpha, \varepsilon, \beta, \gamma}(U \text{ survives}) \geq \frac{1 + r_1}{2} p^{\alpha, \varepsilon, \beta, \gamma}.$$

Let  $U_t$  satisfy (6.1) with  $U_0 = \mu$ . For this  $U$ , at time  $\tau$ , on the event that  $V_\tau > 0$ ,  $U_\tau$  is contained in  $Q_3(0)$  with total mass no greater than 4. We can then decompose it into no more than  $2 \times 3^d$  parts:

$$U_\tau = \sum_{i=1}^{\ell} U_\tau^i, \quad \ell \leq 2 \times 3^d,$$

each of whom has support contained in a unit cube, total mass at most 2, and satisfies Assumption 2.10. To see this last property, the domination in (6.4), the absolute continuity in (6.5), and the finite propagation speed of the super-Brownian motion (see, e.g., Theorem III.1.3 in Perkins (2002)) show that it suffices to prove that if  $U$  is the super-Brownian motion with law  $P_\mu^{\alpha, 0, 0, \gamma}$ , then  $U_\tau$  satisfies Assumption 2.10 a.s. The last claim follows directly from Theorem III.3.4. in Perkins (2002).

By the Markov property of the joint process  $(U, L^U)$  (see Theorem 2.2(d)), Lemma 2.15, and (a scaled version of) Lemma 2.18,

$$P_\mu^{\alpha, \varepsilon, \beta, \gamma}(U \text{ survives}) \leq E \left( \mathbf{1}_{(V_\tau > 0)} \cdot P_{U_\tau}^{\alpha, \varepsilon, \beta, \gamma}(U \text{ survives}) \right).$$

Here we are “throwing away” the killing due to  $L_\tau^U$ . By Lemma 2.19 and translation invariance, the right-hand side is bounded above by

$$E \left( \mathbf{1}_{(V_\tau > 0)} \cdot \sum_{i=1}^{\ell} P_{U_\tau^i}^{\alpha, \varepsilon, \beta, \gamma}(U \text{ survives}) \right) \leq P(V_\tau > 0) \cdot E \left( \sum_{i=1}^{\ell} p^{\alpha, \varepsilon, \beta, \gamma} \right) \leq r_1 p^{\alpha, \varepsilon, \beta, \gamma}.$$

Combining this with the previous inequality and (6.8) we get

$$\frac{1 + r_1}{2} p^{\alpha, \varepsilon, \beta, \gamma} \leq r_1 p^{\alpha, \varepsilon, \beta, \gamma},$$

hence  $p^{\alpha, \varepsilon, \beta, \gamma} = 0$ . □

6.2. *Proof of Lemma 6.3 and Proposition 6.4.* In the arguments below,  $U_t$  is a process satisfying (6.1) with a fixed initial condition  $\mu \in \mathcal{C}$ . The bounds in Lemmas 6.6 – 6.9 below hold for *all*  $\mu \in \mathcal{C}$ , and hence will lead to the uniform bound in Proposition 6.4.

First we note that  $V_t = |U_t|$  satisfies the following SDE for some Brownian motion  $W$ :

$$(6.9) \quad dV_t = \varepsilon V_t dt - \beta U_t(L_t^U) dt + \sqrt{\gamma} \sqrt{V_t} dW_t.$$

By an integration by parts,

$$\beta \int_0^t U_s(L_s^U) ds = \frac{\beta}{2} \int (L_t^U(x))^2 dx.$$

When  $t \leq \tau_3$ , by Cauchy-Schwartz we get that

$$(6.10) \quad \frac{\beta}{2} \int (L_t^U(x))^2 dx \geq \frac{\beta}{2} \cdot \frac{1}{3^d} \left( \int L_t^U(x) dx \right)^2 = p_c \beta \left( \int_0^t V_s ds \right)^2.$$

We now prove Lemma 6.3.

PROOF OF LEMMA 6.3. Suppose otherwise  $P(\tau = \infty) > 0$ , in particular,  $P(\tau_3 = \infty) > 0$ . By (6.9) and (6.10), on the event  $\{\tau_3 = \infty\}$ ,

$$(6.11) \quad V_t \leq 2 + \varepsilon \int_0^t V_s ds - p_c \beta \left( \int_0^t V_s ds \right)^2 + \sqrt{\gamma} \int_0^t \sqrt{V_s} dW_s, \text{ for all } t \geq 0.$$

Define a sequence of stopping times  $\{r_i\}$  by  $r_0 = 0$  and for  $i \geq 0$ ,

$$r_{i+1} = \begin{cases} r_i + 1 & \text{if } V_{r_i+1} \leq 2, \\ \inf\{t \geq r_i + 1 : V_t = 2\} & \text{otherwise.} \end{cases}$$

CLAIM 6.5. *For all  $i$ ,  $r_i < \infty$  almost surely*

Suppose for some  $i$ ,  $r_i < r_{i+1} = \infty$ . Then  $V_t > 2$  for all  $t \geq r_i + 1$ . Therefore (6.11) shows that on  $\{\tau_3 = \infty\}$  the continuous martingale  $\sqrt{\gamma} \int_0^t \sqrt{V_s} dW_s$  approaches  $+\infty$  as  $t \rightarrow \infty$ , an event of probability zero. This proves the claim.

For each  $i$  Proposition 2.5 allows us to bound  $V_t$  above on  $[r_i, r_i + 1]$  by a Feller diffusion with drift  $\varepsilon$  and initial value 2 which does hit 0 in the next one unit of time with probability  $q > 0$ . This shows  $P(V \text{ hits } 0 \text{ on } [r_i, r_i + 1] | \mathcal{F}_{r_i}) \geq q > 0$  and we therefore conclude that  $V$  will hit 0 almost surely, again a contradiction to our supposition.  $\square$

Next we prove Proposition 6.4. Define a continuous random time change  $\eta : \left[0, \int_0^{T_0(V)} V_s ds\right] \rightarrow [0, T_0(V)]$  by

$$(6.12) \quad \eta_t = \inf \left\{ r : \int_0^r V_s ds = t \right\},$$

and let  $\tilde{V}_t = V_{\eta_t}$ . Then  $\tilde{V}_t$  satisfies

$$\tilde{V}_t = 2 + \varepsilon t - \beta \int_0^{\eta_t} U_s(L_s^U) ds + \sqrt{\gamma} B_t, \quad \text{for } t \leq \int_0^{T_0(V)} V_s ds,$$

where  $B_t = \int_0^{\eta_t} \sqrt{V_s} dW_t$  for  $t \leq \int_0^{T_0(V)} V_s ds$  and may be extended, if necessary, to a standard Brownian motion. If

$$(6.13) \quad Y_t = 2 + \varepsilon t - p_c \beta t^2 + \sqrt{\gamma} B_t.$$

then by (6.10)

$$(6.14) \quad \tilde{V}_t \leq Y_t \quad \text{for } t \leq \int_0^{\tau_3 \wedge T_0(V)} V_s ds,$$

since the upper bound on  $t$  implies  $\eta_t \leq \tau_3$ .

We want to bound  $P(V_\tau > 0)$  where  $\tau$  is defined in (6.6). Using the comparison above, noting that by Lemma 6.3  $\tau < \infty$  almost surely, we get that

$$(6.15) \quad \begin{aligned} & P(V_\tau > 0) \\ & \leq P(\tau_3 < T_0(V), \tau_3 < T_4(V)) + P(\tau = T_4(V)) \\ & \leq P(\tau_3 < 1/(4\varepsilon)) + P(1/(4\varepsilon) \leq \tau_3 \leq T_0(V)) + P(T_4(V(\cdot \wedge \tau_3)) < T_0(V)) \\ & \leq P(\tau_3 < 1/(4\varepsilon)) + P(T_1(V) \leq 1/(8\varepsilon) \text{ and } T_0(V) \geq 1/(4\varepsilon)) \\ & \quad + P(T_0(Y) > 1/(8\varepsilon)) + P(T_4(Y) < T_0(Y)), \end{aligned}$$

where in the last line we used that

$$P(T_1(V) > 1/(8\varepsilon), T_0(V) \geq 1/(4\varepsilon) \text{ and } \tau_3 \geq 1/(4\varepsilon)) \leq P(T_0(Y) > 1/(8\varepsilon)).$$

This holds because  $V_0 = 2$ , and hence on the event on the left-hand side

$$\int_0^{\tau_3 \wedge T_0(V)} V_s ds \geq \int_0^{1/(8\varepsilon)} V_s ds \geq 1/(8\varepsilon),$$

which implies  $\eta_{1/(8\varepsilon)} \leq 1/(8\varepsilon)$ , and by (6.14), for all  $t \leq 1/(8\varepsilon)$ ,  $Y_t \geq \tilde{V}_t = V_{\eta_t} > 0$  (since  $T_0(V) \geq 1/(4\varepsilon)$ ). Proposition 6.4 will be proved if we can show that all four probabilities in (6.15) are small.

LEMMA 6.6. *There exists a constant  $C > 0$  such that*

$$P(\tau_3 \leq 1/(4\varepsilon)) \leq C \sqrt{\frac{\alpha}{\gamma}} \exp\left(\frac{\sqrt{\varepsilon}}{8\gamma}\right) + 2\sqrt{\varepsilon} \exp\left(\frac{2\sqrt{\varepsilon}}{\gamma}\right).$$

PROOF. By the domination (6.4), it suffices to show the lemma for  $P^{\alpha, \varepsilon, 0, \gamma}$ , which is then analogous to Lemma 2.1.9 in Mueller and Tribe (1994) where the conclusion for the  $d = 1$  case is proved. We give here a slightly simpler proof for all  $d \leq 3$ .

Following Mueller and Tribe (1994) and using (6.5), we get that

$$\begin{aligned}
(6.16) \quad & P^{\alpha,\varepsilon,0,\gamma}(\tau_3 \leq 1/(4\varepsilon)) \\
& \leq E^{\alpha,0,0,\gamma} \left( \mathbf{1}_{\{\tau_3 \leq 1/(4\varepsilon) \wedge T_{\varepsilon^{-1/2}}(V)\}} \cdot \exp \left( \frac{\varepsilon}{\sqrt{\gamma}} M_{\tau_3}(1) - \frac{\varepsilon^2}{2\gamma} \int_0^{\tau_3} V_s ds \right) \right) \\
& \quad + P^{\alpha,\varepsilon,0,\gamma}(T_{\varepsilon^{-1/2}}(V) < 1/(4\varepsilon)) \\
& \leq \sqrt{P^{\alpha,0,0,\gamma}(\tau_3 \leq 1/(4\varepsilon))} \\
(6.17) \quad & \cdot \sqrt{E^{\alpha,0,0,\gamma} \left( \exp \left( \frac{2\varepsilon}{\sqrt{\gamma}} M_{\tau_3}(1) - \frac{\varepsilon^2}{\gamma} \int_0^{\tau_3} V_s ds \right) \cdot \mathbf{1}_{\{\tau_3 \leq 1/(4\varepsilon) \wedge T_{\varepsilon^{-1/2}}(V)\}} \right)} \\
& \quad + P^{\alpha,\varepsilon,0,\gamma}(T_{\varepsilon^{-1/2}}(V) < \infty).
\end{aligned}$$

A scale function (see, e.g., Proposition VII.3.2 and Exercise VII.3.20 in Revuz and Yor (1999)) for  $V$  when  $\beta = 0$  is given by  $s(x) = \gamma(1 - \exp(-2\varepsilon x/\gamma))/(2\varepsilon)$  and so

$$\begin{aligned}
(6.18) \quad & P^{\alpha,\varepsilon,0,\gamma}(T_{\varepsilon^{-1/2}}(V) < \infty) = \frac{s(2) - s(0)}{s(\varepsilon^{-1/2}) - s(0)} = \frac{1 - \exp(-4\varepsilon/\gamma)}{1 - \exp(-2\sqrt{\varepsilon}/\gamma)} \\
& \leq \frac{4\varepsilon/\gamma}{(2\sqrt{\varepsilon}/\gamma) \exp(-2\sqrt{\varepsilon}/\gamma)} = 2\sqrt{\varepsilon} \exp(2\sqrt{\varepsilon}/\gamma).
\end{aligned}$$

We will use Theorem 1 of Iscoe (1988) to bound  $P^{\alpha,0,0,\gamma}(\tau_3 \leq 1/(4\varepsilon)) \leq P^{\alpha,0,0,\gamma}(\tau_3 < \infty)$ . To do so, we make another scaling: let

$$\tilde{U}_t(\psi) = \frac{U_t \left( \psi \left( \sqrt{\frac{2}{\alpha}} x \right) \right)}{\gamma}, \quad \text{for all } \psi \in C_c^2(\mathbb{R}^d),$$

then by Lemma 2.27  $\tilde{U}_t$  satisfies the assumptions of Theorem 1 in Iscoe (1988), and

$$U_t(Q_3^c(0)) > 0 \iff \tilde{U}_t \left( Q_{3\sqrt{\frac{2}{\alpha}}}^c(0) \right) > 0.$$

Hence by Theorem 1 in Iscoe (1988) and the fact that  $U_0 \in \mathcal{C}$ ,

$$(6.19) \quad P^{\alpha,0,0,\gamma}(\tau_3 < \infty) \leq \frac{\tilde{U}_0 \left( u \left( \left( \frac{3}{2} \sqrt{\frac{2}{\alpha}} \right)^{-1} \cdot \right) \right)}{\left( \frac{3}{2} \sqrt{\frac{2}{\alpha}} \right)^2} \leq C u \left( \frac{2}{3} e_1 \right) \cdot \frac{\alpha}{\gamma},$$

where  $e_1$  is a unit vector and  $u(x)$  is the unique positive (radial) solution of the singular elliptic boundary value problem

$$\Delta u(x) = u^2(x), \quad x \in B(0,1); \quad \text{and } u(x) \rightarrow \infty \text{ as } |x| \rightarrow 1.$$

Next denote by  $\lambda = \frac{2\varepsilon}{\sqrt{\gamma}}$ . Since  $Z(t) = \exp\left(\lambda M_t(1) - \frac{\lambda^2}{2} \int_0^t V_s ds\right)$  is a supermartingale (being a non-negative local martingale),

$$\begin{aligned}
 & E^{\alpha,0,0,\gamma} \left( \exp\left(\frac{2\varepsilon}{\sqrt{\gamma}} M_{\tau_3}(1) - \frac{\varepsilon^2}{\gamma} \int_0^{\tau_3} V_s ds\right) \cdot \mathbf{1}_{\{\tau_3 \leq 1/(4\varepsilon) \wedge T_{\varepsilon^{-1/2}}(V)\}} \right) \\
 &= E^{\alpha,0,0,\gamma} \left( Z(\tau_3 \wedge (4\varepsilon)^{-1}) \exp\left(\frac{\lambda^2}{4} \int_0^{\tau_3} V_s ds\right) \cdot \mathbf{1}_{\{\tau_3 \leq 1/(4\varepsilon) \wedge T_{\varepsilon^{-1/2}}(V)\}} \right) \\
 &\leq \exp\left(\frac{\lambda^2}{4} \cdot \frac{1}{4\varepsilon^{3/2}}\right) E^{\alpha,0,0,\gamma}(Z_0) \\
 (6.20) \quad &= \exp\left(\frac{\sqrt{\varepsilon}}{4\gamma}\right),
 \end{aligned}$$

where optional sampling is used in the next to last line. Now insert (6.18), (6.19), and (6.20) into (6.16) to complete the proof.  $\square$

LEMMA 6.7. *There exists a constant  $C > 0$  such that*

$$P(T_1(V) \leq 1/(8\varepsilon) \text{ and } T_0(V) \geq 1/(4\varepsilon)) \leq C \frac{\varepsilon}{\gamma}.$$

PROOF. Recall that  $V_t$  satisfies (6.9). Applying Proposition 2.5 again, on  $\{T_1(V) < \infty\}$  we may define an  $\mathcal{F}_{T_1(V)+t}$ -adapted solution  $\bar{V}$  of

$$\bar{V}_t = 1 + \varepsilon \int_0^t \bar{V}_s ds + \sqrt{\gamma} \int_0^t \sqrt{\bar{V}_s} dW'_s,$$

where  $W'$  is an  $\mathcal{F}_{T_1(V)+t}$ -Brownian motion and  $\bar{V}_t \geq V_{T_1(V)+t}$  for all  $t \geq 0$ , almost surely on  $\{T_1(V) < \infty\}$ . Therefore

$$P(T_1(V) \leq 1/(8\varepsilon) \text{ and } T_0(V) \geq 1/(4\varepsilon)) \leq P(T_0(\bar{V}) \geq 1/(8\varepsilon)).$$

By Exercise II.5.3. in Perkins (2002) the last term equals

$$1 - \exp\left(\frac{-2\varepsilon}{\gamma(1 - e^{-1/8})}\right) \leq \frac{2\varepsilon}{\gamma(1 - e^{-1/8})}.$$

$\square$

Recall that  $Y$  is defined in (6.13).

LEMMA 6.8. *There exist constants  $C_1, C_2 > 0$  such that for all  $\beta\varepsilon^{-2} \geq 20000$  and  $0 < \varepsilon \leq 1/4$ ,*

$$P(T_0(Y) > 1/(8\varepsilon)) \leq C_1 \exp\left(-C_2 \frac{\beta^2}{\varepsilon^3 \gamma}\right).$$

PROOF. Assume  $\beta, \varepsilon$  are as above and recall that  $p_c = (1/2)3^{-d}$ .

$$\begin{aligned} P(T_0(Y) > 1/(8\varepsilon)) &\leq P(Y_{1/(8\varepsilon)} > 0) \\ &= P\left(B_{1/(8\varepsilon)} \geq \frac{1}{\sqrt{\gamma}} \left(\frac{p_c}{64} \cdot \beta\varepsilon^{-2} - 2 - \frac{1}{8}\right)\right) \\ &\leq P\left(B_1 \geq \frac{\sqrt{8\varepsilon} p_c}{64} \frac{1}{\sqrt{\gamma}} (\beta\varepsilon^{-2} - 10000)\right) \\ &\leq P\left(B_1 \geq \frac{\sqrt{8\varepsilon} p_c}{128} \frac{1}{\sqrt{\gamma}} \cdot (\beta\varepsilon^{-2})\right). \end{aligned}$$

The result follows. □

LEMMA 6.9. *There exists  $C > 0$  such that if  $\varepsilon \leq \min(1/2, p_c\beta)$ , then*

$$P(T_4(Y) < T_0(Y)) \leq C \exp\left(-\frac{1}{8\gamma}\right) + \exp\left(-2p_c\frac{\beta}{\gamma}\right).$$

PROOF. Recall that  $Y_t$  satisfies

$$Y_t = 2 + \varepsilon t - p_c\beta t^2 + \sqrt{\gamma}B_t.$$

Hence if we define  $\tilde{Y}_t$  by

$$\tilde{Y}_t = 2 + \varepsilon t + \sqrt{\gamma}B_t,$$

then  $Y_t \leq \tilde{Y}_t$ . Note also that

$$P(T_4(Y) < T_0(Y)) \leq P(T_3(\tilde{Y}) \leq 1) + P(T_3(\tilde{Y}) \geq 1, T_4(Y) < T_0(Y)).$$

We first estimate  $P(T_3(\tilde{Y}) \leq 1)$ :

$$\begin{aligned} P(T_3(\tilde{Y}) \leq 1) &= P(\max_{t \leq 1}(\varepsilon t + \sqrt{\gamma}B_t) \geq 1) \\ (6.21) \quad &\leq P\left(\max_{t \leq 1} B_t \geq \frac{1-\varepsilon}{\sqrt{\gamma}}\right) \\ &\leq C \exp\left(-\frac{1}{8\gamma}\right) \end{aligned}$$

provided that  $\varepsilon \leq 1/2$ , where  $C > 0$  is some constant independent of  $\varepsilon$  and  $\gamma$ .

We now work with  $P(T_3(\tilde{Y}) \geq 1, T_4(Y) < T_0(Y))$ . Define

$$\bar{Y}_t = Y_1 - p_c\beta t + \sqrt{\gamma}\bar{B}_t,$$

where  $\bar{B}_t = B_{t+1} - B_1$ . If  $\varepsilon \leq p_c\beta$  and  $t \geq 1$ , then

$$\begin{aligned} Y_t &= Y_1 + \varepsilon(t-1) - p_c\beta(t^2-1) + \sqrt{\gamma}\bar{B}_{t-1} \\ &= Y_1 + (t-1)[\varepsilon - p_c\beta(t+1)] + \sqrt{\gamma}\bar{B}_{t-1} \\ &\leq Y_1 + (t-1)[p_c\beta - 2p_c\beta] + \sqrt{\gamma}\bar{B}_{t-1} \\ &= \bar{Y}_{t-1}. \end{aligned}$$

Furthermore, since  $Y_t \leq \tilde{Y}_t$ , on the event  $\{T_3(\tilde{Y}) \geq 1\}$ ,  $Y_1 \leq 3$  and  $T_4(Y) > 1$ . Therefore

$$P(T_3(\tilde{Y}) \geq 1, T_4(Y) < T_0(Y)) \leq P(3 - p_c \beta t + \sqrt{\gamma} \bar{B}_t \text{ hits 4 before 0}).$$

The latter probability can be explicitly calculated using scale functions: if we let

$$s(x) = \int_0^x \exp\left(\int_0^y \frac{2p_c \beta}{\gamma} dz\right) dy = \frac{\gamma}{2p_c \beta} \left(\exp\left(\frac{2p_c \beta}{\gamma} x\right) - 1\right),$$

then

$$\begin{aligned} P\left(3 - \frac{p_c \beta}{2} \cdot t + \sqrt{\gamma} \bar{B}_t \text{ hits 4 before 0}\right) &= \frac{s(3) - s(0)}{s(4) - s(0)} = \frac{\exp\left(\frac{2p_c \beta}{\gamma} \cdot 3\right) - 1}{\exp\left(\frac{2p_c \beta}{\gamma} \cdot 4\right) - 1} \\ &\leq \exp\left(-2p_c \frac{\beta}{\gamma}\right). \end{aligned}$$

□

PROOF OF PROPOSITION 6.4. The hypotheses of the above four Lemmas are satisfied under Assumption 6.2 for small enough  $\varepsilon_0$  and large enough  $\zeta$ . The bounds obtained in all four Lemmas can also be made as small as we like, again by taking  $\varepsilon_0$  small enough and  $\zeta$  large enough. By inserting these bounds into (6.15), we obtain Proposition 6.4. □

6.3. *Proof of extinction for the original equation.* By Proposition 4.3 and Proposition 6.1, in order to show extinction for  $X_t$  defined by the original equation (1.2), it suffices to show that when  $\theta > 0$  is sufficiently small, there exists a scaling as in Lemma 2.27 such that the parameters in the scaled equation satisfy Assumption 6.2. This is the content of the next lemma.

LEMMA 6.10. *For any fixed constants  $0 < \varepsilon_0 < \zeta$ , for all  $\theta > 0$  sufficiently small, there exist a scaling of  $X$ , as in Lemma 2.27 with  $K = 0$ , such that the parameters in the scaled equation satisfy Assumption 6.2.*

PROOF. By Lemma 2.27 we want to find positive constants  $a, b$  and  $c$  such that

$$(6.22) \quad \alpha = ab^2, \quad \varepsilon = a\theta, \quad \beta = \frac{a^2 b^d}{c}, \quad \text{and } \gamma = ac$$

satisfy Assumption 6.2. We will only look at power functions, i.e.,

$$a = \theta^x, \quad b = \theta^y, \quad \text{and } c = \theta^z,$$

and show that for appropriate (real) choices of  $x, y$  and  $z$ , Assumption 6.2 is satisfied provided that  $\theta$  is sufficiently small. We have that

$$\alpha = \theta^{x+2y}, \quad \varepsilon = \theta^{1+x}, \quad \beta = \theta^{2x+dy-z}, \quad \text{and } \gamma = \theta^{x+z}.$$

Looking back at the conditions in Assumption 6.2, we see that it is sufficient that

$$\begin{cases} 1 + x > 2x + dy - z, \\ 1 + x > 0, \quad x + 2y - (x + z) > 0, \quad (1 + x)/2 - (x + z) > 0 \\ (2x + dy - z) - 2(1 + x) < 0, \quad 2(2x + dy - z) - 3(1 + x) - (x + z) < 0, \\ x + z < 0, \quad 2x + dy - z - (x + z) < 0, \end{cases}$$

i.e.,

$$\begin{cases} 1 + z > x + dy, \\ x > -1, \quad 2y > z, \quad 1 - x > 2z \\ dy - z < 2, \quad 2dy < 3z + 3 \\ x + z < 0, \quad x + dy < 2z. \end{cases}$$

There is an abundance of choices, for example,  $x = -3/4$  and  $y = z = 1/2$  will do.  $\square$

**7. A strong form of local extinction.** Theorem 1.3 is a direct consequence of Proposition 4.4 and the following result.

**THEOREM 7.1.** *Assume that  $d = 2$  or  $d = 3$ . If the initial mass distribution  $\mu$  satisfies Assumption 1.1, then for any value of  $\theta$  the epidemic  $X_t$  (the solution to the martingale problem (1.1)) dies out locally, that is, with probability one, for every compact subset  $K \subset \mathbb{R}^d$ ,*

$$(7.1) \quad X_t(K) = 0 \text{ for large enough } t.$$

The remainder of this section will be devoted to the proof of this theorem. Observe at the outset that it suffices to show that the property (7.1) holds when  $K$  is a ball of radius  $\varrho = \varrho(\theta) > 0$  centered at a point with rational coordinates, because any compact  $K$  is covered by finitely many such balls. Moreover, it suffices to consider only balls centered at the origin, because the initial mass distribution  $\mu$  can always be re-centered. Thus, our objective is to prove that the epidemic dies out in  $K = B_\varrho(0)$ .

**7.1. Re-infection at large times .** The proof of Theorem 7.1 will have three parts: First, we will show that (7.1) could fail only if the ball  $B_\varrho(0)$  were re-infected from outside the ball  $B_{3\varrho}(0)$  at indefinitely large times. Second, we will show (in Section 7.2 below) that boundedness of  $EL_\infty(B_{3\varrho}(0))$ , by Proposition 4.1, implies that the mean mass flux through the sphere of radius  $2\varrho$  is finite. Finally, we will show (in Section 7.3) the finite total mean mass flux through the sphere of radius  $2\varrho$  will imply that re-infection of  $B_\varrho(0)$  from outside  $B_{3\varrho}(0)$  at arbitrarily large times cannot occur.

To give precise meaning to the notions of “re-infection from outside” and “mass flux through a boundary” we must bring in the *historical process*  $H$  associated with the spatial epidemic  $X$ . (For a rigorous development of the basic theory, for Dawson-Watanabe processes without interaction, see Dawson and Perkins (1991), for interactive processes including our setting see Perkins (1995), and for an overview of both see Perkins (2002)). Recall that for each time  $t$  the state  $H_t$  is a random measure on the space of continuous paths  $C([0, t], \mathbb{R}^d)$  that projects to  $X_t$  via the time- $t$  evaluation mapping. As in the above references, for  $w \in$

$\mathcal{C} := C([0, \infty), \mathbb{R}^d)$  we set  $w^t(\cdot) = w(\cdot \wedge t)$  and identify  $C([0, t], \mathbb{R}^d)$  with  $\{w \in C([0, \infty), \mathbb{R}^d) : w = w^t\}$ .

Theorem 5.11(a) of Perkins (1995) gives a version of Dawson’s Girsanov Theorem for historical processes. It is then easy to adapt the proof of Theorem 2.2 to see that Theorem 5.11(a) of Perkins (1995) will apply with the drift function  $g$  there equal to  $\theta - L_s^X(w_s)$ . This gives a solution  $H_t$  to a well-posed historical martingale problem so that  $X_t(\varphi) = \int \varphi(w_t)H_t(dw)$  is the unique solution to (1.2). It also shows that the law of  $H$  is absolutely continuous to the law of the historical process associated with super-Brownian motion on the filtration up to time  $t$ , for each  $t > 0$ .

For a fixed  $\varrho > 0$  let

$$\eta(w) = \eta_\varrho(w) = \inf\{t \geq 0 : |w_t| \geq 3\varrho\}$$

be the exit time of the path  $w$  from the interior of  $B_{3\varrho}(0)$ . At time  $t$  color the path  $(w_s)_{s \leq t}$  red if  $\eta \leq t$  and otherwise color it blue. This gives a decomposition

$$(7.2) \quad H_t(\cdot) = H_t^R(\cdot) + H_t^B(\cdot) := H_t(\cdot \cap \{\eta \leq t\}) + H_t(\cdot \cap \{\eta > t\}).$$

Projecting via the time- $t$  evaluation we obtain the decomposition

$$X_t(\cdot) = X_t^R(\cdot) + X_t^B(\cdot) := H_t^R(w_t \in \cdot) + H_t^B(w_t \in \cdot).$$

PROPOSITION 7.2. *For each value  $\theta \in \mathbb{R}$  there exists  $\varrho = \varrho(\theta) > 0$  such that for any initial mass distribution  $\mu$  satisfying Assumption 1.1, the process  $H^B$  in the red/blue decomposition (7.2) will die out with probability one.*

PROOF. Arguing as in Proposition IV.1.4 of Perkins (2002), but using historical processes, one can construct our historical epidemic process  $H$  and the historical process  $\bar{H}$  for a drift- $\theta$  super-Brownian motion,  $\bar{X}$ , on a common probability space so that  $H_0 = \bar{H}_0$  and  $H_t \leq \bar{H}_t$  for all  $t \geq 0$ . We decompose  $\bar{H} = \bar{H}^R + \bar{H}^B$  as in (7.2), thus inducing a corresponding decomposition,  $\bar{X} = \bar{X}^R + \bar{X}^B$ . Then  $\bar{X}^B$  will be the drift- $\theta$  superprocess associated with Brownian motion killed when it exits the interior of  $B_{3\varrho}(0)$ . Therefore if  $Q_x$  is Wiener measure starting at  $x$  and  $\eta_\varrho$  is also the corresponding exit time for the Brownian path, then for  $t > 0$ ,

$$(7.3) \quad E(|\bar{X}_t^B|) = e^{\theta t} \int Q_x(t < \eta_\varrho) d\mu(x) \leq e^{\theta t} |\mu| Q_0(t < \eta_\varrho).$$

(A careful proof of this could use the appropriate version of Proposition 7.4(c) below with  $\eta_\varrho$  in place of  $\tau_k$  and  $\bar{H}$  in place of  $H$ .) For  $\varrho > 0$  sufficiently small (how small will depend on  $\theta$ ) this expectation decays exponentially with  $t$ , by elementary estimates on the transition kernel for killed Brownian motion. (In particular,  $3\varrho > 0$  must be small enough that the first eigenvalue of  $-\Delta/2$  with Dirichlet boundary conditions on  $\partial B_{3\varrho}(0)$  is strictly greater than  $\theta$ .)

It remains to show that the exponential decay of  $E|\bar{X}_t^B|$  implies that  $\bar{X}^B$  dies out almost surely. Let  $Z_t$  denote a Feller branching process with drift  $\theta$ . For  $n \in \mathbb{N}$ , the fact that the

total mass process  $|\overline{X}^B|$  is dominated by the total mass process without killing on  $\partial B_{3\varrho}(0)$  implies

$$(7.4) \quad \begin{aligned} P(|\overline{X}_{n+1}^B| > 0) &\leq E(P_{\overline{X}_n^B}(|\overline{X}_1^B| > 0)) \leq E(P_{|\overline{X}_n^B|}(Z_1 > 0)) = E\left[1 - \exp\left(\frac{-2\theta|\overline{X}_n^B|}{1 - e^{-\theta}}\right)\right] \\ &\leq C(\theta)E(|\overline{X}_n^B|), \end{aligned}$$

where Exercise II.5.3 of Perkins (2002) is used in the next to last line. The exponential decay in the mean on the right-hand side now shows that  $\overline{H}^B$ , and hence the smaller  $H^B$ , dies out a.s. by a Borel-Cantelli argument.  $\square$

For future reference we state a time shifted version of the above. Let  $T > 0$ , define

$$(7.5) \quad \sigma_T = \inf\{t \geq T : |w_t| \geq 3\varrho\},$$

and for  $t \geq T$  set

$$H_t^{B,T}(\cdot) = H_t(\cdot \cap \{\sigma_T > t\}).$$

PROPOSITION 7.3. *For  $\mu$ ,  $\theta$  and  $\varrho(\theta)$  as in Proposition 7.2, the process  $H^{B,T}$  will die out with probability one.*

PROOF. One proceeds just as above but conditional on the past up to  $T$ ,  $\overline{H}_t, t \geq T$  will be the historical process associated with a drift- $\theta$  super-Brownian motion starting at  $\overline{H}_T = H_T$ .  $\square$

Assume for the remainder of the proof that  $\varrho = \varrho(\theta) > 0$  is small enough that the conclusions of Propositions 7.2 and 7.3 hold. Then for any fixed  $T \geq 0$ , all mass in the spatial epidemic  $X_t$  will eventually be descended from the mass in  $X_T$  outside of  $B_{3\varrho}(0)$ . This obviously implies that if local extinction (7.1) fails for  $K = B_\varrho(0)$  then the ball  $B_\varrho(0)$  must be re-infected by mass from outside  $B_{3\varrho}(0)$  at arbitrarily large times.

7.2. *Finite mass flux.* We will control the re-infections of  $B_\varrho(0)$  from outside  $B_{3\varrho}(0)$  by bounding the total ‘‘mass flux’’ (to be made precise below) through  $\partial B_{2\varrho}(0)$ . For any continuous path  $w$  in  $\mathbb{R}^d$  define  $\nu_0 < \tau_1 < \nu_1 < \dots$  to be the successive times of passage between the spheres  $\partial B_{3\varrho}(0)$  and  $\partial B_{2\varrho}(0)$  (that is,  $\nu_0$  is the first hitting time of  $B_{3\varrho}(0)$ ,  $\tau_1$  the first hitting time of  $\partial B_{2\varrho}(0)$  after  $\nu_0$ , and so on). Now for each  $k = 1, 2, \dots$  define  $H_t^k$  to be an associated historical process in which historical mass frozen at time  $\tau_k$  is collected as  $\tau_k$  occurs for  $\tau_k < t$ . For general superprocesses these are the historical random measures constructed by Dynkin (1991) (Theorem 1.5) using log Laplace equations. We will follow Theorem 2.23 and Remark 2.25 of Perkins (1995) which gives a recipe for their construction and associated stochastic analysis, using historical stochastic calculus, and does so in a more general interactive framework which includes our spatial epidemic processes.

$C_b^2(\mathbb{R}^d)$  denotes the space of bounded continuous functions on  $\mathbb{R}^d$  with bounded continuous partials of order 2 or less.

PROPOSITION 7.4. *For each  $k \in \mathbb{N}$ , there is a non-decreasing continuous  $\mathcal{M}(\mathcal{C})$ -valued process,  $H^k$ , and hence an associated random measure on  $[0, \infty) \times \mathcal{C}$  (also denoted by  $H^k$ ), satisfying  $H_0^k = 0$  and the following properties:*

(a)  $w = w^{\tau_k}$ ,  $\tau_k(w) = t$ , and so  $w_t \in \partial B_{2^k}(0)$ , for  $H^k - a.a. (t, w) a.s.$

(b) If  $\psi$  is a bounded measurable function on  $\mathcal{C}$ , then with probability 1 for all  $t \geq 0$ ,

$$\begin{aligned} \int \psi(w^{\tau_k}) \mathbf{1}_{(t > \tau_k)} H_t(dw) &= \int_0^t \int \psi(w^{\tau_k}) \mathbf{1}_{(s > \tau_k)} dM^H(s, w) \\ &\quad + \int_0^t \int \psi(w^{\tau_k}) \mathbf{1}_{(s > \tau_k)} [\theta - L_s^X(w_s)] H_s(dw) ds + H_t^k(\psi), \end{aligned}$$

where  $M^H$  is the orthogonal martingale measure associated with  $H$ .

(c) If  $X_t^k(\cdot) = \int_0^t \int \mathbf{1}_{(w^{\tau_k} \in \cdot)} H^k(ds, dw)$  and  $\varphi \in C_b^2(\mathbb{R}^d)$ , then with probability 1 and for all  $t \geq 0$ ,

$$\begin{aligned} \int \varphi(w_t) \mathbf{1}_{(t > \tau_k)} H_t(dw) &= \int_0^t \int \varphi(w_s) \mathbf{1}_{(s > \tau_k)} dM^H(s, w) \\ &\quad + \int_0^t \int \mathbf{1}_{(s > \tau_k)} \left[ \frac{\Delta \varphi}{2}(w_s) + \varphi(w_s)(\theta - L_s^X(w_s)) \right] H_s(dw) ds \\ &\quad + X_t^k(\varphi). \end{aligned}$$

(d) For any fixed  $t \geq 0$  and bounded Borel  $\psi : \mathcal{C} \rightarrow \mathbb{R}$ , if

$$A_n(t, \psi) = \sum_{i=1}^{\infty} \mathbf{1}_{(i2^{-n} < t)} \int \psi(w^{\tau_k}) \mathbf{1}_{((i-1)2^{-n} \leq \tau_k < i2^{-n})} H_{i2^{-n}}(dw),$$

then  $A_n(t, \psi) \rightarrow H_t^k(\psi)$  in probability as  $n \rightarrow \infty$ . If  $A_n$  also denotes the measure on  $[0, \infty) \times \mathcal{C}$  associated with  $A_n(t, \psi)$ , there is a subsequence  $\{n_j\}$  so that  $A_{n_j}|_{[0, T] \times \mathcal{C}} \rightarrow H^k$  in  $\mathcal{M}([0, T] \times \mathcal{C})$  for all  $T > 0$  a.s.

PROOF. The above result is implicit in Remark 2.25 in Perkins (1995) and carried out for the total mass in Theorem 2.23 of the same reference. We will sketch how the latter construction is easily extended to the measure-valued process  $H^k$ .

Let  $\psi \geq 0$  be a bounded Borel function on  $\mathcal{C}$  and in the setting of Theorem 2.23 in Perkins (1995), set

$$(7.6) \quad C(t, \omega, w) = \psi(w^{\tau_k}) \mathbf{1}_{(t > \tau_k(w))}.$$

The above setting includes our historical epidemic process with the function  $\hat{g}$  on page 9 of this reference equal to  $\theta - L_s^X(\omega, w_s)$  and the integrator  $Z^0$  on page 12 given by

$$dZ^0(s, w) = dM^H(s, w) + \theta H_s(dw) ds - L_s^X(w_s) H_s(dw) ds.$$

Therefore for  $\psi$  fixed, Theorem 2.23 in Perkins (1995) implies (b) and the first conclusion in (d) for some non-decreasing left-continuous process  $H_t^k(\psi)$  satisfying  $H_0^k(\psi) = 0$ . To

derive (c) from (b) (with  $\psi(w) = \varphi(w_{\tau_k})$ ) we need to show that

$$\begin{aligned} \int (\varphi(w_t) - \varphi(w_{\tau_k})) \mathbf{1}_{(t > \tau_k)} H_t(dw) &= \int_0^t \int (\varphi(w_s) - \varphi(w_{\tau_k})) \mathbf{1}_{(s > \tau_k)} dZ^0(s, w) \\ &\quad + \int_0^t \int \frac{\Delta\varphi}{2}(w_s) \mathbf{1}_{(s > \tau_k)} H_s(dw) ds, \end{aligned}$$

and this follows easily from the historical stochastic calculus in Chapter 2 of Perkins (1995).

Consider next the continuity of  $H_t^k(\psi)$  in  $t$  for  $\psi \geq 0$  bounded and Borel. By (IV.48) of Dellacherie and Meyer (1982), it suffices to show that if  $T_n \downarrow T$  are bounded  $(\mathcal{F}_t)$ -stopping times, then

$$\lim_{n \rightarrow \infty} E(H_{T_n}^k(\psi) - H_T^k(\psi)) = 0.$$

Arguing as in (2.44) of Perkins (1995) this reduces to showing

$$(7.7) \quad \lim_{n \rightarrow \infty} E(H_{T_n}(T \leq \tau_k < T_n)) = 0,$$

and

$$(7.8) \quad \lim_{n \rightarrow \infty} E(H_s(\tau_k = T)) = 0 \text{ for each } s > 0.$$

We consider only (7.7) as the proof of (7.8) will then be clear. Using the weak continuity of  $H$  one easily sees that

$$(7.9) \quad \limsup_{n \rightarrow \infty} E(H_{T_n}(T \leq \tau_k < T_n)) \leq E(H_T(|w_T| = 2\varrho, \tau_k \leq T)) \leq E(H_T(|w_T| = 2\varrho) \mathbf{1}_{(0 < T)}),$$

where we used  $\tau_k(w) > 0$ . Theorem III.5.1 of Perkins (2002) and our absolute continuity of  $X$  with respect to super-Brownian motion show that

$$(7.10) \quad P(H_t(|w_t| = 2\varrho) > 0 \text{ for some } t > 0) \leq P(X_t(\partial B_{2\varrho}(0)) > 0 \text{ for some } t > 0) = 0.$$

This implies the right-hand side of (7.9) is zero and so (7.7) is proved, thus giving the continuity of  $H_t^k(\psi)$  for each  $\psi$  as above.

Next we construct  $H^k$  as a measure-valued process. Choose a countable determining class  $\mathcal{D}$  of bounded continuous functions on  $\mathcal{C}$  containing the constant 1. For each  $\psi \in \mathcal{D}$  there is a subsequence  $\{n_j\}$  so that

$$(7.11) \quad \sup_{t \leq T} |A_{n_j}(t, \psi) - H_t^k(\psi)| = 0 \text{ for all } T > 0 \text{ a.s.}$$

This holds by the first part of (d), monotonicity in  $t$  and the a.s. continuity of the limit. By diagonalization we assume the same subsequence works for all  $\psi \in \mathcal{D}$ . It is then easy to check that  $A_{n_j}|_{[0, T] \times \mathcal{C}} \xrightarrow{w} H^k|_{[0, T] \times \mathcal{C}}$  as finite measures on  $[0, T] \times \mathcal{C}$  for all  $T > 0$ . Formally we may use Jakubowski's theorem (Theorem II.4.1 of Perkins (2002)) and the fact that the required compact containment condition follows easily from the modulus of continuity for the historical paths of super-Brownian motion (Theorem III.1.3 of Perkins (2002)) and

the usual absolute continuity argument. Implicit in the above notation is the fact that the limiting random measure  $H^k$  is related to the processes  $H^k(\psi)$  constructed earlier by

$$\int_0^t \int \psi(w) H^k(ds, dw) = H_t^k(\psi) \text{ for all } t \geq 0 \text{ a.s.}$$

This gives the existence of the required process  $H^k$  satisfying properties (b)-(d).

We have  $\tau_k(w) \leq t$ ,  $w = w(\cdot \wedge \tau_k)$ , and so  $w_t \in \partial B_{2\varrho}(0)$  for  $A_n(dt, dw)$ -a.s., and taking weak limits in  $n$  we obtain (a) except with  $\tau_k \leq t$   $H^k$ -a.s. To see that  $\tau_k = t$   $H^k$ -a.s., it suffices to fix  $t \geq \varepsilon > 0$  and show  $H^k((t - \varepsilon, t] \times \{\tau_k < t - \varepsilon\}) = 0$  a.s. This is easily derived from (b) with  $\psi = \mathbf{1}_{(\tau_k < t - \varepsilon)}$  and a bit of historical stochastic calculus.  $\square$

We may repeat the above construction with minor changes for the stopping times  $\nu_k$  in place of  $\tau_k$  and so obtain continuous non-decreasing  $\mathcal{M}(\mathcal{C})$ -valued processes  $\{\hat{H}^k : k \in \mathbb{N}\}$  and their projections  $\{\hat{X}^k : k \in \mathbb{N}\}$  which are  $\mathcal{M}(\mathbb{R}^d)$ -valued processes supported on  $\partial B_{3\varrho}(0)$ . We identify  $X^k$  and  $\hat{X}^k$  with the corresponding random measure on  $[0, \infty) \times \mathbb{R}^d$ .

For future reference we state a truncated version of Proposition 7.4(c). For  $t \geq T > 0$  define

$$\hat{X}^{k,T}(\cdot) = \int_T^t \int \mathbf{1}_{(w_{\nu_k} \in \cdot)} \mathbf{1}_{(T \leq \tau_k)} \hat{H}^k(ds, dw).$$

PROPOSITION 7.5. *If  $T > 0$  and  $\varphi \in C_b^2(\mathbb{R}^d)$ , then with probability 1 for all  $t \geq T$ :*

$$(7.12) \quad \begin{aligned} & \int \varphi(w_t) \mathbf{1}_{(T \leq \tau_k < t)} H_t(dw) \\ &= \int_T^t \int \mathbf{1}_{(T \leq \tau_k < s)} \varphi(w_s) dM^H(s, w) \\ &+ \int_T^t \int \mathbf{1}_{(T \leq \tau_k < s)} \left[ \frac{\Delta \varphi}{2} + \varphi(\theta - L_s^X) \right] (w_s) H_s(dw) ds + (X_t^k - X_T^k)(\varphi), \end{aligned}$$

$$(7.13) \quad \begin{aligned} & \int \varphi(w_t) \mathbf{1}_{(T \leq \tau_k, \nu_k < t)} H_t(dw) \\ &= \int_T^t \int \mathbf{1}_{(T \leq \tau_k, \nu_k < s)} \varphi(w_s) dM^H(s, w) \\ &+ \int_T^t \int \mathbf{1}_{(T \leq \tau_k, \nu_k < s)} \left[ \frac{\Delta \varphi}{2} + \varphi(\theta - L_s^X) \right] (w_s) H_s(dw) ds + \hat{X}_t^{k,T}(\varphi). \end{aligned}$$

PROOF. For (7.12) start with Proposition 7.4(b) with  $\psi(w^{\tau_k}) = \varphi(w_{\tau_k}) \mathbf{1}_{(T \leq \tau_k)}$  and then proceed as in the derivation of (c) above. The fact that  $\tau_k = s$  for  $H^k$ -a.a.  $(s, w)$  is used to get the form of the final term. The proof of (7.13) is similar.  $\square$

The total flux measure on  $[0, \infty) \times \partial B_{2\varrho}(0)$  is  $X^\tau = \sum_{k=1}^\infty X^k$  and similarly we define  $X^\nu = \sum_{k=1}^\infty \hat{X}^k$  on  $[0, \infty) \times B_{3\varrho}(0)$ . At present these measures may be infinite.

As was already noted, our plan is to control the re-infections of  $B_\varrho(0)$  from outside  $B_{3\varrho}(0)$  by bounding the total flux,  $|X^\tau|$ , through  $\partial B_{2\varrho}(0)$ . We next bound this flux in  $L^1$  as a consequence of Proposition 4.1 and Proposition 7.4(c) above.

Color a path  $w$  yellow at time  $t$  iff  $\tau_k < t \leq \nu_k$ , for some  $k \geq 1$ , that is, iff at time  $t$   $w$  is engaged in an excursion from  $\partial B_{2\varrho}(0)$  to  $\partial B_{3\varrho}(0)$ . Let  $H_t^Y$  be the restriction of  $H_t$  to the yellow paths at time  $t$ , that is,  $H_t^Y(A) = \int \mathbf{1}_A(w) \left[ \sum_{k=1}^{\infty} \mathbf{1}_{(\tau_k < t \leq \nu_k)} \right] H_t(dw)$ , and let  $X_t^Y$  be the corresponding time- $t$  projection.

PROPOSITION 7.6.  $E(|X^\tau|) < \infty$  and  $E(|X^\nu|) < \infty$ .

PROOF. We only prove the first conclusion as the proof of the second is similar.

By differencing the decompositions in Proposition 7.4(c) for times  $\tau_k$  and  $\nu_k$  we see that for  $\varphi \in C_b^2(\mathbb{R}^d)$ ,

$$(7.14) \quad \begin{aligned} & \int \varphi(w_t) \mathbf{1}_{(\tau_k < t \leq \nu_k)} dH_t(w) \\ &= \int_0^t \int \varphi(w_s) \mathbf{1}_{(\tau_k < s \leq \nu_k)} dM^H(s, w) \\ &+ \int_0^t \int \mathbf{1}_{(\tau_k < s \leq \nu_k)} \left( \frac{\Delta\varphi}{2}(w_s) + \varphi(w_s)(\theta - L_s^X(w_s)) \right) H_s(dw) ds + X_t^k(\varphi) - \hat{X}_t^k(\varphi). \end{aligned}$$

Let  $0 \leq \varphi_0 \leq 1$  be as above with support in the interior of  $B_{3\varrho}(0)$  and so that  $\varphi_0 = 1$  on  $B_{2\varrho}(0)$ . Then  $\hat{X}_t^k(\varphi_0) = 0$  and  $X_t^k(\varphi_0) = |X_t^k|$  for all  $k, t$ . Take expectations in the above with  $\varphi = \varphi_0$  and then sum over  $k$  to conclude that

$$E(X_t^Y(\varphi_0)) = E\left(\int_0^t \int \frac{\Delta\varphi_0}{2}(w_s) + \varphi_0(w_s)[\theta - L_s^X(w_s)] H_s^Y(dw) ds\right) + E(X^\tau([0, t] \times \mathcal{C})).$$

Rearrange the above and use  $X_t^Y \leq X_t$  and then (1.2) to see that

$$\begin{aligned} E(X^\tau([0, t] \times \mathcal{C})) &\leq E(X_t(\varphi_0)) + E\left(\int_0^t X_s \left( \frac{|\Delta\varphi_0|}{2} + \varphi_0(L_s^X + \theta^-) \right) ds\right) \\ &\leq \mu(\varphi_0) + E(\langle L_t^X, |\Delta\varphi_0| + \varphi_0\theta^+ \rangle). \end{aligned}$$

The right-hand side remains bounded as  $t \rightarrow \infty$  by Proposition 4.1 and so the result follows.  $\square$

7.3. *Local extinction.* Recall that  $\eta(w) = \eta_\varrho(w)$  is the exit time of  $w$  from the interior of  $B_{3\varrho}(0)$ . For any path  $w \in \mathcal{C}$ , if  $\eta \leq t$  and  $|w_t| < 2\varrho$ , then for some  $k \geq 1$ ,  $\tau_k < t \leq \nu_k$ . That is, if you exit from the interior of  $B_{3\varrho}(0)$  before time  $t$  and at time  $t$  are back in the interior of  $B_{2\varrho}(0)$ , then  $t$  must fall in one of the excursions from  $\partial B_{2\varrho}(0)$  to  $\partial B_{3\varrho}(0)$ . Therefore if  $\varphi_1 \in C_c^\infty(\mathbb{R}^d)$ , takes values in  $[0, 1]$ , has support in  $B_{(3/2)\varrho}(0)$ , and  $\varphi_1 = 1$  on  $B_\varrho(0)$ , and

$T > 0$ , then for all  $t \geq T$ ,

$$\begin{aligned} X_t(\varphi_1) &= \int \varphi_1(w_t) \mathbf{1}_{(t < \eta)} H_t(dw) + \sum_{k=1}^{\infty} \int \varphi_1(w_t) \mathbf{1}_{(\tau_k < t \leq \nu_k)} H_t(dw) \\ &= X_t^B(\varphi_1) + \sum_{k=1}^{\infty} \int \varphi_1(w_t) \mathbf{1}_{(\tau_k < T \leq t \leq \nu_k)} H_t(dw) \\ &\quad + \sum_{k=1}^{\infty} \int \varphi_1(w_t) \mathbf{1}_{(T \leq \tau_k < t \leq \nu_k)} H_t(dw) \\ &:= X_t^B(\varphi_1) + \hat{X}_t^{Y,T}(\varphi_1) + X_t^{Y,T}(\varphi_1). \end{aligned}$$

We have decomposed  $X^Y$  according to whether or not the  $k$ th return to  $B_{2\varrho}(0)$  occurs before time  $T$  or after it.

We have already shown (Proposition 7.2) that  $X^B$  dies out a.s. Recall the  $\sigma_T$  defined in (7.5). Clearly  $\tau_k < T \leq t \leq \nu_k$  implies  $\sigma_T > t$  for  $H_t$ -a.a.  $w$  for all  $t \geq T$  a.s. (recall (7.10)), and so by Proposition 7.3,

$$\hat{X}_t^{Y,T}(\varphi_1) \leq \int \varphi_1(w_t) H_t^{B,T}(dw) = 0 \text{ for large } t \text{ a.s. for each } T > 0.$$

Therefore to complete the proof of Theorem 7.1 it suffices to show the following:

PROPOSITION 7.7.  $\lim_{T \rightarrow \infty} P(X_t^{Y,T}(\varphi_1) > 0 \text{ for some } t \geq T) = 0$ .

To prove this result we first recall a standard method used to compute hitting probabilities for a super-Brownian motion  $\bar{X}_t$  with drift  $\theta$ . For  $\lambda > 0$  let  $U(t, x) = U^\lambda(t, x)$  be the unique non-negative solution of

$$(7.15) \quad \frac{\partial U}{\partial t} = \frac{\Delta}{2} U_t + \theta U_t - U_t^2/2 + \lambda \varphi_1, \quad U_0 \equiv 0,$$

which is bounded on  $[0, T] \times \mathbb{R}^d$  for all  $T$  (e.g., see Theorem II.5.11(b) in Perkins (2002)). The duality for superprocesses (e.g., see Theorem II.5.11(c) in Perkins (2002)) implies that for all initial measures  $\nu$ ,

$$E_\nu \left( \exp \left( -\lambda \int_0^t \bar{X}_s(\varphi_1) ds \right) \right) = \exp(-\nu(U_t^\lambda)).$$

It follows that  $U^\lambda(t, x)$  increases as  $\lambda, t \rightarrow \infty$  to a Borel function  $U^\infty(x) \geq 0$  satisfying

$$(7.16) \quad P_\nu(\bar{X}_s(\varphi_1) > 0 \text{ for some } s \geq 0) = 1 - \exp(-\nu(U^\infty)).$$

Next use the fact that  $\bar{X}$  propagates locally at a finite rate (see Theorem III.1.3 of Perkins (2002)) and dies out in small time with high probability if  $|\bar{X}_0|$  is small (recall (7.4)), to see that for  $\varepsilon$  small,

$$\sup_{|x| \geq 2\varrho} P_{\varepsilon \delta_x}(\bar{X}_s(\varphi_1) > 0 \text{ for some } s \geq 0) \leq \frac{1}{2}.$$

It therefore follows from (7.16) that

$$(7.17) \quad \sup_{|x| \geq 2\varrho} U^\infty(x) = C_\varrho < \infty.$$

PROOF OF PROPOSITION 7.7. Fix  $T > 0$ . By differencing the decompositions in Proposition 7.5, we have for  $\varphi \in C_b^2(\mathbb{R}^d)$ , with probability 1 for all  $t \geq T$ ,

$$(7.18) \quad \begin{aligned} & \int \varphi(w_t) \mathbf{1}_{(T \leq \tau_k < t \leq \nu_k)} H_t(dw) \\ &= \int_T^t \int \mathbf{1}_{(T \leq \tau_k < s \leq \nu_k)} \varphi(w_s) dM^H(s, w) \\ &+ \int_T^t \int \mathbf{1}_{(T \leq \tau_k < s \leq \nu_k)} \left[ \frac{\Delta\varphi}{2} + \varphi(\theta - L_s^X) \right] (w_s) H_s(dw) ds + [X_t^k - X_T^k](\varphi) - \hat{X}_t^{k,T}(\varphi). \end{aligned}$$

Fix  $u > T$ . Arguing as in Proposition II.5.7 of Perkins (2002) it is easy to extend (7.18) to time-dependent test functions on  $[0, u] \times \mathbb{R}^d$ , including  $V(t, x) = U^\lambda(u - t, x)$  (see also Theorem II.5.11(b) of Perkins (2002) for the regularity of the above  $V$ ). One gets an additional term involving  $\frac{\partial V}{\partial t}$ , and so with the above choice of  $V$ , the equation (7.15) shows that the function in the square brackets in the second integral in (7.18) becomes

$$\frac{\partial V}{\partial s} + \frac{\Delta V}{2} + \theta V_s - L_s^X V_s = \frac{V_s^2}{2} - \lambda\varphi_1 - L_s^X V_s.$$

Therefore for  $T \leq t \leq u$ ,

$$\begin{aligned} & \int V_t(w_t) \mathbf{1}_{(T \leq \tau_k < t \leq \nu_k)} H_t(dw) \\ &= \int_T^t V_s(w_s) \mathbf{1}_{(T \leq \tau_k < s \leq \nu_k)} dM^H(s, w) + \int_T^t \int \mathbf{1}_{(T \leq \tau_k < s \leq \nu_k)} \\ & \times \left[ \frac{V_s^2}{2} - \lambda\varphi_1 - L_s^X V_s \right] (w_s) H_s(dw) ds + \int_T^t \int V_s(x) [X^k - \hat{X}^{k,T}](ds, dx). \end{aligned}$$

Rearrange the above and sum over  $k$  (using Proposition 7.6) to see that if  $X^{\nu,T} = \sum_{k=1}^\infty \hat{X}_t^{k,T}$ , then for  $T \leq t \leq u$ ,

$$(7.19) \quad \begin{aligned} & X_t^{Y,T}(V_t) + \int_T^t \lambda X_s^{Y,T}(\varphi_1) ds \\ &= M_t^{Y,T}(V) + \int_T^t X_s^{Y,T} \left( \frac{V_s^2}{2} - L_s^X V_s \right) ds + \int_T^t \int V_s(x) [X^\tau(ds, dx) - X^{\nu,T}(ds, dx)], \end{aligned}$$

where  $M_t^{Y,T}(V)$  is a continuous martingale, starting at 0 at time  $T$ , and satisfying  $\langle M^{Y,T}(V) \rangle_t = \int_0^t X_s^{Y,T}(V_s^2) ds$ . Using Proposition 7.6 we see the last term is continuous in  $t$  and it then follows easily that each of the terms in (7.19) is continuous. Now apply

Itô's lemma to  $\exp\left(-X_t^{Y,T}(U_{u-t}^\lambda) - \lambda \int_T^t X_s^{Y,T}(\varphi_1) ds\right)$  and take expectations at  $t = u$ , where  $V_u = U_0 = 0$  and note that  $X_T^{Y,T} \equiv 0$  to deduce that

$$\begin{aligned} & E\left(1 - \exp\left(-\lambda \int_T^u X_s^{Y,T}(\varphi_1) ds\right)\right) \\ &= E\left(\int_T^u \exp\left(-X_t^{Y,T}(U_{u-t}^\lambda) - \lambda \int_T^t X_s^{Y,T}(\varphi_1) ds\right) \right. \\ &\quad \left. \times \int U_{u-t}^\lambda(x) [X^\tau(dt, dx) - X^{\nu,T}(dt, dx) - L_t^X(x) X_t^{Y,T}(dx) dt]\right). \end{aligned}$$

Let  $u, \lambda \uparrow \infty$  and drop the last two negative terms to show that

$$(7.20) \quad P\left(\int_T^\infty X_s^{Y,T}(\varphi_1) ds > 0\right) \leq E\left(\int_T^\infty \int U^\infty(x) X^\tau(dt, dx)\right) \leq C_\rho E(X^\tau([T, \infty) \times \mathbb{R}^d)).$$

The bound (7.17) on  $U^\infty$  for  $|x| \geq 2\rho$  is used in the last inequality. If we sum (7.18) over  $k$  we may argue as in the analysis of (7.19) to see that  $X_t^{Y,T}(\varphi_1)$  is continuous in  $t$ . This and the fact that the upper bound in (7.20) approaches zero as  $T \rightarrow \infty$  by Proposition 7.6 imply the required result.  $\square$

**A. Appendix: Proof of (2.17) for  $d = 2, 3$ .** The main step is to show that for any fixed  $t > 0$ ,

$$\lim_{\varepsilon \rightarrow 0} \limsup_N \max_{x \in \mathbb{Z}^d / \sqrt{N^\alpha \sigma^2}} \frac{\sum_{i=1}^{\lfloor N^\alpha \cdot |\mu| \rfloor} \mathbf{1}_{|x - X_i| \leq \varepsilon} G_{\lfloor N^\alpha t \rfloor}(x \sqrt{N^\alpha \sigma^2} - [X_i \sqrt{N^\alpha \sigma^2}])}{N^{\alpha(2-d/2)}} = 0.$$

The result would then follow easily by using the monotonicity in  $t$ , the SLLN and the local central limit theorems. Using the inequality (19) in Lemma 2 of Lalley and Zheng (2010) one can show that

$$G_{\lfloor N^\alpha t \rfloor}(y) \leq C_1 N^{\alpha(1-d/2)} q^N(y / \sqrt{N^\alpha \sigma^2}) \quad \text{for all } N \text{ and all } y \in \mathbb{Z}^d,$$

where  $q^N(x) = \int_{1/(bN^\alpha)}^{t/b} p_s(x) ds$  for  $x \in \mathbb{R}^d$ , and  $b > 0$  and  $C_1 = C_1(b) > 0$  are both constants. Hence it suffices to show

$$\lim_{\varepsilon \rightarrow 0} \limsup_N \max_{x \in \mathbb{Z}^d / \sqrt{N^\alpha \sigma^2}} \frac{\sum_{i=1}^{\lfloor N^\alpha \cdot |\mu| \rfloor} \mathbf{1}_{|x - X_i| \leq \varepsilon} q^N(x - [X_i \sqrt{N^\alpha \sigma^2}] / \sqrt{N^\alpha \sigma^2})}{N^\alpha} = 0.$$

Let  $h(r) = 1/r$  when  $d = 3$ , and  $h(r) = \log(1/r)$  when  $d = 2$ . Routine calculations show that there exists a constant  $C_2 > 0$  such that for all  $0 \neq |x|$  small and for all  $N$  sufficiently large,

$$(A.1) \quad 1/C_2 h(|x|) \leq q^N(x) \leq C_2 (h(|x|) \wedge h(N^{-\alpha/2})).$$

It follows that there exists  $C_3 > 0$  such that for all  $\varepsilon > 0$  small enough, for all  $N$  large enough, for all  $|z| \leq 2\varepsilon$ ,

$$(A.2) \quad q^N(z) \leq C_3 q^N(v), \quad \text{for all } |v - z| \leq 1/\sqrt{N^\alpha \sigma^2}.$$

Combining this with the bound (2.22) we see that it suffices to show

$$(A.3) \quad \lim_{\varepsilon \rightarrow 0} \limsup_N \sup_{x \in \mathbb{R}^d} Q_\varepsilon^N(x) = 0, \quad \text{where} \quad Q_\varepsilon^N(x) := \frac{\sum_{i=1}^{\lfloor N^\alpha \mu \rfloor} \mathbf{1}_{|x-X_i| \leq \varepsilon} q^N(x-X_i)}{N^\alpha}.$$

Next, for each  $j = 1, \dots, \lfloor N^\alpha \mu \rfloor$ , let

$$\hat{Q}_\varepsilon^N(j) = N^{-\alpha} \sum_{i=1, i \neq j}^{\lfloor N^\alpha \mu \rfloor} \mathbf{1}_{|X_i - X_j| \leq \varepsilon} h(|X_i - X_j|) \wedge h(N^{-\alpha/2}).$$

LEMMA A.1. *There is a  $C_4$  so that for all  $\varepsilon$  small enough and all  $N$  large enough*

$$\sup_{x \in \mathbb{R}^d} Q_\varepsilon^N(x) \leq C_4 \left[ \max_{j \leq \lfloor N^\alpha \mu \rfloor} \hat{Q}_{2\varepsilon}^N(j) + \frac{h(N^{-\alpha/2})}{N^\alpha} \right].$$

PROOF. The upper bound in (A.1) shows that for  $\varepsilon$  small enough and  $N$  large enough (which is assumed in the rest of this proof),

$$Q_\varepsilon^N(x) \leq C_2 N^{-\alpha} \sum_{i=1}^{\lfloor N^\alpha \mu \rfloor} \mathbf{1}_{|x-X_i| \leq \varepsilon} h(|x-X_i|) \wedge h(N^{-\alpha/2}).$$

Fix  $x \in \mathbb{R}^d$  and choose  $j \in \{1, \dots, \lfloor N^\alpha \mu \rfloor\}$  so that  $|X_j - x| = \min_{1 \leq i \leq \lfloor N^\alpha \mu \rfloor} |X_i - x|$ . Then

$$|X_i - X_j| \leq |X_i - x| + |x - X_j| \leq 2|X_i - x|,$$

and so,  $|x - X_i| \leq \varepsilon$  implies  $|X_i - X_j| \leq 2\varepsilon$ , and therefore,

$$(A.4) \quad Q_\varepsilon^N(x) \leq C_2 N^{-\alpha} \sum_{i=1}^{\lfloor N^\alpha \mu \rfloor} \mathbf{1}_{|X_i - X_j| \leq 2\varepsilon} \left[ h(|X_i - X_j|/2) \wedge h(N^{-\alpha/2}) \right].$$

We may assume  $\varepsilon \leq 1/4$ . It follows that in the above summation  $h(|X_i - X_j|/2) \leq 2h(|X_i - X_j|)$  for  $d = 2$  and this is obvious for  $d = 3$ . Therefore by (A.4),

$$Q_\varepsilon^N(x) \leq 2C_2 \left[ \hat{Q}_{2\varepsilon}^N(j) + \frac{h(N^{-\alpha/2})}{N^\alpha} \right],$$

where we have separated out the  $i = j$  term in the summation on the right-hand side of (A.4). The result follows with  $C_4 = 2C_2$  upon taking the max over  $j$  on the right.  $\square$

Therefore to show (A.3), it suffices to establish

$$(A.5) \quad \lim_{\varepsilon \rightarrow 0} \limsup_N \max_{1 \leq j \leq \lfloor N^\alpha \mu \rfloor} \hat{Q}_\varepsilon^N(j) = 0.$$

Let  $C_\mu f_d(r)$  denote the function arising on the right-hand side of (2.20). Let  $r_n = 2^{-n}$  and define

$$M_{n,j} = \#\{i \neq j : |X_i - X_j| \leq r_n\}.$$

If  $K_N = \lfloor N^\alpha |\mu| \rfloor - 1$  and

$$p_n(x) = P(|X_1 - x| \leq r_n) \leq C f_d(r_n)$$

(by (2.20)), then conditional on  $X_j$ ,  $M_{n,j}$  is binomial( $K_N, p_n(X_j)$ ). Therefore a square function inequality for martingales (see Theorem 21.1 in Burkholder (1973)) implies that for any  $q > 0$  there is a  $C'_q$  so that

$$(A.6) \quad E(|M_{n,j} - K_N p_n(X_j)|^q | X_j) \leq C'_q \left( (K_N p_n(X_j)(1-p_n(X_j)))^{q/2} + 1 \right) \leq C_q \left( N^{\alpha q/2} f_d(r_n)^{q/2} + 1 \right).$$

Choose  $n_0$  so that  $r_{n_0} \leq N^{-\alpha/2} < r_{n_0-1}$  and define  $\Lambda_N$  to be the complement of

$$\bigcup_{n=1}^{n_0} \left\{ \max_{j \leq \lfloor N^\alpha |\mu| \rfloor} |M_{n,j} - K_N p_n(X_j)| > N^\alpha f_d(r_n) \right\}.$$

Use (A.6) and Markov's inequality, and then  $f_d(r)^{-1} \leq c(\log(1/r))^3 r^{-1}$  for  $r \in (0, 1/2]$  to see that

$$\begin{aligned} P(\Lambda_N^c) &\leq c \sum_{n=1}^{n_0} N^\alpha |\mu| \cdot N^{-q\alpha} f_d(r_n)^{-q} \left( N^{\alpha q/2} f_d(r_n)^{q/2} + 1 \right) \\ &\leq c N^{\alpha(1-q)} \sum_{n=1}^{n_0} \left( N^{\alpha q/2} f_d(r_n)^{-q/2} + f_d(r_n)^{-q} \right) \\ &\leq c N^{\alpha(1-q)} \sum_{n=1}^{n_0} \left( N^{\alpha q/2} r_n^{-q/2} (\log(1/r_n))^{3q/2} + r_n^{-q} (\log(1/r_n))^{3q} \right). \end{aligned}$$

Recalling the choice of  $n_0$  we can bound the above by

$$\begin{aligned} c N^{\alpha(1-q)} (\log(1/r_{n_0}))^{3q} \left( N^{\alpha q/2} 2^{n_0 q/2} + 2^{n_0 q} \right) &\leq c N^{\alpha(1-q)} (\log N)^{3q} \left( N^{3\alpha q/4} + N^{\alpha q/2} \right) \\ &\leq c N^{\alpha(1-(q/4))} (\log N)^{3q}. \end{aligned}$$

So choose  $q$  large enough so that  $\alpha(1 - (q/4)) < -2$  to conclude that for all  $N$  large enough,

$$(A.7) \quad P(\Lambda_N^c) \leq C N^{-2}.$$

Now for  $\frac{1}{2} > \varepsilon \geq N^{-\alpha/2}$  choose  $n_1 \in \{2, \dots, n_0\}$  so that  $2^{-n_1} \leq \varepsilon < 2^{1-n_1}$ . On  $\Lambda_N$ , for

$1 \leq j \leq [N^\alpha|\mu|]$  we have (use  $h(r)f_d(r) = C_\mu \log(1/r)^{-2}$ )

$$\begin{aligned}
\hat{Q}_\varepsilon^N(j) &= N^{-\alpha} \sum_{i=1, i \neq j}^{[N^\alpha|\mu|]} \mathbf{1}_{|X_i - X_j| \leq \varepsilon} \left( h(|X_i - X_j|) \wedge h(N^{-\alpha/2}) \right) \\
&\leq c \left( \frac{M_{n_0, j}}{N^\alpha} h(N^{-\alpha/2}) + \sum_{n=n_1-1}^{n_0-1} \frac{M_{n, j}}{N^\alpha} h(r_{n+1}) \right) \\
&\leq c \left( \frac{K_N}{N^\alpha} p_{n_0}(X_j) h(N^{-\alpha/2}) + f_d(r_{n_0}) h(N^{-\alpha/2}) \right. \\
&\quad \left. + \sum_{n=n_1-1}^{n_0-1} \left( \frac{K_N}{N^\alpha} p_n(X_j) h(r_n) + f_d(r_n) h(r_n) \right) \right) \\
&\leq c \left( f_d(r_{n_0}) h(r_{n_0}) + \sum_{n=n_1-1}^{n_0-1} f_d(r_n) h(r_n) \right) \\
&\leq c \sum_{n=n_1-1}^{n_0} (\log 1/r_n)^{-2} \\
\text{(A.8)} \quad &\leq cn_1^{-1} \leq c(\log 1/\varepsilon)^{-1}.
\end{aligned}$$

By (A.7), the Borel-Cantelli Lemma and (A.8) we conclude that with probability 1 for all  $N$  large enough we have

$$\max_{1 \leq j \leq [N^\alpha|\mu|]} \hat{Q}_\varepsilon^N(j) \leq c(\log 1/\varepsilon)^{-1} \text{ for } N^{-\alpha/2} \leq \varepsilon < 1/2.$$

This implies (A.5) and we are done.  $\square$

**Acknowledgments.** We are grateful to the Associate Editor and referee for their very careful reading of the paper and constructive suggestions.

## REFERENCES.

- Aldous, D. (1999) ‘‘Brownian excursions, critical random graphs and the multiplicative coalescent’’, *Ann. Probab.*, 25, 812–854.
- Bailey, N.T.J. (1967) ‘‘The simulation of stochastic epidemics in two dimensions’’ in *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1967), Vol. IV: Probability theory*, Berkeley, Calif.: Univ. California Press, pp. 237–257.
- Barlow, M. T., Evans, S. N., and Perkins, E. A. (1991), ‘‘Collision local times and measure-valued processes,’’ *Canad. J. Math.*, 43, 897–938.
- Bramson, M., Durrett, R. and Swindle, G. (1989), ‘‘Statistical Mechanics of Crabgrass’’ *Ann. Probab.*, 17, 444–481.
- Burkholder, D. L., ‘‘Distribution function inequalities for martingales’’ *Ann. Probab.*, 1, 19–42.
- Cox, J.T., and Durrett, R. (1988), ‘‘Limit theorems for the spread of epidemics and forest fires’’ *Stoch. Proc. Appl.*, 30, 171–191.
- Daley, D. J. and Gani, J. (1999), ‘‘Epidemic Modelling’’. Cambridge U. Press.
- Dawson, D. A. and Perkins, E. A. (1991), ‘‘Historical processes,’’ *Mem. Amer. Math. Soc.*, 93, iv+179.
- Dellacherie, C. and Meyer, P.A. (1982), ‘‘Probabilities and Potential B’’, *North Holland Mathematical Studies* 72, North Holland, Amsterdam.
- Dolgoarshinnykh, R. G. and Lalley, Steven P. (2006) ‘‘Critical scaling for the SIS stochastic epidemic’’, *J. Appl. Probab.* 43, 892–898.

- Durrett, R. (1995), “Ten lectures on particle systems,” in *Lectures on probability theory (Saint-Flour, 1993)*, Berlin: Springer, vol. 1608 of *Lecture Notes in Math.*, pp. 97–201.
- Durrett, R. and Perkins, E.A. (1999), “Rescaled contact processes converge to super-Browian motion in two or more dimensions,” *Probab. Th. Rel. Fields*, 114, 309–399.
- Dynkin, E. B. (1991), “Branching particle systems and superprocesses,” *Ann. Probab.*, 19, 1157–1194.
- Evans, S. N. and Perkins, E. (1991), “Absolute continuity results for superprocesses with some applications,” *Trans. Amer. Math. Soc.*, 325, 661–681.
- Feller, W. (1951), “Diffusion processes in genetics,” in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, Berkeley and Los Angeles: University of California Press, pp. 227–246.
- Garsia, A. M. (1972), “Continuity properties of Gaussian processes with multidimensional time parameter,” in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, Berkeley, Calif.: Univ. California Press, pp. 369–374.
- Harris, T.E. (1974), “Contact interactions on a lattice,” *Ann. Probab.*, 2, 969–988.
- Iscio, I. (1988), “On the supports of measure-valued critical branching Brownian motion,” *Ann. Probab.*, 16, 200–221.
- Kermack, W. and McKendrick, A. (1927), “A Contribution to the Mathematical Theory of Epidemics,” *Proc. Roy. Soc. London A*, 115, 700–721.
- Lalley, S. P. (2009), “Spatial epidemics: critical behavior in one dimension,” *Probab. Theory Related Fields*, 144, 429–469.
- Lalley, S. P. and Zheng, X. (2010), “Spatial epidemics and local times for critical branching random walks in dimensions 2 and 3,” *Probab. Theory Related Fields*, 148, 527–566.
- Martin-Löf, A. (1998) The final size of a nearly critical epidemic, and the first passage time of a Wiener process to a parabolic barrier. *J. Appl. Probab.*, 35, 671–682.
- McKendrick, A. G. (1926) “Applications of mathematics to medical problems”, *Proc. Edinburgh Math. Soc.*, 14, 98–130.
- Mollison, D. (1977) “Spatial contact models for ecological and epidemic spread” *J. Roy. Statist. Soc. B*, 39, 283–326.
- Mueller, C. and Tribe, R. (1994), “A phase transition for a stochastic PDE related to the contact process,” *Probab. Theory Related Fields*, 100, 131–156.
- (2011), “A phase diagram for a stochastic reaction diffusion system,” *Probab. Theory Related Fields*, 149, 561–637.
- Müller, C. and Tribe, R. (1995), “Stochastic p.d.e.’s arising from the long range contact and long range voter processes,” *Probab. Theory Related Fields*, 102, 519–545.
- Penrose, M.D. (1993), “On the spread-out limit for bond and continuum percolation,” *Ann. Appl. Probab.*, 3, 253–276.
- Perkins, E. (1995), “On the martingale problem for interactive measure-valued branching diffusions,” *Mem. Amer. Math. Soc.*, 115, vi+89.
- (2002), “Dawson-Watanabe superprocesses and measure-valued diffusions,” in *Lectures on probability theory and statistics (Saint-Flour, 1999)*, Berlin: Springer, vol. 1781 of *Lecture Notes in Math.*, pp. 125–324.
- Pinsky, R. G. (1995), “On the large time growth rate of the support of supercritical super-Browian motion,” *Ann. Probab.*, 23, 1748–1754.
- Revuz, D. and Yor, M. (1999), *Continuous martingales and Brownian motion*, vol. 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, Berlin: Springer-Verlag, 3rd ed.
- Sugitani, S. (1989), “Some properties for the measure-valued branching diffusion processes,” *J. Math. Soc. Japan*, 41, 437–462.
- Tornatore, E., Buccellato, S. M., and Vetro, P. (2006), “stochastic disease model with vaccination,” *Rendiconti del Circolo Matematico di Palermo*, 55, 223–240.
- Walsh, J. B. (1986), “An introduction to stochastic partial differential equations,” in *École d’été de probabilités de Saint-Flour, XIV—1984*, Berlin: Springer, vol. 1180 of *Lecture Notes in Math.*, pp. 265–439.

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

DEPARTMENT OF MATHEMATICS,  
UNIVERSITY OF BRITISH COLUMBIA,  
VANCOUVER, BC V6T 1Z2, CANADA.  
E-MAIL: perkins@math.ubc.ca

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