The Annals of Applied Probability
1993, Vol. 3, No. 2, 339–355

MARKOV MODELS OF STEADY CRYSTAL GROWTH

BY D. J. GATES AND M. WESTCOTT

CSIRO

We consider solid-on-solid models of two- and three-dimensional crystal growth described by Markov rate processes whose states are representations of the entire crystal surface and whose transitions are captures of single atoms by this surface or escapes of atoms. Under natural conditions on the transition probability rates, we prove ergodicity of the Markov process. This implies the existence of states of steady growth and statistically stationary surface structure. Microscopic surface instabilities do not therefore occur under such conditions. For two dimensions, our conditions are very general; for three dimensions, they are less general.

1. Introduction. Crystal growth may be described by Markov processes whose states are configurations of the crystal edge or surface and whose transitions are captures and escapes of particles therefrom ([12]–[15], [17]. [25]). One can give a mathematically exact analysis of steady states of growth in two dimensions ([8], [10], [11]). Two-dimensional models have important application to polymer crystallization (e.g., [1], [7], [16], [22]), where the crystals are flat or *lamellar*. The polymer chain folds itself into a very regular zigzag of segments as it attaches to the edge of a lamella. Here the "particles" are the segments. The analysis in [8], [10] and [11] provides explicit stationary distributions, steady growth rates and other quantities for a physically important class of rates [see (2.9)]. These rates lead to a form of detailed balance and to the dynamic reversibility of the process.

For more general rates, such distributions are unknown. Then it is natural to look for more qualitative descriptions of the growth process, such as the existence of stationary distributions; that is, the ergodicity of the Markov process.

Previously [12], we gave some rather stringent conditions for ergodicity in three dimensions. We also gave converse conditions that ensure transience or null recurrence, implying surface instability. These conditions have simpler two-dimensional counterparts (Theorems 1 and 2 herein). We shall give some much weaker conditions that ensure ergodicity (Theorems 3 and 4). The conditions have a simple physical interpretation: Transitions that tend to smooth the crystal edge are more likely. So the resulting ergodicity is very plausible. Two-dimensional models have application to other growth and aggregation processes (e.g. [3], [4]).

Received October 1990; revised July 1992.
AMS 1991 *subject classifications*. Primary 60J20; secondary 82A60.
Key words and phrases. Markov process, crystal growth, polymer crystal, Foster's theorem, Liapounov function, dynamic reversibility.

Understanding of three-dimensional crystal growth is much more limited. There are no exact solutions for steady growth analogous to the two-dimensional ones. There is no discrete event Markov model where a stationary distribution, under net growth conditions, is known [for continuous state space, such distributions are known, however ([5], [9]). Further, one can prove [12] that an important class of models has no dynamically reversible members. So stationary distributions may be quite complex and difficult to find. These basic deficiencies remain a major barrier to a deeper understanding of the physics of steady growth. Again we resort to finding conditions for ergodicity (Theorems 6 and 7).

2. Two-dimensional models. For the present, we consider only the capture of particles by the surface. A particle may be an atom, a polymer segment, a cell or some other entity depending on the application. A generalization, including the removal or escape of particles from the surface, is discussed in Section 5. Our model is a so-called *solid-on-solid model*, in which particles are represented as unit squares that form stacks with no overhangs. Consecutive columns in the array are labelled $1, 2, \dots, M$ and the stack may be represented by the vector of occupation numbers $\mathbf{n} = (n_1, \dots, n_M)$, where n_i is the height of column i. We are interested only in the upper edge or profile of such stacks, so that stacks differing in absolute height are regarded as equivalent. The state of the stack may therefore be represented by the vector of height differences or step sizes $\mathbf{h} = (h_1, \dots, h_M)$, where $h_i = n_i - n_{i-1}$, and we impose periodic boundary conditions, so that

$$(2.1) \qquad h_{M+i} = h_i \quad \forall\, i \quad \text{and} \quad \sum_{i=1}^{M} h_i = 0.$$

This is equivalent to growing the surface on a closed curve, which has mathematical advantages and has negligible effect when M is large, as it will be in most applications.

The time evolution of the surface is described by a time-homogeneous Markov process $\mathbf{h}(t)$, $t \geq 0$, whose state space H consists of vectors $\mathbf{h} \in Z^M$ satisfying (2.1). Assume $\mathbf{h}(t)$ is standard, so it has a Q matrix ([2], Section II.2) with typical element $q(\mathbf{h}, \mathbf{h}')$, giving the probability rate of the transition $\mathbf{h} \to \mathbf{h}'$. Transitions of the surface occur by addition of a particle to one column, so that

$$h_i \to h_i + 1, \qquad h_{i+1} \to h_{i+1} - 1$$

if the addition is at column i. This takes the state $\mathbf{h}$ to the new state $\mathbf{h}' = (h_1, \dots, h_i + 1, h_{i+1} - 1, \dots, h_M)$. So the transition probability rates $q(\mathbf{h}, \mathbf{h}')$ are zero unless $\mathbf{h}' = \mathbf{h}^i$ for some i, when we write

$$(2.2) \qquad q(\mathbf{h}, \mathbf{h}') = w_i(\mathbf{h}).$$

We assume throughout that there is $K < \infty$ such that $0 < w_i(\mathbf{h}) < K$ for $i = 1, \dots, M$, $\mathbf{h} \in H$. The total instantaneous rate out of $\mathbf{h}$ is denoted by

$$q(\mathbf{h}) = \sum_{i=1}^{M} w_i(\mathbf{h}). \tag{2.3}$$

Because $\sup_{\mathbf{h}} q(\mathbf{h}) < \infty$, $\mathbf{h}(t)$ is stable, conservative and uniquely determined by Q ([2], pages 154 and 251, Theorem II.19.1 and Corollary 2); that is, it is regular [23]. Because the w_i are positive, $\mathbf{h}(t)$ is irreducible, but because H is infinite, the existence of a stationary distribution is not assured a priori.

We assume henceforth that

$$w_i(\mathbf{h}) = w(-h_i, h_{i+1}), \tag{2.4}$$

where w is positive and bounded. Thus the rates are determined by the local topography. Our main purpose in this article is to give conditions on w that ensure the existence of a stationary distribution. First we outline some previous results. These concern the special case (cf. [10], [11])

$$w(x, y) = \begin{cases} \beta_0, & \text{if } x, y \le 0, \\ \beta_1, & \text{if } x \le 0 \text{ and } y > 0, \text{or } x > 0 \text{ and } y \le 0, \\ \beta_2, & \text{if } x, y > 0, \end{cases} \tag{2.5}$$

or, equivalently, if site i has $n = 0, 1, 2$ neighbor columns higher than itself, then

$$w_i(\mathbf{h}) = \beta_n. \tag{2.6}$$

Note that $1 + n$ is the number of interparticle contacts that are made when a new particle is added. So in this special case, the rates depend only on the signs of the steps, not their magnitudes. This type of model has been extensively studied by computer simulation and approximate methods (e.g., [13]–[15], [17], [22], [25]). In [12] we proved some results concerning the ergodicity of a three-dimensional version of this model. In two dimensions these reduce to the following theorems.

THEOREM 1. *If*

$$(M - 1)^2 \beta_0 < \beta_2, \tag{2.7}$$

then the process $\mathbf{h}(t)$ is ergodic. For $M \le 4$ it is ergodic provided

$$\beta_0 < \beta_1 < \beta_2. \tag{2.8}$$

THEOREM 2. *If $\beta_0 > \beta_2$, the process $\mathbf{h}(t)$ is transient; if $\beta_0 = \beta_2$, it is null; if $\beta_0 = \beta_1 = \beta_2$, then the process is null recurrent for $M = 2, 3$ and transient if $M \ge 4$.*

The proofs of these theorems are simplified versions of those in [12]. We shall not give details here because they are not our main concern. Note that

(2.8) is a much weaker condition than (2.7). Just prior to Theorem 4, we show that (2.8) is, in a sense, close to ergodic for all M.

Besides these general results, we gave some very detailed exact results ([10], [11]) in the case

$$\beta_0 + \beta_2 = 2\beta_1, \qquad \beta_0 < \beta_1 < \beta_2. \tag{2.9}$$

This process is ergodic for all M and so may be appended to Theorem 1. But we also found the stationary distribution and growth rate, among other things. Note that (2.9) is consistent with (2.8) and suggests again that (2.7) is excessively strong. We now state our main new result.

THEOREM 3. *Suppose* $w_i(\mathbf{h})$ *satisfies* (2.4), *with*

(i) $w(x, y) \equiv w(y, x)\ \forall\ x, y$.
(ii) $w(x, y)$ *is nondecreasing in* y *for each* x.
(iii) $w(x, y)$ *is ultimately nonconstant in* y *in at least one tail*, *for each* x.

Then the process $\mathbf{h}(t)$ *is ergodic.*

Note that (i) and (ii) imply that $w(x, y)$ is also nondecreasing in x for each y. A related implication can be drawn from (i) and (iii). A proof of the theorem is given in Section 4.

Condition (i) is a natural form of left–right symmetry, which should be satisfied in most applications. It might fail if particles have a preferred direction of approach to the surface. Condition (ii) says that a site with higher neighbours than another has at least as large a capture rate. Neighbours represent new interparticle contact for a captured particle. (In physical theory, energy considerations imply that such contacts are favorable—they lower energy—and so enhance capture rates.) Condition (iii) says that sufficiently high neighbours give a strict capture advantage or that sufficiently low neighbours give a strict disadvantage.

The effect of conditions (ii) and (iii) is to give higher rates to transitions that have a greater smoothing effect on the surface. Smoothing tends to encourage recurrence of the flat state $\mathbf{h} = \mathbf{0}$. The resultant ergodicity is therefore rather plausible.

The rates (2.5), subject to (2.8), satisfy (i) and (ii), but not (iii). Suppose, however, that $w(x, y) = w_{(1)}(x, y) + w_{(2)}(x, y)$, where $w_{(1)}$ is given by (2.5) with $\beta_0 \le \beta_1 \le \beta_2$ and $w_{(2)}$ satisfies (i)–(iii). Then the process $\mathbf{h}(t)$ is ergodic because w satisfies (i)–(iii). Because $w_{(2)}$ can be made arbitrarily small compared to the β's, we can find models, whose rates are arbitrarily close to the β's, that are ergodic under condition (2.8) for all M. In this sense, (2.8) is close to a sufficient condition for ergodicity for all M, not just $M \le 4$.

The conditions of Theorem 2 imply that *exposed* sites are favoured, and are directly opposed to the ergodic conditions. (Physical theory says that when crystal growth is fast, the limited mobility of particles tends to favour exposed sites. Surface instability is then observed.)

THEOREM 4. *Suppose* $w_i(\mathbf{h})$ *satisfies* (2.4) *and, for some* $c > 0$,

$$(2.10) \qquad (x+y)\{w(x,y) - c\} \to \infty \quad \text{as } |x+y| \to \infty.$$

Then the process $\mathbf{h}(t)$ *is ergodic.*

The proof is given in Section 4, and it differs markedly from the proof of Theorem 3. The conditions of Theorems 3 and 4 appear technically rather different. To illustrate their differences and similarities, we consider some special cases of $w(\cdot)$ that are of physical interest.

CASE A. $w(x, y) = \phi(x + y)$.

CASE B. $w(x, y) = \phi(x) + \phi(y)$.

CASE C. $w(x, y) = a + \phi(x)\phi(y)$.

Case A implies that $w_i(\mathbf{h}) = \phi(n_{i-1} - 2n_i + n_{i+1})$, so that the rate of capture at site i depends only on the second difference of column heights; that is, the local curvature of the edge. Similar rates have been used previously in both microscopic models ([5], [9]) and mesoscopic models of dendrite growth ([19], (4.11) and [20]).

To justify Case B, suppose that neighbour columns independently attract a particle to site i with probability rates $\phi(-h_i)$ and $\phi(h_{i+1})$ for left and right neighbours, respectively. Suppose that capture occurs if *either* of the attraction events occurs. Then capture at site i in time dt is the union of independent attraction events, so that

$$(2.11) \quad w_i(\mathbf{h})\, dt = \phi(-h_i)\, dt + \phi(h_{i+1})\, dt - \phi(-h_i)\phi(h_{i+1})(dt)^2,$$

from which Case B follows. Note that (2.9) is of this form.

Rates of the form of Case C occur in the two-dimensional version of the well-known process of Gilmer and Jackson [15]. They obtain Case C from energy equilibrium considerations.

Obviously, Theorem 3 holds in each case if $\phi(\cdot)$ is positive, bounded, nondecreasing and ultimately nonconstant in at least one tail. Call this $\phi \in J$. But for Case A, Theorem 4 holds under the much weaker condition

$$(2.12) \qquad \text{for some } c > 0, \qquad u\{\phi(u) - c\} \to \infty \quad \text{as } |u| \to \infty.$$

On the other hand, for Cases B and C, (2.10) is not satisfied by all $\phi \in J$. For example, in Case C, choose $\phi(\cdot)$ to be a distribution function with $0 < \phi(u) < 1$ for all finite u. For any proposed choice of $c > a$ in (2.10), fix x so that $\phi(x) < c - a$. Then $w(x, y) - c \le 0$ for all y and (2.10) can never hold as $y \to \infty$. Examples where (2.10) does hold are in Case B ϕ is skew symmetric and in Case C $\log \phi$ is skew symmetric.

3. Three-dimensional models. In the three-dimensional solid-on-solid model, particles are regarded as unit cubes and are stacked on the unit squares of a portion of the integer lattice Z^2, a large chessboard, say. As before, we are interested only in the surfaces of such stacks, so that surfaces differing only in absolute height are regarded as equivalent states. The centres of the columns (lying on a shifted Z^2) are labelled (i, j), where $i = 1, \dots, M$ and $j = 1, \dots, N$. If $n_{i,j}$ is the height at site (i, j), we put (note the labelling is different from [12]),

$$g_{i,j} = n_{i,j} - n_{i-1,j}, \qquad h_{i,j} = n_{i,j} - n_{i,j-1}, \tag{3.1}$$

representing height differences between columns at (i, j) and columns to the west and to the south, respectively. We again assume periodic boundary conditions, which here become

$$g_{i+M,j+N} = g_{i,j} \quad \text{for all } i, j, \tag{3.2}$$

and likewise for $h_{i,j}$. It then follows that

$$\begin{aligned} \sum_i g_{i,j} &= 0 \quad \text{for all } j, \\ \sum_j h_{i,j} &= 0 \quad \text{for all } i. \end{aligned} \tag{3.3}$$

Conditions (3.2) imply that, effectively, growth occurs on the surface of a torus. For large M and N, edge effects are expected to be unimportant, so that this topological device should be innocuous, and it has mathematical advantages.

With the periodic boundary conditions, the definitions (3.1) extend to all (i,j). Because the net height difference around any circuit is zero, we have

$$g_{i,j} + h_{i-1,j} - g_{i,j-1} - h_{i,j} = 0. \tag{3.4}$$

Then our state space H_1 comprises all vectors

$$\mathbf{h} = \{g_{i,j}, h_{i,j}: i = 1, \dots, M, j = 1, \dots, N\} \in Z^{MN} \tag{3.5}$$

satisfying (3.2)–(3.4).

The time evolution of the surface is again assumed to be a time-homogeneous standard Markov process $\mathbf{h}_1(t)$, $t \geq 0$, with state space H_1 and rates $q_1(\mathbf{h}, \mathbf{h}')$. Again transitions occur by addition of a particle to one column:

$$\mathbf{h} \to \mathbf{h}',$$

where, for some i, j,

$$\begin{aligned} g'_{i,j} &= g_{i,j} + 1, \qquad & h'_{i,j} &= h_{i,j} + 1, \\ g'_{i+1,j} &= g_{i+1,j} - 1, \qquad & h'_{i,j+1} &= h_{i,j+1} - 1, \end{aligned} \tag{3.6}$$

while other $g_{k,l}$ and $h_{k,l}$ are unchanged. The state $\mathbf{h}'$ defined by (3.6) is denoted $\mathbf{h}^{i,j}$, so $q_1(\mathbf{h}, \mathbf{h}')$ is zero unless $\mathbf{h}' = \mathbf{h}^{i,j}$ for some i, j, when we write $q_1(\mathbf{h}, \mathbf{h}^{i,j}) = w_{i,j}(\mathbf{h})$. We assume throughout that there is $K < \infty$, such that

$0 < w_{i,j}(\mathbf{h}) < K$ for $i = 1, \ldots, M$, $j = 1, \ldots, N$, $\mathbf{h} \in H_1$, and define

$$q_1(\mathbf{h}) = \sum_{i=1}^{M} \sum_{j=1}^{N} w_{i,j}(\mathbf{h}).$$

As before, the Markov process $\mathbf{h}_1(t)$, assumed standard, is stable, conservative, regular and irreducible. By analogy with (2.4), we assume henceforth that

$$w_{i,j}(\mathbf{h}) = w(-g_{i,j}, -h_{i,j}, g_{i+1,j}, h_{i,j+1}), \tag{3.7}$$

where w is positive and bounded, so local topography is again the surface feature determining rates.

Previously, in [12] we considered the special case

$$w(\mathbf{u}) = \beta_n \quad \text{if } n(\mathbf{u}) = n, \tag{3.8}$$

where $\mathbf{u} = (u_1, u_2, u_3, u_4)$ and $n(\mathbf{u})$ is the number of strictly positive components of the vector $\mathbf{u}$. Thus $n(-g_{i,j}, -h_{i,j}, g_{i+1,j}, h_{i,j+1})$ is the number of upsteps facing site (i, j); equivalently, $n + 1$ is the number of cube faces covered when a particle is added to column (i, j) or the number of new interatomic bonds formed. In [12] we proved the following result.

THEOREM 5. *If*

$$(MN - 1)^2 \beta_0 < \min(2K\beta_1, 4\beta_2, 2\beta_3, \beta_4), \tag{3.9}$$

where $K = \min(M, N)$, *the process* $\mathbf{h}_1(t)$ *is ergodic. For* $M = N = 2$, *it is ergodic if*

$$\beta_0 < \beta_2 < \beta_4. \tag{3.10}$$

We also proved transience or null recurrence under different conditions on the β_n. To state our new result, we define $u_. = u_1 + u_2 + u_3 + u_4$.

THEOREM 6. *Suppose* $w_{i,j}(\mathbf{h})$ *satisfies* (3.7) *and, for some* $c > 0$,

$$(u_.)\{w(\mathbf{u}) - c\} \to \infty \quad \text{as } |u_.| \to \infty. \tag{3.11}$$

Then the process $\mathbf{h}_1(t)$ *is ergodic.*

The proof is given in Section 4. Theorem 6 is a natural extension of the two-dimensional Theorem 4. We note from (3.7) that $u_.$ is typically

$$n_{i-1,j} + n_{i,j-1} + n_{i+1,j} + n_{i,j+1} - 4n_{i,j}, \tag{3.12}$$

which is the discrete Laplacian of $n_{i,j}$ and measures the local concavity of the surface (cf. [5], [9], [19], [20] again).

THEOREM 7. *Suppose* $w_{i,j}(\mathbf{h})$ *satisfies* (3.7) *and* $\phi(\cdot)$ *is a positive, bounded, nondecreasing function that is skew symmetric about some level* $c > 0$ *and is* *ultimately nonconstant. Then if w has the form*

$$w(\mathbf{u}) = \phi(u_p + u_q) + \phi(u_r + u_s), \tag{3.13}$$

where $p, q, r, s \in (1, 2, 3, 4)$ are all different, the process $\mathbf{h}_1(t)$ is ergodic. The same holds if w is a positive linear combination of the three distinct cases in (3.13), *not necessarily with the same ϕ's.*

The proof is given in Section 4. In Theorem 7, one possible combination is $u_1 + u_3$, which is typically $n_{i-1,j} + n_{i+1,j} - 2n_{i,j}$, a measure of the surface curvature along a straight line through site (i, j). Similarly a combination $u_1 + u_2$ is typically $n_{i-1,j} + n_{i,j-1} - 2n_{i,j}$, which measures the surface curvature around a corner through site (i, j). So in all the cases covered by (3.13) growth is again driven by hollowness of the surface, but now measured by pairs of neighbours of site (i, j). Perhaps the most physically plausible combined w is

$$w(\mathbf{u}) = \phi(u_1 + u_2) + \phi(u_2 + u_3) + \phi(u_3 + u_4) + \phi(u_4 + u_1). \tag{3.14}$$

Here each corner pair of sites makes a similar contribution to the total rate, and $w_{i,j}(\mathbf{h})$ is invariant with respect to $\pi/2$ rotations of the lattice Z^2.

We make two observations about the proofs of our theorems. First, it is sufficient to consider the embedded (or jump) chain, defined as the discrete-time Markov chain whose transitions are the successive state changes in the original process whenever they occur ([2], page 259 and [18], page 3). It has transition probabilities

$$p_i(\mathbf{h}) = w_i(\mathbf{h})/q(\mathbf{h}).$$

Because the w_i are positive and bounded, the embedded chain is itself irreducible and has the same character (ergodic, null recurrent or transient) as the original process. This is because the process is regular and hence: (i) chain and process are recurrent or not together ([23], Lemma 4.2(iv)); and (ii) chain and process are ergodic or not together ([21], Theorem 3).

Second, the proofs are based upon a general theorem of Foster [6] that involves an unspecified test function or Liapounov function. The success of the approach is governed by one's ingenuity in choosing this function; see the various examples in [12]. This limits the utility of the general theorem and partly accounts for the incompleteness of our results. We state the general theorem in the notation of the two-dimensional process.

PROPOSITION 1. *The process $\mathbf{h}(t)$ is ergodic if we can find a positive function $y(\mathbf{h})$ such that, for all but a finite number of $\mathbf{h} \in H$,*

$$\Sigma(\mathbf{h}) \equiv \sum_i p_i(\mathbf{h}) y(\mathbf{h}^i) \le y(\mathbf{h}) - 1 \tag{3.15}$$

and such that, for the exceptional finite set,

$$\Sigma(\mathbf{h}) < \infty. \tag{3.16}$$

The proposition is a simple extension of Foster's theorem 2 [6] (cf. [24], Theorem 6.1) and the foregoing statement (ii).

4. Proofs of theorems.

PROOF OF THEOREM 3. In Proposition 1, we choose

$$y(\mathbf{h}) = \sum_{i=1}^{M} h_i^2. \tag{4.1}$$

The same test function was used in [12] to prove essentially the second part of Theorem 1. Then

$$y(\mathbf{h}_i) = y(\mathbf{h}) + 2(h_i - h_{i+1} + 1), \tag{4.2}$$

so that

$$\Sigma(\mathbf{h}) = y(\mathbf{h}) + 2 - 2S(\mathbf{h})/q(\mathbf{h}), \tag{4.3}$$

where

$$S(\mathbf{h}) = \sum_{i=1}^{M} w_i(\mathbf{h})(h_{i+1} - h_i). \tag{4.4}$$

To prove (3.15) we need to show that

$$S(\mathbf{h}) > \tfrac{3}{2}q(\mathbf{h}) \tag{4.5}$$

for all but a finite set of $\mathbf{h}$. Because $q(\mathbf{h})$ is bounded, say $q(\mathbf{h}) \le c$, it is sufficient to show that

$$S(\mathbf{h}) > \tfrac{3}{2}c \tag{4.6}$$

for all but a finite set of $\mathbf{h}$. The following lemma is therefore sufficient for the proof of Theorem 3.

LEMMA 1. *For each positive, finite B, $\{\mathbf{h}: S(\mathbf{h}) \le B\}$ is bounded.*

PROOF. To prove the lemma, order the h_i as $t_1 \le \cdots \le t_M$ and define $T_j = [t_j, t_{j+1}]$, $j = 1, \dots, M - 1$.

REMARK 1. If $\mathbf{h} \ne \mathbf{0}$, at least two T_j are nondegenerate.

REMARK 2. Because $\sum_{i=1}^{M} h_i = 0$, at least one T_j contains zero, possibly as an end point.

Let U_i denote the interval between h_i and h_{i+1}, from smaller to larger. Then

$$S(\mathbf{h}) = \sum_{i=1}^{M} w_i(\mathbf{h}) \left\{ \sum_{j=1}^{M-1} (t_{j+1} - t_j) I(U_i \text{ covers } T_j) \right\} \operatorname{sgn}(h_{i+1} - h_i),$$

where $I(A)$ is the indicator function of the event A. So

$$\begin{aligned} S(\mathbf{h}) &= \sum_{j=1}^{M-1} (t_{j+1} - t_j) \left\{ \sum_{i=1}^{M} w_i(\mathbf{h}) \operatorname{sgn}(h_{i+1} - h_i) I(U_i \text{ covers } T_j) \right\} \\ &\equiv \sum_{j=1}^{M-1} (t_{j+1} - t_j) V_j, \quad \text{say}, \end{aligned} \tag{4.7}$$

where

$$V_j = \sum_{\substack{k:\, h_k \le h_{k+1} \\ U_k \supseteq T_j}} w_k(\mathbf{h}) - \sum_{\substack{l:\, h_{l+1} \le h_l \\ U_l \supseteq T_j}} w_l(\mathbf{h}). \tag{4.8}$$

Because $h_{M+1} = h_1$, each sum in (3.8) has the same number of terms and is nonempty. If we have any T_j, U_k and U_l from the sums in (3.8), then

$$h_k, h_{l+1} \le t_j \le t_{j+1} \le h_{k+1}, h_l. \tag{4.9}$$

From Theorem 3(i) and (ii), we deduce that, for such k, l,

$$\begin{aligned} w_k(\mathbf{h}) \equiv w(-h_k, h_{k+1}) &\ge w(-h_k, h_{l+1}) \\ &\ge w(-h_l, h_{l+1}) \equiv w_l(\mathbf{h}), \end{aligned} \tag{4.10}$$

which shows that

$$V_j \ge 0, \qquad j = 1, \dots, M. \tag{4.11}$$

So all the summands in (4.7) are nonnegative, whence

$$0 \le (t_{j+1} - t_j) V_j \le B, \qquad j = 1, \dots, M. \tag{4.12}$$

To use Theorem 3(iii) for $w(x, y)$ ultimately nonconstant in the upper tail, we express it as follows: Given any $\delta > 0$ sufficiently small and any x, y, there exists $b \equiv b(x, y, \delta)$ such that

$$w(x, b) \ge w(x, y) + \delta. \tag{4.13}$$

By Remarks 1 and 2, there is an index $\hat{j} \in (1, \dots, M - 1)$ such that $t_{\hat{j}} \le 0$, $t_{\hat{j}+1} > 0$. Choose any $d > 0$ and define $\tau = \min\{|t_{\hat{j}}|, d\}$, so τ is bounded. Now, given any $\delta > 0$ sufficiently small,

either $t_{\hat{j}+1} \leq b(\tau, 0, \delta)$, in which case it is already bounded,

$$
\begin{aligned}
\textbf{or}\quad w_k(\mathbf{h}) &\geq w(\tau, h_{k+1}) && \text{[by (4.9) and Theorem 3(i) and (ii)]}\\
&\geq w(\tau, b) && \text{[by (4.9), assumption and Theorem 3(ii)]}\\
&\geq w(\tau, 0) + \delta && \text{[by (4.13)]}\\
&\geq w(\tau, h_{l+1}) + \delta && \text{[by (4.9) and Theorem 3(ii)]}\\
&\geq w(-h_l, h_{l+1}) + \delta && \text{[by (4.9) and Theorem 3(i) and (ii)]}\\
&= w_l(\mathbf{h}) + \delta;
\end{aligned}
$$

that is,

$$V_{\hat{j}} \geq \delta. \tag{4.14}$$

So from (4.12) and (4.14),

$$0 < t_{\hat{j}+1} \leq t_{\hat{j}+1} - t_{\hat{j}} \leq B/\delta,$$

and we conclude that $t_{\hat{j}+1}$ is in any case bounded. A similar argument proves that $t_{\hat{j}}$ must also be bounded. Suppose now that $\hat{j} < M - 1$ and consider $j = \hat{j} + 1$. We know that $t_{\hat{j}+1} \equiv t$, say, is bounded. Given $\delta > 0$, sufficiently small,

$$
\begin{aligned}
\textbf{either}\ t_{\hat{j}+2} &\leq b(-t, t, \delta)\\
\textbf{or}\ w_k(\mathbf{h}) &\geq w(-t, h_{k+1})\\
&\geq w(-t, t) + \delta\\
&\geq w(-h_l, h_{l+1}) + \delta\\
&= w_l(\mathbf{h}) + \delta.
\end{aligned}
$$

So as before,

$$0 \leq t_{\hat{j}+2} - t_{\hat{j}+1} \leq B/\delta,$$

whence

$$0 \leq t_{\hat{j}+2} \leq t + B/\delta. \tag{4.15}$$

Therefore, $t_{\hat{j}+2}$ must be bounded. We can continue in the same way to prove that all t_j for $j \geq \hat{j}$ are bounded. To examine the remaining t_j, consider first

$j = \hat{j} - 1$. We know that $t_{\hat{j}} \equiv -t'$, say, is bounded below. Given $\delta > 0$ sufficiently small,

$$\begin{aligned} \textbf{either } |t_{\hat{j}-1}| &\le b(-t', t', \delta) \\ \textbf{or } w_k(\mathbf{h}) &= w(h_{k+1}, -h_k) \\ &\ge w(-t', -h_k) \\ &\ge w(-t', t') + \delta \\ &\ge w(h_{l+1}, -h_l) + \delta \\ &= w_l(\mathbf{h}) + \delta, \end{aligned}$$

and we conclude, as usual, that $t_{\hat{j}-1}$ must be bounded. Similarly, all t_j for $j < \hat{j}$ must be bounded.

The preceding proof has assumed that $w(x, y)$ is ultimately nonconstant in the upper tail. If in fact Theorem 3(iii) holds only for the lower tail, then a very similar line of reasoning holds, using Theorem 3(i) as above to switch arguments of w if necessary. Thus in either case Lemma 1 is proved. □

Proof of Theorem 4. We must show that Lemma 1, used in the proof of Theorem 3, still holds under (2.10). From (2.10) we see that, given any $\tau \ge 0$, there exists $K(\tau)$ such that

$$(x + y)\{w(x, y) - c\} \ge \tau \quad \text{if } |x + y| > K(\tau). \tag{4.16}$$

Now it is obvious from (4.4) that

$$S(\mathbf{h}) \equiv \sum_{i=1}^{M} \{w(-h_i, h_{i+1}) - c\}\{h_{i+1} - h_i\}. \tag{4.17}$$

By (4.16), any negative summand in (4.17) has $|h_{i+1} - h_i|$ bounded by $K(0)$, so the negative terms are all bounded because w is. Hence $S(\mathbf{h}) \le B$ implies that, for some positive finite B',

$$(h_{i+1} - h_i)\{w(-h_i, h_{i+1}) - c\} \le B', \qquad i = 1, \dots, M. \tag{4.18}$$

We deduce from (4.16) and (4.18) that

$$|h_{i+1} - h_i| \le K(B'), \qquad i = 1, \dots, M. \tag{4.19}$$

Because $\sum_{i=1}^{M} h_i = 0$, if $\mathbf{h} \ne \mathbf{0}$ there is at least one index $\hat{i}$ such that $h_{\hat{i}} \le 0$, $h_{\hat{i}+1} > 0$. From (4.19), both $h_{\hat{i}}$ and $h_{\hat{i}+1}$ must be bounded, and we can then deduce successively that $h_{\hat{i}+2}, h_{\hat{i}+3}, \dots, h_{\hat{i}-1}$ are also all bounded. This proves Lemma 1 under (2.10), which immediately gives us Theorem 4. □

Proof of Theorem 6. In Proposition 1 we choose

$$y(\mathbf{h}) = \sum_{i=1}^{M} \sum_{j=1}^{N} \left(g_{i,j}^2 + h_{i,j}^2\right). \tag{4.20}$$

This is the obvious three-dimensional analogue of (4.1), and was previously used in [12] to prove Theorem 5. Using (3.7), we can imitate the derivation of (4.3) to get

$$\Sigma(\mathbf{h}) = y(\mathbf{h}) + 4 - 2S_1(\mathbf{h})/q_1(\mathbf{h}), \tag{4.21}$$

where

$$S_1(\mathbf{h}) = \sum_{i=1}^{M} \sum_{j=1}^{N} w_{i,j}(\mathbf{h})\left(g_{i+1,j} - g_{i,j} + h_{i,j+1} - h_{i,j}\right). \tag{4.22}$$

Because $q_1(\mathbf{h})$ is bounded, it follows as in the proof of Theorem 3 that we must prove Lemma 1 for $S_1(\mathbf{h})$ under (3.11) when $w_{i,j}(\mathbf{h})$ has the form (3.7). A simple modification of the argument leading to (4.19) proves that if $S_1(\mathbf{h})$ is bounded, then

$$|g_{i+1,j} - g_{i,j} + h_{i,j+1} - h_{i,j}| \le K, \qquad i = 1,\dots,M,\ j = 1,\dots,N, \tag{4.23}$$

for some finite K.

To complete the proof, we must show that (4.23) implies the $g_{i,j}$ and $h_{i,j}$ are all bounded. One approach is to note that the expression inside the modulus in (4.23) is the discrete Laplacian of $n_{i,j}$, which can be regarded as a convolution equation in the $n_{i,j}$'s. This can be solved by Fourier transforms to express $n_{i,j}$, hence $g_{i,j}$ and $h_{i,j}$, as a weighted sum of Laplacians, and the desired conclusion follows. However, the details are messy, so we give an alternative argument that we shall, in any case, need again.

The $n_{i,j}$ have at least one global minimum, being a finite set. Without loss of generality, assume one is at site $i = 1$, $j = 1$ and $n_{11} = 0$. Then

$$n_{i,j} \ge 0, \qquad i = 1,\dots,M,\ j = 1,\dots,N. \tag{4.24}$$

Both $g_{21} - g_{11}$ and $h_{12} - h_{11}$ are nonnegative because (1, 1) is a minimum, so by (4.23) both are bounded. Additionally, $g_{21}, h_{12} > 0$, $g_{11}, h_{11} < 0$, so they are also all bounded. Hence $n_{21}, n_{12}, n_{M,1}, n_{1,N}$ are all bounded. Next consider site $i = 2$, $j = 1$. Then:

> **either** $g_{31} - g_{21}$ and $h_{22} - h_{21}$ have the same sign, in which case they are both bounded, by (4.23);
>
> **or** the signs are different, in which case one difference is nonpositive.

Because

$$\begin{aligned} g_{31} - g_{21} &= n_{31} - 2n_{21} + n_{11} \ge -2n_{21}, \\ h_{22} - h_{21} &= n_{22} - n_{21} + n_{2N} \ge -2n_{21}, \quad \text{both by (4.24)}, \end{aligned} \tag{4.25}$$

the boundedness of n_{21} implies that the nonpositive difference is bounded. Then (4.23) says both differences are bounded. Because g_{21} is bounded, g_{31} must also be bounded. Further, by (4.24),

$$-n_{21} \le h_{22} \le n_{21} + (h_{22} - h_{21}), \tag{4.26}$$

$$h_{22} - |h_{22} + h_{11}| \le h_{21} \le h_{22} + |h_{22} - h_{21}|. \tag{4.27}$$

Equation (4.26) shows that h_{22} is bounded, whence h_{21} is bounded from (4.27), and so n_{31} and n_{22} are also bounded. Continuing this argument along the sites with $j = 1$, we can conclude that

$$g_{i,1}, h_{i,2} \text{ and } n_{i,2} \text{ are all bounded for } i = 1, \dots, M. \tag{4.28}$$

Now use the same reasoning as j increases through the sites with any fixed i, because from (4.28), $h_{i,2}$ and $n_{i,2}$ are both bounded. This shows that, for each $i = 1, \dots, M$,

$$h_{i,j+1} \text{ and } g_{i+1,j+1} \text{ are bounded for all } j = 2, \dots, N. \tag{4.29}$$

The desired conclusion follows from (4.28) and (4.29). This establishes (3.15), except for a finite set of $\mathbf{h}$, for the test function (4.20). Condition (3.6) is immediate from (4.21) for bounded $\mathbf{h}$, because $w_{i,j}$ is bounded, and the theorem follows from Proposition 1. □

PROOF OF THEOREM 7. For definiteness, we shall prove the case $p = 1$, $q = 3$, $r = 2$, $s = 4$, whence, from (3.7) and (3.13),

$$w_{i,j}(\mathbf{h}) = \phi(g_{i+1,j} - g_{i,j}) + \phi(h_{i,j+1} - h_{i,j}); \tag{4.30}$$

the other cases are similar. Remembering that ϕ is skew symmetric about $c > 0$, let $f(u) = \phi(u) - c$, so $f(u)$ is bounded, odd, nondecreasing and ultimately nonconstant. We choose the test function (4.20), and see easily from (4.30), (4.21) and (4.22) that, in this case,

$$\begin{aligned} S_1(\mathbf{h}) = \sum_{i=1}^{M} \sum_{j=1}^{N} &\{f(g_{i+1,j} - g_{i,j}) + f(h_{i,j+1} - h_{i,j})\} \\ &\times (g_{i+1,j} - g_{i,j} + h_{i,j+1} - h_{i,j}). \end{aligned} \tag{4.31}$$

As usual, we must prove that $S_1(\mathbf{h})$ bounded implies $g_{i,j}, h_{i,j}$ bounded for all i, j. We shall need the following lemma.

LEMMA 2. *If f is as before and*

$$\{f(x) + f(y)\}(x + y) \le B \tag{4.32}$$

for some $B < \infty$, then:

(a) *If x and y have the same sign, then $|x|, |y| \le c(B) < \infty$.*

(b) *If x and y have opposite sign and one is bounded by D, say, then the other is bounded by a finite function of D and B.*

PROOF. To prove case (a), suppose $x, y > 0$. Then for any $A < \infty$, either $x, y \le A$ or, because f is nondecreasing,

$$\{f(A) + f(0)\}(x + y) \le \{f(x) + f(y)\}(x + y) \le B \quad \text{by (4.32)} \tag{4.33}$$

and $f(A) > 0$ for sufficiently large A because f is ultimately nonconstant. So

$$x, y \le \max_{A > 0} [\min\{A, B/f(A)\}] \equiv c(B) < \infty. \tag{4.34}$$

A similar argument using moduli works if $x, y < 0$.

To prove case (b), note that because f is ultimately nonconstant, then given any $\delta > 0$ sufficiently small and any $u > 0$, there is a $V(u, \delta)$ such that

$$f(u + V) - f(u) \geq \delta. \tag{4.35}$$

Consider $x > 0$, $y < 0$. If, say, $|y| < D$, then either $x < D + V(D, \delta)$ or, because f is odd and nondecreasing,

$$\begin{aligned} \{f(D + V) - f(D)\}\{x - |y|\} &\leq \{f(x) + f(y)\}(x + y) \\ &\leq B \quad \text{by } (4.32). \end{aligned} \tag{4.36}$$

So, by (4.35) and (4.36),

$$x \leq \min\{D + V(D, \delta), D + B/\delta\}. \tag{4.37}$$

If, instead, $x < D$, the same argument proves $|y|$ has the bound (4.37), and similarly if $x < 0$, $y > 0$. This completes the proof of Lemma 2. □

To return to the proof of Theorem 7, note from (4.31) and the properties of f that $S_1(\mathbf{h})$ is termwise nonnegative. So $S_1(\mathbf{h}) \leq B$ implies that, for $i = 1, \ldots, M$, $j = 1, \ldots, N$,

$$\begin{aligned} &\{f(g_{i+1,j} - g_{i,j}) + f(h_{i,j+1} - h_{i,j})\} \\ &\qquad \times (g_{i+1,j} - g_{i,j} + h_{i,j+1} - h_{i,j}) \leq B. \end{aligned} \tag{4.38}$$

We now proceed as in the proof of Theorem 6, assuming that site $i = 1$, $j = 1$ is a global minimum with $n_{11} = 0$. So (4.24) holds and $g_{21} - g_{11}$ and $h_{22} - h_{11}$ are both nonnegative. By (4.38) and Lemma 2(a) both differences are bounded, so we conclude as before that $g_{21}, h_{12}, g_{11}, h_{11}, n_{21}, n_{M1}, n_{1N}$ are all bounded. Now consider site $i = 2$, $j = 1$. Then:

> **either** $g_{31} - g_{21}$ and $h_{22} - h_{21}$ have the same sign, in which case they are both bounded, by (4.38) and Lemma 2(a);
>
> **or** the signs are different.

The same argument as before shows that the nonpositive difference is bounded, whence (4.38) and Lemma 2(b) say they are both bounded. Proceeding exactly as in the previous proof, we deduce that (4.28) and (4.29) hold here also, and so the theorem follows from Proposition 1, as before. □

5. Extensions. A Markov model that incorporates escape events can be specified by transition rates of the form (cf. [8])

$$q(\mathbf{h}, \mathbf{h}') = q^C(\mathbf{h}, \mathbf{h}') + q^E(-\mathbf{h}, -\mathbf{h}'), \tag{5.1}$$

where q^C and q^E are both transition rates of processes with capture events only. In [12] we used (5.1) to extend Theorem 5 to include escapes. Theorems 1 and 2 may be extended in similar fashion. To extend Theorem 3, suppose that (2.2) and (2.4) hold for both q^C and q^E in terms of functions $w^C(x, y)$ and $w^E(x, y)$. Then the process with rate q is ergodic if the functions w^C and

w^E both satisfy the condition of Theorem 3, where the inequalities need be strict for only one w. For example, w^E could satisfy the conditions of Theorem 3 and w^C could be a constant, implying that all captures occur at the same rate, in the manner of Gilmer and Jackson [15]. For further results see [8].

Because small values of **h** are not critically involved in long-term behavior, condition (ii) of Theorem 3 can be weakened considerably. It may, for example, be replaced by the condition: There is a $c > 0$ such that $w(x, y) \leq w(x, y')$ whenever any of the following apply:

1. $x < 0$ and $y \leq c < y'$.
2. $x < 0$ and $c < y < y'$.
3. $x > 0$ and $y < -c \leq y'$.
4. $x > 0$ and $y < y' < -c$.

Our proof essentially follows Theorem 3, but is substantially longer.

REFERENCES

[1] BENNETT, C. H. BÜTTICKER, M., LANDAUER, R. and THOMAS, H. (1981). Kinematics of the forced and overdamped sine-Gordon soliton gas. *J. Statist. Phys.* **24** 419–442.

[2] CHUNG, K. L. (1967). Markov chains with stationary transition probabilities, 2nd ed. Springer, Berlin.

[3] DURRETT, R. (1988). Crabgrass, measles and gypsy moths: An introduction in interacting particle systems. *Math. Intelligencer* **10** 37–47.

[4] EDEN, M. (1961). Two-dimensional growth processes. In *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **4**. Univ. California Press, Berkeley.

[5] EDWARDS, S. F. and WILKINSON, D. R. (1982). The surface statistics of a granular aggregate *Proc. Roy. Soc. London Ser. A* **381** 17–31.

[6] FOSTER, F. G. (1953). On the stochastic matrices associated with certain queueing processes. *Ann. Math. Statist.* **24** 355–360.

[7] FRANK, F. C. (1974). Nucleation controlled growth of a one-dimensional growth of finite length. *J. Crystal Growth* **22** 233–236.

[8] GATES, D. J. (1988). Growth and decrescence of two-dimensional crystals. *J. Statist. Phys.* **52** 245–257.

[9] GATES, D. J. (1990). Discrete and continuous models for surfaces of random aggregation and flow. Preprint.

[10] GATES, D. J. and WESTCOTT, M. (1988). Kinetics of polymer crystallization. I. Discrete and continuum models. *Proc. Roy. Soc. London Ser. A* **416** 443–461.

[11] GATES, D. J. and WESTCOTT, M. (1988). Kinetics of polymer crystallization. II. Growth regimes. *Proc. Roy. Soc. London Ser. A* **416** 463–476.

[12] GATES, D. J. and WESTCOTT, M. (1990). On the stability of crystal growth. *J. Statist. Phys.* **59** 73–101.

[13] GILMER, G. H. (1976). Growth on imperfect crystal surfaces. *J. Crystal Growth* **35** 15–28.

[14] GILMER, G. H. (1980). Transients in the rate of crystal growth. *J. Crystal Growth* **49** 465–474.

[15] GILMER, G. H. and JACKSON, K. A. (1976). Computer simulation of crystal growth. In *Crystal Growth and Materials* (E. Kaldis and H. J. Scheel, eds.). North-Holland, Amsterdam.

[16] HOFFMAN, J. D., DAVIS, G. T. and LAURITZEN, J. I., JR. (1976). The rate of crystallization of linear polymers with chain folding. In *Treatise on Solid-State Chemistry* (N. B. Hanay, ed.) **3**. *Crystalline and Non-Crystalline Solids*. Plenum, New York.

[17] JACKSON, K. A. (1968). On the theory of crystal growth: Growth of small crystals using periodic boundary conditions. *J. Crystal Growth* **34** 507–511.

[18] KELLY, F. P. (1979). *Reversiblity and Stochastic Networks*. Wiley, Chichester.

[19] KURZ, W. and FISHER, D. J. (1984). *Fundamentals of Solidification.* Trans Tech Publications, The Netherlands.

[20] LANGER, J. S. and MÜLLER-KRUMBHAAR, H. (1977). Stability effects in dendritic crystal growth. *J. Crystal Growth* **42** 11–14.

[21] MILLER, R. G., JR. (1963). Stationarity equation in continuous time Markov chains. *Trans. Amer. Math. Soc.* **109** 35–44.

[22] SADLER, D. M. (1987). On the growth of two-dimensional crystals. 2. Assessment of kinetic theories of crystallization of polymers. *Polymer* **28** 1440–1455.

[23] TWEEDIE, R. L. (1975). Sufficient conditions for regularity, recurrence and ergodicity of Markov processes. *Math. Proc. Cambridge Philos. Soc.* **78** 125–136.

[24] TWEEDIE, R. L. (1976). Criteria for classifying general Markov chains. *Adv. in Appl. Probab.* **8** 737–771.

[25] WEEKS, J. D. Gilmer, G. H. (1979). Dynamics of crystal growth. *Adv. in Chem. Phys.* **40** 157–228.

DIVISION OF MATHEMATICS AND STATISTICS
CSIRO
GPO BOX 1965
CANBERRA ACT 2601
AUSTRALIA