

Stanislaw M. Ulam's Contributions to Theoretical Theory

WILLIAM A. BEYER

Los Alamos National Laboratory, Los Alamos, NM 87545, U.S.A.

PETER H. SELLERS

Rockefeller University, New York, New York, U.S.A.

and

MICHAEL S. WATERMAN

Departments of Mathematics and Biological Sciences, University of Southern California, Los Angeles, CA 90089-1113, U.S.A.

(Received: 9 April 1985)

Abstract. S. M. Ulam's contributions to biology are surveyed. The survey covers cellular automata theory, population biology, Fermi–Pasta–Ulam results, pattern recognition, and sequence similarity.

1. Introduction

In Poland before the war, there was a tradition that the scholar should be familiar with other scientific disciplines in addition to his own. In this tradition, Stanislaw M. Ulam was educated in prewar Poland as a mathematician and throughout his life he enjoyed finding problems in physics, astronomy, and biology as well as in mathematics.

In 1934 Ulam left Poland for a few months in Cambridge. On his return to Poland in early 1935, he received an invitation to visit the United States. He spent 1935–1936 at Princeton, 1936–1940 at Harvard, and 1941–1943 at the University of Wisconsin. His work relating to biology was begun at Los Alamos where he moved to next.

His association with the Los Alamos Scientific Laboratory began in early 1944. Chain reactions were an important process at Los Alamos, and these reactions do have a certain vague analogy with the multiplicative processes in biology. In addition, computers were being built and used at Los Alamos. While the analogy between the computer and the brain seems obvious to us today, that was not so in the early days of computing. Ulam, Alan Turing, and John von Neumann shared an interest in this analogy. Also, the birth of the new biology with the Crick and Watson discovery of the structure of DNA in 1953 stimulated more applications of mathematics to biology and Ulam participated in these applications.

Ulam's contributions to theoretical biology are discussed under the sections: (2) cellular automata theory, (3) population biology, (4) Fermi–Pasta–Ulam results and solitons, (5) pattern recognition, and (6) biometric spaces.

2. Cellular Automata Theory

In the late 1940s von Neumann began the development of a general theory of automata. Von Neumann often discussed this work with Ulam who made suggestions and contributions of his own. This theory included self-reproducing automata. Von Neumann's work was first presented in his University of Illinois in December 1949. The work is thoroughly reviewed in his posthumous book [1] edited by Arthur W. Burks. There were five models of self-reproducing automata: the kinematic model, the cellular model, the excitation-threshold-fatigue model, the continuous model, and the probabilistic model. The kinematic model might be described today as a robot without a power source.

The cellular model was suggested in conversation with Ulam. Ulam suggested that this model would be more amenable to logical and mathematical treatment than would the kinematic model. In 1950 [2], Ulam proposed considering an infinite graph of points, each point with a finite number of connections to certain of its 'neighbors'. Each point can occupy one of a finite number of states. The states of the neighbors of a point and the state of the point at time n induce a state of the point at time $n + 1$ by means of a formula. Ulam calls a finite subset F of the graph an automaton of organism if F satisfies the following two conditions. First, the states of F must be periodic or almost periodic functions of time. Second, the neighbors of F must have only a 'weak' influence on F , but F can have a strong influence on its neighbors. (These imprecise statements are made precise by von Neumann.) Ulam proposed that one should find formulae whereby an automaton can reproduce itself in the sense that another automaton, congruent to the original automaton, arises in some other part of the graph, assuming only one automata exists initially. For his own model von Neumann chose 29 states for each point, the points being those with integer coordinates. Each point communicates with its four contiguous neighbors with a delay of at least one unit of time. The points all have the same rules of operation; i.e., the points were homogeneous in space.

On page 30 of Ulam's book, *A Collection of Mathematical Problems*, he posed problems about cellular automata problems in terms of infinite matrices; e.g., does there exist an infinite matrix A of zeros and ones with uniformly bounded row sum so that every possible finite matrix of zeros and ones appears as some power of A ?

Ulam wrote three subsequent papers or reports on cellular automata. The 1962 paper [3] discusses growth models with the plane tessellated into regions which are squares or equilateral triangles. An organism is a single one of the regions which is assigned the value one to denote that it is occupied. An unoccupied region is assigned the value zero. The regions of the plane then evolve in discrete time steps according to various rules involving occupied and unoccupied regions. The state of an organism may be generated to take account of the age of the cell. Examples are given.

The report [4], by Ulam and Schrandt, continued this work. They were able to find self-reproducing automata. They also had models in which the descendents of different systems could collide and 'fights' would ensue for space. In some cases both sets of descendents would die out. They also calculated models of automata in three dimen-

sions. Some very striking pictures appear in this report. These last two reports have been reproduced in the book *Cellular Automata*, edited by A. W. Burks [3, 4]. Schrandt and Ulam realized self-reproducing automata with simpler models than those used by von Neumann.

These cellular automata systems have had an active life. Some of it is summarized by Butler and Ntafos [5]. We quote their abstract:

A major impediment in the application of cellular spaces to the modeling of biological phenomena and other physical processes has been the lack of tools which relate global to local behavior. In this paper, the vector string description is introduced as a tool for the analogies of the cellular spaces studied by S. Ulam. With such a tool, it is quite straightforward to show the existence of global properties which would otherwise be very difficult to prove. For examples, a conjecture by Ulam concerning the density of cells in a 'growth' pattern is shown to be true. We also show that a self-reproduction process occurs which allows regeneration of specific patterns from 'cuttings' of those patterns.

Moore [6] made the surprising discovery that certain patterns cannot be reproduced, and thus is a Garden of Eden configuration.

The subject of cellular automata has had a long and distinguished history since Ulam and von Neumann founded the subject. For example, 95 references to cellular automata have appeared in the *Mathematical Reviews* since 1972. See the book *Cellular Automata* [7] edited by Farmer *et al.*

3. Population Biology

Ulam's interest in population dynamics dates from the beginning of his association with war-time Los Alamos and the work on chain reactions in nuclear fission devices. The development of Monte Carlo methods, introduced by Ulam, Fermi, von Neumann, and Metropolis was a natural calculational tool for the study of the evolution of populations.

We will discuss Ulam's work on population biology under four headings: multiplicative processes, evolution, genealogical distances, and binary reaction systems.

3.1. MULTIPLICATIVE PROCESSES

The four papers [8], one with Hawkins and three with Everett, are watershed papers in the theory of branching processes. Both the theory and history of this subject are given in the book of Harris [9]. In a biological context, simple branching processes can be associated with sexless reproduction. While there were earlier papers on the subject, particularly in reference to the problem of survival of family surnames, the papers by Ulam and his colleagues were the first to set out a comprehensive theory of these processes. For example, Hawkins and Ulam were the first to derive the Abel functional equation for the moment-generating function for the Galton-Watson branching process. Everett and Ulam derived general expressions for the first and second moments for processes in which several types of objects are involved. They also derived an expression for the limiting vector of types and derived the generalized functional equations for the moment-generating function.

In Everett-Ulam [8], Parts I and II, the generating transformation $G(x)$ whose iterates yield the probability distribution for higher generations is introduced and investigated. They investigate both the subcritical and supercritical case. In Everett-Ulam [8], Part III, the set of all possible genealogies for a given branching process starting with one particle of type i is treated abstractly. As Ulam would do naturally, a distance is immediately defined on the set and the set becomes a zero-dimensional metric space satisfying the second axiom of countability. A measure is then defined on this genealogy space. If $x_0 = G(x^0)$ is a fixed-point of $G(x)$, it is shown that the set of genealogies which terminate in death has measure x_i^0 where i is the type of particle which starts the genealogies. Let V be the eigenvector corresponding to the maximal positive eigenvalue $\lambda > 0$ of the supercritical system. The set of all genealogies whose k th generation population approaches the ratios $V_1 : V_2 : \dots : V_i$ has measure $1 - x_i^0$. Hence, almost all genealogies either terminate in death or approach V as a limit.

3.2. EVOLUTION

In contradistinction to particle multiplication in physics, biological populations usually include mating rules. This gives evolutionary trees a different appearance.

Ulam's work on models of evolution was done with Schrandt. These models were first reported on in a transcript of a lecture [10] Ulam gave at the Wistar Institute of Anatomy and Biology. The transcript is amusing reading. (The audience had trouble with the meaning of ' α ' and frequently so informed the speaker.) A smoother presentation was given by Schrandt and Ulam [11] in 1971 where the computational realizations of the models were completed. The parameters of the Schrandt-Ulam evolution models include:

- (a) number of generations,
- (b) population size,
- (c) total number of favorable mutations necessary to produce the desired characteristic,
- (d) probability of a favorable mutation per individual in the population,
- (e) the value γ of a single favorable mutation such that an individual having this mutation would have $k + \gamma$ descendents versus k descendents otherwise.

Additional conditions are then added in the several models. One model, ADAM, allows only asexual reproduction. A second model, called EVE, allows for sexual reproduction. It is observed that sexual reproduction yields faster evolution, and additional work is promised.

In the paper with T. F. Smith, Myron Stein, and Beyer, [12], an investigation is carried out of the reconstruction of evolutionary trees based on 33 protein cytochrome C's from 33 extant plants or animals. Distance between these proteins is calculated by a theory discussed in [13] and hypothetical evolutionary trees constructed by use of linear programming methods. Agreement of the trees with generally accepted evolutionary trees was reasonably good.

3.3. GENEALOGICAL DISTANCES

Pure branching processes can be symbolically represented by a graph-theoretic tree. Ulam was often interested in the problem of genealogical trees (different from graph-theoretical trees) in which mating is required to produce additional branches [14]. In conversation he would call the resulting 'graphs', pair trees. One of his ambitions was to develop their theory.

The paper with Mycielski [15] presents some results on the pairing process and the notion of genealogical distance. The Mycielski-Ulam pair tree is the following. A population P is divided into disjoint generations $A_i, i = 1, 2, 3, \dots$. The size of each generation A_i is a fixed even number n . Each generation is divided randomly into couples and each couple produces exactly two individuals in the next generation. The authors then ask for ways of measuring the genealogical distance between individuals. Who, for example, are close cousins and who are distant cousins? Of course, this will depend on which divisions into couples are realized. The authors discuss three different distances and their properties. For example, the authors place a natural probability measure on the space of realizations of the random matings. They then show that for almost all realizations, the first of their distances are in fact metrics. They also show that the expected value of the distances are always finite.

Kahane and Marr [16] have generalized the Mycielski-Ulam theory to more general pairing processes in which the population size remains constant from generation to generation, but the number of offsprings per pair can vary. Kahane and Marr also obtain stronger bounds on the mean distance.

3.4. BINARY REACTION SYSTEMS

The study of population dynamics with mating rules takes another form in the work of Mary Menzel, P. R. Stein, and Ulam [17] and Stein and Menzel [18]. The first reference [17] deals with quadratic transformations. This was then generalized to cubic transformations [18]. The work on cubic transformations has no biological context, unless someday trisexual mating makes its appearance. Of course, one could argue that each human adult is the product of not only his or her parents, but of his or her culture.

An example of a quadratic transformation arising from mating rules in a population whose size remains constant from generation to generation and which contains three types of individuals is the following. Let x_i be the fraction of the population of type i at the n th generation. Suppose that the following mating rule holds:

type 2	and	type 2	produce	type 1
3		3		1
1		2		1
1		3		2
2		3		2
1		1		3

At the $n + 1$ generation the fractions become

$$x_1 = x_2^2 + x_3^2 + 2x_1x_2, \quad x_2 = 2x_1x_2 + 2x_2x_3, \quad x_3 = x_1^2 \quad (*)$$

with $0 \leq x_i \leq 1$ and $x_1 + x_2 + x_3 = 1$.

Menzel, Stein, and Ulam carried out several numerical studies iterating the transformation (*). In fact, they examined 97 inequivalent systems similar to (*). A few examples of 4-variable binary reaction systems are also studied.

The appendix (due to P. R. Stein) to the work on quadratic transformations deals with a different class of homogeneous quadratic transformations of arbitrary dimension derived more naturally from evolutionary models. For this class of transformation, the limiting behavior of arbitrary vectors under iteration is given explicitly. This rather short appendix needs to be followed up.

4. Fermi-Pasta-Ulam Results and Solitons

It is well known that Fermi-Pasta-Ulam [19] have shown that in a nonlinear system of differential equations of the form

$$x_i = (x_{i+1} + x_{i-1} - 2x_i) + a[(x_{i+1} - x_i)^2 - (x_i - x_{i-1})^2], \quad i = 1, 2, \dots, N, \quad (1)$$

where $x_i = 0$ for $i < 1$ or $i > N$, and $a > 0$ is some positive constant, the distribution of energy among the modes (obtained from the linear part of (1)) may not be asymptotically uniform and may even return once and therefore infinitely often to the original distribution. This result is one of the progenitors of the theory of solitons [20]: Solitons in turn are now playing a role in biology. A good survey article has been written by Scott [21].

Pohl [22] studies the character of cells in the reproductive state and suggests that the cell membrane may contain electrons in certain quantum states which Pohl calls Fermi-Pasta-Ulam-Frohlich resonance.

5. Pattern Recognition

In the period 1964-1982, Ulam occasionally worked on mechanisms the brain might use in pattern recognition. This work has never been published but is discussed in [23] and [24]. He hypothesized that the brain has by some means stored a set of two-dimensional pictures. A picture is thought of as a subset of the plane. A new picture then enters through the senses. What would be an efficient means of searching the memory for pictures which are 'similar' to the new picture? Ulam discussed the possibilities of using Hausdorff distance between two plane sets. A measure of similarity is defined as follows. Let $d(a, b)$ measure the distance between two points a and b in the plane. Let A and B be two closed subsets of the plane. Define the Hausdorff distance by

$$\rho_H(A, B) = \max_{x \in A} \min_{y \in B} d(x, y) + \max_{y \in B} \min_{x \in A} d(x, y).$$

The quantity $\rho_H(A, B)$ is then a real-valued symmetric nonnegative function which satisfies the triangle inequality: $\rho_H(A, C) \leq \rho_H(A, B) + \rho_H(B, C)$. Also $\rho_A(A, B) = 0$ only when $A = B$.

It is suggested there is an evolutionary advantage for the memory to form a class \bar{A} of pictures from each picture by slightly altering A , say by rotation or a slight deformation. Suppose the classes A and B are finite, or compact in the case of infinite classes. Then for two classes \bar{A} and \bar{B} iterate the Hausdorff distance to obtain

$$\rho(\bar{A}, \bar{B}) = \max_{A \in \bar{A}} \min_{B \in \bar{B}} \rho_H(A, B) + \max_{B \in \bar{B}} \min_{A \in \bar{A}} \rho_H(A, B).$$

It may also be useful to consider a distance between sets 'modulo' small sets. A few extra points on a pattern should not change the pattern's class: For example, one might use

$$\bar{\rho}_H(A, B) = \min_{A - C_A} \min_{B - C_B} \rho_H(A, B)$$

where C_A and C_B are sets 'small' compared to A and B and the two minima are taken over all 'small' sets. Obviously, this concept needs to be made more precise.

Other suggestions for distance between sets are also given in [25]. Additional work on these suggestions would be useful.

Around 1965, Schrandt and Ulam carried out some computer experiments at Los Alamos on pattern recognition. An example of a manuscript letter A on a square and an example of a manuscript letter B on a square were each encoded and stored in a computer. Two transformations S and T were chosen, each of which distorted a picture slightly. All $2^7 = 128$ different transformations $W_1 W_2 \dots W_7$, where $W_i = S$ or T , were applied to each of the two example letters A and B , yielding 128 new examples for each. A test letter was then selected and compared (by Hausdorff distance) with each of the samples. A score was then tallied and the letter with the better test score was selected.

6. Biometric Spaces

The concept of a biological metric space introduced by Ulam around 1970 stands apart from the usual examples of mathematical thinking in biology. This metric found immediate uses in genetic studies to answer questions about the relationships between molecules of DNA and proteins. Ulam's metric spaces are strikingly different from well known Euclidean, or locally Euclidean physical metric spaces. Ulam's point-sets are discrete, and the points are sequences of varying lengths. They are combinatorial in nature, as are those topics of biology which are farthest removed from physical theories: evolution, biological automata, and classification problems. The idea that biology is becoming increasingly understood in combinatorial terms was brought out by Ulam articles [25, 26], in which his biological metric was presented as a principal example of biomathematics.

On 3 April 1972, Ulam gave a lecture at the Rockefeller University, entitled 'On Some Mathematical Questions Suggested by Problems in Biology' (unpublished). On this

occasion, speaking to mathematicians and physicists, he defined some biological metric spaces explicitly and, passing quickly over their biological significance, asked the audience to consider them from a purely mathematical viewpoint.

His simplest example was as follows: The metric space is a set of finite sequences of varying lengths, whose terms are taken from a finite alphabet. (For DNA the letters are *A*, *G*, *C* and *T*, but from a mathematical viewpoint any set would do.) The distance between two sequences is defined as the minimum total number (or cost) of mutations or deletions in either sequence needed to make them both equal. A deletion shortens the sequence by the removal of a term, and a mutation changes a term in one sequence to make it the same as a term in the other. For instance, the distance between the words GOOD and BAD is 3, because we have to change 2 letters and delete 1 at least to make the words equal. This has been called *evolutionary distance*, which is somewhat misleading because evolution does not necessarily take the shortest distance. However, from the standpoint of evolutionary studies, it presents a rational basis of measurement. Also, later it was discovered that it could be easily calculated.

Soon it was realized that a dynamic programming method which had been used by Needleman and Wunsch [27] for best matching of two sequences could be modified to compute Ulam's metric [28]. Let $\mathbf{x} = x_1 x_2 \dots x_n$, $\mathbf{y} = y_1 y_2 \dots y_m$, and $D(\mathbf{x}, \mathbf{y})$ denote Ulam's distance between the sequences. Also let $d(a, b)$ denote the positive weight of a mutation from a to b and $d(-, a) = d(a, -)$ be the weight of an insertion and deletion. If $D_{i,j} = D(x_1 \dots x_i, y_1 \dots y_j)$ then the algorithm is

$$D_{i,j} = \min \{D_{i-1,j} + d(x_i, -), D_{i-1,j-1} + d(x_i, y_j), D_{i,j-1} + d(-, y_j)\} .$$

Of course $D_{n,m} = D(\mathbf{x}, \mathbf{y})$. The computation time is proportional to nm . Other biological metrics have been shown to be readily computable by the same general method. See the review [29] for discussion of these and other methods of sequence comparison.

6.1. APPLICATIONS

The most obvious applications of this metric space is to the construction of evolutionary trees of homologous protein sequences as described above [12]. Since the distances between the objects at the vertices of such a tree satisfy the metric axioms, they will avoid the anomalies which will occur, in general, in nonmetric tree constructions.

A second application of the metric, also suggested by Ulam, is as a pattern recognition device. If a DNA sequence is given, the distance algorithm can be used to locate that region of the sequence whose distance from some particular sequential pattern of interest is a minimum. This search can be carried out in such a way as to find all regions in the given sequence whose distance from the pattern is a local minimum with computing time proportional to the length of the pattern multiplied by the length of the sequence [30]. In recent years the use of dynamic programming for pattern recognition has been extensive, not necessarily using distance as the objective function on which

the procedure is based [29]. These uses demonstrate amply the correctness of Ulam's pattern recognition proposal, even though he apparently had no idea of how it would eventually be carried out.

6.2. PROBABILITY DISTRIBUTIONS

Statistical significance is one of the first questions which arises when one tries to evaluate the results of a sequence comparison algorithm. The simplest of these questions can be posed, as Ulam did, with no reference to biological details. The utility to biology, however, is nontrivial. If a located sequence match is in the range of results expected from random sequences of a similar composition, then the match might reasonably be disregarded. If, however, the match is of a quality far above that expected from random sequences, then the sequence analyst might be motivated to look carefully for a biological explanation of his results.

The first attempt to address these issues was made by Chvatal and Sankoff [31] who studied $L_n = L(X, Y)$, the length of the longest common subsequence of two random sequences X and Y each of length n . They assume that the elements of the sequence are chosen randomly and independently from a k -letter alphabet. DNA sequences have $k = 4$ while protein sequences have $k = 20$. Chvatal and Sankoff prove that the expected value of L_n , $E(L_n)$ satisfies $\lim_{n \rightarrow \infty} E(L_n)/n = C_k$. Deken [32] notes that Kingman's subadditive ergodic theorem gives $P(\lim_{n \rightarrow \infty} L_n/n = C_k) = 1$. These authors give bounds on C_k but results on the distribution of L_n seem very difficult. Steele [33] gives additional results on this and related problems.

The relation between L_n and sequence distance is easily derived. Let $D(x, y)$ be the minimum distance with $d(a, -) = d(-, a) = 1$ and $d(a, b) = 2$ if $a \neq b$. Then, where 'indels' denotes 'insertions and/or deletions',

$$D(x, y) = \min \{ 2 \# \text{ mismatches} + \# \text{ indels} \} .$$

Since $2n = \text{total number of letters} = 2 \# \text{ matches} + 2 \# \text{ mismatches} + \# \text{ indels}$,

$$\begin{aligned} D_{n,n} = D(x, y) &= \min \{ 2n - 2 \# \text{ matches} \} \\ &= 2(n - \max \{ \# \text{ matches} \}) \\ &= 2(n - L_n) \end{aligned}$$

and

$$L_n = n - D_{n,n}/2 .$$

Recently, molecular biology has discovered sequence elements that are highly mobile. The result of this and other discoveries is that, while sequences might not be similar when matched end to end, there may be segments of sequence with high similarity. This motivates the study of the longest matching segment between two random sequences. Assume, as above, that X and Y are independent and identically distributed and that $p = P(X_i = Y_j) \in (0, 1)$. Erdos and Renyi in 1970 [34] showed that the length R_n of the

longest head run in n independent tosses of a fair coin satisfies

$$P\left(\lim_{n \rightarrow \infty} R_n / \log_{1/p}(n) = 1\right) = 1.$$

This result has been generalized to show that the length M_n of the longest exact match between $X_1 \dots X_n$ and $Y_1 \dots Y_n$, allowing shifts, doubles the length of the match in the sense that

$$P\left(\lim_{n \rightarrow \infty} M_n / \log_{1/p}(n) = 2\right) = 1.$$

In addition, it can be shown that as the two sequence lengths vary or the two distributions change, the limiting behavior stays identical in a surprising large set. Then a phase transition takes place and $M_n \sim C \log_{1/p}(n)$; with $C < 1$.

These results are closely connected with finding the maximum of *independent* random variables and, not surprisingly, the extreme value distribution appears. If $M_n(k)$ equals the length of the longest match with shifts, then it can be proved that

$$E(M_n(k)) = \log_{1/p}(n^2) + \gamma/\lambda - \frac{1}{2} + r_1(n) + o(1),$$

and

$$\text{Var}(M_n(k)) = \pi^2/6\lambda^2 + \frac{1}{12} + r_2(n) + o(1),$$

where $q = 1 - p$, $\gamma = 0.577 \dots$ is the Euler-Mascheroni constant, and where $r_1(n)$ and $r_2(n)$ are small. The striking feature of the variance being approximately constant with n is related to the extreme value distribution.

Generalizations of these results exist. Different sequence lengths, differently distributed sequences, the case of Markov chains, and the case of more than two sequences have been studied; see [29]. These probability distributions will have an important role in pattern recognition.

7. Conclusion

Obviously biology can be enriched with mathematical ideas, but Stanislaw Ulam has argued that the converse is also true. In his opening remarks in a lecture given at the Rockefeller University in 1983 he put the idea to a group of biologists with the following parody: "Ask not what mathematics can do for you; ask what you can do for mathematics."

A biologist who thinks of mathematics strictly as a technical tool with isolated scientific applications and that his discipline is fundamentally nonmathematical, is implying that, on the whole, biology is beyond the abilities of humans to consistently describe its logic. Ulam's remark suggest the opposite, that biology is a deep and complicated subject which is a rich source of mathematical problems. Its logical structure is a challenge to mathematical ingenuity. The biological contributions of Ulam are an attempt to get at this logical structure. This mathematical subject deserves to be

studied in its own right. And yet, mathematicians who study biology usually focus principally on properties which supply answers to specific biological questions. A bolder approach would be to study these problems with a view to finding theorems which answer questions which have not yet been asked – to look for and discover properties which have not been suspected. Ultimately this can mean not merely better answers to known biological questions, but new insights into biology itself, not to mention mathematics and other fields of application.

References

1. Von Neumann, J., *Theory of Self-Reproducing Automata*, Univ. of Illinois Press, Urbana, 1966.
2. Ulam, S. M., 'Random Processes and Transformations', *Proc. International Congress of Mathematicians, Cambridge, MA, 1950*. Vol. 2, pp. 264–275, 1952.
3. Ulam, S. M., 'On Some Mathematical Problems Connected with Patterns of Growth of Figures', *Proc. Symp. Appl. Math.* **14**, 215–224 (1962); also in A. W. Burks (ed.), *Cellular Automata*, Univ. of Illinois Press, Urbana, 1970, pp. 219–231.
4. Schrandt, R. G. and Ulam, S. M., 'On Recursively Defined Geometrical Objects and Patterns of Growth', Los Alamos Scientific Laboratory, LA-3762, 19 pp., 1967, also in A. W. Burks (ed.), *Cellular Automata*, Univ. of Illinois Press, Urbana, 1970, pp. 232–243.
5. Butler, J. T. and Ntafos, S. C., 'The Vector String Descriptor as a Tool in the Analysis of Cellular Automata Systems', *Math. Biosci.* **35**, 55–84 (1977).
6. Moore, E. F., 'Machine Models of Self Reproduction', *Proc. Symp. Appl. Math.* **14** (1962).
7. Farmer, D., Toffoli, T., and Wolfram, S. (eds.), *Cellular Automata*, Proceedings of an Interdisciplinary Workshop, Los Alamos, New Mexico, U.S.A., North-Holland, Amsterdam, 1983.
8. Hawkins, D. and Ulam, S. M., 'Theory of Multiplicative Processes', Los Alamos Scientific Laboratory, LA-171, 1944, declass. 1946;
Everett, C. J. and Ulam, S. M., 'Multiplicative Systems. Part I.', *Proc. Nat. Acad. Sci.* **34**, 403–405 (1948);
Everett, C. J. and Ulam, S. M., 'Multiplicative Systems in Several Variables', Los Alamos Scientific Laboratory, Part I, LA-683, 1948, Part II, LA-690, 1948, Part III, LA-707, 1948.
9. Harris, T. E., *The Theory of Branching Processes*. Springer-Verlag, Berlin, 1963.
10. Ulam, S. M., 'How to Formulate Mathematically Problems of the Rate of Evolution', in *Mathematical Challenges to the Neo-Darwinian Interpretation of Evolution*, A Wistar Institute Monograph No. 5, 21–33 (1967).
11. Schradt, R. and Ulam, S. M., 'Some Elementary Attempts at Numerical Modeling of Problems Concerning Rates of Evolutionary Processes', Los Alamos Scientific Laboratory, LA-4573-MS, 1971.
12. Beyer, W., Smith, T. F., Stein, M. L., and Ulam, S. M., 'A Molecular Sequence Metric and Evolutionary Trees', *Math. Biosci.* **19**, 9–25 (1974).
13. Beyer, W. A., Smith, T. F., Stein, M. L., and Ulam, S. M., 'Metrics in Biology, An Introduction', Los Alamos Scientific Laboratory, LA-4973, 1972.
14. Ulam, S. M., 'On the Operations of Pair Production, Transmutations and Generalized Random Walks', *Adv. Appl. Math.* **1**, 7–21 (1980).
15. Mycielski, J. and Ulam, S. M., 'On the Pairing Process and the Notion of Genealogical Distance', *J. Comb. Theory* **6**, 227–234 (1969).
16. Kahane, J. and Marr, R., 'On a Class of Stochastic Pairing Processes and the Mycielski–Ulam Notions of Genealogical Distance', *J. Comb. Theory* **A13**, 383–400 (1973).
17. Menzel, M., Stein, P. R., and Ulam, S. M., 'Quadratic Transformations, Part I', Los Alamos Scientific Laboratory, LA-2503, 158 pp. (1959).
18. Stein, P. R. and Ulam, S. M., 'Non-Linear Transformation Studies on Electronic Computers', Los Alamos Scientific Laboratory, LA-DC 5688, 128 pp., 1963, also in *Rozprawy Matematyczne* (Warsaw) **39**, 1–66, (1964) and in A. W. Burks (ed.), *Cellular Automata*, Univ. of Illinois Press, Urbana, 1970, pp. 244–263.
19. Fermi, E., Pasta, J., and Ulam, S. M., 'Studies of Non-Linear Problems', Los Alamos Scientific

- Laboratory, LA-1940, 20 pp., 1955, also in *Enrico Fermi Collected Papers*, Univ. of Chicago Press, Illinois, 1965, pp. 978-988.
20. Lamb, G. L., Jr., *Elements of Soliton Theory*, Wiley, New York, 1980.
 21. Scott, A. C., 'Solitons in Biological Molecules', to appear.
 22. Pohl, H. A., 'Do Cells in a Reproductial State Exhibit a Fermi-Pasta-Ulam-Frohlich Resonance and Limit Electromagnetic Radiation', *J. Biol. Phys.* **8**, 45-75.
 23. Ulam, S. M., 'Speculations About the Mechanisms of Recognition and Discrimination', Los Alamos National Laboratory, Preprint, LA-UR 82-62, 1981.
 24. Ulam, S. M., 'Reflexions on the Brain's Attempts to Understand Itself', Gamow Memorial Lecture, Univ. of Colorado, 3 October 1982, Preprint.
 25. Ulam, S. M., 'Some Ideas and Prospects in Biomathematics', *Ann. Rev. Biophys. Bioeng.* **1**, 277-292 (1972).
 26. Ulam S. M., 'Some Combinatorial Problems Studied Experimentally on Computing Machines', in S. K. Zaremba (ed.), *Applications of Number Theory to Numerical Analysis*, Academic Press, New York, 1972, pp. 1-10.
 27. Needleman, S. B. and Wunsch, C. D., 'A General Method Applicable to the Search for Similarities in the Amino Acid Sequences of Two Proteins', *J. Mol. Biol.* **48**, 443-453 (1970).
 28. Sellers, P. H., 'An Algorithm for the Distance Between Two Finite Sequences', *J. Comb. Theory.* **16**, 253-258 (1974).
 29. Waterman, M. S., 'General Methods of Sequence Comparison', *Bull. Math. Biology* **46**, 473-500 (1984).
 30. Sellers, P. H., 'The Theory and Computation of Evolutionary Distances: Pattern Recognition', *J. Algor.* **1**, 359-373 (1980).
 31. Chvatal, V. and Sankoff, D., 'Longest Common Subsequence of Two Random Sequences', *J. App. Prob.* **12**, 306-315 (1975).
 32. Deken, J., 'Some Limit Results for Longest Common Subsequences', *Discrete Math.* **26**, 17-31 (1979).
 33. Steele, M. J., 'Long Common Subsequences and the Proximity of Two Random Strings', *SIAM J. Appl. Math.* **42**, 731-737 (1982).
 34. Erdos, P. and Renyi, A., 'On a New Law of Large Numbers', *J. Anal. Math.* **22**, 103-111 (1970). Reprinted in *Selected Papers of Alfred Renyi*, Vol. 3, Akademiai Kiado, Budapest, 1976; pp. 1962-1970.