

# Reversible Cellular Automata

Harold V. McIntosh

Departamento de Aplicación de Microcomputadoras,  
Instituto de Ciencias, Universidad Autónoma de Puebla,  
Apartado postal 461, 72000 Puebla, Puebla, México.  
E-mail:mcintosh@servidor.unam.mx

April 10, 1991

## Abstract

A reversible cellular automaton is one whose evolution, and therefore the entire past history of any configuration, can be uniquely deciphered. There are degrees of reversibility, depending upon whether the configurations considered are arbitrary, periodic, or quiescent at infinity; which are subsidiary to more general questions of injectivity and surjectivity, within a general perspective of the ancestry of configurations. Reversibility is examined within this general context, expanded to include the frequency distribution of ancestors and its moments. It is argued that the coupling of zero variance (judged from the maximum eigenvalue of the second moment matrix) with zero frequency for zero ancestors (surjectivity) is the fundamental concept. An ideal theoretic property inherent in decompositions of the de Bruijn matrix suffices to prove the coupling for automata. Surjectivity in different contexts, injectivity, and degrees of multiple valuedness all follow from this central result. Although the article is intended as a review, it is far from a complete historical survey; the presentation is uniformized through the use of graphs, de Bruijn diagrams and matrices wherever possible.

## 1 Introduction

When John von Neumann [1] worked out the theory of universal constructors within the framework of cellular automata in the mid fifties, philosophical questions soon arose concerning ancestorless patterns, which surely could not be “constructed” through the intermediary of other configurations occupying the same system. That such predicaments could arise under certain circumstances was first appreciated by Edward F. Moore [2], whose article of 1962 was reprinted in a collection of influential papers edited by Arthur W. Burks [3] in 1970. Even though the only requirement which von Neumann was interested in satisfying

was that a device with the computational power of a Turing machine should be capable of producing copies of itself — not necessarily that it had to fill space with arbitrary designs — the latter concept is an interesting one which has received further study.

In short order the question turned into the relationship between the *local* mapping in an automaton, which assigns successors to neighborhoods, and the *global* mapping, which produces the succession of configurations from one generation to the next. In contemporary mathematical terms, the interesting questions are whether the global mapping is *injective*, formerly called *one-to-one*, and also whether it is *surjective*, formerly designated *onto*. In the latter case every configuration has at least one ancestor, and there is no Garden of Eden. For finite sets the equivalence of the two characteristics follows from counting arguments which no longer apply to infinite sets.

To pose the problem properly for infinite automata, some thought has to be given to the meaning of infinity, often with respect to boundary conditions which are relaxed or assumed to lose their influence as the automaton grows in size. Three choices usually considered are: 1) cyclic boundary conditions, 2) quiescence at infinity, or 3) a quite general, unspecified arrangement (sometimes with a topology which minimizes the importance of exceedingly remote cells).

In a sense, these are external considerations, subordinate to other, internal, aspects of the problem. Fundamental to the nature of cellular automata is the difference in size between cells and the neighborhoods which determine their evolution, resulting in a permanent discrepancy between the area of a region and that from which it has evolved. Paying attention to the external aspects may try to avoid the difficulty by folding or smoothing the margins, but it is probably better to resolve the internal aspects first, adjusting them to the external requirements at the end.

Although it is quite easy to work out ancestral neighborhoods for one or even a few cells, problems of consistency are bound to arise because the neighborhoods always overlap. Fortunately, especially in one dimension, there are matrix techniques based on the de Bruijn diagram which organize the relationships involved, and the calculation of ancestors can proceed. Most commonly inconsistencies arise which cannot be resolved, leading to the Garden of Eden, and the conclusion that most rules of evolution are not surjective. For such rules, most configurations have multiple ancestors whilst others have none.

Among the exceptions, the freedom of choice which exists in the margin usually leads not just to a single ancestor for each configuration, but to several distinct ancestors. But for a still smaller subset of rules, the freedom is to greater or lesser degree illusory, even to the extent that the state of one of the cells in the ancestral neighborhood — the central cell, for instance — can be deduced once the state of the evolved cell and some of its neighbors is known. The rule of association can be used to run the automaton backward, for which such an automaton would be called reversible.

Historically, various attempts were made to define the problem, and to en-

counter examples of various types of behavior; some of the conflicts inherent in which were discussed by Sven Skyum [4] in a 1975 article aptly titled “Confusion in the Garden of Eden.”

Interestingly, similar studies had grown out of the topic of symbolic dynamics and ergodic theory, originating with Henri Poincaré in the nineteenth century, continued in the first half of the twentieth by George Birkhoff [5], formalized by Walter Gottschalk and Gustav A. Hedlund [6] in a 1955 Colloquium Publication of the American Mathematical Society. In 1969 Hedlund [7] published a thorough summary of the properties of automorphisms and endomorphisms of the shift dynamical system, albeit written in a very terse mathematical language with many topological features.

Indeed, continuous mappings of the shift *are* cellular automata, providing one of the few known instances of an unequivocal application of the theory of cellular automata to another field; ironically the applications to symbolic dynamics and ergodic theory were worked out in exhaustive detail much before the subject of cellular automata became particularly organized, and quite unbeknownst to the majority of its practitioners.

In 1978, about a decade after Hedlund’s fundamental paper, Masakazu Nasu [8] undertook to describe these results with the aid of graph theory, without relying on topology, and incorporating some ideas from automata theory. During this same period a variety of other articles appeared; but in the meantime, a more direct approach to the explicit construction of reversible automata was invented, then applied to some physical problems, by Edward Fredkin, Tommaso Toffoli, and others.

Recent interest in automata has been greatly stimulated by Stephen Wolfram’s computer experiments, first reported in *Reviews of Modern Physics* [9] in 1983; familiarity with his reprint collection [10] of 1986 is assumed, as well as his use of  $(k, r)$  whenever  $k$ -state cells form a neighborhood of radius  $r$  (possibly half-integral) is followed, likewise his  $k$ -ary enumeration of evolutionary rules.

This paper also relies heavily on the properties of nonnegative matrices, an exposition of whose theory can be found in half a dozen or more contemporary textbooks, such as Gantmacher [11], Bellman [12], Varga [13], Seneta [14], Berman and Plemmons [15], or Minc [16].

## 2 Fredkin’s model

Edward Fredkin discovered that the parity, or exclusive or, operation could lead to an interesting replication of patterns in cellular automata [17, 18, 19]; he also devised a model for reversible automata [20]. Reputedly differential equations were his model; first order equations typically have real exponential solutions which are not time reversible, whereas second order equations often have solutions composed of readily reversible pairs of complex conjugates.

## 2.1 Two-generation rule

The proposed rule of evolution involved two generations of cells: let  $x_i^t$  represent the state of the  $i^{\text{th}}$  cell in generation  $t$ . Then if

$$x_i^{t+1} = \varphi(x_{i-1}^t, x_i^t, x_{i+1}^t)$$

were the rule of evolution of any ordinary  $(2, 1)$  automaton, the new rule

$$\begin{aligned} x_i^{t+1} &= \varphi(x_{i-1}^t, x_i^t, x_{i+1}^t) \oplus x_i^{t-1} \\ &= \Phi(x_i^{t-1}, x_{i-1}^t, x_i^t, x_{i+1}^t) \end{aligned} \quad (1)$$

spans two generations, but can be inverted by writing

$$\begin{aligned} x_i^{t-1} &= \varphi(x_{i-1}^t, x_i^t, x_{i+1}^t) \oplus x_i^{t+1} \\ &= \Phi^{-1}(x_i^{t+1}, x_{i-1}^t, x_i^t, x_{i+1}^t), \end{aligned} \quad (2)$$

on account of the algebraic properties of the exclusive or. Any other invertible function of two variables could have been used in place of  $\oplus$ ; for boolean variables the only other candidate is exclusive nor (equivalence).

## 2.2 Cartesian product rule

Toffoli's own scheme [21] increases the dimensionality of the automaton rather than the number of generations spanned by its evolutionary rule. However, the simplest solution to the problem of exhibiting explicitly reversible rules of evolution may well lie in increasing the number of states of the automaton, by forming their cartesian products, whose rule of evolution needs to span but a single generation:

$$\begin{aligned} (x_i, y_i)^{t+1} &= (y_i, \varphi(y_{i-1}, y_i, y_{i+1}) \oplus x_i) \\ &= \Phi((x_{i-1}, y_{i-1})^t, (x_i, y_i)^t, (x_{i+1}, y_{i+1})^t). \end{aligned} \quad (3)$$

Thus a single cell of a  $(4, 1)$  automaton may hold two successive generations of a  $(2, 1)$  automaton by forming pairs whose components are to be interpreted as

$$(x_i^{t-1}, x_i^t);$$

substitution of these values into Eq. 3 reproduces Eq. 1, in its turn reversible by the rule

$$\begin{aligned} (x_i, y_i)^{t+1} &= (\varphi(x_{i-1}^t, x_i^t, x_{i+1}^t) \oplus y_i^t, y_i^t) \\ &= \Phi^{-1}((x_{i-1}, y_{i-1})^t, (x_i, y_i)^t, (x_{i+1}, y_{i+1})^t). \end{aligned} \quad (4)$$

Similar formulas would allow the formation of reversible  $(k^2, r)$  rules from arbitrary  $(k, r)$  rules, the more so by taking advantage of additional alternatives for  $\oplus$ . Nor is it difficult to invent schemes whereby the equivalent of three generations of evolution could be incorporated into  $(k^3, r)$  automata, and so on and on for even longer histories.

To understand possible complications which can arise, consider Fredkin's idea applied to a  $(2, 1/2)$  automaton via a rule in which sequential letters replace subscript indices for improved legibility:

$$\Phi((a, b), (c, d)) = (d, \varphi(b, d) \oplus a).$$

Evolution loses the value of  $c$ , preserves  $d$  intact, yet permits the recovery of  $a$  if  $b$  were known explicitly. Supposing a third cell  $(e, f)$  provides two cells in the second generation,

$$(d, \varphi(b, d) \oplus a) \quad \text{and} \quad (f, \varphi(d, f) \oplus c),$$

from which  $c$  can now be recovered via  $c = \varphi(d, f) \oplus (\varphi(d, f) \oplus c)$  with the aid of the explicitly known  $d$  and  $f$ . Since  $d$  is already known, the cell  $(c, d)$  has been reconstructed, as could be done for all the rest of the cells of a configuration.

In each case evolution "remembers" the second component of the *right* cell, but the *left* ancestor is the one recovered by the reversed rule. Typical of neighborhoods with half integral radius, this asymmetry seems to have no further significance. It is often seen in rules with shifting.

### 2.3 Rules spanning three generations

The next rule depends upon three generations of evolution, following the same convention as before:

$$\Phi((a, b, c), (d, e, f)) = (e, c, \varphi(c, f) \oplus a). \quad (5)$$

Time reversal cannot be achieved as before by a rule with a two cell neighborhood, because the new state is independent of both  $b$  and  $d$ . Again supposing a cell  $(g, h, i)$  to the right of the first two, all three generating the evolved pair

$$(e, c, \varphi(c, f) \oplus a) \quad \text{and} \quad (h, f, \varphi(f, i) \oplus d),$$

the lack of  $b$  prevents recovery of  $(a, b, c)$ , while the lack of  $i$  prevents releasing  $d$  from the combination in which it is bound (admittedly, a degenerate  $\varphi$  would help) in order to recover  $(d, e, f)$ . Not even  $(g, h, i)$  can be obtained, for the lack of  $g$ ; even less information is available going further afield.

*However*, a third cell in the second generation, involving yet a fourth cell from the first generation, would have the form  $(h, i, x)$  wherein the value of  $x$  is immaterial, but  $i$  combines with  $f$  to release  $d$  from  $\varphi(f, i) \oplus d$  revealing  $(d, e, f)$ .

Altogether, we have a reversible  $(8, 1/2)$  rule which requires an  $(8, 1)$  (three cell neighborhood) rule for its reversion. Altogether, the inverse rule appears to be

$$\begin{aligned} \Phi^{-1}((a, b, c), (d, e, f), (g, h, i)) \\ = (\varphi(e, h) \oplus f, a, e). \end{aligned} \tag{6}$$

Evidently considerable ingenuity can be exercised in devising reversible automata, but it would be desirable to have criteria of necessity or sufficiency for the process. One significant aspect of the formulas shown so far is their uniformity in the respect that every state has the same number of counterimages as any other. Additionally, the rules are entirely local; even though the size of the neighborhood of an inverse rule may be different from that of the original rule, it is fixed; moreover distant boundary conditions do not affect it.

### 3 Some relevant diagrams

One approach to encountering reversible automata is to scan all the automata of a given order, to find those for which there are unique ancestors, and to work out their rules of inversion. The task is formidable, considering that there are  $k^{k^{2r+1}}$  different rules for  $(k, r)$  automata; S. Amoroso and Y. N. Patt [22] reported a survey of  $(2, 3/2)$  automata in 1972, which uncovered eight instances of nontrivial reversible automata, whose principle of operation differed from the later ideas of Fredkin. Previously they had known that  $(2, 1/2)$  or  $(2, 1)$  automata lacked nontrivial rules (shift, complement, identity).

Two different algorithms formed the content of their article; one determined the existence of the Garden of Eden (or else concluded that every configuration has ancestors), the other could verify that a configuration had only one ancestor. Formalizing the auxiliary diagrams and tables which were incidental to their presentation would enhance our understanding of the procedures and could facilitate the derivation of additional results.

#### 3.1 De Bruijn diagrams

By a de Bruijn diagram [23] one understands a graph whose nodes are sequences of symbols from some alphabet, for example the set of states of an automaton. The sequences all have the same length, often called the number of stages because of an application to shift register theory [24]; for automata they are partial neighborhoods. In one dimension, a single cell could be dropped from one end or the other of a given neighborhood, although other decompositions are conceivable (dropping or adding *two* cells, say).

Many authors have used de Bruijn diagrams to study cellular automata, among them Nasu [8], Erica Jen [25], Wentian Li [26], and Wolfram [27]. There does not seem to be any evidence of an attempt to formulate two-dimensional de Bruijn diagrams.

The nodes of a de Bruijn diagram are linked according to their overlap, resulting in a map which reveals the possible succession of contents in a moving window as it advances across an even longer sequence. For automata, the most convenient choice of window size — the number of stages — is the length of a full neighborhood, that being the basic unit participating in the evolutionary process.

A de Bruijn diagram is also represented by its connectivity matrix; for this purpose it is convenient to represent the states of a  $(k, r)$  automaton by the digits  $0, 1, \dots, k - 1$ , its partial neighborhoods by  $2r$ -digit numbers radix  $k$ . Multiplication by  $k$  shifts cells, arithmetic modulo  $k^{2r}$  drops the leftmost cell, adding a single digit number adjoins a new rightmost cell; altogether

$$D_{ij} = \begin{cases} 1 & j = \begin{cases} ki \\ ki + 1 \\ \dots \\ ki + (k - 1) \end{cases} \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

As an example,  $(2, 1)$  automata have four different partial neighborhoods  $00, 01, 10,$  and  $10$ , overlapping each other to form eight full neighborhoods of three cells each. Regarding these binary strings as numerical indices produces

$$D = \begin{bmatrix} 1 & 1 & \dots \\ \dots & 1 & 1 \\ 1 & 1 & \dots \\ \dots & \dots & 1 & 1 \end{bmatrix},$$

wherein dots replace zeroes for clarity.

### 3.2 The subset diagram

By its very definition, a de Bruijn diagram serves as a guide to the sequencing of strings of symbols, from which any other set of objects can be given the same sequencing. A trivial example would be to recover the original symbols from the central element of each string, while the purpose of using neighborhoods from an automaton is to obtain the sequence of evolved cells. In the same way, a de Bruijn matrix serves as a skeleton in which the rules of matrix multiplication (for noncommutative elements) ensures that the elements of a product respect the sequence of the corresponding de Bruijn diagram.

In this spirit, alternative labels can be affixed to the links in a de Bruijn diagram; reading one label instead of the other while following along some path associates the labels with one another. In particular, relating the central cell of a neighborhood to the value of the evolved cell associates configurations with their ancestors; whether or not there was any ancestor at all turns into the question of discovering whether the corresponding path exists somewhere in the diagram.

### 3.2.1 Moore's subset construction.

The subset construction [28] was devised to solve this precise problem, so that the transcription of Amoroso and Patt's first algorithm [22] into graphical language is readily accomplished.

The subset diagram of a  $2r$  stage  $k$ -ary de Bruijn diagram has  $2^{k^{2r}}$  nodes; if all configurations of this length possess ancestors, then all configurations of whatsoever length will also do so. If some do not, they can be described; in fact there is a regular expression describing every configuration in the Garden of Eden, which is nothing more than the description of the path leading from the full set to the null set in the subset diagram. Frequently required information can be found without having to use the entire subset diagram; the connected component of the full set is often all that is needed.

As an example, consider the connectivity matrix of the subset diagram belonging to the (2,1) Rule 22. With four partial neighborhoods, the subset diagram has sixteen nodes. Ranking the subsets in order of size places the full subset as number 1 and the empty set as number 16; it is likewise convenient to group intermediate subsets of similar size.

With such an arrangement links between different groups can be discerned, according to variations in the total number of links connecting nodes in the original diagram, which is represented in the subset diagram by the unit classes. Rather than joining to each other, the unit classes will link to the empty set or to larger classes to reflect inhomogenieties in the original linkage.

$$\Sigma = \begin{bmatrix} 1 & 1 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & 1 & 1 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & 1 & \dots & \dots & \dots & 1 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & 1 & \dots & \dots & 1 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & 1 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & 1 & 1 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & 2 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & 1 & \dots & \dots & \dots & \dots & \dots & \dots & 1 & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 1 & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 1 & 1 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 1 & 1 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 1 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 2 \end{bmatrix}$$

$\Sigma^8$  is the lowest power of  $\Sigma$  with an element linking the full subset to the empty set; one of the Garden of Eden configurations thus revealed is the sequence 10101001.

Note that  $\Sigma$  is partially tridiagonal, in that there are nonzero submatrices only along its principal diagonal, superdiagonal, and subdiagonal. This is not a necessary structure inasmuch as the imbalance between the number of links at the nodes of the de Bruijn diagram could have been more severe than it is, nor is it especially helpful in analyzing this particular matrix.



Nevertheless, the rows corresponding to the unit classes are significant, inasmuch as they suffice to determine all the remaining rows.

### 3.2.2 Formal definition.

The indexing shown is one of convenience for visual display, but a more mathematical arrangement would be to use  $n$ -bit binary numbers as indices. The  $i^{\text{th}}$  bit of the index would be 0 or 1 according to whether the  $i^{\text{th}}$  node is included in the subset or not. With such symbolism the remaining rows could be explicitly constructed from those associated with the unit classes.

Even without this element of sophistication, the rule of formation of the subset matrix is easily described. Begin with the observation that the fundamental requirement for the definition of any function is that each argument has just one image. Links in a graph define a function when there is just one link of a given class leading onward from a given node; the subset diagram is created when this condition is violated, by combining multiple links into a single link between subsets.

Even though the linkage of a given graph does not define a function, the subset linkage is always functional and, thanks to the inclusion of the null set as a destination, defined for the entire subset graph. Each class of links defines one function, say  $\Sigma_0$  or  $\Sigma_1$ . The subset diagram describes the union  $\Sigma_0 \cup \Sigma_1$ , which itself is not functional.

Let  $a$  and  $b$  be nodes,  $S$  a subset, and  $|S|$  the cardinality of  $S$ ; then

$$\Sigma_i(S) = \begin{cases} \phi & S = \phi \\ \{b \mid \text{link}_i(a,b)\} & S = \{a\} \\ \bigcup_{a \in S} \Sigma_i(a) & |S| > 1 \end{cases} .$$

### 3.2.3 General characteristics.

From this definition three properties of the subset diagram are evident. First, if there is a chain leading from any subset to the empty set, a similar chain will lead from any smaller (included) subset to the empty set. Conversely, if the unit classes lack links to the empty set, no other subset will have such a link either, and there will be no Garden of Eden.

Such will be the case when every node of the de Bruijn diagram has a link associated with each different kind of cell; that is, when it is row stochastic or itself a function. Similarly, the full subset will link with no others, depriving the rule of a Garden of Eden, if every node has an incoming link for each kind of cell; such rules have column stochastic de Bruijn matrices.

Rules for which the Garden of Eden fails to exist for more subtle reasons are much rarer, correspondingly more interesting, and require much more careful scrutiny of the subset diagram for their discovery.



of a pair diagram are pairs of nodes from the original diagram; labelled links join pairs whenever both members of the pair are joined by links with that same label. Paths in the pair diagram correspond to pairs of paths in the original diagram; they do not even have to originate at the same node. However, the same path taken twice can always be found amongst any others, so that the original diagram is always embedded in the pair diagram.

The pertinent question is whether there are any other, and how many, paths besides. The pair diagram is the appropriate arena for the second of Amoroso and Patt's constructions [22]. One might also refer to Tom Head's recent proposal [29].

The pair diagram for the (2, 1) Rule 22 automaton is subordinate to the maximal matrix which would correspond to Rule 0 or Rule 255, the location of the remainder of whose elements is shown by \*'s in the following matrix:

$$\Pi = \begin{bmatrix} 1 & * & \dots & * & 1 & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & * & 1 & \dots & 1 & * & \dots & \dots & \dots & \dots & \dots \\ \dots & 1 & \dots & 1 & * & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & 1 & 1 & \dots & * & * & \dots & \dots & \dots & \dots \\ \hline \dots & \dots & \dots & \dots & \dots & \dots & * & 1 & \dots & 1 & * & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & 1 & * & \dots & * & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & 1 & * & \dots & * & 1 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & * & * & \dots & \dots & 1 & 1 \\ \hline * & 1 & \dots & 1 & * & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & 1 & * & \dots & * & 1 & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & * & \dots & * & 1 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & * & * & \dots & 1 & 1 & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \hline \dots & \dots & \dots & \dots & \dots & \dots & 1 & * & \dots & 1 & * & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & * & 1 & \dots & * & 1 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & * & 1 & \dots & * & 1 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 1 & 1 & \dots & 1 & 1 \end{bmatrix}.$$

The diagram just described pertains to *ordered* pairs; a variant which is frequently encountered is the *unordered* diagram in which the distinction between  $(a, b)$  and  $(b, a)$  is considered insignificant. A smaller diagram results; it can be gotten by coalescing the nodes of the larger diagram.

## 4 Vector subset diagram

The subset diagram is concerned with the existence of paths in the original diagram, whatever their point of origin, which it discovers by trying them all. However, it records them in terms of "at least one path," displayed according to the collection of nodes which could serve as an origin (row index) versus the collection of possible destinations (column index). The principal information revealed by the subset diagram is: what amount of choice in selecting a terminal right neighborhood remains, having prescribed both the leftmost partial neighborhood and that a configuration has evolved in a certain way? To the degree that none remains, the configuration cannot have had an ancestor.

Paths leading to the empty set reveal this information; other paths or loops indicate variations in the possible partial neighborhoods to be found at any

stage in constructing ancestors. Indeed, Moore’s original intention was locating paths leading to the unit classes — paths capable of isolating one single node from the original diagram, thus casting an automaton into a known state.

In a more accurate reckoning, when counting the paths becomes of interest, specific endpoints are required, and the dichotomy between “some” or “none” is no longer appropriate. One way to retain this additional information is to enlarge the nodes in the subset diagram, making each a “vector” whose components are the original nodes; matrices linking vector nodes preserve the original links. However, we should also consider splitting the original de Bruijn matrix into  $k$  parts, just because we are interested in the differences in what can be linked with one symbol in contrast to what can be linked with another; each part is assigned to its own subset.

For a formal definition of the entire vector subset diagram, define instead a connectivity matrix. Suppose that  $\pi(S)$  is a  $k^{2r} \times k^{2r}$  projection matrix whose elements are zero save for those diagonal elements for which the  $i^{th}$  partial neighborhood belongs to  $S$ . In particular, the projector for the full set is the unit matrix, for the empty set, the zero matrix.

Then, using the matrix  $D$  of Eq. 7, the element of the vector subset matrix indexed by subsets  $S$  and  $T$  is just

$$V_{ST} = \pi(S)D\pi(T).$$

All these references to vectors and matrices can be avoided by establishing a suitable linkage within the cartesian product of the subset diagram with the original diagram, although they reappear when the connectivity matrix of the composite diagram is formed. For Rule 22 (or any other  $(2, 1)$  rule) the result would be a  $64 \times 64$  matrix with  $4 \times 4$  submatrices.

In any event, multiplying vector subset matrices induces a product of submatrices, from which most of the information which one desires can already be extracted. They may as well be used directly, discarding the intermediary of the vector subset matrix.

## 5 Symbolic de Bruijn matrices

Symbolic de Bruijn matrices and matrix fragments can be defined in several ways, each with its individual advantages. A good introduction to the techniques involved can be found in the article of R. C. Backhouse and B. A. Carré [30].

### 5.1 Neighborhood matrix

To form complex symbolic expressions, the language of regular expressions is convenient, whose operations are union, concatenation, and iteration. Another operation which is sometimes useful is the overlapping (whose symbol could be

$\diamond$ ) of two strings; if the words  $ax$  and  $yb$  each consist of letters  $a$  and  $b$ , joined to the shorter words  $x$  and  $y$ , then:

$$ax \diamond yb = \begin{cases} axb & x = y \\ \phi & \text{otherwise} \end{cases}$$

( $\phi$  is the empty word).

With such notation a symbolic de Bruijn matrix

$$D_{ij} = i \diamond j \tag{8}$$

could be defined; the symbolic form of the matrix shown above would be

$$D = \begin{bmatrix} \mathbf{000} & \mathbf{001} & \mathbf{010} & \mathbf{011} \\ \mathbf{100} & \mathbf{101} & \mathbf{110} & \mathbf{111} \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix};$$

dots now represent the empty word. Such matrices would be multiplied by using  $\diamond$  to multiply individual matrix elements. Subsequently the advantage of substituting images — such as the evolved cell — for the symbolic neighborhoods will be seen. The full neighborhood matrix is rarely used, but it serves to illustrate very explicitly the involvement of the row indices and the column indices with the individual matrix elements.

## 5.2 Evolution matrices

The multiplication of ordinary symbolic matrices follows the algebra of regular expressions; that is, multiplication concatenates symbols, union (addition) offers alternatives, the empty set acts like zero.

Moreover the  $n^{\text{th}}$  power of a matrix  $M$  contains all the possible words of length  $n$ , distributed throughout the matrix according to the overlap of the indices of constituent letters as chains are built up to length  $n$ . All this overlap in  $M^n$  is no longer explicit, yet its matrix elements are still indexed by their initial and final partial neighborhoods.

To track the evolution of a configuration consider the following variant of Eq. 8, which describes the evolution of individual neighborhoods:

$$\mathbf{S}_{ij} = \varphi(i \diamond j) \tag{9}$$

which can be decomposed into the sum

$$S = \sum_{i=0}^{k-1} S(k)$$

by defining

$$S(k)_{ij} = \begin{cases} k & \varphi(i \diamond j) = k \\ \phi & \text{otherwise} \end{cases} . \tag{10}$$

As an example, for Rule 22

$$S = \begin{bmatrix} 0 & 1 & \dots & \dots \\ \dots & \dots & 1 & 0 \\ 1 & 0 & \dots & \dots \\ \dots & \dots & 0 & 0 \end{bmatrix},$$

decomposable into the two matrices

$$S(0) = \begin{bmatrix} 0 & \dots & \dots \\ \dots & 0 & \dots \\ \dots & \dots & 0 & 0 \end{bmatrix} \text{ and } S(1) = \begin{bmatrix} \dots & 1 & \dots & \dots \\ \dots & \dots & 1 & \dots \\ 1 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}.$$

Altogether, we see that 0 has five counterimages, 1 only three. From the product  $S(0)S(0) = S(00)$ ,

$$\begin{aligned} S(0)^2 &= \begin{bmatrix} 0 & \dots & \dots \\ \dots & 0 & \dots \\ \dots & \dots & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & \dots & \dots \\ \dots & 0 & \dots \\ \dots & \dots & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 00 & \dots & \dots \\ \dots & \dots & 00 & 00 \\ \dots & \dots & \dots & 00 \\ \dots & 00 & 00 & 00 \end{bmatrix}. \end{aligned}$$

the string **00** is seen to have seven counterimages, and so on for all the other configurations.

### 5.3 Counterimage matrices

Since the evolved cell has been inserted into the symbolic matrix, it is hardly surprising that strings of evolved cells appear as elements of the product matrices. But it is decidedly inconvenient to have to look up all their indices to recover the original string, especially when the same information could have been encoded in the symbolic matrix from the beginning.

Therefore let a subscript such as  $w_t$  denote the  $t^{\text{th}}$  letter of the word  $w$ , beginning with the first;  $(i \diamond j)_{r+1}$  would be the state of the central cell of its neighborhood for integral  $r$ .

This time vary Eq. 8 by defining

$$\mathbf{s}_{ij} = (i \diamond j)_{r+1} \tag{11}$$

which is likewise decomposable into a sum

$$s = \sum_{i=0}^{k-1} s(k)$$

by defining

$$s(k)_{ij} = \begin{cases} (i \diamond j)_{r+1} & \varphi(i \diamond j) = k \\ \phi & \text{otherwise} \end{cases}. \tag{12}$$

Continuing to use Rule 22 as an example,

$$\begin{aligned}
 s(0)^2 &= \begin{bmatrix} 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 \\ \cdot & 0 & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 \\ \cdot & 0 & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 \end{bmatrix} \\
 &= \begin{bmatrix} 00 & \cdot & \cdot & \cdot \\ \cdot & \cdot & 11 & 11 \\ \cdot & \cdot & \cdot & 01 \\ \cdot & 10 & 11 & 11 \end{bmatrix}.
 \end{aligned}$$

There are still seven nonempty matrix elements. Expanded by adjoining the implicit boundary cells to obtain all seven precursors of the sequence 00, the evolution shown below results.

$$\begin{array}{ccccccc}
 0000 & 0110 & 0111 & 1011 & 1101 & 1110 & 1111 \\
 00 & 00 & 00 & 00 & 00 & 00 & 00
 \end{array}$$

There is no reason to insist that the central cell be used in defining the matrices  $s(i)$ ; frequently an ancestor at one side is chosen, particularly if the radius of the neighborhood is half integral.

## 5.4 Counting counterimages

Yet a third variation on the theme of symbolic de Bruijn matrices is actually a numerical matrix, whose purpose is to merely count counterimages, not to display them in some form. This final variant on Eq. 8,

$$D_{ij} = \begin{cases} 1 & \varphi(i \diamond j) \neq \phi \\ 0 & \text{otherwise} \end{cases}. \quad (13)$$

gives an alternative definition of the de Bruijn matrix which in turn is decomposed into a sum

$$D = \sum_{i=0}^{k-1} N(k)$$

by defining

$$N(k)_{ij} = \begin{cases} 1 & \varphi(i \diamond j) = k \\ 0 & \text{otherwise} \end{cases}. \quad (14)$$

Still using Rule 22 as an example,

$$\begin{aligned}
 N(0)^2 &= \begin{bmatrix} 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 \\ \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 \\ \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 \end{bmatrix} \\
 &= \begin{bmatrix} 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 \\ \cdot & \cdot & \cdot & 1 \\ \cdot & 1 & 1 & 1 \end{bmatrix}.
 \end{aligned}$$

Summing (rather than counting) the matrix elements again reveals seven counterimages.

## 6 The statistics of counterimages

Taking the numerical version  $N(i)$  of the symbolic matrices  $s(i)$  allows the number of counterimages to be counted, namely by summing the matrix elements of the product to be evaluated. By invoking the appropriate matrix algebra, various quantities of statistical interest can be obtained; the most important among these are averages and variances.

### 6.1 Averages

In fact there are three ways the sum yielding the number of counterimages can be carried out. If no importance is attached to the boundary cells of the sequence, all the elements should be summed. If the sequence is supposed to be cyclical, then the partial neighborhood with which the sequence starts should be the one with which it ends, and only the diagonal elements should be summed. Finally, if some specific boundaries are required, such as a quiescent partial neighborhood, then only that particular matrix element should be taken into consideration.

All of the summation techniques can be taken into account by choosing appropriate formulas from matrix algebra. Let  $\langle u|$  be a row vector all of whose components are ones,  $|u\rangle$  its transpose, and  $U = |u\rangle\langle u|$  the matrix, all of whose elements are ones. Finally  $|i\rangle$  could be a unit vector whose  $i^{th}$  component is the only one which is non-zero. Then the sum  $s_g(M)$  of all elements of the matrix  $M$  is realized by the formulas

$$s_g(M) = \langle u|M|u\rangle = Tr(UM),$$

the diagonal sum  $s_c(M)$  corresponds to

$$s_c(M) = \sum_i \langle i|M|i\rangle = Tr(IM)$$

and the quiescent sum, assuming that  $q$  is the quiescent state, is given by

$$s_q(M) = \langle q|M|q\rangle = Tr(|q\rangle\langle q|M).$$

The trace formulas are especially convenient.

The average number of ancestors of a configuration of length  $n$  can be obtained from summing all the possible products of matrices representing sequences of cells forming the configuration. Such a collection is simply the result of multiplying out the ordered (because of matrix noncommutativity) product

$$\begin{aligned} M_n &= k^{-n} \sum_i \prod_{j=1}^n N(i_j) \\ &= k^{-n} \left\{ \sum_i N(i) \right\}^n \end{aligned} \tag{15}$$



Given that the de Bruijn matrix  $D_{kl}$  for  $k$  states and  $l$  stages ( $l = 2r$  by our conventions) satisfies the minimal equation

$$D_{kl}^{l+1} = kD_{kl}^l$$

and additionally that  $D_{kl}^l = U$ , from which

$$D_{kl}^{l+m} = k^m U$$

we have the eminently credible results that independently of  $n$ ,

$$\begin{aligned} s_g(M_n) &= (k^{2r})^2 && \text{(general)} \\ s_c(M_n) &= k^{2r} && \text{(cyclic)} \\ s_q(M_n) &= 1 && \text{(quiescent)}. \end{aligned}$$

## 6.2 Second moment

The formulas for higher moments are more interesting. If  $R$  is a certain product of matrices,  $G$  the template for admissible boundary conditions, then the square of the number of ancestors represented by  $R$  is

$$(Tr(GR))^2.$$

However the relation

$$Tr(A)Tr(B) = Tr(A \otimes B),$$

( $\otimes$  designates the tensor product [12]), yields

$$Tr(GR \otimes GR).$$

Identities for the tensor product allow transforming this expression into

$$Tr((G \otimes G)(R \otimes R)),$$

from which the first term can often be extracted as a constant factor from a sum of traces. If the sum in Eq. 15 is now carried out for all values of  $R$ , it is possible to recognize the expansion of

$$N = \sum_{i=0}^{k-1} N(i) \otimes N(i) \tag{16}$$

This sum has to be evaluated in each particular case; the simple closed form of the general average was due to an algebraic identity which does not apply to the higher moments. Nevertheless, even when it has to be analyzed numerically, it constitutes one single matrix encompassing the behavior of all the ancestors. Particularly interesting is the fact that it is just the connectivity matrix of the pair diagram introduced in Sec 3.4.

### 6.3 Higher moments

It is not hard to establish the general formula for the third moment,

$$m_3 = \sum_{i=0}^{k-1} N(i) \otimes N(i) \otimes N(i) \quad (17)$$

nor generally the analogous  $n^{\text{th}}$  moment in terms of the  $n^{\text{th}}$  tensor power sum. However, the first two moments are the most important in view of the role of the average and the variance in statistical analyses. Specifically, configurations of zero variance in their number of ancestors are candidates for reversible rules.

## 7 Moments and frequencies

A distribution of data is characterized by its collection of moments as well as by the individual frequencies with which the data occur; these are really two complementary aspects of the same data set. Here, frequencies  $n(i)$  are associated with the nonnegative integers  $i$ , namely the number of configurations with a given number of ancestors. The  $k^{\text{th}}$  moment of this data is defined by

$$m_k = \sum_{i=0}^r n(i) i^k, \quad (18)$$

where it is understood that  $0^0 = 1$ .

The second volume, second edition of William Feller's popular book on the theory of probability [31] contains some material on moments; an advanced theory can be found in the American Mathematical Society *Survey* written by J. A. Shohat and J. D. Tamarkin [32].

Regarding the right hand side of Eq. 18 as the inner product of two vectors, then displaying the equations for individual moments, reveals a matrix equation

$$\begin{bmatrix} m_0 \\ m_1 \\ m_2 \\ \vdots \\ m_r \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 0 & 1 & 2 & \dots & r \\ 0 & 1 & 4 & \dots & r^2 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 1 & 2^r & \dots & r^r \end{bmatrix} \begin{bmatrix} n(0) \\ n(1) \\ n(2) \\ \vdots \\ n(r) \end{bmatrix}$$

expressing the moments in terms of the frequencies. One desires a reversed relation, expressing frequencies using moments, the more so supposing that the moments could be calculated independently. Naturally this requires the determination of as many moments as frequencies, in order to produce a square matrix for inversion.

The matrix shown is a particular case of a type which is very well known in mathematics, the Vandermonde matrix. Its general form is

$$V = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_0 & x_1 & x_2 & \dots & x_r \\ x_0^2 & x_1^2 & x_2^2 & \dots & x_r^2 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ x_0^r & x_1^r & x_2^r & \dots & x_r^r \end{bmatrix},$$

readily invertible through the use of the Lagrange interpolation polynomials  $\pi_i(x)$ , defined by

$$\begin{aligned} \pi(x) &= \prod_{j=0}^r (x - x_j) \\ \pi_i(x) &= \frac{\pi(x)}{(x - x_i)\pi'(x_i)} \\ &= \sum_{j=0}^r c_{ij}x^j. \end{aligned} \tag{19}$$

The inverse matrix transforms moments into frequencies:

$$\begin{bmatrix} n(0) \\ n(1) \\ n(2) \\ \cdot \\ n(r) \end{bmatrix} = \begin{bmatrix} c_{00} & c_{01} & c_{02} & \dots & c_{0r} \\ c_{10} & c_{11} & c_{12} & \dots & c_{1r} \\ c_{20} & c_{21} & c_{22} & \dots & c_{2r} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ c_{r0} & c_{r1} & c_{r2} & \dots & c_{rr} \end{bmatrix} \begin{bmatrix} m_0 \\ m_1 \\ m_2 \\ \cdot \\ m_r \end{bmatrix}. \tag{20}$$

It is important to appreciate the mechanism by which the product  $V^{-1}V$  yields the identity matrix. There are  $r + 1$  interpolation polynomials to evaluate at  $r + 1$  points; each polynomial vanishes at all the “wrong” points, while assuming unit value at the “right” point, namely the one for which the subscripts match. The matrix  $V^{-1}$  contains rows of coefficients for such polynomials; in the matrix product they form inner products with the columns of  $V$ , composed of all the powers needed to evaluate the polynomial at some particular point. The resultant interplay of right and wrong creates the unit matrix.

Because the ancestral data is all evenly spaced, the interpolation polynomials specialize to the Newton polynomials  $\nu_i(x)$ ,

$$\begin{aligned} \nu(x) &= \prod_{j=0}^r (x - j) \\ &= \sum_{j=0}^{r+1} S_{r+1}^{(j)} x^j \end{aligned} \tag{21}$$

$$\nu_i(x) = \frac{(-1)^{r-i}\nu(x)}{i!(r-i)!(x-i)},$$

whose characteristics have been nicely described, for example in Berezin and Zhidkov's numerical analysis text [33]. The primary polynomial  $\nu(x)$  is symmetric or antisymmetric with respect to  $r/2$  according to its parity, oscillating with diminishing amplitude from the origin to the midpoint, and growing rapidly outside the interval  $(0, r)$ .

The coefficient  $S_{r+1}^{(j)}$  in Eq. 21 is a Stirling number of the first kind, as described in Abramowitz and Stegun's mathematical handbook [34]. The explicit formula cited therein is fairly long; asymptotic formulas, some special values and tables are also presented.

Since we have already seen how to express the moments of the counterimage distribution in terms of tensor powers of the de Bruijn fragments, we have an alternative viewpoint from which to examine the frequency of counterimages.

An important special case is that of zero variance, in which all data equal  $x$ , say; in that case one would have  $m_k = (r+1)x^k$ , for which the inversion formula Eq. 20 would yield  $n(i) = (r+1)\pi_i(x)$ . Supposedly  $x$  would be an integer, so that  $x$ 's actual frequency  $r+1$  would be recovered, all other frequencies vanishing. Nevertheless the formula is valid for any sequence of moments which are powers of  $x$ , such as might arise from an eigenvalue of one of the tensor power sums.

In practice, the high moments of any distribution are dominated by the largest datum; accordingly we might expect the inversion of a moment distribution to involve the maximal frequency plus a tail composed of the lower moments weighted by coefficients from the upper left hand corner of the matrix  $V^{-1}$ . These coefficients approach limiting values for large values of  $r$ , which would be the case of principal interest; moreover they diminish factorially while the moments increase as powers, ensuring convergence. For example, combining Equations 19 and 21 yields the frequency  $n(0)$  corresponding to the Garden of Eden,

$$n(0) = \frac{(-1)^{r-1}}{(r-1)!} \sum_{j=0}^r S_{r+1}^{(j+1)} m_j. \quad (22)$$

We have just seen that  $n(0)$  vanishes (no Garden of Eden) for distributions with zero variance; the converse would be that it *only* vanishes for zero variance, a conclusion whose likelihood is greatly enhanced by the existence and appearance of Eq. 22.

## 8 Some illustrative examples

Having succeeded in finding one single matrix for each automaton which summarizes important properties of the distribution of its ancestors, it is worthwhile

examining some typical cases; because the  $16 \times 16$  matrices of  $(2,1)$  automata are unwieldy, it is convenient to make the same points with  $(2,1/2)$  automata. We first consider an automaton with a Garden of Eden and configurations with arbitrarily large numbers of ancestors (or), then one which has no Garden of Eden but which is not reversible (exclusive or), and finally one which is reversible (right shift). In due course we shall encounter nontrivial reversible automata.

## 8.1 Rule 14

### 8.1.1 Matrices.

Consider Rule 14, boolean or, amongst the  $(2,1/2)$  automata. The matrices  $N(i)$  become

$$N(0) = \begin{bmatrix} 1 & \cdot \\ \cdot & \cdot \end{bmatrix} \text{ and } N(1) = \begin{bmatrix} \cdot & 1 \\ 1 & 1 \end{bmatrix}.$$

Their tensor squares are

$$\otimes^2 N(0) = \begin{bmatrix} 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} \text{ and } \otimes^2 N(1) = \begin{bmatrix} \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & 1 & 1 \\ \cdot & 1 & \cdot & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix};$$

consequently

$$N = \begin{bmatrix} 1 & \cdot & \cdot & 1 \\ \cdot & \cdot & 1 & 1 \\ \cdot & 1 & \cdot & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

### 8.1.2 Eigenvalues.

Evidently this rule gives 0 one ancestor, 1 three. Their squares are 1 and 9, visible respectively as the number of nonzero elements in the individual tensor products. The element sum of  $N$  is 10; its eigenvalues are  $\pm 1$ ,  $1 \pm \sqrt{3}$ , contrasted to eigenvalues 0, 1 for  $N(0)$  and  $(1 \pm \sqrt{5})/2$  for  $N(1)$ .

The  $2^n$  configurations of length  $n$  will have  $2^{n+2}$  ancestors for an average of 4 each, while the sum of the squares of the number of ancestors will eventually grow according to  $2.732^n$ . The growth could be as small as a factor of 2 per cell or as large as a factor of 4, according to whether all configurations have an equal number of ancestors, or all ancestors map into a single configuration.

The quiescent state will have asymptotically  $1.57^n$  ancestors (whose square is  $2.46^n$ ), an increasingly negligible proportion.

Because the mean is always constant and small, the variance,  $m_0^2 \sigma^2 = m_0 m_2 - m_1^2$ , will grow asymptotically at half (square root) the rate of the second moment, or by  $(1.18)^n$ , or 18% per additional cell.

### 8.1.3 Frequencies.

An average splits any set of data; half is less than average, half greater. The frequency distribution for ancestors is uniformly positive with a mean which is always quite small in relation to the range of the data, leaving the frequency distribution highly skewed. The following tabulation of the 512 ancestors of the 256 configurations of length 8 illustrates the point:

ancestors	number of instances	weight	comment
$i$	$n(i)$	$n(i)(i - av)$	
0	142	-284	
1	39	-39	
2	28	0	average
3	16	16	
4	4	8	
5	7	21	
6	2	8	
8	4	24	
9	1	7	
10	2	16	
13	4	44	
15	2	26	
16	2	28	
21	2	38	
89	1	87	
	256	0	total

At one extreme, there is a large number of ancestorless configurations, for example all those containing the string 010 somewhere amongst the eight cells; the other end boasts an outlier - 1\* - with 89 ancestors. Indeed, some 5% of the configurations account for half the ancestors.

The lesson to be learned is that even a few points on the high end must be compensated by numerous points on the low end, the very lowest of which come from the Garden of Eden.

### 8.1.4 Moments.

To avoid the large matrices of Sec. 7, consider just the sixteen ancestors of the eight configurations of three cells, for which in principle a  $17 \times 17$  matrix is required. There are only four nonzero frequencies, shown in the following variant of the previous table:

$i$	$n(i)$
0	1
1	4
2	2
8	1

which gives rise to the following moments (the residuals remain when the powers of the dominant frequency are subtracted; the weights, to be used in Eq. 22, were approximated from Abramowitz and Stegun [34]):

moment	value	residual	$c_{0j}$	$c_{0j}m_j$
$m_0$	8	7	1.000	7
$m_1$	16	8	-3.381	-27
$m_2$	76	12	4.923	59
$m_3$	532	20	-4.159	-83
$m_4$	...	36	2.311	83
$m_5$	...	68	-0.902	-61
$m_6$	...	132	0.256	34
$m_7$	...	260	-0.054	-14
$m_8$	...	516	0.008	4
$m_9$	...	1028	-0.001	-1

Note how the residuals are eventually dominated by the second largest datum; of course the phenomenon would repeat if an attempt were made to separate more and more of the largest data from the rest. Even in this restricted example, at least ten moments are required to approximate  $n(0)$  to the dubious accuracy of  $n(0) = 1$ , so the procedure may not be very practical. Nevertheless it is quite possible in theory, serving as a criterion whether  $n(0)$  vanishes or not, and suggesting that such vanishing might be unlikely unless the variance were exactly zero.

## 8.2 Rule 6

Two further examples are illustrative. First consider Rule 6, evolving via exclusive or.

### 8.2.1 Matrices.

The de Bruijn fragments for Rule 6 are doubly stochastic, which endows them with nice combinatorial properties:

$$N(0) = \begin{bmatrix} 1 & \cdot \\ \cdot & 1 \end{bmatrix} \text{ and } N(1) = \begin{bmatrix} \cdot & 1 \\ 1 & \cdot \end{bmatrix}.$$

The tensor squares are now

$$\otimes^2 N(0) = \begin{bmatrix} 1 & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & 1 \end{bmatrix} \text{ and } \otimes^2 N(1) = \begin{bmatrix} \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & 1 & \cdot \\ \cdot & 1 & \cdot & \cdot \\ 1 & \cdot & \cdot & \cdot \end{bmatrix};$$

with

$$N = \begin{bmatrix} 1 & \cdot & \cdot & 1 \\ \cdot & 1 & 1 & \cdot \\ \cdot & 1 & 1 & \cdot \\ 1 & \cdot & \cdot & 1 \end{bmatrix}.$$

### 8.2.2 Eigenvalues.

Each state has two ancestors, the square of this number is 4;  $N$  has the minimum sum possible which is 8. Eigenvalues are readily calculated;  $N(0)$  is a unit matrix, as is its tensor square which commutes with  $N(1) \otimes N(1)$ .  $N(1)$  has eigenvalues  $\pm 1$ ; its tensor square their products in various combinations, so that altogether  $N$  has eigenvalues 0 and 2; indeed it is essentially (projectively) idempotent since  $N^n = 2^{n-1}N$ . This is the least rate of increase possible for any  $(2, 1/2)$  rule, conforming to the fact that every configuration of Rule 6 has exactly two ancestors.

### 8.2.3 Moments.

Since every configuration has two ancestors, the moment distribution is concentrated on that value, and all the higher moments are powers of the first moment. In particular, the distribution has zero variance.

## 8.3 Rule 12

Finally consider Rule 12, right shift, which intuitively should be a reversible rule.

### 8.3.1 Matrices.

The de Bruijn fragments for Rule 12 are stochastic, but not doubly stochastic.

$$N(0) = \begin{bmatrix} 1 & 1 \\ \cdot & \cdot \end{bmatrix} \text{ and } N(1) = \begin{bmatrix} \cdot & \cdot \\ 1 & 1 \end{bmatrix}.$$

Their tensor squares are

$$\otimes^2 N(0) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} \text{ and } \otimes^2 N(1) = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & 1 & 1 \end{bmatrix};$$

consequently

$$N = \begin{bmatrix} 1 & 1 & 1 & 1 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$



### 8.3.2 Eigenvalues.

Again there are two ancestors per state, giving squares of 4 and the minimal element count of 8 for  $N$ . This time  $N(0)$  and  $N(1)$  are idempotents with eigenvalues 0 and 1 but they do not commute.  $N$  is nevertheless “idempotent” according to  $N^2 = 2N$ ; its element sum is 8 with powers growing at the minimal rate.

The difference is that Rule 12 is reversible, via a left shift, whereas Rule 6 is not reversible at all, each different value of a boundary cell resulting in a completely different ancestor (or no ancestor at all if the left and right boundaries conflict). Rule 12 is indifferent to its right boundary, but the ancestor is uniquely determined by the left boundary. Yet both rules are surjective.

Both these rules have a pair matrix with eight 1’s, the minimal number possible; uniformly distributed throughout the rows and columns for Rule 6, but not for Rule 12. The difference is crucial, and is related to the observation that rules of the Fredkin type are all constructed from groups of variables, some of which are superfluous. Right shift does not depend upon the right cell, so the matrices  $N(i)$  (and their tensor squares) have constant rows, bunching the 1’s in a way that does not occur for Rule 6.

### 8.3.3 Moments.

The moment distribution is the same as for Rule 6, also with zero variance.

## 9 A statistical Gerschgorin theorem

With the knowledge that the moments, especially the average and variance, of the distribution of ancestors of a cellular automaton can be gotten from the powers of readily obtained matrices, attention naturally turns to the matrices themselves. Powers, not to mention other properties, are most readily obtained from the eigenvalues and eigenvectors of any matrix, especially from the eigenvalue of largest absolute value.

### 9.1 Gerschgorin’s theorem

Gerschgorin’s theorem [13], one of the most serviceable estimates for the eigenvalues of a matrix, asserts that the largest row sum of absolute values of the matrix elements forms an upper bound for the absolute values of the eigenvalues. It results from applying the triangle inequality to the definition,

$$MX = \lambda X$$

once it has been written out in terms of the matrix elements,

$$\lambda x_i = \sum_j M_{ij} x_j;$$

namely

$$\lambda \leq \sum_j |M_{ij}| \frac{x_j}{x_i}.$$

Since not all the components of an eigenvector are zero and at least one of them is larger than (or at most equal to) the others, any inequality in which the largest component appears in the denominator can be further improved by increasing all the ratios to 1, leaving

$$\lambda \leq \sum_j |M_{ij}|.$$

Inasmuch as the identity of the largest component is usually unknown, the worst case has to suffice, giving the result which is usually quoted. Sometimes the diagonal element is associated with the eigenvalue, leading to a collection of disks centered on the diagonal elements whose radii are gotten from summing the off diagonal elements in the row.

## 9.2 Positive matrices

Further information is available if all the elements of  $M$  are positive, or if none is negative with an assurance of purely positive components for the eigenvector; then the minimum component can be used to reverse the inequality and reveal the least row sum as a lower bound to the eigenvalue. So it will be for the largest eigenvector of an irreducible nonnegative matrix, according to the classical theory of Perron and Frobenius [14].

Statistical information concerning the eigenvalue can be obtained from the equation just as easily as the traditional bounds; for example, suppose that the equations for each component are summed:

$$\lambda \sum_i x_i = \sum_{ij} M_{ij} x_j.$$

With the definitions

$$\xi_j = \frac{x_j}{\sum_i x_i}$$

and

$$C_j = \sum_i M_{ij}$$

one obtains

$$\lambda = \sum_i C_i \xi_i,$$

provided that the sum of the eigenvector components is not zero; again it is a foregone conclusion that the eigenvector of the largest eigenvalue of a nonnegative matrix can be so chosen.

Inasmuch as the right hand side has the form of an inner product, as long as the quantities involved are real, there is an angle  $\theta$  and vectors  $\xi$  and  $\Xi$  for which

$$\lambda = |\xi| |\Xi| \cos \theta.$$

Even if complex quantities were involved, there would only be a lost phase, which could be restored. In such an equation  $\lambda$  is circumscribed by the variances of the set of column sums, and of the eigenvector components.

### 9.3 The statistical theorem

Supposing that the required means and variances have been calculated, we could write

$$C_j = \gamma + c_j \tag{23}$$

$$\xi_j = \frac{1}{n} + u_j \tag{24}$$

$$\lambda = \gamma + n\sigma_c\sigma_\xi \cos \vartheta. \tag{25}$$

From our point of view, the interesting result is that

$$\gamma = \frac{1}{n} \sum_{i,j=1}^n M_{ij}, \tag{26}$$

and that deviations of  $\lambda$  from this value are expressible in terms of the variances, both of the column sums and the eigenvector components.

The factor  $n$  in Eq. 25 may be disconcertingly large unless it is compensated by considerable uniformity in the column sums and/or in the components of the eigenvector. When either, and especially both, are small, the eigenvalues are estimated by the element sum of the matrix. Since similar estimates apply to *all* eigenvalues (with suitable modifications for complex eigenvectors whose components do not sum to zero), it can be seen that the orientation of the eigenvector relative to the vector of column sums distinguishes between the different eigenvalues.

### 9.4 Examples

In practice it seems that  $\sigma_c$  is a large fraction of  $\gamma$  and that  $\sigma_\xi$  is on the order of the average  $\bar{\xi} = 1/n$ ; thus  $\gamma$  approximates  $\lambda$  to  $\log n$  decimals; the sign of the correction seems to be positive when the relative number of ancestors is fairly uniform, but negative when there is a large preponderance of ancestors for some of the states.

#### 9.4.1 (2,1/2) automata.

As an example, consider the following table of results for (2,1/2) automata, where  $a$  and  $b$  ( $a + b = 4$ ) are the respective numbers of neighborhoods evolving into 0 or 1 ( $\gamma = (a^2 + b^2)/4$ ); the first three columns show empirical results.

ab	min	max	ave	$\gamma$
04	4.000	4.000	4.000	4.000
13	2.533	2.737	2.635	2.500
22	2.000	2.000	2.000	2.000

#### 9.4.2 (2,1) automata.

For (2,1) automata,  $a + b = 8$ , while  $(k^{2r})^2 = 16$  and  $\gamma = (a^2 + b^2)/16$ .

ab	min	max	ave	$\gamma$
08	4.000	4.000	4.000	4.000
17	2.938	3.381	3.102	3.125
26	2.400	2.800	2.580	2.500
35	2.132	2.695	2.253	2.125
44	2.000	2.480	2.155	2.000

#### 9.4.3 (3,1/2) automata.

As a further example, (3,1/2) automata, yield the following data, where  $a$ ,  $b$ , and  $c$  (for which  $a + b + c = 9$ ) are the respective numbers of neighborhoods evolving into 0, 1, or 2;  $(k^{2r})^2 = 9$  with  $\gamma = (a^2 + b^2 + c^2)/9$ .

abc	min	max	ave	$\gamma$
009	9.000	9.000	9.000	9.000
018	7.047	7.519	7.204	7.222
027	5.747	6.110	5.949	5.889
036	5.000	5.531	5.107	5.000
045	4.562	5.266	4.676	4.556
117	5.095	6.002	5.704	5.667
126	4.362	5.283	4.719	4.556
135	3.813	4.615	4.121	3.889
144	3.707	4.854	3.920	3.667
225	3.707	4.854	3.970	3.667
234	3.259	5.000	3.602	3.222
333	3.000	4.002	3.430	3.000

### 9.5 Commentary

The trend seems to be favorable; generally the average value approximates the nominal value, but there are distortions at both ends of the range which can

be attributed to the influence of the bounds to which the tensor squares are subject, but which would not affect a more random collection of matrices.

In order to have  $\lambda = \gamma$  in Eq. 25 it is necessary that  $\sigma_c = 0$ ,  $\sigma_\xi = 0$ , or  $\cos\vartheta = 0$ . In the first case all the column sums will be equal, equal to the eigenvalue, and the matrix will be stochastic. In the second case the vector  $|u\rangle$  will be an eigenvector, the row sums will all be constant, and again the matrix will be stochastic. The final alternative is more interesting, requiring the residuals in the column sums to be orthogonal to the residuals in the eigenvector, for a matrix which is explicitly not stochastic.

## 10 General considerations

Both general properties of positive matrices and general statistical considerations favor sharper results than the ancestral distributions taken from de Bruijn matrices seem to justify. The reasons lie in the difference between positive matrices for which strict conclusions are possible, and nonnegative matrices, for which they *may* hold; and also in the strongly asymmetrical frequency distribution.

So, in spite of the encouragement given by matrix theory and Eq. 22, it is necessary to look elsewhere for actual proofs. An ideal theoretic result, that is, one which depends on collective properties and not on individual matrices themselves, has been published by Michel Dubois-Violette and Alain Rouet [35]; similar reasoning had already been used by Hedlund [7], was recounted by Nasu [8], and was probably the basis of Amoroso and Patt's assertion [22] regarding the balancing of counterimages (they cited an unpublished laboratory report).

Nasu has shown [36] that the arguments, ostensibly valid for de Bruijn diagrams, actually apply to any graphs for which the numbers of incoming and outgoing nodes are themselves balanced. In all cases, the important connection lies between the variance and  $n(0)$ ; otherwise understood as the fact that when every configuration has at least one counterimage, the number of counterimages is uniform.

### 10.1 The de Bruijn fragments

Once there are matrices from which the properties of counterimages may be deduced, the deduction usually takes the form of ascertaining eigenvalues, and occasionally, eigenvectors. For example, according to whether the largest eigenvalue  $\lambda_i$  of  $N(i)$  is greater than, equal, or less than 1, the number of counterimages will inevitably increase, remain stable, or decrease. As it lies in the range from 1 to  $k$ , the rate of increase will range from modest to drastic.

The statistical Gerschgorin theorem, Eq. 26, estimates  $\lambda_i$  as  $|\varphi^{-1}(i)|/k^{2r}$  (— — denotes cardinality), which would yield a value of 1 only for balanced counterimages. Although this estimate is subject to considerable variation, it

gives a good indication of the degree to which a configuration profits from having cells with numerous ancestral neighborhoods.

Setting aside considerations of eigenvalues for the moment, consider the sum  $s_g$ ; its value for each matrix or product of matrices is a nonnegative integer. Its least value  $m$  surely satisfies

$$m \leq \min_{i=0}^{k-1} \langle u | N(i) | u \rangle$$

and could well be as small as zero — the Garden of Eden case.

Consider further a configuration whose cells were in states  $\{i_1, i_2, \dots, i_s\}$ , the number of whose ancestors would therefore be  $m = \langle u | P | u \rangle$  for

$$P = N(i_1)N(i_2) \dots N(i_s).$$

The extended configuration produced by adding one additional cell at the right in state  $i$  will have  $m = \langle u | PN(i) | u \rangle$  ancestors; then the expected number of ancestors averaged over all such extensions is

$$\begin{aligned} \frac{1}{k} \sum_{i=0}^{k-1} \langle u | PN(i) | u \rangle &= \frac{1}{k} \langle u | P \left( \sum_{i=0}^{k-1} N(i) \right) | u \rangle, \\ &= \frac{1}{k} \langle u | PD | u \rangle \\ &= m \end{aligned}$$

because the  $N(i)$  sum up to the de Bruijn matrix, whose eigenvector  $u$  has the eigenvalue  $k$ .

At this point, note that averages lie between the maximum and minimum of their data, reaching either extreme only if all the data are equal. This is the first part of Dubois-Violette and Rouet's [35] result. Such being the present situation, any single (and by induction, finite) extension of a minimal ancestor configuration also has a minimal number of ancestors.

In the worst case, an extended Garden of Eden is still a Garden of Eden, an observation already clear from the subset diagram, not to mention the consequence of multiplying by a zero matrix. The main point of this new result is that it is not possible to gain ancestors by taking longer configurations, then suddenly lose them; at least not for minimal configurations within a de Bruijn diagram. For others, counterexamples exist.

Although the proof shown is right sided, the left sided version is entirely similar. The proof is also completely symmetrical between maxima and minima, with the difference that a minimum is assured whereas a maximum is not.

The foregoing analysis then establishes the set  $A_{\min}$  consisting of all  $P$  for which  $s_g(P)$  takes the minimum value as a two sided ideal. By this we simply mean that

$$P \in A_{\min} \Rightarrow s_q(PQ) = s_g(QP) = m.$$

If  $A_{\max}$  exists, it too must be a two sided ideal; from

$$P \in A_{\min}, \quad Q \in A_{\max}$$

and taking  $n$  as the maximum number of counterimages follows

$$m = g_{\max}(PQ) = n.$$

The existence of ideals makes it hard to find matrices which are not absorbed into the ideal because those lying outside the ideal can only be formed from certain products of  $N(i)$ ; of course when  $m = 0$  the subset diagram provides a map showing just which matrices to choose and which to avoid. Furthermore, multiplication by zero is a drastic action whose role in defining ideals can be readily understood, but one might conjecture that those are the only kinds of ideals that there are. In other words, ideals depending on  $m > 0$  would have to be null or else encompass the whole space.

## 10.2 The second moment

To obtain an ancestor census directly, it necessary to deal with  $k$  different matrices  $N(i)$  together with all their possible products; their compliance with the general properties of nonnegative matrices is frequently complicated by their reducibility. Fortunately the single matrix  $N$  characterizing the second moment subsumes the properties of all these individual matrices and their products; discovering its general properties avoids most of the more detailed calculations, although its own reducibility is still a possibility.

The basic conclusion that the eigenvalues of  $N$  lie in the range from  $k^{2r}$  to  $k^{2r^2}$  is quite easily established.

The lower limit follows from the fact that the diagram of identical pairs,  $\{(a, a)\}$  is a subdiagram of the diagram of arbitrary pairs,  $\{(a, b)\}$ . The de Bruijn matrix of the former, whose maximum eigenvalue is  $k^{2r}$ , is a principal submatrix of the pair matrix, whose maximal eigenvalue must therefore be equal to or greater than  $k^{2r}$ . Since the lower limit corresponds to zero variance, it can only be reached when each of the matrices  $N(i)$  contains exactly  $k^{2r}$  links (and thus  $N$  contains exactly  $k(k^{2r})^2$  links); a necessary but not sufficient condition.

The upper limit follows from Gerschgorin's theorem, taking into account the fact that there are at most  $k^2$  incoming (or outgoing) links from any node in the pair diagram. However, if each node has that many links, all neighborhoods must evolve into the same state, so that the upper limit is associated with automata evolving to a constant field within one generation.

The tables of Sec. 9 confirm these limits, as well as showing a general distribution of eigenvalues reflecting the number of ways that the total number of states of any given cell can be partitioned amongst the individual states.

### 10.3 Zero variance

We have seen that if the number of ancestors of whatever configuration is bounded, all states have the same number of ancestors. Otherwise the number of ancestors would increase as the length of the configuration increased, in agreement with the rate implied by the largest eigenvalue of  $N$ . The possibility is still open that there is no upper limit but that the lower limit might not be so small as zero, implying that every configuration has at least one ancestor. This would create a peculiar but not impossible statistical situation wherein an increasing variance did not imply spreading of the data on both sides of the mean.

For automata, a contradiction can be deduced by confronting a configuration with many ancestors with one with minimal ancestors; many ancestors should lead to still more in a composite configuration, but they cannot because of minimality.

Thus, consider any matrix  $P$ , some matrix  $Q \in A_{\min}$ , and a certain abuse of notation with respect to the evolutionary function  $\varphi$ . Then if  $m > 0$  there are words  $\hat{P} \in \varphi^{-1}(P)$  and  $\hat{Q} \in \varphi^{-1}(Q)$  plus a matrix  $M$ , for which

$$\varphi(\hat{Q}\hat{P}) = QMP.$$

The idea of selecting a fixed  $Q$ , varying  $P$ , and exploiting the fact that  $s_g(QMP) = m$  will not work because a counterimage  $\varphi^{-1}(M)$  of the bridge  $M$  will not necessarily mesh with every  $\hat{P}$ . However if the left hand element of  $P$  is fixed, an accounting can be made. Therefore let

$$\begin{aligned} P &= N(i)\Pi, \\ \hat{N} &\in \varphi^{-1}(N(i)), \\ \hat{P} &= \hat{N}\rho. \end{aligned}$$

Once again

$$\varphi(\hat{Q}\hat{N}\rho) = QMP,$$

whereupon it may be concluded that

$$s_g(\varphi(\hat{N}\rho)) \leq m$$

on account of the fact that these are only some of the items making up the counterimages of the extension  $QMP$ . This inequality contemplates all the choices of  $\rho$  for a fixed  $\hat{N}$ , but there could be as many as  $k^{2r+1}$  choices of  $\hat{N}$ . All together, independently of  $P$ ,

$$s_g(P) \leq mk^{2r+1},$$

which is the upper bound necessary to conclude that whenever  $m > 0$  there are equal numbers of ancestors for all configurations — exactly  $k^{2r}$ . This is the second part of Dubois-Violette and Rouet's [35] result.



## 10.4 The essential result

Somewhat special arguments finally establish the relationship between zero variance and zero ancestors for automata, namely that the number of ancestors is either uniform for all configurations, or else that there are some configurations with no ancestors at all. As special cases, if the distribution of ancestors for a single cell is not balanced, there must be a Garden of Eden (but not conversely); and a reversible rule will lack a Garden of Eden (but not conversely).

Mathematically, evolution is called surjective if every configuration has an ancestor, injective if it has only one. The essential result states that surjectivity requires uniform numbers of ancestors, at first sight incompatible with injectivity and so with reversibility as well. Reversibility can only be reconciled with uniformity if there is an edge effect; that is if multiple ancestors differ only in a remote boundary layer. The simplest layer would reflect the overhang due to the greater size of neighborhoods relative to cells. Of course another source of injectivity would be to eliminate all but one of the alternatives, by requiring periodicity, or quiescence at infinity, for example.

Zero variance is equivalent to the maximum eigenvalue of  $N$  assuming its minimum value,  $k^{2r}$ . This in turn implies that the maximum eigenvalue of the matrices  $N(i)$  is 1, but not conversely. There are consequences in turn for the properties of the subset matrix and the pair matrix, which is the same as the second moment matrix.

## 11 Pairs of states

The second moment matrix is also the connectivity matrix for pairs of states discussed in Sec. 3.4, just as the third moment matrix corresponds to triples of states, and so on. No less than restrictions on possible pairs of sequences implied by zero variance, observation of the same limitations in practice can serve as an indicator of zero variance, and consequently of a surjective rule.

### 11.1 Partitioned pair matrix

The tensor product of the connectivity matrices of two diagrams is the connectivity matrix of their cartesian product, wherein pairs of nodes are linked according to whether both members of the pair are each linked in their own diagram. Sometimes a careful arrangement of the nodes in the tensor product will partially diagonalize the connectivity matrix, rendering some property of the arrangement more evident than it would otherwise have been.

A case in point consists in grouping all the pairs composed of equal elements together in a subset  $\mathcal{A}$ , and the rest into the subset  $\mathcal{B}$ . It then becomes evident whether links in a tensor square describe pairs of links in the original diagram which have a node in common by linking  $\mathcal{A}$  to  $\mathcal{B}$  or the reverse, both nodes in common by linking  $\mathcal{A}$  to  $\mathcal{A}$ , or neither by linking  $\mathcal{B}$  to  $\mathcal{B}$ .

Partition  $D \otimes D$  accordingly

$$D \otimes D = \begin{bmatrix} P & Q \\ R & S \end{bmatrix}$$

and consider the eigenvalue equation with a similarly partitioned eigenvector

$$\begin{bmatrix} P & Q \\ R & S \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}.$$

The following submatrix equations are the result:

$$\begin{aligned} Px + Qy &= \lambda x \\ Rx + Sy &= \lambda y. \end{aligned}$$

If  $\lambda$  is already an eigenvalue belonging to  $P$ , it follows that  $Qy = 0$ ; whether this requires that  $Q = 0$  depends somewhat on  $S$ , but the condition is quite sufficient. If  $\lambda$  is *not* an eigenvalue of  $S$ ,  $S - \lambda I$  is invertible whence  $y = -(S - \lambda I)^{-1}Rx$  and  $Q(S - \lambda I)^{-1}R$  would have to be singular unless  $y$  were 0.

Consideration of a left eigenvector leads to similar conclusions regarding  $\bar{y}R$ , while if  $\lambda$  is an eigenvalue of  $S$  as well as  $P$ ,  $Rx$  must vanish. Broadly speaking, if either of the diagonal submatrices already has the eigenvalue, it had better be decoupled from the other, but the details of the uncoupling can get messy.

In practice, another line of reasoning is slightly more compact. Consider

$$(D \otimes D)^2 = \begin{bmatrix} P^2 + QR & PQ + RS \\ RP + SR & RQ + S^2 \end{bmatrix}$$

If  $P$  were a de Bruijn matrix it would be an irreducible nonnegative matrix and  $P^2 + QR$  would have a maximum eigenvalue strictly larger than  $\lambda^2$  unless  $QR = 0$ ;  $Q = 0$  or  $R = 0$  would be sufficient. In any event, continuing shows that

$$(D \otimes D)^3 = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix};$$

with

$$\begin{aligned} m_{11} &= P^3 + QSR \\ m_{12} &= P^2Q + PRS + QS^2 \\ m_{21} &= RP^2 + SRP + S^2R \\ m_{22} &= RPQ + R^2S + SRQ + S^3 \end{aligned}$$

Gradually the series of requirements

$$\begin{aligned} QR &= 0, \\ QSR &= 0, \\ QS^2R &= 0, \\ &\dots \end{aligned}$$

develops, which can be summed to recover

$$Q(S - \lambda I)^{-1}R = 0. \tag{27}$$

The very slightly stronger result is due to the additional knowledge about  $P$ .

To illustrate this result with respect to the three rules of Sec. 8, rearrange the indices of the matrix  $N$  to group paired indices  $(a, a)$  and then complementary indices  $(a, \bar{a})$ ; we obtain

$$\begin{array}{c} \text{rule 14} \\ \left[ \begin{array}{cc|cc} 1 & 1 & . & . \\ 1 & 1 & 1 & 1 \\ \hline . & 1 & . & 1 \\ . & 1 & 1 & . \end{array} \right] \end{array} \quad \begin{array}{c} \text{rule 6} \\ \left[ \begin{array}{cc|cc} 1 & 1 & . & . \\ 1 & 1 & . & . \\ \hline . & . & 1 & 1 \\ . & . & 1 & 1 \end{array} \right] \end{array} \quad \begin{array}{c} \text{rule 12} \\ \left[ \begin{array}{cc|cc} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ \hline . & . & . & . \\ . & . & . & . \end{array} \right] \end{array} .$$

The off diagonal submatrices for Rule 14, for which  $QR \neq 0$ , force an eigenvalue larger than the minimum, whereas the other two matrices are partially diagonal and maintain the minimum value. Rule 6 is self-complementary, as evidenced by the equality of its diagonal submatrices  $P$  and  $S$ .

## 11.2 Symmetry by permuting states

As the use of the complement in creating these examples shows, a tensor square can be further refined if the basic diagram has an isomorphism, which is to say a permutation of its nodes resulting in the same connectivity matrix; for binary automata such an isomorphism results from complementing the individual cells. For  $k > 2$  there are increasing numbers of isomorphisms,  $k!$  in fact, whose presence can be detected in the full de Bruijn matrix and all the other moment matrices associated with it. Nevertheless if an individual rule lacks some or all of this symmetry, it will be missing from  $N$  and its analogues.

The pair matrix of Sec. 3.4 can be rearranged along these lines, which is to say that it is equivalent to the following matrix, gotten by listing first all the pairs of the form  $(a, a)$ , last all the pairs of the form  $(a, \hat{a})$ , which is the only other permutation of binary values. The remaining pairs are laid out symmetrically in the middle.

As before, the possible pair matrices all conform to a common pattern, which arises for either Rule 0 or Rule 255. Since the pair  $(a, b)$  links to the pair  $(c, d)$  only if  $a$  links to  $c$  and  $b$  to  $d$  via the same symbol, the pairs of pairs actually linked will vary from rule to rule. Furthermore there are some additional symmetry requirements for the formation of pairs: if  $(a, b)$  links to  $(c, d)$ , then  $(b, a)$  must link to  $(d, c)$  and a similar requirement due to transitivity.

With this new ordering, the pair matrix for Rule 22 becomes: (1's belong to the actual pair matrix, while \*'s indicate the extent of the maximal possible matrix)

$$N = \begin{bmatrix} 1 & 1 & \dots & * & \dots & \dots & \dots & * & \dots & \dots & \dots \\ \dots & \dots & 1 & 1 & \dots & \dots & \dots & * & * & \dots & \dots \\ 1 & 1 & \dots & * & \dots & \dots & \dots & \dots & \dots & * & \dots \\ \dots & \dots & 1 & 1 & \dots & \dots & 1 & 1 & \dots & \dots & \dots \\ \hline \dots & \dots & \dots & * & * & \dots & \dots & \dots & \dots & 1 & 1 & \dots \\ * & \dots & 1 & \dots & \dots & \dots & \dots & \dots & \dots & 1 & \dots & \dots \\ \dots & * & 1 & \dots & \dots & * & 1 & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & * & 1 & \dots & \dots & \dots & \dots & * & 1 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & 1 & 1 & \dots & \dots & * & * \\ \dots & * & 1 & \dots & \dots & 1 & * & \dots & \dots & \dots & \dots & \dots \\ * & \dots & 1 & \dots & \dots & \dots & \dots & \dots & 1 & \dots & \dots & \dots \\ \hline \dots & \dots & \dots & 1 & * & \dots & \dots & \dots & 1 & * & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & 1 & 1 & \dots & \dots & \dots & * & * \\ \dots & \dots & \dots & 1 & 1 & \dots & \dots & \dots & * & * & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & * & 1 & \dots & \dots & * & * & 1 \end{bmatrix}$$

The submatrix in the upper left hand corner is always present for whatever Rule (the lower right submatrix is filled out for Rule 150, just as it would be for any other rule symmetric by conjugation), no matter what other elements are deleted from  $N$  on account of the structure of  $N(0)$  and  $N(1)$ . Accordingly, no row sum can exceed 4, so by Gerschgorin's theorem, the largest eigenvalue of  $N$  will be bounded by 4. The limit can only be reached if all row sums are equal, thus only for Rules 0 and 255 (evolution into constants).

On the other hand the diagonal submatrices have a uniform row sum of 2, which must be their largest eigenvalue; alternatively the first and last are just de Bruijn matrices, from which the same result follows. By a theorem in Gantmacher [11], this establishes a lower limit for the greatest eigenvalue of  $N$ , which must exceed the eigenvalues of any principal minor.

Of course there is an especial interest in conditions in which this lowest limit is maintained; so far it has not been possible to characterize the matrix  $N$  sufficiently to decide other than in special cases or by numerical calculation. The general result of Sec. 10 is the best known, that the minimum eigenvalue which already implies surjectivity, implies uniform distribution. The converse is not valid, most uniform distributions lead to higher eigenvalues.

### 11.3 Characterizing surjective rules

Some consequences of zero variance are easily described. First and foremost, no path beginning in the subset  $\mathcal{A}$  can ever return, because that would violate the reasoning leading up to Eq. 27. In other words, if two ancestors of a configuration initially coincide — even for a single partial neighborhood — and later diverge, they can never recombine. Seen from another point of view, no product of the matrices  $N(i)$  can ever have a matrix element as large as 2 while the rule has zero variance.

A similar prohibition applies to a pair of isomorphic ancestors; it is only necessary to replace  $\mathcal{A}$  by the set  $\{(a, f(a))\}$ .

It is also necessary that the matrix  $S$  corresponding to the subset  $\mathcal{B}$  not have an eigenvalue exceeding that of a de Bruijn matrix, which can be checked in any particular case, but not as easily when it is a general proposition. Essentially, the requirement is that mixed ancestors not proliferate more rapidly than single ancestors as the length of a configuration increases, which would clearly be possible if sections of an ancestor could be freely substituted for one another.

## 12 The subset matrix

Whether an automaton has a Garden of Eden can be decided at once from its subset diagram, or the algebraic properties of its subset matrix. For most automata the empty set will be accessible from the full set (implying a Garden of Eden), and there will not be that much interest in the remaining structure of the diagram. Contrarily, for an automaton with zero variance and a surjective rule, the very fact that the empty set can be avoided introduces additional structure into the diagram. For example, what are the largest subsets which lead to the empty set? What are the smallest subsets reachable from the full set? What relationships exist between subsets with similar characteristics?

### 12.1 Maximal node diagram

Recall that nodes in the subset diagram are sets of nodes from the original diagram, linked whenever each node in the tail subset joins some node in the head subset; paths arise from successive links. Call nodes in the original diagram points to distinguish them from nodes in the subset diagram. In the subset diagram of a general rule, multiple links can emanate from a single point, multiple links can also converge to a point; indeed this is the reason subset diagrams were invented, to ensure simple links between subsets.

For a surjective rule, however, only one path can run between a given pair of endpoints, one of whose consequences can be seen by factorizing a path between two nodes. Let the path  $x$  run from nodes  $U$  to  $V$ , continuing on to node  $W_y$  via  $y$  (anticipating a variation of  $y$ ).

Count the paths between points due to  $xy$  in two ways: Take  $U = \{u\}$  as a unit class to eliminate ambiguity in the starting point, then either proceed directly to  $W_y = xy(U)$ , or interrupt the path with a stopover at  $V = x(U)$ . The essential point is that there can be only one interrupted path. Now vary  $y$ , supposing only that it is a path of length  $p$ :

$$\sum_y |W_y| = |V|k^p \tag{28}$$

This equation expresses an average, so that the argument applying to averages of extreme quantities can be used once again: if  $V$  is a node of maximal cardinality, all the terms of the average must assume the same value.

Paths leading from a given unit class will not all reach maximum nodes; those which do will not necessarily do so immediately. Finiteness of the diagrams ensures that there is a short path — shorter than the cardinality of the diagram — leading from any node to the maximal node accessible to it. Once a maximal node has been reached, further links will trace out paths confined to a class of nodes of the same cardinality.

Since the de Bruijn diagram itself is connected, there is only one maximal class, irrespective of the unit class from which it was derived. The reason is that one can find paths which alternate between maximal nodes in the subset diagram and individual points in the de Bruijn diagram, yielding a derivation of any maximal node from any point, and of one maximal node from any other; thus the maximal class is connected, and there are maximal subsets containing any given unit class.

Not all paths lead from unit classes to the maximal class; the fate of paths from other subsets may also vary. When the subset in question is larger (setwise) than maximal (in particular, the full set) we know that its eventual images must remain larger than maximal; their images do not necessarily have to lead into the maximal class, however.

## 12.2 Left and right indices

In any event, the maximal class with its links is an image of the de Bruijn diagram, containing all the same paths as the original, but possibly with fewer nodes. The cardinality of the maximal subsets has been called [7] the *index* of its Rule; strictly the quantity which has been described is a *right index*, reserving the term *left index* for the analogous quantity derived from the reversed subset diagram. The handedness evidently refers to the direction in which a configuration would be extended to get a longer configuration.

The two indices need not be equal. For instance, both indices of  $(2, 1/2)$  Rule 6 (exclusive or) are 1, but for Rule 12 (right shift) the left index is 1 while the right index is 2.

It is worthwhile understanding the significance of an index and the reason that the two can differ. Fundamentally, the index is the number of different partial neighborhoods which have to be grouped together to reinforce each other if an arbitrarily prescribed configuration is to have an ancestor. For the right shift of Rule 12, extension to the right is pointless if the leftmost cell cannot be arbitrarily chosen, a possibility assured by the index of 2. For left extension the rightmost cell is irrelevant because it will be replaced immediately; any choice will do and the index is 1.

Continuation in either direction from whatever initial cell is always possible for Rule 6, so both its indices are 1. The consequence, however, is that the ancestors resulting from such freedom are all essentially distinct. By contrast, Rule 12 has only one ancestor; the necessary flexibility to construct it must be retained in the set of partial neighborhoods made available at each stage.

## 13 Reversing a rule

So far, the rules which are candidates for reversibility are those for which there is no Garden of Eden; one cannot determine an ancestor for a configuration which has none, which precludes a rule which purports to do so. Among those which remain, the equidistribution rule applies. We have seen that there are environments which could restrict the number of ancestors, such as a quiescent infinity or a cyclic automaton. Still the examples constructed in Sec. 2 produce a unique inverse without reference to any boundary condition, which is a general alternative which must be considered. Evidently the apparent variety of ancestors is illusory; the mechanism must be investigated.

### 13.1 Two (2,1/2) examples

Some surjective rules are injective, others are not. As an example of the first kind, consider (2,1/2) Rule 6, and some of the products of ancestor matrices. Recalling the definition of  $s(m)_{ij}$  in Eq. 12, there is either a left or a right ancestor matrix; choosing the right handed version,  $s(m)_{ij} = \mathbf{j}$  when  $\varphi(i, j) = m$ ,  $\phi$  otherwise. Accordingly,

$$\begin{array}{cc} \overline{s(0)} & \overline{s(1)} \\ \left[ \begin{array}{c} \mathbf{0} \ . \\ \ . \ \mathbf{1} \end{array} \right] & \left[ \begin{array}{c} \ . \ \mathbf{1} \\ \mathbf{0} \ . \end{array} \right] \end{array}$$

The four pair products are

$$\begin{array}{cccc} \overline{s(00)} & \overline{s(01)} & \overline{s(10)} & \overline{s(11)} \\ \left[ \begin{array}{cc} \mathbf{00} & \ . \\ \ . & \mathbf{11} \end{array} \right] & \left[ \begin{array}{cc} \ . & \mathbf{01} \\ \mathbf{10} & \ . \end{array} \right] & \left[ \begin{array}{cc} \ . & \mathbf{11} \\ \mathbf{00} & \ . \end{array} \right] & \left[ \begin{array}{cc} \mathbf{10} & \ . \\ \ . & \mathbf{01} \end{array} \right] \end{array}$$

Here the detail of interest is the fact that neither the left nor the right member of the matrix elements in  $s(ij)$  is consistent, nor does any such consistency arise for any of the higher order products (note that row and column indices must also be taken into account when judging consistency). Consequently one concludes that this rule does not have a general purpose inverse.

Repeating the display for (2,1/2) Rule 12 produces

$$\begin{array}{cc} \overline{s(0)} & \overline{s(1)} \\ \left[ \begin{array}{c} \mathbf{0} \ \mathbf{1} \\ \ . \ . \end{array} \right] & \left[ \begin{array}{c} \ . \ . \\ \mathbf{0} \ \mathbf{1} \end{array} \right] \end{array}$$

$$\begin{array}{cccc} \overline{s(00)} & \overline{s(01)} & \overline{s(10)} & \overline{s(11)} \\ \left[ \begin{array}{cc} \mathbf{00} & \mathbf{01} \\ \ . & \ . \end{array} \right] & \left[ \begin{array}{cc} \mathbf{10} & \mathbf{11} \\ \ . & \ . \end{array} \right] & \left[ \begin{array}{cc} \ . & \mathbf{01} \\ \mathbf{00} & \mathbf{01} \end{array} \right] & \left[ \begin{array}{cc} \ . & \ . \\ \mathbf{10} & \mathbf{11} \end{array} \right] \end{array}$$

This time, there is a consistency in the left member of each ancestral pair, so that each evolved pair has a unique ancestor which can be used as the evolved

state of the inverse rule, and all this quite independently of boundary conditions. Note that once consistency is found, it persists for all longer products.

### 13.2 A (2,1) example

The only reversible rules for (2,1) automata are trivial, but nevertheless still instructive. Consider Rule 204, whose evolution is stasis,

$$a = D(0) = \begin{bmatrix} 1 & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}; \quad b = D(1) = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 \end{bmatrix}.$$

These matrices are not permutation matrices nor even stochastic, since neither row sums nor column sums are constant. Nevertheless they are idempotent and generate a semigroup of four matrices of similar appearance, namely  $\{a, ab, ba, b\}$  with the multiplication table

	<i>a</i>	<i>ab</i>	<i>ba</i>	<i>b</i>
<i>a</i>	<i>a</i>	<i>ab</i>	<i>a</i>	<i>ab</i>
<i>ab</i>	<i>a</i>	<i>ab</i>	<i>a</i>	<i>ab</i>
<i>ba</i>	<i>ba</i>	<i>b</i>	<i>ba</i>	<i>b</i>
<i>b</i>	<i>ba</i>	<i>b</i>	<i>ba</i>	<i>b</i>

Consequently, every product has four counterimages, seemingly contradicting the claimed uniqueness. However, we need to form new symbolic matrices, which show the ancestral cell rather than the evolved cell; this ancestral cell is the central cell of the ancestral neighborhood, which can be read off from the indices of the de Bruijn matrix.

For Rule 204, these matrices are

$$s(0) = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \mathbf{0} & \mathbf{0} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}, \quad s(1) = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \mathbf{1} & \mathbf{1} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \mathbf{1} & \mathbf{1} \end{bmatrix};$$

they already display the crucial property: every counterimage of 0 is 0, whilst every counterimage of 1 is 1. Therefore Rule 204 is invertible by a (2,0) automaton, which is not surprising since it is really a radius zero automaton itself. Trivial in the present instance, the property is once again: there exists some word length  $r$  for which every word in the product of  $r$   $\alpha$ 's and  $\beta$ 's has the same central symbol. For this to work  $r$  must be odd, but inversion could be combined with shifting so that shift inverses also exist.



### 13.3 A (2,1) counterexample

Symbolic matrices can also be used to prove that a rule is *not* invertible; consider Rule 30 once more:

$$s(0) = \begin{bmatrix} \cdot & \mathbf{0} & \cdot & \cdot \\ \cdot & \cdot & \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}; \quad s(1) = \begin{bmatrix} \mathbf{0} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \mathbf{0} & \cdot & \cdot \\ \cdot & \cdot & \mathbf{1} & \mathbf{1} \end{bmatrix}.$$

With little effort one can establish that

$$s(1^n) = \begin{bmatrix} \mathbf{0}^n & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \mathbf{1}^{n-1}\mathbf{0} & \mathbf{1}^n & \mathbf{1}^n \end{bmatrix}$$

for  $n > 1$ . Given that the configuration  $1^*$  will always have some ancestors in which 1 remains 1 and others in which 0 becomes 1, there is no neighborhood length for which the ancestral cell is uniquely determined. Therefore Rule 30 cannot be reversed, in spite of the fact that every configuration has a unique ancestor whenever its external environment is specified.

### 13.4 Test procedure

It is evidently possible to check the powers of more and more complex words to see whether there is a possibility of inversion; If  $0^n$  admits inversion for some particular values of  $n$ ,  $1^n$  can be tested, then  $(01)^n$ , and so on, always noting the least common multiple of the admissible values. Eventually enough combinations may be examined to reach a decision: some word is impossible to resolve, or enough words have been analyzed to know that there is a length of neighborhood for which the ancestor is unconditionally unique, and the inverse Rule can be read off according to the sequences of  $\alpha$ s and  $\beta$ s in which the unique values occur—e.g., if the middle letter in each element of the symbolic matrix  $\alpha\beta\beta$  is 0,  $\varphi(0,1,1) = 0$  for inverse evolutionary function  $\varphi$ .

Fortunately there is an upper bound to the lengths of words which must be tested, making eventuality a certainty. Consider the matrix product  $\sigma\xi\tau$  corresponding to a particular counterimage; given that the  $p$ th letter is the same for all the words in the full product, let  $\xi$  be the  $p$ th factor. Decompose  $\xi$  into a sum  $\xi = \gamma + \epsilon$  in which  $\gamma$  contains the symbols which will be lost, and  $\epsilon$  those which will survive;  $\sigma\gamma\tau$  must vanish leaving  $\sigma\epsilon\tau = \sigma\xi\tau$ . There may be symbols common to  $\gamma$  and  $\epsilon$ , but  $\epsilon$  must contain one and only one of them, 0 say.

We are interested in showing that an excessively long  $\sigma$  or  $\tau$  can be shortened; such a demonstration would be akin to shrinking a long path through the vector subset diagram by excising loops, whereby a loop free path would be minimal. But that is not the way to our proof, which is independent of the diagram.

Rather, note that our matrices are sparse, in their numerical version all matrix elements are zeroes or ones, with exactly  $k$  ones per matrix. All their possible products are postulated to have the same property; yet there can only be just so many such matrices of a given dimension.

Some of them are original de Bruijn fragments, others exist only as products, but each of them will have a product of shortest length generating it. Let  $N$  be the maximum length of the shortest products generating the multiplicative closure of the de Bruijn fragments. Now replace both  $\sigma$  and  $\tau$  by minimal equivalents through the process of converting them to numerical form, finding the minimal product, and forming the symbolic product all over again. This will naturally shorten and alter any ancestors being represented; call the new products  $\hat{\sigma}$  and  $\hat{\tau}$ .

The important property of these new chains is

$$\sigma\gamma\tau = 0 \quad \Rightarrow \quad \hat{\sigma}\gamma\hat{\tau} = 0,$$

since the vanishing of the products depends on the placement of their nonzero elements and not their value, even when it is symbolic. In turn new position  $\hat{p}$  is guaranteed for which every matrix element

$$\hat{\sigma}\xi\hat{\tau} = \hat{\sigma}\epsilon\hat{\tau}$$

has the same symbol for its  $\hat{p}$ th letter, all within a neighborhood of no greater length than  $2N + 1$ .

It may still happen that every ancestor has a position with a unique ancestor, but that the positions do not coincide. But then padding may be inserted to extend the neighborhood either to the right or to the left in such a way as to give them all a common centering. Should this be necessary one will simply have an inverse rule which is insensitive to the values of certain cells.

In practice none of this rearrangement has to be carried out; the possibility that it can be done simply shows that there is nothing in a very long or irregular neighborhood whose equivalent would not already have been encountered in a systematic sweep through short neighborhoods.

### 13.5 Use of the pair matrix

Most of these conclusions can be summarized by referring to the pair matrix once again. Nasu [8] outlines such a process, referring also to the concept of *definite automata* [37], which is explained at some length in Zvi Kohavi's textbook [38] on automata theory. He uses the unordered pair matrix, but the conclusions are the same.

If a path is started in the subset  $\mathcal{B}$  of distinct pairs, it may remain within that subset or it may wander into the subset  $\mathcal{A}$  of identical pairs; if it does so there must be no possibility of returning to the original starting point. Otherwise the

pair matrix would not have the minimum eigenvalue required for zero variance and surjectivity.

In the case that the path remains within  $\mathcal{B}$ , it must eventually close into a loop, demonstrating that there must exist configurations with two different ancestors. No matter that some paths may wander into  $\mathcal{A}$ ; the rule cannot be injective. The alternative, that no path originating in  $\mathcal{B}$  remains in  $\mathcal{B}$ , means that it is eventually trapped in  $\mathcal{A}$ . Subsequent return to  $\mathcal{B}$  would force a second visit to  $\mathcal{A}$ , in violation of zero variance. For such automata, cannot be two ancestors, distinct except for their initial and terminal segments. This is the true case of injectivity.

Rules which are not injective, but lack the full multiplicity which is theoretically possible can be detected by using a diagram of triples or higher multiples; whether there are subsets of degenerate multiples which trap the trajectories must be investigated.

## 14 Conclusions

There are two noteworthy aspects of reversibility. One is that there exists a very elaborate and technical theory of injective and surjective mappings which has not dealt extensively with particular examples or applications. On the other hand a variety of empirical methods have been discovered for producing reversible automata, without having had a very extensive theoretical foundation. It would be helpful to reconcile these two tendencies.

The other aspect is the extent to which the theory of reversible automata has been treated in isolation from other cellular automata. This has probably been due to the very special nature of the proofs required, which do not leave much latitude for exceptions or approximations. Nevertheless one doubts that a binary automaton with 1,024 neighborhoods, 511 of which evolve into 0 and 513 of which evolve into 1, is that much different from an automaton whose counterimages balance. In other words, there should be a statistical continuum accounting for nearly reversible rules, almost surjectivity, and the like.

### 14.1 Surjectivity

Surjectivity is the important concept; injectivity is subordinate and can be treated later. The differences between cyclic, quiescent, and general boundary conditions are likewise secondary. In the analysis which we have presented, the results depend upon the frequency distribution of counterimages, and consequently upon its moment problem. The moments can be estimated from fragments of the de Bruijn matrix, determined individually for each automaton.

There are two tantalizing aspects to the problem so formulated. If the matrices involved were irreducible and not just nonnegative, several strong conclusions would be available which have to be established by other means. Likewise

if more information about the moments existed, useful conclusions could be drawn about the existence of gaps or other nonuniformities in the frequency distribution.

In spite of the lack of convenience, the primordial connection between zero variance (balanced distribution of counterimages) and surjectivity (lack of a Garden of Eden) can be established by arguments which still make use of fundamental properties of the de Bruijn matrix. Moreover, the proofs are still quite relevant to the moment problem; for example the knowledge that the variance increases exponentially with the length of a configuration leaves zero variance as the only bounded possibility.

It may still be a surprise, even to those who have calculated numerous subset diagrams and perhaps read Amoroso and Patt's paper [22], that an unbalanced evolutionary rule necessarily has a Garden of Eden.

## 14.2 Injectivity

Once surjectivity has been established, there is no doubt that the uniform multiplicity of counterimages is precisely what would be expected from the size of the partial neighborhoods bounding a finite configuration. Ways exist through which the multiplicity may be rendered ineffective; one is to terminate the configuration with quiescent neighborhoods, another is to close them with a cyclic connection. Fixed boundaries and phase relations in the closure (e.g. the left boundary is the complement of the right boundary) are other alternatives which could be considered.

All the foregoing mechanisms involve selections from a naturally occurring set of boundaries, but there is also the possibility that the boundaries don't really matter. In other words, aside from a finite boundary layer which would be arbitrarily remote from the center of a very long configuration, the different ancestors could just happen to agree. Surely this is what is happening in shift rules or the identity map, and is clearly provided for in rules of Fredkin's type.

Boundaryless injectivity, or one might say, injectivity with a natural boundary, can be tested in one of the pair diagrams. Since the ordered pair diagram is coincidentally the second moment matrix, the condition of zero variance imposes a direct condition. Just as the empty set, implying a Garden of Eden, must be met in a limited number of steps (if at all) in the subset diagram, a pair of ancestors must either coalesce within the confines of the pair diagram or encounter some means of coexisting.

Apparently the decision to use a boundary condition should be postponed as long as possible; the eventual choice may be that much simpler.

## 14.3 Additional comments

The role of topology in treating the evolution of cellular automata is not entirely clear. Hedlund gives it a central role in his fundamental survey [7]; it is basic

to such further articles as those of D. Richardson [39] or Tadakazu Sato and Namio Honda [40], and has been used by Karel Culick II [41]. It does seem to be required for the satisfactory description of the limiting behavior of long time evolution [42, 43, 44].

On the other hand Nasu [8, 36, 45, 46] has been able to obtain extensive results without using topology, at least for one dimensional automata; the present paper likewise demonstrates that no topology is required. We are left with the question of whether there are significantly different results to be obtained from the use of topology or whether topology simply provides a convenient language to express results which could be derived independently. However, this is a topic for a future study, just as is the question of handling two dimensional automata via two dimensional de Bruijn diagrams.

There are also important questions about the applications and the philosophical implications of reversible automata, a preoccupation of Toffoli [20]. Two recent articles addressed to these points are those of Asher Peres [47] and Shinji Takesue [48].

## References

- [1] John von Neumann, *Theory of Self-reproducing Automata* (edited and completed by A. W. Burks), University of Illinois Press, 1966.
- [2] Edward F. Moore, "Machine models of self reproduction," *American Mathematical Society Proceedings of Symposia in Applied Mathematics* **14** 17-33 (1962).
- [3] Arthur W. Burks (editor), *Essays on Cellular Automata*, University of Illinois Press, Urbana, 1970.
- [4] Sven Skyum, "Confusion in the Garden of Eden," *Proceedings of the American Mathematical Society* **50** 332-336 (1975).
- [5] George D. Birkhoff, *Dynamical Systems*, American Mathematical Society, Providence, Rhode Island, 1927.
- [6] Walter Helbig Gottschalk and Gustav Arnold Hedlund, *Topological Dynamics*, American Mathematical Society, Providence, Rhode Island, 1955.
- [7] G. A. Hedlund, "Endomorphisms and automorphisms of the shift dynamical system," *Mathematical Systems Theory* **3** 320-375 (1969).
- [8] Masakazu Nasu, "Local Maps Inducing Surjective Global Maps of One Dimensional Tessellation Automata," *Mathematical Systems Theory* **11** 327-351 (1978).

- [9] Stephen Wolfram, "Statistical mechanics of cellular automata," *Reviews of Modern Physics* **55** 601-644 (1983).
- [10] Stephen Wolfram (Ed.), *Theory and Applications of Cellular Automata*, World Scientific Press, Singapore, 1986 (ISBN 9971-50-124-4 pbk).
- [11] F. R. Gantmacher, *The Theory of Matrices*, volume 2, Chelsea Publishing Company, New York, 1959.
- [12] Richard Bellman, *Introduction to Matrix Analysis*, McGraw-Hill Book Company, New York, 1960.
- [13] Richard S. Varga, *Matrix Iterative Analysis*, Prentice Hall, Inc, Englewood Cliffs, New Jersey, (1962).
- [14] Eugene Seneta, *Non-Negative Matrices*, John Wiley and Sons, New York, 1973 (ISBN 0-470-77605-6).
- [15] Abraham Berman and Robert J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Academic Press, New York, 1979 (ISBN 0-12-092250-9).
- [16] Henryk Minc, *Nonnegative Matrices*, John Wiley and Sons, New York, 1988 (ISBN 0-471-83966-3).
- [17] Martin Gardner, *Wheels, Life, and Other Mathematical Amusements*, W. H. Freeman and Company, New York, 1983. (ISBN 0-7167-1589-9 pbk)
- [18] Brian Hayes, "Computer Recreations - The cellular automaton offers a model of the world and a world unto itself," *Scientific American*, March 1984, pp. 10-16.
- [19] A. K. Dewdney, *The Armchair Universe*, W. H. Freeman and Company, New York, 1988. (ISBN 0-7167-1939-8 pbk)
- [20] Tommaso Toffoli and Norman Margolus, *Cellular Automata Machines*, The MIT Press, Cambridge, Massachusetts, 1987. (ISBN 0-262-20060-0).
- [21] Tommaso Toffoli, "Computation and Construction Universality of Reversible Cellular Automata," *Journal of Computer and System Sciences* **15** 213-231 (1977).
- [22] S. Amoroso and Y. N. Patt, "Decision Procedures for Surjectivity and Injectivity of Parallel Maps for Tessellation Structures," *Journal of Computer and System Sciences* **6** 448-464 (1972).
- [23] Anthony Ralston, "De Bruijn Sequences—A Model Example of the Interaction of Discrete Mathematics and Computer Science," *Mathematics Magazine* **55** 131-143 (1982).

- [24] Solomon W. Golomb, *Shift Register Sequences*, Holden-Day, Inc., San Francisco, 1967.
- [25] Erica Jen, "Scaling of preimages in cellular automata," *Complex Systems* **1** 1045-1062 (1987).
- [26] Wentian Li, "Power spectra of regular languages and cellular automata," *Complex Systems* **1** 107-130 (1987).
- [27] Stephen Wolfram, "Computation theory of cellular automata," *Communications in Mathematical Physics* **96** 15-57 (1984).
- [28] Edward F. Moore, "Gedanken Experiments on Sequential Machines," in C. E. Shannon and John McCarthy (eds), *Automata Studies*, Princeton University Press, Princeton, New Jersey, 1956.
- [29] Tom Head, "One-Dimensional Cellular Automata: Injectivity from Unambiguity," *Complex Systems* **3** 343-348 (1989).
- [30] R. C. Backhouse and B. A. Carré, "Regular Algebra Applied to Path-finding Problems," *Journal of the Institute of Mathematics and its Applications*, **15** 161-186 (1975).
- [31] William Feller, *An Introduction to Probability Theory and its Applications*, volume 2 (second edition), John Wiley and Sons, New York.
- [32] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*, American Mathematical Society, Providence, Rhode Island, 1943.
- [33] I. S. Berezin and N. P. Zhidkov, *Computing Methods*, volume 1, Pergamon Press, Oxford, 1965.
- [34] Milton Abramowitz and Irene A. Stegun (Editors), *Handbook of Mathematical Functions*, U. S. Government Printing Office, Washington D. C., 1964.
- [35] Michel Dubois-Violette and Alain Rouet, "A Mathematical Classification of the One-Dimensional Deterministic Cellular Automata," *Communications in Mathematical Physics* **112** 627-631 (1987), "Addendum," **118** 529 (1988).
- [36] Masakazu Nasu, "Uniformly finite-to-one and onto extensions of homomorphisms between strongly connected graphs," *Discrete Mathematics* **39** 171-197 (1982).
- [37] M. Perles, M. O. Rabin, and E. Shamir, "The Theory of Definite Automata," *IEEE Transactions on Electronic Computers* **EC-12** 233-243 (1963).

- [38] Zvi Kohavi, *Switching and Finite Automata Theory* (second edition), McGraw-Hill Book Company, New York, 1978 (ISBN 0-07-035310-7).
- [39] D. Richardson, "Tessellations with local transformations," *Journal of Computer and System Sciences* **6** 373-388 (1972).
- [40] Tadakazu Sato and Namio Honda, "Certain Relations between Properties of Maps of Tessellation Automata," *Journal of Computer and System Sciences* **15** 121-145 (1977).
- [41] Karel Culick II, "On Invertible Cellular Automata," *Complex Systems* **1** 1036-1044 (1987).
- [42] Lyman P. Hurd, "Formal Language Characterizations of Cellular Automaton Limit Sets," *Complex Systems* **1** 69-80 (1987).
- [43] Lyman P. Hurd, "The Non-wandering Set of a CA Map," *Complex Systems* **2** 549-554 (1988).
- [44] Karel Culick II, Jan Pachl, and Sheng Yu, "On the Limit Sets of Cellular Automata," *SIAM Journal on Computing* **18** 831-842 (1989).
- [45] Masakazu Nasu, "Indecomposable Local Maps of Tessellation Automata," *Mathematical Systems Theory* **13** 81-93 (1979).
- [46] Masakazu Nasu, "An Interconnection of Local Maps Inducing Onto Global Maps," *Discrete Applied Mathematics* **2** 125-150 (1980).
- [47] Asher Peres, "Reversible logic and quantum computers," *Physical Review A* **32** 3266-3276 (1985).
- [48] Shinji Takesue, "Reversible Cellular Automata and Statistical Mechanics," *Physical Review Letters* **59** 2499-2502 (1987).