A TWO-DIMENSIONAL GROWTH PROCESS

MURRAY EDEN
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

1. Introduction

It is the purpose of this paper to examine certain of the properties of populations of cells, in particular, properties relating to the architecture of cell colonies. We imagine that the underlying process in the growth of an organism begins with a single cell (perhaps derived from the fusion of two germ cells), and then continues by a process in which the initial cell divides into daughter cells. These in turn divide into other cells, these divide further, and so on. The rates of division, the rates of growth of the individual cells, the ultimate size of each cell, its life span, the cellular form, and the procedures of differentiation are constrained by the "information" contained in the cells at any time t and by the environment of each cell, that is, the nutrient medium and the cells surrounding it.

That the process of morphogenesis is not entirely determined is fairly obvious. Evidence in favor of the conception that the development of form contains certain indeterminate features is found in studies of twins. Monozygotic twins are initially endowed with identical genetic "information," and if this information were to control the generating process completely, then the twins should be identical in every detail. It is true that by and large there is a greater correlation between morphological properties of monozygotic twins than those of fraternal twins. Still, so-called monozygotic offspring do not exhibit perfect correlations even in rather gross features such as in the dermal ridge count of fingerprints [1], the scute (scale) counts in nine-banded armadillo quadruplets [2], or the skin color patterns in cow twins [3]. There is less correlation in considering finer features such as patterns of retinal venation. There is little evidence available but it is to be anticipated that, on a cellular level, there would be little more correlation between twins than between any other members of a given species population. That is, it seems reasonable to assume that the "blueprint" of structure does not extend down to the position of every cell in the organism. Indeed, workers in the field of neural networks have made an operating principle of the assumption that the connections of individual neurons are essentially random.

This work was supported in part by the U.S. Army (Signal Corps), the U.S. Navy (Office of Naval Research), and the U.S. Air Force (Office of Scientific Research, Air Research and Development Command). The work involving TX-2 computer was supported in part by Lincoln Laboratories of M.I.T., which in turn is operated with support from the U.S. Army, Navy, and Air Force.

It must be pointed out, however, that in this last case there is some evidence that this is not the case, at least in the connections of the optic nerve [4].

The idea that every feature of an organism's structure is not determined in the initial code is a commonplace in embryology. An extensive literature has demonstrated that a specific environment at a specific stage in development can have a crucial effect on differentiation and further development. Thus, whether a certain bit of chick embryo will develop into a feather or into a skin cell will depend on the presence in the environment of a bit of mesoderm [5], or the presence of a bit of cartilage will cause an unorganized clump of cells to produce a kidney tubule [6]. In the light of these considerations it would seem instructive to investigate how probabilistic factors might be introduced into the process of morphogenesis.

The problem that we wish to consider is this. Given a single cell or a homogeneous population of cells, how can such a cluster develop into a structure of lower order of symmetry or no symmetry at all, that is, into a structure of recognizable and characteristic shape, without involving special structural properties of the cells themselves? So far as I know, the only attempt to treat this problem from a mathematical point of view was made by A. M. Turing in 1952 [7]. Turing postulated a one-dimensional ring of cells initially of identical length, each of which was supposed to contain essentially identical concentrations of chemical components and surrounded by a homogeneous medium. In treating this special case, Turing further postulated a specific set of chemical reactions and a set of reaction rate constants. In this way he was able to demonstrate that statistical fluctuations in the concentrations of certain of these hypothetical components would be sufficient to introduce a periodicity in the growth rate of the various cells around the ring so that one might think of the ring as growing into a three- or four-lobed structure. Although Turing made some attempts to extend the model to two-dimensional sheets, he was no longer able to produce an explicit analytical solution.

The model presented below represents a different approach to the problem and in fact considers an even simpler question, namely: starting from a single cell which may divide, its daughter cells divide again and again, what are the structural properties of the resulting colony of cells and how do various possible constraints effect the architecture? It has been necessary to make a large number of simplifying and special assumptions so that the resemblance between the model and the growth of any complicated metazoon or specialized organ or tissue is slight. We may view the successive division of the cells of a colony as a branching process and examine the sequence of configurations through which the colony passes in its growth. If this sequence of configurations can be described by a recursive relation, then a generating process can be formulated for the branching process, and the probability of certain subsets of the set of possible configurations may be computed.

We shall assume that each cell is identical with every other cell, that each cell is connected to at least one other cell, that the location of each cell is specified by a node in some regular lattice. For the purposes of simplicity the model will be restricted to two dimensions. Any migration of cells, differentiation into specialized cell types, variations in cell size, and similar properties of organisms will be neglected. Indeed, cells will be assumed immortal and a very special time-to-division distribution function is used, although these last two specializations could be modified without too much cost in complexity. We shall consider principally the properties of various finite subsets of points in the square lattice, although virtually all the results can be readily extended to hexagonal or triangular lattices and, for that matter, to higher dimensions. However, it has not been possible to generalize the results to sets of points in any less structured topology.

We shall distinguish certain nodes $\nu(x, y)$ by saying that they are *occupied* by a cell $\gamma(x, y)$. We define a nontransitive relation of adjacency by specifying that $\nu(x,y)$ is *adjacent* to $\nu(v, w)$ if and only if $x = v$ and $y = w \pm 1$, or $y = w$ and $x = v \pm 1$. We further specify that two cells are connected if they are adjacent to the same cell or if they can be related to the same cell by a chain of adjacencies. An arrangement of cells will be called a k-configuration c^k if and only if it contains exactly k connected cells. We shall define C^k as the set of k-configurations which are unique under translation, rotation, or reflection.

The enumeration of the configurations of C^k is a known problem and has been referred to by Golomb [8], but it has not yet been solved. An explicit listing of all configurations up to $k = 8$ has been presented in an earlier paper [9]. The methods employed are so laborious that it seems hardly worthwhile to attempt any further extension to higher values of k. However, before introducing any specific procedures for generating such configurations we shall consider certain bounds $C(k)$ on the number of elements in C^k.

J. M. Hammersley has shown that, for a general class of random walks on regular lattices, the number of distinct walks of k steps is approximately $A\gamma^k$ and he has shown how to estimate γ by Monte Carlo methods. Hammersley also has considered percolation processes, a problem similar in many respects to that treated in the present paper [10], [11].

2. A lower bound on elements of C^k

2.1. Distinguish a node in the square lattice as the beginning of a random walk on the lattice subject to the following constraints. If the coordinates of w_j, an element of the walk, are (x, y) then $w_{j+1} = (x + 1, y)$ or $(x, y + 1)$, that is, the random walk extends by going "north" or "east." Obviously such a random walk is nonintersecting and every such walk of length $k - 1$ edges corresponds to some $c \in C^k$. Further, the set of distinct walks $W(C^k) = 2^{k-1}$. However, rotation by 180° and reflection around the main diagonal will reduce the number of distinct configurations by a factor of 4 (except for those configurations which map into themselves under these operations. This is a set of order $2^{k/2}$ and negligible for large values of k; see appendix). Hence $W(C^k) = 2^{k-3}$. A fortiori, the number of configurations in C^k is $C(k) > 2^{k-3}$.

2.2. A somewhat better estimate for a lower bound on C^k can be obtained by generating a larger set of connected configurations from the set W^k. Consider any walk $w \in W^k$. Suppose this walk is bounded by a rectangle of width m and height n. We can describe w uniquely by an ordered n-tuple of integers $A(w) = (r_1, r_2, \cdots, r_n)$. The value of r_j is obtained by counting the number of occupied lattice points in the jth horizontal level (row) of the rectangle.

Consider the configurations generated by translating each node of the subwalk $(r_1, r_2, \cdots, r_i)$ to the east by t lattice points, at the same time retaining the subwalk $(r_{i+1}, \cdots, r_n)$ in its initial position. Such new configurations w_t^* will no longer be walks but they will be connected so long as $t < r_i + r_{i+1} - 2$. It is obvious that each of the t new configurations is unique. Since this translation procedure can be made independently between every pair of rows r_i and r_{i+1}, it follows that the number of unique configurations generated from a particular $w \in W^k$ is

$$w^*(k) = \prod_{i=1}^{n-1} (r_i + r_{i+1} - 1). \tag{1}$$

We can arrange the terms in the product in order of increasing value,

$$w^*(k) = \prod_{\nu=2}^{\max \nu} (\nu - 1)^{n(\nu)}, \tag{2}$$

in which $\max \nu = \max (r_i + r_{i+1})$ and $n(\nu)$ is the number of adjacent row pairs r_i, r_{i+1} for which $r_i + r_{i+1} = \nu$. Now

$$\begin{gathered} \log w^*(k) = \sum_{\nu=2}^{\max \nu} n(\nu) \log(\nu - 1), \\ \log E[w^*(k)] > E[\log w^*(k)] = \sum_{\nu=2}^{\max \nu} E[n(\nu)] \log(\nu - 1). \end{gathered} \tag{3}$$

LEMMA.

$$E[n_k(\nu)] = \frac{(\nu - 1)(k - \nu + 3)}{2^{\nu+1}}. \tag{4}$$

PROOF. By induction. (i) For $\nu = k$ we have $E[n_\nu(\nu)] = (\nu - 1)/2^{\nu-1}$. By our definitions, $n_\nu(\nu)$ stands for the number of walks of ν nodes with two rows. We have already shown that $W(k) = 2^{k-1}$. For $k = \nu + 1$,

$$E[n_{\nu+1}(\nu)] = \frac{2(\nu - 1)}{2^\nu} = \frac{\nu - 1}{2^{\nu-1}}, \tag{5}$$

that is, for three rows either $r_1 + r_2 = \nu$ or $r_2 + r_3 = \nu$ will be counted.

(ii) Consider the set W^{k-1} of 2^{k-2} walks. (a) Adjoin a node to the last row of each walk. The sum of $r_n + r_{n-1}$ is increased by the number of cases in which

$r_n + r_{n-1}$ had previously equaled $\nu - 1$ and decreased by the number of cases in which $r_n + r_{n-1}$ had equaled ν. (b) Start a new row r_{n+1} by adjoining the node above the right-most node of each walk of W^{k-1}. Then $n_k(\nu)$ is increased by the number of last rows whose length is $\nu - 1$. Further, the probability that the last row is of length t is 2^{-t}.

It follows that

$$E[n_k(\nu)] = E[n_{k-1}(\nu)] + \frac{(\nu - 1)2^{k-\nu-2}}{2^{k-1}} \tag{6}$$

$$= E[n_{k-1}(\nu)] + \frac{\nu - 1}{2^{\nu+1}}.$$

By the induction assumption,

$$E[n_{k-1}(\nu)] = \frac{(\nu - 1)(k - 1 - \nu + 3)}{2^{\nu+1}} \tag{7}$$

so that

$$E[n_k(\nu)] = \frac{(\nu - 1)(k - \nu + 2)}{2^{\nu+1}} + \frac{\nu - 1}{2^{\nu+1}} \tag{8}$$

$$= \frac{(\nu - 1)(k - \nu + 3)}{2^{\nu+1}},$$

which is equation (4).

Using this value of $E[n_k(\nu)]$ we have

$$E[\log w^*(k)] = \frac{1}{4} \sum_{\nu=2}^{\max \nu} \frac{k + 2 - (\nu - 1) \log(\nu - 1)}{2^{\nu+1}} \tag{9}$$

$$= \frac{1}{4} \sum_{\alpha=1}^{\max \alpha} \left[(k + 2)\alpha \log \frac{\alpha}{2^\alpha} - \alpha^2 \log \frac{\alpha}{2^\alpha} \right].$$

Since these series converge rapidly, we can neglect the second sum for sufficiently large values of k to obtain

$$\log E[w^*(k)] > E[\log w^*(k)] = \frac{k + 2}{4} (0.775\cdots) \tag{10}$$

$$E[w^*(k)] > (1.57)^{k+2}.$$

Finally, $W^*(k) = W(k)\, E[w^*(k)]$. Thus,

$$W^*(k) > 2^{k-1}(1.57)^{k+2} > (3.14)^{k-1}. \tag{11}$$

It may be remarked that so long as a single row of any W^* is connected, the cells of any other row need not be connected to each other but only to cells in the rows either above or below. As a consequence the number of possible connections of r_i and r_{i+1}, where $r_i \geqq r_{i+1}$, is bounded from below by $\max \left[r_{\cdot i} + r_{\cdot i+1} - 1, \binom{r_i}{r_{i+1}} \right]$. While the use of this condition would certainly raise the lower bound on $C(k)$, the fact that the shifts of adjacent rows are no longer independent of the shifts of their other nearest neighboring rows make the combinatorial formulas intractable.

3. An upper bound on the elements in C^k

The upper bound on $C(k)$ will be estimated using a different combinatorial technique. We shall first devise a procedure for specifying each of the elements of C^k by a *word* $F(c)$ with a finite number of digits and then estimate the size of the set of words representing configurations. We will do this by constructing an algorithm for ordering the cells γ_j of any configuration c_i and then describe a function $F(\gamma_{i1}, \gamma_{i2}, \cdots, \gamma_{ik})$ whose values map uniquely onto C^k.

Choose some $c_i \in C^k$, some $\gamma \in c_i$, and one of the four possible orientations of the square lattice. Distinguish this cell as γ_1. Examine the nodes adjacent to γ_1 beginning with the node "south" of γ_1, and proceeding clockwise. Number any cells occupying these four nodes in the order in which the nodes are examined. Now assign the first four digits of the word $F(c)$ a one or a zero according as whether or not a cell has been found at each of the nodes. Next distinguish γ_2, and examine the nodes adjacent to it beginning with γ_1, proceeding clockwise as before, numbering any new cells to be found at the adjacent nodes and assigning the appropriate binary digit to the next three places of $F(c)$. In this case and in the examination of the neighbors of any succeeding cell in the configuration, only 3 binary digits are required, since to each γ_j there is some γ_l, with $l < h$, such that the designation of γ_l contains a 1 distinguishing its place in the order of $F(c)$, and the orientation of γ_l relative to γ_j has already been established. Such a γ_j will be called the *designator* of γ_l and denoted $d(\gamma_l)$. Thus in the assignment of numbers to the designation of some γ_j, a 1 is entered only for those adjacent nodes occupied by cells which have as yet no designator.

It is obvious that this procedure can be carried out until all of the cells of c have been ordered and a designation given to each. It will be noted that $F(c)$ contains $3k - 2$ binary digits and hence there are 2^{3k-2} possible values of $F(c)$. Since the choice of the initial cell and its orientation in the lattice are arbitrary, each configuration will be specified by $4k$ words (again except for the small set of configurations which can be reflected or rotated into itself). Finally, it is clear from the mode of construction that for c_j distinct from C_m all $F(c_j)$ will differ from the $F(c_m)$. Thus an upper bound on $K(C)$ is $2^{3k}/16k$.

However, if the values of the digits in any $F(c)$ are summed they must equal $k - 1$, since each cell has one designator and the first cell has none. Therefore, we obtain immediately as a better upper bound on $C(k)$,

$$C(k) < \frac{1}{4k} \binom{3k - 2}{k - 1}. \tag{12}$$

By Stirling's approximation,

$$C(k) < \frac{k^{-3/2}}{\pi\sqrt{3\pi}} \left[\frac{27}{4} + \frac{1}{9k^2}\right]^k. \tag{13}$$

A further improvement can be obtained by considering the constraints induced by the configurations on the permissible values of $F(c)$. There is one obvious constraint namely, that $\sum_{j=1}^{\alpha} f_j \geqq \alpha$ for every α as well as $\sum_{j=1}^{k-1} f_j = k - 1$. Any sequence for which the number of cells designated is less than the number of triplets of digits is complete; that is, there are no more cells to order. Such a word will represent a configuration containing fewer than k cells.

We shall invert the process described above and construct sequences of binary digits of length $3k + 1$ and estimate how many of them will specify permissible configurations. We shall call a sequence of length $F\lambda < 3\lambda - 1$ a *partial word* for any $\lambda < k$.

It follows from the algorithm for proceeding from a configuration to the word $F(c)$ that any of the possible sixteen values of the quadruple f_1 of γ_1 are possible and in like manner any of the eight values for the triple f_2 of values specified by γ_2. However, the designations of γ_1 and γ_2 already induce restrictions on the designation of γ_3 and in general the designation of γ_j will depend upon the particular values of other cells designated previous to γ_j, but *not* upon $d(\gamma_j)$. We proceed by cases to consider the constraints upon f_j induced by the designation of $d[d(\gamma_j)]$ and induced by cells other than γ_j that are designated by $d(\gamma_j)$. Of course, other constraints may have been induced by earlier and more distant portions of the configuration. The cases are displayed in figure 1. The nodes of each directed tree distinguish the cells in some configuration. The cell at the root of the tree is the designator of the next cell in the tree and that cell in turn designates the cell or cells corresponding to the twigs of each tree. The boxes marked with an "x" represent nodes which have been designated previously and hence represent constraints on the triples derivable from the neighbors of the second node of the tree. The black boxes represent restrictions on the designations of the cells of the twigs of each tree, each such constraint deriving either from the designation of $d[d(\gamma_j)]$ or from one of the collateral twigs of γ_j.

The first row of diagrams in the chart represents all the possible designations of the middle cell $d(\gamma_j)$ when it is not restricted. It will be noted that when the only new cell γ_j is directly above $d(\gamma_j)$, then there will be no restrictions on the designation of γ_j induced by $d[d(\gamma_j)]$. However, if the single new cell is either "east" or "west" of its designator, then $d[d(\gamma_j)]$ or some earlier cell constrains the designation of γ_j to at most four choices. If $d(\gamma_j)$ designates two additional cells (as shown in the fourth diagram of the first row), then the first twig cell γ_j is unrestricted; however, it induces a restriction on γ_{j+1} (the second twig cell), while $d[d(\gamma)]$ induces a second restriction on γ_{j+1}. Each of the diagrams may be interpreted in a similar fashion. For this set of cases, the expectation of the number of new cells designated equals the expectation of the number of restrictions on further $f_n \in F(c)$.

In the other rows of the chart it is assumed that there are one or more restrictions on $d(\gamma_j)$ induced by some node $\nu_l \neq [d(d\gamma_j)]$, but the designation of ν_l preceded the designation of $d(\gamma_j)$ in $F(c)$. Two subsets of cases are considered, either ν_l is occupied by a cell or it is not. The first of these possibilities is illustrated in the upper diagrams of each row. By the nature of the algorithm, these cells will have been designated before any twig cell. The cases in which a node is restricted but remains unoccupied by a cell, are illustrated in the lower half of each row.

It will be seen that for each row, at least as many restrictions are induced as there are cells added.

It follows that there are 2^7 partial words F^2, less than $2^7 \times 2^{7/3}$ words F^3. However, by reason of the expected number of restrictions on each new triple

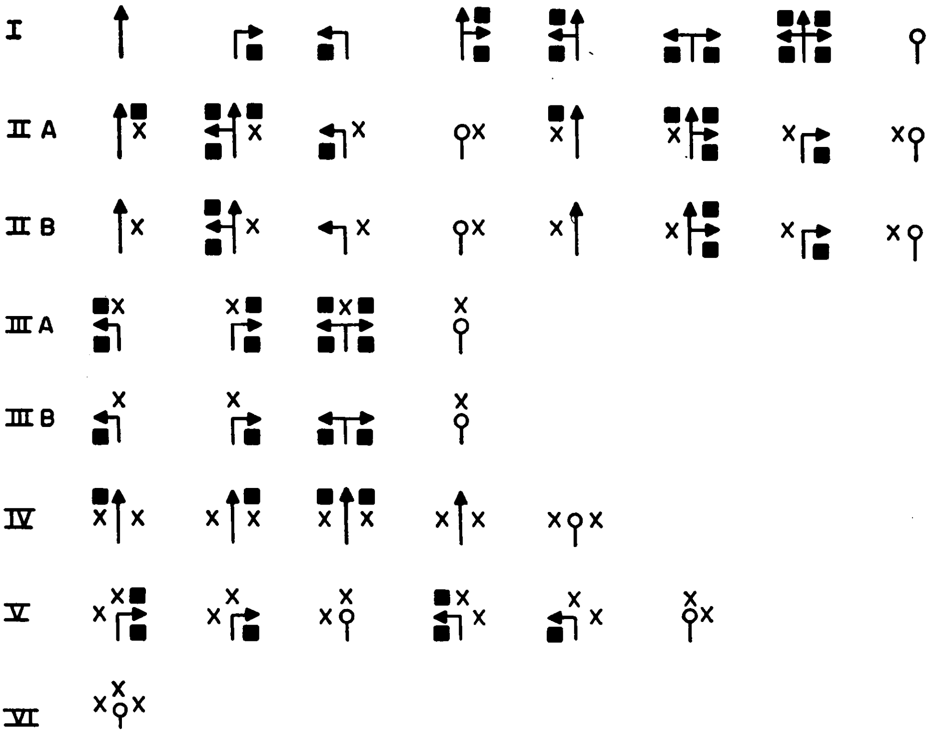

FIGURE 1

A graphical representation of the possible cases
for adjoining new cells to a configuration.
I No restrictions on $d(\gamma)$;
expectation for the number of cells adjoined, $E(\sigma)$, $= 1.5$;
expectation for the number of constraints on these new cells, $E(\rho)$, $= 1.5$.
II One restriction; $E(\sigma) = 1$; $E(\rho) = 1$.
III One restriction; $E(\sigma) = 1$; $E(\rho) = 1.5$.
IV Two restrictions; $E(\sigma) = 0.5$; $E(\rho) = 0.5$.
V Two restrictions; $E(\sigma) = 0.5$; $E(\rho) = 0.75$.
VI Three restrictions; $E(\sigma) = 0$; $E(\rho) = 0$.

of the sequence of partial words, $F^\lambda < 2^{2\lambda+1}$. Now the set $W(C^k)$ is a proper subset of the set of partial words, so that a fortiori, $C(k) < (1/4k)2^{2k+1} = (1/k)2^{2k-1}$.

4. A "symmetrical" growth process

The first model of "growth" which we will consider is perhaps the simplest. Distinguish a single node of the square lattice and assume that it is occupied by a cell γ_1. Assign equal probability to the configurations obtained by adjoining one other cell to the nodes adjacent to γ_1 (there are four of these that are obviously equivalent under rotation). This two-celled configuration has six adjacent nodes. Again assign equal probability to the six 3-configurations obtained by adjoining a single cell. This procedure can be carried out indefinitely, each time adjoining a single cell.

If there is any reasonably close biological counterpart to this process it may be in the growth of bacterial cells or tissue cultures of cells that are constrained from moving. The process described may be likened to the growth of a colony of cells in which growth can proceed only where there is nutrient medium, that is, at the periphery. It may also be commented that *Ulva lactuca,* the common sea lettuce, grows as a sheet two cells thick and apparently grows only at its periphery. Finally, while there is no information concerning its mode of growth, the gamete-forming thalli of *Prasiola stipitata* exhibit a cellular pattern that is almost identical to the patterns developed by the model described above, for small values of k [12].

The probabilities for all k-configurations up to k equal to eight have been tabulated in an earlier paper [9]. In addition, certain estimates have been made concerning the probabilities of certain subsets of C^k up to C^{16}.

THEOREM. *The probabilities of k-configurations isomorphic to the same graph are equal.*

We first define the graph of a configuration by identifying each node occupied by a cell of the configuration by a node of the graph. Any two occupied nodes that are adjacent are connected by an edge of the graph.

We consider configurations having a given cell picked out as the starting cell; the graph of such a configuration has a distinguished node, which we shall call the starting node. We show below that the probability of obtaining such a configuration depends only on its graph. By taking each cell of a configuration in turn as the starting cell (clearly equivalent to taking each node of its graph in turn as starting node) and summing, we get that the total probability of a configuration depends only on its graph.

The probability of a configuration is proportional to "the number of ways that it can be generated." For a configuration c, we shall write $N(c)$ for this number. Examination of the generating process shows that $N(c)$ has the following inductive definition.

(a) If c is the one-celled configuration, $N(c) = 1$.

(b) If c is a k-celled configuration, $k > 1$, then $N(c) = \sum_j M(\gamma_j, c)N(c - \gamma_j)$, where the sum is over all cells $\gamma_j \in c$, such that $c - \gamma_j$ is a configuration and where $M(\gamma_j, c)$ is 1, 2, 3, or 4 according to the number of edges of γ_j which adjoin $c - \gamma_j$. (By the phrase "$c - \gamma_j$ is a configuration" is meant that $c - \gamma_j$ is connected and contains the starting cell.)

Now to each cell γ_j of c there corresponds a node ν_j of G, the graph of c. Thus, $c - \gamma_j$ is a configuration if and only if $G - \nu_j$ is the graph of a configuration, and then the graph of $c - \gamma_j$ is isomorphic with $G - \nu_j$. Also, $M(\gamma_j, c)$ is equal to the number of lines of G joining ν_j with $G - \nu_j$. Thus under the inductive hypothesis that N depends only on the graph for $(k - 1)$-celled configurations, the right side of the formula in (b) is completely determined for any k-celled c by the graph of c. It is trivial to start the induction with $k = 1$.

We may obtain an interesting corollary of this theorem by defining an extension G' of the graph G. This G' is identical with G except that an edge exists between two nodes of G' (permissible under the constraints of the lattice topology) that is not present in G. We also define G^*, the *completion* of G. Namely, G^* is the completion of G if and only if $G^{*\prime} = G^* = [\cdots(G')'\cdots]'$.

It follows immediately from the definition that every G has a completion G^*.

COROLLARY. *To every G there corresponds a G^* such that $p(c^*) \geqq p(c)$.*

Since the probability of a configuration is proportional to "the number of ways it can be generated," and since the number of ways depends on the number of edges which adjoin $c - \gamma_j$, it follows that $p(c') > p(c)$ for every c and c'. Since G^* terminates a chain of G', in which $p[c^{(j')}] > p[c^{(j-1')},]$ the corollary follows.

The subset $\{G^*\}$ is very much smaller than the set $\{G\}$ and since there are many configurations corresponding in general to an arbitrary G, while there is only a single c^* corresponding to any G^*, it is clear that $\{c^*\}$ is a very small subset of C^k. However, it has not been possible to enumerate this set or determine the probabilities associated with each G^*. It will also be noted that G^* is not unique and it has not yet been found possible to order the G^* derivable from an arbitrary G.

5. Remarks on some random variables defined on the set of *k*-configurations

The theorem given above and an examination of the probabilities computed for configurations up to $k = 16$ suggest certain properties as being closely related to the probabilities. One of these properties has been utilized in the proof of the theorem, namely the number of edges of the graph proceeding from each node. If we examine the random variable $M(G)$, where $M(G) = \sum_j M(\gamma_j, c)$, it is true that by and large the configurations with the higher value of this random variable are the more probable. However, it is easy to exhibit counterexamples to the conjecture that $p(c)$ is a monotone function of $M(G)$.

Another random variable is the *perimeter* of c. We define the number of unoccupied nodes adjacent to cell C_j as a_j. Let $\pi(c)$ denote the perimeter of c; then $\pi(c) = \sum_j a_j$. The maximal perimeter of c in C^k is $2k + 2$. Referring again to the graph of c, and defining $L(G)$ as the number of minimal loops in G, then it is easy to verify that $\pi(c) = 2[k + 1 - L(G)]$. For each extension G' of G we have $\pi(c') = \pi(c) - 2$. Clearly, $\pi(c^*)$ is minimal. However, here too one can exhibit counterexamples to the conjecture that $\pi(c_\alpha) > \pi(c_\beta)$ implies $p(c_\beta) > p(c_\alpha)$.

Nevertheless, in order to obtain some insight into the properties of the growth process defined above, the configurations for C^6, C^7, C^8 were ranked in order of decreasing probability and the cumulative distribution function was plotted. In addition, estimates were made of the size of the subsets of C^{10} and C^{16} for each value of (c), as well as estimates of $E[p(c)]$; the value of $\pi(c) = n$ for each of these subsets. These subsets were ordered in decreasing values of $\pi(c)$ and again the cumulative distribution function was plotted. Finally, the measure for each of these functions was normalized to the interval $[0, 1]$. The distribution functions obtained in this way are shown in figure 2. This figure suggests the conjecture that for any c^k generated by the growth process and any ϵ and δ there is some k such that

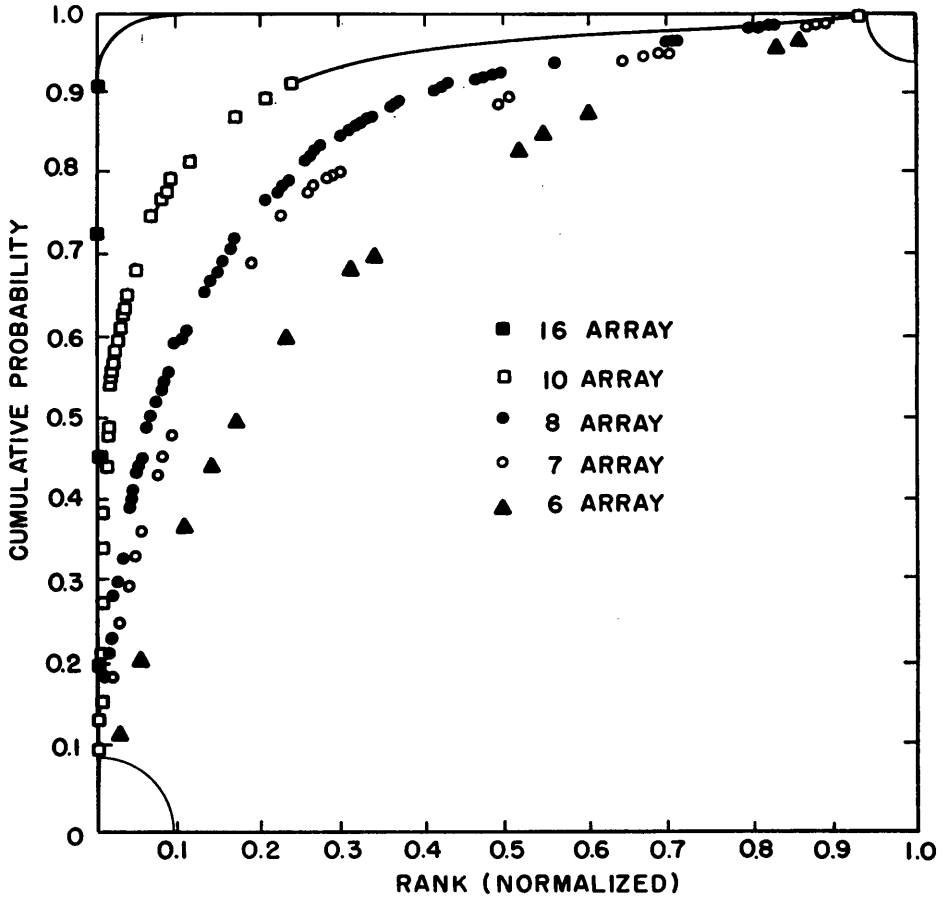

FIGURE 2

Distribution function for configurations with $k = 6, 7, 8, 10, 16$.

(14)

$$P\left\{c_j^k \in C_n^k \middle| c \in C_n^k \quad \text{if and only if} \quad \pi(c) \leqq n \quad \text{and} \quad \frac{N(C_n^k)}{C(k)} < \delta\right\} > 1 - \epsilon,$$

in which $N(C_n^k)$ is the size of the subset C_n^k.

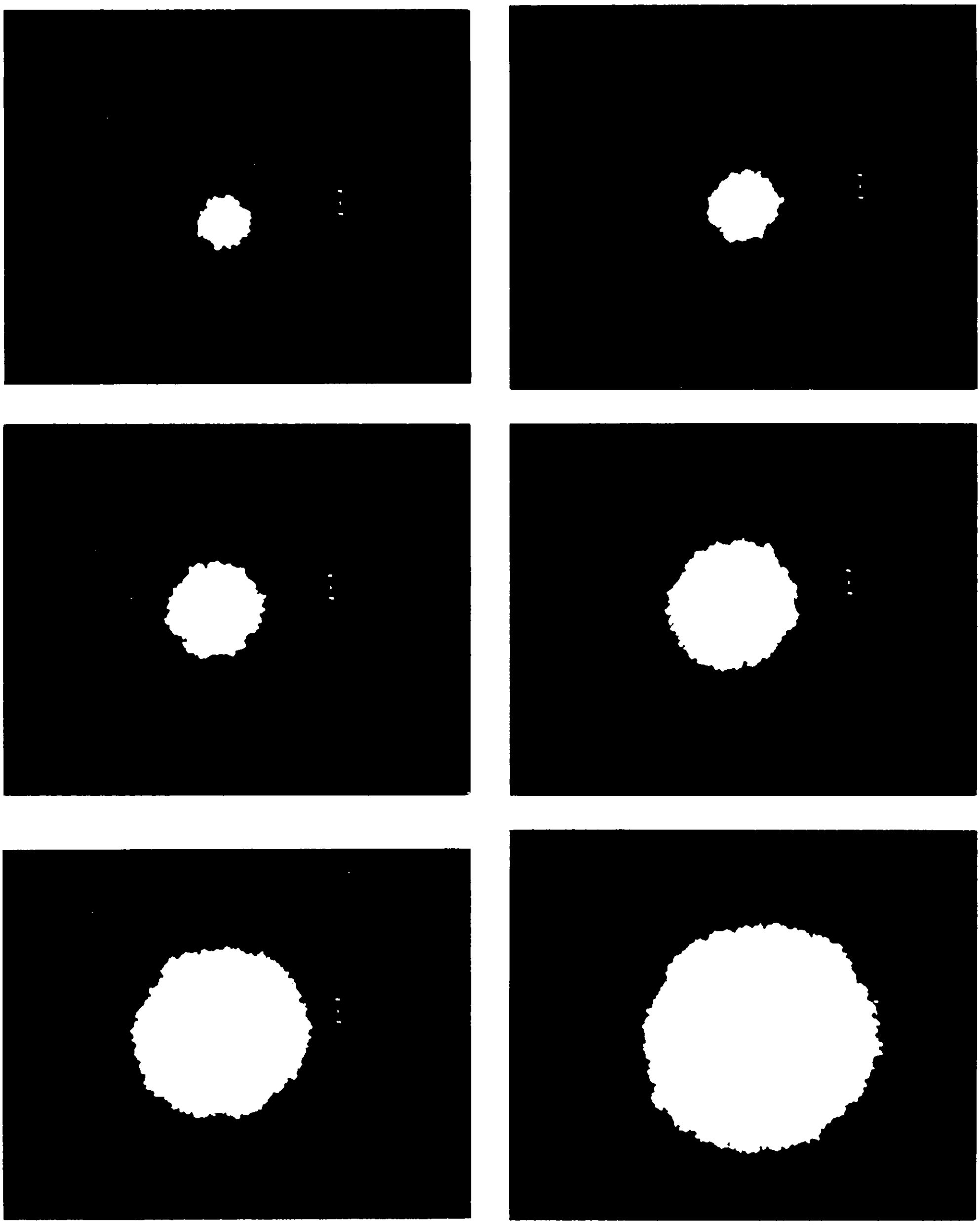

FIGURE 3

Six stages in the growth of a configuration with the condition $p(h) = p(v)$.

6. Monte Carlo procedures applied to the growth process

Since the analytical results obtained for the growth model are so meager a Monte Carlo procedure was devised. The TX-2 computer of Lincoln Laboratories was programmed to compute configurations following the generating rules given above. An example of the growth of a typical configuration is shown in figure 3. In general, the configurations were permitted to grow to a size of 2^{15} cells. Although it is planned to compute moments and other properties of the configuration during the process of growth on the computer, at present the only parameter (aside from the configuration itself) retained in the computer memory was the count of cells and the values of $\pi(k)$ for a number of selected values of k. As yet the samples of configurations computed in this way appear to be too few to justify anything more than a few qualitative statements.

It is to be seen that the colony is essentially circular in outline. Needless to say, there are a number of properties of each growing colony as well as properties of the ensemble that may be worth examining, for example, moments, the eccentricity of the configurations, the "roughness" of the edge. The programming needed in order to investigate such properties has not yet been completed. However it was found that the appropriate instructions controlling the probability of adjoining cells could be altered by a minor change in the computer program. Accordingly configurations were generated in which the probability $p(h)$ of adjoining a cell at a vertical edge was an integral multiple of the probability $p(v)$ of adjoining to a horizontal edge. A series of steps in the development of a figure in which $p(h) = 10\,p(v)$ is displayed in figure 4.

One immediate observation is that the "axial ratio" of the figure generated in this way is a good deal less than the ratio of $p(h)$ to $p(v)$, table I. So far it has not been possible to obtain by analytical means the expectation of the ratio of length of sides of a circumscribing rectangle or its asymptotic value. When a sufficiently large sample of such configurations have been generated by the Monte Carlo procedure, it should be possible to make an estimate of this and similar properties.

It is obvious from the configurations generated that the values of $\pi(c)$ are very much lower than the maximal $\pi(c)$. A log-log plot of $\pi(c)$ as a function of k for the growth of these configurations is illustrated in figure 5. It will be noted that, aside from considerable variability in this property, the curve seems to converge rapidly to the line representing $\pi(c^k) = 8\sqrt{k}$. Even if the value of the coefficient is in doubt it seems reasonable to conjecture that $\pi(c^k) = A\sqrt{k}$. It is clear that the convex hull of these configurations is close to minimal or, roughly speaking, that the "density" is maximal. A few special cases may be of interest. Thus the figure of minimal convex hull is a square, call it c_s^k. In this case $\pi(c_s^k) = 4\sqrt{k}$. If the square is rotated through 45°, call this figure a diamond c_d^k, then $\pi(c_d^k) = 4\sqrt{2k}$. Finally, for a circle c_o^k we have $\pi(c_o^k) = 8\sqrt{k}/\sqrt{\pi}$.

We may estimate the size of the subset of C^k with $\pi(c) = 8\sqrt{k}$ by using the method described for the first upper bound on the size of C^k. We modify the

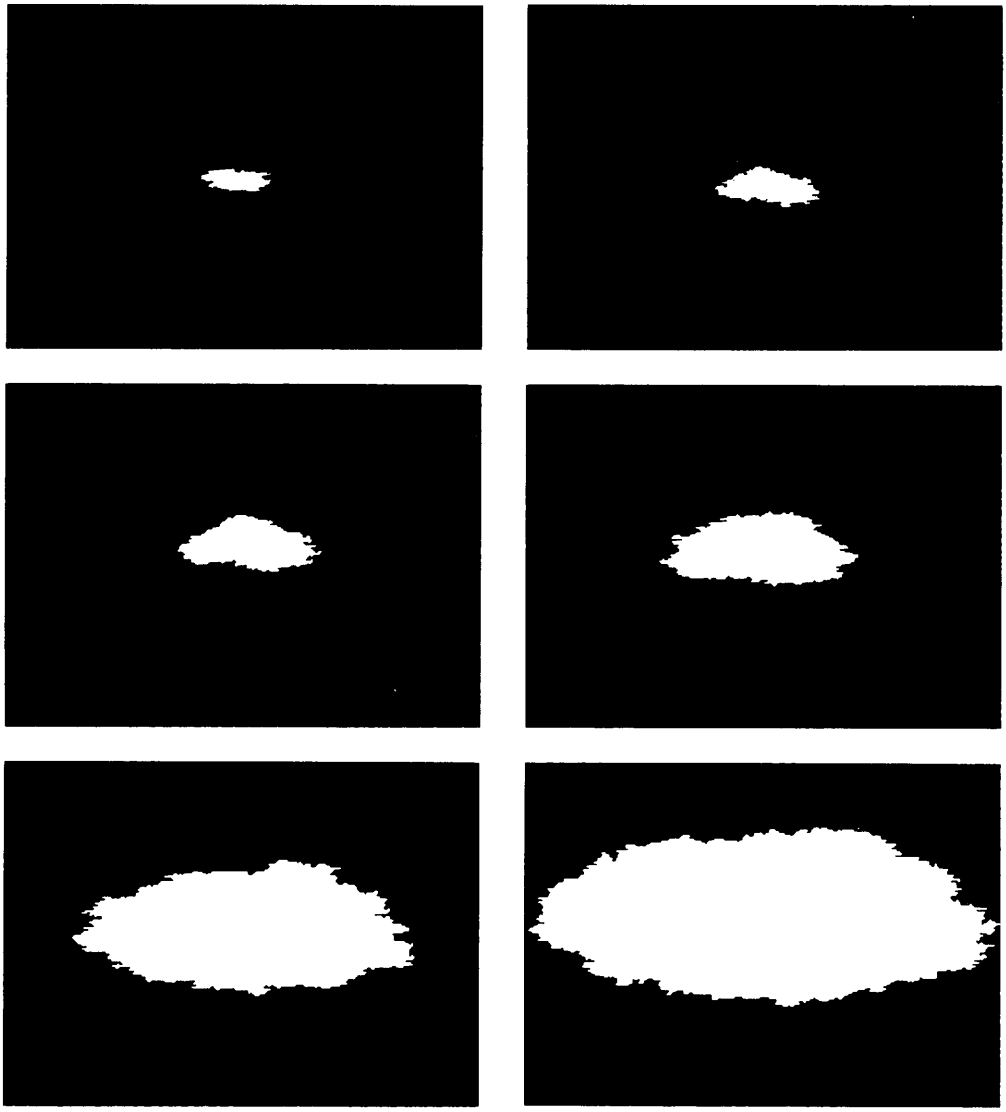

FIGURE 4

Six steps in the growth of a configuration with the condition $p(h) = 10\,p(v)$.

procedure for enumerating the members of C^k by assigning a 1 to the appropriate place in the word $F(c)$ without regard to whether the cell has been designated or not. Thus there may be as many as four 1's in $F(c)$ corresponding to a particular cell of the configuration. Applying this procedure to the configuration with $\pi(c) = 8\sqrt{k}$, at least $k - 8\sqrt{k}$ of the triples in $F(c)$ will be identical, that is (111). Assuming no restrictions on the remaining assignments and assuming

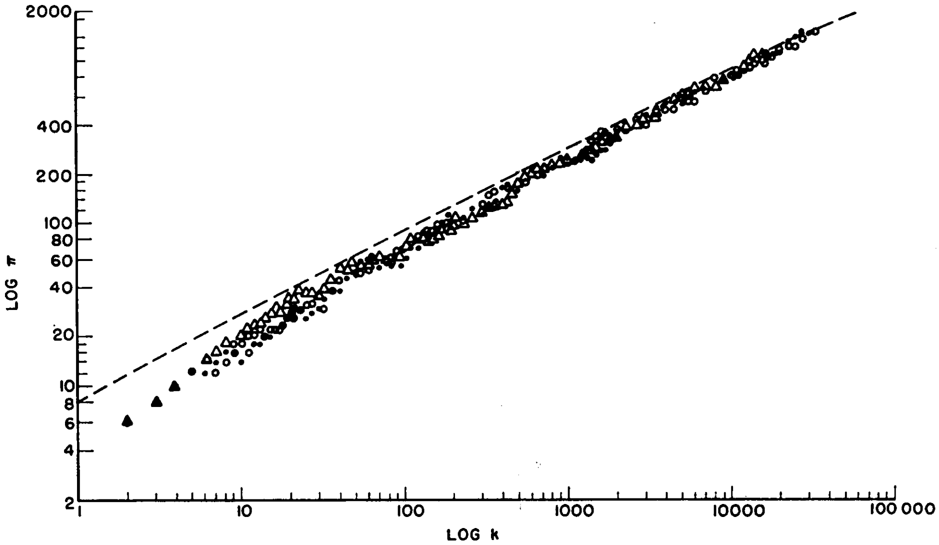

FIGURE 5

Log $\pi(C)$ as a function of log k for three configurations:
dots: $p(h) = p(v)$, open circles: $p(h) = 2\,p(v)$, triangles: $p(h) = 10\,p(v)$.

that any ordering of the $k - 1$ triples is permissible, the number of possible words is

$$8^{8\sqrt{k}}\binom{k}{8\sqrt{k}} \quad \text{or} \quad N[F(c)] < (8 + k)^{8\sqrt{k}}. \tag{15}$$

TABLE I

A MONTE CARLO-PRODUCED CONFIGURATION WITH $p(h) = 10p(v)$

k	Size of Circumscribed Rectangle	"Axial Ratio"
36	1.6 × 3.3	2.0
64	1.7 × 5.4	3
104	1.8 × 5.7	3.2
161	2.5 × 7.5	3
273	3 × 9	3
383	4.6 × 11	2.4
534	5.2 × 12	2.3
740	6 × 14	2.3
957	7 × 17.2	2.5
1518	8 × 20	2.5
2114	8.8 × 24	2.7
3230	11 × 30.5	2.8
5001	14 × 37.6	2.7
6938	16.6 × 41.8	2.5
9770	19.5 × 49.2	2.5
15097	22 × 57	2.6

Recalling the lower bound on $C(k)$ given earlier, we have

$$\frac{N[F(c)]}{C(k)} < \frac{(8+k)^{8\sqrt{k}}}{3.138^{k}} = \left[\frac{(8+k)^{8}}{3.138^{\sqrt{k}}}\right]^{\sqrt{k}}. \tag{16}$$

It is obvious that this ratio converges rapidly to 0 for sufficiently large k.

This result supports the conjecture stated above that when k is sufficiently large then the fraction of all possible configurations that are likely to be generated is vanishingly small. If we consider the biological analogue to be a colony of cells growing wherever there is a nutrient medium we can be quite sure that the colony will have the largely circular morphology exhibited by the model. If the clone of cells has a more complicated morphology, for example, an elliptical shape, it would suggest that there was an additional constraint imposed on the generating process, perhaps one resembling the unequal probabilities which led to the colony of figure 4. Paul Weiss has shown that an anisotropy in the nutrient medium of a tissue culture will produce growth that is reminiscent of this figure [6].

Whether more complicated forms of organization can be described by growth models such as the one presented in this paper remains to be seen. However, the combinatorial difficulties are so large that a direct attack seems out of the question with the tools presently available. On the other hand, the Monte Carlo simulation procedures appear to be a fruitful means of exploring growth processes.

I am indebted to Professor Hale Trotter of Queens University, Kingston, Ontario, for several illuminating discussions as well as for the form of the theorem on equality of probabilities given above. I am also indebted to Mr. Alexander Vanderburgh of Lincoln Laboratories, Lexington, Massachusetts, for assistance in programming and operating the TX-2 computer.

◇ ◇ ◇ ◇ ◇

APPENDIX. INTERNAL SYMMETRIES OF THE k-CONFIGURATIONS

The theorem relating graph isomorphism and probability, that is, the estimates of $C(k)$ and of the probabilities of k-configurations, requires certain correction terms because the effects of k-configurations with internal symmetry have not been taken into account. This source of error has been neglected since the proportion of configurations with internal symmetry is very small for large k.

We sketch below a method for determining an upper bound for the number of symmetrical k-configurations and show that this bound is of the order of $C(k/2)$.

Consider the k-array, k even (a similar proof can be made for k odd). Any k-configuration exhibiting rotational symmetry may be decomposed into two identical $k/2$ configurations adjoined at one or more edges. Since the maximal number of open edges in a $k/2$ configuration is $k + 1$ the number of such symmetrical k-configuration is $S_{\text{rot}}(k) < (k + 1)C(k/2)$. Any k-configuration containing a plane of reflection will have pairs of isomorphic $k/2$, $k/2 - 2$, $k/2 - 4$, $\cdots$ configurations, with the respective planes of symmetry passing through 0, 2, 4, $\cdots$ cells. Taken together,

$$S(k) < 4\left(\frac{k}{2}+1\right)C\left(\frac{k}{2}\right)+4\left(\frac{k}{2}-1\right)C\left(\frac{k}{2}-2\right) + 6\left(\frac{k}{2}-3\right)C\left(\frac{k}{2}-4\right)+\cdots . \tag{17}$$

If, as we have discussed in the text, $C(k) \fallingdotseq A\mu^k$, where A and μ are constants, $3 < \mu < 4; 0 < A < 1$. Then

$$S(k) < 4\left(\frac{k}{2}+1\right)\mu^{k/2}+4\left(\frac{k}{2}-1\right)\mu^{k/2-2}+\cdots < k^2\mu^{k/2}. \tag{18}$$

Since $\mu^{k/2}$ will dominate, $S(k)$ will be of the order of $(\mu^k)^{1/2}$.

REFERENCES

[1] S. B. HOLT, "Genetics of dermal ridges; sib pair correlations for total finger-ridge count," *Ann. Human Genetics*, Vol. 21 (1957), pp. 352–362.

[2] H. H. NEWMAN, *The Biology of Twins (Mammals)*, Chicago, University of Chicago Press, 1917.

[3] J. HANCOCK, *Studies in Monozygotic Cattle Twins*, Dept. of Agriculture, Animal Res. Div. Publ. No. 63, Hamilton, N.Z.

[4] J. Y. LETTVIN, H. R. MATURANA, W. S. MCCULLOUGH, and W. H. PITTS, "What the frog's eye tells the frog's brain," *Proc. I.R.E.*, Vol. 47 (1959), pp. 1940–1951.

[5] J. NEEDHAM, *Biochemistry and Morphogenesis*, Cambridge, Cambridge University Press, 1950.

[6] B. H. WILLIER, P. A. WEISS, and V. HAMBURGER, *Analysis of Development*, Philadelphia, Saunders, 1955.

[7] A. M. TURING, "The chemical basis of morphogenesis," *Philos. Trans. Roy. Soc. London, Ser. B*, Vol. 237 (1952), pp. 37–72.

[8] S. GOLOMB, "Checkerboards and polyominoes," *Amer. Math. Monthly*, Vol. 61 (1954), pp. 675–682.

[9] M. EDEN, "A probabilistic model for morphogenesis," *Symposium on Information Theory in Biology*, New York, Pergamon Press, 1958.

[10] J. M. HAMMERSLEY and K. W. MORTON, "Poor man's Monte Carlo," *J. Roy. Statist. Soc., Ser. B*, Vol. 16 (1954), pp. 23–38.

[11] J. M. HAMMERSLEY, "Bornes supérieures de la probabilité critique dans un processus de filtration," *Le Calcul des Probabilités et ses Applications*, Paris, Centre National de la Recherche Scientifique, 1959, pp. 17–37.

[12] K. M. DREW and I. FRIEDMANN, "Occurrence of motile gametes in *Prasiola stipitata* Suhr," *Nature*, Vol. 180 (1957), pp. 557–558.