Complex Systems **3** (1989) 63–78

Formal Languages and Finite Cellular Automata

Mats G. Nordahl
Institute of Theoretical Physics, S-412 96 Göteborg, Sweden

Abstract. A one-dimensional cellular automaton rule with specified boundary conditions can be considered as acting simultaneously on all finite lattices, which gives a mapping between formal languages. Regular languages are always mapped to regular languages, context-free to context-free, context-sensitive to context-sensitive, and recursive sets to recursive sets. In particular, the finite time sets on finite lattices are regular languages. The limit set on finite lattices (the periodic set) is shown to be neither a regular nor an unambiguous context-free language for certain additive rules with chaotic behavior, and for rules that can simulate one of these additive rules through a finite blocking transformation. The relation between cellular automata on finite and infinite lattices is discussed.

1. Introduction

Cellular automata are simple extended dynamical systems, with discrete space and time, local interactions, and discrete degrees of freedom at each site. They have recently been studied extensively (see, e.g., the collection of reprints in reference [1]), both because they allow simulations of hydrodynamics, spin systems, and other physical systems to be implemented in massively parallel hardware, and since a thorough investigation of the simplest cases imaginable might conceivably reveal some universal properties of extended dynamical systems. Since cellular automata are discrete systems, methods and concepts from discrete mathematics are often useful in the analysis of their behavior.

Formal languages (e.g. [2]) have been used to describe the time evolution of infinite one-dimensional cellular automata [3]. A formal language is any set (usually infinite) of words consisting of symbols from a finite set Σ. This could for example be the set of all finite blocks of symbols appearing in the ensemble of infinite strings allowed at time t in the time evolution of a cellular automaton, starting with a random initial ensemble. These sets, the finite time sets of the cellular automaton, have been shown to be regular languages in the infinite one-dimensional case [3]. A regular language is a language which can be recognized by a finite automaton, a device with a finite number of internal states, and with state transitions labeled by symbols from the set Σ. The Chomsky hierarchy of regular, context-free, context-sensitive, and recursively enumerable (r.e.) languages classifies languages according to the complexity of the devices needed for their recognition. The class of context-free languages, which includes the class of regular languages as a subset, consists of the languages which can be recognized by a push-down automaton, which is an automaton equipped with a stack memory. The context-sensitive languages are recognized by linear bounded automata, i.e. Turing machines restricted to using an amount of work space proportional to the length of the input, and r.e. languages correspond to unrestricted Turing machines (see, e.g., [2] for details).

In this article, we consider cellular automata on finite lattices and characterize their time evolution in terms of formal language theory, and we also attempt to relate the results for finite and infinite lattices. When finite lattices are considered, it is natural to let the cellular automaton mapping act on all finite lattices of different lengths (but with identical boundary conditions) simultaneously. This gives a mapping between formal languages, and we show that regular languages are always mapped to regular, context-free to context-free, context-sensitive to context-sensitive, and recursive sets to recursive sets, which in particular means that the finite time sets on finite lattices are regular languages. We also consider the asymptotic behavior of cellular automata on finite lattices, where results such as non-regularity of the limit set (which in this case consists of all states on temporal cycles) can be shown for some simple rules with chaotic behavior.

One reason for studying finite systems is that one might attempt to relate computation theoretical properties of the infinite lattice cellular automaton mapping to the properties of the mapping on finite lattices, which in some cases are more easily derived (some of the subtleties involved in this approach are discussed in section 4). Another reason is that we would like to gain a better understanding of the transient behavior of these systems in the limit of infinite lattice size (though this paper is mainly focused on the stationary behavior). Transient behavior has been argued to be important for weak turbulence in extended dynamical systems (e.g. [4]).

In section 2 of this paper, the computation theoretical properties of the cellular automaton mapping on finite lattices are investigated. For fixed, periodic, and twisted boundary conditions, the level of the finite time set in the Chomsky hierarchy cannot increase in the time evolution, which in particular means that for a random initial ensemble, the finite time sets are regular languages.

In section 3 the asymptotic behavior of various simple cellular automaton rules on finite lattices is investigated. We prove that the finite lattice limit sets, i.e. the sets of all states on temporal cycles, of the additive rules 60, 90, and 150 (where Wolfram's nomenclature for cellular automaton rules [5] has been used) are neither regular nor unambiguous context-free languages. Furthermore, this result can be extended to all cellular automaton rules that can simulate one of these rules through a finite blocking transformation.

Some examples of periodic sets of class 2 rules on finite lattices are also given.

Section 4, finally, contains a discussion of the relations between cellular automata on finite and infinite lattices.

2. Finite time sets on finite lattices

We now assume that we are given a finite set of symbols Σ with k elements, and a local transformation $\varphi : \Sigma^{2r+1} \rightarrow \Sigma$, where r is the range of the transformation. Together with suitable boundary conditions, which are usually taken to be fixed or periodic, this defines a cellular automaton map $\varphi_{(n)} : \Sigma^n \rightarrow \Sigma^n$ on a finite lattice of arbitrary length n. If all finite lattice lengths with some specified boundary conditions are considered simultaneously, we get a cellular automaton mapping $\phi : \Sigma^* \rightarrow \Sigma^*$ (where Σ^* denotes the set of all finite strings over Σ), which means that we can consider the cellular automaton as acting on any set of finite strings, i.e. any formal language.

If all possible finite sequences are allowed as initial states, so that we start from the set $\Omega^{(0)} = \Sigma^*$ at $t = 0$, we can define the *finite time set* $\Omega^{(t)}$ as the set of sequences allowed at time t,

$$\Omega^{(t)} = \phi^t(\Omega^{(0)}), \tag{2.1}$$

and the limit set of the cellular automaton mapping as

$$\Omega^{(\infty)} = \cap_{t=0}^{\infty} \Omega^{(t)}. \tag{2.2}$$

When we consider all finite lattices simultaneously, the limit set coincides with the *periodic set*, i.e. the set of all configurations that belong to a temporal period, and the maximal invariant set, which is the union of all invariant sets [6]. For periodic boundary conditions, the finite lattice limit set is a subset of the infinite lattice limit set discussed in [3] and [7], and corresponds to those infinite sequences allowed asympotically for periodic initial states. Equivalently, this corresponds to the set of all spatially periodic configurations that lie on a period in time. It would be interesting to know to what extent this subset reflects the structure of the complete limit set for an infinite lattice; this article contains some results in this direction. This situation is rather similar to the case of ordinary chaotic dynamical systems, where important properties such as the topological entropy can be calculated from the set of all periodic orbits [8,9].

If the time evolution of a cellular automaton on an infinite lattice is considered, the finite time set $\Omega^{(t)}$ should be interpreted as the set of all n-blocks of symbols that occur in the ensemble of infinite strings allowed at time t, starting with an ensemble consisting of all possible strings at $t = 0$ [7]. This way of specifying a topological ensemble is analogous to the way a shift-invariant measure on bi-infinite strings is given by a set of block probability distributions satisfying Kolmogorov's consistency conditions [10].

With this random initial ensemble, the finite time sets on an infinite lattice are regular languages [3]. We shall show that this is also the case when the time evolution on the set of all finite lattices is considered. This is a particular case of a more general result shown below, which states that the cellular automaton map preserves trios. Trios are families of formal languages closed under intersection with a regular set, inverse homomorphism, and ε-free forward homomorphism (see [2]). Some examples are the families of regular languages, context-free languages, context-sensitive languages, and recursive sets. This means that a finite number of iterations of the cellular automaton map cannot take us to a higher level in the Chomsky hierarchy, and in particular, starting from an initial ensemble given by $\Omega^{(0)} = \Sigma^*$ (which is a regular language), all finite time sets are regular languages.

We prove this statement by constructing a generalized sequential machine (gsm) which simulates the cellular automaton map on finite strings. A well-known result in formal language theory which states that any ε-free gsm map preserves trios can then be used.

A *generalized sequential machine* is a finite automaton with both input and output associated with its state transitions. The output consists of a finite number of symbols (at least one for an ε-free gsm) from an output alphabet O, and the input is one symbol from an input alphabet I. For fixed boundary conditions given by the symbols s_L and s_R an ε-free gsm simulating the cellular automaton mapping can be constructed as follows: (for simplicity we have restricted ourselves to $r = 1$, but the construction below can easily be generalized to cellular automaton rules with arbitrary r)

Let the set of internal states be

$$K = \{S, \Sigma \times \Sigma, Z\}, \tag{2.3}$$

where S is the start state, Σ is the symbol set of the cellular automaton (which is here used to label states in the gsm), and Z labels an additional state; and let the input alphabet be $I = \Sigma \cup \{z\}$ and the output alphabet be $O = \Sigma \cup \{y, z\}$. If the transition rules of the gsm are chosen as (here a, b, c, and d represent arbitrary symbols in Σ, φ is the local cellular automaton rule, and the transitions between states are labeled by input/output)

$$\begin{array}{cc} S \overset{a/y}{\rightarrow} (s_L a) & S \overset{z/yz}{\rightarrow} Z \\ (ab) \overset{c/\varphi(abc)}{\rightarrow} (bc) & (ab) \overset{z/\varphi(abs_R)z}{\rightarrow} Z \\ Z \overset{a/z}{\rightarrow} Z & Z \overset{z/z}{\rightarrow} Z, \end{array} \tag{2.4}$$

this is an ε-free gsm which maps strings of the form σz (where $\sigma \in \Sigma^*$) to $y\phi(\sigma)z$. This construction is a modification to suit our purposes of a construction by Takahashi [6]. If the cellular automaton map $\sigma \rightarrow \phi(\sigma)$ is written as the composition $\sigma \rightarrow \sigma z \rightarrow f(\sigma z) = y\phi(\sigma)z \rightarrow \phi(\sigma)$, it is evident that it preserves trios; one can easily show that inserting and removing the markers y and z preserves trios, and as was mentioned above, an ε-free gsm map always preserves trios [2].

For periodic boundary conditions we can construct a simulating ε-free gsm in the following way:

Let the set of states be

$$K = \{S, \Sigma, \Sigma \times \Sigma \times \Sigma \times \Sigma, Z\}, \tag{2.5}$$

let the input and output alphabets be the same as in the case of fixed boundary conditions, and let the transition rules be: (there is a certain freedom of choice, since we are only interested in the effects of the gsm on inputs of the form mentioned below)

$$\begin{array}{cc} S \overset{a/y}{\rightarrow} (a) & S \overset{z/yyz}{\rightarrow} Z \\ (a) \overset{b/y}{\rightarrow} (ab, ab) & (a) \overset{z/yaz}{\rightarrow} Z \\ (ab, cd) \overset{e/\varphi(cde)}{\rightarrow} (ab, de) & (ab, cd) \overset{z/\varphi(cda)\varphi(dab)z}{\rightarrow} Z \\ Z \overset{a/z}{\rightarrow} Z & Z \overset{z/z}{\rightarrow} Z. \end{array} \tag{2.6}$$

This ε-free gsm simulates the cellular automaton by mapping every string of the form $\sigma z = \sigma_1 \ldots \sigma_n z$, where $n \geq 3$, to $f(\sigma z) = yy\varphi(\sigma_2 \ldots \sigma_n \sigma_1)z$. The markers y and z can be inserted and removed just as before. We are mainly interested in ensembles invariant under cyclic permutations, such as the finite time sets obtained from an initial ensemble $\Omega^{(0)} = \Sigma^*$, which correspond to translation invariant ensembles in the infinite lattice case. Then the cyclic permutation included in the gsm map above does not make any difference, and we can conclude that in this case trios are preserved also for periodic boundary conditions. We could also undo the cyclic permutation; the map $\sigma_1 \ldots \sigma_n \rightarrow \sigma_n \sigma_1 \ldots \sigma_{n-1}$ can be written as $\sigma_1 \ldots \sigma_n \rightarrow \sigma_n \ldots \sigma_1 \rightarrow \sigma_{n-1} \ldots \sigma_1 \sigma_n \rightarrow \sigma_n \sigma_1 \ldots \sigma_{n-1}$, and since the second step obviously can be performed by a gsm, and the reversal operation preserves regular, context-free, context-sensitive and recursive sets [2], the cellular automaton map with periodic boundary conditions also preserves these properties. This is also true for twisted boundary conditions [11]; in the gsm above it is only necessary to change the output of the transition rule $(ab, cd) \rightarrow Z$ for input equal to z to take the twisted boundary conditions into account. The construction could be generalized to $r > 1$ in a straight-forward manner by expanding the set of internal states so that $r + 1$ symbols were kept in memory, instead of two as above. The output of the gsm map would then be cyclically permuted r steps to the left, but this permutation can of course be reversed by iterating the procedure mentioned above r times.

This in particular means that the finite time sets $\Omega^{(t)}$ for finite lattices defined above are regular languages for all finite t (at least for fixed, periodic, and twisted boundary conditions), and more generally that regular languages are always mapped to regular languages, context-free to context-free, context-sensitive to context-sensitive, and recursive sets to recursive sets. The level in the Chomsky hierarchy thus cannot increase in the time evolution of a cellular automaton (though it of course can decrease, a trivial example is the rule which maps everything to 0). A similar statement is also valid for the cellular automaton mapping on infinite lattices [12]. This can be contrasted to other, more refined, measures of complexity, such as the algorithmic complexity, which is the number of states in the minimal deterministic finite automaton accepting $\Omega^{(t)}$ [3], or the effective measure complexity [13], which measures the rate of convergence of finite length block entropies, that in general increase in the time evolution of a cellular automaton [3,14]. On infinite lattices, the limit set may often be more complicated than the finite time sets, and this is also the case if we consider all finite lattices simultaneously. In the following section we shall see that certain properties of the limit set (periodic set), such as non-regularity, can be derived in the finite lattice case for some simple chaotic cellular automaton rules.

3. Limit sets on finite lattices

In this section, we shall primarily consider cellular automaton rules with $r = 1$ and $\Sigma = \{0, 1\}$, in particular additive rules and rules that can simulate an additive rule through a finite blocking transformation, and characterize their limit sets on finite lattices. It is in general a recursively unsolvable problem even to determine whether the limit set (periodic set) of a cellular automaton is a regular set [15], but in the particular cases we consider, we can show that the limit set is neither a regular nor an unambiguous context-free language.

One of the simplest examples of a cellular automaton rule (with $k = 2$ and $r = 1$) showing chaotic behavior is rule 90 (we label cellular automaton rules according to the conventions of Wolfram [5]), where the value of a site is given by the sum modulo 2 of its nearest neighbors at the preceding time step, $(a_i(t) = a_{i-1}(t-1) + a_{i+1}(t-1)) \bmod 2$. This is an additive rule, which means that the configurations satisfy an additive superposition principle, and many properties of the cellular automaton, such as the cycle structure, can then be determined algebraically. Additive rules, in particular rule 90, on finite lattices were extensively analyzed by Martin, Odlyzko, and Wolfram in reference [11], and some of their results will be useful here.

There are a number of conceivable ways of showing that a language (with an unknown grammar) is not a regular language, or not a context-free language (see e.g. [2,16]). One method frequently used is to prove that certain so called pumping lemmas characterizing regular and context-free languages are violated. For regular languages, the pumping lemma states that any string z in L longer than some constant N (depending on L) can be written as the concatenation of three strings, $z = uvw$, where the string uv^kw belongs to L for any integer $k \geq 0$. This occurs because any regular language is accepted by a finite automaton, and the constant N can then be chosen so that for any string of length larger than N, some state in the accepting automaton must be visited more than once. This gives a closed circuit of states in the automaton, which may be repeated an arbitrary number of times k, always giving acceptable words in L, since the initial and final states are unchanged (and are thus allowed as initial and final states). There is also a pumping lemma for context-free languages.

An alternative approach is to examine the analytic properties of various generating functions associated with the language L [17]. If the number of words of length n in L, i.e. the growth function of L, is denoted $g(n)$, we can for example introduce a generating function (the structure generating function of L [18])

$$G(z) = \sum_{n=0}^{\infty} g(n)\, z^n. \tag{3.1}$$

For a regular language, this generating function is always a rational function, and for an unambiguous context-free language, $G(z)$ is an algebraic function [19,20]. A context-free language is unambiguous if it has a grammar where each string has a unique derivation. Inherently ambiguous context-free languages exist, and some of these are known to have transcendental structure generating functions [21].

Regular and unambiguous context-free languages can thus be excluded by showing that $G(z)$ is transcendental. This can be done in several ways, for example by showing that $G(z)$ has an infinite number of singularities, or by considering the asymptotic behavior of the Taylor coefficients $g(n)$. The Taylor coefficients of a rational function satisfy a linear difference equation, and are thus of the form

$$g(n) = P_1(n)\lambda_1^n + P_2(n)\lambda_2^n + \ldots + P_N(n)\lambda_N^n \tag{3.2}$$

(where the $P_i(n)$ are polynomials in n). Using this expression one can classify the different forms of asymptotic behavior that are allowed for the regular language growth function (e.g. [22]). In typical cases, the number of words of length n in a regular language asymptotically increases exponentially. An oscillating factor may occur if the characteristic equation of the difference equation has complex roots. To illustrate this by a simple example, let us consider the set of Garden of Eden configurations (states that cannot be reached in the time evolution, and thus only can occur as initial states) for rule 90 on finite lattices, with periodic boundary conditions. This set is the complement of the set of configurations allowed at $t = 1$, the finite time set $\Omega^{(1)}$. In the previous section, the finite time sets on finite lattices were shown to be regular languages, and the complement of a regular language is regular, which means that for any cellular automaton rule, the set of Garden of Eden configurations is a regular language. The number of states allowed at time t for rule 90 can be found in reference [11], and also the number of Garden of Eden configurations for a lattice of length n, which equals (for $n \geq 3$)

$$g(n) = (5 + (-1)^n)2^{n-3}, \tag{3.3}$$

which gives a rational structure generating function

$$G(z) = \frac{4z^3(1+3z)}{1-4z^2}. \tag{3.4}$$

The topological entropy of the language is given by the negative logarithm of the smallest positive root of the denominator polynomial of the structure generating function. In this particular case the topological entropy is equal to 1.

If a formal language L is required to define a translation invariant ensemble of infinite strings in a consistent manner, as in the case of the finite time sets $\Omega^{(t)}$ on infinite lattices, then all substrings of a particular string in L are also words in L, and for any string σ in L, at least one of the strings obtained by adding a symbol from Σ to the right (and to the left) of σ necessarily belongs to L. In this case the growth function must be a monotonously increasing function.

For unambiguous context-free languages, the generating function $G(z)$ is algebraic, which means that its Taylor coefficients $g(n)$ satisfy an algebraic recursion relation [23], and that generically their asymptotic behavior is given by: (where κ is a rational number, and λ, c_i, and ω_i are algebraic, with $|\omega_i| = 1$)

$$g(n) \sim c\, n^{\kappa} \lambda^n (\Sigma_i\, c_i\, \omega_i). \tag{3.5}$$

Let us now apply these results to the periodic sets of some simple cellular automaton rules, where the growth function can be obtained algebraically, such as additive rules with $k = 2$ and $r = 1$. We begin by considering rule 90. The growth function $g_P(n)$ of the periodic set, i.e. the total number of states on cycles, was derived for rule 90 (with periodic boundary conditions) in reference [11]. For lattices of odd length n, a fraction $1/2$ of all states lie on cycles, which means that for odd n, $g_P(n) = 2^{n-1}$. For n even, the fraction of configurations on cycles is $1/(2^{D_2(n)})$, where $D_2(n)$ is the maximal power of two that divides n. Thus $g_P(n) = 2^{n-D_2(n)}$ for even n (and if we define $D_2(2m+1) = 1$, this expression is valid for all n). In particular, for lattices of length equal to a power of two the zero configuration is a unique fixed point, which means that $g_P(2^n) = 1$. This gives a generating function

$$\begin{aligned} G(z) &= \sum_{n=3}^{\infty} 2^{n-D_2(n)} z^n \\ &= \sum_{m=0}^{\infty} \frac{z^{2^m}}{1-(2z)^{2^{m+1}}} - z(1+z), \end{aligned} \tag{3.6}$$

which is quite reminiscent of some classical examples of lacunary series (e.g. [24]). On the radius of convergence, $|z| = 1/2$, $G(z)$ is singular for (1/2 times) every 2^{m+1}th root of unity, which gives singularities on a dense set on the radius of convergence, and $|z| = 1/2$ can then be shown to be a natural boundary. The existence of a natural boundary implies that $G(z)$ is a transcendental function. This in turn implies that the limit set of rule 90 on finite lattices (with periodic boundary conditions) is neither a regular nor an unambiguous context-free language.

This could also have been seen from the irregular asymptotic behavior of the coefficients $g_P(n)$, which cannot correspond to the product of an exponential (and/or a power law) and an oscillating factor.

An additive rule closely related to rule 90 is rule 60, where instead $a_i(t) = (a_{i-1}(t-1) + a_i(t-1)) \bmod 2$. The time evolution of rule 90 on a lattice of even length n is equivalent to the evolution of two independent copies of rule 60 on lattices of length $n/2$, which means that the number of states on cycles for rule 60 on a lattice of length n satisfies $g_{P_{60}}(n) = (g_{P_{90}}(2n))^{1/2} = g_{P_{90}}(n)$. Since the structure generating function of the periodic set then is identical to that of rule 90, we conclude that also in the case of rule 60, the periodic set is neither regular nor unambiguous context-free.

There is one other symmetric additive CA rule (with $k = 2$ and $r = 1$) that shows chaotic behavior, namely rule 150, for which $a_i(t) = (a_{i-1}(t-1) + a_i(t-1) + a_{i+1}(t-1)) \bmod 2$. If the methods of reference [11] are applied to this case, one finds (using periodic boundary conditions) that for lattice lengths n not divisible by three, all states lie on cycles, while if n is a multiple of three, the fraction of configurations on cycles is $1/(2^{D_2(n)})$. This gives a generating function

$$\begin{aligned} G(z) &= \sum_{n=1}^{\infty}(2^{3n+1}z^{3n+1} + 2^{3n+2}z^{3n+2}) + \sum_{n=1}^{\infty} 2^{3n-D_2(n)}z^{3n} \\ &= \frac{16z^4(1+2z)}{1-8z^3} + \sum_{m=0}^{\infty} \frac{2^{2m+1}z^{3\cdot 2m}}{1-(2z)^{3\cdot 2m+1}}. \end{aligned} \tag{3.7}$$

An argument similar to that above shows that $G(z)$ is transcendental, which means that the periodic set of rule 150 (with periodic boundary conditions) is neither a regular nor an unambiguous context-free language.

Periodic sets that are regular languages are typically found for class 2 cellular automata. These are cellular automaton rules that asymptotically simulate a shift map (or a power of the shift map, if the asymptotic behavior is periodic) for almost all initial states. Often both right and left shift maps can be simulated, even though one of these corresponds to a dominating fraction of the initial states. The limit set, which includes behavior occurring with vanishing probability, then typically includes both these simulation sets, and often (for asymmetric rules) a set of measure zero of infinite sequences with a boundary separating different asymptotic behavior to the right and to the left. The limit sets and attractors of class 2 rules on infinite lattices are discussed more extensively in reference [25], here we just intend to give a few examples of finite lattice periodic sets.

The simplest examples are given by cellular automaton rules with bounded transient length on an infinite lattice. Then all spatially periodic infinite sequences allowed at arbitrarily large times lie on temporal cycles, and the (finite lattice) periodic set consists of all blocks of symbols in the infinite lattice limit set $\Omega^{(\infty)}$ that correspond to part of some closed circuit of states in the automaton accepting $\Omega^{(\infty)}$. The periodic set is thus equivalent to the union of the irreducible subsets of $\Omega^{(\infty)}$. These are the subsets where any state can be reached from any other, assuming that the finite automaton is constructed in such a way that all states are allowed as initial states, which is often convenient when ensembles of infinite strings are discussed (irreducibility then corresponds to the stationarity requirement on ensembles of reference [13]). In the cases where $\Omega^{(\infty)}$ is a finite time set, it is necessarily irreducible, and we conclude that the periodic set P on finite lattices then is identical to $\Omega^{(\infty)}$. This applies to many of the simplest class 1 and class 2 rules; for $k = 2$ and $r = 1$ the rules 0, 1, 2, 3, 4, 5, 8, 10, 12, 19, 24, 29, 34, 36, 38, 42, 46, 72, 76, 108, 138, 200 (and the trivial surjective cases 15, 51, 170, 204 and 240), and their reflection and/or conjugation equivalents, all reach their limit set at $t \leq 2$ (see the table in reference [26]), and it can be checked that this is also the maximal length of transient behavior in these cases.

A rather trivial example of a CA rule with unbounded transients on an infinite lattice (and linearly growing maximal transient length on finite lattices) is rule 128, where the block 111 is mapped to 1, and all other length three blocks are mapped to 0. On finite lattices with periodic boundary conditions all configurations except 1^* are eventually mapped to the fixed point 0^*, which means that the periodic set is $P = 0^* \cup 1^*$ (which corresponds to the irreducible subsets of the infinite lattice limit set, which can be found in [3]), and the structure generating function of P is then given by

$$G(z) = \frac{2z^3}{1-z} \tag{3.8}$$

(which obviously corresponds to zero topological entropy). The notation a^* used above stands for an arbitrary number (including zero) of repetitions of the string a, but when cellular automata with periodic boundary conditions are considered, we assume that strings of overall length smaller than $2r + 1$ have been excluded.

A slightly less trivial example is given by rule 44, where the blocks 101, 011, and 010 are mapped to 1, and all other length three blocks to 0. On an infinite lattice, almost all initial configurations eventually approach a stable state consisting of single symbols 1, isolated by at least two symbols 0. The infinite lattice limit set, which consists of those sequences that have predecessors arbitrarily far back in time, can be explicitly constructed (see [25] for details). In particular, the block 0110 uniquely determines its predecessor blocks in a backward light cone (a periodic pattern of the form $(011)^*$ is obtained inside the light cone), which means that on a finite lattice, this block can only occur in the limit set in a configuration of the form $(011)^*$, up to cyclic permutations (when periodic boundary conditions are considered, we shall tacitly assume that all cyclic permutations of words are included in a language, an operation which is known to preserve trios [2]). On an infinite lattice, the block 111 may occur as a boundary between different forms of asymptotic behavior to the right and the left, a phenomenon which cannot occur on a finite lattice. Furthermore, the block 1111 is forbidden at $t \geq 2$. The periodic set can then be shown to be the union of the fixed point set $0^* \cup (1000^*)^*$, and all states of the form $(011)^*$, which have period three. This means that $P_{44} = 0^* \cup (011)^* \cup (1000^*)^*$ (and cyclic permutations). The fixed point set has a growth function $g(n)$ which obeys

$$g(n) = g(n-1) + g(n-3), \quad g(3) = 4, g(4) = 5, g(5) = 6, \tag{3.9}$$

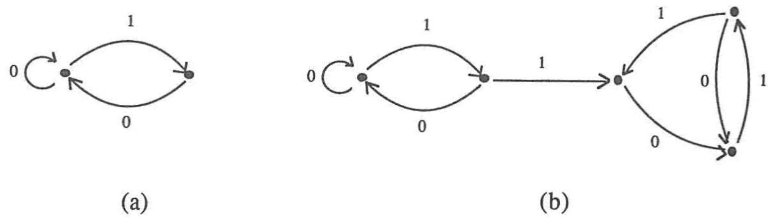

Figure 1: Finite automata giving (a) the attractor (b) the complete limit set of CA rule 56.

and if we include the period three states, we obtain a rational structure generating function for the complete periodic set,

$$G(z) = \frac{z^3(z^2+z+4)}{1-z-z^3} + \frac{3z^3}{1-z^3} \tag{3.10}$$

The periodic set and the infinite lattice limit set both have a topological entropy equal to that of the fixed point set, $s(0) = \log_2 1.4656 = 0.5515$.

As a final example of class 2 rules we consider rule 56, where the blocks 101, 100, and 011 are mapped to 1, and all other length three blocks to 0. Here almost all initial states are attracted to a simulation of a right shift on sequences containing only isolated symbols 1. The attractor (also discussed in [27]) is represented as a finite automaton in figure 1(a). The complete limit set (see [25] for details) is shown in figure 1(b), in a representation where all nodes in the finite automaton are allowed start nodes. The limit set also includes a set of sequences corresponding to a left shift, and a set of measure zero of sequences with different asymptotic behavior to the right and the left.

The periodic set on finite lattices with periodic boundary conditions corresponds to the union of the two irreducible subsets of figure 1(b), which yields a structure generating function (note that this time the two sets are not disjoint, since a right shift cannot be distinguished from a left shift on sequences of the form ...01010101....)

$$G(z) = \frac{z^3(3z+4)}{1-z-z^2} + \frac{z^3(10z^3+6z+3)}{1-z^2-z^3} - \frac{2z^4}{1-z^2}. \tag{3.11}$$

The topological entropy of the limit set is in this case identical to that of the attractor, $s(0) = \log_2 1.618 = 0.696$, both on finite and infinite lattices.

Finally, the results for additive rules can be extended to all cellular automaton rules that can simulate one of the additive rules discussed above through a finite blocking transformation. A cellular automaton rule R_1 (with $k = 2$) simulates another, R_2, if the evolution of R_1 on sequences consisting of the finite blocks of symbols B_0 and B_1 corresponds to the evolution of R_2 on sequences over Σ under the mapping $B_0 \to 0$, $B_1 \to 1$. A mapping $\Sigma \to \Delta^*$ (where Δ is some set of symbols) is a homomorphism of languages if the image of $a \in \Sigma$ is a single string $h(z) \in \Delta^*$. Many classes of formal languages are closed under the inverse of a homomorphism; this requirement was for example included in the definition of a trio (see section 2), which means that, e.g., the regular languages are closed under inverse homomorphisms. It can also be shown that the class of unambiguous context-free languages is closed under inverse homomorphisms (to check whether a string a is an element of $h^{-1}(L)$, a push-down automaton could construct $h(a)$ and determine whether it is an element of L [2]). We shall also need the operation of taking the intersection of a language with a regular set. The class of regular languages is closed under this operation (this property was also included in the definition of a trio), and this is also the case for the unambiguous context-free languages (this can be shown by constructing a push-down automaton which runs the accepting automata of the two languages in parallel [2]).

If a cellular automaton rule R_1 simulates one of the additive rules mentioned above through a finite blocking transformation $B_0 \to 0$, $B_1 \to 1$, the limit set of the additive rule can be obtained from the limit set (periodic set) P_1 of the rule R_1. We first form the intersection $P_1 \cap B$ of P_1 and the regular language B consisting of all possible strings made out of the blocks B_0 and B_1, $B = (B_0^* B_1^*)^*$. We then get the limit set of the additive rule through the inverse of the homomorphism $0 \to B_0$, $1 \to B_1$. Both of the operations used, i.e. inverse homomorphism and intersection with a regular set, are closure operations of the classes of regular and unambiguous context-free languages. Then, since the limit set of the additive rule is neither regular nor unambiguous context-free, this is necessarily the case also for the limit set of R_1.

We thus conclude that any cellular automaton rule that can simulate one of the rules 60, 90, or 150 has a periodic set which is neither a regular nor an unambiguous context-free language. This in particular applies to the ($k = 2$ and $r = 1$) rules 18, 22, 26, 94, 122, 146, 154, and 164, which are known to be capable of simulating rule 90 (e.g. [26]), and to the rules related to one of these through conjugation and/or reflection, and also to rule 105, which can simulate rule 150. Examples of cellular automaton rules where the argument does not apply, but where the irregular behavior of the total number of states on cycles indicates that the result should still be valid, are the non-linear surjective rules 30 (discussed in [28]) and 45, and rules 73 and 110.

4. Discussion: The relation between finite and infinite lattices

Some of the results of the previous section might seem slightly surprising, in view of the fact that the additive rules considered (rules 60, 90, and 150) are surjective on an infinite lattice, which means that all sequences are allowed at all times. The limit set of one of these rules on an infinite lattice is thus the set Σ^* of all possible sequences, which is obviously a regular language.

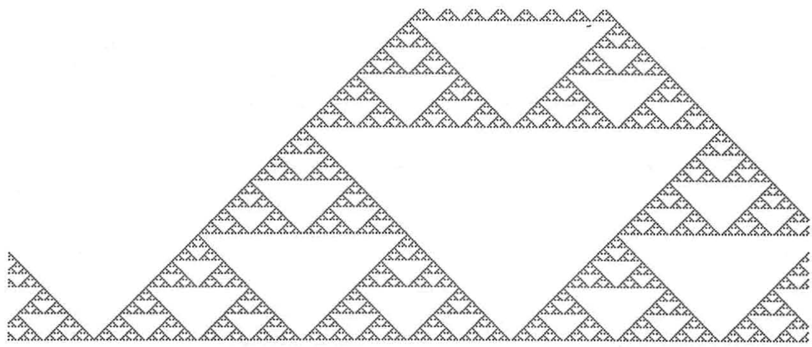

Figure 2: Part of an infinite evolution pattern of CA rule 90, illustrating how the spatial periods of the predecessors of the state ...01010101... diverge with distance back in time.

In such cases, where the limit set $\Omega^{(\infty)}$ of a cellular automaton on an infinite lattice is a regular language, the infinite, spatially periodic states in $\Omega^{(\infty)}$ correspond to all closed circuits of states in the finite automaton accepting the limit set, and they are thus obtained from a regular language. But the periodic set on finite lattices, on the other hand, which is a subset of the limit set in the infinite lattice case, corresponds only to those spatially periodic states that are actually on a temporal cycle. A spatially periodic state could also have predecessors arbitrarily far back in time (and thus be included in the limit ensemble), where the spatial period necessarily increased without bound as its ancestors were traced back in time. The subset of spatially periodic states on cycles need not have a description in terms of a regular language, even though the union of that set, and those periodic states in the limit set corresponding to infinite transients, is a regular language. This is what happens for the additive rules in question, where the periodic states on infinite transients simply correspond to the complement of the periodic set. Both these sets are non-regular, though their union is a regular language.

It is not hard to find explicit examples of this phenomenon. Consider for example rule 90, and a spatially periodic infinite state of the form ...010-10101010101.... This state is not on a temporal cycle, since it is mapped to the fixed point ...00000000000000... in one time step, but it still has predecessors arbitrarily far back in time (after all, the rule is surjective, so this is true for all states). One possible choice of predecessors for the sequence $(10)^\infty$ is given by the infinite Sierpinski gasket partially shown in figure 2, which is constructed so that the sequence $(10)^\infty$ has predecessors given by $(10^{2^{n+1}-1})^\infty$ at instances 2^{n-1} steps back in time, which shows that asympotically the spatial period diverges linearly as predecessors are traced back in time. Since each infinite sequence has exactly 4^t predecessors t steps back in time for rule 90, a simple counting argument shows that the linearly increasing spatial period of the predecessors is a generic feature in this case.

The relation between the computation theoretical properties of the periodic set and the infinite lattice limit set thus depends on the nature of the CA rule considered; only for rules where every spatially periodic state is on a temporal cycle can we immediately relate the two, as was discussed in section 3. We are however not aware of any example with a regular periodic set and a non-regular limit set on an infinite lattice.

It can finally be noted that even though the periodic set does not have the same computation theoretical properties as the infinite lattice limit set e.g. for rule 90, it does have topological entropy 1, so that the topological entropies are identical. This was also the case for the class 2 rules considered. One may in general conclude that the topological entropy of the periodic set on finite lattices is smaller than or equal to that of the infinite lattice limit set, since it corresponds to a subset of the initial states on an infinite lattice.

References

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

[2] J.E. Hopcroft and J.D. Ullman, *Introduction to Automata Theory, Languages and Computation* (Reading, MA: Addison-Wesley, 1979).

[3] S. Wolfram, "Computation Theory of Cellular Automata", *Communications in Mathematical Physics*, **96** (1984) 15–57.

[4] K. Kaneko, "Pattern Dynamics in Spatiotemporal Chaos. Pattern Selection, Diffusion of Defect and Pattern Competition Intermittency", *Physica* **D34** (1989) 1–41.

[5] S. Wolfram, "Statistical Mechanics of Cellular Automata", *Reviews of Modern Physics*, **55** (1983) 601–644.

[6] H. Takahashi, "The Maximum Invariant Set of an Automaton System", *Information and Control*, **32** (1976) 307–354.

[7] L.P. Hurd, "Formal Language Characterizations of Cellular Automaton Limit Sets", *Complex Systems*, **1** (1987) 69–80.

[8] D. Auerbach, P. Cvitanovic, J.P. Eckmann, G. Gunaratne, and I. Procaccia, "Exploring Chaotic Motion Through Periodic Orbits", *Physical Review Letters*, **58** (1987) 2387–2389.

[9] P. Cvitanovic, "Invariant Measurement of Strange Sets in Terms of Cycles", *Physical Review Letters*, **61** (1988) 2729–2732.

[10] M. Denker, C. Grillenberger, and K. Sigmund, *Ergodic Theory on Compact Spaces*, Lecture Notes in Mathematics 527 (Springer-Verlag, Berlin Heidelberg, 1976).

[11] O. Martin, A.M. Odlyzko, and S. Wolfram, "Algebraic Properties of Cellular Automata", *Communications in Mathematical Physics*, **93** (1984) 219–258.

[12] M.G. Nordahl, in preparation.

[13] P. Grassberger, "Towards a Quantitative Theory of Self-Generated Complexity", *International Journal of Theoretical Physics*, **25** (1986) 907–938.

[14] K. Lindgren and M.G. Nordahl, "Complexity Measures and Cellular Automata", *Complex Systems* **2** (1988) 409–440.

[15] H. Takahashi, "Undecidable Questions About the Maximum Invariant Set", *Information and Control*, **33** (1977) 1–34.

[16] W.M. Conner, "The Dimension of a Formal Language", *Information and Control*, **29** (1975) 1–10.

[17] P. Flajolet, "Ambiguity and Transcendence", in ICALP 85, Proceedings of the International Colloqium on Automata, Languages and Programming, *Lecture Notes in Computer Science* **194** (Springer-Verlag, Berlin Heidelberg, 1985) 179–188.

[18] B.B. Mandelbrot, "On Recurrent Noise Limited Coding", in *Brooklyn Polytechnic Institute Symposium on Information Networks*, ed. E. Weber, (Interscience, New York, 1955) 205–221.

[19] N. Chomsky and M.P. Schützenberger, "The Algebraic Theory of Context-Free Languages", in *Computer Programming and Formal Systems* (North-Holland, Amsterdam, 1963) 118–161.

[20] W. Kuich, "Entropies of Context-Free Languages", *Information and Control*, **16** (1970) 173–200.

[21] R. Kemp, "A Note on the Density of Inherently Ambiguous Context-Free Languages", *Acta Informatica*, **14** (1980) 295–298.

[22] O.H. Ibarra and B. Ravikumar, "On Sparseness, Ambiguity and Other Decision Problems for Acceptors and Transducers" in STACS 86, *Lecture Notes in Computer Science* **210** (Springer-Verlag, Berlin Heidelberg, 1986) 171–179.

[23] L. Comtet, "Calcul pratique des coefficients de Taylor d'une fonction algébrique", *L'Enseignement Mathmatique*, **10** (1964) 267–270.

[24] E. Hille, *Analytic Function Theory*, vol. 1 (Ginn and Company, Boston MA, 1959).

[25] M.G. Nordahl, "Limit Sets and Attractors of Class Two Cellular Automata", *Institute of Theoretical Physics*, Gteborg, preprint (1988).

[26] L.P. Hurd, in appendix to reference [1].

[27] W. Li, "Power Spectra of Regular Languages and Cellular Automata", *Complex Systems*, **1** (1987) 107–130.

[28] S. Wolfram, "Random Sequence Generation by Cellular Automata", *Advances in Applied Mathematics*, **7** (1986) 123–169.