SOME PECULIAR PROPERTIES OF A PARTICLE SYSTEM WITH SEXUAL REPRODUCTION

Richard Durrett, U.C.L.A. and Lawrence Gray, Univ. of Minnesota

Abstract. In this paper we study a growth model on $\mathbb{Z}^2$ in which particles reproduce sexually: in order for a new particle to appear at a vacant site, two neighboring sites must be occupied. The fact that the birth mechanism requires two parent particles causes the example to exhibit a number of phenomena which are unheard of in models with asexual reproduction (i.e. the "additive processes" of Harris (1974)). For example, if we perturb the system by adding spontaneous births at a small rate, then for certain parameter values we obtain a process with "positive rates" which has two stationary distributions.

§1. Introduction. In this paper we investigate a growth model on $\mathbb{Z}^2$ which has sexual reproduction and contrast its behavior with related models which have asexual reproduction. The first step will be to give an informal description of the models under consideration. A more precise definition of these processes can be found in Section 2.

In each model the state of the system at time t is ξ_t, a subset of $\mathbb{Z}^2$. We interpret ξ_t as the set of sites occupied by particles at time t. Particles *die* (or in other words, occupied sites are vacated) according to the simplest possible rule:

(i) if $x \in \xi_t$, then $\xi_t \to \xi_t \setminus \{x\}$ at rate 1.

The *birth* mechanism (by which vacant sites become occupied) is determined by a set of birth rates b_x, $x \in \mathbb{Z}^2$, which depend on the state of the system:

(ii) if $x \notin \xi_t$, then $\xi_t \to \xi_t \cup \{x\}$ at rate $b_x(\xi_t)$,

where $b_x(\xi_t)$ depends only on $|\xi_t \cap \{x + e_1, x + e_2\}|$. (We use $|\cdot|$ to denote cardinality and e_1 and e_2 for the natural unit basis vectors.) We are mainly interested in two choices of the birth rates b_x:

Example 1. Sexual reproduction.

$$b_x(\xi) = \lambda \quad \text{if } |\xi \cap \{x + e_1, x + e_2\}| = 2$$
$$= 0 \quad \text{otherwise.}$$

Research supported in part by N.S.F Grant DMS 83-01080

Example 2. Asexual reproduction.

$$b_x(\xi) = \lambda \quad \text{if } |\xi \cap \{x + e_1, x + e_2\}| = 1 \text{ or } 2$$
$$= 0 \quad \text{otherwise.}$$

These examples can be thought of as two different ways to generalize to two dimensions the so-called "one-sided contact process on $\mathbb{Z}$", which is one of the most studied of all interacting particle systems. In the contact process, a birth can occur at $x \in \mathbb{Z}$ only if $x + 1$ is occupied. In our Example 1, a birth can occur at $x \in \mathbb{Z}^2$ only if $x + e_1$ *and* $x + e_2$ are occupied, while in Example 2, the condition is that $x + e_1$ *or* $x + e_2$ be occupied. It turns out that "or" is easier to analyze than "and". The reason for this is that Example 2 is additive in the sense of Harris (1978) while Example 1 is not. Additive processes have associated with them certain "dual processes" which greatly facilitate analysis of their behavior (see Griffeath (1979) for a thorough exposition). Dual processes for non-additive models have only recently been defined (Gray (1986)) and are more difficult to manage. The contrast will become clear when we describe the dual processes for Examples 1 and 2 in Section 2.

In order to set the stage for our results, we need to recall some general facts which hold for a certain class of interacting particle systems. A system is called *attractive* if the birth and death rates b_x and d_x satisfy

$$b_x(\xi) \geq b_x(\eta) \quad \text{and} \quad d_x(\xi) \leq d_x(\eta) \quad \text{whenever} \quad \eta \subset \xi \subset \mathbb{Z}^2 .$$

In our examples, the death rates are identically 1 and the birth rates b_x are non-decreasing functions of the number of occupied sites in the set $\{x + e_1, x + e_2\}$, so the above condition is satisfied. As first observed by Holley (1972), attractive systems have certain useful monotonicity properties. Let ξ_t^0 and ξ_t^1 denote the state at time t when the initial states are $\emptyset$ and $\mathbb{Z}^2$ respectively. Then for all $A \subset \mathbb{Z}^2$,

$$P(\xi_t^0 \cap A \neq \emptyset) \text{ increases and } P(\xi_t^1 \cap A \neq \emptyset) \text{ decreases as } t \to \infty.$$

Thus ξ_t^0 and ξ_t^1 converge weakly as $t \to \infty$ to limits which we call ξ_∞^0 and ξ_∞^1, and these limits are stationary distributions of the process. We will often use the symbols " ξ_∞^0 " and " ξ_∞^1 " for random variables which have these distributions instead of for the distributions themselves.

In Examples 1 and 2, $\xi_t^0 = \emptyset$ for all t, so ξ_∞^0 is trivial and all the attention is focussed on ξ_∞^1 . Let

$$\rho(\lambda) = \lim_{t\to\infty} P(0 \in \xi_t^1) = P(0 \in \xi_\infty^1).$$

If $\rho(\lambda) = 0$, then $\xi_\infty^1 = \xi_\infty^0$, and it is not hard to show that for all initial configurations, $\xi(t) \Rightarrow \delta_\emptyset$ as $t \to \infty$. (The notation means "$\xi(t)$ converges weakly to the point mass at $\emptyset$". In this context, weak convergence is equivalent to the convergence of the finite dimensional distributions. In other words, it means convergence of the distribution of $\xi(t) \cap \{x\}$ for each x.) On the other hand, if $\rho(\lambda) > 0$, then ξ_∞^1 is non-trivial and distinct from ξ_∞^0.

Let $\lambda_c = \inf\{\lambda\colon \rho(\lambda) > 0\}$. According to the preceding paragraph, $\xi_\infty^1 = \xi_\infty^0 = \delta_\emptyset$ if $\lambda < \lambda_c$, and $\xi_\infty^1 \neq \xi_\infty^0$ if $\lambda > \lambda_c$. Thus, one of the most basic questions to ask concerning a given system is whether or not $0 < \lambda_c < \infty$. The answer to this question is "yes" for both of our examples.

The proof that $\lambda_c > 0$ is easy in both cases. If we think of the particles as potential parents of new particles, then the birth rate per particle is bounded by λ in Example 1 and by 2λ in Example 2. Thus if $\lambda < 1$ in the first example or if $\lambda < \frac{1}{2}$ in the second, the population dies off at a faster rate on the average than it replenishes itself, and it is easy to see that $\xi_\infty^1 \Rightarrow \delta_\emptyset$.

It is considerably harder to show that $\lambda_c < \infty$. For Example 2, we can reduce this to a known result by observing that if the process is projected onto the line $y = -x$, then it dominates the so-called two-sided contact process on $\mathbb{Z}$. It is known (Holley and Liggett (1978)) that $\lambda_c \le 4$ in the two-sided contact process on $\mathbb{Z}$, so $\lambda_c \le 4 < \infty$ for Example 2.

The fact that $\lambda_c < \infty$ in Example 1 follows from our first main result:

Theorem 1. In the sexual contact process (Example 1), $\rho(\lambda) \to 1$ as $\lambda \to \infty$.

Our proof uses what is commonly known as a contour argument. This method was developed for use in continuous time processes by Gray and Griffeath (1982) and Bramson and Gray (1984). (In other contexts, the technique can be traced back to the so-called "Peierls' argument" in Peierls (1936). For discrete time processes, Toom (1979) is most responsible for its application and refinement.) We will follow the basic pattern set in Gray and Griffeath (1982). We will define a contour in such a way that

$$P(0 \notin \xi_\infty^1) \le P(\text{at least one contour exists}).$$

We will estimate the right side by using the expected number of contours, which will in turn be bounded by a series expansion which converges for sufficiently large λ. The sum of the series will go to 0 as $\lambda \to \infty$, proving Theorem 1.

As is usual for contour methods, an upper bound for λ_c emerges from the computations, but this bound is typically crude. In our case we will be able to conclude that

$$\lambda_c \le 4(3 + \sqrt{5})^2 - 2 < 108$$

in Example 1, whereas computer simulations (done by Tom Liggett) suggest that $10 < \lambda_c < 20$. For comparison, note that for the one-sided contact process, the estimate obtained in Gray and Griffeath (1982) is $\lambda_c < 14$, while the actual value is believed to be a little larger than 3.

Once it has been shown that $0 < \lambda_c < \infty$, it is natural to ask how ξ_t behaves as $t \to \infty$ starting from simple initial distributions other than the ones concentrated at $\emptyset$ or $\mathbb{Z}^2$. Let ξ_t^p be the state at time t when the initial distribution is product measure with parameter p (i.e., the events $\{x \in \xi_0^p\}$, $x \in \mathbb{Z}^2$, are independent and have probability p). Using duality techniques it can be shown that Example 2 (like all additive models) satisfies

(1) if $p > 0$, then $\xi_t^p \Rightarrow \xi_\infty^1$ as $t \to \infty$.

We will indicate the easy proof at the end of Section 2.

Somewhat surprisingly, (1) is false for Example 1. It is known from the theory of oriented percolation (see Durrett (1984)) that there exists a p^* strictly between 0 and 1 such that for all $p < p^*$ there is a.s. a sequence of vacant sites $\{x_k\}$ in the initial configuration ξ_0^p such that $x_{k+1} =$
$x_k + e_1$ or $x_k + e_2$ for all k. It is easy to see from the description of the birth mechanism that these sites can never become occupied and hence that a birth can never occur at any of the sites $x_k - e_1$ or $x_k - e_2$. Therefore, after any particles present at these two sites die, they will be forever vacant. Iterating this argument we find that any site of the form $x_k - n_1 e_1 - n_2 e_2$, $n_1, n_2 \ge 0$, will eventually remain unoccupied, and it is not difficult to conclude from translation invariance that for Example 1,

(2) if $p < p^*$, then $\xi_t^p \Rightarrow \delta_\emptyset$ as $t \to \infty$.

If we define the critical probability $p_c(\lambda) = \sup\{p: \xi_t^p \Rightarrow \delta_\emptyset\}$, then (2) implies that $p_c(\lambda) \geq p^*$ for all λ. We conjecture that

(a) $p_c(\lambda) \searrow p^*$ as $\lambda \nearrow \infty$; and

(b) if $p > p_c(\lambda)$ then $\xi_t^p \Rightarrow \xi_\infty^1$ as $t \to \infty$

but we have no idea how to prove either of these results.

Another interesting question concerns the behavior of the system when $\lambda = \lambda_c$. If we examine the proof given above that (1) does not hold in Example 1, we see that by the same argument, $\xi_t \Rightarrow \delta_\emptyset$ for any initial distribution in which the vacant sites percolate in the sense of oriented percolation. (We would like to thank Michael Aizenman for this observation.) This suggests the following:

Conjecture. In Example 1, there exists $\varepsilon^* > 0$ such that if $\rho(\lambda) \equiv P(0 \in \xi_\infty^1) < \varepsilon^*$, then $\rho(\lambda) = 0$.

Since $\rho(\lambda)$ is known to be right continuous (a simple compactness argument shows that ξ_∞^1 is weakly right continuous in λ), it would follow from the conjecture that

$$\rho(\lambda) \to \rho(\lambda_c) > 0 \text{ as } \lambda \searrow \lambda_c .$$

We have been able to prove this conjecture when $\mathbb{Z}^2$ is replaced by the binary tree $\mathcal{T}$. We label the sites of $\mathcal{T}$ with finite strings of 0's and 1's which begin with a 0 (the "root" is labelled with a 0, the two nodes next to the root are labelled 00 and 01, and so on, with n + 1 binary digits required to label the nodes at level n above the root). The process is defined as in Example 1 with a constant death rate equal to 1 and a birth rate at the site x which depends on the sites x1 and x0 that neighbor x at the next higher level: the birth rate at x is λ if both x0 and x1 are occupied, and the rate is 0 otherwise. Our proof of the conjecture in this case is simple and short and doesn't fit in anywhere else, so we give it here.

Let ξ_t^μ denote the state at time t if the initial distribution is $\mu = \xi_\infty^1$. Then

$$\frac{d\, P(0 \in \xi_t^\mu)}{dt} = -P(0 \in \xi_t^\mu) + \lambda P(0 \notin \xi_t^\mu \text{ and } 00, 01 \in \xi_t^\mu).$$

Since ξ_∞^1 is stationary, $\rho(\lambda) = P(0 \in \xi_t^\mu)$, so it follows that

$$0 = -\rho(\lambda) + \lambda P(0 \notin \xi_t^{\mu} \text{ and } 00, 01 \in \xi_t^{\mu}).$$

By a result of Harris (1977), the events $\{0 \notin \xi_t^{\mu}\}$ and $\{00, 01 \in \xi_t^{\mu}\}$ are negatively correlated. Furthermore it is easy to see that the behavior of the process at the site 00 is independent of the behavior at 01 and that $P(00 \in \xi_t^{\mu}) = P(01 \in \xi_t^{\mu}) = P(0 \in \xi_t^{\mu}) = \rho(\lambda)$. Therefore,

$$\begin{aligned}(3)\quad 0 &\le -\rho(\lambda) + \lambda P(0 \notin \xi_t^{\mu}) P(00, 01 \in \xi_t^{\mu}) \\ &= -\rho(\lambda) + \lambda P(0 \notin \xi_t^{\mu}) P(00 \in \xi_t^{\mu}) P(01 \in \xi_t^{\mu}) \\ &= -\rho(\lambda)[1 - \lambda\rho(\lambda)(1 - \rho(\lambda))].\end{aligned}$$

If $\lambda < 4$, then $1 - \lambda\rho(\lambda)(1 - \rho(\lambda)) > 0$, so it follows that $\rho(\lambda) = 0$ and $\xi_\infty^1 = \delta_\emptyset$, or in other words, $\lambda_c \ge 4 > 0$. Now assume that $\infty > \lambda > \lambda_c$ (the proof that $\lambda_c < \infty$ is similar to the proof of Theorem 1). Then $\rho(\lambda) > 0$, so (3) implies that $1 \le \lambda\rho(\lambda)(1 - \rho(\lambda))$, from which it follows by right continuity that $1 \le \lambda_c\rho(\lambda_c)(1 - \rho(\lambda_c))$. This last inequality implies that $\rho(\lambda_c) > 0$. Since $\rho(\lambda)$ is non-decreasing, the statement in the conjecture follows, with $\varepsilon^* = \rho(\lambda_c)$. □

For small p, we can think of product measure as a perturbation of the absorbing state $\emptyset$. Statements (1) and (2) imply that under such a perturbation, the equilibrium $\delta_\emptyset$ is an "attracting fixed point" in the sexual contact process in Example 1 and an "unstable fixed point" in the asexual contact process in Example 2. We may expect similar statements to hold under another kind of perturbation, namely that of adding a small quantity $\beta > 0$ to all the birth rates ("spontaneous births at rate β"). An argument similar to the one used for (1) shows this to be true for Example 2. In fact, if we let $\xi^{0,\beta}$ and $\xi^{1,\beta}$ be the states at time t for the system with spontaneous births at rate β and initial states $\emptyset$ and $\mathbb{Z}^2$ respectively, and if we let $\xi_\infty^{0,\beta}$ and $\xi_\infty^{1,\beta}$ be the weak limits of $\xi_t^{0,\beta}$ and $\xi_t^{1,\beta}$ as $t \to \infty$, then in Example 2,

(4a) $\xi_\infty^{0,\beta} = \xi_\infty^{1,\beta}$ is the unique stationary distribution for all $\beta > 0$.

There is a strong sense in which $\xi_\infty^{1,\beta}$ is "larger" than ξ_∞^{1} for all $\beta > 0$ (they can be defined jointly so that $\xi_\infty^{1,\beta} \supset \xi_\infty^{1}$), so it follows from (4a) and a compactness argument that

(4b) as $\beta \searrow 0$, $\xi_\infty^{0,\beta} \Rightarrow \xi_\infty^1$.

(See Holley (1972) or Liggett (1985), Chapter 3, for the missing details, particularly concerning the meaning of the word "larger".)

For $\lambda > \lambda_c$, $\delta_\emptyset = \xi_\infty^0 \neq \xi_\infty^1$, so (4b) says that $\delta_\emptyset$ is unstable under this type of perturbation in Example 2. In contrast, $\delta_\emptyset$ is stable in Example 1 under the same type of perturbation:

Theorem 2. For any $\lambda > 0$ in the sexual contact process (Example 1), $\xi_\infty^{0,\beta} \Rightarrow \delta_\emptyset$ as $\beta \searrow 0$.

Since $\xi_\infty^{1,\beta} \Rightarrow \xi_\infty^1$ as $\beta \searrow 0$, it follows from Theorem 2 that for $\lambda > \lambda_c$ and β sufficiently small, $\xi_\infty^{0,\beta} \neq \xi_\infty^{1,\beta}$. Our proof of Theorem 2 actually yields an estimate for how small "sufficiently small" is:

Corollary 1. If $\lambda > \lambda_c$ and $6\beta^{1/4}\lambda^{3/4} < 1$, then the sexual contact process with parameter λ and spontaneous births at rate β has $\xi_\infty^{0,\beta} \neq \xi_\infty^{1,\beta}$.

The reader should note that the lower bound on the allowable perturbation does not go to 0 as $\lambda \to \lambda_c$. We think this is more evidence for the conjecture made above.

Systems like the one in Corollary 1 which have strictly positive translation invariant birth and death rates with finite range interaction, and which are known to be non-ergodic (i.e. to have two distinct stationary distributions) are relatively rare. One well-known set of examples in provided by the two dimensional stochastic Ising models. It is instructive to compare their behavior under perturbation with that of our Example 1. Consider birth and death rates defined as follows:

if $x \notin \xi_t$ then $\xi_t \to \xi_t \cup \{x\}$ at rate $\exp(2h + (k - 2)\beta)$,
if $x \in \xi_t$ then $\xi_t \to \xi_t \setminus \{x\}$ at rate 1

where $\beta > 0$, $h \in \mathbb{R}$ and

$$k = |\xi_t \cap \{x + e_1, x + e_2, x - e_1, x - e_2\}|.$$

Let $\xi_\infty^{0,\beta,h}$ and $\xi_\infty^{1,\beta,h}$ be the limiting (stationary) distributions as $t \to \infty$ starting from the initial states $\delta_\emptyset$ and $\mathbb{Z}^2$ respectively, and define

$$m^0(\beta, h) = P(0 \in \xi_\infty^{0,\beta,h}) \text{ and } m^1(\beta, h) = P(0 \in \xi_\infty^{1,\beta,h}).$$

It is known (see Durrett (1981) and the references therein) that there exists $\beta_c > 0$ such that

(5a) $\xi_\infty^{0,\beta,h} \neq \xi_\infty^{1,\beta,h}$ iff $h = 0$ and $\beta > \beta_c$;

(5b) if $\beta > \beta_c$, then $1/2 < m^1(\beta, 0) = 1 - m^0(\beta, 0)$;

(5c) $m^i(\beta, h) \to m^1(\beta, 0)$, $i = 0$ or 1, as $h \searrow 0$, and

$m^i(\beta, h) \to m^0(\beta, h)$, $i = 0$ or 1, as $h \nearrow 0$.

The reader should especially note (5a) and (5c), which say that non-ergodicity is very unstable in the Ising model: one of the two stationary distributions disappears when h is perturbed away from 0. In contrast, the non-ergodicity of Example 1 is stable under arbitrary types of perturbations, as long as they are finite range in nature. (The finite range condition ensures the existence and uniqueness of the process.) This fact is a second consequence of Theorem 2:

Corollary 2. Let (c_x) be an arbitrary collection of finite range birth rates with $0 \le c_x \le 1$. If $\lambda > \lambda_c$ and $\beta > 0$ is so small that $P(0 \in \xi_\infty^{0,\beta}) < P(0 \in \xi_\infty^1)$, then adding βc_x to the birth mechanism in Example 1 preserves the non-ergodicity of the system.

Proof. Let $(\tilde{\xi}_t)$ stand for the perturbed system, with $\tilde{\xi}_t^0$ and $\tilde{\xi}_t^1$ representing the states at time t if the initial states are $\delta_\emptyset$ and $\mathbb{Z}^2$. Then in the same sense described immediately following (4a), the perturbed system $(\tilde{\xi}_t)$ is "larger" than the unperturbed system (ξ_t) and "smaller" than the system in which the birth rates are perturbed by adding β. Thus if β and λ satisfy the hypotheses,

$$\begin{aligned}\overline{\lim_{t\to\infty}} P(0 \in \tilde{\xi}_t^0) &\le \overline{\lim_{t\to\infty}} P(0 \in \xi_t^{0,\beta}) \\ &= P(0 \in \xi_\infty^{0,\beta}) \\ &< P(0 \in \xi_\infty^1)\end{aligned}$$

$$= \lim_{t\to\infty} P(0 \in \xi_t^1)$$

$$\le \underline{\lim}_{t\to\infty} P(0 \in \tilde{\xi}_t^1).$$

A standard compactness argument completes the construction of two distinct stationary distributions for (ξ_t). □

We hope that the results above have convinced the reader that "sex makes life interesting." We offer one final piece of evidence for this viewpoint by describing some unusual properties of an example in which the birth rate has both sexual and asexual components:

Example 3. The death rate is still 1. The birth rate b_x is defined as:

$$\begin{aligned} b_x(\xi) &= \lambda(1+\alpha) \text{ if } |\xi \cap \{x+e_1, x+e_2\}| = 2 \\ &= \lambda\alpha \text{ if } |\xi \cap \{x+e_1, x+e_2\}| = 1 \\ &= 0 \quad \text{otherwise,} \end{aligned}$$

where $\alpha > 0$. (The case $\alpha = 0$ is Example 1.)

Let ξ_t^* be the state at time t if the initial state is the singleton {0}. Since this initial state has only finitely many occupied sites, it is possible for the process to reach the state $\emptyset$, in which case we say it "dies out". If $\xi_t^* \ne \emptyset$ for all $t \ge 0$, then we say it "survives for all t". Let Ω_∞ be the event that ξ_t^* survives for all t, and let $\lambda_f = \inf\{\lambda: P(\Omega_\infty) > 0\}$. (In Example 3, the value of λ_f depends of course on α.) At one time it was tempting to conjecture that $\lambda_f = \lambda_c$ for "reasonable" growth models. In many examples this is known to be true (e.g., all one-dimensional attractive nearest neighbor systems and many "nearest particle" systems -- see Durrett and Griffeath (1983), Gray (1985) and Liggett (1983)) -- and it is strongly believed to hold for many other systems such as our Example 2. Thus it is a little surprising to realize that the process in Example 1 goes extinct a.s. starting from *any* finite initial state, so that $\lambda_f = \infty$ in Example 1 (proof: observe that if $\xi_0 \subset [-N,N]^2$, then $\xi_t \subset [-N,N]^2$ for all t. It is easy to see that since the death rate is 1, the easy half of Borel-Cantelli implies that sooner or later, all the sites in $[-N,N]^2$ will be vacant and the process will die out. In fact for any finite initial state there is an $\varepsilon > 0$ such that $P(\xi_t \ne \emptyset) \le (1-\varepsilon)^t \to 0$ as $t \to \infty$.)

Since λ_f is infinite in Example 1, it is tempting to weaken the conjecture "$\lambda_f = \lambda_c$" to "$\lambda_f = \lambda_c$ whenever both are finite." But Example 3 provides a counterexample even to this modified version, as shown by our third main result and its corollaries:

Theorem 3. In Example 3, suppose $\lambda \geq 1$ and $\alpha < 1/(144\lambda)$. If $\varepsilon \leq \alpha^3$ and $p < \varepsilon/(1 + \varepsilon)$, then $P(0 \in \xi_t^p) \to 0$ exponentially fast as $t \to \infty$. (See above for the definition of ξ_t^p.)

Consequently, if we let ξ_t^A be the state at time t when the initial state is the set A, we have

Corollary 3. In Example 3, if $\lambda \geq 1$ and $\alpha < 1/(144\lambda)$, then for any finite A,

$P(\xi_t^A \neq \emptyset) \to 0$ exponentially fast as $t \to \infty$.

Proof. If the initial measure were product measure with density p, then the probability that A is occupied at time 0 would be $p^{|A|}$, so $P(\xi_0^A \subset \xi_0^p) > 0$. It follows from the Markov property and Corollary 2 that

$P(x \in \xi_t^A) \to 0$ exponentially fast as $t \to \infty$, uniformly in $x \in \mathbb{Z}^2$.

It follows from this that for any constant $c > 0$,

$P(\xi_t^A \cap [-ct, ct]^2 \neq \emptyset) \to 0$ exponentially fast as $t \to \infty$.

Standard arguments concerning the rate of spread of the set of occupied sites in population models (based on the simplest large deviations for random walks) imply that c can be chosen (depending on λ) so that

$P(\xi_t^A \cap ([-ct, ct]^2)^c \neq \emptyset) \to 0$ exponentially fast as $t \to \infty$,

and Corollary 3 follows. □

Comparison with the two-sided contact process (as done for Example 2) reveals that for Example 3, $\lambda_f \leq 4/\alpha$, since it is known for the 2-sided contact process that $\lambda_f = \lambda_o \leq 4$. Combining this with Corollary 3 gives:

Corollary 4. If $\alpha < 1/144$ then $1/(144\alpha) \leq \lambda_f(\alpha) \leq 4/\alpha$.

The constants in these inequalities are crude, but we do see that $\lambda_f(\alpha) = O(\alpha^{-1})$, which is more than enough information to conclude that since $\lambda_c(\alpha)$ is bounded above by the critical value of Example 1, $0 < \lambda_c(\alpha) <$

$\lambda_f(\alpha) < \infty$ in Example 3 for sufficiently small $\alpha > 0$. In other words, the system has two phase transitions.

The rest of the paper is devoted to proving the statements made above. In Section 2 we describe the objects which are key to the proofs, namely dual processes. The recipe that we give for constructing duals can be applied to any attractive system. If it is used on additive processes like Example 2, then it leads to essentially the same duals that are found in Harris (1978) and Griffeath (1979). In such cases, the state space of the dual process is the set of finite subsets of $\mathbb{Z}^d$. When we apply our recipe to non-additive systems like Example 1, however, the result is one level more complicated: the dual process takes values in the space of finite collections of finite subsets of $\mathbb{Z}^d$. This dark cloud has a silver lining: the exotic behavior exhibited by Example 1 is only possible because its dual is complicated.

In Section 3 we prove Theorem 1, in Section 4 we prove Theorem 2 and Corollary 1, and in Section 5 we prove Theorem 3. All of these proofs are "contour arguments". This type of argument is notoriously unpleasant, so very few people have used them enough to know that they are also very simple. To oversimplify just a little, there are always three main steps:

Step 1. Introduce a lot of notation to define the contour.

Step 2. Show that a certain positive fraction of the turns in the contour must be associated with low probability events which are essentially independent and derive exponential bounds for the length of a contour with a given number of turns.

Step 3. Do a seemingly endless number of multiple integrals and then sum a geometric series.

To emphasize our point that contour arguments are simple, we have divided the proofs below into the three steps outlined above, and have tried to do the arguments in parallel as much as possible. We have done this with the aim of making contour arguments into a science (rather than an art) but we have not completely succeeded. If we had, we would be able to tell which properties of Example 1 hold for the variant system in which the birth rate is λ if (and only if) both sites in at least one of the two sets $\{x + e_1, x + e_2\}$ or $\{x - e_1, x - e_2\}$ are occupied. Apart from reaching the trivial conclusion that $\lambda_c < \infty$, which follows by comparison with Example 1, we have not been able to do this.

§2. Defining the dual processes. In this section we will show that every attractive set-valued process has two dual processes. Roughly speaking, these are obtained by starting at time t with the question "Is x occupied?" or "Is x vacant?" and working backwards to time 0. To show how this works, we will first describe a simple, rigorous method for constructing the examples in Section 1. Dual processes will then arise in a natural way from this construction.

For each $x \in \mathbb{Z}^2$ let $S_n(x)$ and $T_n(x)$, $n \geq 1$, be independent Poisson processes with rates 1 and λ respectively. Thus if we let $S_0(x) = T_0(x) = 0$, then the increments $S_n(x) - S_{n-1}(x)$ and $T_n(x) - T_{n-1}(x)$, $n \geq 1$, are independent exponentially distributed random variables, with means 1 and $1/\lambda$ respectively. Given these raw materials, the construction of the process is easy. We begin by labelling certain points in the space-time graph $\mathbb{Z}^2 \times [0,\infty)$, using the Poisson processes:

(6) mark the points $(x, S_n(x))$ with δ's (for death), and interpret the δ as telling us to kill a particle at the site x at time $S_n(x)$, if one is present;

(7) mark the points $(x, T_n(x))$ with λ's (for life), and interpret the λ as a birth at x at time $T_n(x)$, provided the necessary conditions are met, i.e.

in Example 1, $x \notin \xi_{T_n(x)^-}$ and both $x + e_1$ and $x + e_2 \in \xi_{T_n(x)^-}$,

in Example 2, $x \notin \xi_{T_n(x)^-}$ and either $x + e_1$ and $x + e_2 \in \xi_{T_n(x)^-}$.

Having marked points in the space-time graph, we can compute the evolution of the process according to the rules for interpreting the δ's and λ's given in (6) and (7).

Since there are infinitely many sites, there is no first time at which a δ or λ appears, so it is not immediately clear that the recipe above specifies a unique process. To prove this we observe that we can make $P(S_1(x) > \varepsilon, T_1(x) > \varepsilon)$ as close to 1 as we like by choosing $\varepsilon > 0$ sufficiently small. Then trivial results about percolation in $\mathbb{Z}^2$ imply that for each $x \in \mathbb{Z}^2$ there is a finite set ("random island") $C(x)$ containing x which has the property that its boundary $\partial C(x) = \{y \notin C(x)$: there exists $z \in C(x)$ with $|y - z| = 1\}$ consists completely of points y with $S_1(y) > \varepsilon$ and $T_1(y) > \varepsilon$. According to our prescription, there will be no births or deaths on $\partial C(x)$ during the time interval $[0, \varepsilon]$, so each of the islands $C(x)$ can be treated as an isolated, finite system for which the evolution is uniquely determined during $[0, \varepsilon]$. It follows that the entire process is uniquely defined for $0 \le t \le \varepsilon$, and hence by induction for all $t \ge 0$. This construction is due to Harris (1978). The reader should note that it allows us to define for each starting set $A \in \mathbb{Z}^2$ and starting time s, a process $\xi(t; s, A)$, $t \ge s$, which starts with $\xi(s; s, A) = A$ and evolves according to the rules specified above. It is important to realize that this can be done simultaneously for all A and s, all on the probability space determined by the Poisson point locations $S_n(x)$ and $T_n(x)$.

Thus we have (a rather clumsy) method for computing the evolution of the process. If we are interested in the state of a particular site x at a particular time t, however, it is more efficient to work backwards from the point (x, t) in space-time, and identify the δ's and λ's which could have affected the state of x at time t. Consider the outcome drawn in Figure 1,

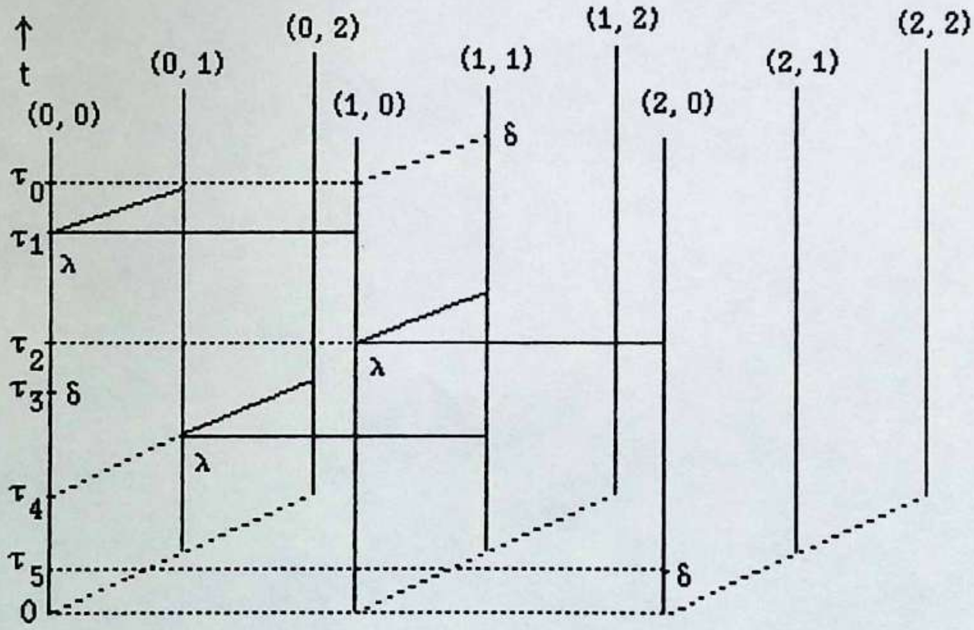


Figure 1

and suppose for the moment that it is being used to compute the evolution in Example 2. Next to each λ, we have drawn a pair of lines connecting the corresponding site to its potential parent sites. If we start at time t with the question "Is the site x = (0, 0) occupied?" and work backward we have the following answers:

In this time interval	the answer is yes if any of these sites is occupied
$(\tau_1, t]$	(0,0)
$(\tau_2, \tau_1]$	(0, 0), (0, 1), (1, 0)
$(\tau_3, \tau_2]$	(0, 0), (0, 1), (1, 0), (1, 1), (2, 0)
$(\tau_4, \tau_3]$	(0, 1), (1, 0), (1, 1), (2, 0)
$(\tau_5, \tau_4]$	(0, 1), (1, 0), (1, 1), (0, 2), (2, 0)
$[0, \tau_5]$	(0, 1), (1, 0), (1, 1), (0, 2)

Table 1

If instead, we take the same outcome pictured in Figure 1 and use it to compute the evolution of Example 1, we get much different answers:

In this time interval	the answer is yes if any of these sets of sites is completely occupied
$(\tau_1, t]$	{(0, 0)}
$(\tau_2, \tau_1]$	{(0, 0)} or {(1, 0), (0,1)}
$(\tau_3, \tau_2]$	{(0, 0)}
	or {(1, 0), (0,1)}
	or {(0, 1), (1, 1), (2,0)}
$(\tau_4, \tau_3]$	{(1, 0), (0,1)}
	or {(0, 1), (1, 1), (2,0)}
$(\tau_5, \tau_4]$	{(1, 0), (0,1)}
	or {(0, 1), (1, 1), (2, 0)}
	or {(1, 0), (1, 1), (0, 2)}
	or {(1, 1), (0, 2), (2, 0)}
$[0, \tau_5]$	{(1, 0), (0,1)}
	or {(1, 0), (1, 1), (0, 2)}

Table 2

From the sample paths above it should be clear that

(a) in Example 2 we can use the second column in Table 1 to define a dual process $\tilde{\xi}_s$, $s \in [t, 0]$, which has the interpretation that if some point in $\tilde{\xi}_s$ is occupied at time s then $x = (0,0)$ will be occupied at time t,

(b) in Example 1 we can use the second column of Table 2 to define a dual process χ_s, $s \in [t, 0]$, whose state at any time is a collection of finite subsets of $\mathbf{Z}^2$ and has the interpretation that if one of the sets in χ_s is completely occupied, then $x = (0,0)$ will be occupied at time t.

A little thought reveals that the behavior of the dual processes can be described as follows (remember that one works backward in time when tracking the evolution of the dual process):

(8) when a δ occurs at site x at time s, then
in Example 2, x disappears from $\tilde{\xi}_s$
in Example 1, all sets containing x disappear from χ_s;

(9) when a λ occurs at site x at time s, then
in Example 2, $x + \mathbf{e}_1$ and $x + \mathbf{e}_2$ are added to $\tilde{\xi}_s$
in Example 1, the set $(A \setminus \{x\}) \cup \{ x + \mathbf{e}_1, x + \mathbf{e}_2\}$ is added to χ_s for each $A \in \chi_s$ such that $x \in A$.

To prepare for a remark below we would like to observe that if we rewrote $\tilde{\xi}_s$ as a collection of singletons $\tilde{\chi}_s = \{\{x\} : x \in \tilde{\xi}_s\}$, then the occurrence of a δ at x affects χ_s and $\tilde{\chi}_s$ in the same way, that is, all sets containing x are removed. However, the occurrence of a λ at x affects $\tilde{\chi}_s$ differently than it does χ_s: if $\{x\} \in \tilde{\chi}_s$, then two new sets, $\{x + \mathbf{e}_1\}$ and $\{x + \mathbf{e}_2\}$, are added to $\tilde{\chi}_s$.

To define the dual process rigorously we need a number of definitions. Let $\mathbb{Z}^d_f$ denote the finite subsets of $\mathbb{Z}^d$. A function π from $[s, t]$ into $\mathbb{Z}^d_f$ is said to fill a set B at time t starting from a set A at time s if π is right continuous, has a finite number of discontinuities, and satisfies:

(10) $\pi(s) = A$ and $\pi(t) \supset B$; and

(11) if $s \le u < v \le t$, then $\xi(v; u, \pi(u)) \supset \pi(v)$.

Taking $u = s$ in (11) gives $\xi(v; s, A) \supset \pi(v)$, so $\pi(v)$ is a subset of the set of occupied sites. Applying (11) at $u > s$ shows that π is "self-sustaining" in the sense that if $\pi(u)$ is occupied at time u then $\pi(v)$ will be occupied at time v. An example should help explain the definition. For the sample path in Figure 1, one way to fill $\{(0, 0)\}$ at time t starting from $\{(0, 2), (1, 1), (1, 0)\}$ at time 0 is to use the following π (for the process of Example 1):

time	$\pi(s)$
$[0, \tau_4)$	$\{(0, 2), (1, 1), (1, 0)\}$
$[\tau_4, \tau_1)$	$\{(0, 2), (1, 1), (0, 1), (1, 0)\}$
$[\tau_1, \tau_0)$	$\{(0, 2), (1, 1), (0, 1), (1, 0), (0, 0)\}$
$[\tau_0, t]$	$\{(0, 2), (0, 1), (1, 0), (0, 0)\}$.

The reader should note that in the definition above we only require $\xi(v; u, \pi(u)) \supset \pi(v)$ and that in the example worked out above we could have made π "smaller". Given this state of affairs we will find it convenient to restrict our attention to π's which satisfy the additional condition:

(12) π is minimal (that is, if $\pi'(u) \subset \pi(u)$ for all $u \in [s, t]$ and if π' satisfies both (10) and (11), then $\pi' = \pi$),

and we call such a π a path from (A, s) to (B, t). For examples of paths, we again turn to Figure 1. For the process of Example 1, the only two paths from any (A, 0) to ({0}, t) are:

time	$\pi^1(s)$	$\pi^2(s)$
$[0, \tau_4)$	$\{(0, 2), (1, 1), (1, 0)\}$	$\{(0, 1), (1, 0)\}$
$[\tau_4, \tau_1)$	$\{(0, 1), (1, 0)\}$	$\{(0, 1), (1, 0)\}$
$[\tau_1, t]$	$\{(0, 0)\}$	$\{(0, 0)\}$.

We are now ready for the formal definition of our first kind of dual process. Although we have been concentrating on the processes in Examples 1 and 2 for illustrations, this definition works quite generally for attractive processes: for $s \leq t$ and $B \in \mathbb{Z}_f^d$, let

$$\chi^1(s; t, B) = \{A \in \mathbb{Z}_f^d : \text{ there is a path from } (A, s) \text{ to } (B, t)\}.$$

Our insistence on defining the dual in terms of paths rather than just in terms of functions which only satisfy (10) and (11) makes $\chi^1(s; t, B)$ a collection of sets A which are minimal in the sense that if the process starts at time s in a state A' which is strictly smaller than A, then B will not be completely occupied at time t.

Comparison of Table 2 with the descriptions of the paths π^1 and π^2 shows that for Example 1, $\chi^1(0; t, \{(0, 0)\})$ has the same state as the previously defined χ_0. The reader will find it a good exercise to work out the paths to (t, {(0, 0)}) from different times $s \in [0, t]$, and then check from Table 2 that for all such s, $\chi^1(s; t, \{(0, 0)\}) = \chi_s$. Another good exercise is to check that in Example 2, $\chi^1(s; t, \{(0, 0)\}) = \tilde{\chi}_s$. We recommend that the reader do these before proceeding further.

The process defined above is what we call the "occupancy dual" since it is obtained by asking the question "Is B completely occupied at time t?" Given this description it should be clear that we can also define a "vacancy dual" $\chi^0(s; t, B)$ by asking the question "Is B completely vacant at time t?", and that if we do this, then the result is just the occupancy dual of the complementary process η, where:

$$\eta(t; s, A) = \xi(t; s, A^c)^c.$$

If we use this definition and compute the vacancy dual of Example 2 for the outcome in Figure 1, we essentially get the occupancy dual again. The only change is that the heading of the second column in Table 1 is changed to "the answer is yes if <u>all</u> of these sites are vacant" or in terms of the dual processes, $\chi^0(s; t, \{(0, 0)\}) = \{\tilde{\xi}_s\}$, that is, it is the set whose only element is $\tilde{\xi}_s$, whereas the occupancy dual $\chi^1(s; t, \{(0, 0)\})$ for Example 2 is the collection of singletons whose union is $\tilde{\xi}_s$.

If we compute the vacancy dual of Example 1, however, what we get looks quite different from the occupancy dual:

time	$\chi^0(s; t, \{(0, 0)\})$
$(\tau_1, t]$	$\{(0, 0)\}$
$(\tau_2, \tau_1]$	$\{(0, 0), (0, 1)\}, \{(0, 0), (1, 0)\}$
$(\tau_3, \tau_2]$	$\{(0, 0), (0, 1)\}, \{(0, 0), (1, 0), (2, 0)\}, \{(0, 0), (1, 0), (1, 1)\}$
$(\tau_4, \tau_3]$	$\{(0, 1)\}, \{(1, 0), (2, 0)\}, \{(1, 0), (1, 1)\}$
$(\tau_5, \tau_4]$	$\{(0, 1), (0, 2)\}, \{(0, 1), (1, 1)\}, \{(1, 0), (2, 0)\}, \{(1, 0), (1, 1)\}$
$[0, \tau_5]$	$\{(0, 1), (0, 2)\}, \{(0, 1), (1, 1)\}, \{(1, 0)\}$

Table 3

The reader should note that the paths that go with the vacancy dual are "minimal cut sets" for the occupancy dual. In other words, if π^0 is a path corresponding to the vacancy dual then for all paths π^1 corresponding to the occupancy dual and for all times $u \in [s, t]$, $\pi^0(u) \cap \pi^1(u) \neq \emptyset$ and no smaller function satisfying conditions (10) and (11) has this property. A little thought reveals that the vacancy dual of Example 1 evolves as follows:

(13) when a δ occurs at site x at time s, x disappears from all the sets in $\chi^0(s; \cdot, \cdot)$

(14) when a λ occurs at site x at time s, all sets $A \in \chi^0(s; \cdot, \cdot)$ which contain x are replaced by two sets, $A \cup \{x + e_1\}$ and $A \cup \{x + e_2\}$.

To prepare for the proof in the next section, the reader should note that if in the outcome in Figure 1 there had been a δ at the site (1, 0) at some time $\tau_6 \in (0, \tau_5)$, then we would have (from Tables 2 and 3 and the rules for computing duals):

$$\chi^0(0; t, \{(0, 0)\}) = \{\emptyset, \{(0, 1), (0, 2)\}, \{(0, 1), (1, 1)\}\}$$
$$\chi^1(0; t, \{(0, 0)\}) = \emptyset.$$

From the state of the vacancy dual, we see that the site (0, 0) will be vacant at time t if the set $\emptyset$ is completely vacant at time 0 (which is of course always the case). On the other hand, the state of the occupancy dual tells us that there is no set A whose total occupancy at time 0 would lead to the occupancy of the site (0, 0) at time t. In either case, we conclude that (0, 0) is vacant at time t.

We conclude this section with a sketch of the proof promised in Section 1 that for the asexual contact process (Example 2),

$$\text{if } p > 0, \text{ then } \xi_t^p \Rightarrow \xi_\infty^1 \text{ as } t \to \infty.$$

In terms of the occupancy dual of Example 2, we have

$$P(x \in \xi_\infty^1) = \lim_{t\to\infty} P(\chi^1(0; t, x) \neq \emptyset) \text{ and}$$

$$P(x \in \xi_t^p) = P(\chi^1(0; t, x) \text{ contains } \{z\} \text{ for at least one site z which is occupied by the initial state } \xi_0^p)$$

$$\geq (1 - (1 - p)^N)\, P(\chi^1(0; t, x) \text{ contains at least N singletons})$$

A standard fact about transient Markov processes implies that $P(\chi^1(0; t, x)$ contains at least N singletons) has the same limit as $t \to \infty$ as $P(\chi^1(0; t, x) \neq \emptyset)$, for all finite N. Thus $P(x \in \xi_t^p) \geq (1 - (1 - p)^N)$ $P(x \in \xi_\infty^1)$ for all finite N, and the proof is completed by letting $N \to \infty$. □

§3. Proof of Theorem 1. In this section we will prove Theorem 1. To begin with we observe that

(15) $P(x \notin \xi(t; 0, \mathbb{Z}^2)) = P(\emptyset \in \chi^0(0; t, \{x\}))$
$= P(\text{a vacancy path exists from } (\emptyset, 0) \text{ up to } (\{x\}, t)),$

where by a "vacancy path" we mean the object one obtains by applying the definition of a path in Section 2 to $\eta(t; s, A)$ instead of $\xi(t; s, A)$. For an example of a vacancy path, we turn to the modification of the outcome in Figure 1 which was made at the end of Section 2, namely the one in which we inserted a new time τ_6 at which a δ occurred at the site (1, 0). Then we have

<u>time</u>	<u>$\pi(s)$</u>
$[0, \tau_6)$	$\emptyset$
$[\tau_6, \tau_5)$	$\{(1, 0)\}$
$[\tau_5, \tau_3)$	$\{(1, 0), (2, 0)\}$
$[\tau_3, \tau_2)$	$\{(0, 0), (1, 0), (2, 0)\}$
$[\tau_2, \tau_1)$	$\{(1, 0), (0, 0)\}$
$[\tau_1, t]$	$\{(0, 0)\}$

We are now ready for the steps outlined in the first section.

Step 1.

As we mentioned in the introduction, the first step in the proof is defining a "contour" in such a way that

$$P(0 \notin \xi_\infty^1) \quad \leq \quad P(\text{contour exists}).$$

Start by extending the Poisson processes defined in the beginning of Section 2 to all of $\mathbb{R}$ in the time direction (they were originally restricted to $[0, \infty)$). Thus the random variables $S_n(x)$ and $T_n(x)$ are now indexed by $n \in \mathbb{Z}$. This allows us to define $\xi(t; s, A)$ for all real $s \leq t$. It is clear from the homogeneity in time that

(16) $P(x \notin \xi(t; 0, \mathbb{Z}^2)) \;=\; P(x \notin \xi(0; -t, \mathbb{Z}^2)$

$= \;\; P(\text{there is a vacancy path from } (\emptyset, -t) \text{ up to } (\{x\}, 0))$

where now the last event is increasing in t. Introducing some obvious notation for the union of all these events (over $t > 0$) we have

(17) $P(x \notin \xi^1_\infty) =$ P(there is a vacancy path from $(\emptyset, -\infty)$ up to $(\{x\}, 0)$).

Let $\pi(t)$, $t \le 0$, be a vacancy path from $(\emptyset, -\infty)$ up to $(\{x\}, 0)$. We have found that negative times are a nuisance, so at this point it will be convenient for us to reverse time. (This is a small dilemma faced by all who work with graphically defined dual processes.) Let $\sigma(t)$ be the right continuous modification of $\pi(-t)$, $t \ge 0$. With each possible σ we will associate a contour which starts at $(\{(0, 0)\}, 0)$, traces around the outside of the set $\Sigma = \{(x, t): x \in \sigma(t)\}$, and returns to $(\{(0, 0)\}, 0)$. (Note that Σ is necessarily bounded since $\pi(t) = \emptyset$ for large negative values of t.)

In describing a contour of length n, we will use

(i) a code consisting of an alternating sequence $a_1, b_1, a_2, \ldots, a_n, b_n, a_{n+1}$ of letters a_m taken from the two element set {u, d} (standing for up and down) and numbers $b_m \in \{1, 2, 3, \ldots\}$;

(ii) a sequence of nonnegative numbers t_i, $1 \le i \le n + 1$, t_i being the amount of time we travel in the direction a_i.

The contour for a given path is constructed according to the rules given below. The numbers in the definition are the ones we will use to code the types of turns made by the contour. Keep in mind that "up" and "down" refer to directions associated with the "upside down" (time-reversed) path σ. In particular, the contour will start out by moving up from the point ((0, 0), 0).

Rules for constructing a contour: Starting at the point ((0, 0), 0) and moving upward, the contour moves according to the following rules until it returns to the point ((0, 0), 0): If the contour is at (x, t) and moving up, then it continues upward until the first time $u > t$ which is in one of the two sets $D_x = \{S_n(x), n \in \mathbb{Z}\}$ (the deaths at x) or $L_x = \{T_n(x), n \in \mathbb{Z}\}$ (the births at x). If $u \in L_x$, then by definition the path σ must either already contain $x + e_1$ or $x + e_2$, or it must expand to include one of those two sites. In either case, $\sigma(u)$ contains either $x + e_1$ or $x + e_2$ (or both).

(Rule 1) If $u \in L_x$ and if $x + e_1 \in \sigma(u)$, then the contour jumps to

$x + e_1$ and then continues upward;

(Rule 2) If $u \in L_x$ and if $x + e_1 \notin \sigma(u)$, then the contour jumps to $x + e_2$ and continues upward.

If, on the other hand, $u \in D_x$, then by definition, $x \notin \sigma(u)$. Thus

(Rule 3) If $u \in D_x$, the contour stays at x and starts to move down.

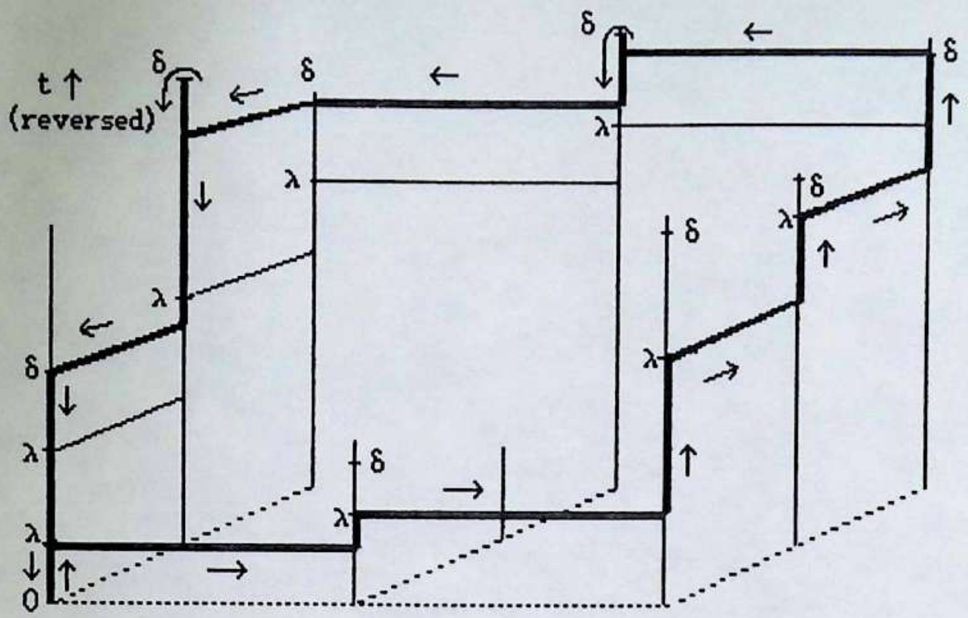
Now suppose the contour is at (x, t) and starting to move downward. If $x = (0, 0)$ then it continues downward until it reaches ((0, 0), 0), and the construction is then ended. Now suppose $x \neq (0, 0)$. Then according to a procedure which we will describe at the end of this section, we choose a site y from the set $\{x - e_1, x - e_2\}$ and a time $s \leq t$, such that $y \in \sigma(s^-)$. The choice will be made in such a way that if $s < t$, then $s \in D_y$. Accepting for now that this is possible, the contour moves downward from the point (x, t) to the point (x, s) and then jumps to y. There are four possibilities.

(Rule 4) If $s < t$ and $y = x - e_1$, or **(Rule 5)** $s < t$ and $y = x - e_2$, the contour starts to move downward after the jump.

(Rule 6) If $s = t$ and $y = x - e_1$, or **(Rule 7)** $s = t$ and $y = x - e_2$, then the contour moves upward after the jump.

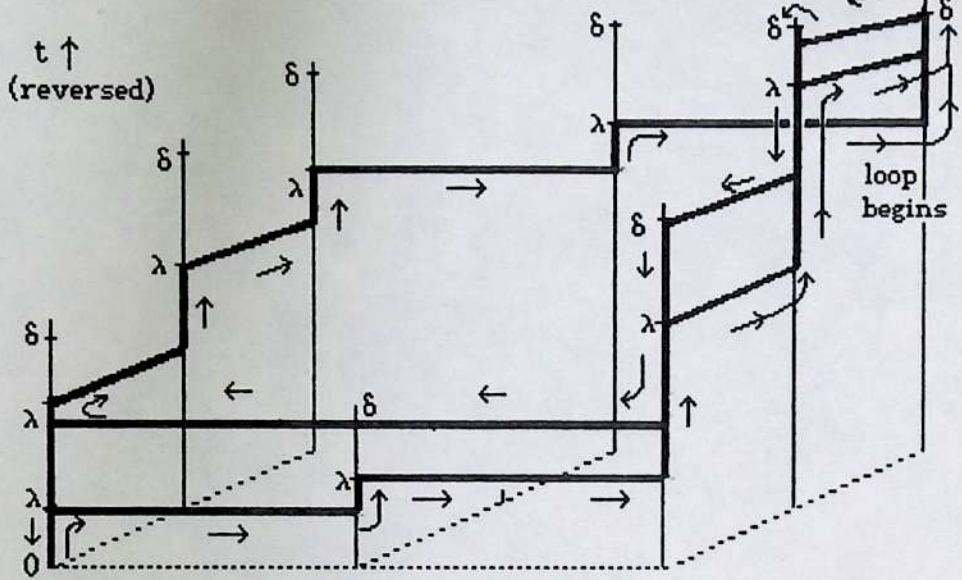
Notice that in the last two rules, the movement downward from (x, t) to (x, s) has distance 0, since $s = t$. This is why we use the phrase "starts to move downward" in rules 3 - 5. We will have more to say about this when we define the code for the contour.

In order to complete the construction of the contour, we need to discuss the procedure for choosing the site y and the time s in rules 4 - 7. We had originally thought that it would be sufficient to choose $s = \sup \{s \leq t: \exists\, y \in \{x - e_1, x - e_2\} \cap \sigma(s^-)\}$, and then to choose y arbitrarily from the set $\{x - e_1, x - e_2\} \cap \sigma(s^-)$. Using this procedure we arrive at the contour shown in Figure 2 (remember that we have reversed time, so that up and down are reversed from the way they were in Figure 1).



The coding sequence is:
u1u1u2u2u36u3d47u3d5d
(see text below for explanation of combined numbers such as 36 and 47)
Figure 2

Unfortunately there are certain exceptional situations in which such a simple method gets the contour into a loop and the construction never ends (see Figure 3, which is a modification of Figure 2). To avoid this problem, the choice of s and y must be made more carefully. Fortunately, the details of how this choice is made have nothing to do with the rest of the proof. Therefore, we will merely state here the facts concerning the existence of an acceptable procedure and leave the technicalities to the end of the section. It turns out that this more complicated procedure leads to the same contour as the simple method described above in many cases. In particular this is true of the example in Figure 2, and the reader should not feel uncomfortable in using this example as an illustration of the ideas in the main part of the proof of Theorem 1.



The coding sequence is u1u1u2u2(u37u3d5d46u2u2u1u1)$^\infty$

Figure 3

Proposition 1. In the construction of the contour, rules 4 through 7, there exists a procedure for choosing the site y from the set $\{x - e_1, x - e_2\}$ and a time $s \le t$ such that $y \in \sigma(s^-)$ and such that $s \in D_y$ if $s < t$. This choice has the property that the contour never passes over a point in space-time more than once in a given direction.

Having constructed the contour, we are ready to define the code used in its description. Looking back through the rules above, we see that in the sequence of alternating letters and numbers which codes the turns made by the contour, there are only seven possible values for any triple of the form (a_k, b_k, a_{k+1}), with the code letters representing a direction, followed by a rule number, followed by another direction:

u 1 u	d 4 d
u 2 u	d 5 d
u 3 d	d 6 u
d 7 u .	

For future calculations, it will be useful to remove any letter "d" which corresponds to a downward segment of length 0 in the contour (see rules 6 and 7). Thus we remove any "d" which precedes either a "6" or a "7". Once this is done, combine any "6" or "7" with the number before it in the sequence to form a two-digit number. This produces an alternating sequence of letters and numbers in which the possible values of a triple (a_k, b_k, a_{k+1}) are:

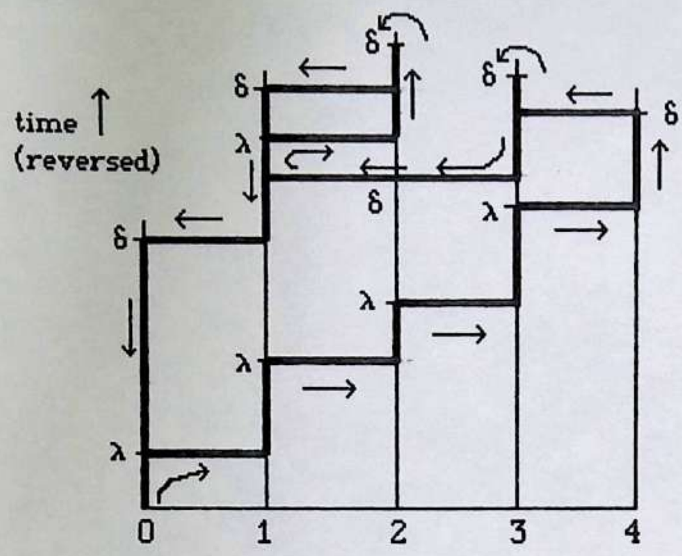
u 1 u	d 4 d	u 36 u	u 37 u
u 2 u	d 5 d	d 46 u	d 47 u
u 3 d	d 56 u	d 57 u .	

The resulting sequence of alternating letters and numbers is the code for our contour. As described earlier, the contour is completely determined by this code together with the sequence of (strictly positive) distances t_i travelled by the contour in the direction a_i.

Remark. Before we move on to the next step, it may be helpful to the reader to see what our contour would look like for the one-sided contact process on $\mathbb{Z}$. This process is built from the same ingredients as the model we are working on (i.e., independent Poisson processes $\{S_n(x)\}$ and $\{T_n(x)\}$ for each $x \in \mathbb{Z}$ with rates 1 and λ respectively), but now the points $T_n(x)$ are interpreted as "there will be a birth at x at time $T_n(x)$ if at time $T_n(x)^-$ $x + 1$ is occupied and x is vacant (see Griffeath (1979) for more explanation and a survey of what was known then about this process).

We now imitate the contour construction given above in this new context. The contour starts at the point $x = 0$, $t = 0$, and moves according to the following rules until it returns to its starting point:

If the contour is at (x, t) and moving upward, then it continues upward until the first time $u > t$ which is in D_x or L_x, where these two sets are defined as before. **(Rule 1)** If $u \in L_x$, then it must be that $x + 1 \in \sigma(u)$. The contours jumps to $x + 1$ and continues upward. **(Rule 2)** If $u \in D_x$, then the contour stays at x and (starts to) move downward. If the contour is at (x, t) and moving downward, then it continues downward until the first time $s \le t$ at which $x - 1$ is in $\sigma(s^-)$. **(Rule 3)** If $s < t$, the contour jumps to $x - 1$ and continues to move down. **(Rule 4)** If $s = t$, then the contour jumps to $x - 1$ and moves upward. As before, if the contour is at $x = (0, 0)$ and moving down, then it continues down until it reaches time 0 and then ends. The example drawn in Figure 4 should explain the definitions. Those readers that have seen contour methods before will note that our new contours are in this case just shrunken versions of the ones found in Gray and Griffeath (1982).



The coding sequence is: u1u1u1u1u24u2d34u1u2d3d3d
Figure 4

Step 2.

Let ω be a realization of the system of Poisson processes that underlie our model, and assume that ω is such that there exists a vacancy path from $(\varnothing, -\infty)$ up to $((0, 0), 0)$. Let

$N(\omega)$ = the length of the corresponding contour = number of turns

$\Delta(\omega) = (a_1, b_1, a_2, \ldots, b_N, a_{N+1})$ = the corresponding coding sequence, with $N = N(\omega)$

$\tau(\omega) = (t_1, t_2, \ldots, t_N)$ = the lengths of the time segments travelled by the contour; in particular, t_i is the length of the segment covered by the contour which is associated with the direction a_i.

In the definition of τ we left out the last time segment t_{N+1} since its length can be determined from the others: if we set $\varepsilon_i = 1$ if $a_i = u$ and $\varepsilon_i = -1$ if $a_i = d$, then $\Sigma\varepsilon_i t_i = 0$ if the summation is taken from $i = 1$ to $N + 1$. We are now ready for

The Basic Estimate. Let Δ be a coding sequence of length N and let A be any Borel subset of $\mathbb{R}^N$. Define $\mu_\Delta(A) = P(\Delta(\omega) = \Delta, \tau(\omega) \in A)$. Then μ_Δ is absolutely continuous with respect to N-dimensional Lebesgue measure ν_N and

$$(18) \qquad \frac{d\mu_\Delta}{d\nu_N} \quad \le \quad \lambda^{N_1+N_2} \exp\left(-\sum_{i\in U} t_i (\lambda+1)\right)$$

where $N_j = |\{i: b_i = j\}|$, $j = 1, 2$, and $U = \{i: a_i = u\}$.

Proof. The result is easy to explain but tedious to prove. To see why (18) is true we observe that

if b_i is in	the turn is due to
$\{1,2\}$	birth at $\mathbf{x}$
$\{3, 36, 37\}$	death at $\mathbf{x}$
$\{4, 46, 47\}$	death at $\mathbf{x} - \mathbf{e}_1$
$\{5, 56, 57\}$	death at $\mathbf{x} - \mathbf{e}_2$

and if we are going up then we are forced to move or change direction if we encounter a birth or a death.

The first task in proving (18) is to check that the contour cannot make two different turns that are both associated with the same Poisson point. We will use the fact (Proposition 1) that the contour never traverses the same segment twice travelling in the same direction. Since the contour only makes turns involving births when it is travelling upward, we do not need to worry about using the same birth twice. Similarly, since the contour always starts to travel down after making a turn involving a death, our assumption also rules out using the same death twice.

The argument in the preceding paragraph shows that each turn in the contour is associated with a unique point. Therefore, the coding which describes a contour of length n determines the locations of n distinct Poisson points. This fact accounts for the absolute continuity with respect to Lebesgue measure and for the factor $\lambda^{N_1+N_2}$ in (18). The exponential term follows from the fact that, according to the rules that define the contour, no Poisson points can be located within any of the segments along which the contour travels in an upward direction. For more details, we refer to Gray and Griffeath (1982), Lemma 8, where a similar estimate is proved. □

We are at last ready for the (trivial) observation which makes the proof of Theorem 1 work:

(19) there exists $\varepsilon > 0$ such that $N_1 + N_2 \le (1 - \varepsilon)N$.

It is clear that in our attempts to control the estimate in (18), the type 1 and type 2 turns are our enemies, while the other types are our friends. The statement in (19) is just what we need to ensure that our friends are more powerful than our enemies (see Step 3 below).

To prove (19), we use the simple fact that if the contour exists, it begins and ends at the same point. Since each type of turn is associated with a jump of a specific size and direction , this fact leads to the some equations involving the N_i 's. We tabulate the jumps associated with each type of turn in the next table:

Type of turn	Jump made by contour
u1u	(1, 0)
u2u	(0, 1)
u3d	(0, 0)
d4d	(-1, 0)
d5d	(0, -1)
u36u	(-1, 0)
u37u	(0, -1)
d46u	(-2, 0)
d47u	(-1, -1)
d56u	(-1, -1)
d57u	(0, -2)

Table 4

Then

$$N_1 - N_4 - N_{36} - 2N_{46} - N_{47} - N_{56} = 0$$
$$N_2 - N_5 - N_{37} - N_{47} - N_{56} - 2N_{57} = 0.$$

It follows that

$$N_1 + N_2 = N_4 + N_5 + N_{36} + N_{37} + 2(N_{46} + N_{47} + N_{56} + N_{57}) ,$$

and since $\Sigma N_i = N$ it follows that (19) holds with $\varepsilon = 1/3$. We can improve this slightly by noting that each turn of type 3 reverses the direction of the contour from up to down, while turns of types 46, 47, 56, or 57 reverse the direction of the contour from down to up. Since the contour starts in the upward direction and ends downward, it follows that

$$N_3 = N_{46} + N_{47} + N_{56} + N_{57} + 1,$$

implying that

$$N_1 + N_2 + 1 = \sum_{i\geq 3} N_i$$

Therefore (19) holds with $\varepsilon = 1/2$:

(20) $N_1 + N_2 \leq N/2.$

Step 3.

The last step in the proof is to integrate the basic estimate and sum over all possible coding sequences Δ. We define $\mathcal{D}(\Delta)$ to be the domain of integration associated with Δ. In other words, $\mathcal{D}(\Delta)$ is the subset of $\mathbb{R}^N$ consisiting of all sequences of lengths $(t_1, \ldots, t_N)$ which are possible for a contour with coding sequence Δ. Then

(21) $$P(\Delta(\omega) = \Delta) = \mu_\Delta(\mathbb{R}^N)$$

$$\leq \lambda^{N_1+N_2} \int_{\mathcal{D}(\Delta)} dt_1 \cdots dt_N \exp\left(-\sum_{i\in U} (\lambda + 1)\, t_i\right).$$

Let $V = \{1, \ldots, N\} \setminus U$ and observe that since we have left out t_{N+1}

(22) $$\sum_{i\in U} t_i \geq \sum_{i\in V} t_i$$

Let $k = |U|$. Now we make a change of variables: change t_i, $i \in V$, to $v_1, \ldots, v_{N-k}$; change t_i, $i \in U \setminus \{\max U\}$, to $u_1, \ldots, u_{k-1}$; and let $r = u_1 + \ldots + u_k$ Then by the definition of r and (22), the domain of integration of the following multiple integral contains $\mathcal{D}(\Delta)$:

(23)

$$\int_0^\infty dr \exp(-r(1 + \lambda)) \int_{\Sigma u_i \leq r} du_1 \cdots du_{k-1} \int_{\Sigma v_i \leq r} dv_1 \cdots dv_{N-k} .$$

It follows that the quantity in (23) is an upper bound for the right side of (21). Using elementary calculus, the reader can check that the multiple integral in (23) equals

$$(1 + \lambda)^{-N} \binom{N-1}{k-1},$$

which is bounded above by $2^{N-1} (1 + \lambda)^{-N}$, so we have shown that if $\lambda \geq 1$,

$$\begin{aligned}
(24)\ P(\Delta(\omega) = \Delta) \quad &\leq \quad 2^{N-1}\ \lambda^{N_1+N_2}\ (1 + \lambda)^{-N} \\
&\leq \quad 2^{N-1}\ \lambda^{N/2}\ (1 + \lambda)^{-N} \quad (\text{since } N_1 + N_2 \leq N) \\
&= \quad 1/2\ \ b(\lambda)^N
\end{aligned}$$

where $b(\lambda) = 2\ \lambda^{1/2}(1 + \lambda)^{-1}$. The last detail is to compute a bound on the number of different possible codings of length N. Since there are 11 different types of turns, a trivial bound is 11^N. If we multiply the right side of (24) by 11^N and sum on N, we obtain a bound on the probability that a contour of any length exists, and hence by (17) a bound on $P(x \notin \xi^1_\infty)$. Since $b(\lambda) \to 0$ as $\lambda \to \infty$, Theorem 1 follows. □

The estimates made at the end of the proof show that $\lambda > \lambda_c$ if $b(\lambda)$ is small enough that $\Sigma\ (11\,b(\lambda))^N < 2$. This leads to an upper bound on λ_c somewhere between 4000 and 5000. We can improve this by using two standard tricks. The first is to note that it is not the series $\Sigma\ (11\,b(\lambda))^N$ itself which is so important, but instead its tail. The reason is that in order to prove that the distribution ξ^1_∞ is non-trivial, it is sufficient to prove that $P(\ \xi^1_\infty$ contains no site in $\{(0, 0),\ (-1, 0)\ \ldots,\ (-M + 1, 0)\}) < 1$ for some M. If this event occurs, then a contour of length at least M exists (it will start at (0, 0) and end at (-M + 1, 0), a difference which does not affect the estimate), so the probability is bounded by summing $(11\,b(\lambda))^N$ over $N \geq M$. It follows that $\lambda > \lambda_c$ if $b(\lambda) < 1/11$, or in other words, $\lambda_c < 482$.

The second improvement involves the quantity 11^N. It is easy to see that many of the coding sequences that we have counted are impossible since certain types of turns cannot follow certain other types of turns. For example, a type 4 turn, which ends with the contour heading down can only be followed by types 4, 5, 46, 56, 47, or 57 turns, which are the only ones that follow a downward direction. If we group the types of turns according to the associated directions,

u1u	u3d	d4d	d46u
u2u		d5d	d56u
u36u			d47u
u37u			d57u

we see from standard graph theory ideas that the number of coding sequences of length N that respect these restrictions is equal to the sum of the entries of the matrix

$$\begin{bmatrix} 4 & 1 \\ 4 & 2 \end{bmatrix}^N$$

which is bounded by a constant times (the largest eigenvalue)^N. The largest eigenvalue is $3 + \sqrt{5}$. Therefore, we can replace 11 by $3 + \sqrt{5}$, leading to our final bound

$$\lambda_c < 108.$$

Proof of Proposition 1. The construction will be in two stages. First we will define a temporary procedure for choosing the site y that ensures that the contour will have a finite length. This procedure will sometimes result in a contour that visits the same point more than once while travelling in the same direction, but it will avoid infinite loops. The second stage will be to simply cut out repeated portions of the temporary contour. This last modification will produce the desired contour and determine the procedure for choosing the site y and the time s in rules 4 - 7.

We will now give the temporary method for choosing the site y and the time s in rules 4 - 7. Assume that the contour has been formed up until it has reached a site $x \neq (0, 0)$ at time t and that from there it has started to move downward, as in rules 3 - 5. Assume further that somewhere in the part of the contour that has already been formed, there has been a jump from one of the sites $x - e_1$ or $x - e_2$ to the site x. Let (y, s') be the point in space-time from which such a jump most recently took place in the formation of the contour. (The expression "most recently" does not have reference to the time parameter in the dual process, but instead to motion along the contour as it is constructed.) Now let $s = t \wedge \min\{u > s' : u \in D_y\}$, and let the contour move downward from (x, t) to (x, s) and then jump to (y, s). If $s = t$, then the contour moves upward from (y, s) as in rules 6 and 7, otherwise the contour starts to move downward from (y, s). It can be easily checked inductively that if this procedure is always followed, the contour will only visit sites x such that either $x = (0, 0)$, or x was previously jumped to from one of the sites $x - e_1$ or $x - e_2$. Thus the procedure can be applied throughout the construction and determines our temporary choice of the site y and the time s. Our temporary contour is now well-defined in a manner that is consistent with the rules. It has the property that in rules 4 - 7, if $s < t$, then $s \in D_y$.

It remains to modify the temporary contour to avoid visiting the same point twice while travelling in the same direction. Before we can do this, we will need to know that $s' \le s$ in the construction given in the preceding paragraph. Suppose that the contour has been constructed up to the point (x, t) where it starts to move downward, and assume inductively that the procedure for choosing the site y and the time s in rules 4 - 7 has satisifed the condition that $s' \le s$ in all previous parts of the contour where downward motion was involved (it is not necessary to our argument that any such previous parts exist, so this is a case where the inductive step and the initial step in the induction can be handled together). The contour consists of three types of components: the first type of component is an upward segment at some site z, ending in a jump to $z + e_1$ or $z + e_2$; the second type of component is an upward segment which ends when the contour reverses direction and starts downward, and the third type of component is a downward segment at a site of the form $z + e_1$ or $z + e_2$, followed by a jump to z. By the inductive hypothesis, any components of the third type must end at a time which is upwards (larger than) the time at which a previous component of the first type ended. Consequently the journey made by the contour between (y, s') and (x, t) has resulted in a net upward movement in the time direction, so that $t > s'$. It follows that $s \ge s'$.

We will now show that the temporary contour, as defined above, can never get into an infinite loop. First note that it follows from the construction that if the contour jumps from a site y to a site $x = y + e_1$ or $y + e_2$ at the time point s', then the next time the contour visits y, it will do so by jumping from (x, s) to (y, s) at some time s; as shown in the preceding paragraph, $s \ge s'$. After this last jump, the contour will move upward. If the next jump is another jump to one of the sites $y + e_1$ or $y + e_2$, then of course, this will occur at some time $s'' > s \ge s'$. It can be deduced from this fact that an infinite loop is impossible, since such a loop would have at least one site (for example, the one closest to the origin) where this behavior would be violated. The reader is invited at this point to try out the temporary procedure on the situations shown in Figures 2 and 3.

The last step is to remove some parts of the temporary contour so that the same point will not be visited twice while travelling in the same direction. Let (x, t) be such a point in the temporary contour. Consider the first and last visits made by the contour to (x, t), both travelling in the same direction. If the piece of the contour that runs between those two visits is removed, we are left with a new contour which still conforms to the rules, but which now visits (x, t) only once in the given direction. It is not hard to see that such a procedure can be carried out in a systematic way to obtain a new contour which visits each point only once while travelling in a given direction. Since it is only important that we obtain some contour which follows the rules and which satisifies the conclusions of Proposition 1, we will not spend any time choosing a specific method. The construction is now complete. □

§ 3. Proof of Theorem 2. In this section we will prove Theorem 2. Before we can begin, we need to augment the construction in Section 2 to allow for spontaneous births. For each $x \in \mathbb{Z}^2$ we let $U_n(x)$, $n \geq 1$, be a Poisson process with density β, independent of the processes $S_n(x)$ and $T_n(x)$. There is now a third rule in the description of the process, corresponding to spontaneous births at rate β:

(7') we mark the points $B_x = \{(x, U_n(x): n \geq 1\}$ with β's (for birth) and interpret the β as telling us that there will be a birth at x if x is vacant.

Thus there are now two different ways in which births occur: (7) describes births that only occur when certain neighboring sites are occupied, and (7') describes births that occur without any conditions on the neighboring sites. It is clear from the remarks in Section 2 that we can use this new graphical representation to define a process $\xi(t; s, A)$, $t \geq s$, for each starting set $A \subset \mathbb{Z}^2$ and each starting time s. This process satisfies $\xi(s; s, A) = A$ and evolves according to rules (6), (7) and (7').

Now that we have constructed the process, we can define its 1's dual $\chi^1(s; t, \{z\})$ by the recipe given in Section 2. This gives us the following:

(25) $P(z \in \xi(t; 0, \emptyset)) = P(\emptyset \in \chi^1(0; t, \{z\})$

$= P($there is a path from $(\emptyset, 0)$ up to $(\{z\}, t))$.

Step 1.

As in Section 3, it will be convenient for us to turn the picture upside down (reverse time) when we define the contour. Thus, as before, we extend our graphical construction to $t \leq 0$ and note that

(26) $P(z \in \xi(t; 0, \emptyset)) = P(z \in \xi(0; -t, \emptyset)$

$= P($there is a path from $(\emptyset, -t)$ up to $(\{z\}, 0))$.

and

(27) $P(z \in \xi^0_\infty) = P($there is a path from $(\emptyset, -\infty)$ up to $(\{z\}, 0))$.

Let $\pi(t)$, $t \leq 0$, be a path from $(\emptyset, -\infty)$ up to $(\{z\}, 0)$, and let $\sigma(t)$ be the right continuous modification of $\pi(-t)$, $t \geq 0$. Again we will define the contour by specifying

(i) an alternating sequence $a_1, b_1, a_2, \ldots, b_n, a_{n+1}$ of letters $a_m \in \{u, d\}$ and numbers b_m .

(ii) a sequence of nonnegative numbers t_i, $1 \le i \le n+1$, where t_i is the amount of time we travel in direction a_i.

This time the contour is described by the following four rules: if the contour is at (z, t) and moving upward, then it continues upward until the first time $u \ge t$ at which $x \notin \sigma$, and

(Rule 1) if $u \notin B_x$, then we know that $u \in L_x$ (since deaths are never part of the path of a 1's dual), and hence both $x + e_1$ and $x + e_2 \in \sigma(u)$. The contour jumps to $x + e_1$ and continues to go up.

(Rule 2) If on the other hand $u \in B_x$, then the contour stays at x and starts to go down.

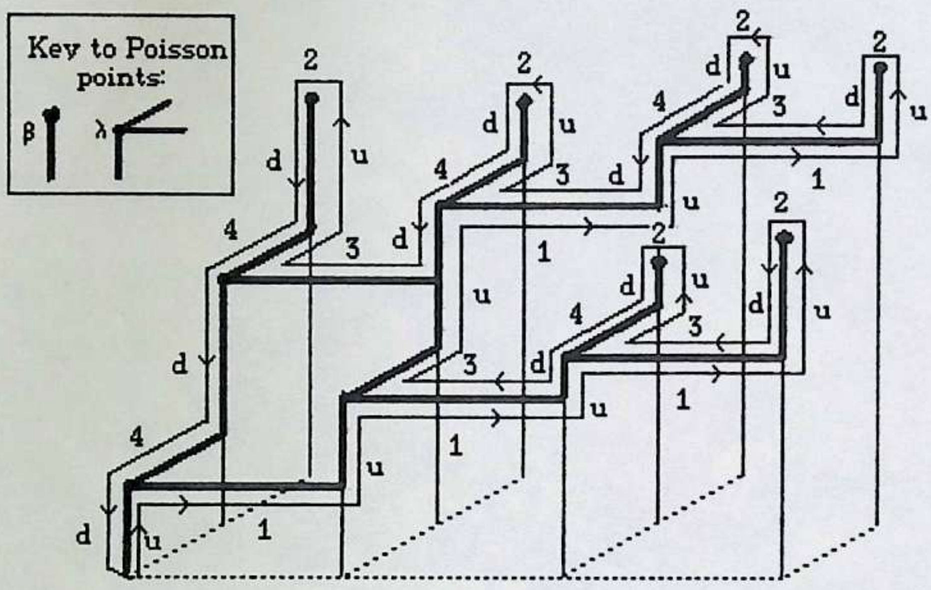
Important Remark. It will be convenient for us to allow paths which may ignore several points in B_x before reaching a time u at which $x \notin \sigma(u)$. This actually violates our minimality condition (12) for paths. However, there are two places in the remainder of this paper in which we will want to include some paths which ignore one or more of the points marked with β's. Thus we will be counting some contours that travel around non-minimal paths π in our computations. This increases our upper bounds for the probability that a contour exists, so these bounds remain valid. We will alert the reader to the two places in which this remark is used when the time comes.

If the contour is at ((0,0), t) and moving downward, then it continues downward until it reaches time 0 and then ends. If the contour is at a site $x \ne (0,0)$ and moving downward, then it continues downward until the first time $s < t$ at which the contour jumped from either $x - e_1$ or $x - e_2$, and

(Rule 3) if the jump was from $(x - e_1 , s) \in L_{x - e_1}$ (Rule 1), the contour jumps first to $x - e_1$ and then to $x - e_1 + e_2$ and then moves upward.

(Rule 4) If instead, the jump was from $(\mathbf{x} - \mathbf{e}_2 , s) \in L_{\mathbf{x} - \mathbf{e}_2}$ (Rule 3), the contour jumps to $\mathbf{x} - \mathbf{e}_2$ and then continues downward.

Note the lack of symmetry between rules 3 and 4. The example shown in Figure 5 should help explain all the rules given above. From the picture it should be clear that the contour cannot traverse the same segment twice going in the same direction, so we won't need to make any modifications such as were needed in Section 3.



The coding sequence is:
u1u1u1u2d3u2d4d3u1u1u2d3u2d4d3u2d4d3u2d4d4d

Figure 5

Looking back through the rules we see that the only possible values for the triple a_i, b_i, a_{i+1} are

u1u
u2d
d3u
d4d

and that, in contrast to the contour defined in Section 3, all the movements have positive length. That's the good news. The bad news, which we will encounter in Step 2, is that two different turns may possibly correspond to the same Poisson point (see Rules 3 and 4). This is inconvenient but not fatal (something similar happened in the contour arguments for oriented percolation in Durrett (1984)).

Step 2.

Let ω be a realization of the Poisson point locations and define $N(\omega)$, $\Delta(\omega)$, $\tau(\omega)$ and μ_Δ as in Section 3. Because of the fact that more than one turn in the contour can be associated with the same Poisson point, this measure is not absolutely continuous with respect to N-dimensional Lebesgue measure. To cope with this, we need some notation. Let

$$\varepsilon_i = 1 \quad \text{if } a_i = u$$
$$= -1 \quad \text{if } a_i = d;$$

$$s_i = \sum_{j=1}^{i} \varepsilon_i t_i = \text{the time corresponding to the end of the ith vertical movement.}$$

$$J(\omega) = \{j: \text{for some } i < j,\ s_i = s_j\}.$$

If $j \in J$, then the turn a_j, b_j, a_{j+1} corresponds to a Poisson point the contour visited previously (we are relying on the fact that except on a null set, no two Poisson points occur at exactly the same time). Let

$$\mu_{\Delta,J}(A) = P(\Delta(\omega) = \Delta,\ J(\omega) = J,\ \tilde{\tau}(\omega) \in A\},$$

where $\tilde{\tau}(\omega) \in \mathbb{R}^{N-|J|}$ is the vector obtained by listing the t_i, $i \in \{1, \ldots, N\} \setminus J$ in order. With this notation introduced, we can state

The Basic Estimate. If $M = N - |J|$ then $\mu_{\Delta,J}(A)$ is absolutely continuous with respect to ν_M (i.e., M-dimension Lebesgue measure), and

(27)

$$\frac{d\mu_{\Delta,J}}{d\nu_M} \le \beta^{N_2} \lambda^{M-N_2} \exp(-\sum_{i\in U} t_i)$$

where $N_i = |\{j: b_j = i\}|$ and U is the set of indices of upward segments.

Proof. To prove this we only need to observe that when

b_i =	this is due to a
1	λ at x
2	β at x
3	λ at $x - e_1$
4	λ at $x - e_2$

and that (i) if we want the contour to have a given sequence of lengths, then the Poisson points corresponding to the t_i for $i \notin J$ must occur at specified times, (ii) $\{j: b_j = 2\} \cap J = \emptyset$, since we can only hit β's once, and (iii) there must be no δ's at x along any of the upward segments. The reader should not have much difficulty in matching up these observations with the various terms that appear on the right of (27). □

Having proved the Basic Estimate we come at last to the trivial observation which makes the proof work:

(28) $N_2 \geq N/4$ and $M - N_2 \leq 3N/4$.

To prove this fact we begin by observing that according to the list of possible values of a_i, b_i, a_{i+1} given earlier, the only possible transitions $b_i \to b_{i+1}$ are given by the graph drawn in Figure 6.

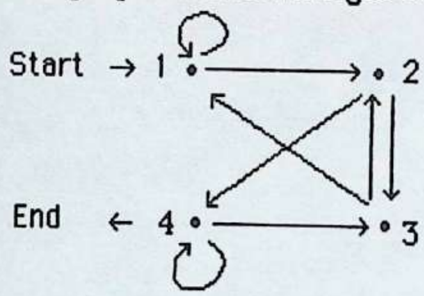


Figure 6

As we have indicated in the drawing, except for the short contour with coding u2d, all contours always begin with $b_1 = 1$ and end with $b_N = 4$. We see immediately that (i) starting from 1 we must visit 2 before 3, (ii) between any two visits to 2 there must be exactly one visit to 3, and (iii) after the last visit to 2, there can be no more visits to 3. These three observations imply that

(29) $N_2 = N_3 + 1$.

Now we list the vectors associated with each kind of jump made by a contour:

Type of turn	Jump made by contour
u1u	(1, 0)
u2d	(0, 0)
d3u	(-1, 1)
d4d	(0, -1)

Table 5

Since any contour which exists must begin and end at the same point, it follows that $N_1 = N_3$ and $N_3 = N_4$, so

$$N_2 - 1 = N_1 = N_3 = N_4 .$$

It is now easy to deduce (28). □

Step 3.

The last step in the proof is again to integrate the Basic Estimate and sum over Δ and J. The details are almost the same as in Section 3, but some minor changes occur because the Basic Estimate has a slightly different form. If we let $\mathcal{D}(\Delta, J)$ = the domain of integration associated with coding Δ and "excluded set" J then (27) implies

$$(30)\quad P(\Delta(\omega) = \Delta, J(\omega) = J) = \mu_{\Delta,J}(\mathbb{R}^N)$$

$$\leq \beta^{N_2} \lambda^{M-N_2} \int_{\mathcal{D}(\Delta, J)} dt_1 \ldots dt_M \exp(-\sum_{i\in U} t_i)$$

If we let V be the indices corresponding to downward segments, then since we have left out some of the downward segments (the set J only includes indices corresponding to downward segments) , we have

$$\sum_{i\in U} t_i \geq \sum_{i\in V-J} t_i$$

Let $k = |U|$. Now we make the following changes of variables: change t_i, $i \in U$, to $u_1, \ldots, u_k$; change t_i, $i \in V \setminus J$, to $v_1, \ldots, v_{M-k}$, and let r =

$u_1 + \ldots + u_k$. Then the integral in (30) is less than

(31)

$$\int_0^\infty dr \exp(-r) \int du_1 \ldots \int du_{k-1} 1_{(\Sigma u_i \le r)} \int dv_1 \ldots \int dv_{M-k} 1_{(\Sigma v_i \le r)}$$

which by calculations similar to those in Section 3 is

$$\le 2^{M-1} .$$

At this point we have shown

(32) $P(\Delta(\omega) = \Delta, J(\omega) = J) \le \beta^{N_2} \lambda^{M-N_2} 2^{M-1}.$

and we know that $N_2 \ge N/4$ and $M - N_2 \le 3N/4$. If we substitute these bounds and sum over all $J \subset \{1, 2, \ldots, N\}$, we can use the facts that $\beta < 1$ and that

$$\sum_{M=0}^{N} 2^M \binom{N}{M} = 3^N .$$

to convert (32) to

(33) $P(\Delta(\omega) = \Delta) \quad \le \quad \beta^{N/4} (\lambda \vee 1)^{3N/4} 3^N .$

Now recall that in the coding of Δ, each number can only be followed by two possible numbers, so $|\mathcal{C}_N| = 2^{N-1}$, with N ranging over the possible values 1, 3, 5, ... Thus when we sum (33) over all Δ we have

(34) P(contour exists)

$$\le \sum_{N=1}^{\infty} 2^{N-1} \beta^{N/4} (\lambda \vee 1)^{3N/4} 3^N$$

$$= (6 \beta^{1/4}(\lambda \vee 1)^{3/4}) / 2 (1 - 6 \beta^{1/4}(\lambda \vee 1)^{3/4})$$

provided that $6 \beta^{1/4}(\lambda \vee 1)^{3/4} < 1$. As $\beta \to 0$, the right side of (34) goes to 0 like a constant times $\beta^{1/4}$, and Theorem 2 is proved. □

We now want to obtain the estimate in Corollary 1. For this we need an estimate on the probability that there exists a contour with total upward travel time ≥ T for values of T > 0. Let

$$L(\omega) = \sum_{i \in U} t_i = \text{total upward travel time.}$$

If we restrict the integration in (31) to values of r ≥ T, we get

$$(35) \quad P(\Delta(\omega) = \Delta, J(\omega) = J, L(\omega) \geq T)$$

$$\leq \beta^{N_2} \lambda^{M-N_2} \int_T^\infty dr \exp(-r) \frac{r^{M-1}}{(k-1)!\,(M-k)!} .$$

Again using elementary calculus and simple estimates as in Section 3, the reader can check that for $\gamma > 0$ such that

$$\int_T^\infty dr \ \exp(-r)\, r^{M-1} \leq \exp(-\gamma T) \int_0^\infty dr \exp(-(1-\gamma)r)\, r^{M-1}$$

$$\leq \exp(-\gamma T)\, M! \,/\, (1-\gamma)^M$$

Substituting this estimate into (35) gives

$$(36) \quad P(\Delta(\omega) = \Delta, J(\omega) = J, L(\omega) \geq T)$$

$$\leq \exp(-\gamma T)\, \beta^{N_2} \lambda^{M-N_2}\, 2^{M-1} \,/\, (1-\gamma)^M.$$

As in the end of the proof of Theorem 2, we can sum over J and Δ, this time using the fact that

$$\sum_{M=0}^{N} \frac{2^M}{(1-\gamma)^M} \binom{N}{M} = \left(1 + \frac{2}{1-\gamma}\right)^N = \left(\frac{3-\gamma}{1-\gamma}\right)^N ,$$

to obtain

(37) P(there exists a contour with $L(\omega) \geq T$) $\leq \exp(-\gamma T)\, a / 2(1-a)$,

provided that $a = \beta^{1/4} (\lambda \vee 1)^{3/4} (6 - 2\gamma) / (1 - \gamma) < 1$. This last condition can always be fulfilled by choosing $\gamma > 0$ sufficiently small if $6\beta^{1/4}(\lambda \vee 1)^{3/4} < 1$.

To get from (37) to the Corollary, we first note that since our current graphical representation is an augmentation of the one used in Section 3, we can still use it to construct the process without spontaneous births (simply ignore the points labelled with β's). Let χ^1 be the 1's dual of the process without spontaneous births, and let $\chi^{1,\beta}$ be the 1's dual of the process with spontaneous births at rate β. If we fix $\lambda > \lambda_c$, then the event

$$A = \{\chi^1(-t;\, 0, \{z\}) \neq \emptyset \text{ for all } t \geq 0\}$$

has positive probability, since

$$P(A) = P(z \in \xi^1_\infty) > 0 \text{ for } \lambda > \lambda_c.$$

We also define two more events. Fix $T > 0$, and define

$$B = \{\text{no contour exists with length } L(\omega) \geq T\}$$

$$C = \{\text{no contour exists with length } L(\omega) \leq T\}.$$

These two events refer to the contour constructed in this section in terms of the 1's dual $\chi^{1,\beta}$. All three events are in the same probability space determined by the graphical representation of this section. By (37), $P(B) \to 1$ as $T \to \infty$, so it is legitimate to assume that T is large enough so that $P(A \cap B) > 0$. We claim that for such T,

(38) $P(A \cap B \cap C) > 0$.

Before proving (38), let us first see how Corollary 1 follows. On $B \cap C$, $z \notin \xi(0; -t, \emptyset)$ for all $t \geq 0$ for the process with spontaneous births at rate β. On A, $z \in \xi(0; -t, \mathbb{Z}^2)$ for all $t \geq 0$ for the process without spontaneous births, so the same statement holds true a fortiori for the process with spontaneous births at rate β. Therefore, by letting $t \to \infty$, we can conclude that

$$P(z \in \xi_{\infty}^{1,\beta}) - P(z \in \xi_{\infty}^{0,\beta}) \geq P(A \cap B \cap C).$$

Corollary 1 follows from this and (38).

To prove (38), we will need to first condition on the locations of all the λ's and δ's. Let $\mathcal{G}$ be the σ-algebra determined by these locations. Note that $A \in \mathcal{G}$, since we do not need to know the locations of the β's in order to determine whether or not A occurs. Further note that once we condition on $\mathcal{G}$, the indicators of the events B and C become decreasing functions of the points associated with the β's in the sense that adding β points never increases the values of these functions. This is the sense needed to apply a slightly generalized form of Harris' (1960) correlation inequality (which was the forerunner of the well-known FKG inequality). That inequality implies that B and C are positively correlated, conditioned on $\mathcal{G}$, so we have

(39) $$P(A \cap B \cap C \mid \mathcal{G}) = 1_A\, P(B \cap C \mid \mathcal{G}) \geq 1_A\, P(B \mid \mathcal{G})\, P(C \mid \mathcal{G}) \text{ a.s.}$$

(This is the first place in which we use the fact, mentioned in the "Important Remark" towards the beginning of this section, that contours are allowed which travel around non-minimal paths. The indicator of B would not be a decreasing function of the β points if we restricted our attention to minimal paths.)

We claim that the quantity $P(C \mid \mathcal{G})$ which appears on the right side of (39) is strictly positive a.s. This claim relies on the fact that for almost all realizations of the locations of the λ's and the δ's, there exists a bounded region of the space-time graph such that C occurs if this region does not contain any points labelled with a β. (The proof of this last fact is very similar to the proof given in Section 2 that the state of the process $\xi(t;\, s,\, A)$ at a site z can be determined for almost all ω by looking at a bounded region of space-time). Since the locations of the β's are Poisson distributed, independently of $\mathcal{G}$, our claim follows.

Thus the expected value of the right side of (39) is strictly positive iff $E(1_A\, P(B \mid \mathcal{G})) > 0$. But $E(1_A\, P(B \mid \mathcal{G})) = P(A \cap B)$ since $A \in \mathcal{G}$, so (38) follows from the fact that T has been chosen to make $P(A \cap B) > 0$. □

§ 5. Proof of Theorem 3. In this section we will outline the proof of Theorem 3. The details are very similar to those of Section 4, so we will only cover the main points.

First, we must once again augment the graphical construction to allow for the new types of births. For each $x \in \mathbb{Z}^2$, let A_x be a collection of points which is Poisson distributed with parameter α ("α" for asexual reproduction), with the usual independence assumptions, and add the following rule to the description of the process:

(7'') we mark the points in A_x with α's and interpret the α as telling us that there will be a birth at x if x is vacant and at least one of the neighbors $x + e_1$ or $x + e_2$ is occupied.

As before, this rule, together with our previous rules (6), (7) and (7'), allow us to define the process $\xi(t; s, A)$ and its two dual processes $\chi^1(s; t, B)$ and $\chi^0(s; t, B)$. Note that we are still allowing spontaneous births at rate β. This will be convenient for us in the proof of Theorem 3. The process in Example 3 is obtained by setting $\beta = 0$.

As in Section 4, we will be looking at contours which travel around paths (not necessarily minimal -- recall the "Important Remark" from Section 4) of the l's dual χ^1. With the length of the contour $L(\omega)$ defined as before, we will show that

(37') if $\lambda \geq 1$, $\alpha < 1/144\lambda$, and $\beta \leq \alpha^3$, then there are constants C and γ in $(0, \infty)$ such that for all $T \geq 0$,

$$P(\text{there exists a contour of length } L(\omega) \geq T) \leq C \exp(-\gamma T).$$

(Throughout this section, we will use a numbering system for displayed statements that parallels as much as possible the numbers used in Section 4. We hope this will help the reader to make comparisons. In some cases, it will lead to numberings which are out of sequence.)

Let us first see how Theorem 3 follows from (37'). Consider a sort of hybrid process with starting time $t_0 < 0$ and initial state $\varnothing$, in which we only allow spontaneous births to occur at times $t < 0$ and then set $\beta = 0$ for times $t \geq 0$. Thus, if we restrict our attention to times $t \geq 0$, we simply have the process in Example 3, with an initial state at time $t = 0$ which is determined by what happened in the hybrid process at negative times. Since we are allowing spontaneous births at negative times, the state at time 0 dominates product measure with density p for some p >

0. It is not hard to see that as $t_0 \to -\infty$, we can allow p to go to $\beta/(\beta+1)$.

Now consider the state of the hybrid process at time $T > 0$. We can define its 1's dual in the usual way. If a site x is occupied at time T, then there must be a path from $(\emptyset, s)$ to $(\{x\}, T)$ for some time $s < 0$ (there cannot be such a path for $s \geq 0$ since spontaneous births are not allowed at non-negative times). The contour around such a path must have length $L(\omega) \geq T$. Since any such path will also be a (not necessarily minimal path) for the process in which we allow spontaneous births at all times, it follows from (37') that if α , β and λ satisfy the stated conditions, $P(x \in \xi_t^p) \leq C \exp(-\gamma t)$, for all x. Theorem 3 follows. □

To prove (37'), we use the by now familiar three-step procedure. As stated earlier, we will economize on space and only indicate the modifications needed to make the argument of Section 4 suit the current situation.

Step 1.

The addition of the points marked with α's leads to four new types of turns in the contour. Recall that if the contour is at x and moving upward, it continues until the first time u such that $x \notin \sigma(u)$. It may now happen that $u \in A_x$. In this case,

(Rule 5) if $x + \mathbf{e}_1 \in \sigma(u)$, the contour jumps to $x + \mathbf{e}_1$ and continues going upward; otherwise **(Rule 6)** if $x + \mathbf{e}_1 \notin \sigma(u)$, then the contour jumps to $x + \mathbf{e}_2$ and continues upward.

If the contour is at x and moving downward, as before it moves downward to a time s at which a jump was made by the contour to x from either $x - \mathbf{e}_1$ or $x - \mathbf{e}_2$. This birth may possibly be due to the new kind of Poisson points, so we have the following new rules:

(Rule 7) if the jump was due to rule 4, so that $(x - \mathbf{e}_1, s) \in A_{x - \mathbf{e}_1}$, then the contour jumps to $x - \mathbf{e}_1$ and continues downward; otherwise **(Rule 8)** if the jump was due to rule 5 , so that $(x - \mathbf{e}_2, s) \in A_{x - \mathbf{e}_2}$, then the contour jumps to $x - \mathbf{e}_2$ and continues downward.

The reader should try modifying the example in Figure 5 by changing some of the λ's to α's. It is again easily checked that the contour cannot traverse the same segment twice while going in the same direction.

If we look at rules 1-4 in Section 4 and rules 5-8 above, we see that the only possible values for the triples a_i, b_i, a_{i+1} in the coding sequence are

u1u	u2d	d3u	d4d
u5u			d7d
u6u			d8d

The contour can start with a 1, 5, or 6, and each number can have exactly four followers, so the number of codings of length N is bounded by $3 \cdot 4^{N-1}$.

Step 2.

Using the notation introduced in the last section we can state the new

Basic Estimate:

(27')

$$\frac{d\mu_{\Delta,J}}{d\nu_M} \leq \beta^{N_2} \alpha^{N_{58}} \lambda^{M-(N_2+N_{58})} \cdot \exp(-\sum_{i\in U} t_i)$$

where $N_{58} = |\{j: b_j \in \{5, 6, 7, 8\}, j \notin J\}|$. The proof of the Basic Estimate is the same as in Section 4 and is left to the reader.

The form of the estimate above shows that we need to control the number of 1's, 3's and 4's. As in Section 4, we can argue that

(29') $\quad N_2 = N_3 + 1.$

(Use Figure 6 again, but this time, the point labelled with a "1" should be labelled with "1, 5 or 6", and the point labelled "4" should be labelled "4, 7, or 8".) This equation gives us control over N_3.

To control N_1 and N_4, we will use the "conservation equations" which express the fact that the contour must begin and end at the same point. Consider the following table, analogous to Table 5 of the last section:

Type of turn	Jump made by contour
u1u	(1, 0)
u2d	(0, 0)
d3u	(-1,1)
d4d	(0,-1)
u5u	(1,0)
u6u	(0,1)
d7d	(-1,0)
d8d	(0,-1)

Table 5'

We see from the table above that

$$N_1 - N_3 + N_5 - N_7 = 0 \text{ and } N_3 - N_4 + N_6 - N_8 = 0.$$

Substituting in (29'), we get

$$N_1 + N_5 = N_2 + N_7 - 1 \quad \text{and} \quad N_4 + N_8 = N_2 + N_6 - 1.$$

Combining all these equations yields

$$(28') \quad M - (N_2 + N_{58}) \leq N_1 + N_3 + N_4$$

$$\leq N_1 + N_3 + N_4 + N_5 + N_8$$

$$\leq 3N_2 + N_6 + N_7$$

$$(28'') \quad N = N_1 + \ldots + N_8 \leq 4N_2 + 2N_6 + 2N_7.$$

We also have the observation that

$$(28''') \quad N_{58} \geq N_6 + N_7.$$

which follows from the fact that (i) turns of type 6 always involve Poisson points that have not been previously visited; (ii) turns of type 7 either involve Poisson points that have not been previously visited, or ones that were first visited during a turn of type 5.

Step 3.

Computations similar to those in Section 4 show that when we integrate the Basic Estimate we get

(32') $P(\Delta(\omega) = \Delta, J(\omega) = J)$

$$\leq \beta^{N_2} \alpha^{N_{58}} \lambda^{M-(N_2 + N_{58})} 2^{M-1}$$

If we assume that $1/\alpha \geq \lambda \geq 1$, it follows from (28') - (28''') that the right side of (32') is bounded above by

$$\beta^{N_2} \alpha^{N_6 + N_7} \lambda^{3N_2 + N_6 + N_7} 2^{M-1}$$

Now we sum over J as in Section 4 and use (28''') and the hypothesis that $\beta \leq \alpha^3$:

(33') $P(\Delta(\omega) = \Delta)$

$$\leq \alpha^{3N_2} \alpha^{N_6 + N_7} \lambda^{3N_2 + N_6 + N_7} 3^N \quad \leq \quad (\alpha\lambda)^{N/2} 3^N .$$

Now recall that there are at most $3 \cdot 4^{N-1}$ codings of length N. Thus if we sum (33') over N, we obtain

(34') P(contour exists) $\leq \; 3\,(12\sqrt{\alpha\lambda}\,) \,/\, 4\,(1 - 12\sqrt{\alpha\lambda}\,)$,

provided $12\sqrt{\alpha\lambda} < 1$, i.e. $\alpha < 1/144\lambda$.

Getting from here to the exponential estimate (37') is analogous to getting from the proof of Theorem 2 to the exponential estimate in Corollary 1. The relevant inequalities are found in (35) - (37). The details are left to the reader. □

Bibliography

1. Bramson, M. and Gray, L. (1985). The survival of branching annihilating random walk. Z. Wahrscheinlichkeitstheorie verw. Gebiete **68**, 447-460.

2. Durrett, R. (1981). An introduction to infinite particle systems. Stoch. Proc. Appl. **11,** 109-150.

3. Durrett, R. (1984). Oriented percolation in two dimensions. Ann. Probab. **12**, 999-1040.

4. Durrett, R. and Griffeath, D. (1983). Supercritical contact processes on $\mathbb{Z}$. Ann. Probab. **11**, 1-15.

5. Gray, L. (1986). Duality for general attractive spin systems with applications in one dimension. Ann. Probab. **14**, 371-396.

6. Gray, L. and Griffeath, D. (1982). A stability criterion for attractive nearest neighbor spin systems on $\mathbb{Z}$. Ann. Probab. **10**, 67-85.

7. Griffeath, D. (1979). Additive and Cancellative Interacting Particle Systems. Lecture Notes in Math. **724**. Springer-Verlag, New York.

8. Harris, T. (1960). A lower bound for the critical probability in a certain percolation process. Proc. Cambridge Philos. Soc. **56**, 13-20.

9. Harris, T. (1978). Additive set-valued Markov processes and graphical methods. Ann. Probab. **6,** 355-378.

10. Holley, R. (1972) An ergodic theorem for interacting particle systems with attractive interactions. Z. Wahrscheinlichkeitstheorie verw. Gebiete **24**, 325-334.

11. Holley, R. and Liggett, T. (1978). The survival of contact processes. Ann. Probab. **6**, 198-206.

12. Liggett, T. (1983). Attractive nearest particle systems. Ann. Probab. **11**, 16-33.

13. Liggett, T. (1985). Interacting Particle Systems. Springer-Verlag, New York.

14. Peierls, R. (1936). On Ising's model of ferromagnetism. Proc. Cambridge Philos. Soc. **32**, 477-481.

15. Toom, A. (1979). Stable and attractive trajectories in multicomponent systems. Adv. in Probability **6**, 549-575. Dekker, New York.