Synchronizing to Periodicity: The Transient Information and Synchronization Time of Periodic Sequences

David P. Feldman[1] and James P. Crutchfield[2]

[1] *College of the Atlantic, 105 Eden St., Bar Harbor, ME 04609 and Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501*
Electronic Address: dpf@santafe.edu

[2] *Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501*
Electronic Address: chaos@santafe.edu

(Dated: October 13, 2003)

We analyze how difficult it is to synchronize to a periodic sequence whose structure is known, when an observer is initially unaware of the sequence's phase. We examine the transient information $\mathbf{T}$, a recently introduced information-theoretic quantity that measures the uncertainty an observer experiences while synchronizing to a sequence. We also consider the synchronization time τ, which is the average number of measurements required to infer the phase of a periodic signal. We calculate $\mathbf{T}$ and τ for all periodic sequences up to and including period 23. We show which sequences of a given period have the maximum and minimum possible $\mathbf{T}$ and τ values, develop analytic expressions for the extreme values, and show that in these cases the transient information is the product of the total phase information and the synchronization time. Despite the latter result, our analyses demonstrate that the transient information and synchronization time capture different and complementary structural properties of individual periodic sequences — properties, moreover, that are distinct from source entropy rate and mutual information measures, such as the excess entropy.

Santa Fe Institute Working Paper 02-08-043 arxiv.org e-print nlin.AO/0208040

PACS numbers: 02.50.Ey, 02.50.Ga, 05.45.-a, 05.45.Tp, 89.75.Kd
Keywords: complexity; entropy; synchronization; transient information

I. INTRODUCTION

Imagine you are about to begin observing a sequence of events. You know the sequence is periodic; you even know the particular pattern that will repeat. However, you do not know what phase the sequence is in. How many observations, on average, would you have to make before you know with certainty the sequence's phase? And, as you obtain this certainty, how uncertain are you about the phase? Will the answers to these questions be the same for all sequences of a given period or are there differences between such sequences? What structural properties of a sequence determine the difficulty of synchronization? Here, we answer these questions.

A natural place to begin is with information theory, since it has long been used to analyze the statistical properties of sequences that arise in a variety of settings, including dynamical systems, time-series analysis, statistical mechanics, signal processing, and cryptography [1, 2]. One of the central quantities in these analyses is the *entropy rate* h_μ, the long-time measure of unpredictability of sequences produced by an information source. Although useful and important for quantifying randomness, h_μ does not capture a sequence's structural properties: its correlation, memory, or statistical complexity. Fortunately, today there are scores of measures of these latter properties. For example, on the information theoretic side, an oft-used measure of memory is the *excess entropy* $\mathbf{E}$ — the time-averaged, "all-point" mutual information between a sequence's past and future [3–7].

For any sequence of period P, it is well known that the entropy rate h_μ vanishes and the excess entropy $\mathbf{E} = \log_2 P$, since a periodic sequence is asymptotically predictable and since $\log_2 P$ bits of information are needed to store in which of the P possible phases a symbol in the sequence is. Thus, all periodic sequences of the same period have the same entropy rate and excess entropy, and so these quantities are unable to capture structural differences between distinct sequences of the same period. In particular, h_μ and $\mathbf{E}$ cannot help answer the questions posed above.

We turn instead to a recently introduced measure, the transient information $\mathbf{T}$ [5, 8], which captures information-theoretic differences between sequences of a given period. We will also make use of the synchronization time τ, defined as the average number of observations needed to synchronize to a periodic information source. We shall see that there are indeed significant differences in the synchronization properties of periodic sequences. Furthermore, while $\mathbf{T}$ and τ are related for extreme cases — minimum and maximum $\mathbf{T}$ and τ sequences — there are wide ranges of $\mathbf{T}$ and τ values for periodic sequences of the same period.

In the following section we review information sources, entropy rate, and excess entropy. In Sec. III we define $\mathbf{T}$ and τ more carefully. After a brief discussion of combinatorics and symmetry types of periodic sequences in Sec. IV, in Sec. V we discuss methods used to calculate $\mathbf{T}$ and τ. In Sec. VI we present the results of exhaustively calculating $\mathbf{T}$ and τ up to period 23. In Secs. VII

and VIII, we investigate these results, developing analytic expressions for the minimum and maximum $\mathbf{T}$ and τ values of a given period, exploring relationships and bounds between the two synchronization measures. Finally, in Sec. IX we summarize and interpret the results, pointing out several applications and open questions.

II. ENTROPY RATE AND EXCESS ENTROPY

A. Information Sources

We shall be concerned with a one-dimensional infinite sequence of variables:

$$\overleftrightarrow{S} \equiv \ldots S_{-2}S_{-1}S_0S_1\ldots \ . \tag{1}$$

Here, the S_t's are random variables that range over a finite set $\mathcal{A}$ of alphabet symbols. In general, $\mathcal{A} = \{0, 1, \ldots, k-1\}$, although in the following, we will restrict ourselves to binary sequences. We denote a block or *word* of L consecutive variables by $S^L \equiv S_1S_2\ldots S_L$. We follow the convention that a capital letter refers to a random variable, while a lowercase letter denotes a particular value of that variable. Thus, $s^L = s_1s_2\cdots s_L$, denotes a particular symbol block of length L.

We assume that the underlying *information source* is described by a shift-invariant measure μ on infinite sequences $\cdots s_{-2}s_{-1}s_0s_1s_2\cdots; s_t \in \mathcal{A}$ [9]. The measure μ induces a family of distributions, $\{\Pr(s_{t+1}, \ldots, s_{t+L}) : s_t \in \mathcal{A}\}$, where $\Pr(s_t)$ denotes the probability that at time t the random variable S_t takes on the particular value $s_t \in \mathcal{A}$ and $\Pr(s_{t+1}, \ldots, s_{t+L})$ denotes the joint probability over blocks of L consecutive symbols. We assume that the source is stationary; $\Pr(s_{t+1}, \ldots, s_{t+L}) = \Pr(s_1, \ldots, s_L)$.

We are interested in periodic sequences; a sequence is *periodic* of period P if $s_i = s_{i+P}$ for all i. The *prime period* of the sequence is the smallest such P. We can specify a periodic sequence by giving the smallest word — the *prime word* — that is exactly repeated. The prime word is aperiodic, necessarily. For example, $\ldots 1100110011001100\ldots$ is period 8 (and period 16 and so on), but has prime word 1100 and prime period 4.

B. Entropy Growth and Entropy Rate

Here, we give a brief review of the information-theoretic description of sequences, concentrating on those that are periodic. For more detail about information-theoretic measures of uncertainty and structure in the context of general one-dimensional random sequences, see, e.g., Refs. [5, 10, 11] and references therein.

The *total Shannon entropy* of length-L blocks is defined by:

$$H(L) \equiv -\sum_{s^L \in \mathcal{A}^L} \Pr(s^L)\log_2 \Pr(s^L) \ , \tag{2}$$

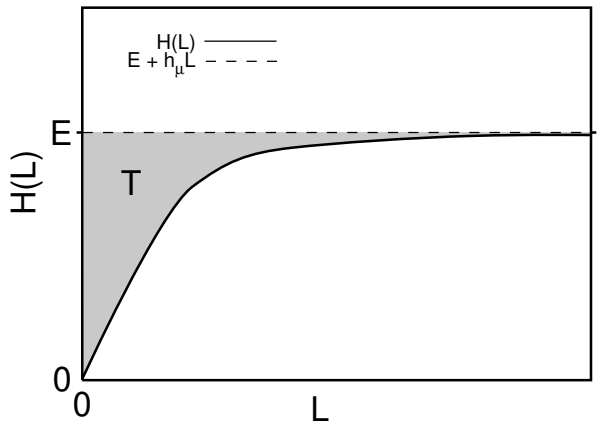

FIG. 1: Total Shannon entropy growth: A schematic plot of $H(L)$ versus L. The entropy $H(L)$ of L-blocks increases monotonically and approaches the asymptote $\mathbf{E} + h_\mu L$, where $\mathbf{E}$ is the excess entropy and h_μ is the source entropy rate. For periodic sequences, as shown here, $h_\mu = 0$. The shaded area is the transient information $\mathbf{T}$.

where $L > 0$. The sum runs over all possible blocks of L consecutive symbols. The units of $H(L)$ are *bits*. The entropy $H(L)$ measures the average uncertainty in identifying one sequence in the set of length-L sequences. Equivalently, $H(L)$ tells us how many yes-no questions, on average, are needed to determine the value of S^L. Another way to state this is that $H(L)$ sets a lower bound on the size of the code, measured in bits, needed to encode successive outcomes of the random variable S^L. For more on this coding interpretation of the Shannon entropy, see, e.g., Ref. [2].

The general behavior of $H(L)$ as a function of L for periodic information sources is shown schematically in Fig. 1. Note that we define $H(0) \equiv 0$.

The *entropy rate* h_μ is the rate of increase with respect to L of the total Shannon entropy in the large-L limit:

$$h_\mu \equiv \lim_{L\to\infty} \frac{H(L)}{L} \ , \tag{3}$$

where μ denotes the measure over infinite sequences that induces the L-block joint distribution $\Pr(S^L)$; the units are *bits per symbol*. One can also define a finite-L approximation to h_μ,

$$\begin{aligned} h_\mu(L) &= H(L) - H(L-1) \ , & (4)\\ &= H[S_L|S_{L-1}S_{L-2}\ldots S_0] \ , & (5) \end{aligned}$$

where $H[X|Y]$ is the entropy of the random variable X conditioned on the random variable Y: $H[X|Y] \equiv -\sum_{x,y} \Pr(x,y)\log_2 \Pr(\mathrm{x|y})$. One can then show [2] that

$$h_\mu = \lim_{L\to\infty} h_\mu(L) \ . \tag{6}$$

Note that Eq. (4) shows that $h_\mu(L)$ may be viewed as the slope of $H(L)$. For more on this point of view, see Ref. [5].

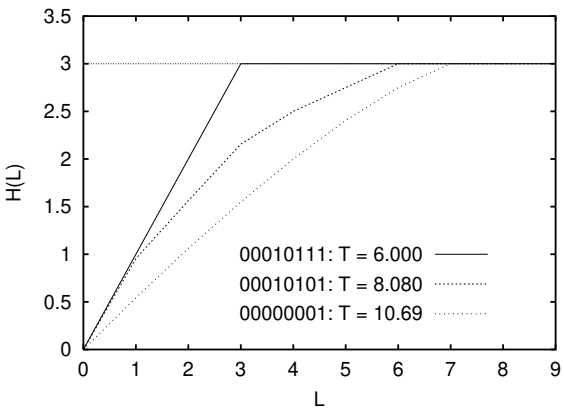

FIG. 2: Shannon entropy growth curves $H(L)$ for three different period-8 sequences: 00010111, 00010101, and 00000001.

Eq. (6) gives us another interpretation of h_μ. The entropy rate quantifies the irreducible randomness or unpredictability of the sequences; h_μ measures the randomness that persists even after the statistics of longer and longer blocks are taken into account. Since all periodic sequences are, ultimately, predictable, the entropy rate for all periodic sequences is $h_\mu = 0$. For a period-P sequence $h_\mu(P) = 0$ and $h_\mu(L) \geq 0$, for $L = 1, 2, \ldots, P-1$. This is illustrated in Fig. 2 for three different period-8 sequences; when the slope of $H(L)$ vanishes, $h_\mu(L) = 0$.

C. Excess Entropy: Apparent Memory

The entropy rate measures the randomness of sequences. A complementary quantity is the excess entropy $\mathbf{E}$, which measures the deviations of finite-L estimates of h_μ from its asymptotic value:

$$\mathbf{E} \equiv \sum_{L=1}^{\infty} [h_\mu(L) - h_\mu] \ , \tag{7}$$

The units of $\mathbf{E}$ are *bits*. Using this definition, one can show that the excess entropy is the subextensive part of $H(L)$ (see, e.g., [4, 5, 7, 11–13]); that is,

$$\mathbf{E} = \lim_{L\to\infty} [H(L) - h_\mu L] \ . \tag{8}$$

This establishes a geometric interpretation for $\mathbf{E}$, as shown in Fig. 1; it is the y-intercept of the line to which $H(L)$ asymptotes. That is, when $\mathbf{E}$ is finite, $H(L)$ asymptotes to $\mathbf{E} + h_\mu L$.

Another way to understand excess entropy is through its expression as a mutual information: the mutual information between the past and future semi-infinite halves of the chain of random variables:

$$\mathbf{E} = \lim_{L\to\infty} I[S_{-L+1}, S_{-L+2}, \cdots S_0; S_1 S_2 \cdots S_L] \ , \tag{9}$$

when the limit exists [5, 7]. Eq. (9) says that $\mathbf{E}$ measures the amount of historical information stored in the present that is communicated to the future.

Our focus here is on periodic sequences, for which it is easy to show from one or another of the above definitions that

$$\mathbf{E} = \log_2 P \ . \tag{10}$$

That is, the excess entropy is the total phase information stored in the periodic source. As noted above, the entropy of a random variable gives the average number of yes-no questions (i.e., bits), needed to determine the outcome of the variable. Since the observer starts out equally ignorant of the initial phase, it follows that $\mathbf{E}$ is a lower bound on the amount of information (in bits) an observer must extract in order to know in which phase the source is — that is, in order for it to be synchronized to the source.

The entropy growth curves $H(L)$ for three different period-8 sequences are shown in Fig. 2. Note that all curves asymptote to the same line, $\mathbf{E}+h_\mu L$. Since $h_\mu = 0$ (periodic sequences) and $\mathbf{E} = \log_2 P = \log_2 8 = 3$, the $H(L)$ curves asymptote to the horizontal line $H(L) = 3$ bits, when $L \geq 8$.

Despite the fact that these three period-8 sequences have the same h_μ and $\mathbf{E}$, Fig. 2 shows that their entropy growth curves $H(L)$ are not the same; for example, they reach the linear asymptote at different L's. Do all period-8 curves have different $H(L)$ behaviors? And do these different $H(L)$ behaviors matter, or do h_μ and $\mathbf{E}$ suffice to characterize periodic sequences?

III. MEASURES OF SYNCHRONIZATION

A. Transient Information

To address these questions, we consider the *transient information* $\mathbf{T}$, which was introduced by us in Ref. [5] and is defined by summing the deviations of $H(L)$ from its linear asymptote:

$$\mathbf{T} \equiv \sum_{L=0}^{\infty} [\mathbf{E} + h_\mu L - H(L)] \ . \tag{11}$$

Note that the units of $\mathbf{T}$ are *bits* $\times$ *symbols* and that $\mathbf{T} \geq \mathbf{E}$. The transient information is the shaded area in Fig. 1. We refer to $\mathbf{T}$ as *transient*, since, unlike h_μ and $\mathbf{E}$, it is dominated by nonasymptotic quantities: how $H(L)$ behaves before it reaches the linear scaling form $\mathbf{E}+h_\mu L$. By inspection, one sees that the three period-8 sequences in Fig. 2 all have different $\mathbf{T}$ values.

As noted above, for finite-$\mathbf{E}$ processes $H(L)$ scales as $\mathbf{E} + h_\mu L$ for large L [5]. When this scaling form is attained, we say that the observer is *synchronized* to the source. In other words, when

$$\mathbf{T}(L) \equiv \mathbf{E} + h_\mu L - H(L) = 0 \ , \tag{12}$$

we say the observer is synchronized at length-L sequences. The quantity $\mathbf{T}(L)$ provides a measure of the departure from synchronization. Note that $\mathbf{T}(L)$ is nonnegative. Looking at Eq. (11), we see that the transient information $\mathbf{T}$ may be written as a sum of the $\mathbf{T}(L)$'s:

$$\mathbf{T} = \sum_{L=0}^{\infty} \mathbf{T}(L) \ . \tag{13}$$

To give the transient information a more precise interpretation, let us consider in more detail the synchronization scenario we have in mind. An observer begins making measurements of a sequence, seeing one symbol at a time. The observer knows the periodic sequence it is about to start seeing, but it doesn't know the phase. That is, the observer knows the period and the prime word that will be repeated. The task for the observer is to make measurements and determine the sequence's phase. Exact prediction is possible from this point onwards, though *not* before. How uncertain is the observer during this synchronization process?

To answer this question, we must introduce some additional notation. Before the observer is synchronized, its knowledge is characterized by a distribution over the P possible phases of the periodic sequence. Equivalently, the sequence's phases may be viewed as the periodic source's internal states. Let φ denote a particular phase, and let Φ denote the set of all phases. Clearly, $|\Phi| = P$. Next, let $\Pr(\varphi|s^L)$ denote the probability, as inferred by the observer, that the sequence is in phase φ, given that it has just seen the particular sequence of symbols s^L. The entropy of the distribution of φ measures the observer's average uncertainty of the phase Φ. Averaging this uncertainty over the possible length-L observations, we obtain the *average state-uncertainty*:

$$\mathcal{H}(L) \equiv -\sum_{s^L} \Pr(s^L) \sum_{\varphi\in\Phi} \Pr(\varphi|s^L) \log_2 \Pr(\varphi|s^L) \ . \tag{14}$$

The quantity $\mathcal{H}(L)$ can be used as a criterion for synchronization. The observer is synchronized to the sequence when $\mathcal{H}(L) = 0$ — that is, when it is completely certain about the phase, or internal state, $\varphi \in \Phi$ of the source generating the sequence. And so, when the condition in Eq. (12) is met, we see that $\mathcal{H}(L) = 0$, and the uncertainty associated with the next observation is 0.

However, while the observer is still unsynchronized $\mathcal{H}(L) > 0$. The average total uncertainty experienced by the observer during the synchronization process is the total *synchronization information* $\mathbf{S}$:

$$\mathbf{S} \equiv \sum_{L=1}^{P} \mathcal{H}(L) \ . \tag{15}$$

For the periodic sequences under consideration here, it turns out that [5]

$$\mathbf{S} = \mathbf{T} \ . \tag{16}$$

That is, the transient information $\mathbf{T}$ is equal to the total synchronization information $\mathbf{S}$.

Eq. (16) is a special case of a theorem recently proved by us in Ref. [5] and discussed further in Ref. [8]. Here, we briefly sketch the main argument behind Eq. (16). At $L = 0$, no measurements have been made and the observer posits that each phase is equally likely. The state-uncertainty is thus $\log_2 P = \mathbf{E}$. After L observations have been made, the observer has gained, on average, $H(L)$ bits of information about the phase of the process. As a result, the average state-uncertainty $\mathcal{H}(L)$ is now $\mathbf{E} - H(L)$. Plugging this last observation into Eq. (15), Eq. (16) follows from the definition of $\mathbf{T}$, Eq. (11).

As a result of Eq. (16), the transient information $\mathbf{T}$ provides a direct measure of how difficult or confusing it is to synchronize to a periodic sequence. If a periodic sequence has a large $\mathbf{T}$, then on average an observer will be highly uncertain about its phase during the synchronization process. The transient information measures a structural property of a sequence — a property captured neither by the entropy rate nor by the excess entropy. If the observer is in a position where it must take immediate action, it does not have the option of waiting for full synchronization. In this circumstance, the synchronization information $\mathbf{S}$ provides an information-theoretic average-case measure of the error incurred by the observer during the synchronization process.

B. Synchronization Time

In addition to measuring the total uncertainty experienced during synchronization, we can ask a related question: On average, *how many* measurements must be made before the observer is certain in which state the source is? We call the number τ of measurements, averaged over the P possible starting phases, the *synchronization time*.

For example, consider the periodic sequence $(0001)^\infty$. There are four possible phases in which one might begin to observe this sequence. If the first symbols parsed are $1-0-0-0$, then the observer is synchronized after the first observation (a 1). If the observer initially sees $0-1-0-0$, it is synchronized after two symbols. If the observer sees either $0-0-1-0$ or $0-0-0-1$, it is synchronized after three measurements. Each of these initial measurement sequences is equally likely. Thus, on average, it will take

$$\tau = \frac{1+2+3+3}{4} = 2.25 \tag{17}$$

measurements before the observer is synchronized to $(0001)^\infty$.

Below, we conduct a survey of the transient information $\mathbf{T}$ and synchronization time τ of periodic sequences. We shall see that $\mathbf{T}$ and τ capture fundamentally different properties of a sequence. We shall also see that within

different sequences of a given period there are wide variations of $\mathbf{T}$ and τ. First, however, we pause to consider some combinatorial properties and symmetry classes of periodic sequences.

IV. ENTROPIC SYMMETRY TYPES AND COMBINATORICS OF PERIODIC SEQUENCES

How many periodic sequences of a given prime period are there? Aperiodic length-P words that are related by an application of a cyclic permutation group C_P give the same periodic sequence. Thus, for our purposes — considering infinitely repeated periodic sequences — two length-P words related via the group operation C_P are equivalent; they yield identical infinite sequences. For example, the aperiodic words 001, 010, and 100 all yield the same, infinite, period-3 sequence.

However, we are interested in looking for distinct $H(L)$ versus L behaviors, and this induces another set of equivalences on the period-P sequences. Because $H[X]$ depends only on the distribution of X, and not on the values X assumes, the entropy of a random variable is unchanged under a one-to-one mapping of alphabet symbols [2]. For the binary cases we are interested in here, this means that the mapping $0 \leftrightarrow 1$ leaves $H(L)$ unchanged; for our purposes, the sequence $(011)^\infty$ is identical to $(100)^\infty$. In terms of group action, swapping 1 and 0 is the action of the symmetric group S_1. So, any two sequences that are based on binary aperiodic length-P words related to each other by some combination of S_1 and C_P will have identical $H(L)$ versus L behavior.

The number L_2 of aperiodic binary words inequivalent under $S_1 \times C_P$ is given by:

$$L_2(P) = \frac{1}{P} \sum_{d|P, d \text{ odd}} \mu(d) 2^{P/d} \,. \tag{18}$$

The sum runs over all odd positive divisors d of n, and $\mu(m)$ is the Möbius function: $\mu(m) = 0$, if m is the product of nondistinct primes; $+1$, if m is the product of an even number of distinct primes; and -1, otherwise. This result was first obtained by Fine [14] in 1958 and simplified a few years later by Gilbert and Riordan [15]. These words are also related to a version of the famous necklace problem from combinatorics: the words inequivalent under $S_1 \times C_P$ are exactly those 2-color necklace sequences with prime period P where interchanging bead color is allowed, but turning the necklace over is not.

These aperiodic words are also related to the so-called Lyndon words. A *Lyndon word* is aperiodic and the lexicographically least among its rotations. An *unlabeled Lyndon word* is aperiodic and lexicographically least among its rotations and relabelings. Eq. (18), then, also counts the number of binary unlabeled Lyndon words. For more discussion, see Ref. [16] and sequence #A000048 of Ref. [17]. Titsworth [18] has pointed out that Eq. (18) also gives the number of distinct finite-state machines needed to produce all periodic binary sequences of prime period P.

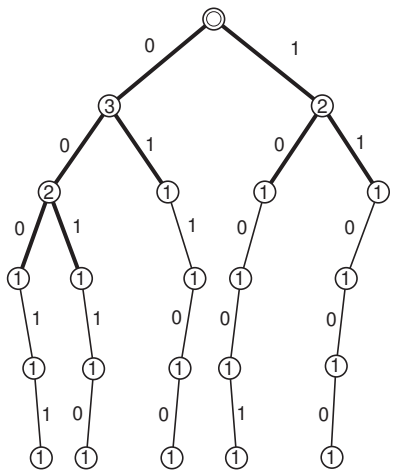

FIG. 3: Parse tree for the period-5 sequence with prime word 00011. Unvisited paths are not shown. The synchronizing words are shown with bold paths from the start node.

Prime Word	$\mathbf{T}$	τ
000001	6.97905	3.33333
000011	5.37744	2.83333
000101	5.58496	3.33333
000111	4.83659	2.66667
001011	4.83659	2.66667

TABLE I: The transient information $\mathbf{T}$ and synchronization time τ of all five distinct period-6 sequences.

V. METHODS

We calculated the transient information $\mathbf{T}$ and synchronization time τ for all distinct periodic sequences (unlabeled Lyndon words) up to and including period $P = 23$. To enumerate all of these words of a given period, we used the efficient algorithms available at Ref. [16].

To calculate $\mathbf{T}$ and τ for a given length-P word, we begin by determining the frequency of occurrence of all its subwords. This is done by parsing the word and its $P-1$ cyclic permutations into a tree, whose paths from the root to any node are the subwords and each of whose nodes contain the number of subwords found that lead to it from the root.

An example is shown in Fig. 3. Consider the sequence whose prime word is the length-5 word 00011. We start at the tree root, indicated by the double circle at the top, and follow the leaves labeled with the appropriate symbol. Each time we cross a node we increment the count there by one. We repeat this procedure for all 5

cyclic permutations of 00011.

Once the parse tree is built, $\mathbf{T}$ can be easily and exactly calculated. First, the node counts are turned into probabilities by normalizing — dividing each count by the period P. Reading across the tree at level ℓ gives the probability of subwords of length ℓ. In the $P = 5$ example considered here, $\Pr(00) = 2/5$, $\Pr(01) = 1/5$, $\Pr(10) = 1/5$, and $\Pr(11) = 1/5$. From these probabilities, the block entropies $H(L)$, and, in turn, $\mathbf{T}$, follow directly by using Eq. (11) and noting that $\mathbf{E} = \log_2 P$ and $h_\mu = 0$.

Calculating τ is only a bit more involved. After building the parse tree, we reparse each cyclic permutation of the word. As we reparse and proceed down the tree, we monitor the node counts. For each cyclic permutation of the prime word, we follow the corresponding path from the root. When we come to the first node that has a count of 1, we are synchronized. That is, for the ith cyclic permutation of the word, the level ℓ at which we first encounter a node count of 1 is the synchronization time τ_i for that permutation. The average synchronization time τ is then simply:

$$\tau = \frac{1}{P} \sum_{i=0}^{P-1} \tau_i \ . \tag{19}$$

For the example in Fig. 3, we have

$$\tau = \frac{1}{5}(3 + 3 + 2 + 2 + 2) = 2.4 \ . \tag{20}$$

As for $\mathbf{T}$, the above method yields an exact value for τ.

For example, in Table I we show the results of calculating the transient information $\mathbf{T}$ and the synchronization time τ for all five distinct period-6 sequences. (The reader may find it helpful to verify these results.)

VI. EMPIRICAL RESULTS

We empirically investigated the behavior of $\mathbf{T}$ and τ by first exhaustively enumerating all distinct periodic sequences up to and including period $P = 23$ and then exactly calculating these two quantities for each sequence using the methods of the previous section. The results are summarized in Figs. 4, 5, 6, 7, and 8. In Table II we show, for periods 2 through 23, the number of distinct periodic sequences, the number of distinct $\mathbf{T}$ and τ values, and the mean, minimum, and maximum $\mathbf{T}$ and τ values.

Based on these results, one can make a host of observations. First, note that $\mathbf{T}$ and τ are *not* the same for all sequences of a given period. This can be seen in our period-6 results of Table I, as well as in Figs. 4, 5, and 6. Thus, within a given period, there are many different synchronization behaviors. As noted before, these differences are not accounted for by the entropy rate h_μ and the excess entropy $\mathbf{E}$, since $\mathbf{E}$ and h_μ are the same for all sequences with the same period.

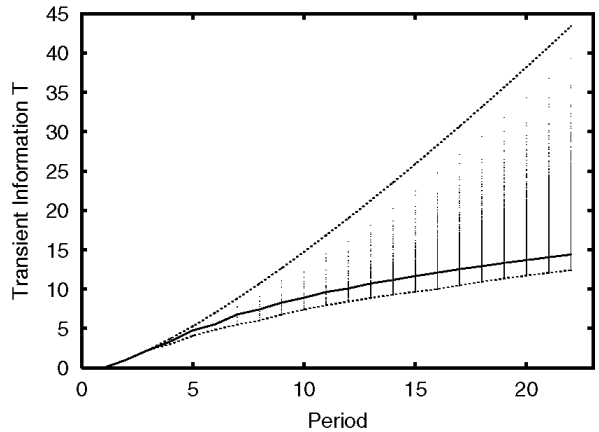

FIG. 4: Scatter plot of the transient information $\mathbf{T}$ for all distinct periodic sequences as a function of their prime period P, up to $P = 22$. The dashed lines are the maximum and minimum $\mathbf{T}$ values computed analytically; see text. The solid line gives the estimated average $\mathbf{T}$ value at the given period.

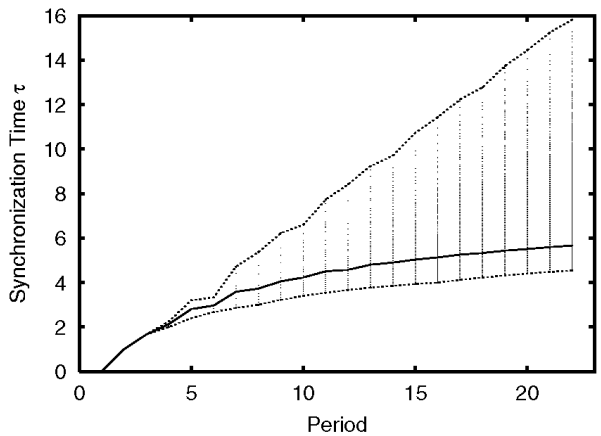

FIG. 5: Scatter plot of the synchronization time τ for all distinct periodic sequences as a function of their period P, up to $P = 22$. The dashed lines are the maximum and minimum τ values computed analytically; see text. The solid line gives the estimated average τ value at the given period.

Second, there are words with the same period that have identical $\mathbf{T}$ or identical τ values. This can be seen in Fig. 6, in which we have plotted the count data shown in Table II. The number of distinct $\mathbf{T}$ and τ values clearly grows less quickly than the total number of distinct periodic sequences. Note also that the number of distinct $\mathbf{T}$ and τ values grow at different rates. This provides direct evidence that the transient information and the synchronization time measure different qualities. For example, in Table I, note that two sequences have the same τ but different $\mathbf{T}$'s. The converse also occurs, but much less frequently. There are sequences with the same $\mathbf{T}$ but different τ values, but this does not occur until $P = 12$. As can be seen from both Table II and Fig. 6, τ tends

Period P	Distinct Periodic Sequences	Distinct $\mathbf{T}$ Values	$\langle \mathbf{T} \rangle$	$\mathbf{T}_{\min}$	$\mathbf{T}_{\max}$	Distinct τ Values	$\langle \tau \rangle$	$\tau_{\min}$	$\tau_{\max}$
3	1	1	2.25163	2.25163	2.25163	1	1.66667	1.66667	1.66667
4	2	2	3.34436	3.00000	3.68872	2	2.12500	2.00000	2.25000
5	3	3	4.73957	4.07291	5.27291	3	2.80000	2.40000	3.20000
6	5	4	5.52293	4.83659	6.97905	3	2.96667	2.66667	3.33333
7	9	8	6.77183	5.48662	8.78940	8	3.58730	2.85714	4.71429
8	16	13	7.42624	6.00000	10.6907	10	3.72656	3.00000	5.37500
9	28	21	8.27733	6.76598	12.6728	12	4.05556	3.22222	6.22222
10	51	35	8.89152	7.39483	14.7274	22	4.22549	3.40000	6.60000
11	93	53	9.63275	7.94889	16.8480	29	4.50733	3.54545	7.72727
12	170	90	10.0802	8.42155	19.0290	28	4.57157	3.66667	8.41667
13	315	145	10.7162	8.88705	21.2656	49	4.80293	3.76923	9.23077
14	585	261	11.1637	9.29398	23.5540	60	4.89634	3.85714	9.71429
15	1091	484	11.6530	9.66732	25.8906	64	5.03367	3.93333	10.7333
16	2048	610	12.0846	10.0000	28.2725	78	5.13293	4.00000	11.4375
17	3855	1091	12.5298	10.4930	30.6969	104	5.25312	4.11765	12.2353
18	7280	1878	12.9133	10.9359	33.1614	104	5.33134	4.22222	12.7778
19	13797	3205	13.3158	11.3449	35.6640	132	5.43299	4.31579	13.7368
20	26214	5015	13.6772	11.7168	38.2027	143	5.50798	4.40000	14.4500
21	49929	10355	14.0405	12.0774	40.7759	153	5.59134	4.47619	15.2381
22	95325	16031	14.3815	12.4083	43.3818	191	5.66223	4.54545	15.8182
23	182361	27322	14.7179	12.7191	46.0192	207	5.73570	4.60870	16.7391

TABLE II: The number of distinct periodic sequences, the number of distinct $\mathbf{T}$ and τ values, and the average, minimum, and maximum $\mathbf{T}$ and τ values as a function of period, up to $P = 23$.

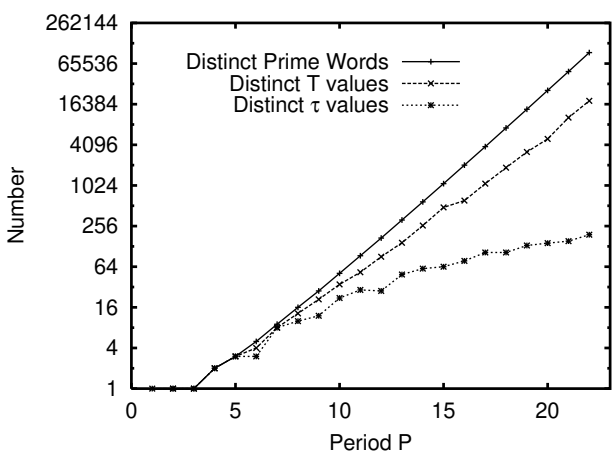

FIG. 6: The number of distinct periodic sequences, distinct $\mathbf{T}$ values, and distinct τ values as a function of period P. Note the logarithmic scale on the vertical axis.

to be substantially more degenerate than $\mathbf{T}$; τ is somehow a coarser measure of a sequence's synchronization properties.

Third, note that for a given period there is a considerable range of transient information and synchronization time values. This can be seen in Figs. 4 and 5, where the minimum and maximum $\mathbf{T}$ and τ values are shown as dashed lines. This suggests that there are significant differences in synchronization behaviors among different sequences with the same period. We will return to the question of minimum and maximum $\mathbf{T}$ and τ sequences in some detail in Sec. VII.

Another way to view the range of $\mathbf{T}$ values is found in Fig. 7, in which we plot the distribution of $\mathbf{T}$ values for period-16 and period-23 sequences. Note the asymmetry; most $\mathbf{T}$ values are closer to the minimum $\mathbf{T}$ than the maximum $\mathbf{T}$. Similar observations can be made about the distribution of τ values shown in Fig. 8.

Finally, note from Fig. 6 and Table II that the number of distinct τ values is not strictly monotonic with increasing periodicity. Curiously, the number of distinct τ values decreases as the period goes from 11 to 12 and it remains constant in going from $P = 5$ to $P = 6$, while it increases for all other P increases.

VII. MINIMUM, MAXIMUM, AND AVERAGE T AND τ

Here we develop analytical expressions for the extreme behaviors of the transient information and the synchronization time, and we determine how these extreme val-

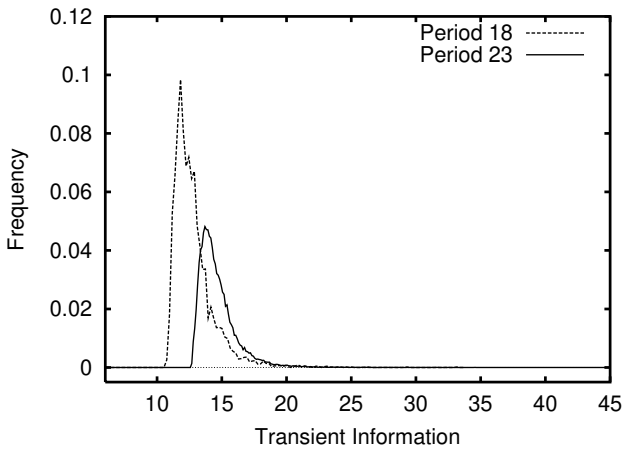

FIG. 7: Normalized histogram of the distribution of **T** values for period-18 and period-23 sequences.

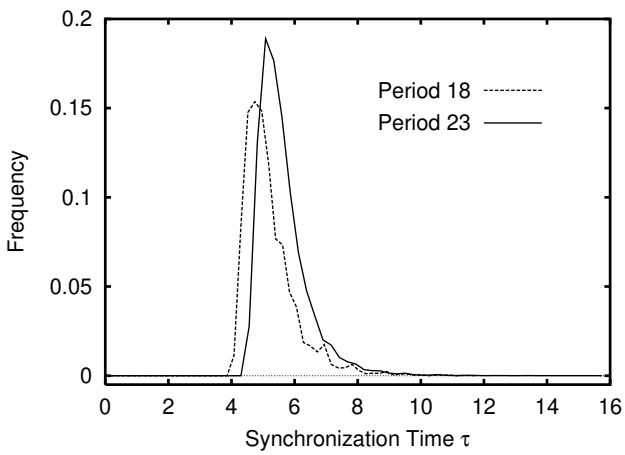

FIG. 8: Normalized histogram of the distribution of τ values for period-18 and period-23 sequences.

ues grow asymptotically with period. This analysis allows us to derive a number of important conclusions about which complementary structural features these quantities capture.

First, some notational preliminaries. For a given period, we denote by $\langle \mathbf{T} \rangle$ the average transient information, averaged over all distinct prime words. Similarly, $\mathbf{T}_{\min}$ and $\mathbf{T}_{\max}$ denote the minimum and maximum **T** values at a given period, respectively. We use similar notation for the average, maximum, and minimum τ values.

A. Maximum-T Sequence

A sequence's transient information **T** is large if $H(L)$ approaches its asymptotic value of $\mathbf{E} = \log_2 P$ slowly. This can be deduced graphically from Figs. 1 and 2. The $H(L)$ curve that grows most slowly is that in which the distribution over subwords is maximally nonuniform, since the entropy of a random variable decreases as its distribution departs from uniformity. For a given period P, the sequence with the maximum value of **T** is one consisting of $P-1$ 0's followed by an isolated 1, since this sequence is the one whose distribution over subwords is as nonuniform as possible. This behavior is seen in one of the period-8 sequences shown in Fig. 2.

A direct calculation for the maximal-**T** sequence yields:

$$\mathbf{T}_{\max} = \frac{1}{P} \sum_{n=2}^{P} n \log_2 n \; . \tag{21}$$

To develop an expression for $\mathbf{T}_{\max}$ as a function of P we approximate the sum by an integral. After some work, one obtains,

$$\mathbf{T}_{\max} \approx \frac{1}{\ln(16)P} \big[2 - P - P^2 + 4\ln(2) + P^2 \ln(P) + (1+P)^2 \ln(1+P) \big] \; . \tag{22}$$

For large P, this scales as

$$\mathbf{T}_{\max} \approx \frac{1}{2} P \log_2 P \; . \tag{23}$$

The values obtained from Eq. (21) agree with the exact numerical results shown in Table II. In Fig. 9 we plot the approximation of Eq. (22) and compare them with our exact numerical results. As expected, the approximation is quite good and improves at larger P.

The synchronization time τ for the maximum-**T** word can be calculated directly by noticing that one is synchronized as soon as one sees a 1, or after $P-1$ measurements, whichever comes first. One obtains, for large P:

$$\tau_{\max \mathrm{T}} \approx \frac{P-1}{2} \; , \tag{24}$$

B. Minimum-T Sequences

The minimum-**T** word is particularly easy to analyze when the period is a power of 2. We begin with this special case and use it to derive a simple form for $\mathbf{T}_{\min}$ as a function of the period.

The sequences with the minimum transient information $\mathbf{T}_{\min}$ correspond to those whose $H(L)$ curves rise most quickly to **E**. This is achieved by a sequence whose distribution over subwords, at each subword length, is most nearly uniform and so gives high entropy: the measurements are most informative about the phase.

When the period $P = 2^M$, where M is an integer, the distribution over subwords for the $\mathbf{T}_{\min}$ sequence is exactly uniform. Words with this property are known as *de Bruijn* sequences. A binary de Bruijn sequence of order k is a binary circular string containing every binary substring of length k exactly once. For example, the lexicographically smallest order-3 binary de Bruijn sequence

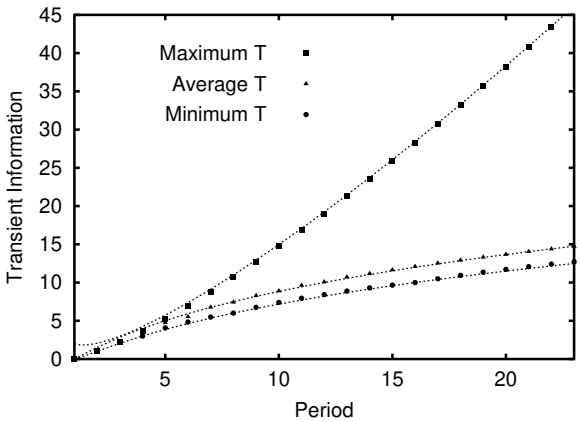

FIG. 9: The minimum, maximum, and average transient information values plotted as a function of period. The solid points are our exact results; the dashed lines are the approximations developed in the text: Eq. (22) for $\mathbf{T}_{\max}$, Eq. (26) for $\mathbf{T}_{\min}$, and Eq. (34) for $\langle \mathbf{T} \rangle$.

is 00010111. Note that this was one of the period-8 sequences whose $H(L)$ curve was plotted in Fig. 2. In that figure, the rapid convergence of $H(L)$ to $\log_2 8 = 3$ is clearly seen.

Because every subword appears with equal probability, $H(L) = L$, until $L = \log_2 P$. That is, $H(L)$ converges linearly to the excess entropy $\mathbf{E} = \log_2 P$. In other words, the source appears to be a fair coin until synchrony is achieved. At that time, exact predictability becomes possible. This linear convergence can be seen in Fig. 2. The transient information is thus given by the area between $H(L)$ and $\log_2 P$;

$$\mathbf{T}_{\min} \;=\; \sum_{n=1}^{\log_2 P} n \;. \tag{25}$$

Evaluating the sum, one has:

$$\mathbf{T}_{\min} \;\approx\; \frac{1}{2}\left(\log_2^2 P + \log_2 P\right) \;. \tag{26}$$

Equivalently, since $\mathbf{E} = \log_2 P$, we may write this as:

$$\mathbf{T}_{\min} \;\approx\; \frac{1}{2}\,\mathbf{E}\,(\mathbf{E}+1) \;. \tag{27}$$

This result is exact for all P that are a power of 2 and serves as an excellent lower bound for P's that are not.

For periods that are not a power of 2, the minimum-$\mathbf{T}$ words are those for which the distribution over subwords is as uniform as possible. For example, for $P = 10$ there are two prime words with minimum $\mathbf{T}$: 0000101111 and 0001011101. Each has a $\mathbf{T}$ of 7.39483. Note that for each word, the distribution over subwords of length 1 is uniform, while the distribution over words of length 2 is as uniform as can be, given that there are 10 occurrences of subwords of length 2; for each, two length-2 subwords occur with a frequency of 3/10 and two occur with a frequency of 2/10.

We have obtained an analytic expression for $\mathbf{T}_{\min}$ for sequences of general period P. The main idea, as stated above, is to distribute the frequencies of the subwords as uniformly as possible. How uniform this distribution is depends on how 2^L (the number of subwords of length-L) divides into the period P. Following this line of reasoning, we obtain the following results. Let

$$n_p(L) \;\equiv\; \left\lfloor \frac{P}{2^L} \right\rfloor \tag{28}$$

and

$$N_{\text{odd}}(L) \;\equiv\; P \bmod 2^L \;, \tag{29}$$

$$N_{\text{even}}(L) \;\equiv\; 2^L - N_{\text{odd}}(L) \;. \tag{30}$$

Here, $\lfloor x \rfloor$ denotes the *floor* of x — the largest integer smaller than x. Similarly, $\lceil x \rceil$ is the *ceiling* of x — the smallest integer larger than x.

The $H(L)$ curve converges to $\mathbf{E} = \log_2 P$ by length L^*, where

$$L^* \;\equiv\; \left\lceil \frac{\log_2 P}{2} \right\rceil \;+\; 1 \;. \tag{31}$$

Then, we find that the transient information is given by:

$$\mathbf{T}_{\min} \;=\; (L^* + 1)\log_2 P - \phi \;, \tag{32}$$

where

$$\phi \;=\; \sum_{n=1}^{L^*} \left[-N_{\text{odd}} \frac{n_p(L)+1}{n} \log_2 \left(\frac{n_p(L)+1}{n} \right) - N_{\text{even}}(L) \frac{n_p(L)}{n} \log_2 \left(\frac{n_p(L)}{n} \right) \right] \;. \tag{33}$$

Values obtained using Eq. (32) agree with the exact results of Table II.

C. Average T

In Fig. 9 we have also plotted $\langle \mathbf{T} \rangle$, the average value of the transient information, as a function of the period P. These average values were also given in Table II. The average values appear to be well approximated by

$$\langle \mathbf{T} \rangle \;=\; \frac{1}{2}\log_2^2 P + \log_2 P \;. \tag{34}$$

This is an empirical fit. Note that $\langle \mathbf{T} \rangle$ and $\mathbf{T}_{\min}$ of Eq. (26) are the same to leading order in P. This is not surprising, given the asymmetry in the distribution of $\mathbf{T}$ values evident in Fig. 7.

D. Minimum-τ Sequences

For a given period P, we found that the sequences with the minimum transient information $\mathbf{T}_{\min}$ are the same as the sequences with the minimum synchronization time $\tau_{\min}$. As noted above, these sequences are those for which the distribution of subwords is most nearly uniform at each subword length. Using this observation, it is possible to derive an analytic expression for $\tau_{\min}$ among the sequences of period P.

As discussed above when considering the minimum-$\mathbf{T}$ word, if $P = 2^M$, where M is an integer, then the minimum-τ word is such that all subwords appear with equiprobability. In particular, each subword of length $\log_2 P$ occurs exactly once. Thus, regardless of the phase in which one begins observing such a sequence, there is no ambiguity about the phase of the sequence after making $\log_2 P$ observations. For observations shorter than $\log_2 P$, there will always be some uncertainty about the phase of the process. Thus, the synchronization time is:

$$\tau_{\min} = \log_2 P = \mathbf{E} \,. \tag{35}$$

This result is exact when $P = 2^M$ and is an excellent approximation for other periods.

As mentioned above, $\mathbf{E}$ is the total phase information stored in the periodic source; it is simply the Shannon entropy of the P possible phases. Recall that the entropy of a random variable sets a lower bound on the number of yes-no questions needed, on average, determine the value of that variable. The synchronization time τ is equivalent to this average number of yes-no questions — measuring a binary symbol entails asking a yes-no question of the source. Eq. (35) thus shows that the periodic word with the minimum τ saturates the lower bound set by the entropy.

As was the case for $\mathbf{T}_{\min}$, one can derive an expression for $\tau_{\min}$ for periods that are not a power of 2. As with the minimum-$\mathbf{T}$ word, in the minimum-τ word the distribution over the subwords is as uniform as possible. In the case where $P = 2^M + 1$, $P - 2$ phases will synchronize in M observations. The remaining two phases, though, require an additional observation to synchronize; an additional symbol must be seen to distinguish between these two phases. This was the case, for example, for the period-5 case considered in Fig. 3. Here, $P = 5 = 2^2 + 1$ and $M = 2$. Three of the phases synchronize after two observations, while two phases synchronize after three observations.

Generalizing this observation to $P = 2^M + N$, we obtain the following result:

$$\tau_{\min} = \frac{1}{P}\left[\, 2N(M+1) + (P-2N)M \,\right] \,, \tag{36}$$

where

$$M = \left\lfloor \log_2 P \right\rfloor \,, \tag{37}$$

and

$$N = P \bmod 2^M \,. \tag{38}$$

Eq. (36) can also be written as [17, seq. A061717]:

$$\tau_{\min} = \frac{1}{P}\left\lceil \log_2 P^P \right\rceil \,. \tag{39}$$

The minimum-τ values given by Eq. (36) agree with the exact numerical results shown in Table II. For large P, Eq. (39) scales as $\log_2 P$, in agreement with Eq. (35).

E. Maximum-τ Sequence

Recall that the maximum-$\mathbf{T}$ sequence consisted of P–1 0's followed by one 1. For example, the $\mathbf{T}_{\max}$ sequence is 000000001 for $P = 9$. However, this is *not* the maximum-τ sequence. The sequence 000000001 has an $H(L)$ that grows very slowly, leading to its large $\mathbf{T}$. However, once one sees a 1, one is synchronized. Thus, it is possible for an observer to get lucky, see a 1 after one or two observations and be synchronized quickly. As a result, the expected synchronization time for this sequence is relatively low, only $\tau = 4.88889$.

In contrast, the sequence with the largest synchronization time for period 9 is 000010001 and has $\tau = 6.2222$. This sequence has the maximum τ due the presence of an additional 1 which prevents an observer from synchronizing after a single observation; this leads to a larger τ.

The maximum-τ word for periods 4-16 are given in Table III. Note that a maximum-τ word need not be unique, although we find that it is unique for all periods examined, except for $P = 6$.

For P odd, we noticed the maximum-τ word takes on a particularly simple form: two 1's separated by $(P-3)/2$ zeros. This allows us to determine a simple expression for $\tau_{\max}$ in this case. We find:

$$\tau_{\max} = \frac{3}{4}P - \frac{1}{2} - \frac{1}{4P} \,, \quad P \text{ odd} \,. \tag{40}$$

If the period is even, it is also possible to write down an expression for $\tau_{\max}$. If P is divisible by 4, then

$$\tau_{\max} = \frac{3}{4}P - \frac{1}{2} - \frac{1}{P} \,, \quad \text{P divisible by 4} \,. \tag{41}$$

However, if P is divisible by 2, but not by 4, then:

$$\tau_{\max} = \frac{3}{4}P - \frac{1}{2} - \frac{4}{P} \,, \quad \text{P divisible by 2 and not 4} \,. \tag{42}$$

Note that, in all cases, to leading order in P, $\tau_{\max}$ grows linearly with P:

$$\tau_{\max} \sim \frac{3}{4}P \,. \tag{43}$$

Also, in all cases the preceding results agree with the exact values obtained by enumeration and shown in Tables

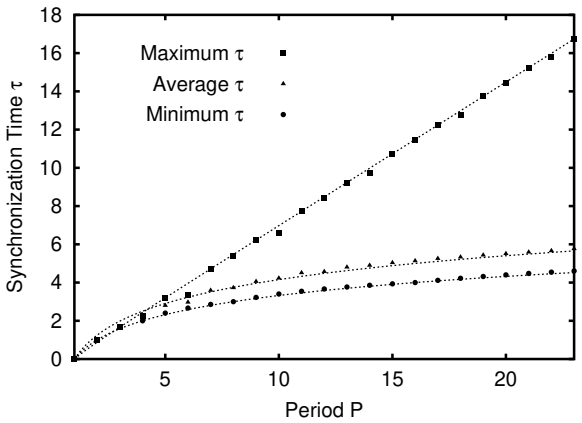

FIG. 10: The minimum, maximum, and average synchronization time plotted as a function of period. The solid points are the exact results from the enumeration; the dashed lines are the approximations developed in the text. The approximation for $\tau_{\max}$ is that of Eq. (40), which is exact for odd periods. The approximation for $\tau_{\min}$ is that of Eq. (35), which is exact for periods that are a power of 2. The approximation for $\langle \mathbf{T} \rangle$ is the empirical fit of Eq. (45).

II and III. Eqs. (40) through (42) were obtained by noting various regularities in the enumeration data and not via a first-principles calculation of the subword statistics.

It is also possible to calculate the transient information $\mathbf{T}$ for the word that maximizes τ for the case in which P is odd. Recall that the $\tau_{\max}$ word in this case consisted of two 1's separated by $(P-3)/2$ zeros. A brute force, direct counting approach yields, for large P:

$$\mathbf{T}_{\max\,\tau} \sim \frac{1}{4} P\mathbf{E} + \frac{P}{2} \,. \tag{44}$$

F. Average Synchronization Time $\langle \tau \rangle$

Figure 10 also shows $\langle \tau \rangle$, the average synchronization time at a given period, as a function of the period. The dashed line in this figure is an empirical fit:

$$\langle \tau \rangle \approx \frac{5}{4} \log_2 P \,. \tag{45}$$

As was the case with the transient information, the average and minimum synchronization times appear to grow asymptotically at the same rate.

VIII. RELATIONSHIPS BETWEEN τ AND T

We have argued that there are important structural differences between distinct periodic sequences at a given period. Moreover, the structural properties captured by $\mathbf{T}$ and τ, though both related to synchronization, can be different. Perhaps the differences are not surprising. Based on their definitions these quantities have different interpretations: $\mathbf{T}$ measures the total uncertainty experienced while synchronizing; τ measures the expected number of observations needed in order to synchronize. To clarify the differences and similarities, in this section we analyze more directly the relationship between $\mathbf{T}$ and τ. We shall see that, while there are relationships between $\mathbf{T}$ and τ for the extremal synchronization behaviors, there is a fairly wide range of $\mathbf{T}$ and τ combinations possible among different words of the same period.

Period	$\tau_{\max}$	$\tau_{\max}$ word
4	2.25	0001
5	3.2	00101
6	3.3333	000101, 000001
7	4.71429	0001001
8	5.375	00100101
9	6.222	000010001
10	6.6	0001001001
11	7.72727	00000100001
12	8.41667	001010010101
13	9.23077	0000001000001
14	9.71429	00001000010001
15	10.7333	000000010000001
16	11.4375	0010101001010101

TABLE III: Maximum synchronization-time words and $\tau_{\max}$ values as a function of period P.

Function	Minimum	Maximum
$\mathbf{T}$	$\frac{1}{2}\log_2^2 P = \frac{1}{2}\mathbf{E}^2$	$\frac{P}{2}\log_2 P = \frac{1}{2}\mathbf{E}2^{\mathbf{E}}$
τ	$\log_2 P = \mathbf{E}$	$\frac{3}{4}2^{\mathbf{E}}$

TABLE IV: Summary of minimum and maximum $\mathbf{T}$ and τ behaviors expressed as a function of period P and as a function of the excess entropy $\mathbf{E}$.

A. Extreme Synchronizations

We start by summarizing the behaviors of the minimum and maximum values of the transient information $\mathbf{T}$ and the synchronization time τ to leading order in the period P in Table IV. Note the enormous range in $\mathbf{T}$ and τ values. For example, if $P = 256$, $\mathbf{T}_{\min} \approx 32$, while $\mathbf{T}_{\max} \approx 1024$. Similarly, for $P = 256$, $\tau_{\min} \approx 8$ while $\tau_{\max} \approx 192$.

Note that for both the minimum and maximum cases, $\mathbf{T}$ and τ are related to each other by:

$$\mathbf{T} = a\tau\mathbf{E} \,, \tag{46}$$

where a is a constant that does not depend on the period P. For the minimum-($\mathbf{T}$, τ) case $a = 1/2$ and for the maximum case $a = 2/3$. Recall, however, that the maximum-$\mathbf{T}$ word is not the same as the maximum-τ word. If one uses $\mathbf{T}_{\max}$ and $\tau_{\max\,\mathrm{T}}$ from Eq. (24) in Eq. (46), one finds $a = 1$. And, if one uses $\mathbf{T}_{\max\,\tau}$ and $\tau_{\max}$, one find $a = 1/3$.

In all cases, note that the ratio of $\mathbf{T}$ to $\mathbf{E}$ gives a quantity proportional to τ. This ratio may thus be viewed as setting a characteristic time for synchronization.

B. Nonextreme Synchronizations

It turns out, however, that these simple relations mask a wide diversity of synchronization behaviors. Fig. 11, a scatter plot of $\mathbf{T}$ and τ for all period-18 sequences, shows that the relationship between $\mathbf{T}$ and τ for individual sequences is quite a bit more complicated. For example, recall from the previous section that the maximum-$\mathbf{T}$ sequence does not correspond with the maximum-τ sequence. In the Fig. 11, the maximum-$\mathbf{T}$ value is shown by a solid triangle and the maximum-τ value by a solid square. The minimum-$\mathbf{T}$ and -τ sequences are identical, however. This corresponds to the single point in the figure's lower left corner.

The wide range of points in the scatter plot of Fig. 11 makes it clear that $\mathbf{T}$ and τ do indeed measure different properties: $\mathbf{T}$ and τ are not simply rescaled versions of each other. Any simple functional relationship is precluded by the diffuse scatter of points. Said another way, if one lists sequences of a given period in order of increasing $\mathbf{T}$, this order will not be the same as listing the sequences in order of increasing τ.

Although there is considerable spread evident in Fig. 11, there are bounds limiting the range of $\mathbf{T}$ values at each τ.

In fact, one can develop an approximate form for the upper bound in Fig. 11 as follows. Given an arbitrary τ value and the period P, we are interested in determining the largest possible $\mathbf{T}$. Call this maximum value $\mathbf{T}_{\mathrm{upper}}$. Recall that the synchronization time $\tau_{\max\,\mathrm{T}}$ for the maximum $\mathbf{T}$ word scales as $P/2$ for large P, as shown in Eq. (24). This lets us write

$$\mathbf{T}_{\max} = \frac{1}{2}P\mathbf{E} = \mathbf{E}\tau_{\max\,\mathrm{T}} \,. \tag{47}$$

To maximize $\mathbf{T}$ for a given τ, assume that the given τ is that which maximizes $\mathbf{T}$. Then, using the above equation, we see that $\mathbf{T}_{\mathrm{upper}} \propto \mathbf{E}\tau$. That is, the slope of the linear upper bound is $\mathbf{E}$. To get the full equation of the line, we require that it go through the minimum-($\tau, \mathbf{T}$) point on the lower left hand corner of Fig. 11. Using Eqs. (26) and (35) for the values of $\mathbf{T}_{\min}$ and $\tau_{\min}$, respectively, we then obtain

$$\mathbf{T}_{\mathrm{upper}} \approx \mathbf{E}\left[\tau - \frac{1}{2}(\mathbf{E}-1)\right] . \tag{48}$$

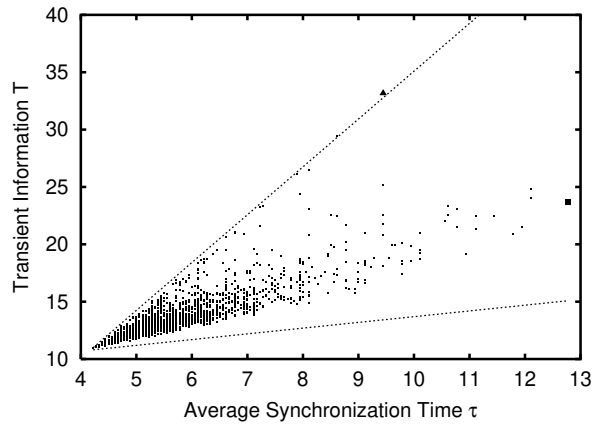

FIG. 11: A scatter plot of $\mathbf{T}$ versus τ for all period-18 sequences. The upper bound is that of Eq. (48) and the lower bound is that of Eq. (50). The solid triangle is the ($\tau, \mathbf{T}$) point for the $\mathbf{T}_{\max}$ sequence; the solid square denotes the $\tau_{\max}$ sequence.

This upper bound is plotted in Fig. 11, with $\mathbf{E} = \log_2 18 \approx 4.1699\ldots$. This upper bound is only approximate because the minimum-($\tau, \mathbf{T}$) values used in its calculation are the asymptotic values for large P. The bound is also approximate because the relationship in Eq. (47), which is true for the maximum-$\mathbf{T}$ sequence (the solid triangle in the figure), is assumed to hold for all sequences. Despite these approximations, the linear upper bound appears to fit the data quite well, as shown in Fig. 11.

A similar line of reasoning allows us to form a lower bound $\mathbf{T}_{\mathrm{lower}}$ for the smallest possible $\mathbf{T}$, given a value of τ and the period P. From Eqs. (27) and (35), for large P we know that $\mathbf{T}_{\min} \approx \frac{1}{2}E^2$ and $\tau_{\min} \approx \mathbf{E}$. This gives us

$$\mathbf{T}_{\min} = \frac{1}{2}\mathbf{E}\tau_{\min} \,. \tag{49}$$

To minimize $\mathbf{T}$, we assume that the given τ is the minimum τ. This tells us that the slope of a linear lower bound is $\frac{1}{2}\mathbf{E}$. Requiring the line to go through the minimum $\tau, \mathbf{T}$ point yields

$$\mathbf{T}_{\mathrm{lower}} = \frac{1}{2}\tau + \frac{1}{2}E^2 \,. \tag{50}$$

This is only an approximate bound, for the same reasons that $\mathbf{T}_{\mathrm{upper}}$ is approximate. Eq. (50) is plotted in Fig. 11. It is indeed a lower bound, but the bound is not very tight.

IX. DISCUSSION AND CONCLUSION

We began by introducing two measures of how difficult it is to synchronize to a sequence: the transient information $\mathbf{T}$, an information-theoretic measure of the total

uncertainty experienced by an observer during the synchronization process; and the synchronization time τ, the number of symbols an observer expects to measure before it is synchronized. We exactly calculated $\mathbf{T}$ and τ for all synchronization-distinct periodic sequences up to and including period 23. We also derived a handful of analytic expressions and approximations for the minimum, maximum, and average $\mathbf{T}$ and τ as a function of period.

Our results show that there are many structural differences between periodic sequences, even within sequences of the *same period*. These differences are simply not captured by commonly used information theoretic measures, such as the entropy rate and the excess entropy. Are all periodic sequences of a given period the same from an information-theoretic standpoint? The answer we gave here is "no"; the recently introduced transient information, captures the information-theoretic differences between sequences of a given period. Are there structural differences between different periodic sequences of the same period? We showed that the answer is "yes" and argued that the differences are significant. Are there distinct classes of periodic sequence? Again, we provided a positive answer, partly by contrasting properties captured by the transient information with a sequence's group theoretic properties and its synchronization time. The latter measures the average number of observations needed to synchronize to the sequence and, for a given sequence at a given period, this number was neither the same nor simply proportional to the transient information. Nonetheless, at a coarse level, the leading-order approximations we developed provide simple and direct links between the three different concepts of phase memory $\mathbf{E}$, synchronization time τ, and the total uncertainty experienced during synchronization $\mathbf{T}$.

There are a number of areas in which our results may find application. For one, recognizing the phases of long periodic sequences of the type investigated here is a key technology for current and future large-scale satellite-to-satellite communication systems. These multiagent systems require extremely accurate and robust satellite-to-satellite synchronization [19] generally for coordination in signal processing and navigation, and specifically for estimating inter-satellite signal delays and distances. Synchronization between two satellites is achieved by one transmitting a very long ($P \sim 1,000$) and known binary periodic sequence. The receiver then infers the phase using a hierarchical correlation algorithm adapted to account for noise corruption of individual symbols during transmission. It turns out that the long periodic *acquisition sequences* are aperiodic sequences formed from relatively "prime" short aperiodic sequences. In fact, the synchronization properties of the component sequences and the composite acquisition sequence determine bounds on the computational effort and noise robustness that can be achieved by the receiver's detection system. We conjecture that the transient information of the acquisition sequence is a key parameter determining these properties.

More generally, these results bear on any situation in which an intelligent agent needs to learn the phase of a periodic component in its environment's behavior; cf. Ref. [8]. This arises in multiagent systems if an agent needs to know where it is in a physical environment that varies periodically in space or time, if it needs to adjust its behavior in response to the repetitive behavior of other agents, or if a collective decision requires coordinated information processing.

Finally, we mention several open questions and directions for future work. First, do these results hold if the synchronization process is noisy and bits flip occasionally? They should extend relatively directly to a class of "noisy-periodic" processes in which one or several symbols in a periodic sequence are random. However, can these results be extended to the full class of finite-memory sources? For a further discussion of the synchronization scenario in a setting not restricted to periodic sequences see Refs. [5] and [8].

Second, one may view those sequences at a given period with the same $\mathbf{T}$ as forming an equivalence class. Recall that distinct periodic sequences were defined by the group $S_k \times C_P$. These are the equivalence classes of zero entropy rate, finite excess entropy sequences. But what new algebra or symmetry group of periodic sequences characterizes those sequences that are equally hard to synchronize to, as measured by $\mathbf{T}$ or τ? If we can characterize identical $\mathbf{T}$ or τ sequences in terms of a symmetry group or algebra, it should be possible to obtain an analytic expression for the number of distinct $\mathbf{T}$ or τ values at a given period.

Third, as noted above, Titsworth [18] pointed out that Eq. (18) gives the number of distinct finite-state machines needed to generate all sequences of a given prime period. The picture is that starting in a different state of such a machine gives the cyclic permutations, while exchanging the output symbols gives the permutation $0 \leftrightarrow 1$. Titsworth does not comment on the structure of the synchronizing states of these machines. To account for synchronization one wants each finite-state machine to have a unique start state that corresponds to the condition of an observer not knowing in which phase a periodic sequence starts. It turns out that ϵ-machines [20–22] capture this synchronizing structure in their transient states, and these in turn determine $\mathbf{T}$. Moreover, this is true for both periodic and random sources. Based on these observations we conjecture that the number of distinct $\mathbf{T}$ values for a given period is the number of distinct ϵ-machines needed to recognize all periodic sequences of a given prime period. That is, we conjecture that those sequences with the same $\mathbf{T}$ value have the same transient ϵ-machine structure, up to edge relabeling.

Fourth, we have recently illustrated several ways that the excess entropy can be extended to two dimensions, and argued that these extentions capture structural properties of two-dimensional patterns [23]. A natural direction for future work is to generalize the transient information to two dimensions and use it to probe the properties

of different two-dimensional periodic patterns.

In summary, we analyzed what is arguably the simplest class of sequences: zero entropy rate, finite-period sequences. The sequences represent behaviors that are often ignored in many fields as being too simple. As we demonstrated, however, their structural complexity and synchronization properties are rich and subtle. Within a given period, there are large differences in structural complexity, as measured by the transient information $\mathbf{T}$ and the synchronization time τ. Thus, even simple periodic systems are surprisingly complex.

Acknowledgments

The authors thank James Massey for sending his lecture notes. This work was supported at the Santa Fe Institute under the Computation, Dynamics, and Inference Program via SFI's core grants from the National Science and MacArthur Foundations. Direct support was provided from DARPA contract F30602-00-2-0583. DPF thanks the Department of Physics and Astronomy at the University of Maine for its hospitality.

[1] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, Champaign-Urbana, 1962.

[2] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.

[3] J. P. Crutchfield and N. H. Packard. Symbolic dynamics of noisy chaos. *Physica D*, 7:201–223, 1983.

[4] P. Grassberger. Toward a quantitative theory of self-generated complexity. *Intl. J. Theo. Phys.*, 25(9):907–938, 1986.

[5] J. P. Crutchfield and D. P. Feldman. Regularities unseen, randomness observed: Levels of entropy convergence. *Chaos*, 15:25–54, 2003.

[6] K. Lindgren and M. G. Norhdal. Complexity measures and cellular automata. *Complex Systems*, 2(4):409–440, 1988.

[7] W. Li. On the relationship between complexity and entropy for Markov chains and regular languages. *Complex Systems*, 5(4):381–399, 1991.

[8] J. P. Crutchfield and D. P. Feldman. Synchronizing to the environment: Information theoretic constraints on agent learning. *Advances in Complex Systems*, 4:251–264, 2001.

[9] R. M. Gray. *Entropy and Information Theory*. Springer-Verlag, New York, 1990.

[10] W. Ebeling. Prediction and entropy of nonlinear dynamical systems and symbolic sequences with LRO. *Physica D*, 109:42–52, 1997.

[11] W. Bialek, I. Nemenman, and N. Tishby. Predictability, complexity, and learning. physics/0007070v2, 2000.

[12] R. Shaw. *The Dripping Faucet as a Model Chaotic System*. Aerial Press, Santa Cruz, California, 1984.

[13] I. Nemenman. *Information Theory and Learning: A Physical Approach*. PhD thesis, Princeton University, 2000.

[14] N. J. Fine. Classes of periodic sequences. *Illinois Journal of Mathematics*, 2:285–302, 1958.

[15] E. N. Gilbert and J. Riordan. Symmetry types of periodic sequences. *Illinois J. Math.*, 5:657–665, 1961.

[16] Combinatorial object server. `http://www.theory.csc.uvic.ca/~cos/root.html`.

[17] N. J. A. Sloan. The on-line encyclopedia of integer sequences. `http://www.research.att.com/~njas/sequences/index.html`. Individual integer sequences in the encyclopedia are referenced by Annnnnn.

[18] R. C. Titsworth. Equivalence classes of periodic sequences. *Illinois J. Math.*, pages 266–270, 1964.

[19] J. L. Massey. Noisy sequence acquisition with minimum computation. Lecture Notes, Ulm Winter School, 2000.

[20] J. P. Crutchfield and K. Young. Inferring statistical complexity. *Phys. Rev. Lett.*, 63:105–108., 1989.

[21] J. P. Crutchfield. The calculi of emergence: Computation, dynamics, and induction. *Physica D*, 75:11–54, 1994.

[22] C. R. Shalizi and J. P. Crutchfield. Computational mechanics: Pattern and prediction, structure and simplicity. *Journal Statistical Physics*, 104:819–881, 2001.

[23] D.P. Feldman and J.P. Crutchfield. Structural information in two-dimensional patterns: Entropy convergence and excess entropy. *Physical Review E*, 67:051104, 2003.