Complex Systems 3 (1989) 421–456

Enumeration of Preimages in Cellular Automata

Erica Jen
Center for Nonlinear Studies and Theoretical Division,
Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Abstract. For a given rule and arbitrary spatial sequence, the preimages of the sequence are defined to be the set of tuples that are mapped by the rule onto the sequence. The enumeration of preimages provides information on the probability distribution associated with an automaton rule, determining, for example, the probability of occurrence of a sequence after one iteration of the rule operating on a random initial condition. It is shown here that formulae can be obtained for the exact number of preimages under an arbitrary one-dimensional nearest-neighbor automaton rule for any spatial sequence. The qualitative features of these formulae are determined by the combinatorial structure of the automata rule tables. The formulae are analytically and computationally useful for a wide variety of problems, including the identification of gardens-of-Eden (spatial sequences with no preimages), the evaluation of quantities that require knowledge of the probabilities of occurrence for all possible spatial sequences, and the characterization of statistical features such as the tendency to produce long runs of 1's and 0's.

1. Introduction

This paper derives formulae for the number of preimages for arbitrary spatial sequences generated by one iteration of a a one-dimensional cellular automaton on an infinite lattice. The formulae provide information on the statistical and dynamical features of these systems.

The *preimages* of any spatial sequence S are defined to be the set of sequences that are mapped to S by the automaton. The number of preimages for S determines its probability of occurrence when the automaton rule acts upon an initial condition in which all spatial sequences appear with equal frequency.

Preimages have been the subject of many studies, including those focussed on the relationship among preimages, surjectivity, and reversibility [1–3] the existence of gardens-of-Eden (sequences with no preimages) [4,5], and the nature of limit sets for cellular automata [6]. The enumeration of preimages was also considered in [7], where it was shown that the number of preimages for arbitrary sequences satisfies a system of recurrence relations with non-constant coefficients depending on the automaton rule. The recurrence relations were used in that paper to establish that the total number of preimages for certain well-defined subsets of spatial sequences scales with sequence length.

In this paper, it is shown that the full system of recurrence relations can be reduced to two uncoupled systems of linear recurrence relations with constant coefficients. Although uncoupled, the two systems are interdependent in that the solution of one serves as the initial value for the other. The systems can be solved to give formulae for the number of preimages for arbitrary spatial sequences.

The preimage formulae differ from rule to rule, and yet exhibit qualitative features that permit classification into general types. For example, the formulae for elementary (nearest-neighbor, binary site-valued) rules can be loosely characterized as belonging to one of six general types, with the particular type being determined by the combinatorial structure of the automaton's rule table.

Once the preimage formula for an automaton has been found, it is then possible to characterize its one time-step probability distribution. For example, the formulae provide analytically and computationally useful expressions for the evaluation of quantities (such as the one time-step spatial metric entropy) that require knowledge of the probabilities of occurrence for spatial sequences. The formulae are also useful for studying the features of the formal languages generated by automata rules, such as the identification of rules that generate only a finite number of excluded blocks (blocks of site values with no preimages).

The paper is organized as follows. Section 2 presents the system of recurrence relations for the number of preimages for arbitrary spatial sequences, and outlines the method of its solution. Section 3 discusses qualitative features of the different types of preimage formulae, and investigates the connection between the type of formula and the combinatorial structure of the automaton's rule table. In section 4, some examples are given of the recurrence relations for specific rules and their solution. Section 5 considers the use of the formulae to characterize the probability distribution and formal language features of the automaton. Concluding remarks are made in section 6, and the appendix lists the preimage formulae for all 32 symmetric elementary rules.

2. General form of recurrence relations

The general form of a one-dimensional cellular automaton on an infinite lattice is given by

$$x_i^{t+1} = f(x_{i-r}^t, \cdots, x_i^t, \cdots, x_{i+r}^t), \quad f : F_k^{2r+1} \to F_k,$$

where x_i^t denotes the value of site i at time t, f represents the "rule" defining the automaton, and r is a non-negative integer specifying the radius of the rule. The site values are restricted to a finite set of integers $F_k = \{0, 1, \cdots, k-1\}$, and are computed synchronously (in parallel) at each time step.

Let $S = s_n \cdots s_1$ be an arbitrary sequence of length n and R be an arbitrary cellular automaton rule of radius r. Denote the number of preimages of S under the rule R by $N(S)$. In [7], recurrence relations were derived for $N(S)$ in terms of the number of preimages of its subsequences, with the preimages categorized according to their leading symbols of length $2r$. Given such a count of preimages for a subsequence of length k, for example, it is clearly possible to determine the number of preimages for the sequence of length $k+1$ with a new site value (either 0 or 1) appended.

First consider the recurrence relations for $N(S)$ moving from right to left in the sequence. For any integer $0 \leq m \leq k^{2r} - 1$, with $m = \sum_{i=0}^{2r-1} m_i k^{2r-1-i}$, denote by $M = m_0 \cdots m_{2r-1}$ the symbol corresponding to its k-ary representation. The symbols M thus range over all possible blocks of length $2r$. In the case of elementary rules, for example, the set of symbols M is given by $\{00, 01, 10, 11\}$. The preimages for sequence S will then be grouped and counted according to their leading symbols, with the total number of preimages being the sum over all M of the number of preimages beginning with M; that is,

$$N(S) = \sum_{m=0}^{k^{2r}-1} L_M^n,$$

where, for any j,

$$L_M^j = \text{number of preimages beginning on the left with symbol } M \text{ for sequence } s_j \cdots s_1.$$

For elementary rules, for example, the number of preimages of a sequence S of length n

$$N(S) = L_{00}^n + L_{01}^n + L_{10}^n + L_{11}^n.$$

Now define an indicator function

$$\begin{aligned} I_j(x) &= 1 \text{ if } s_j = x, \\ &= 0 \text{ if } s_j \neq x, \end{aligned}$$

and for $0 \leq i \leq k-1$,

$$\begin{aligned} x_i &= f(m_0, \cdots, m_{2r-1}, i), \\ M_i &= m_1 \cdots m_{2r-1} i, \end{aligned}$$

that is, M_i is the symbol M shifted one component to the left, and with the element i appended on the right. For $M = 01$, for example, $M_0 = 10$ and $M_1 = 11$. Then it follows that

$$L^j_M = \sum_i L^{j-1}_{M_i} I_j(x_i), \tag{2.1}$$

and the above system of recurrence relations can be used to express $N(S)$ in terms of starting values L^1 that are easily computed from the definition of the automaton rule.

For example, consider Rule 126 defined by

$$\{000, 111\} \to 0, \quad \{001, 010, 011, 100, 101, 110\} \to 1.$$

Then from (2.1), the recurrence relations for the number of preimages for any sequence S can be written as

$$\begin{aligned}
L^j_{00} &= L^{j-1}_{00} I_j(0) + L^{j-1}_{01} I_j(1), \\
L^j_{01} &= (L^{j-1}_{10} + L^{j-1}_{11}) I_j(1), \\
L^j_{10} &= (L^{j-1}_{00} + L^{j-1}_{01}) I_j(1), \\
L^j_{11} &= L^{j-1}_{11} I_j(0) + L^{j-1}_{10} I_j(1).
\end{aligned}$$

The recurrence relations can equivalently be written for $N(S)$ moving left to right in the sequence S.

The problem now is to solve the system (2.1) for an arbitrary rule and string S. The solution is obtained by decomposing the full system into two linear subsystems coupled in the sense that the solution to one subsystem serves as the initial value for the other, and vice versa. The general technique is to define one subsystem (I) for the special case of a string of all 0's, and the other (II) for the case of a string of all 1's. Consider an arbitrary string S, then, as consisting of a series of alternating blocks of 0's and 1's. Subsystem (I) is then a linear system of recurrence relations that can be solved for the number of preimages of the first block of 0's. The solution then provides the initial values for subsystem (II) (again a linear system of recurrence relations) to be solved for the first block of 1's, and so on.

3. Solution of recurrence relations

Systematic solution of the recurrence relations for the 88 distinct elementary rules yields preimage formulae whose exact forms and parameters depend on the particular rules, but whose qualitative features permit classification into six types. The types, together with the labels and examples of the rules belonging to each type, are listed in this section. Note that the preimage formulae for all 88 distinct elementary rules have been derived [9], and are available upon request.

The notation used is the following. Let S be an arbitrary string.

(i) From right to left, divide S into "blocks" of consecutive 0's and 1's, and let

$$\begin{aligned} a_i &= \text{number of consecutive 0's in the } i\text{th block}, \\ b_i &= \text{number of consecutive 1's in the } i\text{th block}. \end{aligned}$$

For example, with $S = 00101001100100$, set $a_1 = a_2 = a_3 = 2$, $a_4 = 1$, $a_5 = 2$ and $b_1 = 1, b_2 = 2, b_3 = b_4 = 1$. The convention is used that all sequences begin on the right with a_1 0's, with $a_1 \geq 0$.

(ii) From right to left, divide S into blocks of consecutive isolated 1's, and let

$$d_i = \text{number of consecutive isolated 1's in the } i\text{th block.}$$

For example, with $S = 00101001100100$, set $d_1 = 1$, $d_2 = 2$.

(iii) $I(x)$ is defined to be the indicator function such that $I(1) = 1$ and $I(x) = 0$ otherwise.

Then preimage formulae for all elementary automata rules have been shown to belong to one of the following six types:

(A) constant;

(B) products of integers representing lengths l_i of blocks of;

(C) products of integers representing the l_i^{th} terms in;

(D) terms in sequences satisfying telescoping recurrence;

(E) terms in sequences whose values vary periodically;

(F) solutions to other linear recurrence relations (with non-constant coefficients) that do not obviously simplify.

For each of the above types, examples, together with a list of the rule numbers of automata belonging to that type, are given below. The rules listed are the 88 rules distinct under reflection and symmetry operations.

(A) constant;

Rules 0, 15, 30, 45, 51, 60, 90, 105, 150, 170, 204

Example: For linear (i.e., automata with interaction rules f that are linear in the site values) and other automata for which the uniform measure is invariant,

$$N(S) = 4.$$

(B) products of integers representing lengths l_i of blocks of consecutive "units," where units are either 1's and 0's, or combinations of 1's and 0's;

Rules 11, 46, 138, 12, 24, 34, 44, 14, 38, 42, 35, 140

Example: For Rule 12, the number of preimages for a sequence S is given by

$$\begin{aligned} N(S) &= 2(a_1+2), \quad \text{if } n=1, \\ &= 2(a_1+1)(a_n+1)\prod_{i=2}^{n-1} a_i, \quad \text{if } n>1 \text{ and all } b_i=1, \\ &= 0, \quad \text{otherwise.} \end{aligned}$$

where the a_i's are defined above and n is the number of distinct blocks of consecutive 0's.

Example: For Rule 35, the number of preimages $N(S)$ is given by

$$\begin{aligned} N(S) &= [2I(b_{n-1})+2-I(a_n)][3-I(a_1)]\prod_{j=2}^{n-1}[1+I(b_{j-1}) \\ &\quad -I(a_j)] \text{ if } a_1, a_n \neq 0, \\ &= [1+I(b_{n-1})][3-I(a_1)]\prod_{j=2}^{n-1}[1+I(b_{j-1})-I(a_j)] \\ &\quad \text{if } a_1 \neq 0, a_n = 0 \\ &= [1+I(b_n)][2-2I(a_2)+I(b_1)]\prod_{j=3}^{n}[1+I(b_{j-1})-I(a_j)] \\ &\quad \text{if } a_1 = a_n = 0 \\ &= [2I(b_{n-1}]+2-I(a_n))[2-2I(a_2) \\ &\quad + I(b_1)]\prod_{j=3}^{n-1}[1+I(b_{j-1})-I(a_j)], \\ &\quad \text{if } a_1 = 0, a_n \neq 0, \end{aligned}$$

where the a_i's, b_i's, and the indicator function $I(x)$ are defined above.

(C) products of integers representing the l_ith terms in Fibonacci-like sequences, where the l_i's again are lengths of blocks of consecutive units;

Rules 1, 2, 3, 17, 18, 19, 36, 72, 126, 128, 136, 200, 4, 32, 168, 7, 8, 16, 28, 50, 56, 76, 57, 162, 184, 156

Example: For Rule 126, $N(S)$ is given by

$$\begin{aligned} N(S) &= 2F_{b_1+3}, \quad \text{if } n=1, \\ &= 2F_{b_n+1}F_{b_1+1}\prod_{i=2}^{n-1} F_{b_i-1} \quad \text{otherwise,} \end{aligned}$$

where the b_i's are defined above, n is the number of distinct blocks of 1's, and $F_k = F_{k-1} + F_{k-2}$ with $F_0 = 0, F_1 = 1$.

Example: For Rule 50, $N(S)$ is given by

$$N(S) = 2F_{d_1+2} \times \prod_{i=2}^{\alpha} F_{d_i}[1,1] + I(b_n)F_{d_1+2}F_{d_\alpha-1}\prod_{i=1}^{\alpha-1} F_{d_i}[0,1]$$

where the d_i's are defined above, α is the number of distinct blocks of consecutive isolated 1's (i.e., blocks in which $b_i = 1$), and $F_k \equiv F_k[0,1] = F_{k-1} + F_{k-2}$.

(D) terms in sequences satisfying telescoping recurrence relations;
Rules 5, 6, 9, 33, 37, 54, 122, 94, 132, 146, 160, 23, 232, 62, 110, 108, 40, 130, 73
Example: For Rule 62, $N(S)$ is given by

$$\begin{aligned} N(S) &= F_{b_n+4}[0, F_{b_{n-1}+1}[0,\cdots, F_{b_2+1}[0, F'_{b_1+1}, F'_{b_1-1}],\cdots], \\ &\quad F_{b_{n-1}-1}[0,\cdots, F_{b_2+1}[0, F'_{b_1+1}, F'_{b_1-1}],\cdots]], \end{aligned}$$

where $F_k[f_0, f_1, f_2]$ satisfies the recurrence relation $F_k = F_{k-2} + F_{k-3}$ with $F_i = f_i$ for $i = 0, 1, 2$, and $F'_k = F_k[0,0,1]$.
Example: For Rule 33, the number of preimages $N(S)$ is given by

$$\begin{aligned} N(S) &= F_{a_n+3}[0, I(b_{n-1})F_{a_{n-1}}[0, I(b_{n-2})F_{a_{n-2}}[0,\cdots, \\ &\quad I(b_2)F_{a_2}[0, I(b_1)F_{a_1-1}, I(b_1)F_{a_1-1} + F_{a_1-2}]\cdots]], \\ &\quad I(b_{n-1})F_{a_{n-1}}[0, I(b_{n-2})F_{a_{n-2}}[0,\cdots, \\ &\quad I(b_2)F_{a_2}[0, I(b_1)F_{a_1-1}, I(b_1)F_{a_1-1} + F_{a_1-2}]\cdots]] \\ &\quad +F_{a_{n-1}-1}[0, I(b_{n-2})F_{a_{n-2}}[0,\cdots, \\ &\quad I(b_2)F_{a_2}[0, I(b_1)F_{a_1-1}, I(b_1)F_{a_1-1} + F_{a_1-2}]\cdots]], \end{aligned}$$

where the a_i's, b_i's, and the indicator function $I(x)$ are defined above.

(E) terms in sequences whose values vary periodically depending on the lengths l_i of blocks of consecutive units;
Rules 22, 104, 164, 41, 152, 10, 25, 26, 27, 172, 134
Example: For Rule 152, $N(S)$ is given by

$$N(S) = 2(v^{(n)}_{a_n} + v^{(n)}_{a_n-1}) + \sum_{i=0}^{a_n-2} v^n_i,$$

where $v^{(n)}_j = v^{(n)}_0$ if j is even, and $v^{(n)}_j = v^{(n)}_1$ if j is odd, and

$$v^{(n)} \equiv [v^{(n)}_0, v^{(n)}_1] = [v^{(n-1)}_{a_{n-1}}, I(b_{n-1})\sum_{i=0}^{a_{n-1}-1} v^{n-1}_i],$$

and

$$v^{(0)} = [1, I(b_1)(a_1+1)],$$

where $I(x) = 1$ iff $x = 1$.

Example: For Rule 22, $N(S)$ is given by

$$N(S) = v_{b_n}^{(n)} + \{F_{a_n+1} + F_{a_n-1} + F_{a_n-2}\}[0, v_{b_n-1}^{(n)}, v_{b_n-1}^{(n)} + v_{b_n+1}^{(n)}],$$

where $v_j^{(n)} = v_k^{(n)}$, $k = 0, 1, 2$ and $j \equiv k \pmod 3$, and the vector $v^{(n)}$ is defined to be

$$\begin{aligned} v^{(n)} &= [v_{b_{n-1}}^{(n-1)}, F_{a_n-1}[0, v_{b_{n-1}-1}^{(n-1)}, v_{b_{n-1}-1}^{(n-1)} + v_{b_{n-1}+1}^{(n-1)}], \\ &\quad F_{a_n-2}[0, v_{b_{n-1}-1}^{(n-1)}, v_{b_{n-1}-1}^{(n-1)} + v_{b_{n-1}+1}^{(n-1)}], \end{aligned}$$

with $v^{(0)} = \{F'_{a_1-2}, F'_{a_1-3}, 1\}_{b_1}$ and $F'_k \equiv F_k[2, 3, 4] = F_{k-1} + F_{k-3}$, and the b_i's as defined above.

(F) solutions to other linear recurrence relations (with non-constant coefficients) that do not obviously simplify.

Rules 13, 43, 58, 78, 142, 178, 29, 74, 77 (10)

Example: For Rule 178, $N(S)$ is given by

$$\begin{aligned} N(S) &= X_n + I(b_{n-1})X_{n-1} + I(a_n b_{n-1} b_{n-2})X_{n-2} \\ &\quad + I(a_n)K(b_{n-1})X_{n-3}, \end{aligned}$$

where

$$\begin{aligned} X_n &= K(a_n)K(b_{n-1})X_{n-1} + [I(b_{n-2}b_{n-1})K(a_n) \\ &\quad + I(a_n a_{n-1})K(b_{n-2})]X_{n-2} \\ &\quad + I(a_n a_{n-1} b_{n-2} b_{n-3})X_{n-3}, \end{aligned}$$

with initial values $X_1 = I(a_1) + K(a_1)$, $X_2 = K(b_1)X_1 + I(b_1) + I(a_1 a_2)$, and $X_3 = K(b_2)X_2 + I(b_1 b_2)X_1 + I(a_2 a_3)[K(b_1)X_1 + I(b_1)]$, where $I(x) = 1$ iff $x = 1$ and $K(x) = 1$ iff $x = 1$ or 2.

The type of preimage formula associated with a rule is determined by combinatorial features of its rule table. The easiest way to see the connection is by looking at the De Bruijn graph associated with a rule. This graph is constructed by assigning a node to each of the 4 possible 2-tuples 00, 01, 10, and 11, drawing a directed edge between each tuple xy and y (i.e., between each pair such that the second component of one tuple is the same as the first component of the other), and then labelling the edge with the value assigned by the rule to the 3-tuple xy. Figure 1a shows the De Bruijn graph associated with the generic rule

$$000 \to a_0, 001 \to a_1, \cdots, 111 \to a_7,$$

and the De Bruijn graph for Rule 12 is given in figure 1b.

In discussing such graphs, the terminology used will be as follows. A *path* is any sequence of connected edges in the graph. A *circuit* is a path

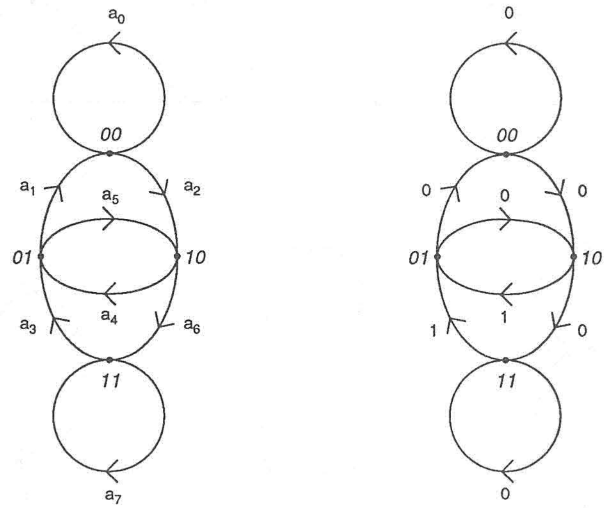

Figure 1: De Bruijn graphs for (a) a general elementary automaton rule; (b) Rule 12 defined by $\{000, 001, 100, 101, 110, 111\} \to 0$, $\{010, 011\} \to 1$.

that begins and ends at the same node. The *length* of a circuit is the number of edges in the shortest path traversing the circuit. A *constant-valued circuit* is a circuit with all edges carrying the same value. Two circuits are *equal-valued* if they are constant-valued and carry the same value. Two circuits are *disjoint* if they contain no node in common. A *simple* constant-valued circuit is a circuit that is disjoint from all other circuits carrying the same value.

It follows that the total number of preimages for a sequence $S = s_0 s_1 \cdots$ is given by the total number of distinct paths such that the values of the edges traversed by the path are given by s_0, s_1, and so on. In particular, suppose (as an exercise) that the problem is to determine the number of preimages for a sequence of n 0's, and that the set of edges labelled with 0 are connected as shown in the various diagrams of figure 2. Suppose further that the preimages to be counted are only those starting with the components represented by x and ending with those represented by y. Then the desired number of preimages is equivalent to the number of distinct paths of length n that can be constructed starting from the node x and ending with y. Specifically, the number of preimages is given in each case by

(a) $N(S) = 1$.

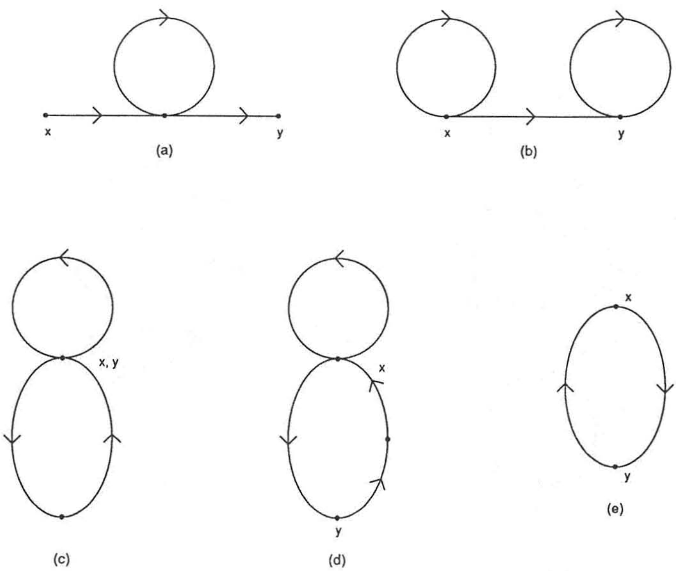

Figure 2: De Bruijn graphs for which the number of paths of length n beginning at node x and ending at node y are computed.

(b) $N(S) = n$. An admissible path is constructed by traversing the loop $x - x$ for some $n_1 \geq 0$ steps, then following the edge $x - y$, and finally traversing the loop $y - y$ for exactly $n - (n_1 + 1)$ steps. There are n possible positions for the edge $x - y$ to occur in the path, and therefore n possible paths.

(c) $N(S) = F_n$, where $F_n = F_{n-1} + F_{n-2}$ with $F_0 = F_1 = 1$. In this case, the number of paths of length n that begin at x and end at y is the sum of the number of paths of length $n - 1$ that end at y and can be lengthened by one traversal of the loop $y - y$, and the number of paths of length $n - 2$ that end at y and can be lengthened by one traversal of the circuit $y - z - y$. Thus the number of paths satisfies a recurrence relation of degree 2 with $F_0 = F_1 = 1$ (in fact, the Fibonacci relation).

(d) $N(S) = F_{n-1}$, where $F_k = F_{k-1} + F_{n-3}$ with $F_0 = F_1 = F_2 = 1$. The reasoning is similar to that above.

(e) $N(S) = 1$ if n odd,
$= 0$ if n even.

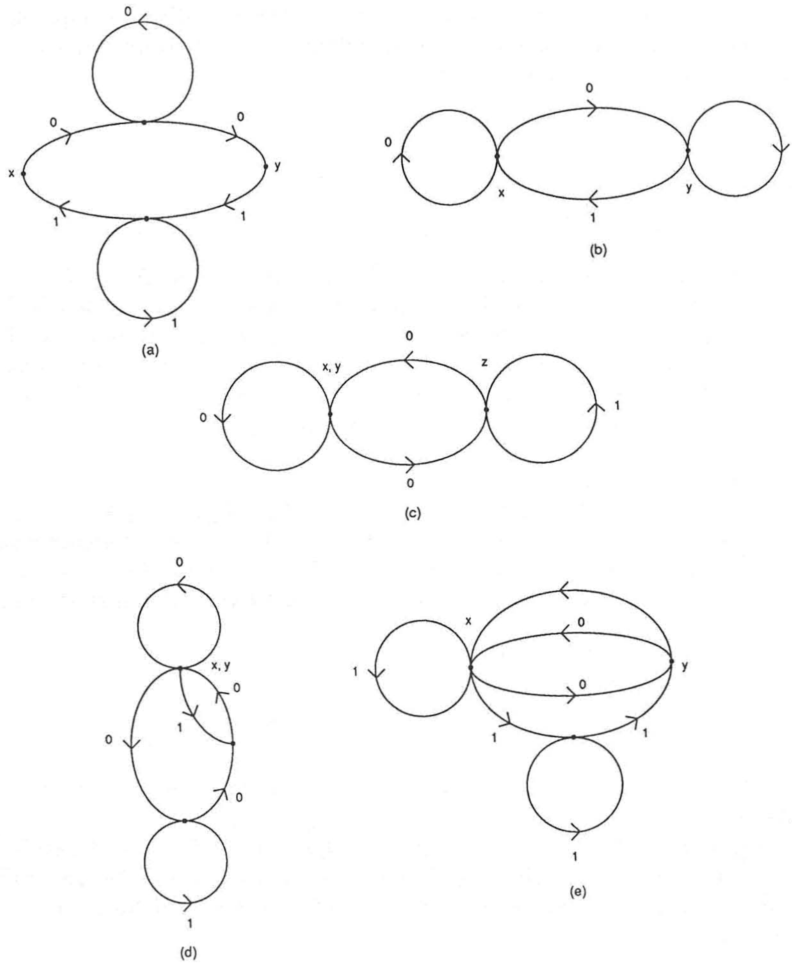

Figure 3: De Bruijn graphs for which the number of paths of length n, beginning at node x and ending at node y, and consisting of edges with the value 0 together with edges with value 1, are computed.

Now consider in addition edges labelled with 1's, as in the diagrams of figure 3. Suppose that the problem is to find the number of preimages of a sequence consisting of a_2 0's followed by b_1 1's followed by a_1 0's, and again that the only preimages to be counted are those starting with the components x and ending with the components y. By extending the reasoning above, the number of preimages is found to be

(a) $N(S) = 1$.

(b) $N(S) = a_2 a_1$ if $b_1 = 1$,
$= 0$ otherwise.

(c) $N(S) = F_{a_2-1} F_{a_1-1}$, where $F_k = F_{k-1} + F_{k-2}$, and $F_0 = F_1 = 1$. The number of paths that start at x, traverse the edges labelled 0 for a_1 steps, and then traverse the loop $z - z$ for b_1 steps is F_{a_1-1} with $F_0 = F_1 = 1$. The number of paths that start at z, traverse the edges labelled 0 for a_2 steps, and end at y is F'_{a_2-1} where $F'_k = F'_{k-1} + F'_{k-2}$ with $F'_0 = F'_1 = F_{a_1-1}$, or therefore $F_{a_2-1} F_{a_1-1}$.

(d) $N(S) = F_{a_2-2} F_{a_1-1}$ if $b_1 > 1$,
$= F_{a_2}[0, F_{a_1}, F_{a_1} + F_{a_1-1}]$ if $b_1 = 1$, where $F_k = F_{k-1} + F_{k-3}$. The reasoning in both cases is the same as in (c). The difference when $b_1 = 1$ is that the the initial values of F_k cannot be factored out so as to express the solution as the simple product (without sums) of terms.

(e) $N(S) = 1$ if a_1, a_2 both odd,
$= b_1 - 1$ if a_1, a_2 both even,
$= 0$ otherwise.

Note that the above five cases exhibit the qualitative features of types (A)-(E) listed above.

In general, as the preceding examples suggest, conditions can be derived relating a rule's type of preimage formula to the structure of its De Bruijn graph. Combinatorial arguments establish that the preimage formula for an elementary rule

(I) is constant if every node has the feature that either all incoming or all outgoing edges carry distinct values (i.e., either the graph or the graph with the direction of each edge reversed is deterministic).

(II) consists of "pure" terms (i.e., terms that do not involve solutions to recurrence relations) only if all constant-valued circuits are disjoint.

(III) contains Fibonacci-like terms if there exist at least 2 distinct non-disjoint constant-valued circuits $C_1, C_2, \cdots C_n$ of lengths $l_1, l_2, \cdots l_n$ (i.e., the circuits contain a node common to all). Then the terms satisfy the recurrence relation $F_k = F_{k-l_1} + F_{k-l_2} + \cdots F_{k-l_n}$.

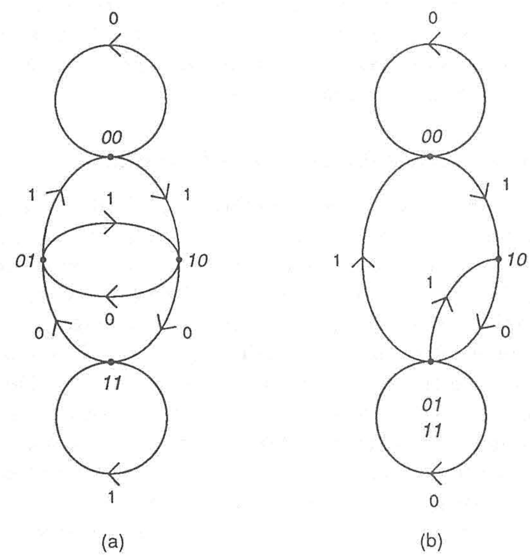

Figure 4: (a) Rule 50, for which blocks of consecutive isolated 1's are the basic counting units; (b) Rule 178, for which no constant-valued circuit of length greater than 1 exists.

Condition (III) can be generalized to treat cases such as Rule 50 (see the examples following the description of Type C) in which the fundamental unit is a block containing an isolated 1, rather than a block of arbitrary length of consecutive 1's. The De Bruijn graph of Rule 50 (shown in figure 4a) indicates that with a redefinition of the relevant circuits $C_1, C_2, \cdots$ to permit the occurrence of an isolated 1, the rule satisfies the condition for a preimage formula with recurrence relation terms.

A further distinction can be made among rules whose formulae represent products of simple terms (as in Types B and C), telescoping recurrence terms (as in Type D), and terms in periodic series (Type E). A preimage formula

(IV) consists of products of Fibonacci-like terms if the conditions of (c) hold and in addition, denoting by x^* a node common to the constant-valued circuits $C_1, C_2, \cdots C_n$, the number of edges separating x^* from any edge labelled with the same value is at least $l_{\max} - 1$, where $l_{\max}$ is the maximum circuit length. Otherwise the formula belongs to Type D.

(V) belongs to Type E (contains periodic terms) iff there exists a simple constant-valued circuit of length greater than 1. The length of the circuit is then the period of the sequence.

Note that the rules classified here as belonging to Type F are characterized by the absence in their De Bruijn graphs of any constant-valued circuits of length greater than 1. Such is the case, for example, in the graph for Rule 178, shown in figure 4b. A re-definition of variables (as in the case of Rule 50 discussed above) may, however, permit the solution in closed-form of the recurrence relations for rules of this type.

4. Examples of solving for preimage formulae

In this section, cellular automata exemplifying each of the major types will be treated in detail. Aspects to be discussed include the method of solution for each example and the relationship between the combinatorial features of each rule table and its type of preimage formula. As the discussion in the previous section made clear, the number of preimages for a sequence is equal to the number of appropriately defined paths in the reduced De Bruijn graph associated with the rule. For counting purposes, the De Bruijn graph given in conjunction with each example will be the reduced graph obtained by combining nodes equivalent in the sense that all paths through the nodes carry the same symbol sequences. (See [10] for discussion of equivalent representations of De Bruijn graphs for automata rules.)

Case a. As a trivial but nonetheless interesting example, first consider Rule 30 defined by

$$\{000, 101, 110, 111\} \rightarrow 0, \quad \{001, 010, 011, 100\} \rightarrow 1. \tag{4.1}$$

The recurrence relations for this rule are given by

$$\begin{aligned}
L^j_{00} &= L^{j-1}_{00} I_j(0) + L^{j-1}_{01} I_j(1), \\
L^j_{01} &= (L^{j-1}_{10} + L^{j-1}_{11}) I_j(1), \\
L^j_{10} &= L^{j-1}_{00} I_j(1) + L^{j-1}_{01} I_j(0), \\
L^j_{11} &= (L^{j-1}_{11} + L^{j-1}_{10}) I_j(0).
\end{aligned}$$

Solving the above for the special case of a string of all 0's, and then for the case of a string of all 1's, the system can be rewritten as the two systems

$$\begin{aligned}
U^j_{00} &= U^{j-1}_{00}, \\
U^j_{01} &= 0, \\
U^j_{10} &= U^{j-1}_{00}, \\
U^j_{11} &= U^{j-1}_{11} + U^{j-1}_{10},
\end{aligned} \tag{4.2}$$

$$\begin{aligned}
V^j_{00} &= V^{j-1}_{01}, \\
V^j_{01} &= V^{j-1}_{10} + V^{j-1}_{11}, \\
V^j_{10} &= V^{j-1}_{01}, \\
V^j_{11} &= 0.
\end{aligned} \tag{4.3}$$

With $U^1_{00} = U^1_{10} = 1, U^1_{11} = 2, U^1_{01} = 0$ as initial values, system (2) can be solved to yield $U^{a_1}_{00} = 1, U^{a_1}_{01} = 0, U^{a_1}_{10} = I(a_1), U^{a_1}_{11} = 3 - I(a_1)$, implying that for any sequence of a_1 0's, $N(S) = 4$. Moreover, for any $n > 1$, it is easy to show that

$$U_{00}^{a_n} = V_{00}^{b_{n-1}}, U_{01}^{a_n} = 0, U_{10}^{a_n} = I(a_n)V_{01}^{b_{n-1}},$$

$$U_{11}^{a_n} = V_{10}^{b_{n-1}} + [1 - I(a_n)]V_{01}^{b_{n-1}},$$

which together with the fact that $V_{11}^{b_{n-1}} = 0$, implies that the total number of preimages for a sequence ending with 1's on the left is not changed by adding 0's. Similarly, it can be shown that $N(S)$ is unaffected by adding 1's to the left of a sequence of 0's, and thus for all sequences S, it follows that $N(S) = 4$. The constancy of the number of preimages is of course clear from figure 5a since exactly one edge with the value 1 and one edge with the value 0 emanate from every node in the graph.

As a further detail, the decomposition of the preimages by leading components exhibits a structure described by

$$V_{00}^{b_n} = v_{b_{n-1}-1}^{(n-1)}, V_{01}^{b_n} = v_{b_{n-1}}^{(n-1)}, V_{10}^{b_n} = v_{b_{n-1}-2}^{(n-1)}, V_{11}^{b_n} = 0,$$

where $v_j^{(n)} = v_k^{(n)}$, $k = 0, 1, 2$ and $j \equiv k \pmod 3$, with the vector $v^{(n)}$ being one of the four vectors [0,3,1], [0,1,3], [0,4,0], or [0,0,4]. Initially $v^{(0)} = [0, 3, 1]$, and the vector evolves, as the sequence grows right to left, according to the rules:

$$\begin{array}{ccccc}
[0,3,1] & & & & \\
& \searrow b_i \equiv 1 & & & \\
b_i \equiv 2 \uparrow\downarrow b_i \equiv 2 & & [0,4,0] & \underset{b_i \equiv 1,2}{\overset{b_i \equiv 2}{\rightleftharpoons}} & [0,0,4] \\
& \nearrow b_i \equiv 1 & & & \\
[0,1,3] & & & &
\end{array}$$

The above should be interpreted to mean that the vector changes in the fashion indicated by the arrow depending on each new value of $b_i \pmod 3$, where b_i is the number of 1's in the next-to-the last block of 1's. (If b_i assumes a value not associated with an arrow, the vector remains unchanged.) For a sequence with a total of b_n blocks of 1's, in other words, the vector $v^{(n)}$ is determined by the values of $b_1, b_2, \cdots b_{n-1}$, and the number of preimages according to leading left components is then given by the b_nth component of the vector. For instance, in a sequence with $b_i \equiv 0, 2 \pmod 3$ for all i, the vector $v^{(n-1)}$ is given by [0,3,1] if the number of $b_i \equiv 2 \pmod 3$ is even, and by [0,1,3] if the number of $b_i \equiv 2 \pmod 3$ is odd, and the number of preimages of the sequence that begin with 01 is given by the kth component of $v^{(n-1)}$ where $b_n \equiv k \pmod 3$.

Case b. Next consider Rule 12 defined by

$$\{000, 001, 100, 101, 110, 111\} \to 0, \quad \{010, 011\} \to 1. \tag{4.4}$$

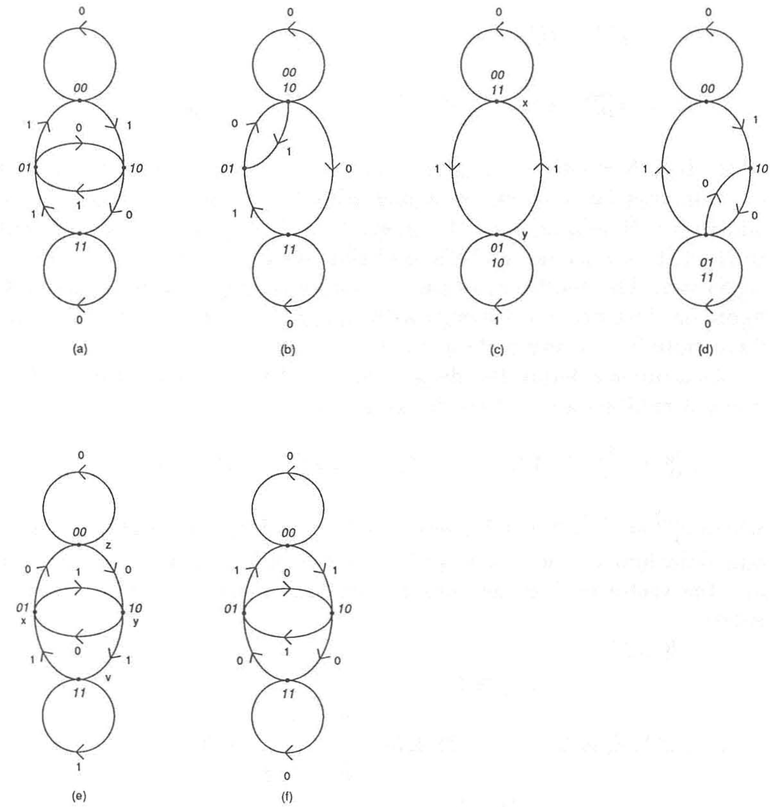

Figure 5: Reduced De Bruijn graphs for Rules (a) 30, (b) 12, (c) 126, (d) 18, (e) 232, (f) 22.

The recurrence relations for this rule are given by

$$\begin{aligned}
L_{00}^{j} &= (L_{00}^{j-1} + L_{01}^{j-1})I_j(0), \\
L_{01}^{j} &= (L_{10}^{j-1} + L_{11}^{j-1})I_j(1), \\
L_{10}^{j} &= (L_{00}^{j-1} + L_{01}^{j-1})I_j(0), \\
L_{11}^{j} &= (L_{11}^{j-1} + L_{10}^{j-1})I_j(0).
\end{aligned}$$

Solving the above for the special case of a string of all 0's, and then for the case of a string of all 1's, the system can be rewritten as the two systems

$$U_{00}^{j} = U_{00}^{j-1} + U_{01}^{j-1},$$

$$\begin{aligned} U_{01}^{j} &= 0, \\ U_{10}^{j} &= U_{00}^{j-1} + U_{01}^{j-1}, \\ U_{11}^{j} &= U_{11}^{j-1} + U_{10}^{j-1}, \end{aligned} \tag{4.5}$$

$$\begin{aligned} V_{00}^{j} &= 0, \\ V_{01}^{j} &= V_{10}^{j-1} + V_{11}^{j-1}, \\ V_{10}^{j} &= 0, \\ V_{11}^{j} &= 0. \end{aligned} \tag{4.6}$$

The definition of the rule implies

$$U_{00}^{1} = U_{10}^{1} = U_{11}^{1} = 2, \quad U_{01}^{1} = 0.$$

With these as initial values, system (4.5) yields for a string of a_1 0's,

$$U_{00}^{a_1} = U_{10}^{a_1} = 2, U_{01}^{a_1} = 0, U_{10}^{a_1} = 2a_1.$$

For a string that consists of b_1 1's appended on the left to the first block of 0's, system (4.6) yields (using the above as starting values)

$$V_{00}^{b_1} = V_{11}^{b_1} = V_{10}^{b_1} = 0, \; V_{01}^{b_1} = 2I(b_1)(a_1+1),$$

where $I(x) = 1$ iff $x = 1$. The values above are then used as starting values for system (4.5) with a_2 0's, and continuing the process yields the general solution

$$U_{00}^{a_u} = U_{10}^{a_n} = 2(a_1+1)\prod_{i=2}^{n-1} a_i, \; U_{01}^{a_n} = 0, \; U_{11}^{a_n} = 2(a_1+1)(a_n-1)\prod_{i=2}^{n-1} a_i,$$

for the case that $b_i = 1$ for all i. Since the total number of preimages is equal to the sum of the individual terms above, it follows that the number of preimages is given by

$$\begin{aligned} N(S) &= 2(a_1+2), \quad \text{if } n = 1, \\ &= 2(a_1+1)(a_n+1)\prod_{i=2}^{n-1} a_i, \quad \text{if } n > 1 \text{ and all } b_i = 1, \\ &= 0, \text{ otherwise.} \end{aligned} \tag{4.7}$$

It is straightforward to show that the same formula holds true in the case when either a_1 or a_n is equal to 0.

The De Bruijn graph for Rule 12 (shown in figure 5b) indicates that all constant-valued circuits are disjoint, and therefore the rule satisfies the condition for a Type A preimage formula.

Case c. Consider Rule 126, defined by

$$\{000, 111\} \to 0, \quad \{001, 010, 011, 100, 101, 110\} \to 1. \tag{4.8}$$

The recurrence relations for the number of preimages $N(S)$ were given in the previous section. System (2.2) for this rule can be decomposed into the two systems

$$\begin{aligned} U_{00}^{j} &= U_{00}^{j-1}, \\ U_{01}^{j} &= 0, \\ U_{10}^{j} &= 0, \\ U_{11}^{j} &= U_{11}^{j-1}. \end{aligned} \tag{4.9}$$

$$\begin{aligned} V_{00}^{j} &= V_{01}^{j-1}, \\ V_{01}^{j} &= V_{10}^{j-1} + V_{11}^{j-1}, \\ V_{10}^{j} &= V_{00}^{j-1} + V_{01}^{j-1}, \\ V_{11}^{j} &= V_{10}^{j-1}. \end{aligned} \tag{4.10}$$

The definition of the rule implies

$$U_{00}^{1} = 1,\ U_{01}^{1} = U_{10}^{1} = 0,\ U_{11}^{1} = 1.$$

With these as initial values, system (4.9) yields for a string of a_1 0's,

$$U_{00}^{a_1} = U_{11}^{a_1} = 1,\ U_{01}^{a_1} = U_{10}^{a_1} = 0,$$

For a string that consists of b_1 1's appended on the left to the first block of 0's, system (10) yields (using the above as starting values)

$$V_{00}^{b_1} = V_{11}^{b_1} = F_{b_1 - 1}[0,1],\ V_{01}^{b_1} = V_{10}^{b_1} = F_{b_1}[0,1],$$

where $F_k = F_{k-1} + F_{k-2}$ (e.g., the kth Fibonacci number) with starting values $F_0 = x, F_1 = y$. The values above are then used as starting values for system (4.9) with a_2 0's, and continuing the process yields the general solution

$$V_{00}^{b_n} = V_{11}^{b_n} = V_{01}^{b_n - 1},\ V_{01}^{b_n} = V_{10}^{b_n} = F_{b_n} F_{b_1 - 1} \prod_{i=2}^{n-1} F_{b_i - 1},$$

where the terms $F_k[0,1]$ have now been abbreviated as F_k. Since the total number of preimages is equal to the sum of the individual terms above, it follows that

$$N(S) = 2F_{b_n+1} F_{b_1+1} \times \prod_{i=2}^{n-1} F_{b_i - 1}, \tag{4.11}$$

independent of the values of the a_i's.

The De Bruijn graph for Rule 126 is shown in figure 5c. Given any sequence of consecutive b_i 1's (bordered in general by 0's on the left and right), with $1 < i < n$, a preimage for the sequence is constructed by concatenating the nodes of a path that starts at node x, then uses some combination of traversals of loops x-y-x and loop y-y, and finally ends at node x. The number of preimages is the number of possible such paths. Let F_k denote the number of paths of length k that start and end at x. Statement (III) from section 3 states that F_k satisfies the recurrence relation $F_k = F_{k-1} + F_{k-2}$. Moreover, the "distance" between the node x (labelled 0) and node y is 1, and therefore the distance is $\leq k-1$, where k is the degree of the recurrence relation. Thus, statement (IV) from section 3 implies that the total number of paths that can be constructed for an arbitrary sequence satisfies a formula expressible as the product of Fibonacci terms, and the total number of preimages for the entire string is given by the above expression for $N(S)$.

Case d. Next consider Rule 18, defined by

$$\{000, 011\} \to 0, \quad \{001, 010, 100, 101, 110, 111\} \to 1. \tag{4.12}$$

The two systems corresponding to the cases of all 0's and all 1's are given by

$$\begin{aligned} U_{00}^j &= U_{00}^{j-1}, \\ U_{01}^j &= U_{11}^{j-1}, \\ U_{10}^j &= 0,) \\ U_{11}^j &= 0. \end{aligned} \tag{4.13}$$

$$\begin{aligned} V_{00}^j &= V_{01}^{j-1}, \\ V_{01}^j &= V_{10}^{j-1}, \\ V_{10}^j &= V_{00}^{j-1} + V_{01}^{j-1}, \\ V_{11}^j &= V_{10}^{j-1} + V_{11}^{j-1}. \end{aligned} \tag{4.14}$$

It is easy to see that all b_i must be either 1 or 2 in order for the number of preimages to be nonzero. Since

$$U_{00}^1 = U_{01}^1 = 1, \; U_{10}^1 = U_{11}^1 = 0,$$

system (3.3) yields for a string of a_1 0's,

$$U_{00}^{a_1} = 1, = 1, \; U_{01}^{a_1} = U_{11}^{a_1} = F_{a_1-1}[2,3], \; U_{10}^{a_1} = F_{a_1-1}[2,3],,$$

For a string that consists of b_1 1's appended on the left to the first block of 0's, system (3.4) yields (using the above as starting values)

$$V_{00}^{b_1} = I(b_1)F_{a_1-1}[2,3], \; V_{01}^{b_1} = V_{11}^{b_1} = 0, \; V_{10}^{b_1} = I(b_1) + I(b_1-1)F_{a_1-1}[2,3],$$

where F_k again denotes the kth Fibonacci number, and $I(x)$ is an indicator function such that $I(1) = 1$ and $I(x) = 0$ otherwise. The values above are then used as starting values for system (4.13) with a_2 0's, and continuing the process yields the general solution

$$U_{00}^{a_n} = I(b_{n-1})U_{01}^{a_n-1}, \; U_{10}^{a_n} = U_{01}^{a_n-1},$$

$$U_{01}^{a_n} = U_{11}^{a_n} = F_{a_n}[I(b_{n-1}b_{n-2})U_{01}^{a_n-2} + I(b_{n-1}-1)U_{01}^{a_n-1}], \tag{4.15}$$

where F_k denotes $F_k[0,1]$. A closed-form expression can be given for the above system by defining k_L, k_R to be the leftmost and rightmost values of i such that $b_i = 2$. Then

$$U_{01}^{a_n} = F_{a_n^*} F_{a_1^*} \prod_{i \in A \cup B} F_{a_i},$$

where

$$\begin{aligned} F_{a_1^*} &= F'_{a_1 - 1} F_{a_3} \cdots F_{a_{k_R}}, \quad k_R \textit{ even}, \\ &= F_{a_2} F_{a_4} \cdots F_{a_{k_R}}, \quad k_R \textit{ odd}, \\ F_{a_n^*} &= F_{a_n} F_{a_{n-2}} \cdots F_{a_{k_L + 3}}, \quad n - k_L \textit{ odd}, \\ &= 0, \quad n - k_L \textit{ even}, \end{aligned}$$

and

$$A \equiv \{ i \mid b_{i-1} = 2 \},$$

$$B \equiv \{ i \mid \exists k_1, k_2 \textit{ with } k_1 < k_2, \textit{ and } b_{k_1} = b_{k_2} = 2,$$

$$b_j = 1 \text{ for all } k_1 < j < k_2,\ k_1 \textit{ odd and } i = k_1 + 3, k_1 + 5, \cdots, k_2 \}.$$

If there exists k_1, k_2 satisfying the conditions above but with $k_2 - k_1$ even, then B is defined to be the set $\{0\}$. The other values U_{00}, U_{10}, U_{11} are determined from equation (4.15). The total number of preimages is thus given by

$$\begin{aligned} N(S) &= (2F_{a_n} + F_{a_{n-1}}) \prod_{j=2, j \textit{ even}}^{n-k_L-3} F_{a_{n-j}} F_{a_1^*} \prod_{i \in A \cup B} F_{a_i}, \quad n - k_L \textit{ odd}, \\ &= \prod_{j=1, j \textit{ odd}}^{n-k_L-3} F_{a_{n-j}} F_{a_1^*} \prod_{A,B} F_{a_i}, \quad n - k_L \textit{ even}, \end{aligned}$$

for strings with all b_i equal to 1 or 2, and $N(S) = 0$ otherwise.

The De Bruijn graph of Rule 18 is shown in figure 5e, and clearly indicates a Fibonacci-structure for the preimage formula. The presence of the 0-valued loop x-x complicates the counting procedure for this graph, and in particular induces a dependence on the odd versus even nature of the number of blocks of 0's.

To illustrate the use of the preimage formula, let

$$S = 1100010101100000011010.$$

Then $n = 7$ (since by assumption all strings begin and end with 0), and $k_L = 6$, $k_R = 2$. Hence

$$F_{a_1^*} = F_{a_2} = F_1.$$

Also, the set A is defined to be $A \equiv \{a_3,\ a_4\}$, and B to be the set $B \equiv \{a_6\}$, and $N(S)$ is given by

$$\begin{aligned} N(S) &= (2F_{a_n} + F_{a_n - 1}) \prod_{j=2, j\ even}^{n-k_L-3} F_{a_{n-j}} F_{a_1^*} \prod_{i \in A \cup B} F_{a_i}, \\ &= F_{-1} F_1 F_5 F_1 F_3, \\ &= 10. \end{aligned}$$

Case e. Next consider Rule 232, defined by

$$\{000, 001, 010, 100\} \to 0, \quad \{011, 101, 110, 111\} \to 1. \tag{4.16}$$

Then from (2.1), the special cases of a string of all 0's and a string of all 1's can be written, as in the preceding examples, as

$$\begin{aligned} U_{00}^j &= 0, \\ U_{01}^j &= U_{11}^{j-1}, \\ U_{10}^j &= U_{01}^{j-1}, \\ U_{11}^j &= U_{10}^{j-1} + U_{11}^{j-1}. \end{aligned} \tag{4.17}$$

$$\begin{aligned} V_{00}^j &= V_{00}^{j-1} + V_{01}^{j-1}, \\ V_{01}^j &= V_{10}^{j-1}, \\ V_{10}^j &= V_{00}^{j-1}, \\ V_{11}^j &= 0. \end{aligned} \tag{4.18}$$

For a string S beginning on the right with a_1 0's, the system (4.17) yields

$$\begin{aligned} U_{00}^{a_1-1} &= 0,\ U_{01}^{a_1-1} = F_{a_1-2}[2,3,4], \\ U_{10}^{a_1-1} &= F_{a_1-3}[2,3,4], U_{11}^{a_1-1} = F_{a_1-1}[2,3,4], \end{aligned}$$

with $F_k[x_0, x_1, x_2]$ defined as the value satisfying the recurrence relation $F_k = F_{k-1} + F_{k-3}$ with initial values $F_0 = x_0, F_1 = x_1, F_2 = x_2$. Then for a string of b_1 1's, system (4.18) yields (using the above as starting values)

$$\begin{aligned} V_{00}^{b_1} &= F_{b_1}[0, F_{a_1-2}[2,3,4], F_{a_1-2}[2,3,4] + F_{a_1-3}[2,3,4]], \\ V_{01}^{b_1} &= F_{b_1-2}[0, F_{a_1-2}[2,3,4], F_{a_1-2}[2,3,4] + F_{a_1-3}[2,3,4]], \\ V_{10}^{b_1} &= F_{b_1-1}[0, F_{a_1-2}[2,3,4], F_{a_1-2}[2,3,4] + F_{a_1-3}[2,3,4]], \\ V_{11}^{b_1} &= 0. \end{aligned}$$

Continuing the process yields the general solution

$$\begin{aligned} U_{00}^{a_n} &= 0, \\ U_{01}^{a_n} &= U_{11}^{a_n-1}, \\ U_{10}^{a_n} &= U_{11}^{a_n-2}, \end{aligned}$$

$$\begin{aligned}U_{11}^{a_n} &= F_{a_n}[0, F_{b_{n-1}-1}[0, F_{a_{n-1}-2}[0, \cdots, F_{a_2-2}[0, F_{b_1-1}[0, F'_{a_1-2}, F'_{a_1-2} \\ &\quad + F'_{a_1-3}], \\ &\quad F_{b_1-2}[0, F'_{a_1-2}, F'_{a_1-2} + F'_{a_1-3}]], \cdots,]], \\ &\quad F_{b_{n-1}-2}[0, F_{a_{n-1}-2}[0, \cdots, F_{a_2-2}[0, F_{b_1-1}[0, F'_{a_1-2}, F'_{a_1-2} \\ &\quad + F'_{a_1-3}], \\ &\quad F_{b_1-2}[0, F'_{a_1-2}, F'_{a_1-2} + F'_{a_1-3}]], \cdots,]]],\end{aligned}$$

where the terms $F_k[2,3,4]$ have now been abbreviated as F'_k, and all other terms represent solutions to the recurrence relation with specified initial values. It follows that

$$\begin{aligned}N(S) &= F_{a_n+2}[0, F_{b_{n-1}-1}[0, F_{a_{n-1}-2}[0, \cdots, F_{a_2-2}[0, F_{b_1-1}[0, F'_{a_1-2}, \\ &\quad F'_{a_1-2} + F'_{a_1-3}], \\ &\quad F_{b_1-2}[0, F'_{a_1-2}, F'_{a_1-2} + F'_{a_1-3}]], \cdots,]], \\ &\quad F_{b_{n-1}-2}[0, F_{a_{n-1}-2}[0, \cdots, F_{a_2-2}[0, F_{b_1-1}[0, F'_{a_1-2}, \\ &\quad F'_{a_1-2} + F'_{a_1-3}], \\ &\quad F_{b_1-2}[0, F'_{a_1-2}, F'_{a_1-2} + F'_{a_1-3}]], \cdots,]]], \qquad (4.19)\end{aligned}$$

and thus the total number of preimages for any string is obtained as the solution to a telescoping recurring relation.

Again, the De Bruijn graph for Rule 232 (shown in figure 5e) serves as the basis of a combinatorial proof of the formula for $N(S)$. The graph can be decomposed into two connected subgraphs, one generating strings of all 0's and the other generating strings of all 1's. Each subgraph consists of two loops, one of order 1 (v-v and z-z) and one of order 3 (edges x-z-y-x and x-y-v-x). Thus, for each case, the number of paths G_k that can be constructed corresponding to a string of length k can be shown, using an argument analogous to that used for Rule 126 above, to satisfy the relation $G_k = G_{k-1} + G_{k-3}$.

Case f. Finally consider Rule 22 defined by

$$\{000, 011, 101, 110, 111\} \to 0, \quad \{001, 010, 100\} \to 1. \qquad (4.20)$$

The full system is given by

$$\begin{aligned}L_{00}^j &= L_{00}^{j-1} I_j(0) + L_{01}^{j-1} I_j(1), \\ L_{01}^j &= (L_{10}^{j-1} + L_{11}^{j-1}) I_j(1), \\ L_{10}^j &= (L_{00}^{j-1} + L_{01}^{j-1}) I_j(1), \\ L_{11}^j &= L_{11}^{j-1} I_j(0) + L_{10}^{j-1} I_j(1).\end{aligned}$$

With

$$L_{00}^1 = L_{01}^1 = L_{10}^1 = 1, L_{11}^1 = 2.$$

as initial values, the above system yields for a string of a_1 0's,

$$U_{00}^{a_1} = 1, U_{01}^{a_1} = F_{a_1-2}[2,3,4], U_{10}^{a_1-3} = F_{a_1}[2,3,4], U_{11}^{a_1} = F_{a_1-1}[2,3,4],$$

where $F_k[x_0, x_1, x_2]$ satisfies the relation $F_k = F_{k-1}+F_{k-3}$, with $F_0 = x_0, F_1 = x_1, F_2 = x_2$. For a string that consists of b_1 1's appended on the left to the first block of 0's, (using the above as starting values)

$$\begin{aligned} V_{00}^{b_1} &= \{F'_{a_1-2}, F'_{a_1-3}, 1\}_{b_1}, \\ V_{01}^{b_1} &= V_{00}^{b_1-2}, \\ V_{10}^{b_1} &= V_{00}^{b_1-1}, \\ V_{11}^{b_1} &= 0, \end{aligned}$$

where $F_k[2,3,4]$ has been abbreviated as F'_k, and the expression on the right hand side is defined as follows:

$$\{x_0, x_1, x_2\}_k = x_j, \quad j = 0,1,2, \quad k \equiv j \pmod 3.$$

Continuing the process leads to

$$N(S) = v_{b_n}^{(n)} + \{F_{a_n+1} + F_{a_n-1} + F_{a_n-2}\}[0, v_{b_n-1}^{(n)}, v_{b_n-1}^{(n)} + v_{b_n+1}^{(n)}],$$

where $v_j^{(n)} = v_k^{(n)}$, $k = 0,1,2$ and $j \equiv k \pmod 3$, and the vector $v^{(n)}$ is defined to be

$$\begin{aligned} v^{(n)} = {} & (v_{b_{n-1}}^{(n-1)}, F_{a_n-1}[0, v_{b_{n-1}-1}^{(n-1)}, v_{b_{n-1}-1}^{(n-1)} + v_{b_{n-1}+1}^{(n-1)}], \\ & F_{a_n-2}[0, v_{b_{n-1}-1}^{(n-1)}, v_{b_{n-1}-1}^{(n-1)} + v_{b_{n-1}+1}^{(n-1)}]), \end{aligned}$$

with $v^{(0)} = \{F'_{a_1-2}, F'_{a_1-3}, 1\}_{b_1}$ and $F'_k \equiv F_k[2,3,4] = F_{k-1} + F_{k-3}$.

The De Bruijn graph for Rule 22 is shown in figure 5f. Statement (V) in section 3 asserts that the periodic structure in the number of preimages arises from the existence of the simple circuit of length 3 and all edges with the value 1.

5. Applications of preimage formulae

The formulae for preimages derived in section 3 provide information on the statistical and dynamical features of cellular automata. Examples of uses of preimage formulae include the analysis of the complexity of automata rules, the calculation of specific statistical quantities that require knowledge of probabilities of sequences, and the characterization of the probability distribution of spatial sequences generated by one iteration of the rule. In this section, some applications of the preimage formulae are discussed.

5.1 Gardens-of-Eden

A finite-complement regular language is defined to be a regular language with a finite number of excluded blocks. Since blocks excluded after one iteration of a cellular automaton rule are termed "gardens-of-Eden" [1,4,5], a rule that generates a finite-complement regular language at time $t = 1$ has the property that there exists an $L < \infty$ such that any garden-of-Eden sequence of length greater than L contains a garden-of-Eden sequence of length $\leq L$. In [10], Wolfram describes a test for determining whether a rule generates a finite-complement regular language. The test involves determining the size Ξ of the minimal DFA graph associated with the language, and checking for excluded blocks of length $\leq 2\Xi$. If the only excluded blocks are those of length $< \Xi$, the language is a finite-complement language. If there exist excluded blocks of length greater than Ξ and less than 2Ξ, as well as of length $< \Xi$, then the language cannot be a finite-complement language. Since the size of the minimal DFA can be as great as 15 [see table 1 in [1]), the test requires in some cases that all strings of size up length 30 be checked.

The formulae derived in section 3 provide an analytical check of whether an automaton rule generates a finite-complement regular language after one iteration. In many cases, the answer is obvious. For instance, equation (4.11) clearly indicates that a sequence is a garden-of-Eden for Rule 126 iff $b_i = 1$ for some $1 < i < n$. Therefore, in this case, every garden-of-Eden contains the subsequence 010, and the sequences generated by one iteration of the rule constitute a finite-complement language.

In this section, the preimage formulae derived in section 3 will be used to provide proofs that a rule does or does not generate a finite-complement language for less obvious cases. The first case to be considered is Rule 22, defined in (4.16).

Proposition 1. *The sequences generated by one iteration of Rule 22 do not constitute a finite-complement language.*

Proof. Consider the infinite set of sequences $S_k \equiv 10\bar{S}0101$, where $\bar{S}$ denotes a sequence of $k = 3q$ 1's, for any value of q. Clearly, S_k does not contain any sequence S_j for $j < k$. Show that $N(S_k) = 0$ for all $k = 3q$, where $N(S)$ is the expression (19). For any S_k, the associated parameters are given by $a_1 = 0, a_2 = a_3 = a_4 = b_1 = b_2 = b_4 = 1$, $b_3 = k$. Then

$$\begin{aligned}
v^{(1)} &= [1, F_{-2}[2,3,4], F_{-3}[2,3,4]] = [1,1,1], \\
v^{(2)} &= [1, F_0[0,1,2], F_{-1}[0,1,2]] = [1,0,1], \\
v^{(3)} &= [0, F_0[0,1,2], F_{-1}[0,1,2]] = [0,0,1], \\
v^{(4)} &= [0, F_0[0,1,1], F_{-1}[0,1,1]] = [0,0,0],
\end{aligned}$$

and thus $N(S_k) = 0$ for all k, and the rule is not a finite-complement language. ∎

The next case to be considered is rule 54, defined by

$$\{000, 011, 110, 111\} \rightarrow 0, \quad \{001, 010, 100, 101\} \rightarrow 1.$$

In [10], on the basis of computational searches as described earlier in the section, Rule 54 is included in the set of rules that do not generate finite-complement languages after one iteration; it will be shown here that the reverse is true.

Proposition 2. *The sequences generated by one iteration of Rule 54 constitute a finite-complement language.*

Proof. The number of preimages for a sequence S under Rule 54 is given by

$$\begin{aligned} N(S) \;=\;& F_{b_n+3}[0, F_{b_{n-1}-2}[0, J(a_{n-1})F_{b_{n-2}} \cdots, J(a_2)F_{b_1}[0, 2 + J(a_1), 1 \\ & + J(a_1)], \cdots,] \\ & + J(a_n)F_{b_{n-1}}[0, J(a_{n-1})F_{b_{n-2}} \cdots, J(a_2)F_{b_1}[0, 2 + J(a_1), 1 \\ & + J(a_1)], \cdots,], \\ & J(a_n)F_{b_{n-1}}[0, J(a_{n-1})F_{b_{n-2}} \cdots, J(a_2)F_{b_1}[0, 2 + J(a_1), 1 \\ & + J(a_1)], \cdots,]], \end{aligned}$$

where $F_k = F_{k-3} + F_{k-2}$ and $J(x) = 1$ if $x > 1$ and $J(x) = 0$ otherwise.

It is easy to show using the above expression that the four sequences 10110, 10101, 01101, and 10111101 are gardens-of-Eden. Let S_n be any other sequence such that $N(S) = 0$ and $a_n \neq 0$. Show that S_n must contain one of the four sequences.

Assume the contrary. First consider the case that $J(a_i) = 0$ for $i = n, n-1$. For any m, define S_m to be the sequence that matches S_n in all entries starting from the right up to the mth block of consecutive 1's. From the preimage formula,

$$N(S_n) = F_{b_n+3}[0, F_{b_{n-1}-2}[0, Z, 0], 0]$$

where $Z = F_{b_{n-2}-2}[0, F_{b_{n-3}-2}[0, \cdots, 0], 0]$. Since S_n is assumed to contain no smaller garden-of-Eden, it follows that $Z \neq 0$ since otherwise $F_{b_{n-1}+3}[0, Z, 0] = 0$, and therefore $N(S_{n-1}) = 0$. But then $b_n + 3$ must assume one of the values -1, 0, or 2 in order for $N(S_n)$ to be equal to 0, and this contradicts the definition of the parameters b_i's. It follows therefore that $F_{b_{n-1}-2}[0, Z, 0] = 0$, and b_{n-1-2} must assume one of the values -1, 0, or 2; in other words, $b_{n-1} = 1$, 2, or 4. If $b_{n-1} = 1$, then the sequence begins on the left with the values 10101; if $b_{n-1} = 2$, then it begins with 101101; if $b_{n-1} = 4$, then it begins with 10111101.

Next suppose that $J(a_i) = 1$ for either $i = n$ or $i = n - 1$. It is easy to show using reasoning similar to that above that S_n must contain either the sequence 01101 or 10111101.

Finally, consider the case $J(a_n) = J(a_{n-1}) = 1$. Then

$$N(S_n) = F_{b_n+3}[0, X + Y, Y],$$

where

$$\begin{aligned} X &= F_{b_{n-1}-2}[0, Z_1 + Z_2, Z_2], \\ Y &= F_{b_{n-1}}[0, Z_1 + Z_2, Z_2], \\ Z_1 &= F_{b_{n-2}-2}[0, F_{b_{n-3}-2}[0, \cdots, 0], 0], \\ Z_2 &= F_{b_{n-2}}[0, F_{b_{n-3}-2}[0, \cdots, 0], 0], \end{aligned}$$

In order for $N(S)$ to be equal to zero, it must be true that $X = Y = 0$. It follows then that $b_{n-1} = 2$ and $Z_2 = 0$. The fact that $Z_2 = 0$ implies that $b_{n-2} = 2$ and either $J(a_{n-2}) = 0$ or $F_{b_{n-3}-2}[0, \cdots, 0] = 0$. If the former is true, then the case reduces to one of the cases discussed above. If the latter is true, then the argument can be continued either until $J(a_i) = 1$ for some i, or until the end of the sequence S_n is reached. If $J(a_i) = 1$ for all i, then in order for S_n to be a garden-of-Eden not containing any of the four smaller gardens-of-Eden, it must be true that $F_{b_1}[0, 3, 2] = 0$ or $b_1 = 0$, which is a contradiction. Therefore S_n cannot be a garden-of-Eden distinct from the four given sequences, and therefore the language is a finite-complement regular language.

Finally, note that the preimage formulae identify, in addition to the gardens-of-Eden, the sequences that are *not* excluded after one iteration of the rule, and therefore presumably are useful in the characterization of the regular languages that can be generated as the one time-step image of an automaton [11]. ∎

5.2 Computations of spatial measure entropy

The closed-form formulae of section 3 are useful for the calculation of quantities such as spatial metric entropy that require knowledge of the probabilities of occurrence for the entire set of spatial sequences of length n. In general, such calculations are limited by bounds on memory and are restricted to block length $n \leq 20$ [see, for example, table 6 of [1]). With the closed-form formulae, the only fundamental limits on such calculations are those of computation time.

The spatial measure entropy $h_\mu^{(x)}$ provides a measure of the "information content" of cellular automata spatial configurations, and is defined as the limit as $L \to \infty$ of the quantity

$$-\frac{1}{L} \sum p_i \log p_i, \tag{5.1}$$

where the sum is taken over all sequences of length L, each with probability of occurrence p_i. In the case that the initial condition is chosen randomly, the probability p_i of any sequence S is exactly determined by the number of preimages $N(S)$, and thus the one time-step spatial measure entropy can be evaluated for any finite length L.

As an example, consider Rule 12 defined by (4.4). The number of preimages for any string is shown in section 3 to be

$$\begin{aligned} N(S) &= 2(a_1+2), && \text{if } n=1, \\ &= 2(a_1+1)(a_n+1)\textstyle\prod_{i=2}^{n-1} a_i, && \text{if } n>1 \text{ and all } b_i = 1, \\ &= 0, && \text{otherwise.} \end{aligned}$$

Finite estimates of the spatial measure entropy can of course be obtained by generating all 2^L sequences of length L, and substituting the above expression into (5.1). A computationally more efficient formula is obtained by using techniques of generating functions and partitions of integers. In particular, decompose the set of all sequences of length L into (i) the sequence of all 0's, (ii) sequences containing exactly one 1, and (iii) sequences with more than one 1. Denote by $E^{(i)}$, $E^{(ii)}$, and $E^{(iii)}$, respectively, the associated terms contributed by each set of sequences to the expression in (5.1). In other words,

$$\begin{aligned} E^{(i)} &= 2(L+2)\log(L+2), \\ E^{(ii)} &= \sum_{p=0}^{L-1} 2(p+1)(q+1)\log[2(p+1)(q+1)] \\ E^{(iii)} &= \sum_{p=0}^{L-3}\sum_{q=0}^{L-p-3} 2(p+1)(q+1) \sum_{n=1}^{\frac{[m+1]'}{2}} \sum_{\gamma_n(m-n+1)} \frac{n!}{k_1!k_2!\cdots k_{m-n+1}!} \\ &\quad \times 1^{k_1}2^{k_2}\cdots(m-n+1)^{k_{m-n+1}} \\ &\quad \times \log[2(p+1)(q+1)1^{k_1}2^{k_2}\cdots(m-n+1)^{k_{m-n+1}}], \end{aligned}$$

where $m = L-p-q-3$ represents the maximum possible length of an "inner" sequence bounded on the left and right by p and q 0's, respectively, the term $[x]'$ is defined to be the largest odd integer $\leq x$, and $\gamma_d(c)$ is defined to be the set of compositions (unordered partitions) of the integer c into d parts; i.e., each member of $\gamma_d(c)$ is an unordered set of non-negative multiplicities $\{k_1, k_2, \cdots, k_d\}$ with $\sum_{i=1} k_i = d$, and $\sum_{i=1} ik_i = c$. The final term above thus represents a sum over all possible blocks of p 0's on the left and q 0's on the right, with the number of 1's in the middle varying from 1 to approximately half the length of the sequence (since 1's must be isolated). For each such sequence, the contribution to the entropy is given by the number of its preimages weighted by the number of ways to distribute the 0's. Then further calculations yield

$$\begin{aligned} E_L &= 2(L+2)\log(L+2), \\ &\quad + \sum_{p=0}^{L-1} 2(p+1)(q+1)\log[2(p+1)(q+1)] \\ &\quad + \sum_{p=0}^{L-1}\sum_{q=0}^{L-p-1} 2(p+1)(q+1)2^{L-p-q-3}\log[2(p+1)(q+1)] \end{aligned}$$

$$+\sum_{p=0}^{L-3}\sum_{q=0}^{L-p-3} 2(p+1)(q+1)\{(L-p-q-2)\log[L-p-q-2]$$
$$+\sum_{j=1}^{L-p-q-5} 2^{l-p-q-j-6}(L-p-q-j+3)j\log j$$
$$+2(L-p-q-4)\log(L-p-q-4)\},$$

as an approximation, using exact probabilities for sequences of length L, for the spatial entropy $h_\mu^{(x)}$. The above, its awkward appearance notwithstanding, represents a considerable computational saving. It should be noted in this context that the value $h_\mu^{(x)} = .51771$ for Rule 4 given in table 6 of [1] was obtained using the assumption that the rule generates spatial sequences with no correlations beyond 3 sites. Since higher-order correlations do in fact exist, the true value of the entropy is considerably lower. (For $L = 36$, for example, the above yields .5111 for the entropy.) The values obtained for E_L can be used to bound the true entropy from above, since the expression (5.1) monotonically decreases as a function of sequence length L.

5.3 Characterization of probability distributions

The formulae can be used to characterize the form of the probability distribution for spatial sequences generated by one iteration of a cellular automaton. As a prototypical example, consider Rule 12 whose preimage formula is given by (4.7) and whose distribution analysis, while simpler than most other rules, provides a broad outline for the general case. For a fixed sequence length L, define X to be a random variable with $X = N(S)$. The objective is to characterize the probability distribution of X as S varies over all possible 2^L sequences.

First let n, the number of blocks of 0's, be fixed, and let S^* be any sequence of length L with n blocks of 0's. Then the set of a_i's can be considered the "spacings" [12] between n random variables uniformly distributed over the integers $(0, L)$. It follows [13] that the random variable

$$Y = \log N(S^*) = \sum \log(a_i)$$

is asymptotically normal, implying that $N(S^*)$ is asymptotically log-normal.

The variable n ranges from 1 to $\frac{L}{2}$, with a probability density function given by

$$\begin{aligned} f(x) &= \frac{c[L-(x-1), x]}{2^L}, \\ &= \frac{C_{L-x,x-1}}{2^L}, \end{aligned} \tag{5.2}$$

where $c[p, q]$ is defined to be the number of compositions of p with exactly q parts (i.e., the number of ways to express p as the sum of q positive integers when order counts), and $C_{p,q}$ is the usual binomial coefficient. It follows therefore that the random variable X can be considered a "mixture" [14] of asymptotically log-normal random variables, with the mixture parametrized by the variable n with density function given by (5.2).

5.4 Identification of sequences with maximal probability

The preimage formulae can be used to find the sequences with the maximum probability under one iteration of the rule. Even cursory examination of computer simulations shows that rules vary greatly in this respect, with some rules favoring spatial sequences consisting of long runs of constant value, and others appearing to prefer sequences in which 1's and 0's are more evenly sprinkled. The propensity to generate sequences of a certain type is linked to qualitative features of the rule's dynamical behavior, such as the tendency of rules like Rule 126 to generate large "upside-down" triangles (whose bases consist of long runs of one value).

First consider Rule 12 whose preimage formula is given in (4.7). A straightforward calculation shows that, for a fixed length L, the expression $N(S)$ is maximized for $b_i = 1$ for all i and $a_1 = a_n = \alpha - 1$, $a_i = \alpha$ for $i \neq 1, n$, where α satisfies the equation

$$\alpha = e^{\frac{\alpha+1}{\alpha}}.$$

Since the a_i's are constrained to be integers, it follows that the maximal probability is attained for sequences of isolated 1's separated by blocks of 0's of length a_i with $2 \leq a_1, a_n \leq 3$ and $3 \leq a_i \leq 4$ for $i \neq 1, n$. (Note that the sequence of length L with maximal number of preimages is uniquely defined iff $L \equiv 1 \pmod 4$. The sequence in this case consists of a core of $\frac{L-5}{4}$ blocks of three 0's separated by single 1's, with a block of two 0's at each end; the number of preimages in this case is $2 \times 3^{\frac{L+3}{4}}$. For other values of L, multiple maximal sequences exist.)

The same reasoning implies that the sequence of alternating 1's and 0's has maximal preimages for Rule 76. On the other hand, for rules such as Rule 126 discussed in section 3 with preimage formula given by (4.7), it is easy to see from the expression for the number of preimages that the sequence of all 1's has $2F_L$ preimages, and this represents for fixed L, the maximal number of preimages.

In general, for rules whose preimage formulae involve products, telescopes, or alternating series of "pure" terms, the number of preimages is maximized for (short) sequences with blocks of uniform length. For rules whose preimage formulae involve products, telescopes, or alternating series of recurrence-type terms, preimages are maximized for the sequence consisting of a single block of either 1's or 0's (for those rules whose preimage formulae depend only on either the a_i's or the b_i's, defined in section 3), or a single block of alternating 1's and 0's (for those rules depending upon the d_i's, also defined in section 3). Thus the circuit structure of an automaton rule table can be used to determine immediately the sequences of maximal probability under one iteration of the rule.

Finally note that the preimage formulae determine the scaling of the maximal probability with sequence length. In the case, for instance, of Rules 12 and 76, the maximal probability scales as α^n, where n is the sequence length. In the case of Rule 22, on the other hand, the maximal probability scales as does the n−th Fibonacci number; i.e., as $c_1 r_1^n + c_2 r_2^n$.

5.5 Enumeration of preimages with constraints

The preimage formulae can be used with slight modification to enumerate for an arbitrary sequence the number N^* of its preimages that satisfy specified constraints. Examples of the types of constraints that can be incorporated are the enumeration of preimages in which:

(1) no 100 occurs;

(2) no 1000 occurs.

In the first case, with the constraint given in terms of a string (100) of length 3, the general system can be modified so as to preclude the occurrence of 10 followed by 00. In the second case, with the constraint in terms of a string (1000) of length 4, the first-order recurrence relations must be first re-expressed as second-order relations, and then modified so as to preclude the sequential appearance of 10, 00, and 00.

For example, consider the problem of finding N^* for preimages satisfying (1) or (2) for Rule 12. The full set of recurrence relations for this rule is given in section 4b. The constraint in (1) is satisfied by preimages enumerated by the system

$$\begin{aligned}
L_{00}^j &= (L_{00}^{j-1} + L_{01}^{j-1}) I_j(0), \\
L_{01}^j &= (L_{10}^{j-1} + L_{11}^{j-1}) I_j(1), \\
L_{10}^j &= L_{01}^{j-1} I_j(0), \\
L_{11}^j &= (L_{11}^{j-1} + L_{10}^{j-1}) I_j(0).
\end{aligned}$$

since the above relations omit from counting the preimages in which 10 occurs followed by 00. The solution to the above system is easily computed to be

$$\begin{aligned}
N^*(S) &= 6, && \text{if } S \text{ begins, ends with } 0, \\
&= 3, && \text{if } S \text{ begins with } 0, \text{ ends with } 1, \\
&= 4, && \text{if } S \text{ begins with } 1, \text{ ends with } 0, \\
&= 4, && \text{if } S \text{ begins, ends with } 1.
\end{aligned}$$

In the case of (2), the system is first re-expressed in second-order terms as

$$\begin{aligned}
L_{00}^j &= (L_{00}^{j-1} + L_{01}^{j-1}) I_j(0), \\
L_{01}^j &= (L_{10}^{j-1} + L_{11}^{j-1}) I_j(1), \\
L_{10}^j &= (L_{00}^{j-1} + L_{01}^{j-1}) I_j(0),
\end{aligned}$$

$$\begin{aligned} &= (L_{00}^{j-2} + L_{01}^{j-2} + L_{01}^{j-1})I_j(0), \\ L_{11}^j &= (L_{11}^{j-1} + L_{10}^{j-1})I_j(0). \end{aligned}$$

In order to satisfy the constraint, the third relation in the system is rewritten as

$$L_{10}^j = (L_{01}^{j-2} + L_{01}^{j-1})I_j(0),$$

which has the solution

$$N^*(S) = c\prod_{i=2}^{n-1}[1 + I(a_i)],$$

where $I(x) = 1$ iff $x = 1$, and

$$\begin{aligned} c &= 10, && \text{if } S \text{ begins, ends with 0 and } a_n > 1, \\ &= 15, && \text{if } S \text{ begins, ends with 0 and } a_n = 1, \\ &= 5, && \text{if } S \text{ begins with 0, ends with 1}, \\ &= 4, && \text{if } S \text{ begins with 1, ends with 0 and } a_n > 1, \\ &= 6, && \text{if } S \text{ begins with 1, ends with 0 and } a_n = 1, \\ &= 2, && \text{if } S \text{ begins, ends with 1} \end{aligned}$$

6. Summary

This paper is concerned with the enumeration of preimages for one-dimensional cellular automaton rules on infinite lattices. For a given rule and arbitrary spatial sequence of values, the preimage of the sequence is defined to be the set of tuples that are mapped by the rule onto the sequence. The number of preimages of a sequence can be interpreted as determining the a priori probability of occurrence of the sequence after one iteration of the rule applied to an initial condition with uniform measure.

Recurrence relations are presented here for finding the number of preimages of general spatial sequences. These relations group and count preimages according to their endtuples, and then, for any sequence, express the number of its preimages beginning (either on the left or right) with a particular endtuple in terms of the number of preimages beginning with other endtuples for its subsequences.

The preimage formulae for nearest-neighbor rules are found to be essentially one of the following types:

(A) constant;

(B) products of integers representing lengths l_i of blocks of consecutive "units," where units are either 1's and 0's, or combinations of 1's and 0's;

(C) products of integers representing the l_ith terms in Fibonacci-like sequences, where the l_i's are defined as above;

(D) terms in telescoping Fibonacci-like sequences where, on any level, the initial values are given by the l_ith terms in the sequences on "lower" levels;

(E) terms of periodic sequences whose elements are determined either by the lengths l_i or by the l_ith terms in Fibonacci-like sequences;

(F) solutions to more general recurrence relations.

Applications of the preimage formulae include the identification of excluded blocks (blocks of site values with no preimages), identification of rules that generate only a finite number of excluded blocks, computation of quantities such as spatial metric entropy that require knowledge of the probability (or equivalently, the number of preimages) of all possible sequences, and enumeration of preimages with specified constraints.

Finally, the preimage formulae are of interest because they directly relate the structure of a cellular automaton's rule table to the one time-step behavior generated by the automaton. As mentioned above, the combinatorial properties of the De Bruijn directed graph associated with a rule table determine the automaton's type of preimage formula, which in turn determines a number of central features of the rule's one time-step probability distribution. The combinatorial structure of the rule table can be shown, for example, to determine the propensity of the automaton after one iteration on a random initial condition to generate "runs" of consecutive 1's or 0's.

Appendix A.

Listed below are the preimage expressions for the 32 symmetric elementary rules. The notation is as follows:

(i) For a string S, let a_i be the number of 0's in the ith block of consecutive 0's counting from the right, and let b_i be the number of 1's in the ith block of consecutive 1's. It is assumed unless otherwise stated that S begins and ends with 0's, implying that either a_1 or a_n is equal to 0 if that is not the case.

(ii) From right to left, divide S into blocks of consecutive isolated 1's, and let d_i be the number of consecutive 1's in the ith block, and α be the number of distinct blocks of consecutive isolated 1's.

(iii) $F_k[x_0, x_1, \cdots, x_{n-1}]$ denotes the kth term in an nth degree recurrence relation with initial values given by $F_0 = x_0, F_1 = x_1, \cdots, F_{n-1} = x_{n-1}$.

(iv) Let $v = \{u_0, u_1, \cdots, u_{n-1}\}$. Then $v_k \equiv \{u_0, u_1, \cdots, u_{n-1}\}_k$ is defined to be the term u_j where $k \equiv j \pmod{n}$.

(v) $I(x)$ is an indicator function defined so that $I(x) = 1$ if $x = 1$ and $I(x) = 0$ otherwise.

To conserve space, the formulae are given only for the cases where the sequence begins and ends with a 0, and contains at least one 1; the other cases differ only slightly.

The numbers to the left of each preimage expression are the rule numbers given according to the labelling scheme of [1]. The full table of formulae for the 88 distinct rules is available upon request [9].

4,32 $N(S) = F_{a_1+3}[F_{a_n+2} + F_{a_n}\prod_{i=2}^{n-1} F_{a_i}$
where $F_k \equiv F_k[0,1,1,1] = F_{k-1} + F_{k-2} + F_{k-4}$.

18,72 $N(S) = (2F_{a_n} + F_{a_n-1})\prod_{j=2, j\ even}^{n-k_L-3} F_{a_{n-j}} F_{a_1^*} \prod_{i\in A\cup B} F_{a_i},$
$n - k_L$ *odd*,
$= \prod_{j=1, j\ odd}^{n-k_L-3} F_{a_{n-j}} F_{a_1^*} \prod_{A,B} F_{a_i},\ n - k_L$ *even*,
$= 0$ if some $b_i > 2$.
where $F_k \equiv F_k[0,1] = F_{k-1} + F_{k-2}$, and k_L, a_1^*, A, B defined in example (4d) of text.

22,104 $N(S) = v_{b_n}^{(n)} + \{F_{a_n+1} + F_{a_n-1} + F_{a_n-2}\}[0, v_{b_n-1}^{(n)}, v_{b_n-1}^{(n)} + v_{b_n+1}^{(n)}],$
where $v^{(n)} = [v_{b_{n-1}}^{(n-1)},\ F_{a_n-1}[0, v_{b_{n-1}-1}^{(n-1)}, v_{b_{n-1}-1}^{(n-1)} + v_{b_{n-1}+1}^{(n-1)}],$
$F_{a_n-2}[0, v_{b_{n-1}-1}^{(n-1)}, v_{b_{n-1}-1}^{(n-1)} + v_{b_{n-1}+1}^{(n-1)}],$
$v^{(0)} = \{F'_{a_1-2},\ F'_{a_1-3},\ 1\}_{b_1},$
where $F'_k \equiv F_k[2,3,4] = F_{k-1} + F_{k-3}$

36 $N(S) = 2F_{a_1+1}F_{a_n+1}\prod_{i=2}^{n-1} F_{a_i-1}$
where $F_k \equiv F_k[0,1] = F_{k-1} + F_{k-2}$

50 $N(S) = 2F_{d_1+2}\prod_{i=2}^{\alpha-1} F_{d_i+1} + I(b_n)F_{d_1+2}F_{d_\alpha-1}\prod_{i=1}^{\alpha-1} F_{d_i}[0,1]$
where $F_k \equiv F_k[0,1] = F_{k-1} + F_{k-2}$.

54 N(S) $= F_{b_n+3}[0, F_{b_{n-1}-2}[0, J(a_{n-1})F_{b_{n-2}} \cdots, J(a_2)F_{b_1}[0, 2$
$+J(a_1), 1 + J(a_1)], \cdots,]$
$+J(a_n)F_{b_{n-1}}[0, J(a_{n-1})F_{b_{n-2}} \cdots, J(a_2)F_{b_1}[0, 2$
$+J(a_1), 1 + J(a_1)], \cdots,],$
$+J(a_n)F_{b_{n-1}}[0, J(a_{n-1})F_{b_{n-2}} \cdots, J(a_2)F_{b_1}[0, 2$
$+J(a_1), 1 + J(a_1)], \cdots,]],$
where $F_k = F_{k-3} + F_{k-2}$, and $J(x) = 1 - I(x)$.

76 $N(S) = F_{d_1+2}[2F_{d_\alpha+2} + I(b_{n-1})F_{d_{\alpha-1}})\prod_{i=2}^{n-1} F_{d_j}$
where $F_k \equiv F_k[1,1] = F_{k-1} + F_{k-2}$.

90,150,204 $N(S) = 4$.

94,122 $N(S) = F_{b_n+3}[0, I(a_n)F_{b_{n-1}-2}[0, I(a_{n-1})F_{b_{n-2}-2}[0, \cdots,$
$F_{b_1}[0, I(a_1) + 1, I(a_1 + 1), 1]], \cdots,]]$
$+F_{b_{n-1}-3}[0, I(a_{n-1})F_{b_{n-2}-2}[0, \cdots, F_{b_1}[0, I(a_1)$
$+1, I(a_1 + 1), 1]], \cdots,]]$
where $F_k = F_{k-3} + F_{k-4}$.

108 $N(S) = F_{b_n+4}[(1 - I(b_{n-1}))F_{b_{n-1}+1}[(1 - I(b_{n-2}))F_{b_{n-2}}[\cdots,$
$(1 - I(b_2))F_{b_1+1}[(1 - I(b_1)), 0, 2 + I(a_1)]], \cdots,]]$
where $F_k = F_{k-2} + F_{k-3}$.

126,300 $N(S) = 2F_{b_n+1}F_{b_1+1}\prod_{i=2}^{n-1} F_{b_i-1}$
where $F_k \equiv F_k[0,1] = F_{k-1} + F_{k-2}$.

128,254 $N(S) = F_{a_1+2}F_{a_n+2}\prod_{i=2}^{n-1} F_{a_i-1}$
where $F_k \equiv F_k[0,0,1] = F_{k-1} + F_{k-2} + F_{k-3}$.

132,222 $N(S) = F_{a_n+3}[0, I(b_{n-1})F_{a_{n-1}}[0, I(b_{n-2})F_{a_{n-2}}[0, \cdots,$
$I(b_2)F_{a_2}[0, I(b_1)F_{a_1-1}, I(b_1)F_{a_1-1} + F_{a_1-2}]\cdots]],$
$I(b_{n-1})F_{a_{n-1}}[0, I(b_{n-2})F_{a_{n-2}}[0, \cdots,$
$I(b_2)F_{a_2}[0, I(b_1)F_{a_1-1}, I(b_1)F_{a_1-1} + F_{a_1-2}]\cdots]]$
$+F_{a_{n-1}-1}[0, I(b_{n-2})F_{a_{n-2}}[0, \cdots,$
$I(b_2)F_{a_2}[0, I(b_1)F_{a_1-1}, I(b_1)F_{a_1-1} + F_{a_1-2}]\cdots]]$.

146,182 $N(S) = F_{a_n+3}[0, I(b_n)F_{a_{n-1}-2}[0, I(b_{n-1})F_{a_{n-2}-2}[0, \cdots,$
$F_{a_1}[0, I(b_1) + 1, I(b_1 + 1), 1]], \cdots,]]$
$+F_{a_{n-1}-3}[0, I(b_{n-1})F_{a_{n-2}-2}[0, \cdots, F_{a_1}[0, I(b_1) + 1,$
$I(b_1 + 1), 1]], \cdots,]]$
where $F_k = F_{k-2} + F_{k-3}$.

160,250 $N(S) = F_{a_n+2}[0, 0, (I(b_{n-1}) + 1)F_{a_{n-1}-2}[0, 0, \cdots (I(b_1) + 1)$
$F_{a_1-3} + I(b_1)F_{a_1-4},$
$(2I(b_1) + 1)F_{a_1-3} + I(b_1)F_{a_1-4}], \cdots]]$
$+I(b_{n-1})F_{a_{n-1}-3}[0, 0, \cdots (I(b_1) + 1)F_{a_1-3} + I(b_1)F_{a_1-4},$
$(2I(b_1) + 1)F_{a_1-3} + I(b_1)F_{a_1-4}], \cdots]],$
$(2I(b_{n-1}) + 1)F_{a_{n-1}-2}[0, 0, \cdots (I(b_1) + 1)$
$F_{a_1-3} + I(b_1)F_{a_1-4},$
$(2I(b_1) + 1)F_{a_1-3} + I(b_1)F_{a_1-4}], \cdots]]$
$+I(b_{n-1})F_{a_{n-1}-3}[0, 0, \cdots (I(b_1) + 1)F_{a_1-3} + I(b_1)F_{a_1-4},$
$(2I(b_1) + 1)F_{a_1-3} + I(b_1)F_{a_1-4}], \cdots]],$

164,218 $N(S) = F_{a_n+3}[0, f_1^{(n-1)}, f_1^{(n-1)} + f_2^{(n-1)}, f_1^{(n-1)} + f_2^{(n-1)} + f_3^{(n-1)}]$
where $f_1^{(n-1)} = v_{b_{n-1}}^{(n-1)}, f_3^{(n-1)} = v_{b_{n-1}+1}^{(n-1)},$
$f_2^{(n-1)} = F_{a_{n-1}-2}[0, f_1^{(n-2)}, f_1^{(n-2)} + f_2^{(n-2)}, f_1^{(n-2)}$
$+f_2^{(n-2)} + f_3^{(n-2)}]$
$v^{(n-1)} = \{F_{a_{n-2}-3}[0, f_1^{(n-3)}, f_1^{(n-3)} + f_2^{(n-3)}, f_1^{(n-3)}$

$$+f_2^{(n-3)} + f_3^{(n-3)}]$$
$$F_{a_{n-2}-1}[0, f_1^{(n-3)}, f_1^{(n-3)} + f_2^{(n-3)}, f_1^{(n-3)} + f_2^{(n-3)}$$
$$+f_3^{(n-3)}]\},$$
and $v^1 = \{F'_{a_1-4}, F'_{a_1-2}\}, f_2^{(1)} = F'_{a_1-2}, F_k = F_{k-1} + F_{k-4}$,
and $F'_k \equiv F_k[2,3,4,5]$.

178 $N(S) = X_n + I(b_{n-1})X_{n-1} + I(a_n b_{n-1} b_{n-2})X_{n-2}$
$+I(a_n)K(b_{n-1})X_{n-3}$,
where $X_n = K(a_n)K(b_{n-1})X_{n-1} + [I(b_{n-2}b_{n-1})K(a_n)$
$+I(a_n a_{n-1})K(b_{n-2})]X_{n-2}$
$+I(a_n a_{n-1} b_{n-2} b_{n-3})X_{n-3}$,
with initial values $X_1 = I(a_1) + K(a_1)$,
$X_2 = K(b_1)X_1 + I(b_1) + I(a_1 a_2)$,
and $X_3 = K(b_2)X_2 + I(b_1 b_2)X_1 + I(a_2 a_3)[K(b_1)X_1 + I(b_1)]$,
and $I(x) = 1$ iff $x = 1$ and $K(x) = 1$ iff $x = 1$ or 2.

200,236 $N(S) = F_{a_1} F_{a_n+2} \prod_{i=2}^{n-1} F_{a_i}$,
where $F_k \equiv F_k[0,1] = F_{k-1} + F_{k-2}$

232 $N(S) = F_{a_n+2}[0, F_{b_{n-1}-1}[0, F_{a_{n-1}-2}[0, \cdots,$
$F_{a_2-2}[0, F_{b_1-1}[0, F'_{a_1-2}, F'_{a_1-2} + F'_{a_1-3}]$
$F'_{b_1-2}, F'_{a_1-2} + F'_{a_1-3}]], \cdots,]]$
$F_{b_{n-1}-2}[0, F_{a_{n-1}-2}[0, \cdots,$
$F_{a_2-2}[0, F_{b_1-1}[0, F'_{a_1-2}, F'_{a_1-2} + F'_{a_1-3}],$
$F_{b_1-2}[0, F'_{a_1-2}, F'_{a_1-2} + F'_{a_1-3}]], \cdots,]]],$
where $F_k = F_{k-1} + F_{k-3}$, and $F'_k = F_k[2,3,4]$.

Acknowledgments

The work was supported in part by the Applied Mathematics Program of the Department of Energy under contract no. KC-07-01-01.

References

[1] S. Wolfram, *Theory and Applications of Cellular Automata* (World Scientific Press, Singapore, 1986).

[2] J. Milnor, "Notes on surjective cellular automaton-maps," Institute for Advanced Study preprint (June 1984).

[3] K. Sutner, "On σ-automata," *Complex Systems*, **2** (1988) 1.

[4] S. Amoroso and G. Cooper, "The garden-of-Eden theorem for finite configurations," *Proc. Amer. Math. Soc.*, **26** (1970) 158.

[5] S. Skyum, "Confusion in the garden of Eden," *Proc. Amer. Math. Soc.*, **50** (1975) 332.

[6] K. Culik and Y. Sheng, *On the Limit Sets of Cellular Automata* (University of Waterloo Research Report CS-87-47).

[7] E. Jen, "Scaling of preimages in cellular automata," *Complex Systems*, **1** (1987) 1048.

[8] E. Jen, "Cylindrical cellular automata," *Commun. Math. Phys.*, **118** (1988) 569.

[9] E. Jen, *Table of Preimage Formulae for Elementary Rules* (Los Alamos National Laboratory Report No. LA-UR-88-3359, 1988).

[10] S. Wolfram, "Computation theory of cellular automata," *Commun. Math. Phys.*, **96** (1984) 15.

[11] L. Hurd, "Formal language characterization of cellular automaton limit sets," *Complex Systems*, **1** (1987) 69.

[12] R. Pyke, "Spacings," *J. Roy. Stat. Soc. Ser. B*, **7** (1962) 395.

[13] L. Holst, "Asymptotic normality of sum-functions of spacings," *Ann. Prob.*, **7** (1979) 1066.

[14] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2 (John Wiley and Sons, New York, 1966).