Journal of Statistical Physics, Vol. 74, Nos. 1/2, 1994

On Critical Phenomena in Interacting Growth Systems. Part I: General

Andrei Toom[1]

Received January 22, 1993; final August 4, 1993

Components which are placed in a finite or infinite space have integer numbers as possible states. They interact in a discrete time in a local deterministic way, in addition to which all the components' states are incremented at every time step by independent identically distributed random variables. We assume that the deterministic interaction function is translation-invariant and monotonic and that its values are between the minimum and the maximum of its arguments. Theorems 1 and 2 (based on propositions which we give in a separate Part II), give sufficient conditions for a system to have an invariant distribution or a bounded mean. Other statements, proved herein, provide background for them by giving conditions when a system has no invariant distribution or the mean of its components' states tends to infinity. All our main results use one and the same geometrical criterion.

KEY WORDS: Random process; local interaction; critical phenomena; invariant distribution; growth; eroder; convexity.

Most interacting random processes (IRP) considered till now have compact sets of states of every component. (These sets are usually finite, often they have two elements.) IRP with compact sets of states have at least one invariant distribution (see, for example, Proposition 2.5 on p. 25 of ref. 14) and its uniqueness (which may be called "ergodicity") is often at the center of attention. It was proved long ago[15,16] that any system of a wide class has only one invariant distribution, provided the noise is strong enough. On the other hand, refs. 10–13, 7, and 2 used the contour method and its ramified version to present systems which have more than one invariant distribution with any small enough noise. (See also a survey[14] which covers some of these developments.) Together these two groups of results

[1] Incarnate Word College, San Antonio, Texas 78209.

0022-4715/94/0100-0091$07.00/0

showed the existence of "critical phenomena," that is, transition from a unique to nonunique invariant distribution due to a continuous change of parameters.

Throughout our paper, every component of a system has $\mathbf{Z}$, the set of integer numbers, as the set of its possible states. $\mathbf{Z}_M$ is the ring of residues modulo M. Components of the system, denoted s, are placed in the Space, which is $\mathbf{Z}^d$ in the infinite and $\mathbf{Z}_M^d$ in the finite case. At every step of the discrete time t every component calculates its new state $x(s, t)$, applying one and the same *transition function* $\mathbf{f}: \mathbf{Z}^n \mapsto \mathbf{Z}$ to the states of its "neighbors" at the $\mathbf{m}$ previous moments of time, after which the component's state grows by a random variable ζ; all these random increments are mutually independent.

Now, before asking how many invariant distributions a system has, it is appropriate to ask if it has any at all. It is this, more preliminary question which we partially answer here. Another, but related question, which we also partially answer, is how the system's averages behave, which are the expectations of $x(s, t)$, as $t \to \infty$. If the initial distribution is space-uniform, these expectations do not depend on s, and we denote them $\mathbf{E}_t$. Theorems 1 and 2, our main results, give sufficient conditions (which involve the geometry of interaction) for a system to have at least one invariant distribution in spite of the presence of a random noise and for a system's average not to grow beyond a constant. They are based on propositions, which we put in a separate Part II.[15] Other results include Propositions 1–3, which provide background for our theorems, being their converses. Proposition 4 uses the same geometrical criterion to tell whether a deterministic system is an "eroder," that is, turns every initial condition with a finite set of nonzero components into "all zeros" after a finite time. Taken together, our results show the existence of "critical phenomena." In our case this means transition from existence to nonexistence of an invariant distribution of from bounded to unbounded growth due to a continuous change of parameters.

1. DEFINITIONS AND FORMULATIONS

Denote

$$\text{Time} = \{t \in \mathbf{Z} \mid t \geqslant -\mathbf{m}\}$$

$$V = \text{Space} \times \text{Time} = \{(s, t) \mid s \in \text{Space},\ t \in \text{Time}\}$$

Elements of V are called *points* and subsets of V are called *point-sets*. The origin $(0, 0)$ of V is denoted $\mathcal{O}$. If $v = (s, t) \in V$, then $s(v) = s$ and $t(v) = t$. Points with $t \in [-\mathbf{m}, -1]$ are called *initial*, those with $t \geqslant 0$ are called *inner*. V_{init} and V_{inner} denote the sets of initial and inner points. Every inner points (s, t) has *n neighbors*

$$N_i(s, t) = (s - \Delta s_i, t - \Delta t_i), \qquad \text{where} \quad \Delta t_i \in [1, \mathbf{m}] \quad \text{for all } i = 1,..., n$$

The neighborhood $N(v)$ of an inner point v is the set of its neighbors. Initial points have no neighbors. To every point $v = (s, t) \in V$ there corresponds a variable $x(v) = x(s, t) \in \mathbf{Z}$, the state of this point. The set $X = \mathbf{Z}^V$ is the configuration space whose elements x are called *configurations*. There are mutually independent *hidden variables* $h(v)$ for all inner points v, every one of which is distributed as a random integer variable ζ. Thus we have the hidden distribution on the configuration space of hidden variables. To any transition function, any neighbor vectors, any distribution of ζ, and any M in the finite case, there corresponds a process, i.e., the distribution on X induced by the hidden distribution with the map defined in the following inductive way:

$$x(v) = \begin{cases} y(v) & \text{if} \quad t(v) < 0 \\ \mathbf{f}(x(N(v))) + h(v) & \text{if} \quad t(v) \geqslant 0 \end{cases} \tag{1}$$

where the distribution of the initial variables $y(\cdot)$ serves as a parameter.

We may also interpret our systems as linear operators $P: \mathscr{M} \mapsto \mathscr{M}$ which act on the set $\mathscr{M}$ of probability measures on the configuration space $C_{\mathbf{m}}$ which corresponds to any $\mathbf{m}$ consecutive moments of time. We may call $\mu \in \mathscr{M}$ *invariant* if $P(\mu) = \mu$ and interpret some of our results as those about the existence of an invariant distribution. In the deterministic case P turns into a deterministic operator D which acts on $C_{\mathbf{m}}$.

We assume that our transition function is bounded by the minimum and the maximum of its arguments

$$\forall x_1,..., x_n\text{: } \min\{x_1,..., x_n\} \leqslant \mathbf{f}(x_1,..., x_n) \leqslant \max\{x_1,..., x_n\} \tag{2}$$

translation-invariant in the set of states

$$\forall x_1,..., x_n, \quad c\text{: } \mathbf{f}(x_1 + c,..., x_n + c) = \mathbf{f}(x_1,..., x_n) + c \tag{3}$$

and monotonic

$$x \preccurlyeq y \Rightarrow \mathbf{f}(x) \leqslant \mathbf{f}(y) \tag{4}$$

where $x = (x_1,..., x_n)$, $y = (y_1,..., y_n)$, and $x \preccurlyeq y$ means that $x_i \leqslant y_i$ for all i. We call (2)–(4) our "standard assumptions." All our theorems and propositions are true under these assumptions, but some are true under weaker ones. In the most important cases we specify what we actually need to assume.

The central idea of our paper is to connect the behavior of our systems with their geometrical properties. The following definitions make this possible. Call a set $S \subseteq \{1,..., n\}$ a *drag* if whenever $\mathbf{f}(x_1,..., x_n) = \max\{x_1,..., x_n\}$, there is such $i \in S$ that $x_i = \max\{x_1,..., x_n\}$. To every drag S there corresponds an $\mathcal{O}$-drag $S_{\mathcal{O}} = \{(-\Delta s_i, -\Delta t_i) \mid i \in S\}$. ($\mathcal{O}$-drags are analogs of zero-sets of ref. 12.) Thus we have the family $\mathcal{D}$ of drags which are subsets of $\{1,..., n\}$ and the family $\mathcal{D}_{\mathcal{O}}$ of $\mathcal{O}$-drags, or drags of $\mathcal{O}$, which are subsets of $N(\mathcal{O})$. Due to uniformity, we may call v-drags the sets resulting from $\mathcal{O}$-drags by a shift at the vector v.

Note that in the infinite case V is a subset of $\mathbf{Z}^{d+1}$ and therefore a subset of $\mathbf{R}^{d+1}$. Using this, we can consider any $\mathcal{O}$-drag as a subset of $\mathbf{R}^{d+1}$. Let $\mathrm{conv}(S)$ denote the convex hull of any set $S \subset \mathbf{R}^{d+1}$. For any real number k and any set $S \subset \mathbf{R}^{d+1}$ we denote $k \cdot S = \{k \cdot v \mid v \in S\}$ and $\mathrm{ray}(S) = \bigcup \{k \cdot S \mid k \geqslant 0\}$. The following subset of $\mathbf{R}^{d+1}$ plays the key role in our paper:

$$\sigma = \bigcap \{\mathrm{ray}(\mathrm{conv}(S)) \mid S \in \mathcal{D}_{\mathcal{O}}\}$$

This formula certainly makes sense, because the family of drags is non-empty, since the set $\{1,..., n\}$ always is a drag. (And the empty set never is a drag.) Of course, σ always contains the origin $\mathcal{O}$. It it contains nothing else, we write $\sigma = \{\mathcal{O}\}$; otherwise, $\sigma \neq \{\mathcal{O}\}$. In the analogous way we can define antidrags, $\mathcal{O}$-antidrags, and the set anti-σ simply by changing the signs of all states.

Theorem 1. For any system in which both σ and anti-σ equal $\{\mathcal{O}\}$, there is a positive constant ε, for which the following holds: whenever

$$\mathrm{Prob}(\zeta \geqslant k) \leqslant \varepsilon^k \qquad \text{and} \qquad \mathrm{Prob}(\zeta \leqslant -k) \leqslant \varepsilon^k \qquad \text{for all} \quad k = 1, 2, 3,...$$

the system has at least one invariant distribution.

To prove Theorem 1, we need only (2) of our "standard assumptions."

The following Theorem 2 and Propositions 2 and 3 are about the initial condition "all zeros": $y(v) = 0$ for all initial v. In this case the expectation $\mathbf{E}(x(s, t))$ does not depend on s, and we call it $\mathbf{E}_t$. In the infinite case we ask whether $\mathbf{E}_t$ remains bounded forever (from above or from below or both). In the finite case we denote T_E the time (which may be infinite) when $\mathbf{E}_t$ first exceeds a given boundary E and ask how T_E behaves as a function of the system's size M.

Theorem 2. Take such a transition function and neighbor vectors that $\sigma = \{\mathcal{O}\}$. Take the initial condition "all zeros." Then there is such a positive constant ε that whenever $\mathrm{Prob}(\zeta \geqslant k) \leqslant \varepsilon^k$ for all positive k:

1. In the infinite system $\mathbf{E}_t$ never exceeds a constant.
2. In the finite systems T_E has a lower bound which grows as an exponent in M for large enough E.

To prove Theorem 2, we need nothing more than the right inequality of (2) of our "standard assumptions."

Behavior, described by Theorems 1 and 2, is an analog of nonuniqueness of an invariant distribution (which may be called "nonergodicity") of systems with a finite set of states of every component. It is unclear what might serve as an analog of "ergodicity" in our case. It seems that this analog should state that every system of some rich class has no invariant distribution or displays an unbounded growth, provided the noise is strong enough. Propositions 1–3 are steps in this direction.

Proposition 1. Let $\text{Prob}(\zeta = 0) < 1$. Then in the finite case there is no invariant distribution.

If the distribution of ζ has a mean, $\mathbf{E}_t$ does not exceed a linear function of time. We say that $\mathbf{E}_t$ has a linear lower bound with a constant $C > 0$ if $\mathbf{E}_t \geqslant C \cdot t$ for all t. Existence of a linear lower bound implies $T_E \leqslant \text{const} \cdot E$.

Proposition 2. Take the initial condition "all zeros" and let $\text{Prob}(\zeta < 0) = 0$ and $\text{Prob}(\zeta = 0) < 1/n$. Then $\mathbf{E}_t$ tends to infinity when $t \to \infty$ and has a linear lower bound with one and the same constant for the infinite and finite systems with all M.

Proposition 3. Take such a transition function and neighbor vectors that $\sigma \neq \{\mathcal{O}\}$. Take the initial condition "all zeros." Let $\text{Prob}(\zeta < 0) = 0$ and $\text{Prob}(\zeta > 0) > 0$. Then $\mathbf{E}_t$ tends to infinity when $t \to \infty$ and has a linear lower bound with one and the same constant for the infinite and finite systems with all M.

The next and last one of our main propositions is about the deterministic case $\zeta \equiv 0$. (We do not single out other deterministic cases $\zeta \equiv k$, $k \neq 0$.) In this case for any initial condition y the resulting process is concentrated in one configuration called *trajectory* and denoted y^{tr}. Call a configuration (and thereby a trajectory) *finitary* if it has a finite set of nonzero components. Say that a deterministic system is an *eroder* if for any finitary initial condition the resulting trajectory is also finitary.

Proposition 4. An infinite deterministic system is an eroder if and only if both σ and anti-σ equal $\{\mathcal{O}\}$.

Let us present several examples to which we shall refer. In all of them $\Delta t_i = 1$ for all $i = 1,\dots, n$. Thus, to specify neighbor vectors we need only to present their space components.

Example 0. Take any space and let $n=1$, that is, there is only one neighbor vector with an arbitrary space component. Let $\mathbf{f}(x)=x$. (There is essentially no interaction here.)

Example 1. Space is $\mathbf{Z}$ or $\mathbf{Z}_M$ and $n=2$, that is, there are two neighbor vectors. Assume that the two neighbor vectors are not collinear and let $\mathbf{f}(x, y)=\min(x, y)$. (If we take collinear neighbor vectors in this example, it becomes almost as trivial as the previous one.)

Example 2. Space is $\mathbf{Z}^2$ or $\mathbf{Z}_M^2$ and $n=4$, that is, there are four neighbor vectors. Their space components are

$$\Delta s_1=(0,0), \qquad \Delta s_2=(0,-1), \qquad \Delta s_3=(-1,0), \qquad \Delta s_4=(-1,-1)$$

and

$$\mathbf{f}(x_1, x_2, x_3, x_4)=\min\{\max\{x_1, x_2\}, \max\{x_3, x_4\}\}$$

In the next examples Space is $\mathbf{Z}^2$ or $\mathbf{Z}_M^2$ and $n=3$. Space components of the three neighbor vectors are

$$\Delta s_1=(0,1), \qquad \Delta s_2=(1,0), \qquad \Delta s_3=(0,0)$$

(This is known as the NEC neighborhood, i.e., North, East, Center; see, e.g., ref. 1; see also refs. 7, 3, and 4.) It remains to define the transition function in every example.

Example 3. $\mathbf{f}(x, y, z)$ equals the median of its three arguments, where the median of $2k-1$ numbers equals the kth one among them if they are sorted in the increasing order.

Example 4. $\mathbf{f}(x, y, z)$ equals the arithmetic mean of its three arguments, rounded to the nearest integer.

Example 5.

$$\mathbf{f}(x, y, z)=\begin{cases}\max\{x, y, z\} & \text{if at least two among } x, y, z \\ & \text{equal } \max\{x, y, z\} \\ \max\{x, y, z\}-1 & \text{otherwise}\end{cases}$$

All our examples satisfy our standard assumptions. Let us examine which drags they have. First note that if S is a drag and $S \subseteq S' \subseteq \{1,..., n\}$, then S' is also a drag. Thus the family of drags is completely described as soon as we list all the minimal drags. (A minimal drag is a drag, all of whose proper subsets are not drags. The same about $\mathcal{O}$-drags.) Of course, we can take intersection over minimal $\mathcal{O}$-drags rather than over all $\mathcal{O}$-drags in the definition of σ. In Example 0 the set σ is a half-line emanating from $\mathcal{O}$ in the direction of the (unique) neighbor vector. In Example 1 there are two minimal drags: $\{1\}$ and $\{2\}$, whence $\sigma = \{\mathcal{O}\}$. In Example 2 also there are two minimal drags: $\{1, 2\}$ and $\{3, 4\}$, whence again $\sigma = \{\mathcal{O}\}$. In Examples 3–5 there are three minimal drags: $\{1, 2\}$, $\{1, 3\}$, and $\{2, 3\}$. Since our three neighbor vectors are not coplanar, $\sigma = \{\mathcal{O}\}$ in all these cases. To obtain more examples in which $\sigma \neq \{\mathcal{O}\}$, it is sufficient to change our Examples 3–5 just in one respect: take Δs_1, Δs_2, and Δs_3 that belong to a straight line. Thus we see that systems with one and the same transition function, but different neighbor vectors, can behave in quite different ways.

Informally, the following assumption means that the transition function does not give preference to greater or smaller values:

$$\forall x_1,..., x_n, \quad c\colon \mathbf{f}(c - x_1,..., c - x_n) = c - \mathbf{f}(x_1,..., x_n) \tag{5}$$

Whenever it holds, every drag is an antidrag and vice versa. Although we never formally use (5), it helps us to show that Theorems 1 and 2 are not trivial, because they apply to Examples 3–5. Examples 0, 3, and 4 satisfy (5). Example 5 may even seem to give preference to greater values.

Note 1. Under our standard assumptions, a set $S \subseteq \{1,..., n\}$ is a drag if and only if $\mathbf{f}(z_1,..., z_n) = 0$, where every z_i equals 0 if $i \in S$ and equals 1 otherwise.

Note 2. For any transition function $\mathbf{f}$ we can define another transition function $\mathbf{f}^+$ by the following rule, where Max denotes $\max\{x_1,..., x_n\}$:

$$\mathbf{f}^+(x_1,..., x_n) = \begin{cases} Max - 1 & \text{if the set } \{i \mid x_i < Max\} \text{ is a drag} \\ Max & \text{otherwise} \end{cases}$$

Applied to Example 4, this rule gives Example 5. Generally, $\mathbf{f}^+$ is always monotonic, has the same drags as $\mathbf{f}$, and $\mathbf{f} \leqslant \mathbf{f}^+$. Substitution of an arbitrary $\mathbf{f}$ by $\mathbf{f}^+$ helps us to avoid assuming (4) when proving Theorems 1 and 2 and some other statements.

Let us introduce some notations which we shall use. Choose any norm on $\mathbf{R}^{d+1}$, define distance $\operatorname{dist}(v, w) = \operatorname{norm}(v - w)$, and let $\operatorname{diam}(S)$ denote the diameter of any set S. Denote

$$\Delta = \max\{\operatorname{norm}(\Delta s_i, \Delta t_i) \mid i = 1,..., n\}$$

For any real function x on a nonempty set S we denote

$$\max(x \mid S) = \max\{x(e) \mid e \in S\}$$

$$\min(x \mid S) = \min\{x(e) \mid e \in S\}$$

$$\operatorname{dev}(x) = \{v \in S \mid x(v) \neq 0\}$$

2. PROOF OF PROPOSITION 1

The following argument does not use (4). Note that in the finite case our system is a Markov chain with a countable set of states. Suppose that it has an invariant distribution μ. We may assume that μ is concentrated in one recurrent class. Take some state a in this class. Let D: $C_m \mapsto C_m$ be our deterministic operator. The sequence $D^n(a)$ must be periodic starting from some place, because all the components of its members have states in the range between the minimal and the maximal components of a. Let T denote the length of its period. Choose a member s of this period and let S denote the set of shifts of s in the set of states. (We say that s' is a shift of s by the number k if all the components of s' are k greater than the corresponding components of s.) Consider another Markov chain, whose set of states is S, in which the probability of going from a state b to a state c equals the following conditional probability of the original process: the probability that $x = c$ on condition that x is the first state belonging to S where we get having started from b. The restriction of μ to S gives us a nondegenerate invariant distribution of the new process. Note that the new process is isomorphic to a uniform random walk on an integer line. If $\text{Prob}(\zeta = k) = 1$ for some $k \neq 0$, this walk is a nonzero shift, which is evidently transient. Otherwise there are two different numbers, say k_1 and k_2, such that $\text{Prob}(\zeta = k_1) > 0$ and $\text{Prob}(\zeta = k_2) > 0$. Let us prove that in this case our random walk is transient also. Consider the old process starting from s. The first, deterministic, part of our operator transforms it into the state $D(s)$. Now, when the random part is applied, we go with positive probabilities to $D(s)$ shifted by k_1 and k_2. In the same way, after T steps, we go with positive probabilities to at least two different shifts of $D^T(s) = s$. ∎

3. PROOF OF PROPOSITION 2

Of all our standard assumptions we need only the left inequality of (2) here. The well-known monotonicity arguments reduce the proof to the case

$$\mathbf{f}(x_1, \ldots, x_n) = \min\{x_1, \ldots, x_n\} \tag{6}$$

and

$$\text{Prob}(\zeta = 0) = 1 - \varepsilon \qquad \text{and} \qquad \text{Prob}(\zeta = 1) = \varepsilon \qquad \text{where} \quad \varepsilon > 1 - 1/n$$

Given (6), every single number from 1 to n is a drag, whence σ consists only of the origin whenever there are at least two noncollinear neighbor vectors. Example 1 belongs to this case. We may call (6) "the percolation process," due to the following representation. Let $\bar{V}$ be the directed graph which has V as its set of vertices, and edges that go to every point from its neighbors. Reconstruct $\bar{V}$ into another directed graph $\bar{V}'$ by stretching every its vertex v into a "new" directed edge and assign to this edge v a random time delay that equals $h(v)$. "Old" edges, which came from $\bar{V}$, have zero delay. Define the time delay of a path in $\bar{V}'$ as the sum of delays along it. Note that $x(v)$ equals the minimum of delays of directed paths from initial points to v, which is well known as the time of first-passage percolation from V_{init} to v. Thus, the theory of first-passage percolation may be used here, but in fact straightforward estimations are sufficient.

Take an inner point $v = (s, t)$. For any C the probability that $x(v) \leqslant C$ equals the probability that there is a path from V_{init} to v whose delay is not more than C. The number of these paths does not exceed n^t. Now let length(Π) of a path Π denote the number of "new" edges in it. In our case every path's length is not less than $t/\mathbf{m}$. Delay of a path of length l is distributed as $B_{l,\varepsilon}$, the classical Bernoullian sum of l independent random variables, every one of which equals 1 with probability ε and 0 with probability $1-\varepsilon$. It is well known that for any $\delta > 0$ and any $\gamma < 1$ there is such $\varepsilon < 1$ that $\text{Prob}(B_{l,\varepsilon} \leqslant \gamma \cdot l) \leqslant \text{const} \cdot \delta^l$ for all l. Applying this to our case, one sees that

$$\text{Prob}(\text{delay}(\Pi) \leqslant \gamma \cdot \text{length}(\Pi)) \leqslant \text{const} \cdot \delta^{\text{length}(\Pi)} \leqslant \text{const} \cdot \delta^{t/\mathbf{m}}$$

Thus

$$\text{Prob}(x(v) \leqslant \gamma \cdot t/\mathbf{m}) \leqslant \text{const} \cdot \delta^{t/\mathbf{m}} \cdot n^t = \text{const} \cdot (\delta^{1/\mathbf{m}} \cdot n)^t$$

Choose such δ that $\delta^{1/\mathbf{m}} \cdot n \leqslant 1/2$. Then for any $\gamma < 1$ there is such $\varepsilon < 1$ that

$$\text{Prob}(x(v) \leqslant \gamma \cdot t/\mathbf{m}) \leqslant \text{const} \cdot (\tfrac{1}{2})^t$$

whence $\mathbf{E}_t \geqslant \gamma \cdot t/\mathbf{m} - \text{const}$. Since $\mathbf{E}_t > 0$ for all $t > 0$, we may cross out the constant. ∎

4. ABOUT THEOREMS 1 AND 2

Using the well-known convexity, compactness, and monotonicity arguments (explained, for example, on pp. 25–31 of ref. 14), our Theorems 1 and 2 can be deduced from Propositions 1 and 2 of our Part II.[15] Here we preface arguments of Part II with some explanations.

The general idea comes from the contour method: to estimate the probability of the event $x(p) \geqslant q$ by covering it by several events and estimating the sum of their probabilities by combinatorial means. We shall call "inquest" the process of constructing these events. We start by looking at the point v to see what "responsibility" it bears for having a positive state. If $x(p) = h(p)$, the "responsibility" remains completely with this point, and we do not need to go further. Otherwise $x(p)$ "inherited" at least part of its positive value from its neighbors, and we "interrogate" them in the same way, and so on. Thus we trace our event back to its ultimate causes and split it into several cases. In each case several hidden variables must be positive to ensure our event. If k denotes the sum of their values, the crucial point of our argument is to obtain an exponential (in k) estimation for the number of corresponding cases; we achieve this by presenting a geometrical construction (a "branching contour") whose complexity does not exceed const $\cdot\, k$. This construction is based on the Linear Lemma (see below), which shows that the condition $\sigma = \{\mathcal{O}\}$ given, the further into the past our "inquest" goes, the greater is the space region involved into it.

Let us explain in more detail how to estimate $\mathrm{Prob}(x(p) > 0)$ in Example 1 if $\mathrm{Prob}(\zeta = 0) = 1 - \varepsilon$ and $\mathrm{Prob}(\zeta = 1) = \varepsilon$. In this case we do not need to discriminate between positive values of states. In other words, we may assume that the process is defined by

$$x(v) = \begin{cases} y(v) & \text{if} \quad t(v) < 0 \\ \max\{\mathbf{f}(x(N(v))),\, h(v)\} & \text{if} \quad t(v) \geqslant 0 \end{cases} \tag{7}$$

rather than by (1). This process was examined long ago[9,10]; in the survey[14] it is described as Example 1.2 and explained on pp. 9–10 and 74–78. The percolation graph $\bar{V}'$ in this case is planar and therefore has a dual one. Every path in the dual graph serves as a "fence" which separates our point p from initial points. (One fence is shown in Fig. 1.1 on p. 10 in ref. 14.) $x(p) > 0$ iff there is a fence, along which all the hidden variables are positive. In the infinite case it is easy to prove that the number of those fences, which include k hidden variables, does not exceed 27^k. This leads to the estimation

$$\mathrm{Prob}(x(p) \geqslant 1) \leqslant \sum_{k=1}^{\infty} (27\varepsilon)^k$$

In the same vein $x(p) \geqslant q$ if at least q fences separate our point p from initial points, whence

$$\mathrm{Prob}(x(p) \geqslant q) \leqslant \left(\sum_{k=1}^{\infty} (27\varepsilon)^k \right)^q$$

which proves the first assertion of Theorem 2 for Example 1. Substituting (7) for (1) works for Example 2 also: with a minor modification it turns it into Example 1.5 of ref. 14 (which comes originally from an earlier paper[11]). Although Example 2 does not allow such a straightforward percolation representation, our combinatorial constructions come to contours in this case also.

5. THE LINEAR LEMMA

The following lemma clarifies the geometrical meaning of the condition $\sigma = \{\mathcal{O}\}$. Remember that a linear function L on a linear space with the origin $\mathcal{O}$ is called homogeneous if $L(\mathcal{O}) = 0$.

Lemma 1. Condition $\sigma = \{\mathcal{O}\}$ is equivalent to the following: There are such a number $r \leqslant d+1$ and such r homogeneous linear functions $L_1, \ldots, L_r$ on R^{d+1} that

$$L_1 + \cdots + L_r \equiv -t \tag{8}$$

and

$$\text{for every } i \text{ the set } \{e \in N(\mathcal{O}) \mid L_i(e) \leqslant 0\} \text{ is an } \mathcal{O}\text{-drag} \tag{9}$$

Proof. First, assume that there are such homogeneous linear functions $L_1, \ldots, L_r$ on R^{d+1} that (8) and (9) hold. Assume that there is some nonzero point $v \in \sigma$ and come to a contradiction. Note that $t(v) < 0$. For $i = 1, \ldots, n$ denote $S_i = \{w \in N(\mathcal{O}) \mid L_i(w) \leqslant 0\}$. From (9), S_i is an $\mathcal{O}$-drag, whence $v \in \mathrm{ray}(\mathrm{conv}(S_i))$, which means that there is $k > 0$ for which $k \cdot v \in \mathrm{conv}(S_i)$. Since values of L_i on all elements of S_i are nonpositive, its value on $k \cdot v$ is nonpositive also. Therefore, the value of L_i on v is nonpositive. Thus, the values of all our functions $L_1, \ldots, L_r$ on v are nonpositive, which contradicts (8).

Now assume that $\sigma = \{\mathcal{O}\}$ and prove the existence of functions L_i in question. For any finite $S \subset R^{d+1}$ the set $\mathrm{ray}(\mathrm{conv}(S))$ can be represented as an intersection of several half-spaces. (A half-space is a set in R^{d+1}, where some homogeneous linear function is nonnegative.) Apply this to $\mathcal{O}$-drags, and one has several drag-half-spaces whose intersection is $\{\mathcal{O}\}$. (A drag-half-space is a half-space whose intersection with $N(\mathcal{O})$ is an $\mathcal{O}$-drag.) For every one of these drag-half-spaces we introduce a homogeneous linear function f_i which is nonpositive on it and only on it. We know that the origin is the only point in R^{d+1} where all f_i are nonpositive. This allows us to apply Theorem 21.3 of ref. 8 (a version of Helly's theorem) to the hyperplane $\Pi = \{(s, t) \mid t = -1\}$, restrictions of our functions to it, and any nonempty closed convex set $C \subseteq \Pi$. We take $C = \Pi$. Thus, there exist such nonnegative real numbers λ_i, at most $d+1$ of which are positive, that for some $\varepsilon > 0$

$$\forall v \in C: \sum_i \lambda_i \cdot f_i(v) \geqslant \varepsilon$$

Since the left part is linear and bounded from below by a positive constant on C, it has to be a positive constant on C:

$$\forall v \in C: \sum_i \lambda_i \cdot f_i(v) = \delta = \text{const} \geqslant \varepsilon$$

whence

$$\forall v \in R^{d+1}: \sum_i \lambda_i \cdot f_i(v) = -\delta \cdot t$$

Thus the functions $L_i = \lambda_i \cdot f_i$ fit (8) and (9). ∎

Whenever the functions in question exist (that is, $\sigma = \{\mathcal{O}\}$), we choose the smallest available value of r and some functions $L_1, ..., L_r$. We also denote

$$\Lambda = \max\{L_i(v) \mid i = 1, ..., r, \text{norm}(v) \leqslant 1\}$$

Let us see what the Linear Lemma gives for our examples. In Example 0, $\sigma \neq \{\mathcal{O}\}$, whence the functions do not exist. When the functions exist, the smallest possible value of their number r is 2. This takes place in Example 1, and we can take $L_1 = s - t$, $L_2 = -s$, where s is the space coordinate. In Example 2 also $r = 2$ and we can take $L_1 = s_1 - t$, $L_2 = -s_1$, where s_1 is the first space coordinate. Generally, the constructions of our Part II come to the traditional (that is, nonbranching) contours whenever $r = 2$. (See ref. 11, which concentrates on this case.)

In Examples 3–5, $r = 3$ and we take $L_1 = s_1$, $L_2 = s_2$, $L_3 = -s_1 - s_2 - t$. Whenever $r \geqslant 3$, constructions of our Part II ramify. However, Example 3 is also more easy to manage than the general case. (I thank one of my referees who noticed this.) Here again we can dump all the positive values of $x(p)$ together, when speaking about $\text{Prob}(x(v) > 0)$, that is, use (7) instead of (1). This turns our Example 3 into Example 1 of ref. 12, whereby existence of some critical behavior in the infinite case is assured: when t tends to ∞, the percentage of zeros tends to a limit which equals 0 for large values of ε, but is positive for small values of ε. The former fact is well known (see, for example, Proposition 2.17 in ref. 14), the latter one follows from Theorem 5 in ref. 12. In the same way the main result of ref. 2 follows from the more general Proposition 2 of our Part II.

To prove our theorems for Examples 4 and 5, the constructions developed in our Part II seem to be necessary, such as they are.

6. SEEDS AND STEMS

Take any initial condition y, all of whose values are zeros and ones. From (2), the resulting trajectory's values are also zeros and ones. Let us call the sets

$$\{v \in V_{\text{init}} \mid y(v) = 1\} \qquad \text{and} \qquad \{v \in V \mid y(v) = 1\} \tag{10}$$

a *seed* and a *stem*, which grows from it, if the former is finite and the latter is infinite. We are going to prove that seeds exist. For any A, $B \subset \mathbf{R}^{d+1}$ and $c \in \mathbf{R}^{d+1}$ denote

$$A + c = \{a + c \mid a \in A\} \qquad \text{and} \qquad A + B = \{a + b \mid a \in A, b \in B\}$$

Call $A + c$ the *shift* of A by the vector c. Say that $A \subset \mathbf{R}^{d+1}$ is *obtuse* for $B \subset \mathbf{R}^{d+1}$ if

$$\forall c \in \mathbf{R}^{d+1}\colon (A + c) \cap B = \varnothing \Rightarrow (A + c) \cap \operatorname{conv}(B) = \varnothing$$

Note that any sets A, B, $C \subset \mathbf{R}^{d+1}$ given, if A is obtuse for B, then any shift of $A + C$ is obtuse for any shift of B. Also it is known that for any bounded $A \subset \mathbf{R}^{d+1}$ there is a bounded obtuse set, because the set $-k \cdot \operatorname{conv}(A)$ is obtuse for A whenever $k \geqslant d + 1$ (see Lemma 11 in ref. 12). Hence for any finite collection of bounded sets and any positive number R there is a bounded set which is obtuse for all of these sets and contains a sphere with a radius R. Now take any large enough bounded set P whch is obtuse for all drags, and so shifted that the time coordinates of all its points are less than $-\mathbf{m}$. Define $S = (P - \rho) \cap V_{\text{init}}$ and prove that S is a seed. For this purpose let us prove by contradiction that the resulting stem contains the set $T = (P - \rho) \cap V$. Assume that T contains some points where the resulting trajectory equals 0 and take one of them, v, whose time is minimal. To make the state of v equal to zero, some drag of v, say $v + U$, must be filled with zeros. Let us prove that this is impossible. Note that $P - \sigma$ cannot intersect $v + U$ because of the way we chose v. But, since P is obtuse for all drags (including U), $P - \sigma$ also is, whence $P - \sigma$ cannot intersect $v + \operatorname{conv}(U)$. But, on the other hand, ray(conv(U)) contains σ, whence $\operatorname{conv}(U)$ intersects all half-lines emanating from the origin, which belong to σ, whence $v + \operatorname{conv}(U)$ intersects all half-lines emanating from v, which belong to $v - \sigma$. Thus we get a contradiction, because any segment whose ends are v and any element of P is a piece of such a half-line and it is long enough to intersect $v + \operatorname{conv}(U)$. ∎

7. PROOF OF PROPOSITION 3

From now on, shifts of seeds will be called seeds also. Take any system S whose $\sigma \neq \{\mathcal{O}\}$. Of course, σ contains a point $(s, -1)$ at which all the components of s are rational: $s = (s_1,..., s_d)$, where $s_1,..., s_d \in \mathcal{Q}$. Let k be the least common denominator of $s_1,..., s_d$. Let us consider another system S' with the same transition function, same time components of the neighbor vectors, and the following space components of the neighbor vectors: $\Delta s'_i = k \cdot (\Delta s_i - s_i \cdot \Delta t_i)$. Note that $\mathbf{E}_t$ is the same for both systems for all t and that σ of the new system contains the point $(0, -1)$. Thus we have reduced the general case to the case when $(0, -1) \in \sigma$.

Now S is a system whose σ contains the point $(0, -1)$. Existence of seeds assures us that we can minorate this system (in the sense of pp. 27–29 of ref. 14) by a *wary* process, which is defined as follows. [Wary processes do not fit into our general rule (1).] The configuration space is the same as that of the original process and the initial condition is "all zeros" also. But the neighborhoods $W(\cdot)$ of the wary process are larger that those of the original process. To define them, choose a large enough constant C and for every inner point v define $W(v)$ as follows:

$$W(v) = \{w \mid \text{dist}(w, v) \leqslant C,\ t(w) < t(v)\}$$

Whenever $t > 0$, define the map as follows:

$$x(s, t) = \begin{cases} x(s, t-1) + h(s, t) & \text{if} \quad x(s, t-1) \leqslant \min\{x(w) \mid w \in W(s, t)\} \\ x(s, t-1) & \text{otherwise} \end{cases}$$

Here every $h(s, t)$ is independent of all the other hidden variables, and

$$h(s, t) = \begin{cases} 1 & \text{with probability } \varepsilon \\ 0 & \text{with probability } 1 - \varepsilon \end{cases}$$

Here ε is a positive parameter. The idea behing these definitions is to choose ε less than the probability that hidden variables of all points of a seed would be positive in the original process. Note also that in any wary process states of elements of the space never decrease as time goes on. Speaking about wary processes, we may use the same notation $\mathbf{E}_t$ and term "linear lower bound" as before. We have reduced Proposition 3 to the following: In every wary process (with $\varepsilon > 0$) the expectation $\mathbf{E}_t$ has a linear lower bound. To prove this, it is sufficient to find a constant T for which

$$\mathbf{E}(x(s, kT)) \geqslant \text{const} \cdot k \qquad \text{for all} \quad k = 0, 1, 2,...$$

Now note that $\mathbf{E}(x(s, kT)) \geqslant \mathbf{E}(x'(s, k))$, where $x'(s, k)$ are variables of the same wary process with $\varepsilon' = 1 - (1 - \varepsilon)^T$. (This is because the opportunities to grow accumulate during T time steps.) Thus, it is sufficient to find some $\varepsilon < 1$ with which our wary process has a linear lower bound. But this is certainly possible because every wary process is minorated by the percolation process (Section 3) with the same ε and neighbors. ∎

8. PROOF OF PROPOSITION 4. CASE $\sigma = \{\mathcal{O}\}$

Call an initial condition *nonnegative* if all its components are nonnegative.

Proposition 5. Take any infinite deterministic system with $\sigma = \{\mathcal{O}\}$. For any finitary nonnegative initial condition y:

1. The resulting trajectory y^{tr} is also finitary.
2. $\mathrm{diam}(\mathrm{dev}(y^{\mathrm{tr}})) \leqslant \mathrm{const} \cdot (\mathrm{diam}(\mathrm{dev}(y)) + \max(y))$

Each of these assertions implies Proposition 4 in one direction. However, we prove both, because assertion 1 is more easy to prove, but assertion 2 is more precise. In both cases we need only the right inequality of (2) of our standard assumptions.

8.1. Proof of Assertion 1 of Proposition 5

This can be proved by induction based on the following: Consider the infinite case. Assume $\sigma = \{\mathcal{O}\}$. Take any finitary nonnegative initial y. Then:

(a) $\max(y \mid V^{\mathrm{tr}}) = \max(y \mid V_{\mathrm{init}})$.

(b) The set $\{v \mid y^{\mathrm{tr}}(v) = \max(y \mid V_{\mathrm{init}})\}$ is finite.

Here (a) comes from the right inequality of (2). Let us prove (b). Take any v where $y^{\mathrm{tr}}(v) = \max(y)$. Any drag of v must contain a point u where $y^{\mathrm{tr}}(u) = y^{\mathrm{tr}}(v)$. Thus for every $i = 1,..., r$ there is a point $u \in N(v)$ where $y^{\mathrm{tr}}(u) = y^{\mathrm{tr}}(v)$ and $L_i(u) \leqslant L_i(v)$. Applying this argument inductively, we obtain an initial point w_i where $y(w_i) = \max(y)$ and $L_i(w_i) \leqslant L_i(v)$ for every $i = 1,..., r$. Choose any $w_0 \in \mathrm{dev}(y)$. Then, using (8),

$$t(v) = -\sum_{i=1}^{r} L_i(v) \leqslant -\sum_{i=1}^{r} L_i(w_i) = \sum_{i=1}^{r} L_i(w_0 - w_i) - \sum_{i=1}^{r} L_i(w_0)$$

$$\leqslant r \cdot \Lambda \cdot \mathrm{diam}(\mathrm{dev}(y)) + t(w_0) < r \cdot \Lambda \cdot \mathrm{diam}(\mathrm{dev}(y)) \quad ∎$$

8.2. Proof of Assertion 2 of Proposition 5

Choose some $i \in \{1, ..., r\}$ and let L_i be one of the functions provided by Lemma 1. Given any infinite increasing sequence K of real numbers $k_1 < k_2 < k_3 < \cdots$, let a (i, K)-*stairs* denote the following configuration z:

$$\forall v \in V\colon z(v) = \begin{cases} 0 & \text{if} \quad L_i(v) < k_1 \\ j & \text{if} \quad k_j \leqslant L_i(v) < k_{j+1} \end{cases}$$

(A similar definition was used in ref. 6 for the one-dimensional case.) Call a configuration x a *supertrajectory* if $x(v) \geqslant \mathbf{f}(x(N(v)))$ for every inner point v.

Lemma 2. Assume

$$\forall j\colon k_{j+1} - k_j \geqslant 2 \cdot \Delta \cdot \Lambda \tag{11}$$

Then the (i, K)-stairs is a supertrajectory for any $i = 1, ..., r$.

Proof. Let z be our (i, K)-stairs and denote $\operatorname{var}(z \mid S) = \max(z \mid S) - \min(z \mid S)$. First prove that $\operatorname{var}(z \mid N(v)) \leqslant 1$ for every inner v. Assume the contrary: there are such $u_1, u_2 \in N(v)$ that $z(u_2) - z(u_1) \geqslant 2$. Then there is such j that $L_i(u_1) < k_j$ and $L_i(u_2) \geqslant k_{j+1}$, whence

$$L_i(u_2) - L_i(u_1) > k_{j+1} - k_j \geqslant 2 \cdot \Delta \cdot \Lambda$$

But this is impossible, since

$$L_i(u_2) - L_i(u_1) = L_i(u_2 - v) - L_i(u_1 - v) \leqslant 2 \cdot \Delta \cdot \Lambda$$

Now assume that $z(v) < \mathbf{f}(z(N(v)))$ for some v. Since $\operatorname{var}(z \mid N(v)) \leqslant 1$, all the values of z at $N(v)$ are in the range $[j, j+1]$ for some j. Thus, from the right inequality of (2), $\mathbf{f}(z(N(v))) \leqslant j+1$. Then, from our assumption, $z \leqslant j$. From (9) the set $\{e \in N(v) \mid L_i(e) \leqslant L_i(v)\}$ is a v-drag. Hence we see that $z(v)$ cannot be less than j, because it would mean that the empty set is a v-drag. Thus $z(v) = j$, and the set $\{e \in N(v) \mid z(e) = j\}$ is a v-drag, whence $\mathbf{f}(z(N(v))) = j$, which contradicts our assumption. ∎

Now to prove assertion 2. Call $\max(t \mid \operatorname{dev}(x))$ the *lifetime* of any configuration x. All we need to prove is that lifetime of any nonnegative initial y does not exceed

$$\text{const} \cdot (\operatorname{diam}(\operatorname{dev}(y)) + \max(y \mid V_{\text{init}})) \tag{12}$$

For any real number c denote

$$\operatorname{Int}(c) = \begin{cases} \text{the least integer which is not less than } c & \text{if} \quad c > 0 \\ 0 & \text{if} \quad c \leqslant 0 \end{cases}$$

For any point v_0 and any constants $C>0$ and C' the configuration

$$x(v)=\mathrm{Int}(C\cdot L_i(v-v_0)+C')$$

is a stairs. This stairs satisfies (11), provided $C\leqslant(2\cdot\Delta\cdot\Lambda)^{-1}$. Now take any finitary nonnegative initial configuration y, choose any $v_0\in\mathrm{dev}(y)$, and define another configuration

$$x(v)=\min\{\mathrm{Int}(C\cdot L_i(v-v_0)+C')\mid i=1,\dots,r\}$$

where

$$C=(2\cdot\Delta\cdot\Lambda)^{-1}\qquad\text{and}\qquad C'=C\cdot\Lambda\cdot\mathrm{diam}(\mathrm{dev}(y))+\max(y\mid V_{\mathrm{init}})$$

Let us prove that $y(v)\leqslant x(v)$ for every initial v. This is evident if $y(v)=0$. Now let $y(v)>0$. Then

$$\begin{aligned}x(v)\geqslant C\cdot L_i(v-v_0)+C'\geqslant C'-C\cdot\Lambda\cdot\mathrm{norm}(v-v_0)&\\ \geqslant C'-C\cdot\Lambda\cdot\mathrm{diam}(\mathrm{dev}(y))&\\ \geqslant\max(y\mid V_{\mathrm{init}})\geqslant y(v)&\end{aligned}$$

Now the relation $y^{\mathrm{tr}}\preccurlyeq x$ in general can be proved by induction, using the standard monotonicity arguments. (Remember that Note 2 allows us to assume that the transition function is monotonic.) Thus our estimate (12) for the lifetime of y follows from the same estimate for the lifetime of x and it remains to prove the latter one. Take any point v where $x(v)>0$. For all $i=1,\dots,r$, $\mathrm{Int}(C\cdot L_i(v-v_0)+C_i)>0$, whence $C\cdot L_i(v-v_0)+C_i>0$, whence $L_i(v-v_0)>-C_i/C$. Summing the last inequalities and using (8) proves that

$$t(v)<t(v-v_0)\leqslant\frac{r\cdot C'}{C}=r\cdot\Lambda\cdot\mathrm{diam}(\mathrm{dev}(y))+2\cdot r\cdot\Delta\cdot\Lambda\cdot\max(y\mid V_{\mathrm{init}})\quad\blacksquare$$

9. PROOF OF PROPOSITION 4. CASE $\sigma\neq\{\mathcal{O}\}$

Proposition 6. Take any infinite deterministic system with $\sigma\neq\{\mathcal{O}\}$. For any natural number T there is such a finitary initial condition y that for every t there is such s that $y^{\mathrm{tr}}(s,t)=T$.

Proposition 6 completes the proof of Proposition 4. For the case $T=1$ we have already proved it when we proved the existence of seeds (Section 6). To prove it in general, take any seed S and a large enough constant R and define T seeds $S_1,\dots,S_T$ as follows: S_i is the intersection of

V_{init} with the (vector) sum of S and the sphere with the center $\mathcal{O}$ and the radius $(T-i)\cdot R$. (Thus S_T coincides with S.) Now define a finitary initial condition y as follows:

$$y(v)=\begin{cases}\max\{i \mid v\in S_i\} & \text{if } \ v\in S_1\\ 0 & \text{otherwise}\end{cases}$$

Now every S_i is filled with states equal to i or greater, and is surrounded by a thick enough layer of states equal to $i-1$ or greater. Let us prove that in the resulting trajectory the ith stem will be filled with states equal to i or greater. This is based on the fact that

$$(\forall S\in\mathscr{D}\colon \max\{x_i \mid i\in S\}=\max\{x_1,..., x_n\})\Rightarrow \mathbf{f}(x_1,..., x_n)=\max\{x_1,..., x_n\} \tag{13}$$

for any $x_1,..., x_n\in\{0, 1\}$. ($\mathscr{D}$ is the family of drags.) To prove this by contradiction, take any $x_1,..., x_n\in\{0, 1\}$, assume $\max\{x_i \mid i\in S\}=\max\{x_1,..., x_n\}$ for all $S\in\mathscr{D}$, but $\mathbf{f}(x_1,..., x_n)<\max\{x_1,..., x_n\}$. Hence and from (2)

$$\mathbf{f}(x_1,..., x_n)=0 \qquad \text{and} \qquad \max\{x_1,..., x_n\}=1 \tag{14}$$

But from Note 1 the set $S=\{i \mid x_i=0\}$ is a drag, whence

$$\max\{x_1,..., x_n\}=\max\{x_i \mid i\in S\}=0$$

which contradicts (14). ∎

REFERENCES

1. Charles H. Bennett and G. Grinstein, Role of irreversibility in stabilizing complex and nonergodic behavior in locally interacting discrete systems, *Phys. Rev. Lett.* **55**(7):657–660 (1985).
2. Piotr Berman and Janos Simon, Investigations of fault-tolerant networks of computers, in *Proceedings of the 20-th Annual ACM Symposium on the Theory of Computing* (1988), pp. 66–77.
3. Maury Bramson and Lawrence Gray, A useful renormalization argument, in *Random Walks, Brownian Motion, and Interacting Particle Systems* (Birkhäuser Boston, Boston, Massachusetts, 1991).
4. B. Derrida, J. L. Lebowitz, E. R. Speer, and H. Spohn, Dynamics of an anchored Toom interface, *J. Phys. A: Math. Gen.* **24**:4805–4834 (1991).
5. Roland Dobrushin, Markov processes with a large number of locally interacting components—Existence of a limit process and its ergodicity, *Problems Information Transmission* 7(2):149–194 (1971).
6. G. A. Gal'perin, One-dimensional local monotone operators with memory, *Sov. Math. Dokl.* **17**(3):688–692 (1976).

7. Joel L. Lebowitz, Christian Maes, and Eugene R. Speer, Statistical mechanics of probabilistic cellular automata, *J. Stat. Phys.* **59**:117–168 (1990).
8. R. T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, New Jersey, 1970).
9. O. N. Stavskaya and I. I. Piatetski-Shapiro, Uniform networks of spontaneously active elements, *Problemy Kibernetiki* **20**:91–106 (1968) [in Russian].
10. Andrei Toom, A family of homogeneous nets of formal neurons, *DAS* **9**:1338–1341.
11. Andrei Toom, Non-ergodic multidimensional systems of automata, *Problems Information Transmission* **10**:239–246 (1974).
12. Andrei Toom, Stable and attractive trajectories in multicomponent systems, in *Multicomponent Random Systems*, R. Dobrushin and Ya. Sinai, eds. (Marcel Dekker, New York, 1980), pp. 549–576.
13. Andrei Toom, Estimations for measures that describe the behavior of random systems with local interaction, in *Interactive Markov Processes and their Application to the Mathematical Modelling of Biological Systems* (Pushchino, 1982), pp. 21–33 [in Russian].
14. A. Toom, N. Vasilyev, O. Stavskaya, L. Mityushin, G. Kurdyumov, and S. Pirogov, Discrete local Markov systems, in *Stochastic Cellular Systems: Ergodicity, Memory, Morphogenesis*, R. Dobrushin, V. Kryukov, and A. Toom, eds. (Manchester University Press, 1990).
15. Andrei Toom, On critical phenomena in interacting growth systems. Part II: Bounded growth, *J. Stat. Phys.*, submitted.
16. Leonid Vaserstein, Markov processes over denumerable products of spaces, describing large systems of automata, *Problems Information Transmission* **5**(3):47–52 (1969).