Complex Systems **5** (1991) 381–399

On the Relationship between Complexity and Entropy for Markov Chains and Regular Languages

Wentian Li*
Santa Fe Institute, 1660 Old Pecos Trail, Suite A,
Santa Fe, NM 87501, USA

Abstract. Using the past-future mutual information as a measure of complexity, the relation between the complexity and the Shannon entropy is determined analytically for sequences generated by Markov chains and regular languages. It is emphasized that, given an entropy value, there are many possible complexity values, and vice versa; that is, the relationship between complexity and entropy is not one-to-one, but rather many-to-one or one-to-many. It is also emphasized that there are structures in the complexity-versus-entropy plots, and these structures depend on the details of a Markov chain or a regular language grammar.

1. Introduction

It has been recognized recently that in order to describe complex dynamical behavior such as the evolution of life on earth—in which the system becomes more sophisticated instead of more random—one needs to define a quantity called *complexity* [2, 7, 15, 19]. The meaning of this quantity should be very close to certain measures of difficulty concerning the object or the system in question: the difficulty of constructing an object, the difficulty of describing a system, the difficulty of reaching a goal, the difficulty of performing a task, and so on. The definition of complexity cannot be unique, simply because there are many different ways to quantify these difficulties, and also because there are many different tasks concerning which the difficulties are to be quantified.

For example, suppose we are interested in the complexity of an infinitely long one-dimensional symbolic sequence. We can find at least two tasks that can be performed on the sequence—to reconstruct the sequence and decode the sequence—and we can ask the question of how difficult it is to reconstruct or decode the sequence. By "reconstruct the sequence" it is not

*Current address: Physics Department, Rockefeller University, 1230 York Avenue, New York, NY 10021.

clear whether one wants to reconstruct the sequence to be exactly correct or approximately correct. Since the difficulty for performing these two tasks can be quite different for the same sequence, one has to specify which task is to be performed. Suppose we have finally decided to perform one task, for example, to reconstruct the sequence exactly. We still have to choose which quantity is appropriate for measuring the difficulty of performing this particular task.

Different proposals of the measure of complexity are in fact different ways to measure this difficulty, or the amount of effort put in, or the cost, of generating the sequence. One can measure the length of the shortest program that generates the sequence (*algorithmic complexity* [3, 11, 18]); or the number of instructions to generate one symbol in the sequence from the *shortest* program (one of the *computational complexities*); or the number of instructions to generate one symbol in the sequence from a short, but not necessarily the shortest, program (*logical depth* [2, 4]); and the list goes on. If we choose the task of generating the sequence to be approximately correct, many probability-based or information-theoretic quantities can be used to measure the difficulty for reconstructing the sequence—for example, the average amount of information stored in order to reproduce the next symbol on the sequence correctly "on average" [7, 20].

It has been pointed out repeatedly that algorithmic complexity does not fit our intuitive notion of complexity [2, 7, 9]. One vague argument is that algorithmic complexity measures the difficulty of the task—to reconstruct a sequence exactly—that is not very "meaningful." To say that this task is not very meaningful is partly based on our daily experience that we can pick up the main feature of a picture or the basic meaning of a text by quickly scanning it without paying attention to every detail. A similar argument for algorithmic complexity not being a "good" measure of complexity is that the effort put in to generate a sequence to be exactly correct does not contribute to the "important features" of that sequence. Rather, they contribute too much to the "trivial" details.

This observation states that an intuitively satisfactory definition of complexity should measure the amount of effort put in that generates *correlations* in a sequence. Of course, one cannot be sure that all the effort is spent on generating correlations. As a result, a measure of correlation typically provides a lower bound of a measure of complexity [7], and might be a reasonable estimate of the complexity. But the two quantities may not be equal to each other. This idea was indeed discussed, explicitly or implicitly, by Bennett [2], Chaitin [5], and Grassberger [7], among others.

A question naturally arises: how is an intuitively satisfactory definition of complexity related to a definition of complexity that describes the difficulty of a task for specifying every detail exactly, such as algorithmic complexity? Since it is well known that algorithmic complexity is equivalent to entropy as a measure of randomness, our question becomes: what is the relationship between the complexity and the entropy? Before answering this question, we should ask two other questions. First, does such a relationship exist? Second,

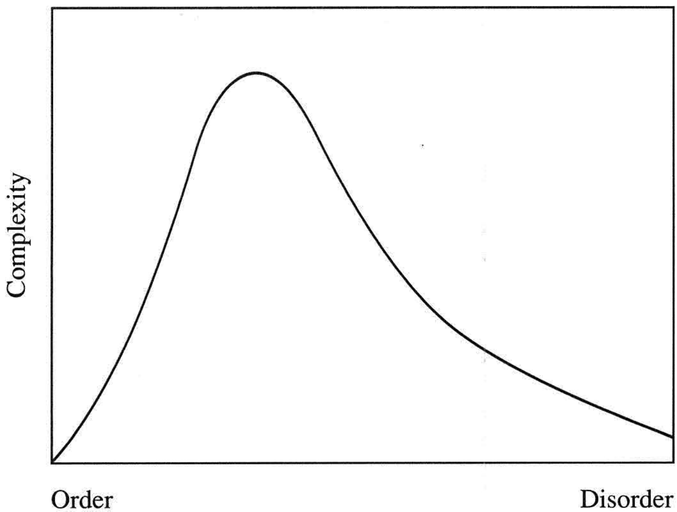

Figure 1: Proposed schematic relationship between complexity and entropy (after reference [9].)

does such a relationship depend on the sequence studied, or the definition of complexity used?

Several authors speculate that the typical relationship between complexity and entropy is a unimodal one: the complexity values are small for small and large entropy values, but large for intermediate entropy values [7, 9]. Hogg and Huberman [9], in particular, drew a function schematically as described above (see figure 1). More recently, Crutchfield and Young [6] and Langton [12] each plotted complexity versus entropy from some real numerical simulations. Although the two groups used different definitions of complexity, different definitions of entropy, and different ensemble of sequences, their plots are strikingly similar.

I will briefly discuss their results in the last section. But the main goal of this paper is to show that it is possible to determine analytically a relationship between complexity and entropy with a specific choice of the complexity and a specific choice of the sequence ensemble. It will be shown that there is no one-to-one relationship between complexity and entropy, but rather a one-to-many or many-to-one relationship between complexity and entropy, which strongly depends on the sequence ensemble being studied. The conclusion from this study seems to be that there is no universal relationship between complexity and entropy independent of the underlying sequences.

The paper is organized as follows: section 2 re-introduces the definition of complexity used here and specifies which sequence ensemble I am going to use; section 3 determines analytically the relationship between complexity and entropy for one-step two-symbol Markov chains (and the result will be presented graphically); section 4 presents the complexity-entropy relationship for two ensembles of sequences generated by regular languages; and section 4 discusses some complexity-entropy relationships studied by others.

2. Definitions

Since measures of the amount of correlation existing in the sequence provide at least lower bounds of the complexity, I will use the "past-future mutual information" as the measure of complexity (for a discussion of mutual information, see [16, 14]). This quantity was used by Shaw in his study of the symbolic sequences generated by dynamical systems [17], and later used by Grassberger as a measure of complexity (he called it *effective measure complexity* [7]). The definition is as follows. Partition an infinitely long sequence in half. The two half-infinite sequences can be considered two patches on the sequence whose lengths are M and N, respectively, in the limit as M and N go to infinity. The block entropy of a patch S^N with length N is

$$H(S^N) = -\sum_{S^N} P(S^N) \log_K P(S^N), \tag{1}$$

where $P(S^N)$ is the probability of a certain block configuration with length N, and the base of the logarithm (K) is the total number of symbols (e.g., $K = 2$ for a binary sequence). The past-future mutual information C is

$$C = \lim_{N,M\to\infty} [H(S^M) + H(S^N) - H(S^{M+N})]. \tag{2}$$

The Shannon entropy is the entropy per symbol in the infinite block-length limit:

$$h = \lim_{N\to\infty} [H(S^{N+1}) - H(S^N)]. \tag{3}$$

Both the past-future mutual information C and the Shannon entropy h have nice geometric interpretations in the block entropy $H(S^N)$ plot: h is the value of the slope of the limiting straight line that fits $H(S^N)$ as $N \to \infty$; and C is the value of the y-intercept of this straight line [17] (see figure 2). The proof is very simple. The limiting straight line that fits $H(S^N)$ is

$$\lim_{N\to\infty} H(S^N) = C + hN + \epsilon(N), \tag{4}$$

where $\epsilon(N)$ consists of all terms that increase less than linearly with N. Then the trivial equations

$$\begin{aligned} (C + h(N+1)) - (C + hN) &= h \\ (C + hM) + (C + hN) - (C + h(N+M)) &= C \end{aligned} \tag{5}$$

guarantee that the slope and the y-intercept are indeed the Shannon entropy and the past-future mutual information. If one already knows the Shannon entropy h, the past-future mutual information C can be derived from equation (4):

$$C = \lim_{N\to\infty} (H(S^N) - hN), \tag{6}$$

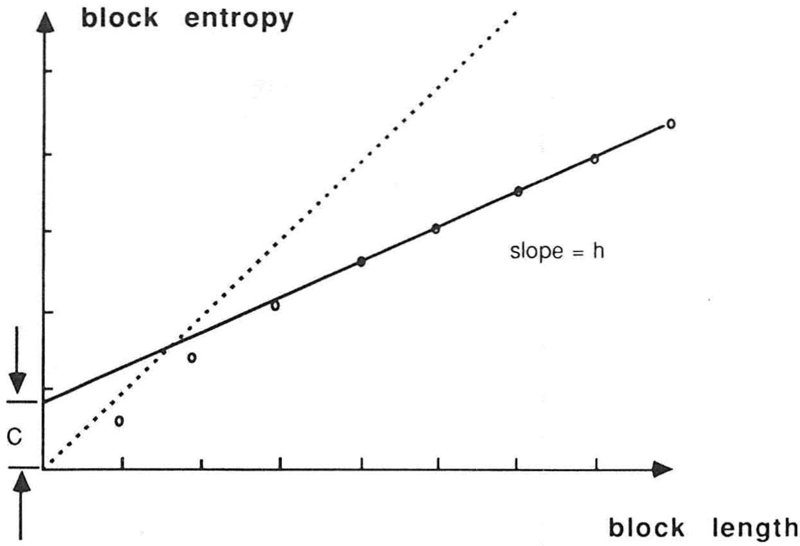

Figure 2: Illustration of the block entropy $H(S^N)$ as a function of the block length N. The slope of the limiting straight line is equal to the Shannon entropy h; and the y-intercept of the line is equal to the past-future mutual information, or the complexity C (plotted according to a similar figure in reference [17].)

a formula that was once used as the definition of C [7].

If the asymptotic behavior of $H(S^N)$ is not linear—for example, if $H(S^N)$ increases as a power-law function with the scaling exponent smaller than 1 or as a logarithmic function—then our definition leads to zero h and infinite C. This is the case when there are long-range correlations present in the sequences. For example, this occurs in the symbolic dynamics at the accumulation point of the period-doubling bifurcations [7]. When this happens, in principle one should use other definitions of complexity to further distinguish different divergent Cs, a subject being actively studied [1].

In this paper, I deliberately avoid the divergence of C by studying sequences with short-range correlations. In these sequences, the correlation as a function of the distance between two symbols decays exponentially, and the amount of correlation at long distances is negligible. As a result, the past-future mutual information is always finite. Because of the short-range correlation, the $H(S^N)$ may converge to the limiting straight line at a finite N, and both C and h can be easily calculated. These seemingly trivial calculations nevertheless give us a handle with which to study exactly the complexity-entropy relationship.

The sequences with short-range correlations are modeled by Markov chains (in the study of stochastic processes) and by regular languages (in formal language theory).[1] In both Markov chains and regular languages, the probability of a given symbol is completely determined by the preceding symbol (for one-step Markov chains) or by the preceding patch of symbols (for regular languages). Once we know all the transition probabilities, we are able to calculate the block entropy. I will start with one-step Markov chains in the next section.

3. Complexity-entropy relationship for one-step Markov chains

Since a Markov chain specifies the transition probabilities of having new symbols knowing the past history, it is straightforward to relate the block entropy of block length $N+1$ with that of block length N:

$$\begin{aligned} H(S^{N+1}) \;-\; & H(S^N) \\ = & -\sum_{S^{N+1}} P(S^{N+1}) \log_K P(S^{N+1}) + \sum_{S^N} P(S^N) \log_K P(S^N) \\ = & -\sum_{s_N} P(s_N) \left[\sum_{s_{N+1}} T(s_N \to s_{N+1}) \log_K T(s_N \to s_{N+1}) \right] \\ = & -\sum_{s} P(s) \log_K \prod_{s'} T(s \to s')^{T(s \to s')} \end{aligned} \tag{7}$$

where s_N is the last symbol of the block S^N and s_{N+1} is the last symbol of the block S^{N+1}. Since the earlier history does not count, the transition probability $T(s_N \to s_{N+1})$ is independent of N (the notation $T(s \to s')$ is used to indicate this fact). Similarly, the density of a symbol s should not depend on its location in the block, that is, $P(s_N) = P(s)$. The reader should be able to derive equation (7) easily, or consult any information theory book [10].

Because $T(s \to s')$ and $P(s)$ are not functions of block length N, $H(S^{N+1}) - H(S^N)$ does not depend on N. In other words, $H(S^N)$ approaches the limiting straight line as soon as N is larger than 1:

$$H(S^N) = H(S^1) + (N-1)h, \tag{8}$$

where $H(S^1) = -\sum_s P(s) \log_K P(s)$. This leads directly to the expression for C:

$$C = H(S^1) - h = -\sum_s P(s) \log_K \frac{P(s)}{\prod_{s'} T(s \to s')^{T(s \to s')}}. \tag{9}$$

If the maximum value $H(S^1) = 1$ is always reached (e.g., 0 and 1 appear with equal probability for binary sequences), the upper bound of C is $1 - h$, which is a straight line in the C-versus-h plot.

---

[1]The original purpose of formal language theory was to characterize a sequence as being *grammatically* correct [8], whereas here we are more interested in being *statistically* correct.

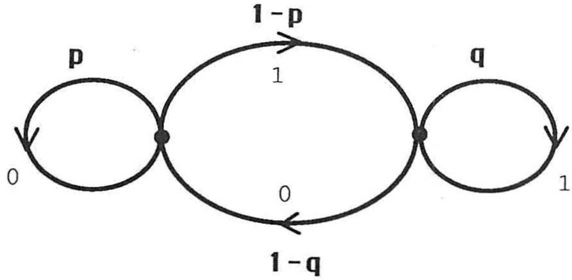

Figure 3: Graphic representation of the one-step two-symbol Markov chain. Each site in the sequence of the Markov chain is a link in this graph, and each sequence is a path. The transition probabilities are also included.

To be more specific, I will work out the example of the one-step two-symbol Markov chains. The transition probabilities are

$$\begin{aligned} T(0 \to 0) &= p \\ T(0 \to 1) &= 1-p \\ T(1 \to 0) &= 1-q \\ T(1 \to 1) &= q, \end{aligned} \tag{10}$$

which can also be represented as a transition matrix $\mathbf{T}$:

$$\mathbf{T} = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix} \end{matrix} \tag{11}$$

where $T_{ij} \equiv T(i \to j)$.

This Markov chain is graphically illustrated in the simple directed graph shown in figure 3. The densities of each symbol compose the left eigenvectors of the transition matrix $\mathbf{T}$ with respect to the largest eigenvalue 1:

$$P(0) = \frac{1-q}{T}, \quad P(1) = \frac{1-p}{T}, \tag{12}$$

where $T = 2 - p - q$ is the normalization factor. From these densities, we have

$$H(S^1) = \log_2(T) - \frac{1}{T}\log_2(1-q)^{1-q}(1-p)^{1-p}; \tag{13}$$

and from these densities as well as the transition probabilities, we have

$$h = -\frac{1-q}{T}\log_2 p^p(1-p)^{1-p} - \frac{1-p}{T}\log_2 q^q(1-q)^{1-q}. \tag{14}$$

By equation (9), the complexity is equal to

$$C = \log_2(T) + \frac{p(1-q)}{T}\log_2\frac{p}{1-q} + \frac{q(1-p)}{T}\log_2\frac{q}{1-p}. \tag{15}$$

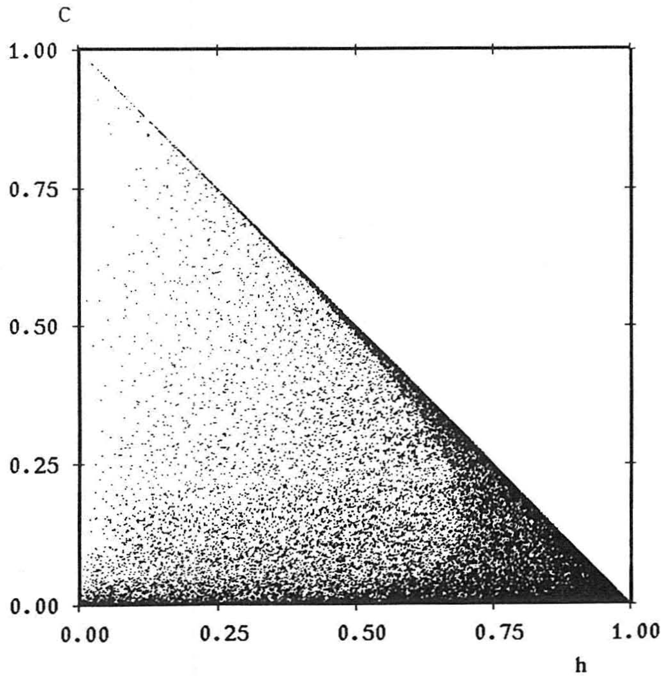

Figure 4: Plot of C versus h for sequences generated by the one-step two-symbol Markov chain. The transition probabilities p and q are chosen randomly.

Both C and h are functions of the parameters p and q. Fixing p and q, one gets a single point on the complexity-versus-entropy plot. For example, $p = q = 0$ gives a point at $h = 0$ and $C = 1$ (it is the periodic sequence $\ldots 010101 \ldots$). Different p, q may give the same h value but different C values (or vice versa), and we have the case of a many-to-one (or one-to-many) relationship between complexity and entropy. Figure 4 shows the C-versus-h plot when p and q are chosen randomly, and figure 5 shows the same plot when q is fixed and p is randomly chosen.

One main feature of this plot is that there are two branching areas with dense dots: one toward the origin ($h = 0$ and $C = 0$, corresponding to the all-zero or all-one sequences), and another toward the upper-left corner ($h = 0$ and $C = 1$, corresponding to the period-two sequences). Zero C is the case when the block entropy $H(S^1)$ is equal to the increment of block entropy $h = H(S^{N+1}) - H(S^N)$ (for all $N \geq 1$), so the limiting straight line passes through the origin. When this happens, the slope of the straight line can still be any value between 0 and 1, so h can be arbitrary even though C is kept at zero.

Similarly, zero h does not prevent C from being of any value. Whenever the sequence is periodic, $H(S^N)$ reaches a plateau. The slope of the limiting

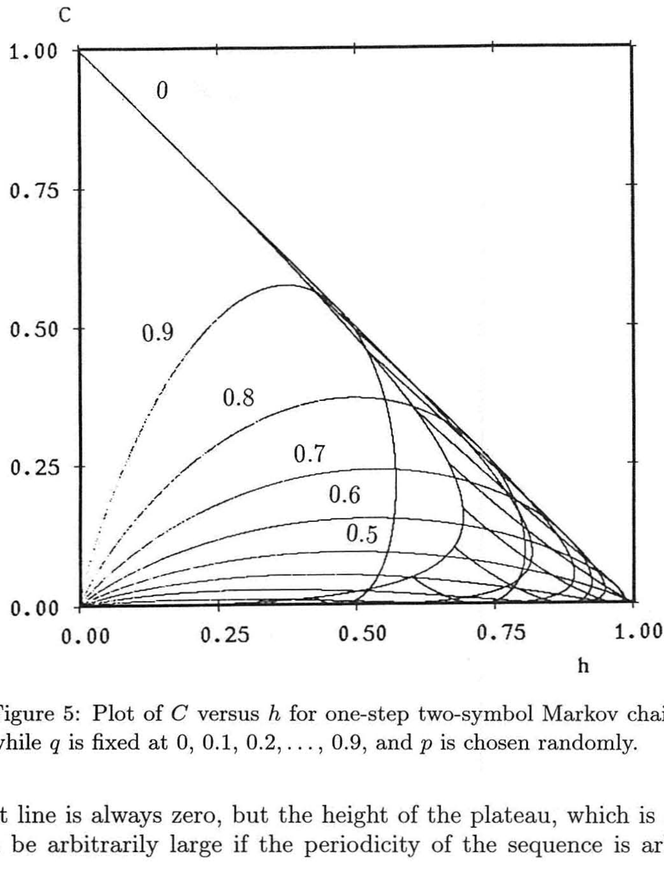

Figure 5: Plot of C versus h for one-step two-symbol Markov chains while q is fixed at 0, 0.1, 0.2, ..., 0.9, and p is chosen randomly.

straight line is always zero, but the height of the plateau, which is equal to C, can be arbitrarily large if the periodicity of the sequence is arbitrarily long.

By examining the above two extreme cases, it is easily understood why the complexity-entropy relationship can be quite arbitrary. In fact, the slope and the y-intercept are two independent parameters of a straight line. The only reason that the two should be related after all is because the underlying sequences can have certain structures such that the rise of $H(S^N)$ as N is increased is restricted in certain ways. In the next section, I will examine the complexity-entropy relationship when the underlying sequences are beyond the one-step Markov chains.

4. Complexity-entropy relationship for some regular languages

Regular languages are very similar to higher-order Markov chains except that different symbols in the corresponding Markov chain become the same symbol in the regular language [13]. Figure 6 shows an example of a regular language. Any path on the graph represents a grammatically correct word, and the collection of all paths that are followed according to the indicated transition probabilities represents an ensemble of sequences. Notice that

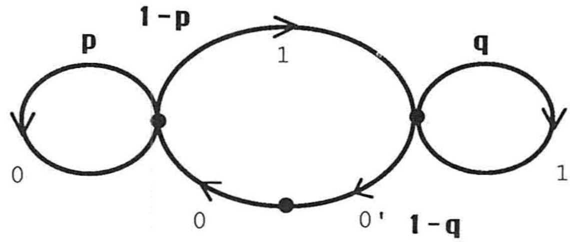

Figure 6: The first example of the regular language grammar discussed in section 4.

there are two 1s and three 0s on the graph. Both 1s inject into the same node so they represent the same history. Similarly, two out of the three 0s inject into the same node and they should also be the same. Nevertheless, one of the 0s (it is marked by $0'$) represents a different past history, though it is "invisible" in the sequence (i.e., it will not be marked as $0'$ in the sequence generated by this regular language grammar).

The procedures to calculate C and h for sequences generated by this regular language grammar is the following: (1) calculate the $H_M(S^N)$ (subscript M indicates the Markov chain) when the sequence is considered to consist of three symbols (0, $0'$, and 1); (2) determine how the introduction of the extra symbol ($0'$) overestimates the block entropy, where the overestimation is $\delta(N) = H_M(S^N) - H(S^N)$; (3) determine δ, which is either $\delta = \lim_{N\to\infty} \delta(N)$, or $\delta = \delta(N)$ (when $N \geq N_0$) if the limit is reached at some finite N; (4) finally, we have $h = h_M$ and $C = C_M - \delta$, where h_M and C_M are the Shannon entropy and past-future mutual information for the corresponding Markov chain, and h and C are those for the regular language.

To see how this works, consider the regular language grammar in figure 6. The transition matrix for the corresponding Markov chain is

$$\mathbf{T} = \begin{array}{c} \\ 0 \\ 0' \\ 1 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 0' & 1 \end{array} \\ \begin{pmatrix} p & 0 & 1-p \\ 1 & 0 & 0 \\ 0 & 1-q & q \end{pmatrix} \end{array}. \tag{16}$$

By calculating the left eigenvector corresponding to the eigenvalue 1, we have the densities of the three symbols:

$$P(0) = \frac{1-q}{T}, \quad P(0') = \frac{(1-p)(1-q)}{T}, \quad P(1) = \frac{1-p}{T}, \tag{17}$$

where $T \equiv 3 - 2p - 2q + qp$ is the normalization factor. The Shannon entropy and the past-future mutual information are

$$h_M = -\frac{1-q}{T} \log_2 p^p (1-p)^{1-p} - \frac{1-p}{T} \log_2 q^q (1-q)^{1-q} \tag{18}$$

and

$$\begin{aligned} C_M &= \log_2(T) + \frac{1-q}{T}\log_2 p^p + \frac{1-p}{T}\log_2 q^q \\ &\quad -\frac{1}{T}\log_2(1-p)^{1-p}(1-q)^{1-q} \end{aligned} \tag{19}$$

In order to determine $\delta(N)$, a brute-force counting of the degenerate blocks is carried out. For $N = 2$, there is only one block in the regular language, 00, that has more than one corresponding block in the Markov chain, 00 and 0′0. We have

$$\begin{aligned} P_M(00) &= P(0)p = \frac{(1-q)p}{T} \\ P_M(0'0) &= P(0') = \frac{(1-q)(1-p)}{T} \\ P(00) &= P_M(00) + P_M(0'0), \end{aligned} \tag{20}$$

again, with the subscript M indicating the corresponding Markov chain. It is easy to show that

$$\begin{aligned} \delta(2) &= -P_M(00)\log_2 P_M(00) - P_M(0'0)\log_2 P_M(0'0) \\ &\quad + P(00)\log_2 P(00) \\ &= -\frac{1-q}{T}(\log_2 p^p(1-p)^{1-p}) \end{aligned} \tag{21}$$

For $N = 3$, there are four blocks in the Markov chain (000, 0′00, 001, 0′01) that are actually two blocks (000, 001) in the regular language. The densities of these blocks are

$$\begin{aligned} P_M(000) &= P(0)p^2 = \frac{(1-q)p^2}{T} \\ P_M(0'00) &= P(0')p = \frac{(1-q)(1-p)p}{T} \\ P_M(001) &= P(0)p(1-p) = \frac{1-q)p(1-p)}{T} \\ P_M(0'01) &= P(0')(1-p) = \frac{(1-q)(1-p)^2}{T} \\ P(000) &= P_M(000) + P_M(0'00) \\ P(001) &= P_M(001) + P_M(0'01). \end{aligned} \tag{22}$$

And the overestimation of the block entropy at $N = 3$ is

$$\begin{aligned} \delta(3) &= -P_M(000)\log_2 P_M(000) - P_M(0'00)\log_2 P_M(0'00) \\ &\quad - P_M(001)\log_2 P_M(001) - P_M(0'01)\log_2 P_M(0'01) \\ &\quad + P(000)\log_2 P(000) + P(001)\log_2 P(001) \\ &= -\frac{1-q}{T}\log_2 p^p(1-p)^{1-p} = \delta(2) \end{aligned} \tag{23}$$

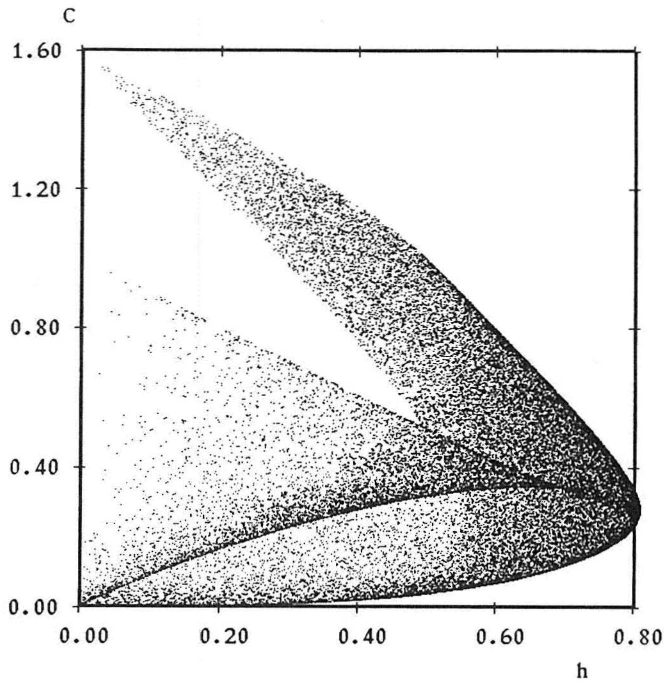

Figure 7: Plot of C versus h for sequences generated by the regular language grammar in figure 6 with randomly chosen p and q.

Similarly, by counting the degenerate situation for larger blocks, it can be shown that $\delta(N) = \delta(2)$, that is, the limit value of $\delta = \lim_{N\to\infty} \delta(N)$ is reached at $N = 2$. With the value of δ, the complexity of this regular language is

$$\begin{aligned} C &= C_M - \delta \\ &= \log_2(T) + \frac{2(1-q)}{T}\log_2 p^p + \frac{1-p}{T}\log_2 q^q - \frac{q}{T}\log_2(1-p)^{1-p} \\ &\quad -\frac{1}{T}\log_2(1-q)^{1-q} \end{aligned} \tag{24}$$

Figure 7 shows the complexity-entropy relation with p and q randomly chosen. One of the branches of dense dots at the upper-left corner represents the period-three sequence (...100100100...). It is curious that there are blank regions in the plot with no dots: one is below the period-three branch, and another is near the maximum entropy and minimum complexity point. Figures 8 and 9 show the "skeletons" of the relation presented in figure 7, with either q or p being fixed. By comparing figure 7 with figure 4, one can easily conclude that the complexity-entropy relationship strongly depends on the sequence ensemble being studied. Different grammars generally lead to

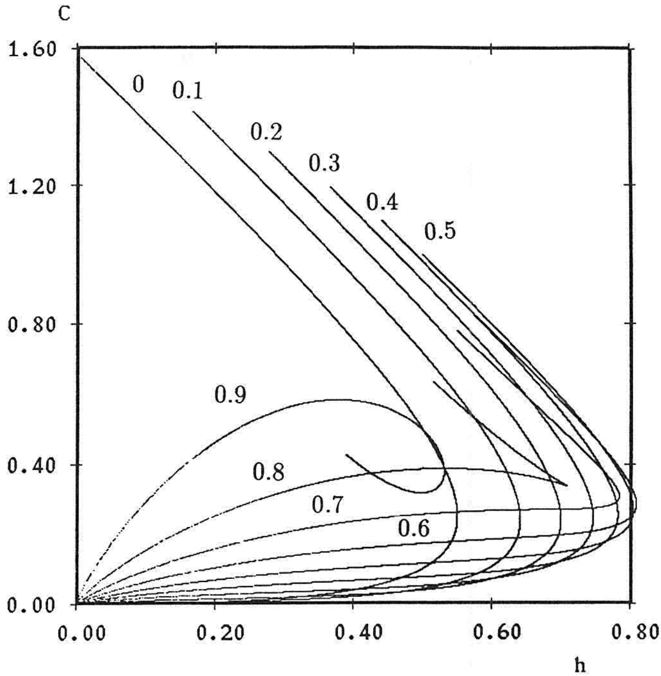

Figure 8: Similar to figure 7, but q is fixed at 0, 0.1, 0.2, ..., 0.9, and p is randomly chosen.

different complexity-entropy relationships.

To end this section, I will calculate the complexity-entropy relationship for another regular language, in which the limiting value of $\delta = \lim_{N\to\infty} \delta(N)$ is reached only when N is infinity. This regular-language grammar is shown in figure 10. By tuning the parameter p from 1 to 0, one can transform the generated sequence from the homogeneous all-zero sequence to the periodic sequence $\ldots 101010\ldots$.

We first determine the C_M and h_M for the corresponding Markov chain (with three symbols: 0, 0′, and 1). The Markov transition matrix is

$$\mathbf{T} = \begin{array}{c} \\ 0 \\ 0' \\ 1 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 0' & 1 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 \\ p & 0 & 1-p \\ 0 & 1 & 0 \end{pmatrix} \end{array}. \tag{25}$$

The densities of the three symbols are

$$P(0) = \frac{p}{2}, \quad P(0') = \frac{1}{2}, \quad P(1) = \frac{1-p}{2}. \tag{26}$$

We have

$$h_M = -\frac{1}{2}\log_2 p^p (1-p)^{1-p} \quad \text{and} \quad C_M = 1. \tag{27}$$

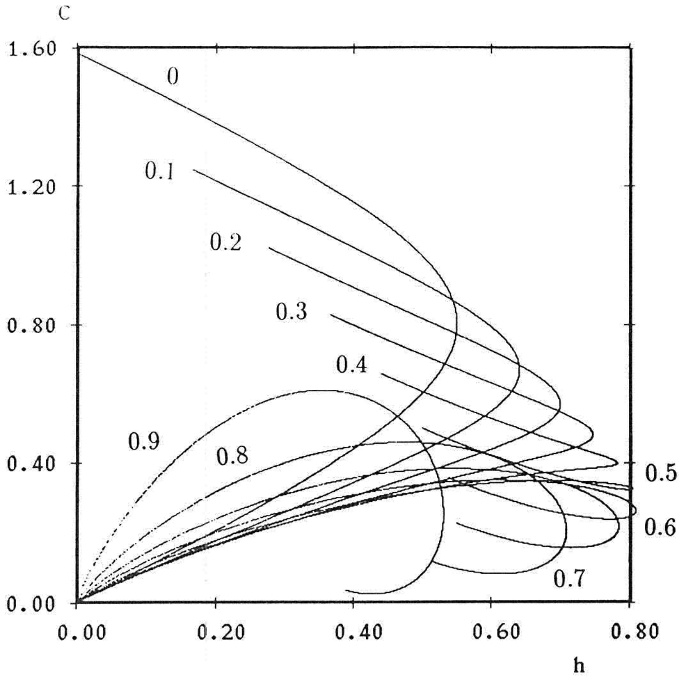

Figure 9: Similar to figure 7, but p is fixed at 0, 0.1, 0.2, ..., 0.9, and q is randomly chosen.

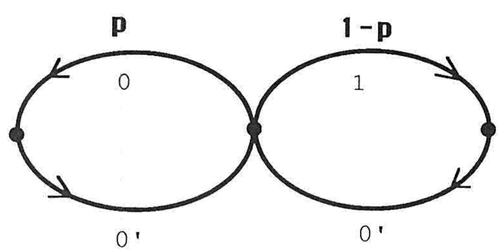

Figure 10: The second example of the regular language grammar discussed in section 4. It can produce either a homogeneous all-zero sequence (when p=1) or a periodic sequence ...1010... (when p=0).

The overestimation of the block entropy at $N = 2$ is

$$\begin{aligned}\delta(2) &= -P_M(00')\log_2 P_M(00') - P_M(0'0)\log_2 P_M(0'0) \\ &\quad +P(00)\log_2 P(00) \\ &= P(00') + P(0'0) = P(0) + P(0')p = \frac{p}{2} + \frac{p}{2} = p; \end{aligned} \tag{28}$$

at $N = 3$ it is

$$\delta(3) = \frac{p}{2}\log_2 \frac{(1+p)^{1+p}}{p^p}; \tag{29}$$

and for all N, it is

$$\begin{aligned}\delta(2n) &= p^n \\ \delta(2n+1) &= \frac{p^n}{2}\log_2 \frac{(1+p)^{1+p}}{p^p}. \end{aligned} \tag{30}$$

If $0 \le p < 1$, the complexity value defined in the infinite N limit is simply 1 because

$$C = \lim_{N\to\infty}(C_M - \delta(N)) = C_M = 1. \tag{31}$$

Nevertheless, if we first take the limit of $p \to 1$, we have

$$\lim_{N\to\infty}\lim_{p\to 1}(C_M - \delta(N)) = 0. \tag{32}$$

This result is consistent with our intuition because $p \to 1$ leads to the all-zero sequence.

It is interesting to note that it is possible to have a complexity value that is neither 1 nor 0 if the $p \to 1$ limit and the $N \to \infty$ limit are taken conspiratorially such that $N \sim 1/\log(p)$. This provides a perfect example for illustrating that the complexity as defined in the infinite block-length limit may not fit the intuition derived from finite sizes. And in many cases, a single quantity as a measure of complexity is not enough. We might need one measure at one length scale, and overall have a "spectrum of complexity measures."

5. Discussions

From the few examples presented in this paper, it can be seen that the complexity-entropy relationship is typically one-to-many or many-to-one instead of one-to-one. On the other hand, the dots in the complexity-versus-entropy plot usually do not fill the plane uniformly. This means that there are certain structures in the complexity-entropy relationship that depend on the sequence ensemble from which the complexity-entropy relationship is derived. The goal of studying the complexity-entropy relationship is to see how this structure changes as a function of the sequence ensemble.

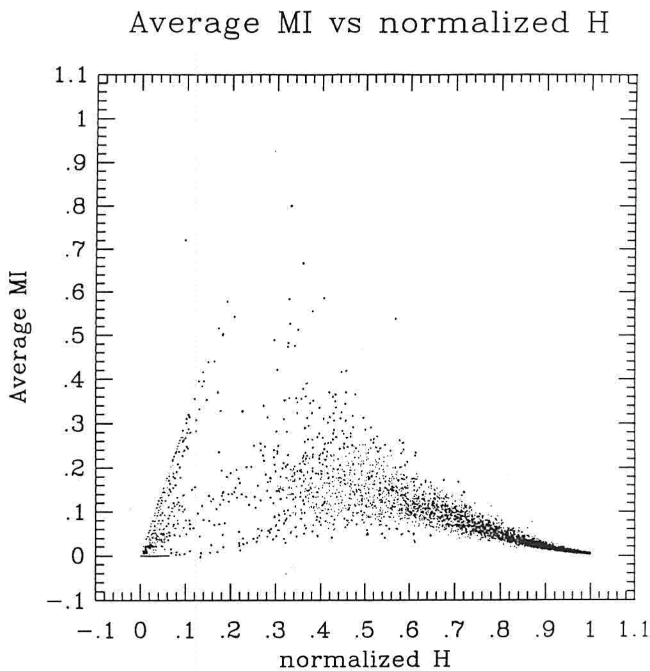

Figure 11: The nearest-neighbor mutual information versus single-site entropy $H(S^1)$ for configurations generated by two-dimensional eight-symbol cellular automata (reproduced with permission from reference [12, dissertation].)

In his study of cellular automata, Langton [12] plotted the nearest-neighbor mutual information (setting $M = N = 1$ instead of the infinity in equation (2)) as a function of the single-site entropy ($H(S^1)$). His plot is reproduced in figure 11. For periodic sequences, the joint-block entropy $H(S^{1+1})$ is the same as the single-site entropy $H(S^1)$, so the mutual information is equal to the single-site entropy. Thus it is not surprising that the left boundary of the region with dots is a straight line with slope 1.[2] If h is used as the x axis in place of $H(S^1)$, the left boundary will approach the y axis.

Crutchfield and Young [6] plotted something very similar, though in a completely different context. They studied the symbolic dynamics of the logistic map $x_{t+1} = \lambda x_t(1 - x_t)$ and generated binary symbolic sequences at hundreds of different λ values between 3 and 4. For each symbolic sequence, they constructed a regular-language grammar that can reproduce the se-

[2]The slope in figure 11 is 3 because Langton studies the eight-symbol cellular automata, and he normalizes the single-site entropy by the maximum value, which is 3, but does not normalize the mutual information. The slope becomes 1 if both entropy and mutual information are normalized.

quence statistically. The logarithm (or the "$p \log p$"-type logarithm) of the size of this regular-language grammar is the complexity measure they use. Since the symbolic sequences at certain λ values can have extremely long-range correlations that may not be characterizable by a regular language, a cutoff of the maximum size of the regular-language grammar is applied. This cutoff introduces a ceiling to the complexity value.

In Crutchfield and Young's complexity-versus-entropy plot there is again a left boundary of the region with dots, which is a straight line with slope 16. This value of slope is purely due to the choice of the x variable, which is $H(S^{16})/16$. This is because the logarithm of the grammar size is closely related to the past-future mutual information, which for periodic sequences is equal to the plateau value of the block entropy $H(S^N)$. If we use $x = H(S^N)/N$ and $y = C = H(S^N)$, it is clear that $y = Nx$.

In conclusion, when the past-future mutual information is used as a measure of complexity, its relation to the Shannon entropy for sequences of short-range correlations can be determined analytically. It is observed that the complexity-entropy relation depends on the specific structure of the short-range correlations (which is captured by the "grammar"). Though only three examples of the sequences with short-range correlations are studied in this paper, the method is obviously applicable to other cases. It will be interesting to see whether other regular-language grammars can produce complexity-entropy relations that are dramatically different from what is derived in this paper.

Acknowledgments

This study was inspired by the "AFTER-ALIFE" talk at Los Alamos given by Rob Shaw in March 1990. Besides him, I would like to thank Ditza Auerbach, Seth Lloyd, and Mats Nordahl for discussions. I also thank an anonymous referee whose comments prompted me to rewrite the draft, and Melanie Mitchell for proofreading the paper. My work at the Santa Fe Institute is funded by grants from the MacArthur Foundation, the National Science Foundation (PHY-87-14918), and the Department of Energy (DE-FG05-88ER25054).

References

[1] Ditza Auerbach, "Dynamical Complexity of Strange Sets," in *Measures of Complexity and Chaos*, edited by N. B. Abraham, A. M. Albano, A. Passmante, and P. E. Rapp (New York, Plenum Press, 1989); Ditza Auerbach and Itamar Procaccia, "Grammatical Complexity of Strange Sets," *Physical Review A*, **41**(12) (1990) 6602–6614.

[2] Charles H. Bennett, "On the Nature and Origin of Complexity in Discrete, Homogeneous, Locally-interacting Systems," *Foundations of Physics*, **16**(6) (1986) 585–592; Charles H. Bennett, "Information, Dissipation, and the Definition of Organization," in *Emerging Syntheses in Science*, edited by David

Pines (Reading, MA, Addison-Wesley, 1987); Charles H. Bennett, "Logical Depth and Physical Complexity," in *The Universal Turing Machine: a Half-Century Survey*, edited by Rolf Herken (Oxford, Oxford University Press, 1988); Charles H. Bennett, "How to Define Complexity in Physics, and Why," in *Complexity, Entropy and the Physics of Information*, edited by Wojciech H. Zurek (Reading, MA, Addison-Wesley, 1990).

[3] Gregory Chaitin, "On the Length of Programs for Computing Finite Binary Sequences," *Journal of the Association for Computing Machinery*, **13** (1966) 547–569; Gregory Chaitin, "Randomness and Mathematical Proof," *Scientific American*, **232** (1975) 47–52; Gregory Chaitin, *Algorithmic Information Theory* (New York, Cambridge University Press, 1987).

[4] Gregory Chaitin, "Algorithmic Information Theory," *IBM Journal of Research and Development*, **21** (1977) 350–359, 496.

[5] Gregory Chaitin, "Toward a Mathematical Definition of 'Life,'" in *The Maximum Entropy Formalism*, edited by Levine and Tribus (Cambridge, MIT Press, 1979); Gregory Chaitin, *Information Randomness and Incompleteness* (Singapore, World Scientific, 1987).

[6] James P. Crutchfield and Karl Young, "Inferring Statistical Complexity," *Physical Review Letters*, **63** (1989) 105–108; James P. Crutchfield and Karl Young, "Computation at the Onset of Chaos," in *Complexity, Entropy, and the Physics of Computation*, edited by Wojciech H. Zurek (Reading, MA, Addison-Wesley, 1990).

[7] Peter Grassberger, "Toward a Quantitative Theory of Self-generated Complexity," *International Journal of Theoretical Physics* **25**(9) (1986) 907–938; Peter Grassberger, "Problems in Quantifying Self-organized Complexity," *Helvetica Physica Acta*, **62** (1989) 498–511.

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

[9] T. Hogg and B. A. Huberman, "Order, Complexity, and Disorder," in *Mondes en Développement*, No. 54–55 (1986); B. A. Huberman and T. Hogg, "Complexity and Adaptation," *Physica D*, **22** (1986) 376–384.

[10] A. I. Khinchin, *Mathematical Foundations of Information Theory* (New York, Dover Publications, 1957).

[11] A. N. Kolmogorov, "Three Approaches to the Definition of the Concept 'Quantity of Information,'" *Problems of Information Transmission*, **1** (1965) 1–7.

[12] Chris Langton, "Computation at the Edge of Chaos: Phase Transition and Emergent Computation," *Physica D*, **42** (1990) 12–37; Chris Langton, "Computation at the Edge of Chaos" (Doctoral dissertation, University of Michigan, 1990).

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

[14] W. Li, "Mutual Information Functions versus Correlation Functions in Binary Sequences," in *Measures of Complexity and Chaos*, edited by N. B. Abraham, A. M. Albano, A. Passmante, and P. E. Rapp (New York, Plenum Press, 1989); W. Li, "Mutual Information Functions versus Correlation Functions," *Journal of Statistical Physics*, **60**(5/6) (1990) 823–837.

[15] Seth Lloyd and Heinz Pagels, "Complexity as Thermodynamic Depth," *Annals of Physics*, **188** (1988) 186–213.

[16] C. E. Shannon, "The Mathematical Theory of Communication," *The Bell System Technical Journal* **27** (1948) 379-423; C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (Urbana, University of Illinois Press, 1962).

[17] Rob Shaw, *The Dripping Faucet as a Model Chaotic System* (Santa Cruz, CA, Aerial Press, 1984).

[18] R. J. Solomonoff, "A Formal Theory of Inductive Inference," *Information and Control*, **7** (1964) 1–22, 224–254.

[19] Stephen Wolfram, "Computational Theory of Cellular Automata," *Communications in Mathematical Physics*, **96** (1984) 15–57.

[20] Domenico Zambella and Peter Grassberger, "Complexity of Forecasting in a Class of Simple Models," *Complex Systems*, **2** (1988) 269–303.