Complex Systems **2** (1988) 409–440

Complexity Measures and Cellular Automata

Kristian Lindgren
Physical Resource Theory Group, Chalmers University of Technology, S-412 96 Göteborg, Sweden

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

Abstract. Various observables measuring the complexity of an ensemble of patterns are discussed, in particular statistical quantities related to the convergence of block entropies, and computation theoretical quantities related to a grammatical description of the ensemble. These measures of complexity are applied to one-dimensional cellular automata, by characterizing the time evolution of the probability measure on configuration space in terms of stochastic finite automata. In particular it is found that the effective measure complexity increases linearly in time for an additive rule with a random initial state with density $p \neq 1/2$. Some results on the convergence of block entropies for regular languages are shown, and context-free languages are also discussed. These results are used in an attempt to interpret the critical exponents observed by Grassberger in the convergence of block entropies for certain chaotic cellular automaton rules.

1. Introduction

Many of the systems encountered in physics, biology, and other fields consist of large numbers of fairly simple components that can produce very complex behavior when acting together. Even in simple model systems such as cellular automata [1] and chaotic low-dimensional dynamical systems [2] quite complex behavior can be seen, both in the sense that the individual patterns or trajectories generated may be effectively random, or may show signs of complicated structure in the form of long-range correlations, and in the sense that the ensemble of allowed patterns may be very hard to describe. In general, for an observer facing a complex situation, whether it is a physicist attempting to understand a new phenomenon, the brain confronted with maybe 10^9 bits of sensory data each second, a child learning language, or someone listening to Schönberg's woodwind quintet for the first time, a reasonable strategy would be to attempt, in one way or another, to model the situation. This involves extracting generic features, and separating them from the noise, or the specific information contained in individual patterns. The result might be a grammar for the allowed patterns in the discrete case [3], or an approximate model of the equations of motion in the continuous case [4,5]. Sometimes the grammar need not be explicitly known, but may rather be implicitly contained in a pattern recognizing structure, such as a parser.

In this article we shall consider a number of observables that measure either information or complexity. In a sense this is far more trivial than the general problem of pattern recognition and inductive inference, but the problems are related, since some of the quantities we consider as measures of complexity are properties of a model, such as a grammar for a set of strings. Some of these observables will in particular be applied to the spatial patterns generated by simple cellular automaton rules. Cellular automata have been considered as simple models of extended dynamical systems, and have been used both in attempts to explain phenomena such as 1/f noise [6], and as tools for simulating physical systems [7]. Since they are discrete systems, they are naturally analyzed using methods and concepts from computation theory, such as formal languages, and our main results will concern the relation between the statistical and computation theoretical properties of the patterns generated in cellular automaton time evolution.

Before describing this in more detail, however, we would like to give a short taxonomy of methods of measuring information, or randomness, and complexity.

1.1 Randomness

> Sequences are commonly considered random if no pattern can be discerned in them. But whether a pattern is found depends on how it is looked for. *Stephen Wolfram*

The patterns which we shall mostly be concerned with in this article are one-dimensional sequences of symbols, in particular sequences produced in the time evolution of one-dimensional cellular automata. One way of viewing the concept of randomness [8] is to consider a string of symbols random if no procedure can detect any regularity which could be used to give it a shorter description. Various measures of randomness, or information, can then be thought of as asymptotic degrees of compressibility of the string, given an allowed class of algorithms to detect structure in it, and perhaps also some amount of *a priori* given information, which could be restricted by a function of string length (e.g. by allowing a certain number of queries to an oracle set). Some examples of quantities that measure randomness are: (our list does not pretend to be exhaustive)

(a) The various Rényi entropies $s(\alpha)$ [9] that can be calculated given measurements of block frequencies. These include the topological entropy ($\alpha = 0$) and the measure entropy ($\alpha = 1$) of the sequence. Among the Rényi entropies the measure entropy s_μ to some extent plays a distinguished rôle, since it has a straightforward interpretation in terms of how much an infinite sequence can be compressed from the knowledge of all block frequencies. In fact, universal coding algorithms exist which asymptotically achieve this degree of data compression for any stationary ergodic source without any advance knowledge of source probabilities [10–12]. Many more sophisticated statistical definitions of randomness also exist [13–15].

(b) The Kolmogorov complexity [15,16–18] of a string x, which is the size of a minimal program that generates x on a universal Turing machine. Infinite strings of maximal Kolmogorov complexity pass every conceivable statistical test for randomness [13], and they in particular have measure entropy equal to one. Conversely, for almost all infinite sequences produced by a stationary stochastic source the Kolmogorov complexity is equal to the measure entropy [19,20]. Of course, the Kolmogorov complexity is in general an uncomputable quantity (though its average over a suitably chosen ensemble of strings apparently can be deduced), and it is thus an extreme case of allowing arbitrary *a priori* information in the framework above. Time- and space-bounded versions of the Kolmogorov complexity have also been proposed [21,22], where one instead considers the minimal program generating x in polynomial time, or using polynomial space on the worktape of the Turing machine.

(c) The notion of complexity relative to a class of sources (e.g. all finite-state machine defined sources, which includes the measures generated at finite times by cellular automata starting from random initial states) recently introduced in coding theory by Rissanen [23,24], which combines features from a) and b). This measure of information is defined, for a finite string x, as the minimum over the class of sources of the difference between a source complexity term and the logarithm of the probability of generating the string x.

(d) The "effective information content" Θ of a sequence, which was proposed by Wolfram [8]. Here the class of algorithms used to detect structure in the string is reduced (in some unspecified way) to a feasible, i.e. polynomial-time computation. Note that this differs from the time-bounded Kolmogorov complexities; in that case we are considering the shortest program running e.g. in polynomial time, here we are concerned with the shortest specification that can be *found* in polynomial time. This is similar in spirit to the effective entropy for ensembles of finite strings introduced by Yao [25], where one considers the minimal code, in terms of average code length, that can be produced in polynomial time by a probabilistic Turing machine.

Even if these quantities tend to agree on large classes of sequences, they are certainly not completely equivalent. This can for example be seen if we attempt to use them as evolution criteria for infinite cellular automata. Then all Rényi entropies decrease in time (or at least do not increase, at each time step $\Delta s(\alpha) \leq 0$, see section 3), and this is also the case for the Kolmogorov complexity, since evolving the cellular automaton a finite number of steps forward in time only requires a finite addition to the minimal program for the initial state, which makes no difference in the limit of infinite strings. For quantities of type d), where one is restricted to polynomial-time computations, this need not necessarily be the case. Even if our limited class of algorithms can detect some structure in the sequence at a certain time, it might not be able to accomplish this at the next time step, since finding a predecessor of a configuration one time step back can be an NP-complete problem in two and more space dimensions [26]. Even in one dimension, where a predecessor configuration a fixed number of time steps back certainly can be found in time polynomial in the length of the sequence (though presumably exponential in the number of steps back in time) by explicitly constructing the regular language of predecessors, the number of different predecessor configurations in general increases exponentially with the length of the sequence. Thus, if we need to find a particular predecessor with a short description this might take exponential time, and it then seems plausible that any polynomial-time regularized quantity in certain cases could increase in the time evolution of cellular automata.

1.2 Complexity

Alors entre l'ordre et le désordre, règne un moment délicieux...
Paul Valéry

This section should begin with a remark on semantics, previously emphasized by Grassberger [27], Hubermann [28], and others. Several of the quantities mentioned above went under the name of "complexity", and the word was then used as being synonymous to "information" or "randomness". Physicists generally seem to prefer to reserve the word "complex" for structures that are neither random nor regular, but (loosely speaking) show signs of intricate, perhaps hierarchical organization. In the following we shall use the word in this sense.

One way to make this notion of complexity more precise is to regard complexity as a property of *ensembles* of patterns, rather than the individual configurations themselves [27]. A natural approach would then be to define the complexity as the size of a minimal description of the ensemble of patterns. As an example, in this way a random pattern could be associated to the ensemble of all possible patterns, which has a very simple description, at least in the case of strings considered here.

Some examples of quantities that have been suggested as measures of complexity are the following:

(a) Various statistical quantities related to the convergence rate of the finite length block entropies $S_n(\alpha)$ rather than their actual values [27,29,30], such as the effective measure complexity introduced by Grassberger. These quantities essentially measure the amount of long-range correlations contained in a pattern, or an ensemble of patterns. We shall discuss these measures more extensively in section 2.

(b) At least one quantity related to the Kolmogorov complexity has been suggested, Bennett's "logical depth" of a pattern [31], which is the time required to produce it from the minimal program. Clearly this quantity is small both for very regular patterns and for completely random patterns, where the minimal program is essentially a copy of the pattern. Even though it presumably shares the property of uncomputability with the Kolmogorov complexity, one might still hope that the generalized Kolmogorov complexity classes mentioned above, such as the classes of all patterns produced from logarithmic size programs in polynomial or exponential time, could be characterized in alternative (e.g. statistical) ways. This could have interesting implications for biological systems.

(c) In the particular case when the pattern are trees rather than strings, a measure of complexity has been introduced by Hubermann and Hogg [28]. Even in the case of a probability distribution on a set of strings this could be relevant, since the probability distribution could be decomposed into pure states, and in some cases these could show an approximately ultrametric hierarchical organization [32]. The complexity of trees also turns out to be measured by the rate of relaxation for ultradiffusion in the hierarchical space described by the tree [33].

(d) Another class of complexity measures are those related to a description of the ensemble of patterns. For sequences of symbols this description could be a grammar of a formal language (e.g. [34]), or a weighted grammar if a measure on the ensemble is considered. Different classes of formal languages can be characterized by their accepting automata, and the complexity could for example be measured by the number of nodes in the automaton [35], or by the entropy of a probability distribution on the nodes [27]. We shall discuss this further in the next section. This approach requires that a grammatical description of the ensemble is known. For one-dimensional cellular automata this description can in principle be calculated at any finite time (though this might not be computationally feasible), if the ensemble of initial states is known. In most physical situations one would however encounter the difficult problem of inferring this description from data. Furthermore these concepts are considerably less developed in higher dimensions.

Finally one can note that this state of affairs leaves some room for further developments. Most of these measures of complexity are of limited applicability. The quantities mentioned in a) and d) might seem very general (at least if appropriate generalizations to higher dimensions could be constructed), but unfortunately they are all divergent in very complex environments, and they are not easily computed in practice.

In the following section we first review, interpret, and in some cases generalize the definitions of various statistical and computation theoretical measures of complexity, and discuss the relations between them. We then briefly discuss whether the computation theoretical quantities can be computed when the grammar is not known from the outset, and finally we prove some results on the generic form of the convergence of finite length block entropies for measures corresponding to regular languages. The convergence of block entropies for context-free languages is also discussed. Section 3 deals with cellular automata at finite time, and starts with a characterization of the exact time development of the measure on the space of infinite sequences in terms of probabilistic finite automata. We then apply some of the concepts from section 2 to the time evolution of cellular automata, in particular to an additive cellular automaton rule starting from a random initial state, but with a density of ones different from 1/2. It is shown that in this case the effective measure complexity increases linearly in time. Section 4, finally, contains a discussion of the attractors and limit sets of cellular automata. We attempt to interpret the critical exponents observed numerically by Grassberger [36] in the convergence of block entropies for certain chaotic cellular automaton rules. These numerical results indicate that the attractor in these cases does not correspond to a regular or an unambiguous context-free language.

2. Entropies and complexity measures

We shall now consider some of the quantities mentioned in the introduction more in detail. Let us begin by defining the Rényi entropies. Suppose that we have a finite alphabet of symbols Σ, and for each integer n a probability distribution on the strings of length n in Σ^* (Σ^* denotes the set of all finite strings over Σ). This could for example correspond to the block probabilities of a single infinite string, or an ensemble of infinite strings. The Rényi entropy of order α is then defined by [9]

$$s(\alpha) = \overline{\lim}_{m\to\infty} \frac{1}{m} S_m(\alpha), \tag{2.1}$$

where the block entropies $S_m(\alpha)$ are given by

$$S_m(\alpha) = \frac{1}{1-\alpha}\log(\sum_{|\sigma|=m} p^\alpha(\sigma)). \tag{2.2}$$

Here the sum is over all strings σ of length m, and the base of the logarithm is a, the number of symbols in Σ. For $\alpha = 0$, (2.1) is the topological entropy, and in the limit $\alpha \to 1$ the measure entropy s_μ. Using a coding procedure, the measure entropy can be interpreted as the minimal average code length per symbol in the limit of infinite strings. This means that we are attempting to minimize $l = \sum p(\sigma_k)N_k$, where N_k is the length of the code word that corresponds to σ_k. For $\alpha \neq 1$, we are instead minimizing an average "code length of order α" $l_n(\alpha)$ [37], where

$$l_n(\alpha) = \frac{\alpha}{1-\alpha}\log(\sum_k p(\sigma_k)e^{\frac{1-\alpha}{\alpha}N_k}) \geq S_n(\alpha). \tag{2.3}$$

For $\alpha < 1$, this is equivalent to the minimization of a total cost, where an exponential cost function has been associated to each code word.

The Rényi entropies measure the information content, or randomness, of a sequence. A sequence with a certain entropy can still be more or less complex, and one way of capturing this concept is to consider the convergence rate of the block entropies $S_m(\alpha)$. For the moment we restrict ourselves to the case $\alpha = 1$. The *effective measure complexity* (EMC) introduced by Grassberger [27] is then defined as

$$\eta = \lim_{m\to\infty}(S_m - ms_\mu), \tag{2.4}$$

which can also be expressed as

$$\eta = -\sum_{m=1}^{\infty} m\Delta^2 S_{m+1}, \tag{2.5}$$

where $\Delta S_m = S_m - S_{m-1} \geq 0$ and $\Delta^2 S_m = \Delta S_m - \Delta S_{m-1} \leq 0$. The total information contained in the correlations of a sequence can be divided into independent contributions $k_n = -\Delta^2 S_n$ from block entropies of different lengths n [38], which shows that η can be interpreted as the product of an average correlation length η/k_{corr} and the total correlational information $k_{corr} = \sum k_n$. The effective measure complexity can also be written as an average Kullback information (relative information)

$$\eta = \lim_{n\to\infty}\lim_{m\to\infty}\sum_{\beta_m} p(\beta_m)\sum_{\sigma_n} p(\sigma_n|\beta_m)\log\frac{p(\sigma_n|\beta_m)}{p(\sigma_n)}, \tag{2.6}$$

where β_m is a string of length m preceding σ_n in the sequence. This represents the average information gain when the distribution $p(\sigma)$ is replaced by $p(\sigma|\beta)$, or equivalently the average information stored in a semi-infinite sequence about its continuation. When n additional symbols σ_n are added to a semi-infinite sequence β, an average information $n \cdot s_\mu$ is gained. The remaining $n(1-s_\mu)$ bits of information are contained in the structure of the ensemble, and can be divided into one part which is the correlational information contained in σ_n, and one part which (if convergent) is equal to η in the limit of large n. The EMC is divergent if the block entropies converge slower than $1/n$, a phenomenon which can occur in more complex environments, since when strong long-range correlations are present, a semi-infinite sequence could store an infinite amount of information about its continuation. We show at the end of this section that in the less complex situation where the ensemble of strings corresponds to a regular language, with a measure generated by a finite automaton, the block entropies in many cases converge as

$$\frac{1}{n}S_n \sim s + \frac{\eta}{n} + c\,\mathrm{e}^{-\gamma n}. \tag{2.7}$$

In particular the EMC is always finite if the automaton has a non-zero stationary probability distribution on its nodes. If the block entropies S_m converge as above, the differences ΔS_m converge exponentially. This rate of convergence has been used as a measure of the complexity of trajectories in dynamical systems by Györgyi and Szépfalusy [29,30] who defined a quantity

$$\gamma = -\lim_{m\to\infty}\frac{1}{m}\log(\Delta S_m - s_\mu). \tag{2.8}$$

These quantities can be generalized to arbitrary Rényi entropies; we can for example define $\eta(\alpha)$ in terms of the convergence rate of $S_n(\alpha)$,

$$\eta(\alpha) = \lim_{m\to\infty}(S_m(\alpha) - ms(\alpha)), \tag{2.9}$$

and similarly for $\gamma(\alpha)$. One could also contemplate other definitions, not necessarily equivalent to (2.9), that reduce to η when $\alpha \to 1$, such as a correspondence of (2.6), where we form a weighted average of the relative information of order α,

$$\eta_1(\alpha) = \lim_{n\to\infty}\lim_{m\to\infty}\frac{1}{\alpha-1}\sum_{\beta_m}\frac{p^\alpha(\beta_m)}{\sum_{\beta'_m}p^\alpha(\beta'_m)}\log\Big(\sum_{\sigma_n}p^\alpha(\sigma_n|\beta_m)p^{1-\alpha}(\sigma_n)\Big). \tag{2.10}$$

We now turn to quantities that involve a grammatical description of the ensemble of patterns. This description could be the grammar of a formal language (given here as an accepting automaton) if only the set of allowed sequences is considered, or a measure generated for example by a finite automaton. The reader wishing an introduction to formal languages and computation theory could turn e.g. to [34], and to the article [35] for applications to cellular automata. Languages with associated probability distributions are for example encountered in the literature on syntactic pattern recognition (e.g. [39]). Very briefly, a formal language is any subset of Σ^*, the set of finite strings over the symbol set Σ. The Chomsky hierarchy of regular, context-free, context-sensitive and unrestricted languages classifies (recursive) languages according to the complexity of the automata needed for their recognition.

Regular languages correspond to finite automata, which have a finite number of internal states (nodes in a transition graph) forming a set Q, and a transition function $\delta : Q \times \Sigma \to Q$, which means that the arcs in the transition graph are labelled by symbols in Σ. A string $\sigma = \sigma_1 \ldots \sigma_k$ belongs to the language L accepted by the automaton if there is a path labelled by $\sigma_1 \ldots \sigma_k$ from the start node S to a node in the set F of allowed final nodes. As long as no consideration is given to the measure, *deterministic* finite automata, where a certain sequence can only be obtained in one way, are equivalent to *nondeterministic*, where several arcs leaving a node may carry the same symbol, since for any regular language L a unique minimal deterministic automaton accepting L can be constructed. In general this procedure cannot be performed when a probability distribution on L is taken into account; the class of measures where the underlying finite automaton is nondeterministic is larger than that where we only consider deterministic automata. This yields two classes of measures generated by finite automata. The semi-group measures investigated in [40] correspond to deterministic automata with a unique start node, and since they can be written in terms of a finite-dimensional transition matrix, they are in many ways similar to Markov measures. In particular arbitrary Rényi entropies can be calculated from the eigenvalues of the matrix obtained by raising each element of the transition matrix to the power α. Each such semi-group measure is the image of a Markov chain under some cellular automaton rule (which is finite-to-one on the space of infinite sequences) [40], but in general, even for additive rules, the measures generated in cellular automaton time evolution belong to the larger class we now describe.

A measure of this class is given by a finite automaton with transition probabilities as well as symbols assigned to the state transitions, and equipped with an initial probability distribution on its nodes. For each symbol σ_k, $k = 1 \ldots a$, we then have a transition matrix $\mu(\sigma_k)$, and the matrix obtained by summing over symbols, $\mu = \sum \mu(\sigma_k)$, is an ordinary stochastic transition matrix with row sums equal to unity. This yields a probability distribution on the strings of length n in L for any n; if the vector α_i represents the initial probability distribution, and we allow all nodes as final nodes, we have

$$p(\sigma = \sigma_1 \ldots \sigma_n) = \alpha^{\mathrm{T}} \mu(\sigma_1) \ldots \mu(\sigma_n) \beta, \tag{2.11}$$

where $\beta = (1, 1, \ldots, 1)$. This set of probability distributions extends to a unique shift-invariant measure on the set Σ^Z of bi-infinite strings over Σ if the Kolmogorov consistency conditions (e.g. [41,42]) are fulfilled. It can easily be checked that this is the case if α_i is an equilibrium probability distribution on the automaton. Measures belonging to this larger class can be classified according to the nature of the underlying finite automaton, which is either deterministic (but with an initial probability distribution on nodes), for example in the case of the finite time measures for surjective rules, or nondeterministic. In the first case the measure entropy can be calculated exactly as

$$s_\mu = \sum_j p(j) \sum_{\sigma\in\Sigma} p(\sigma|j) \log \frac{1}{p(\sigma|j)} \tag{2.12}$$

(where $p(j)$ is the equilibrium probability distribution on the nodes of the automaton), while in the much more subtle second case, it appears that entropies can only be calculated exactly in particular cases [43].

Context-free and higher languages can also be represented by transition graphs, but the graphs are then infinite. Context-free languages are accepted by pushdown automata, i.e. automata with a stack. The use of the stack can be restricted to only push and pop operations (and no move) [34], which means that one can represent this structure as an infinite self-similar tree of finite automata, see figure 1. The nodes here represent all states with a certain stack content, and their internal structure depends only on the symbol at the top of the stack. The arrows thus summarize a number of transitions in both directions, and the branching ratio of the tree depends on the number of symbols in the stack alphabet (which need not equal Σ). An acceptable word must terminate with the stack empty, i.e. at the top node (or we could equivalently use acceptance by final state). This means that the entropies of context-free languages should behave similarly to the auto-correlation function for diffusion on a self-similar tree; we shall discuss this further in section 4.

Similarly, an arbitrary Turing machine can be simulated by a two-counter machine [34], which has two stacks where except for a start symbol at the bottom of the stack, only one symbol is used. This structure could be drawn as (one quadrant of) a two-dimensional array of finite automata, see figure 1. When arbitrary stack moves are allowed, the transitions become rather complicated, and this representation is less useful.

Given a mechanism with internal states which describes our ensemble, various quantities measuring the complexity of the ensemble can be introduced [27,35]. These are essentially entropies of the equilibrium probability distribution on the nodes, or internal states. We should distinguish two cases; we could either consider automata that only reproduce the set of allowed sequences, or we could consider probabilistic automata that are required to reproduce the measure. In the first case it suffices to consider deterministic automata (for regular languages) and one can define the algorithmic complexity as the number of nodes in the minimal deterministic automaton accepting the language [35],

$$\nu = \inf_A \log N(A), \tag{2.13}$$

(A is the class of deterministic automata accepting the language). If there is a measure on the ensemble of infinite strings, this induces a probability

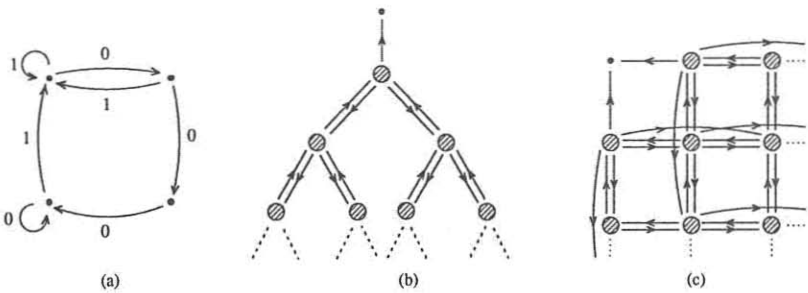

Figure 1: Accepting automata for (a) a regular language, (b) context-free languages, (c) unrestricted languages.

distribution on the nodes of any automaton accepting the language, and Grassberger defined the *set complexity* as the entropy of this distribution [27],

$$\sigma = \inf_A \Big(\sum_j p(j) \log \frac{1}{p(j)}\Big). \tag{2.14}$$

This quantity may be finite for many infinite automata accepting context-free and higher languages, but on the other hand, the restriction to deterministic automata is not as meaningful in that case, since it excludes certain languages.

If we now consider the class A' of probabilistic automata that reproduce a certain measure (and the underlying finite automata are now in general nondeterministic, as mentioned above), one can define

$$\nu' = \inf_{A'} \log N(A'), \tag{2.15}$$

$$\tau_1 = \inf_{A'} \Big(\sum_j p(j) \log \frac{1}{p(j)}\Big), \tag{2.16}$$

$$\begin{aligned} \tau_2 &= \inf_{A'} \Big(\sum_j p(j) \log \frac{1}{p(j)} - \lim_{n\to\infty} \sum_{\sigma_n} p(\sigma_n) \sum_j p(j|\sigma_n) \log \frac{1}{p(j|\sigma_n)}\Big) \\ &= \inf_{A'} \Big(\sum_j p(j) \sum_{\sigma_n} p(\sigma_n|j) \log \frac{p(\sigma_n|j)}{p(\sigma_n)}\Big), \end{aligned} \tag{2.17}$$

where ν' and τ_1 are the cases $\alpha = 0$ and $\alpha = 1$ of the Rényi entropy $\tau(\alpha)$ of the probability distribution on nodes. The quantity τ_1 represents the average information contained in the state of the automaton; in τ_2, which is the *true measure complexity* (TMC) introduced by Grassberger [27], we have subtracted a term which represents the amount of information that asymptotically remains in the automaton. This term vanishes if asymptotically one for almost all sequences can uniquely determine at which node the sequence started, which means that the probability distribution $p(j|\sigma_n)$ singles out one particular node. We conjecture that this is always the case when the underlying finite automaton is deterministic, while in the nondeterministic case the number of paths that corresponds to a sequence increases with length, and this statement need not be valid (counterexamples can easily be found, e.g. some of the automata discussed in section 3). The quantity τ_2 can also be written as an average Kullback information in analogy with (2.6). There is an obvious inequality $\nu' \geq \tau_1 \geq \tau_2$, and it is also evident that the information about the future sequence carried by the state of the automaton is larger than or equal to the information the preceding part of the sequence contains about its continuation, i.e. the effective measure complexity η [27]. One also can show more formally that $\tau_2 \geq \eta$:

Using $p(\sigma_n) = \sum p(\sigma_n|j)p(j)$, and $p(\sigma_n|j)p(j) = p(j|\sigma_n)p(\sigma_n)$ for any node j, we can rewrite the block entropy S_n as

$$\begin{aligned} S_n &= \sum_{\sigma_n} p(\sigma_n) \log \frac{1}{p(\sigma_n)} \qquad (2.18) \\ &= \sum_j p(j) \sum_{\sigma_n} p(\sigma_n|j) \log \frac{1}{p(\sigma_n|j)} + \\ &\quad + \sum_j p(j) \log \frac{1}{p(j)} - \sum_{\sigma_n} p(\sigma_n) \sum_j p(j|\sigma_n) \log \frac{1}{p(j|\sigma_n)} = \\ &= \sum_j p(j) \sum_{\sigma_n} p(\sigma_n|j) \log \frac{1}{p(\sigma_n|j)} + \tau_2. \end{aligned}$$

Here the first term is smaller than or equal to $n \cdot s_\mu$, since

$$\begin{aligned} 0 &\leq \lim_{m\to\infty} \sum_{\beta_m} p(\beta_m) \sum_{\sigma_n, j} p(\sigma_n|j)p(j|\beta_m) \log \frac{p(\sigma_n|j)}{p(\sigma_n|\beta_m)} \qquad (2.19) \\ &= \lim_{m\to\infty} \sum_{\beta_m} p(\beta_m) \sum_{\sigma_n} p(\sigma_n|\beta_m) \log \frac{1}{p(\sigma_n|\beta_m)} \\ &\quad - \sum_j p(j) \sum_{\sigma_n} p(\sigma_n|j) \log \frac{1}{p(\sigma_n|j)} \\ &= n \cdot s_\mu - \sum_j p(j) \sum_{\sigma_n} p(\sigma_n|j) \log \frac{1}{p(\sigma_n|j)}, \end{aligned}$$

and thus $S_n \leq n \cdot s_\mu + \tau_2$, which in turn implies that $\tau_2 \geq \eta$. The Kullback information used in this derivation is always well-defined, since $p(\sigma_n|\beta_m) = 0$ implies that $p(\sigma_n|j)p(j|\beta_m) = 0$ for any j. We then have the following set of inequalities:

$$\nu' \geq \tau_1 \geq \tau_2 \geq \eta. \tag{2.20}$$

This in particular shows that the EMC is finite for any measure given by a finite automaton, provided that it has a non-zero equilibrium probability distribution. We have not been able to find a useful extension of (2.20) to arbitrary Rényi entropies.

Several of the quantities mentioned above involved a grammatical description of the ensemble. If the initial ensemble is known, and the dynamics is given for example by a cellular automaton rule, one could in principle evolve the description of the initial state forward in time to obtain the grammar (or measure) at any finite time. But this might not be computationally feasible, and in a more realistic physical situation one is confronted with a set of patterns without any *a priori knowledge* of a grammatical description. The problem of inferring the grammar of a formal language has been studied in computer science (e.g. [3]); here we just intend to make some brief remarks about the possibility of calculating quantities like (2.13).

Suppose that a sequence $w_1, w_2, \ldots$ of strings from an unknown formal language L is presented to us, and that we make a sequence of guesses $G_n(w_1, \ldots, w_n)$ about the grammar, assuming that L belongs to some class U of languages. The class U is said to be *identifiable in the limit* [44] if there is an inference method M such that for any acceptable presentation $w_1, w_2, \ldots$ of a language in U (a minimal requirement is that the presentation includes all strings in L at least once), the correct grammar is obtained after a finite number of conjectures. This is a rather weak notion of convergence, since the algorithm M need not know whether it has converged to the right answer. In a similar way, we could call an integer-valued function $f(w_1, w_2, \ldots)$ of an infinite sequence computable in the limit, if there is an algorithm with finite subsequences $w_1, \ldots, w_n$ as input which converges to the right value for some finite n. It is known [44] that the class of regular languages is not identifiable in the limit from positive presentations (i.e. when no examples from the complement of L are given), and we have modified this argument to show that the regular language algorithmic complexity (2.13) is not computable in the limit in this case. But for ensembles of infinite strings this is not really relevant, since the argument depends on the fact that all finite languages are included among the regular languages as allowed hypotheses, and a finite language cannot be the set of permitted n-blocks of an ensemble of infinite strings. This indicates that a different space of hypotheses should have been chosen. Furthermore, these difficulties are only encountered for a negligible subset of all presentations; all context-free languages (and thus all regular languages) are known to be identifiable in the limit with probability 1 [45].

In practice, one would need a computationally feasible procedure for estimating the grammar. Assuming that we are given both a positive sample of strings from L, and a negative sample of strings not in L, a natural approach would be to find the minimal finite automaton (or regular expression) compatible with the samples. This has been shown to be an NP-complete problem [46,47]. For the more difficult (and more relevant to us) problem of only positive presentations, polynomial-time algorithms exist if one considers certain subclasses of the regular languages, e.g. the k-reversible languages [48]. In cases where the EMC can be calculated numerically, a lower bound on the algorithmic complexity ν' of the automaton A' giving the measure is obtained from the inequality (2.20). This does not necessarily yield any information about the algorithmic complexity n of the deterministic automaton A corresponding to the set of sequences, since A is obtained from A' by a combination of a reduction (when transition probabilities are removed, nodes may become equivalent), and a conversion from a nondeterministic to a deterministic automaton. (It seems likely that a lower bound on n could be obtained from $\gamma(0)$, which measures the difference between the first and second eigenvalue of the transition matrix of A.) Instead of the algorithmic complexity, one could attempt to calculate the entropies σ, τ_1 and τ_2, which should behave significantly better when the grammar is approximated. The set complexity has been calculated for the symbolic dynamics of iterated one-dimensional maps by Grassberger [49].

We now turn to the question of the rate of convergence of finite length block entropies for measures generated by finite automata. For the Rényi entropies with $\alpha = 0, 2, 3, 4, \ldots$ we can prove that generically the block entropies converge as

$$\frac{1}{n} S_n \sim s + \frac{\eta}{n} + c\, \mathrm{e}^{-\gamma n} \tag{2.21}$$

for all measures generated by finite automata, for other values of α our results are incomplete.

To prove the statement (2.21), let us first introduce some concepts from the theory of formal power series in noncommuting variables, following the book by Salomaa and Soittola [50]. A formal power series is a formal sum

$$s = \sum_{w_i \in M} c_i w_i, \tag{2.22}$$

where the coefficients c_i in our case are real numbers (and in a more general case belong to a semiring A), and the w_i are the elements of a monoid M (a monoid is an object that satisfies every axiom for a group except the existence of an inverse for every element). The monoid that will actually enter here is the *free monoid* generated by a finite set Σ, which is simply the set Σ^* of all finite strings of symbols in Σ. The product $w_1 * w_2$ of two strings w_1 and w_2 in Σ^* is formed by concatenation, $w_1 * w_2 = w_1 w_2$.

We shall primarily be concerned with the important class of formal power series called *rational series*, which are rational functions in noncommuting variables, i.e. one component of the formal solution to a linear system of equations in noncommuting variables. As we shall see, these series are closely related to the regular languages. There are various operations that preserve the rationality of formal power series, a less obvious one which we shall need later is the *Hadamard product* $r \odot r'$ of two series, which is defined by termwise multiplication of coefficients,

$$(\sum_i c_i w_i) \odot (\sum_i d_i w_i) = \sum_i (c_i d_i) w_i. \tag{2.23}$$

An important theorem which characterizes the class of rational series is the Representation Theorem of Schützenberger [50,51], which states the following:

Any rational series can be written as

$$r = \sum_{w \in M} (\alpha^{\mathrm{T}} \mu(w) \beta) w, \tag{2.24}$$

and conversely any formal power series of the form above is rational.

Here $\mu : M \to A^{m \times m}$ is a representation of the monoid M in terms of $m \times m$ matrices with elements in the semiring of coefficients A. This means that the matrices $\mu(w)$ satisfy $\mu(w_1)\mu(w_2) = \mu(w_1 w_2)$ for any w_1, w_2 in M. Furthermore α and β are constant vectors of length m. The connection between rational series and regular languages is evident if we let $\mu(w = \sigma_1 \ldots \sigma_n) = \mu(\sigma_1) \ldots \mu(\sigma_n)$, where $\mu(\sigma)$ for $\sigma \in \Sigma$ are the transition matrices given by a finite automaton accepting the regular language, and we let the vectors α and β be given by the start node and the allowed final nodes (e.g. $\alpha^{\mathrm{T}} = (1, 0, \ldots, 0)$ and $\beta^{\mathrm{T}} = (1, 1, \ldots, 1)$ if all nodes belong to the set F of final nodes). We can also see that Schützenberger's Representation Theorem applies to the measures defined by finite automata according to (2.11); the series

$$s = \sum_{\Sigma^*} p(w) w \tag{2.25}$$

is then a rational series, and taking the Hadamard product of s with itself n times we find that

$$s^{(n)} = \sum_{\Sigma^*} p^n(w) w \tag{2.26}$$

is a rational series for any integer $n \geq 1$ (or rather $n \geq 0$, if the coefficient of w is defined as 0 when $w \notin L$). For a rational series r, the generating function $G(z)$ obtained by replacing each symbol from $\sum$ in r by the commuting variable z,

$$G(z) = \sum_m (\sum_{|w|=m} c(w)) z^m, \tag{2.27}$$

is an ordinary rational function, which means that

$$a_m^{(n)} = \sum_{|w|=m} p^n(w) \tag{2.28}$$

are the Taylor coefficients of a rational function. These necessarily satisfy a linear difference equation [52], which means that they are of the form

$$a_m^{(n)} = P_1(m)\gamma_1^m + P_2(m)\gamma_2^m + \ldots + P_{N(n)}\gamma_{N(n)}^m, \tag{2.29}$$

where the $P_i(m)$ are polynomials in m. The matrices $\mu^{(n)}(w)$ that enter if $s^{(n)}$ is written as in (2.24) are obtained by taking tensor products, $\mu^{(n)}(w) = \mu(w) \otimes \ldots \otimes \mu(w)$, and have dimension $(\dim \mu(w))^n$, which means that the number of terms $N(n)$ increases exponentially with n. For the Rényi entropies with $\alpha = n$ ($n \neq 1$) this implies that

$$\frac{1}{m} S_m \sim s + \frac{\eta}{m} + C\,\mathrm{e}^{-\gamma m} \tag{2.30}$$

in nondegenerate cases, where P_1 is a constant. In a degenerate case, the corrections to S_m/m may behave as $\log(m)/m$, and we can give a simple example illustrating this for the topological entropy. The number of words in the regular language $L = 0^*1^* = \{0^n1^p \text{ for arbitrary } n,\, p \geq 0\}$ clearly increases linearly with length, which means that the topological entropy vanishes, since S_m is proportional to $\log m$. It has also been noted in [53] that the growth function for a regular language, i.e. the number of words of length m, asymptotically increases exponentially, polynomially, or approaches a constant. Nonstationary ensembles of this kind may occur as cellular automaton limit sets (see the examples of limit sets in [35] and [54]), but the automata defining the ensemble then have transient parts, which are removed if a measure on the ensemble is introduced according to (2.11), since a stationary probability distribution on nodes is needed to define the measure.

When the characteristic equation giving rise to (2.29) has complex roots, oscillatory behavior could be seen in the convergence of block entropies, particularly in the topological case. This was observed in [55] for the topological entropy $\Delta S_n(0)$ in symbolic dynamics of the logistic equation.

The argument above does not extend to arbitrary Rényi entropies in its present form, except in restricted cases. For the semi-group measures in [40] there is a unique start node for each sequence, and entropies can be calculated by raising each element of the transition matrix to the power α. In this case (2.30) is valid for all α. We also expect this to be the case for all measures of the form (2.11) with a deterministic underlying finite automaton, since the probability of a sequence of any length is then given by the sum of a fixed number of terms (equal to the number of nodes), and asymptotically this sum is dominated by its largest term for almost all sequences, which asymptotically singles out a unique start node.

For unambiguous context-free languages (CFLs), where each word has a unique derivation, the argument above could be modified by replacing rational by algebraic series [50]. But since we have not discussed what classes of measures could be associated to CFLs, we shall here primarily restrict ourselves to the topological entropy. In that case a classical theorem [56] states that the structure generating function ((2.27) with $c(w) = 1$ if $w \in L$,

$c(w) = 0$ otherwise) of an unambiguous CFL is an algebraic function, and from the asymptotic behavior of the Taylor coefficients of algebraic functions it then follows that asymptotically the number of words in L increases as [57]

$$g(n) \sim c n^{\kappa} \gamma^{n} (\sum_i c_i \omega_i^n), \tag{2.31}$$

where κ is rational, γ is an algebraic number, and in the oscillating factor, c_i and ω_i are algebraic with $|\omega_i| = 1$. The topological block entropies then converge as $\log(m)/m$. A simple example is given by the context-free language of all finite strings consisting of an equal number of the symbols 0 and 1. Then (for n even)

$$g(n) = \binom{n}{n/2} \sim \sqrt{\frac{2}{\pi n}} 2^n + O(n^{-3/2} 2^n). \tag{2.32}$$

The asymptotic behavior of the topological entropy for context-free languages will be compared to the auto-correlation function for diffusion on a self-similar tree in section 4, where we also briefly discuss measures corresponding to context-free languages.

3. Cellular automata at finite time

Let us now apply some of these concepts to the generation of complexity in the time evolution of infinite one-dimensional cellular automata. The cellular automaton mapping on infinite sequences is induced by a local transformation $\phi : \Sigma^{2r+1} \to \Sigma$, where Σ is a finite set of symbols, in our examples $\Sigma = \{0, 1\}$, and r is the range of the transformation. Many cellular automata are irreversible systems, and this is reflected in a decrease of spatial entropies [58,59,20]. In fact all spatial Rényi entropies satisfy $\Delta s(\alpha) \leq 0$ at each time step; we can easily generalize the proof involving the measure entropy given in [20]:

$$\begin{aligned} \Delta s(\alpha) = {} & \lim_{m\to\infty} \quad (\frac{1}{m} S_m(\alpha, t+1) - \frac{1}{m} S_m(\alpha, t)) \\ = {} & \lim_{m\to\infty} \quad (\frac{1}{m+2r}(S_m(\alpha, t+1) - S_{m+2r}(\alpha, t)) \\ & \qquad + \frac{2r}{m(m+2r)} S_m(\alpha, t+1)) \end{aligned} \tag{3.1}$$

and

$$\begin{aligned} S_m(\alpha, t+1) &= \frac{1}{1-\alpha} \log(\sum_{|\sigma|=m} p^{\alpha}(\sigma_m, t+1)) \\ &= \frac{1}{1-\alpha} \log(\sum_{|\sigma|=m} (\sum_{\phi(\sigma_{m+2r})=\sigma} p(\sigma_{m+2r}, t))^{\alpha}) \\ &\leq \frac{1}{1-\alpha} \log(\sum_{|\sigma|=m+2r} p^{\alpha}(\sigma_{m+2r}, t)) = S_{m+2r}(\alpha, t), \end{aligned} \tag{3.2}$$

which means that the first term in (3.1) is ≤ 0, and since the second term vanishes as $m \to \infty$, we obtain

$$\Delta s(\alpha) \leq 0. \tag{3.3}$$

If the initial state has well-defined block probabilities and entropy, this will be the case at every subsequent time as well, and (3.3) is then valid at every time step.

The quantities defined as measures of complexity in the previous section can also be used to characterize cellular automaton time evolution, but before we can do this we need to discuss the time evolution of the probability measure on the space of sequences. Previous work on cellular automata has often studied the generation of complexity by considering the set of all allowed sequences at time t, starting from arbitrary initial configurations. For finite time t, this set corresponds to a regular language [35,60], which in principle can be constructed explicitly. As was described above, one way to measure the generated complexity is to count the number of nodes in the minimal deterministic finite automaton accepting the language (finite time set) $\Omega(t)$. In practice the algorithmic complexity of the finite time sets increases very rapidly for chaotic rules, and cannot be calculated by explicit construction of $\Omega(t)$ except for the first few time steps. Furthermore, this way of measuring complexity does not take statistical aspects of the time evolution into account, and gives equal weight to common behavior and phenomena occurring with vanishing probability. A treatment of the time evolution of the measure would be a complement to this approach and show several phenomena more clearly.

One example is given by class 2 rules, which by definition asymptotically simulate a shift map for almost all initial states. The set of sequences on which the rule simulates a shift map is a simple regular language, which can easily be constructed from the cellular automaton rule [35], and it seems reasonable to call this set an attractor. Even though the attractor is a simple regular language, and the individual patterns generated seem to become less complex with time as transients disappear, the algorithmic complexity in general increases polynomially, often linearly, in time for class 2 cellular automata. One reason for this is that the limit set, which includes configurations occurring with vanishing probability, may be more complicated than the attractor [54]. This is illustrated even in the simple case of rule 128 discussed later in this section.

Another example is the case of surjective rules, where the set of all possible sequences is a fixed point of the time evolution. Certain of these rules can still show chaotic behavior, and if we instead measure the complexity generated starting from restricted classes of sequences, or even from random sequences with a density of ones p different from $1/2$, the behavior will be more similar to other chaotic rules. This is exemplified below for an additive rule, where we show that the effective measure complexity increases linearly in time when the initial state has $p \neq 1/2$.

The time evolution of the measure for cellular automata has been studied by Gutowitz, Victor and Knight [42] in an approximation ("local structure theory") where the block probabilities above some length n are obtained by Bayesian extension, which means that the length m correlational informations k_m vanish above that length. This approximation seems to work very well in many cases. Here we shall instead consider the exact time evolution of the measure, which in some cases will reveal phenomena, such as a diverging effective measure complexity, that cannot be seen in a truncation to a fixed block length.

If the ensemble of initial states is given by a probabilistic finite automaton A in the way described in section 2, the measures at finite time can be calculated by a procedure very similar to that of applying the cellular automaton mapping to an ordinary finite automaton described in [35]. For $r = 1$ we can label the nodes in the resulting automaton $\phi(A)$ by connected pairs of arcs in A, and the allowed transitions and transition probabilities in $\phi(A)$ are then given by the arc-to-arc transitions in the original automaton. To define a measure, we also need an initial probability distribution on the nodes of $\phi(A)$. This distribution should in general be the equilibrium distribution for the automaton to consistently define a measure on bi-infinite sequences, and one can check by explicit calculation that an equilibrium distribution on A is in fact mapped to an equilibrium distribution on $\phi(A)$. If a node is labelled by two arcs where the first starts at node i, the new initial probability distribution should be $p(i)$ multiplied by the transition probabilities of the arcs to give the correct measure, and this turns out to be exactly the equilibrium distribution on the new automaton. For surjective rules, we obtain a measure where the underlying finite automaton is deterministic, for other rules it is in general nondeterministic. One can sometimes simplify the resulting measure by identifying equivalent nodes just as for ordinary finite automata [35] (nodes are equivalent if they have identical transitions, i.e. with equal probabilities and labelled by identical symbols, to all equivalence classes of nodes), but we know of no general procedure to determine whether two measures of this kind are equal. Let us now illustrate this by a few simple examples:

As a first example, let us consider rule 4 (we number CA rules according to the convention of Wolfram [58]). This rule maps 010 to 1, and all other length three blocks to 0, and thus it reaches a fixed point at $t = 1$ for any initial state, since $\phi^2 = \phi$. One can easily calculate the limit language and the corresponding measure in this case; starting from an uncorrelated initial state with a density of ones equal to p one obtains the automata shown in figure 2.

This measure is an example of a nondeterministic automaton where an exact expression for the measure entropy can be written down, since the occurrence of the symbol 1 in a sequence implies a unique state for the automaton. The conditional probabilities $p(10\ldots0|1)$ satisfy (where the sequence is read from right to left, and we define $a_n = p(10^n|1)$)

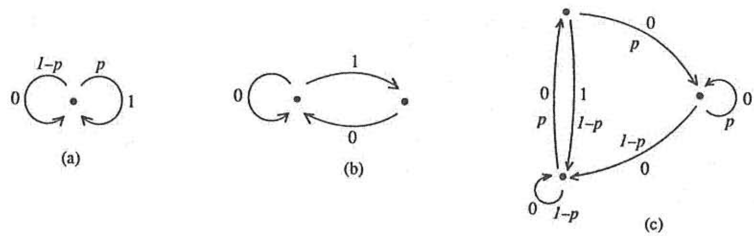

Figure 2: Probabilistic automata for rule 4, giving (a) the initial state, (b) the limit set, (c) the invariant measure.

$$\begin{aligned} a_n &= (1-p)a_{n-1} + \sum_{k=3}^{n-1}(1-p)p^{k-1}a_{n-k} \\ &\Rightarrow \; a_n - a_{n-1} + p(1-p)a_{n-2} - p^2(1-p)a_{n-3} = 0 \end{aligned} \tag{3.4}$$

and the measure entropy can be expressed in terms of the solution to this difference equation as the following infinite series:

$$s_\mu = \frac{\sum_{n=1}^{\infty} a_n \log \frac{1}{a_n}}{\sum_{n=1}^{\infty}(n+1)a_n}. \tag{3.5}$$

As another simple example we consider rule 128, which maps 111 to 1 and all other length three blocks to 0. In this case we have arbitrarily long transients, and all block probabilities approach zero except for blocks consisting only of zeroes. The measures at the first few time steps are shown in figure 3 (the finite time sets can be found in [35]). At time t the measure is given by an automaton with $2t+1$ nodes, where the non-zero elements of the transition matrices $\mu(0)$ and $\mu(1)$ are given by $\mu(0)_{i,2} = 1-p$ for $i = 1 \ldots 2t+1$, $\mu(0)_{i,i+1} = \mu(0)_{2t+1,1} = p$ for $i = 2 \ldots 2t$, and $\mu(1)_{11} = p$. These nondeterministic automata could be converted to infinite deterministic automata of the form shown in figure 3(d), since once again the symbol 1 determines the state of the automaton. An expression for the measure entropy could in principle be given, since the transition probabilities obey a recursion relation $p(10^n|1) = p(10^{n-1}|1) - p^{2t+1}(1-p)p(10^{n-(2t+2)}|1)$, but numerical results are more easily obtained directly from the block probabilities given by the automaton.

Even though the algorithmic complexity ν' obviously diverges with t in this case, the entropy of the probability distribution on the nodes τ_1 remains finite. If we calculate the equilibrium probability distribution on the automaton giving the measure at time t, we find that

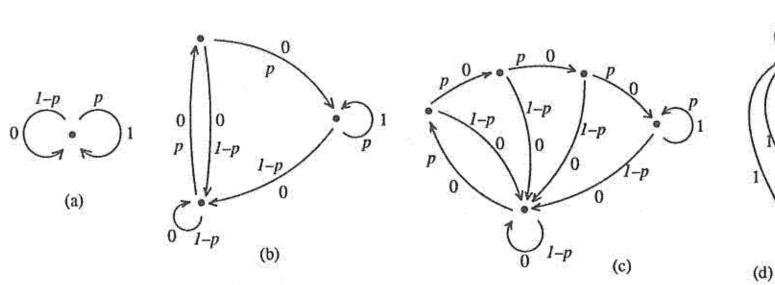

Figure 3: Automata giving the measures for rule 128, (a) at $t = 0$, (b) at $t = 1$, (c) at $t = 2$, (d) the structure of an equivalent infinite deterministic automaton.

$$\tau_1(t) = \frac{1-p^{2t}}{1-p}(p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}) = \frac{1-p^{2t}}{1-p} s_\mu(0), \tag{3.6}$$

which asymptotically approaches a constant. The inequality (2.20) then shows that τ_2 and η also remain bounded, and since it cannot be uniquely determined at which node a sequence with an initial segment of more than one zero started, τ_2 is strictly smaller than τ_1 at every finite t. Class 2 rules in general appear to be characterized by asymptotically finite τ_1, τ_2 and η. This seems reasonable, since for class 2 rules we expect that information in general only spreads over a finite distance, and the EMC should then remain finite.

If we now turn to the additive rule 90, where the value of a cell is given by the sum modulo 2 of its neighbors at the preceding time step, then if $p \neq 1/2$ the measures at finite time are given by the de Bruijn graphs shown for the first few time steps in figure 4. Starting with $p = 1/2$ one clearly remains at the invariant measure.

These graphs have 2^{2t} nodes at time t, and if each node is labeled by a binary address $a_1 a_2 \ldots a_{2t}$ we have the following transitions:

$a_1 a_2 \ldots a_{2t} \to a_2 \ldots a_{2t} 1$ with probability p,

labeled by $s = \phi^t(a_1 a_2 \ldots a_{2t} 1)$

and

$a_1 a_2 \ldots a_{2t} \to a_2 \ldots a_{2t} 0$ with probability $1-p$,

labeled by $1 - s = \phi^t(a_1 a_2 \ldots a_{2t} 0)$,

where ϕ is now considered as a mapping $\phi : \Sigma^{l+2r} \to \Sigma^l$ for arbitrary l. One can show that there are no pairs of equivalent nodes in this graph. The equilibrium probability distribution should satisfy

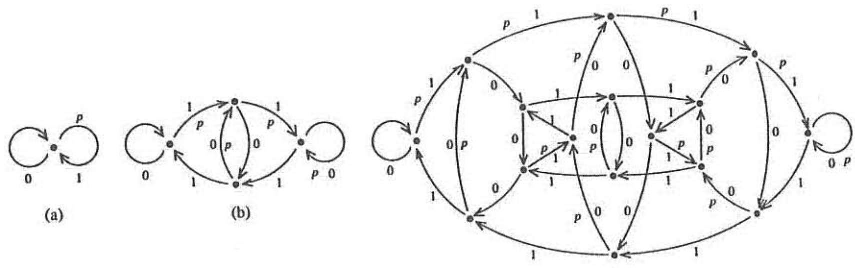

Figure 4: Automata giving the measures for rule 90 at (a) $t = 0$, (b) $t = 1$, and (c) $t = 2$.

$$p(a_1 \dots a_{2t}) = \begin{cases} p(p(0a_1 \dots a_{2t-1}) + p(1a_1 \dots a_{2t-1})) \text{ if } a_{2t} = 1 \\ (1-p)(p(0a_1 \dots a_{2t-1}) + p(1a_1 \dots a_{2t-1})) \text{ if } a_{2t} = 0, \end{cases} \tag{3.7}$$

which is solved by

$$p(a_1 \dots a_{2t}) = p^{n_1(a_1 \dots a_{2t})}(1-p)^{2t - n_1(a_1 \dots a_{2t})}, \tag{3.8}$$

where the function $n_1(w)$ counts the number of ones in w. If we then calculate the entropy of this distribution at time t, we find that the quantity τ_1 introduced above increases linearly in time when $p \neq 1/2$,

$$\begin{aligned} \tau_1(\alpha, t) &= \frac{1}{1-\alpha} \log\left(\sum_{a_1 \dots a_{2t}} p^\alpha(a_1 \dots a_{2t}) \right) \\ &= 2t \frac{1}{1-\alpha} \log(p^\alpha + (1-p)^\alpha) \\ &= 2t\, s(\alpha, 0). \end{aligned} \tag{3.9}$$

We have not yet shown that these automata actually minimize the entropy for the probability distribution on nodes, but we shall now calculate the effective measure complexity η, which turns out to equal (3.9), and the inequality (2.20) then shows that this is the minimum value. This means that we should calculate the rate of convergence for the finite length block entropies S_n at arbitrary times. We first take a single step forward in time and consider $t = 1$, and the calculation is done for rule 60 (where $a_i(t) = (a_{i-1}(t-1) + a_i(t-1)) \bmod 2$) instead of rule 90. Since an uncorrelated initial state is used, the results for rule 90 can be obtained by considering two independent copies of the time evolution of rule 60. For rule 60, each infinite sequence has exactly two predecessors, and the block probabilities at $t = 1$ are

$$p(\sigma = \sigma_1 \ldots \sigma_m) = \sum_{\sigma' : \phi(\sigma') = \sigma} p^{n_1(\sigma')}(1-p)^{m+1-n_1(\sigma')}, \tag{3.10}$$

and the block entropies at $t = 1$ are given by

$$S_m(\alpha) =$$

$$\frac{1}{1-\alpha}\log(\frac{1}{2}\sum_{j=0}^{m+1}\binom{m+1}{j}(p^j(1-p)^{m+1-j} + (1-p)^j p^{m+1-j})^\alpha). \tag{3.11}$$

Using Stirling's formula, we find that the expression summed over has two maxima (when $\alpha > 0$ and $p \neq 1/2$) at

$$j = \frac{p^\alpha}{p^\alpha + (1-p)^\alpha}(m+1) + \frac{1}{2}\frac{p^\alpha - (1-p)^\alpha}{p^\alpha + (1-p)^\alpha} + O(1/m), \tag{3.12}$$

and at the value of j obtained by interchanging p and $(1-p)$. The sum can then be approximated by two Gaussian integrals, and after some amount of calculation it is found that

$$\frac{1}{m}S_m(\alpha, t=1) = (1 + \frac{1}{m})\log(p^\alpha + (1-p)^\alpha) = (1 + \frac{1}{m})s(\alpha, t=0) \tag{3.13}$$

to order $1/m$. This can be extended to arbitrary times by first noting that for $t = 2^n$, where $a_i(t) = (a_{i-t}(0) + a_i(0)) \bmod 2$, we can use (3.13) to obtain

$$\eta(\alpha, t = 2^n) = 2^n s(\alpha, t = 0). \tag{3.14}$$

For the additive rules we are discussing, $s(\alpha)$ is constant in time, which implies that the change in $\eta(\alpha)$ in one time step satisfies

$$\Delta\eta(\alpha) = \lim_{m\to\infty}(S_m(\alpha, t+1) - S_{m+2r}(\alpha, t) + 2rs(\alpha)) \leq s(\alpha). \tag{3.15}$$

This shows that the value of $\eta(\alpha)$ at $t = 2^n$ is the maximum value allowed, and this value can only be reached if $\Delta\eta(\alpha)$ is maximal at every step in time. At an arbitrary time t, we then have

$$\begin{aligned}\eta(\alpha, t) = &\quad t \cdot s(\alpha, 0) \quad \text{for } \alpha > 0 \text{ and } p \neq \frac{1}{2} \\ = &\quad 0 \qquad\quad \text{for } \alpha = 0 \text{ or } p = \frac{1}{2}.\end{aligned} \tag{3.16}$$

For rule 90 we get

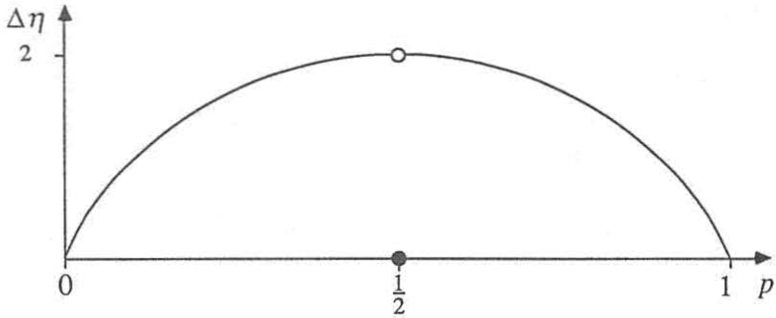

Figure 5: The increase of the effective measure complexity in one time step for rule 90, shown as a function of the density of the initial state.

$$\begin{aligned} \eta(\alpha,t) = &\ 2t \cdot s(\alpha,0) \quad \text{for } \alpha > 0 \text{ and } p \neq \frac{1}{2} \\ = &\ \quad 0 \qquad\quad \text{for } \alpha = 0 \text{ or } p = \frac{1}{2}, \end{aligned} \tag{3.17}$$

which equals (3.9), so that in particular $\tau_1 = \tau_2 = \eta$ at any t in this case. The dependence of $\Delta\eta$ on the initial density p is illustrated in figure 5.

The discontinuity at $p = 1/2$ might seem peculiar at first, but it can be understood in the following way. Suppose an infinite sequence ...01100101 given by the automaton is known up to a certain point. When $p \neq 1/2$, the state of the automaton can always be determined with probability 1, but if p is increased towards 1/2, a larger segment of the sequence is necessary to estimate the state. This means that η increases as p approaches 1/2, since the information is contained in longer blocks. When p is exactly 1/2, all correlations disappear, the automaton collapses to a single node, and the EMC changes discontinuously to zero.

If we finally consider the time evolution of an arbitrary cellular automaton rule with s symbols per site and range r, starting from an uncorrelated initial state, the nondeterministic automaton giving the measure at time t has at most s^{2rt} nodes, which means that

$$\nu' = \log_s N(A) \leq 2rt, \tag{3.18}$$

and from (2.20) it then follows that the effective measure complexity is bounded by

$$\eta \leq 2rt. \tag{3.19}$$

For rules with a positive sum of left and right Lyapunov exponents, we expect that asymptotically the EMC, which measures a product of average correlation length and total correlational information, in general will increase linearly.

4. Asymptotic behavior of cellular automata

For class 2 cellular automaton rules, some observations on the asymptotic behavior were made in the previous section. In that case, the attractor [61], the set of sequences that dominates the asymptotic behavior and represents the behavior for typical initial states, is in general considerably simpler than the limit set, i.e. the intersection of all finite time sets. This could conceivably be the case for some chaotic cellular automaton rules as well, even though the algorithmic complexity of their finite time sets grows very rapidly in time, indicating that the limit set is more complicated in this case (limit sets of cellular automata in general need not even correspond to recursively enumerable sets [60]). If the attractor happened to correspond to a regular or a context-free language, it might be possible to deduce its grammatical structure from the results of simulations, using various inductive inference methods [3].

Some general conclusions on the structure of the attractor (invariant measure) can be drawn from numerical data such as Fourier spectra [61] or finite length block entropies. Here we shall discuss what can be learned about the structure of the attractor from numerical data for the block entropies of rule 22. This cellular automaton rule, which maps 100, 010 and 001 to 1, and other length 3 blocks to 0, has been extensively studied as an example of a rule showing chaotic behavior, both by direct simulations [36,62], and in the local structure theory approximation [42]. These results indicate that this rule is strongly mixing, and that there is rapid convergence to an invariant measure. The space-time patterns generated contain significant long-range correlations; in [36], algebraic decay towards zero was found for the finite length block entropies ΔS_n (i.e. for $\alpha = 1$),

$$\Delta S_n \approx c \cdot n^{\beta-1} \tag{4.1}$$

both for spatial and temporal block entropies, with different critical exponents $\beta < 1$. Figure 6 shows the finite length spatial block entropies $S_n(\alpha)$ for $\alpha = 0$, 1 and 2 up to length $n = 18$ obtained in a fairly small scale simulation of rule 22. Periodic boundary conditions were used on a lattice of length 50 000 (the data shown were taken at $t = 10000$). The data for $\alpha = 1$ and $\alpha = 2$ are well fitted by

$$S_n(\alpha) \approx c(\alpha) \cdot n^{\beta} \tag{4.2}$$

with $\beta = 0.95 \pm 0.01$, which is in agreement with the results in [36] for $\alpha = 1$. The critical exponent β does not appear to depend significantly on α in this regime. A more extensive numerical investigation of this phenomenon is in

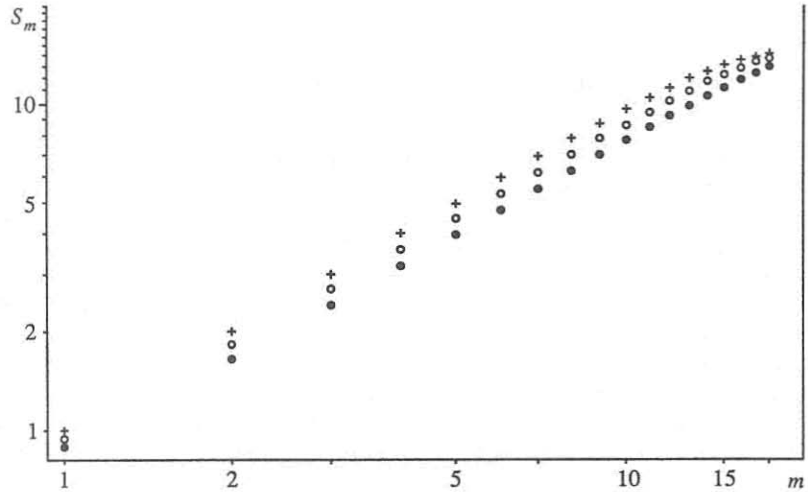

Figure 6: Finite length block entropies $S_m(\alpha)$ for rule 22, where (+) stands for $\alpha = 0$, (o) for $\alpha = 1$, and (•) for $\alpha = 2$.

progress. At any finite time, the block entropies ΔS_n decay exponentially if n is sufficiently large, but assuming that the behavior in (4.2) characterizes the attractor, we can immediately conclude from (2.30) that the attractor cannot correspond to a regular language. More precisely, this means that finitely generated measures (corresponding to regular languages) of the form discussed in section 2 are excluded. If more complicated measures, possibly without finite specification, on a regular language L are considered, any allowed behavior for the block entropies could be reproduced, since the measure could be (approximately) concentrated to a subset of L corresponding to a non-regular language.

Our remarks on context-free languages are more speculative. For unambiguous context-free languages, we know from (2.31) that in the topological case, finite length topological block entropies generically converge as $\log(m)/m$. The topological entropy both converges more slowly and is more sensitive to finite size effects than the measure entropy (as is evident in figure 6), and our numerical data do not admit any firm conclusions in this case. It is still of interest to understand the scaling behavior (2.31) better, in particular since one may expect that the simplest measures associated to CFLs should show qualitatively similar behavior for $\alpha = 0$ and $\alpha = 1$. This would be the case for any measure given by a stochastic push-down automaton with a unique start node (and such measures could then be excluded by the results for $\alpha = 1$ and $\alpha = 2$). Probability distributions on CFLs have been constructed in this way [39], but to define a measure on infinite strings one should impose Kolmogorov's consistency conditions, which gives rise to strong restrictions. In fact, for a deterministic CFL (the deterministic CFLs form a subset of the unambiguous CFLs where the accepting push-down automaton is deterministic, i.e. two transitions from a node cannot be labeled by identical symbols), no string in L can be a prefix of any other, and it is thus impossible to define a measure on infinite strings in this case. Measures corresponding to context-free and higher languages should be further investigated.

A qualitative understanding of the context-free language growth function (2.31) can be obtained by considering diffusion on the tree of finite automata drawn in figure 1(b) as a representation of a push-down automaton. We first consider a case with a fixed number of allowed transitions from each state, and with an unambiguous (but not necessarily deterministic) CFL, so that each path beginning with the start symbol S on the stack, and ending with empty stack, i.e. at the absorbing top node ϕ, gives a unique word in the language. If the k allowed transitions at each node are given equal probabilities, the probability that a diffusing particle, initially located at the start node with S on the stack, is absorbed at the top node at discrete time $t = n$ is equal to

$$p(S \to \phi, n) = \frac{g(n)}{k^n}, \tag{4.3}$$

where $g(n)$ is the number of words of length n in the language. In a rough approximation we could neglect the internal structure of the finite automata in the tree and instead introduce effective transition probabilities, obtained by averaging over L, from an automaton to itself and those connected to it. The transition probabilities from a node (automaton) then only depend on the top element of the stack at that node, so the tree is self-similar.

Diffusion on the backbone of a tree (as opposed to diffusion among the leaves at a certain level as in [33]) has been treated by several authors [63-65], e.g. in connection with chaotic transport in Hamiltonian systems. The Markov tree model in [65] of transport across cantori (see also [66] for a discussion of 1/f noise in this context) is in fact almost identical to our model of a push-down automaton accepting a context-free language; we only need to modify the transition probabilities. The self-similarity of the tree can then be used (as in [65]) to derive a system of algebraic equations for the Laplace transforms of the transition probabilities $p(a \to \phi)(t)$, (where nodes are labeled by stack content, so that a is a node with a single symbol a from the stack alphabet on the stack, and ϕ is the absorbing top node). The time dependence found by transforming the solution of this system back generically corresponds to asymptotic power law relaxation (possibly multiplied by an exponential factor) on the tree, which in turn corresponds to the form (2.31) for the CFL growth function. For a more general automaton, where the number of transitions at a node is not necessarily constant, this approximation can be modified by including an absorption probability at each node with fewer transitions than the maximum, so that all words of equal length still have the same weight. When diffusion on the tree is considered, an effective absorption probability at each node can be included without changing the functional form of the solution.

If, on the contrary, the scaling behavior of (4.1) and (4.2) was found for the topological entropy, this would give a growth function

$$g(n) = c\, e^{\gamma n^{\beta}} \tag{4.4}$$

with $\beta < 1$ (for entropy different from zero this would be multiplied by an exponential factor). This correspond to Kohlrausch relaxation on the tree. It does not appear to be known whether a context-free language can have a growth function with asymptotic behavior of the form given by (4.4) [57]. This could only happen for an inherently ambiguous language, where several paths in the tree correspond to the same word.

In any case, block entropies vanishing according to (4.1) appear not to be characteristic of unambiguous context-free languages, but it should then be noted that our understanding of measures associated with context-free and higher languages is very limited.

References

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

[2] P. Cvitanović, ed., *Universality in Chaos*, (Adam Hilger, Bristol, 1984).

[3] D. Angluin and C. H. Smith, "Inductive Inference: Theory and Methods", *ACM Computing Surveys*, **15** (1983) 237–269.

[4] J. P. Crutchfield and B. S. McNamara, "Equation of Motion from a Data Series", *Complex Systems*, **1** (1987) 417–452.

[5] J. D. Farmer and J. J. Sidorowich, "Predicting Chaotic Time Series", *Physical Review Letters*, **59** (1987) 845–848.

[6] P. Bak, C. Tang, and K. Wiesenfeld, "Self-Organized Criticality: An Explanation of 1/f Noise", *Physical Review Letters*, **59** (1987) 381–384.

[7] T. Toffoli and N. Margolus, *Cellular Automata Machines*, (MIT Press, Cambridge MA, 1987).

[8] S. Wolfram, "Origins of Randomness in Physical Systems", *Physical Review Letters*, **55** (1985) 449–452.

[9] A. Rényi, "On Measures of Entropy and Information", in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, (University of California Press, Berkeley-Los Angeles, 1961) 547–561.

[10] J. Ziv and A. Lempel, "Compression of Individual Sequences via Variable-Rate Encoding", *IEEE Transactions on Information Theory*, **IT-24** (1978) 530–536.

[11] J. Rissanen, "A Universal Data Compression System", *IEEE Transactions on Information Theory*, **IT-29** (1983) 656–664.

[12] P. Grassberger, "Estimating the Information Content of Symbol Sequences and Efficient Codes", University of Wuppertal preprint WU B 87-11 (1987).

[13] P. Martin-Löf, "The Definition of Random Sequences", *Information and Control*, **9** (1966) 602–619.

[14] D. E. Knuth, *The Art of Computer Programming*, vol. 2, (Addison-Wesley, Reading MA, 1981).

[15] G. J. Chaitin, *Algorithmic Information Theory*, (Cambridge University Press, 1987).

[16] R. J. Solomonoff, "A Formal Theory of Inductive Inference", *Information and Control*, **7** (1964) 1–22.

[17] A. N. Kolmogorov, "Three Approaches to the Definition of the Concept 'Quantity of Information'", *Problemy Peredachii Informatsii*, **1** (1965) 3–11, (in Russian).

[18] G. J. Chaitin, "On the Length of Programs for Computing Finite Binary Sequences", *Journal of the Association for Computing Machinery*, **13** (1966) 547–569.

[19] A. K. Zhvonkin and L. A. Levin, "The Complexity of Finite Objects and the Development of the Concepts of Information and Randomness by Means of the Theory of Algorithms", *Russian Mathematical Surveys*, **25** (1970) 83–124.

[20] K. Lindgren, "Correlations and Random Information in Cellular Automata", *Complex Systems*, **1** (1987) 529–543.

[21] J. Hartmanis, "Generalized Kolmogorov Complexity and the Structure of Feasible Computations", *Proceedings of the 24th IEEE Symposium on the Foundations of Computer Science*, (1983) 439–445.

[22] M. Sipser, "A Complexity Theoretic Approach to Randomness", *Proceedings of the 15th ACM Symposium on the Theory of Computing* (1983), 330–335.

[23] J. Rissanen, "Universal Coding, Information, Prediction, and Estimation", *IEEE Transactions on Information Theory*, **IT-30** (1984) 629–636.

[24] J. Rissanen, "Complexity of Strings in the Class of Markov Sources", *IEEE Transactions on Information Theory*, **IT-32** (1986) 526–532.

[25] A. C. Yao, "Theory and Applications of Trapdoor Functions", *Proceedings of the 23rd IEEE Symposium on the Foundations of Computer Science*, (1982) 80–91.

[26] N. H. Packard and S. Wolfram, "Two-Dimensional Cellular Automata", *Journal of Statistical Physics*, **38** (1985) 126–171.

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

[28] B. A. Hubermann and T. Hogg, "Complexity and Adaptation", *Physica*, **22D** (1986) 376–384.

[29] G. Györgyi and P. Szépfalusy, "Calculation of the Entropy in Chaotic Systems", *Physical Review* **A31** (1985) 3477–3479.

[30] P. Szépfalusy and G. Györgyi, "Entropy Decay as a Measure of Stochasticity in Chaotic Systems", *Physical Review* **A33** (1986) 2852–2855.

[31] C. H. Bennett, "On the Nature and Origin of Complexity in Discrete, Homogeneous, Locally-Interacting Systems", *Foundations of Physics*, **16** (1986) 585–592.

[32] G. Parisi, "Facing Complexity", *Physica Scripta*, **35** (1987) 123–124.

[33] C. P. Bachas and B. A. Hubermann, "Complexity and Ultradiffusion", *Journal of Physics A: Mathematical and General*, **20** (1987) 4995–5014.

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

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

[36] P. Grassberger, "Long-Range Effects in an Elementary Cellular Automaton", *Journal of Statistical Physics*, **45** (1986) 27–39.

[37] L. L. Campbell, "A Coding Theorem and Rényi's Entropy", *Information and Control*, **8** (1965) 423–429.

[38] K.-E. Eriksson, K. Lindgren and B. Å. Månsson, *Structure, Context, Complexity, Organization*, (World Scientific, Singapore, 1987).

[39] K. S. Fu, *Syntactic Pattern Recognition and Applications*, (Prentice-Hall, Englewood Cliffs NJ, 1982)

[40] B. Kitchens and S. Tuncel, "Finitary Measures for Subshifts of Finite Type and Sofic Systems", *Memoirs of the American Mathematical Society*, **58** (1985) no. 338.

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

[42] H. A. Gutowitz, J. D. Victor and B. W. Knight, "Local Structure Theory for Cellular Automata", *Physica*, **28D** (1987) 18–48.

[43] B. Marcus, K. Petersen and S. Williams, "Transmission Rates and Factors of Markov Chains", in *Conference in Modern Analysis and Probability, Contemporary Mathematics* **26**, (American Mathematical Society, Providence RI, 1984), 279–293.

[44] E. M. Gold, "Identification in the Limit", *Information and Control*, **10** (1967) 447–474.

[45] J. J. Horning, "A Study of Grammatical Inference", Ph.D. dissertation, Computer Science Dept., Stanford University, 1969.

[46] E. M. Gold, "Complexity of Automaton Identification from Given Data", *Information and Control*, **37** (1978) 302–320.

[47] D. Angluin, "On the Complexity of Minimum Inference of Regular Sets", *Information and Control*, **39** (1978) 337–350.

[48] D. Angluin, "Inference of Reversible Languages", *Journal of the Association for Computing Machinery*, **29** (1982) 741–765.

[49] P. Grassberger, "On Symbolic Dynamics of One-Humped Maps of the Interval", University of Wuppertal preprint WU B 87-3 (1987).

[50] A. Salomaa and M. Soittola, *Automata-Theoretic Aspects of Formal Power Series*, (Springer-Verlag, New York, 1978).

[51] M. P. Schützenberger, "On a Theorem of R. Jungen", *Proceedings of the American Mathematical Society*, **13** (1962) 885–890.

[52] K. Mahler, "On the Taylor Coefficients of Rational Functions", *Proceedings of the Cambridge Philosophical Society*, **52** (1956) 39–48.

[53] O. H. Ibarra and B. Ravikumar, "On Sparseness, Ambiguity and other Decision Problems for Acceptors and Transducers" in STACS 86, 3rd Annual Symposium on Theoretical Aspects of Computer Science, Lecture Notes in Computer Science 210, (Springer-Verlag, Berlin Heidelberg, 1986), 171–179.

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

[55] J. P. Crutchfield and N. H. Packard, "Symbolic Dynamics of One-Dimensional Maps: Entropies, Finite Precision, and Noise", *International Journal of Theoretical Physics*, **21** (1982) 433–466.

[56] 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.

[57] 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.

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

[59] D. A. Lind, "Applications of Ergodic Theory and Sofic Systems to Cellular Automata", *Physica*, **10D** (1984) 36–44.

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

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

[62] P. Grassberger, "Some More Exact Enumeration Results for 1D Cellular Automata", *Journal of Physics A: Mathematical and General*, **20** (1987) 4039–4046.

[63] S. Grossmann, F. Wegner and K. H. Hoffmann, "Anomalous Diffusion on a Self-Similar Hierarchical Structure", *Journal de Physique Lettres*, **46** (1985) L575–L583.

[64] A. Erzan, S. Grossmann and A. Hernández-Machado, "Diffusion on Random Hierarchical Structures", *Journal of Physics A: Mathematical and General*, **20** (1987) 3913–3934.

[65] J. D. Meiss and E. Ott, "Markov Tree Model of Transport in Area-Preserving Maps", *Physica*, **20D** (1986) 387–402.

[66] T. Geisel, A. Zacherl and G. Radons, "Generic 1/f Noise in Chaotic Hamiltonian Dynamics", *Physical Review Letters*, **59** (1987) 2503–2506.