The Annals of Applied Probability
1999, Vol. 9, No. 2, 331–351

THE TWO-STAGE CONTACT PROCESS[1]

BY STEPHEN M. KRONE

University of Idaho

We introduce a multitype interacting particle system, which is a natural generalization of the contact process. Here, individuals in the population have two life stages, young and adult. Only adults can give birth and each new offspring is young. Transition from young to adult occurs at constant rate, and individuals die at rates that depend on their life stage. Important to the analysis of this process is the construction of a multitype dual process.

1. Introduction. The contact process was introduced by Harris (1974). It is a continuous-time Markov process whose state at time t is $\xi_t \in \{0,1\}^{\mathbb{Z}^d}$; that is, for each $x \in \mathbb{Z}^d$, $\xi_t(x) \in \{0,1\}$ represents the state of site x at time t. The dynamics are as follows. Let $\mathscr{N}(x) = \{y \in \mathbb{Z}^d\colon 0 < \|y-x\|_p \le r\}$ be the set of neighbors of site x. [In general, one can use any L^p norm with $1 \le p \le \infty$ in the definition of $\mathscr{N}(x)$.] If the current configuration is ξ, then the contact process has the following transition rates at x:

$$
\begin{aligned}
&0 \to 1\text{: rate } \lambda n_1(x),\\
&1 \to 0\text{: rate } 1,
\end{aligned}
\tag{1.1}
$$

where $n_1(x) = n_1(x,\xi) = \#\{y \in \mathscr{N}(x)\colon \xi(y) = 1\}$. A site in state 0 is thought of as being "vacant," while a site in state 1 is "occupied." Perhaps the most important result about the contact process is the existence of a critical birth rate $\lambda_c \in (0,\infty)$, such that

$$
\begin{aligned}
\mathbb{P}\big(\xi_t^0 \not\equiv 0\ \forall\ t\big) &= 0 \quad \text{if } \lambda \le \lambda_c,\\
\mathbb{P}\big(\xi_t^0 \not\equiv 0\ \forall\ t\big) &> 0 \quad \text{if } \lambda > \lambda_c,
\end{aligned}
$$

where ξ_t^0 is the contact process starting with the origin in state 1 and all other sites in state 0. We refer the reader to Liggett (1985) and Durrett (1988) for information on the contact process.

We define the two-stage contact process to be a continuous-time Markov process whose state at time t is $\xi_t \in \{0,1,2\}^{\mathbb{Z}^d}$ and whose dynamics are as follows. If the current configuration is ξ, then the transition rates at site x

Received March 1998; revised May 1998.
[1]Supported in part by NSF Grant DMS-96-26764 and a University of Idaho Seed Grant.
AMS 1991 *subject classification*. 60K35.
Key words and phrases. Interacting particle system, contact process, multitype, dual process, survival.

are given by

$$\begin{aligned} &0 \to 1\colon \text{rate } \lambda n_2(x),\\ &1 \to 2\colon \text{rate } \gamma,\\ &1 \to 0\colon \text{rate } 1+\delta,\\ &2 \to 0\colon \text{rate } 1, \end{aligned} \tag{1.2}$$

where $n_2(x) = n_2(x, \xi) = \#\{y \in \mathscr{N}(x)\colon \xi(y) = 2\}$ and $\mathscr{N}(x)$ is as above. Here, we interpret a site in state 0 as being "vacant," state 1 as being occupied by a "young" individual and state 2 as being occupied by an "adult." (A site will be called "occupied" if it is in state 1 or 2.) Only adults can give birth, and they do so onto a given neighboring site at rate λ. (Births onto occupied sites are suppressed, as in the contact process.) Each new offspring is young and the transition from young to adult occurs at rate γ. Adults die at rate 1, whereas young individuals die at rate $1+\delta$, the added rate $\delta \geq 0$ being thought of as some type of "infant mortality." If $\gamma = \infty$, we recover the contact process with state space $\{0,2\}^{\mathbb{Z}^d}$.

For disjoint sets $A, B \subseteq \mathbb{Z}^d$, we will let $\xi_t^{A(1),\,B(2)}$ denote the two-stage contact process whose initial state is given by

$$\xi_0^{A(1),\,B(2)}(x) = \begin{cases} 1, & \text{if } x \in A,\\ 2, & \text{if } x \in B,\\ 0, & \text{if } x \notin A \cup B. \end{cases} \tag{1.3}$$

We will sometimes write $A(1) \cup B(2)$ for this initial configuration. Similarly, write $\xi_t^{A(1)}$ (resp., $\xi_t^{A(2)}$) when the process starts with all sites in A occupied by 1's (resp., 2's) and all other sites vacant.

Another interpretation of this model (in fact, what originally led us to consider the two-stage contact process) is as follows. Consider a stochastic spatial model of metapopulation dynamics which takes into account demographic as well as environmental extinctions. Here, each site represents an island or habitat patch and can be in one of three states: 0 = "vacant," 1 = "partially occupied," and 2 = "fully occupied." The dynamics are given by the following:

1. 1's and 2's are killed at rate δ_e (environmental extinctions);
   1's have an additional death rate δ (demographic effects);
2. 1's become 2's at rate γ;
3. A 0 at x becomes a 1 at rate $\lambda n_2(x)$.

A site in state 1 represents a small colony at risk of going extinct due to demographic (rate δ) as well as environmental (rate δ_e) factors. If it survives this vulnerable period, it can grow to a full colony (rate γ). Only full colonies can send migrants to another site. A more realistic model would have the colony size fluctuating according to some Markov chain. A routine calculation would then give the probability that the colony reaches full size before going extinct. In some populations [cf. Hanski and Zhang (1993)], the within-colony dynamics operate on a much faster time scale than the metapopulation dynamics (migration). This assumption of a faster time scale for the local dynamics simplifies the model. At the metapopulation time scale, we do not see the fluctuations in colony size, only the transitions between "partially full," "full" and "vacant." A site in state 2 represents a full colony. Such a colony is assumed to be large enough to be in no danger of demographic extinction; it can only experience extinction due to environmental effects. A full colony sends a small number of migrants to a neighboring vacant site at rate λ. The small number of emigrants does not diminish the size of the colony enough to increase its risk of extinction. When $\delta_e = 1$ (which can be assumed without any harm by simply changing the time scale), the above model is just the two-stage contact process.

The above model considers the noncompetitive population dynamics for a single species. One can also imagine a host of other spatial ecological models, perhaps involving competition between several species, in which the birth rate for one type depends on the local density of another type (where types might model life stages, various stages of infection, etc.). Embedded in such a model would be the birth dynamics of the two-stage contact process or something similar. One of our hopes is that an understanding of the relatively simple two-stage contact process will help us deal with some of these more complex models in spatial ecology.

2. Graphical representation and dual process.

2.1. *Graphical representation.* As for the contact process, the two-stage contact process can be constructed graphically via families of independent Poisson processes. To get this graphical representation, we begin with the space–time diagram $\mathbb{Z}^d \times \mathbb{R}$. This is augmented to give a *percolation diagram* $\mathscr{P}$ as follows. For each $x \in \mathbb{Z}^d$, consider the arrival times of independent families of Poisson processes $\{U_n^x\colon n \geq 1\}$ with rate 1, $\{V_n^x\colon n \geq 1\}$ with rate δ, $\{W_n^x\colon n \geq 1\}$ with rate γ, and, for $y \in \mathscr{N}(x)$, $\{T_n^{x,y}\colon n \geq 1\}$ with rate λ. At the space–time points (x, U_n^x) in $\mathscr{P}$, put an $\times$ to indicate that a death will occur if x is occupied by a 1 or a 2; at (x, V_n^x) put a δ to indicate that a death will occur if x is occupied by a 1; at (x, W_n^x) put a 2-dot ($\bullet$) to indicate that if there is a 1 at x it will turn into a 2; finally, from $(y, T_n^{x,y})$ to $(x, T_n^{x,y})$ draw an arrow to indicate that, if y is occupied by a 2, there will be a birth of a 1 at x (if it is not already occupied).

For $s < t$, we say there is a *path* from (x, s) to (y, t) in $\mathscr{P}$ if there is a sequence of times $s = s_0 < s_1 < s_2 < \cdots < s_n < s_{n+1} = t$ and a corresponding sequence of spatial locations $x = x_0, x_1, x_2, \ldots, x_n = y$ such that:

1. For $i = 1, 2, \ldots, n$, there is an arrow from x_{i-1} to x_i at time s_i.
2. The vertical segments $\{x_i\} \times (s_i, s_{i+1})$, $i = 0, 1, \ldots, n$, do not contain any $\times$'s.

[That is, a path is a chain of upward vertical and directed horizontal edges in $\mathscr{P}$ which leads from (x, s) to (y, t) without passing (vertically) through an $\times$.] For such a path, we will call (x, s) the *initial point* and (y, t) the *terminal point*. We say the above path is *active* if it also satisfies the following:

3. Each of the vertical segments $\{x_i\} \times (s_i, s_{i+1})$, $i = 1, \ldots, n-1$, contains at least one 2-dot which is below any δ's that happen to be in the segment.
4. If the top vertical segment $\{x_n\} \times (s_n, s_{n+1})$ contains any δ's, then it contains a 2-dot which lies below them all.

(Thus a path is active if it contains enough 2-dots to allow passage through any δ's in the path and to allow all the arrows above the lowest arrow in the path to be used. Whether or not the lowest arrow can be used will depend on whether the path is "compatible with the initial configuration," in a sense to be described next.)

In the case where $s = 0$, we will say the above path from $(x, 0)$ to (y, t) has a *strong* initial point (resp., a strong terminal point) if there is a 2-dot in the lowest vertical segment $\{x\} \times (0, s_1)$ (resp., in the highest vertical segment $\{y\} \times (s_n, t)$) which is below any δ's that happen to be in the segment. An initial point (resp., a terminal point) which is not strong is called *weak*. We say that a path from $(x, 0)$ to (y, t) is *compatible with the initial configuration* $\xi_0^{A(1), B(2)} = A(1) \cup B(2)$ if it has a strong initial point in $A \cup B$ or a weak initial point in B. In other words, such a path is "strong enough" to allow the type coming in at $(x, 0)$ to start up the path and use the first arrow in the path. Note that, if there is an active path from $(x, 0)$ to (y, t) which is compatible with the initial configuration, then site y will be occupied at time t, although it might not use this particular path to get its type.

An example of a realization of the percolation diagram (when $d = 1$), together with some paths, is given in Figure 1. Since we have finite-range births, an idea due to Harris (1972) allows one to get around the fact that

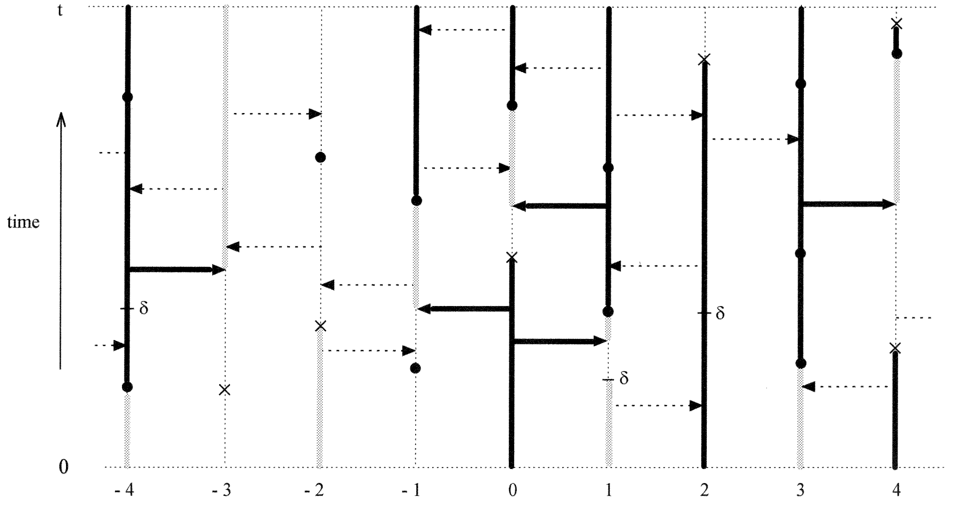

FIG. 1. *Realization of process starting with sites* $-4, -2, 1, 3$ *in state* 1 *and sites* 0, 2, 4 *in state* 2. *The solid gray rule denotes type* 1 ("*blue*"); *the solid black rule denotes type* 2 ("*red*").

there are an infinite number of sites, and hence no first birth, to construct the process for all times starting from any initial configuration. To get the configuration at time t when the initial configuration is $A(1) \cup B(2)$, imagine that there is blue fluid entering the bottom of the percolation diagram at all sites in A and red fluid entering at all sites in B. The fluids flow up the diagram in the direction of increasing time. (All existing fluid will be at height t at time t.) In addition to flowing upward, red fluid can flow across arrows in the indicated directions, but when it comes out at the tip of the arrow it turns blue (unless the arrow lands on a line that is already carrying fluid, in which case the flow across the arrow is stopped). Blue fluid cannot cross an arrow. If blue fluid encounters a δ, it is blocked; δ's have no effect on red fluid. If fluid of either color encounters an $\times$, it is blocked. Finally, if blue fluid passes through a 2-dot, it turns red; red fluid is not affected by 2-dots. The color of the fluid (if any) at (y, t) determines the type $\xi_t^{A(1), B(2)}(y)$; 1 if blue, 2 if red and 0 if there is no fluid. Clearly the resulting process is the two-stage contact process. In Figure 1, $A = \{-4, -2, 1, 3\}$ is the initial set of 1's and $B = \{0, 2, 4\}$ is the initial set of 2's; all other sites start in state 0. At the top of the graph we can read off the types at time t: $\xi_t^{A(1), B(2)}(y) = 1$ for $y = -3$, $\xi_t^{A(1), B(2)}(y) = 2$ for $y \in \{-4, -1, 0, 1, 3\}$ and all other sites are in state 0 at time t.

2.2. *Dual process*. To construct the dual process $\hat{\xi}_t$ for the two-stage contact process, we will begin by defining dual paths in the time-honored fashion by running time backwards in $\mathscr{P}$ and traversing the arrows in the direction which is opposite to their orientation. The collection of dual paths out of a given site will determine a set of "ancestors" at any time. These ancestors will then be arranged according to the order in which they determine the type of that site. Since we are dealing with a multitype process in which individuals can change their type while moving along a path, it turns out that we will also need to have a way of assigning types to the ancestors at any time. This can be thought of as coloring the dual paths in a way that is related to the coloring of the forward paths. The next step is to show how to combine the duals from several sites to compute probabilities of interest. In Section 2.3 we will see how to alter this dual to get a slightly simpler dual process.

For $s < t$, we say there is a *dual path* from (y, t) down to (x, s) if there is a sequence of times $0 = s_0 < s_1 < s_2 < \cdots < s_n < s_{n+1} = t - s$ and a corresponding sequence of spatial locations $y = x_0, x_1, x_2, \dots, x_n = x$ such that:

1′. For $i = 1, 2, \dots, n$, there is an arrow in $\mathscr{P}$ from $(x_i, t - s_i)$ to $(x_{i-1}, t - s_i)$.
2′. The vertical segments $\{x_i\} \times (t - s_{i+1}, t - s_i)$, $i = 0, 1, \dots, n$, do not contain any $\times$'s.

The above dual path is said to be *active* if it also satisfies:

3′. Each of the vertical segments $\{x_i\} \times (t - s_{i+1}, t - s_i)$, $i = 1, \dots, n - 1$, contains a 2-dot which is below any δ's that happen to be in the segment.
4′. If the top vertical segment $\{x_0\} \times (t - s_1, t)$ contains any δ's, then it must contain a 2-dot which lies below them all.

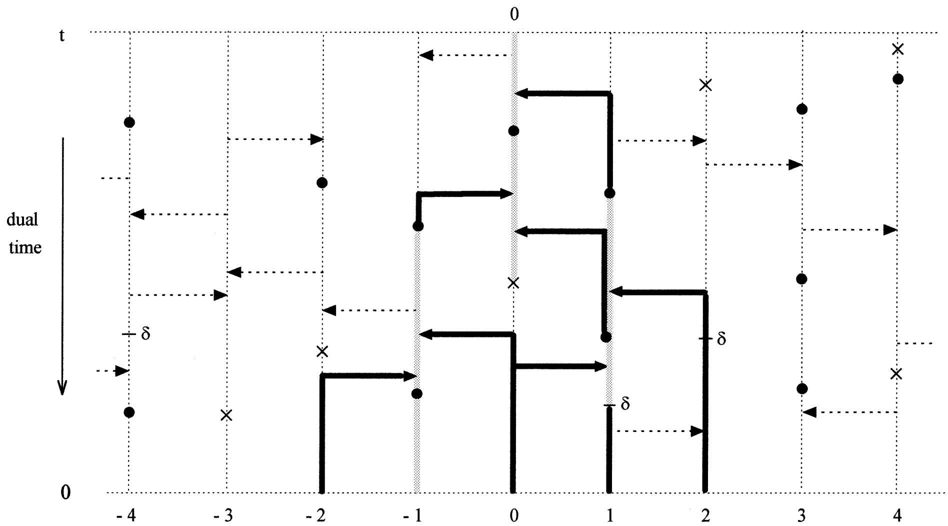

FIG. 2. *Dual process out of site* 0, *starting in state* 2. *The solid black rule denotes type* 1 ("*red*"); *the solid gray rule denotes type* 2 ("*blue*").

Some examples of dual paths can be seen in Figure 2. This graph will be explained in more detail as we proceed.

Note that there is an active path from (x, s) to (y, t) in $\mathscr{P}$ if and only if there is an active dual path from (y, t) to (x, s). In a dual path from (y, t) down to (x, s), we refer to (y, t) as the *initial* (*dual*) *point* and (x, s) as the *terminal* (*dual*) *point*. A dual point has a strong terminal point if there is a 2-dot in the lowest vertical segment $\{x\} \times (s, t - s_n)$ which is below any δ's that happen to be in the segment. A terminal point which is not strong is called weak. In Figure 2, the space–time point $(-1, 0)$ is a strong terminal point for a dual path and $(1, 0)$ is a weak terminal point. We say that a dual path is *compatible with the configuration* $A(1) \cup B(2)$ if it has a strong terminal point in $A \cup B$ or a weak terminal point in B.

Clearly, there can be more than one dual path coming down out of (y, t). The *dual process* $\{\hat{\xi}_s^{(y,t)}: 0 \le s \le t\}$ starting at (y, t) will consist of an ordered arrangement of all such dual paths out of (y, t), together with a way to characterize the type of each dual path at any given time. (Time s in the dual process corresponds to $t - s$ in $\mathscr{P}$.) For $0 \le s \le t$, let

$$\mathscr{A}\big(\hat{\xi}_s^{(y,t)}\big) \equiv \{x: \text{there is an active dual path from } (y, t) \text{ down to } (x, t-s)\}$$

be the set of *ancestors* of (y, t) at (dual) time s. These are the locations s units of time in the past from which (y, t) might have gotten its type. Note that a site $x \in \mathscr{A}(\hat{\xi}_s^{(y,t)})$ might be the terminal point for several dual paths in $\hat{\xi}_s^{(y,t)}$. [In Figure 2, $\mathscr{A}(\hat{\xi}_t^{(0,t)}) = \{-2, -1, 0, 1, 2\}$.] Next, we arrange the set of ancestors according to the *order* in which they determine the state at (y, t). For this we follow Neuhauser (1992), where a different kind of multitype contact process was considered.

To order the ancestors at time 0 (dual time t), start at (y, t) and go down the percolation diagram until the first time an $\times$ is encountered. If no $\times$'s are encountered, then y is the first ancestor (at time 0). If we encounter an $\times$, go back up the diagram until the first time the tip of an arrow is encountered. Follow this arrow backward to the vertical line the arrow is attached to. Then go down this line until another $\times$ is encountered and repeat the above procedure. Stop when time 0 is reached. The corresponding site is the first ancestor and there is a unique path from this first ancestor up to (y, t) that only uses parts of the percolation diagram traversed during the above procedure. A moment's thought shows that, if this path is compatible with the initial configuration (equivalently, if the corresponding dual path is compatible with the initial configuration), then it determines the state at (y, t). The state will be 2 if the upward path has a strong terminal point and is 1 if the terminal point is weak. If the resulting path from the first ancestor is not compatible with the initial configuration, then we need the second ancestor. To find this, go up, starting at the first ancestor along its path until the tip of an arrow is encountered. Follow this arrow backward and then go down the vertical line from which it originated, repeating the algorithm used to find the first ancestor. When you reach time 0, you will be at the site of the second ancestor. Continue in this way until you have all the ancestors (i.e., there are no more arrow tips to reach).

There are only a finite number of ancestors for any finite t. It can happen that a given ancestral site will appear several times in the ordering, but each has a unique forward path (and a unique dual path) associated with it. To determine the type at (y, t), go through the ordered set of ancestors until you find the first one which is compatible with the initial configuration. Call this ancestor the *determining ancestor* and the corresponding path the *determining path* for (y, t). If none of the ancestors is compatible with the initial configuration, then the state at (y, t) will be 0. If there is a determining ancestor, give it the type (color) assigned by the initial configuration and follow the determining path up to (y, t), changing the type along the path according to the rules of the process. The color at the top of the determining path gives the type at (y, t). An example of the ancestral hierarchy can be obtained from the realization of the dual process in Figure 2. In this example, the ordered set of ancestors is $1, 0, 2, -1, -2, 0, 1, 0, 2$.

Now that we have described the ancestors of a given point (y, t), we show *how to assign types to the ancestors* to get a dual process with values in $\{0, 1, 2, \{1, 2\}\}^{\mathbb{Z}^d}$. This can be thought of as a collection of colored paths growing down out of (y, t), where the *dual coloring* is related to, but not the same as, that of the forward process. The type (color) assigned to a point on a dual path will be determined by the "minimum type" needed at that point to go successfully "up" the path. For example, after a dual birth (i.e., when the dual path goes backward across an arrow) we color the vertical segment coming down out of the base of the arrow with red. This is because we need the forward path to be red (type 2) at that point in order to use the birth arrow. We will change the color along a dual path from red to blue if the path crosses a 2-dot. This is because the minimum color needed below a 2-dot is blue (type

1); if we have a 1 or 2 at this point, it will turn into a 2 after going up through the 2-dot, and can then give birth across any arrow it encounters. A red dual path cannot branch; that is, we cannot follow an arrow backward out of a red segment. For, if we did, the corresponding forward path would have a birth that does not become strong enough to use the next arrow. A blue dual path can branch; for example, if it became blue by going down through a 2-dot, then the corresponding forward path can accept a birth and change to a 2 at the 2-dot before reaching the next birth arrow. If a dual path crosses a δ, it turns red. For, if the type going forward were not at least a 2, it would be killed by the δ. As before, dual paths (of any color) are killed by $\times$'s.

A quick review of the above dual coloring shows that, in the dual, red is the "weak" color (cannot branch, etc.) and blue is the "strong" color. For this reason, red in the dual will be called type 1 and blue will be called type 2. So the coloring for weak and strong is different in the forward process and the dual. The types are the same in both processes in the sense that 1 designates weak and 2 designates strong.

Thus, to get the *dual process out of* (y, t), we start (y, t) with some color corresponding to its type (which will be determined by what we are looking for—more on this later). Proceed with this color down the percolation diagram in the direction of increasing dual time. The path stops if it hits an $\times$, no matter what color the path is at that time. If the path encounters a 2-dot, the color becomes blue (type 2). A blue path turns red (type 1) when it crosses a δ and a red path is unchanged by a δ. If a blue path encounters the tip of an arrow, it produces a red path starting at the space–time point from which the arrow originated and proceeding down, and the original blue path continues down as well. If a red path encounters the tip of an arrow, it just continues down and does not give birth across the arrow. If two of these paths come together (as the result of a birth), the newly formed path will be red and the path it runs into can be either color. If they are both red, they coalesce into a single red path; a 2 coming up at that point would be able to use both red paths. If the other path is blue, they remain two separate paths at the same site (this gives rise to type $\{1, 2\}$ at the site). If such a pair of paths encounters the tip of an arrow, only the blue path branches. If the pair of paths encounters a 2-dot, they coalesce into one blue path (i.e., $\{1, 2\}$ becomes 2). Also, if $\{1, 2\}$ encounters a δ it turns into a single red (1) path. Finally, if a (red) dual birth lands on a site in state $\{1, 2\}$, the red paths coalesce as before, resulting in state $\{1, 2\}$ again.

Once we specify the initial state at (y, t), this determines a $\{0, 1, 2, \{1, 2\}\}^{\mathbb{Z}^d}$-valued process $\hat{\xi}_s^{(y,t)}$, where 0 means vacant. As usual [cf. Durrett (1988)], it is possible to define a dual process $\hat{\xi}_s^y$ for all $s \geq 0$ in terms of its own percolation diagram and which has the same distribution as our original dual on the interval $[0, t]$:

$$\left\{\hat{\xi}_s^y : 0 \leq s \leq t\right\} =_{\text{law}} \left\{\hat{\xi}_s^{(y,t)} : 0 \leq s \leq t\right\}.$$

We will employ this version of the dual freely and without further comment. Figure 2 gives a realization of the dual process $\hat{\xi}_t^{0(2)}$ out of site 0, starting with state 2. The dual process $\hat{\xi}_t^{0(1)}$, starting with state 1 at site 0, would not have the dual path that starts with the top arrow. The rest would be the same (except for the initial color difference).

When *combining duals out of different sites*, the rule is to coalesce two lines when they meet if they have the same color (just as for different dual paths out of the same point); if a red line encounters a blue line, they stay separate (only the blue line can branch) until they are both killed by an $\times$, or cross a δ and coalesce into a single red line, or cross a 2-dot and coalesce into a single blue line. We write $\hat{\xi}_t^{C(1),\,D(2)}$ for the dual process coming down out of $C \cup D$, where all dual paths starting in C begin with type 1 (red), and all dual paths starting in D begin with type 2 (blue). This process again takes values in $\{0, 1, 2, \{1, 2\}\}^{\mathbb{Z}^d}$ since each site can accommodate a red and a blue line. Of course, each point of $C \cup D$ will have its own distinct ancestral hierarchy embedded in the overall dual $\hat{\xi}_t^{C(1),\,D(2)}$.

We are now ready to describe the role played by the *initial configuration for the dual process*. We want to be able to write all probabilities of interest, such as

$$\begin{aligned}(2.1)\qquad &\mathbb{P}\big(\xi_t^{A(1),\,B(2)}(x) = 0 \text{ for all } x \in E,\\ &\quad \xi_t^{A(1),\,B(2)}(y) = 1 \text{ for all } y \in F,\ \xi_t^{A(1),\,B(2)}(z) = 2 \text{ for all } z \in G\big),\end{aligned}$$

for any finite disjoint E, F, G, in terms of the dual process. Using an inclusion–exclusion argument, it is not hard to see that such probabilities can be written in terms of probabilities of the form

$$\begin{aligned}(2.2)\qquad &\mathbb{P}\big(\xi_t^{A(1),\,B(2)}(x) = 2 \text{ for some } x \in C\\ &\quad \text{or } \xi_t^{A(1),\,B(2)}(y) \neq 0 \text{ for some } y \in D\big),\end{aligned}$$

where C and D are finite and disjoint. Thus it is enough to be able to write (2.2) in terms of the dual.

To get $\xi_t^{A(1),\,B(2)}(x) \neq 0$, we need at least one active path in $\mathscr{P}$ from $(A \cup B, 0)$ to (x, t), which is compatible with the initial configuration $A(1) \cup B(2)$. This is equivalent to having at least one active path in $\hat{\xi}_t^{x(2)}$ which is compatible with $A(1) \cup B(2)$. We start the dual in state $x(2)$, that is, with a 2 at site x and 0's elsewhere, because any forward path which brings a 1 or 2 to site x will be enough to cause x to be occupied at time t. In other words, all dual paths which branch out of the top vertical segment can be used. If site x is occupied at time t, the actual type of the particle there can be found by using the determining path. (This last statement will not be true for the "suppressed dual" in Section 2.3.)

To get $\xi_t^{A(1),\,B(2)}(x) = 2$, we need to be more restrictive. Here, we need to have x occupied, but the determining path must carry type 2 to x; that is, it must cross a 2-dot in the segment leading up to (x, t). This can be achieved by starting the dual in state $x(1)$; $\hat{\xi}_t^{x(1)}$ only uses branches out of the top segment which are below the top 2-dot (if there is one). Certainly x will be occupied at time t if there is at least one active dual path in $\hat{\xi}_t^{x(1)}$ which is compatible with $A(1) \cup B(2)$. To see that the type at x will actually be 2, the ancestral ordering plays a key role. Note that a dual path which follows a lower branch out of the top vertical segment has higher priority than a path which follows a higher branch. Thus, if the path going through the lower branch is occupied, the path going through the higher branch will not be used. In particular, any path in $\hat{\xi}_t^{x(1)}$ will have priority over a path that would branch from a spot above the highest 2-dot in the top segment (and hence could possibly give x type 1). Thus $\xi_t^{A(1),\,B(2)}(x) = 2$ if and only if there is at least one active path in $\hat{\xi}_t^{x(1)}$ which is compatible with $A(1) \cup B(2)$.

Finally, a slight extension of this argument shows that the event

$$\left\{\xi_t^{A(1),\,B(2)}(x) = 2 \text{ for some } x \in C \text{ or } \xi_t^{A(1),\,B(2)}(y) \neq 0 \text{ for some } y \in D\right\}$$

will occur if and only if there is at least one active path in $\hat{\xi}_t^{C(1),\,D(2)}$ which is compatible with $A(1) \cup B(2)$.

DEFINITION 2.1. We write $\hat{\xi}_t^{C(1),\,D(2)} \sim A(1) \cup B(2)$ if there is at least one active dual path in $\hat{\xi}_t^{C(1),\,D(2)}$ which is compatible with $A(1) \cup B(2)$. This is equivalent to

$$\begin{aligned} &\hat{\xi}_t^{C(1),\,D(2)}(x) = 2 \text{ or } \{1,2\} \text{ for some } x \in A \text{ or} \\ &\hat{\xi}_t^{C(1),\,D(2)}(y) \neq 0 \text{ for some } y \in B. \end{aligned} \tag{2.3}$$

We can now summarize the above discussion with the following duality equations.

THEOREM 2.2. *If $A, B, C, D \subseteq \mathbb{Z}^d$ with $A \cap B = \varnothing$ and $C \cap D = \varnothing$, then*

$$\begin{aligned} &\mathbb{P}\left(\xi_t^{A(1),\,B(2)}(x) = 2 \textit{ for some } x \in C \textit{ or } \xi_t^{A(1),\,B(2)}(y) \neq 0 \textit{ for some } y \in D\right) \\ &\quad = \mathbb{P}\left(\hat{\xi}_t^{C(1),\,D(2)} \sim A(1) \cup B(2)\right) \\ &\quad = \mathbb{P}\Big(\hat{\xi}_t^{C(1),\,D(2)}(x) = 2 \\ &\qquad\qquad \textit{or } \{1,2\} \textit{ for some } x \in A \textit{ or } \hat{\xi}_t^{C(1),\,D(2)}(y) \neq 0 \textit{ for some } y \in B\Big). \end{aligned} \tag{2.4}$$

In particular,

$$\mathbb{P}\left(\xi_t^{A(1),\,B(2)}(x) \neq 0 \textit{ for some } x \in C\right) = \mathbb{P}\left(\hat{\xi}_t^{C(2)} \sim A(1) \cup B(2)\right); \tag{2.5}$$

$$\mathbb{P}\left(\xi_t^{A(1),\,B(2)}(x) = 2 \textit{ for some } x \in C\right) = \mathbb{P}\left(\hat{\xi}_t^{C(1)} \sim A(1) \cup B(2)\right). \tag{2.6}$$

2.3. *The suppressed dual.* We now show that, when computing probabilities of the form (2.2), we can use a slightly simpler dual obtained by suppressing any branchings (dual births) that land on an occupied site. Although this will remove some of the dual paths, it will not remove any that matter. This is because the only thing needed to get an event of the form (2.2) is for *at least* one dual path to make it down to time 0 and be compatible with the initial configuration. The ancestral ordering is obtained in the same way as before, and the relative order of the ancestors which are not removed will be the same as before. This new dual will be denoted by $\tilde{\xi}_t$. It is constructed from the same percolation diagram as before, but the rules change slightly. In this case, whenever a (red) dual path lands on a site which is already occupied by another dual path, this new dual birth is suppressed. For this reason, we will refer to $\tilde{\xi}_t$ as the *suppressed dual*. All other transitions occur just as for $\hat{\xi}_t$. Thus there is no need for the state $\{1, 2\}$ in this dual process. In other words, $\tilde{\xi}_t$ takes values in $\{0, 1, 2\}^{\mathbb{Z}^d}$, just as for the two-stage contact process. Note that the set of ancestors $\mathscr{A}(\tilde{\xi}_t^{A(1), B(2)})$ in the suppressed dual is the same as the set of ancestors $\mathscr{A}(\hat{\xi}_t^{A(1), B(2)})$ in the original dual; we do not lose any ancestors by suppressing dual births onto occupied sites. The suppressed dual process corresponding to the dual in Figure 2 would be obtained by suppressing the third dual birth out of site 0. Otherwise, it looks the same.

It is important to realize that starting the suppressed dual in state 2 at some site x is used only to determine whether or not x will be *occupied* at time t. This, by itself, will not tell us the type of the particle. In the original dual, we could determine whether the occupied site was in state 1 or 2 by simply following the determining path. In the suppressed dual, this is not the case, since we are removing certain paths. For example, in Figure 2, if the initial configuration at the bottom of the graph has a 2 at site 1 and 0's at all other sites, then the determining path in the original dual, $\hat{\xi}_t^{0(2)}$, shows that site 0 will have state 2 at time t. If we tried the same thing with the suppressed dual, it would appear that the state at site 0 should be 1. The problem is that, in simplifying our original dual to get the suppressed dual, we had to throw out certain information. The suppressed dual, $\tilde{\xi}_t^{0(2)}$, cannot (by itself) distinguish between the two types of particles at the top. This, however, is where the inclusion–exclusion argument comes to the rescue. It tells us we need only consider probabilities of the form (2.2); all others can be obtained from these. For example,

$$\mathbb{P}(\xi_t(x) = 1) = \mathbb{P}(\xi_t(x) \neq 0) - \mathbb{P}(\xi_t(x) = 2).$$

From the above discussion, it follows that the duality equation (2.4) holds with $\tilde{\xi}_t$ in place of $\hat{\xi}_t$ and state $\{1, 2\}$ removed:

$$\begin{aligned} &\mathbb{P}\big(\xi_t^{A(1), B(2)}(x) = 2 \text{ for some } x \in C \text{ or } \xi_t^{A(1), B(2)}(y) \neq 0 \text{ for some } y \in D\big) \\ &\quad = \mathbb{P}\big(\tilde{\xi}_t^{C(1), D(2)} \sim A(1) \cup B(2)\big) \\ &\quad = \mathbb{P}\big(\tilde{\xi}_t^{C(1), D(2)}(x) = 2 \text{ for some } x \in A \\ &\qquad\quad \text{or } \tilde{\xi}_t^{C(1), D(2)}(y) \neq 0 \text{ for some } y \in B\big) \end{aligned} \tag{2.7}$$

holds for $A, B, C, D \subseteq \mathbb{Z}^d$ with $A \cap B = \varnothing$ and $C \cap D = \varnothing$. In fact, we have

$$\big\{\hat{\xi}_t^{C(1), D(2)} \sim A(1) \cup B(2)\big\} = \big\{\tilde{\xi}_t^{C(1), D(2)} \sim A(1) \cup B(2)\big\}.$$

It is easy to see (cf. discussion in Section 2.2) that the suppressed dual $\tilde{\xi}_t$ has dynamics given in terms of the following transition rates at x:

$$
\begin{aligned}
&0 \to 1\colon \text{rate } \lambda n_2(x),\\
&1 \to 2\colon \text{rate } \gamma,\\
&2 \to 1\colon \text{rate } \delta,\\
&1 \to 0\colon \text{rate } 1,\\
&2 \to 0\colon \text{rate } 1.
\end{aligned}
\tag{2.8}
$$

In particular, when $\delta = 0$, the suppressed dual has the same distribution as the two-stage contact process, so we get the following result.

THEOREM 2.3. *When $\delta = 0$, the two-stage contact process is self-dual in the sense that $\tilde{\xi}_t^{C(1), D(2)}$ has the same distribution as $\xi_t^{C(1), D(2)}$. In particular*, (2.7) *can be written as*

$$
\begin{aligned}
&\mathbb{P}\big(\xi_t^{A(1), B(2)}(x) = 2 \textit{ for some } x \in C\\
&\qquad \text{or } \xi_t^{A(1), B(2)}(y) \neq 0 \textit{ for some } y \in D\big)\\
&\quad = \mathbb{P}\big(\xi_t^{C(1), D(2)}(x) = 2 \textit{ for some } x \in A\\
&\qquad\qquad \text{or } \xi_t^{C(1), D(2)}(y) \neq 0 \textit{ for some } y \in B\big).
\end{aligned}
\tag{2.9}
$$

The process $\tilde{\xi}_t$ given by (2.8) is rather interesting in its own right. It can be thought of as a contact process with the birth mechanism having On–Off periods. A new offspring starts with its birth mechanism in the Off state. After an Exp(γ) time, its birth mechanism switches to On, provided it is still alive. While On, births occur as for a rate λ contact process. This On period lasts for an Exp(δ) period of time, after which it switches to Off again. This On–Off toggling keeps occurring until the individual at that site is killed. Note that deaths occur at rate 1 for each type. One could also interpret this process in terms of metapopulations.

3. General properties.

3.1. *Monotonicity.* Define the natural partial order on the space of configurations by

$$\xi \le \eta \text{ if and only if } \xi(x) \le \eta(x) \quad \text{for all } x \in \mathbb{Z}^d. \tag{3.1}$$

THEOREM 3.1. *The two-stage contact process is monotone in the sense that, if $\lambda_1 \le \lambda_2$, $\gamma_1 \le \gamma_2$, and $\delta_1 \ge \delta_2$, then we can construct on the same probability space two such processes $\xi_t^{(1)}$ and $\xi_t^{(2)}$ with parameters $(\lambda_1, \gamma_1, \delta_1)$ and $(\lambda_2, \gamma_2, \delta_2)$, respectively, such that*

$$\xi_0^{(1)} \le \xi_0^{(2)} \Rightarrow \xi_t^{(1)} \le \xi_t^{(2)} \;\; \forall\; t \ge 0. \tag{3.2}$$

In particular, if we consider one parameter at a time, this implies that the two-stage contact process is monotone increasing in λ, γ and the initial configuration and monotone decreasing in δ; that is, it satisfies the following:

1. *Attractive*. If $\xi_t^{(1)}$ and $\xi_t^{(2)}$ are two versions of the process with the same rates, then $\xi_0^{(1)} \le \xi_0^{(2)} \Rightarrow \xi_t^{(1)} \le \xi_t^{(2)}$, $\forall\ t \ge 0$.
2. *Increasing in* λ. If $\xi_t^{(1)}$ and $\xi_t^{(2)}$ have birth rates λ_1 and λ_2, respectively, with $\lambda_1 \le \lambda_2$ (other rates being the same), then $\xi_0^{(1)} \le \xi_0^{(2)} \Rightarrow \xi_t^{(1)} \le \xi_t^{(2)}$, $\forall$ $t \ge 0$.
3. *Increasing in* γ. If $\xi_t^{(1)}$ and $\xi_t^{(2)}$ have maturation rates γ_1 and γ_2, respectively, with $\gamma_1 \le \gamma_2$ (other rates being the same), then $\xi_0^{(1)} \le \xi_0^{(2)} \Rightarrow$ $\xi_t^{(1)} \le \xi_t^{(2)}$, $\forall\ t \ge 0$.
4. *Decreasing in* δ. If $\xi_t^{(1)}$ and $\xi_t^{(2)}$ have death rates δ_1 and δ_2, respectively, with $\delta_1 \le \delta_2$ (other rates being the same), then $\xi_0^{(1)} \ge \xi_0^{(2)} \Rightarrow \xi_t^{(1)} \ge \xi_t^{(2)}$, $\forall$ $t \ge 0$.

PROOF OF THEOREM 3.1. Let $\lambda_1 \le \lambda_2$, $\gamma_1 \le \gamma_2$, and $\delta_1 \ge \delta_2$ be given. Construct a percolation diagram with ordinary arrows, 2-dots, δ's and $\times$'s at rates λ_1, γ_1, δ_2 and 1, respectively. To this we add, independently, green arrows, 2-dots and δ's at rates $\lambda_2 - \lambda_1$, $\gamma_2 - \gamma_1$ and $\delta_1 - \delta_2$, respectively. Define $\xi_t^{(1)}$ to be the process which uses all the ordinary objects and the green δ's (and refer to this, for the moment, as the "basic system"); $\xi_t^{(2)}$ is the process which uses all the ordinary objects in addition to the green arrows and 2-dots (refer to this as the "enhanced system"). It is easy to see that these processes have the desired rates. If $\xi_0^{(1)} = A(1) \cup B(2)$ and $\xi_0^{(2)} = C(1) \cup D(2)$, then $\xi_0^{(1)} \le \xi_0^{(2)}$ implies $A \subseteq C \cup D$ and $B \subseteq D$. Thus any path which is compatible with $A(1) \cup B(2)$ will also be compatible with $C(1) \cup D(2)$.

Next note that any path in the basic system will also be a path in the enhanced system. (The enhanced system contains all the arrows of the basic system and the same set of $\times$'s.) A path which is active in the basic system will also be active in the enhanced system, and strong terminal (resp., initial) points in the basic system will also be strong in the enhanced system.

Now, for all t, $\xi_t^{(1)}(x) \ne 0$ if and only if there is an active path in the basic system from $(A \cup B, 0)$ to (x, t) which is compatible with $A(1) \cup B(2)$, and this implies that there is an active path to (x, t) in the enhanced system which is compatible with $C(1) \cup D(2)$, and hence $\xi_t^{(2)}(x) \ne 0$. A similar argument shows that $\xi_t^{(1)}(x) = 2$ implies $\xi_t^{(2)}(x) = 2$. Here the desired condition is that there is at least one active path in the appropriate system which is compatible with the initial configuration and has a strong terminal point. If there is such a path, then the type at (x, t) will be 2 since the determining path, if not this one, would have to come into the top vertical segment below this one, and hence would also have a strong terminal point. The theorem is now proved. $\square$

THEOREM 3.2. *The two-stage contact process is additive in the sense that, for any disjoint sets* $A, B, C, D \subseteq \mathbb{Z}^d$,

$$\xi_t^{(A \cup C)(1), (B \cup D)(2)}(x) = \xi_t^{A(1), B(2)}(x) \vee \xi_t^{C(1), D(2)}(x). \tag{3.3}$$

PROOF. The right-hand side will equal 2 if and only if $\tilde{\xi}_t^{x(1)} \sim A(1) \cup B(2)$ or $\tilde{\xi}_t^{x(1)} \sim C(1) \cup D(2)$, and this happens if and only if $\tilde{\xi}_t^{x(1)} \sim (A \cup C)(1) \cup (B \cup D)(2)$, which happens if and only if the left-hand side of (3.3) equals 2.

Similarly, both sides of (3.3) will be nonzero if and only if $\tilde{\xi}_t^{x(2)} \sim A(1) \cup B(2)$ or $\tilde{\xi}_t^{x(2)} \sim C(1) \cup D(2)$. To get the values 0 and 1, just use the above two cases. □

3.2. *Critical parameters.* For the rest of the paper, we will write $\mathbb{P}_{\lambda,\gamma,\delta}$ for the law of the process with parameters λ, γ, δ whenever we want to emphasize the parameters. When $\delta = 0$, we will abbreviate this by $\mathbb{P}_{\lambda,\gamma}$. For fixed γ and δ, let

$$\lambda_c(\gamma,\delta) \equiv \inf\left\{\lambda: \mathbb{P}_{\lambda,\gamma,\delta}\left(\xi_t^{0(2)} \not\equiv 0 \ \forall\ t\right) > 0\right\} \tag{3.4}$$

be the unique value such that

$$\mathbb{P}_{\lambda,\gamma,\delta}\left(\xi_t^{0(2)} \not\equiv 0 \ \forall\ t\right) = 0 \quad \text{if } \lambda < \lambda_c(\gamma,\delta); \tag{3.5}$$

$$\mathbb{P}_{\lambda,\gamma,\delta}\left(\xi_t^{0(2)} \not\equiv 0 \ \forall\ t\right) > 0 \quad \text{if } \lambda > \lambda_c(\gamma,\delta). \tag{3.6}$$

Here, $\xi_t^{0(2)}$ is the two-stage contact process starting with one adult at the origin and all other sites empty. This critical value is well defined due to monotonicity in λ. We will write $\lambda_c(\gamma) = \lambda_c(\gamma, 0)$ for the critical birth rate when $\delta = 0$. The following lemma shows, among other things, that the definition of $\lambda_c(\gamma, \delta)$ does not depend on the state of the single initial particle. Both parts of the lemma are obtained by simply conditioning on the outcome of the first jump.

LEMMA 3.3. *For any λ, γ, δ,*

$$\mathbb{P}_{\lambda,\gamma,\delta}\left(\xi_t^{x(1)} \not\equiv 0 \ \forall\ t\right) = \frac{\gamma}{\gamma + 1 + \delta}\mathbb{P}_{\lambda,\gamma,\delta}\left(\xi_t^{x(2)} \not\equiv 0 \ \forall\ t\right) \tag{3.7}$$

and

$$\mathbb{P}_{\lambda,\gamma,\delta}\left(\tilde{\xi}_t^{x(1)} \not\equiv 0 \ \forall\ t\right) = \frac{\gamma}{\gamma + 1}\mathbb{P}_{\lambda,\gamma,\delta}\left(\tilde{\xi}_t^{x(2)} \not\equiv 0 \ \forall\ t\right). \tag{3.8}$$

The next result shows that the critical birth rate, $\lambda_c(\gamma, \delta)$, is finite for any fixed δ, as long as γ is sufficiently large.

THEOREM 3.4. *For fixed $\delta \geq 0$, if γ and λ are sufficiently large, then $\mathbb{P}_{\lambda,\gamma,\delta}(\xi_t^{x(2)} \not\equiv 0 \ \forall\ t) > 0$.*

PROOF. We use a simple comparison with (1-dependent) oriented site percolation, similar to Harris's (1974) original proof that the critical birth rate for the contact process is finite. For a given $\varepsilon > 0$, begin by choosing $T > 0$ so small that $\mathbb{P}_{\lambda,\gamma,\delta}$ (there is a death at x in $[0, 2T]$) $< \varepsilon/2$. Here, "death" refers to an $\times$ or δ in the percolation diagram. We say *x is good at time nT* if (1) there are no deaths at x during $[nT, (n+2)T]$, and (2) during $[(n+1)T, (n+2)T]$, there is a 2-dot at x followed by birth arrows from x to each site in $\mathcal{N}(x)$. By choosing λ and γ sufficiently large, we can guarantee that (2) will occur with probability at least $1 - \varepsilon/2$, and hence $\mathbb{P}_{\lambda,\gamma,\delta}(x$ is good at time $nT) \geq 1 - \varepsilon$. Finally, let $\mathscr{L}$ denote the renormalized lattice $\mathscr{L} = \{(m, n) \in \mathbb{Z} \times \mathbb{Z}_+ : m + n \text{ is odd}\}$ and say that $(m, n) \in \mathscr{L}$ is *open* if site $(m, 0, \dots, 0) \in \mathbb{Z}^d$ is good at time nT. For sufficiently small ε, the above procedure guarantees percolation of the open sites in $\mathscr{L}$, and this implies survival of $\xi_t^{x(2)}$ with positive probability. □

We remark that the above proof can be used to get an upper bound on the critical birth rate, but, as for the contact process, this bound is not good.

The following definition helps to delimit the survival region in parameter space.

DEFINITION 3.5. For each $\delta \geq 0$, define

$$\gamma_*(\delta) = \inf\left\{\gamma > 0 : \mathbb{P}_{\lambda, \gamma, \delta}\left(\xi_t^{x(2)} \not\equiv 0 \ \forall\ t\right) > 0 \text{ for some } \lambda < \infty\right\}, \tag{3.9}$$

$$\lambda_*(\delta) = \inf\left\{\lambda > 0 : \mathbb{P}_{\lambda, \gamma, \delta}\left(\xi_t^{x(2)} \not\equiv 0 \ \forall\ t\right) > 0 \text{ for some } \gamma < \infty\right\}. \tag{3.10}$$

These values are well defined because of monotonicity in γ and λ, respectively. So, for a given δ, survival cannot occur if $\gamma < \gamma_*(\delta)$, no matter how large λ is, and survival cannot occur if $\lambda < \lambda_*(\delta)$, no matter how large γ is. Note that $\gamma_*(\delta) = \inf\{\gamma > 0 : \lambda_c(\gamma, \delta) < \infty\}$.

The next result shows that these thresholds are not trivial.

PROPOSITION 3.6. *For any* $\delta \geq 0$,

$$\lambda_*(\delta) = \lambda_c(\infty) < \infty, \tag{3.11}$$

$$\gamma_*(0) \leq \gamma_*(\delta) < \infty. \tag{3.12}$$

Moreover, in the case of one-dimensional nearest-neighbor births,

$$\gamma_*(\delta) \geq \tfrac{1}{4}\left(\sqrt{17 + 14\delta + \delta^2} - 3 - \delta\right) \quad \forall\ \delta \geq 0. \tag{3.13}$$

[Note that the right-hand side of (3.13) is increasing in δ and strictly positive.]

PROOF. Finiteness of $\lambda_*(\delta)$ and $\gamma_*(\delta)$ follows immediately from Theorem 3.4, and monotonicity in δ gives the first inequality in (3.12).

To get the equality in (3.11), first use monotonicity in γ to get

$$\lambda_c(\gamma, \delta) \geq \lambda_c(\infty) \quad \text{for each } \gamma, \tag{3.14}$$

and hence

$$\lambda_*(\delta) \geq \lambda_c(\infty), \tag{3.15}$$

where we recall that $\lambda_c(\infty)$ denotes the critical birth rate for the contact process given in (1.1); that is, $\gamma = \infty$. [Note that $\lambda_c(\infty, \delta) = \lambda_c(\infty)$ for all $\delta \geq 0$.] The inequality $\lambda_*(\delta) \leq \lambda_c(\infty)$ follows from a rescaling argument which is similar to the proof of Theorem 3.4. The idea is that, if $\lambda > \lambda_c(\infty)$, then by taking γ sufficiently large the process behaves like a (supercritical) contact process in finite space–time boxes with high probability. Thus the process will survive with positive probability for sufficiently large γ, giving the desired inequality. We omit the details and refer the reader to Durrett (1991) for a number of examples of similar "perturbation arguments."

To show that $\gamma_*(\delta)$ is bounded away from 0 in the one-dimensional nearest-neighbor case (i.e., $\mathscr{N}(x) = \{x - 1, x + 1\}$ is the neighborhood set for site $x \in \mathbb{Z}$), we have to work a bit harder. We do this by studying the rightmost occupied site r_t for the two-stage contact process η_t with $\lambda = \infty$. In η_t, as soon as a site makes a transition from state 1 to state 2, it gives birth to 1's at the two neighboring sites if they are not already occupied. Note that no site neighboring a 2 can be vacant; if such a site experienced a death, the 2 would immediately replace the killed individual with a 1. In particular, the rightmost particle must be in state 1.

Clearly, if this process does not survive for a given value of γ, then neither will the two-stage contact process with the same γ and any $\lambda < \infty$. To get bounds on the transition rates for the rightmost 1 in η_t, we must consider the three sites at the right edge, that is, the right edge and the two sites to its left. There are seven possible three-site boundaries at the right edge:

$$\underbrace{221 \quad 121}_{\text{stable}} \quad \underbrace{211 \quad 111 \quad 011 \quad 101 \quad 001}_{\text{vulnerable}}.$$

The first two states are "stable" in the sense that, if a death occurs at the rightmost 1, the neighboring 2 will immediately replace it; the right edge cannot move left when in this configuration. The other five states are "vulnerable" in the sense that a death at the right edge (rate $1 + \delta$) causes the right edge to move to the left by *at least* one position. Each of the stable states will turn into a vulnerable state if the 2 adjacent to the rightmost 1 is killed (rate 1). The vulnerable states 211, 111 and 011 turn into stable states if the middle site makes a transition from 1 to 2 (rate γ), with no change in r_t. The right edge will move one unit to the right whenever the rightmost 1 turns into a 2 (rate γ). In fact, if a vulnerable state has a $1 \to 2$ transition at r_t, the right edge moves right by 1 and the boundary becomes the stable state 121; if a stable state has a $1 \to 2$ transition at r_t, the right edge moves right by 1 and the boundary becomes the stable state 221.

Thus there are three types of transitions:

(a) Stable to vulnerable (rate 1), or vulnerable to stable (rate γ) (with no change in r_t for either of these);
(b) r_t to $r_t + 1$ (rate $\gamma + \delta$);
(c) r_t to site less than or equal to $r_t - 1$ (rate $1 + \delta$) if vulnerable.

We can ensure that the next jump in r_t is to the left if we get an $\times$ at the middle position of the three-site boundary (ensuring that the right edge is vulnerable) before the right edge receives a 2-dot, and then get an $\times$ or δ at the right edge, before getting a 2-dot at either the middle position or the position of the rightmost 1. Thus, the probability that the next jump in r_t is to the left is bounded below by

$$\frac{1}{1+\gamma}\frac{1+\delta}{1+\delta+2\gamma}.$$

To get the right edge to drift to the left (and hence get η_t to die), it is enough that the above probability be greater than $1/2$. This is achieved when $0 < \gamma < \frac{1}{4}(\sqrt{17 + 14\delta + \delta^2} - 3 - \delta)$, and hence (3.13) is proved. □

An "artist's conception" of the survival region in parameter space, illustrating the roles of $\gamma_*(\delta)$ and $\lambda_*(\delta)$, appears in Figure 3.

By attractiveness, we can infer the existence of an upper invariant measure $\nu^{(2)}_{\lambda,\gamma,\delta}$,

$$\mathbb{P}_{\lambda,\gamma,\delta}\left(\xi_t^{\mathbb{Z}^d(2)} \in \cdot\right) \Rightarrow \nu^{(2)}_{\lambda,\gamma,\delta}(\cdot)$$

(weak convergence of probability measures on the space of configurations, which is equivalent to convergence of finite-dimensional distributions) as $t \to \infty$. To see why this is so, first note that it is enough to show $\mathbb{P}_{\lambda,\gamma,\delta}(\xi_t^{\mathbb{Z}^d(2)} \in \Gamma) \to \nu^{(2)}_{\lambda,\gamma,\delta}(\Gamma)$ for sets $\Gamma \in \{0,1,2\}^{\mathbb{Z}^d}$ of the form

$$\Gamma = \{\xi\colon \xi(x) = 2 \text{ for some } x \in C \text{ or } \xi(y) \neq 0 \text{ for some } y \in D\}.$$

The indicator function 1_Γ is increasing ["f increasing" means $\xi \le \eta \Rightarrow f(\xi) \le f(\eta)$], and so attractiveness implies that $t \mapsto \mathbb{P}_{\lambda,\gamma,\delta}(\xi_t^{\mathbb{Z}^d(2)} \in \Gamma)$ is a decreasing function, hence converges to a limit as $t \to \infty$. [Cf. Liggett (1985) for a discussion of these ideas in the setting of attractive spin systems.]

As usual, when $\delta = 0$, we will write $\nu^{(2)}_{\lambda,\gamma}$ instead of $\nu^{(2)}_{\lambda,\gamma,0}$. Since the two-stage contact process is a Feller process, standard arguments show that

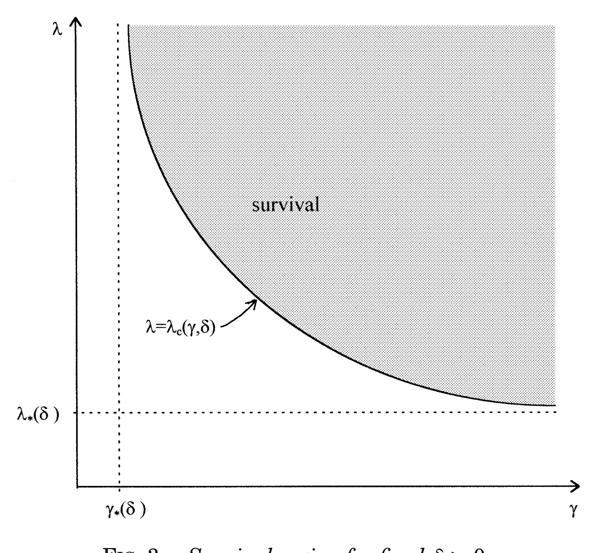

FIG. 3. *Survival region for fixed* $\delta \ge 0$.

$\nu^{(2)}_{\lambda,\gamma,\delta}$ is an invariant measure. It is obviously translation invariant. Let δ_0 denote the point mass on the all 0 configuration; this is trivially an invariant measure. By attractiveness, all invariant measures μ must lie between δ_0 and $\nu^{(2)}_{\lambda,\gamma,\delta}$ in the sense that

$$\delta_0(f) \le \mu(f) \le \nu^{(2)}_{\lambda,\gamma,\delta}(f) \quad \text{for all increasing } f,$$

where $\mu(f) \equiv \int_{\{0,1,2\}^{\mathbb{Z}^d}} f(\xi)\mu(d\xi)$.

Duality and (3.8) can be used to relate the density of 2's and the density of occupied sites in the upper invariant measure,

$$\nu^{(2)}_{\lambda,\gamma,\delta}\{\xi\colon \xi(x) = 2\} = \frac{\gamma}{\gamma+1}\nu^{(2)}_{\lambda,\gamma,\delta}\{\xi\colon \xi(x) \ne 0\}. \tag{3.16}$$

One can distinguish between two types of survival. We say the two-stage contact process has *finite survival* at (λ,γ,δ) if $\mathbb{P}_{\lambda,\gamma,\delta}(\xi_t^{0(2)} \not\equiv 0\ \forall\ t) > 0$ and we say it has infinite survival at (λ,γ,δ) if $\nu^{(2)}_{\lambda,\gamma,\delta} \ne \delta_0$. In the case where $\delta = 0$, it follows easily from self-duality that these two notions of survival are equivalent for the two-stage contact process. In particular, when $\gamma > \gamma_*(0)$ and $\lambda > \lambda_c(\gamma)$, $\nu^{(2)}_{\lambda,\gamma} \ne \delta_0$; if $\lambda < \lambda_c(\gamma)$, then the upper invariant measure equals δ_0.

What about other initial configurations? If we start the two-stage contact process with configuration $\mathbb{Z}^d(1)$, then we claim that the stationary measure is still $\nu^{(2)}_{\lambda,\gamma,\delta}$. We will only sketch the proof. For example, to get the stationary density of 2's, use duality to write

$$\begin{aligned}\mathbb{P}_{\lambda,\gamma,\delta}\left(\xi_t^{\mathbb{Z}^d(1)}(x) = 2\right) &= \mathbb{P}_{\lambda,\gamma,\delta}\left(\tilde{\xi}_t^{x(1)} \sim \mathbb{Z}^d(1)\right)\\ &= \mathbb{P}_{\lambda,\gamma,\delta}\left(\tilde{\xi}_t^{x(1)} \sim \mathbb{Z}^d(1) \,|\, \tilde{\xi}_t^{x(1)} \not\equiv 0\right)\mathbb{P}_{\lambda,\gamma,\delta}\left(\tilde{\xi}_t^{x(1)} \not\equiv 0\right).\end{aligned}$$

Since the all-0 configuration is absorbing for $\tilde{\xi}_t^{x(1)}$, the last probability decreases to $\nu^{(2)}_{\lambda,\gamma,\delta}\{\xi\colon \xi(x) = 2\}$ as $t \to \infty$. A simple Markov argument shows that, if the process $\tilde{\xi}_t^{x(1)}$ survives, then its size goes to infinity as $t \to \infty$. To get $\tilde{\xi}_t^{x(1)} \sim \mathbb{Z}^d(1)$, we need for at least one of the terminal points to be strong. The states of the various terminal points are not independent, but they are essentially so for points which are far enough apart. Clearly, as the number of terminal points increases to infinity, the probability of having at least one strong goes to 1, and this yields the desired result. A similar argument shows that the limit starting from any translation-invariant initial distribution with an infinite number of occupied sites is $\nu^{(2)}_{\lambda,\gamma,\delta}$.

REMARK. The arguments in this section can be used to show that the On–Off contact process $\tilde{\xi}_t$ has many of the same properties as the two-stage contact process. For example, it has the same monotonicity properties (cf. Theorem 3.1) and there are obvious analogues, $\tilde{\lambda}_c(\gamma,\delta)$, $\tilde{\lambda}_*(\delta)$ and $\tilde{\gamma}_*(\delta)$, of the critical parameters. We do not pursue this further here.

4. Some open problems.

We conclude by mentioning several open problems for the two-stage contact process.

1. Are finite survival and infinite survival equivalent for the two-stage contact process with $\delta > 0$ or are there two distinct critical values?

From (2.7), it follows that $\nu^{(2)}_{\lambda,\gamma,\delta}\{\xi\colon \xi(x) \neq 0\} = \mathbb{P}_{\lambda,\gamma,\delta}(\tilde{\xi}_t^{x(2)} \not\equiv 0\ \forall\ t)$, so infinite survival of the two-stage contact process is equivalent to finite survival of the On–Off contact process given by (2.8). Thus the question becomes one of whether or not the critical birth rate for finite survival is the same for the two-stage contact process and the On–Off contact process when $\delta > 0$.

2. In the one-dimensional nearest-neighbor case, can we characterize the critical birth rate in terms of edge speed?

Define the leftmost and rightmost occupied sites for $\xi_t^{A(2)}$ by

$$l_t^A \equiv \inf\{x\colon \xi_t^{A(2)}(x) \neq 0\} \quad \text{and} \quad r_t^A \equiv \sup\{x\colon \xi_t^{A(2)}(x) \neq 0\} \tag{4.1}$$

and let

$$\tau^{A(1),B(2)} \equiv \inf\{t\colon \xi_t^{A(1),B(2)} \equiv 0\}$$

denote the lifetime of $\xi_t^{A(1),B(2)}$. Let $\alpha_t(\lambda,\gamma,\delta) \equiv \mathbb{E}_{\lambda,\gamma,\delta}(r_t^{(-\infty,0]})$. Then

$$\alpha(\lambda,\gamma,\delta) \equiv \lim_{t\to\infty} \frac{\alpha_t(\lambda,\gamma,\delta)}{t} = \inf_{t>0} \frac{\alpha_t(\lambda,\gamma,\delta)}{t} \in [-\infty,\infty), \tag{4.2}$$

and

$$\lim_{t\to\infty} \frac{1}{t} r_t^{(-\infty,0]} = \alpha(\lambda,\gamma,\delta), \quad \mathbb{P}_{\lambda,\gamma,\delta}\text{-a.s.} \tag{4.3}$$

This follows from the subadditive ergodic theorem [cf. Liggett (1985)] in the usual way. Just let

$$\begin{aligned} r_{s,t} = \max\Bigl\{&x \in \mathbb{Z}\colon \text{there is an active path} \\ &\text{from } (y,s) \text{ to } (x,t) \text{ for some } y \le r_s^{(-\infty,0]}\Bigr\} - r_s^{(-\infty,0]}, \end{aligned} \tag{4.4}$$

for $0 \le s \le t$. In other words, $r_{s,t}$ is obtained by restarting the process at time s, with all sites $\le r_s^{(-\infty,0]}$ in state 2 and all other sites in state 0, and then seeing how far the right edge moved by time t. It is clear that $r_{0,t} = r_t^{(-\infty,0]}$ and that $r_{0,t} \le r_{0,s} + r_{s,t}$. The rest of the proof of (4.2) and (4.3) is as in Liggett (1985).

We would like to show that

$$\lambda_c(\gamma,\delta) = \inf\{\lambda\colon \alpha(\lambda,\gamma,\delta) > 0\} = \sup\{\lambda\colon \alpha(\lambda,\gamma,\delta) < 0\}.$$

This would follow as in Durrett [(1988), pages 53–58] were it not for the lack of a complete coupling theorem. It is easy to see from the percolation diagram that it is *not* true that $\xi_t^{0(2)}(x) = \xi_t^{\mathbb{Z}(2)}(x)$, $\forall\ x \in [l_t^0, r_t^0]$, on $\{\tau^{0(2)} > t\}$. If we soften the requirement by just asking for occupied sites, we might hope that, on $\{\tau^{0(2)} > t\}$,

$$\{x\colon \xi_t^{0(2)}(x) \neq 0\} = \{x\colon \xi_t^{\mathbb{Z}(2)}(x) \neq 0\} \cap [l_t^0, r_t^0].$$

Unfortunately, this is also *not* true. The problem is that when two active paths intersect, the resulting composite paths need not be active. Presumably, the way around this is with a renormalized bond construction similar to the one in Durrett and Schonmann (1987). This would give an asymptotic version of the above formula. The problem here is in "tying paths together" with other paths so that one obtains active composite paths. As in Durrett and Schonmann (1987), there are "conditioning problems," and these do not seem to be as easy to overcome as for their discrete-time growth models or for the contact process.

3. Does the complete convergence theorem hold?

Is it true that

$$\mathbb{P}_{\lambda,\gamma,\delta}\big(\xi_t^{A(1),B(2)} \in \cdot\big) \Rightarrow \mathbb{P}_{\lambda,\gamma,\delta}\big(\tau^{A(1),B(2)} = \infty\big)\,\nu^{(2)}_{\lambda,\gamma,\delta} + \mathbb{P}_{\lambda,\gamma,\delta}\big(\tau^{A(1),B(2)} < \infty\big)\,\delta_0,$$

for all disjoint A, B? The comments for problem 2 apply here.

4. Describe the critical birth rate for the two-stage contact process in terms of the critical birth rate for the contact process.

It is certainly unreasonable to ask for the exact value of $\lambda_c(\gamma, \delta)$, since the exact critical value is not even known for the contact process. However, it might be possible to express $\lambda_c(\gamma, \delta)$ in terms of $\lambda_c(\infty)$, γ and δ.

5. Describe the structure of the survival region.

Is Figure 3 accurate? The answer to this problem would follow if we knew the answer to problem 4.

6. Find better bounds on $\gamma_*(\delta)$.

In particular, a lower bound is needed for the case $d \geq 2$. The argument we gave in the previous section works only for one-dimensional, nearest-neighbor births.

Acknowledgments. I thank Claudia Neuhauser and Rinaldo Schinazi for discussions on this work. Thanks also go to Claudia Krone for her patient assistance with the graphs. Finally, I thank an anonymous referee who caught a number of minor errors in the original manuscript.

REFERENCES

DURRETT, R. (1988). *Lecture Notes on Particle Systems and Percolation*. Wadsworth, Belmont, CA.

DURRETT, R. (1991). A new method for proving the existence of phase transitions. In *Spatial Stochastic Processes* (K. Alexander and J. Watkins, eds.) 141–169. Birkhäuser, Boston.

DURRETT, R. and SCHONMANN, R. (1987). Stochastic growth models. In *Percolation Theory and Ergodic Theory of Infinite Particle Systems* (H. Kesten, ed.) 85–119. Springer, New York.

HANSKI, I. and ZHANG, D.-Y. (1993). Migration, metapopulation dynamics and fugitive coexistence. *J. Theoret. Biol.* **163** 491–504.

HARRIS, T. (1972). Nearest neighbor Markov interaction processes on multidimensional lattices. *Adv. Math.* **9** 66–89.

HARRIS, T. (1974). Contact interactions on a lattice. *Ann. Probab.* **2** 969–988.

LIGGETT, T. (1985). *Interacting Particle Systems*. Springer, New York.

NEUHAUSER, C. (1992). Ergodic theorems for the multitype contact process. *Probab. Theory Related Fields* **91** 467–506.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF IDAHO
MOSCOW, IDAHO 83844
E-MAIL: krone@uidaho.edu