

Infinite Excess Entropy Processes with Countable-State Generators

Nicholas F. Travers
James P. Crutchfield

SFI WORKING PAPER: 2011-11-052

SFI Working Papers contain accounts of scientific work of the author(s) and do not necessarily represent the views of the Santa Fe Institute. We accept papers intended for publication in peer-reviewed journals or proceedings volumes, but not papers that have already appeared in print. Except for papers by our external faculty, papers must be based on work done at SFI, inspired by an invited visit to or collaboration at SFI, or funded by an SFI grant.

©NOTICE: This working paper is included by permission of the contributing author(s) as a means to ensure timely distribution of the scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the author(s). It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author's copyright. These works may be reposted only with the explicit permission of the copyright holder.

www.santafe.edu



SANTA FE INSTITUTE

Infinite Excess Entropy Processes with Countable-State Generators

Nicholas F. Travers^{1,2,*} and James P. Crutchfield^{1,2,3,4,†}

¹*Complexity Sciences Center*

²*Mathematics Department*

³*Physics Department*

University of California at Davis,

One Shields Avenue, Davis, CA 95616

⁴*Santa Fe Institute*

1399 Hyde Park Road, Santa Fe, NM 87501

(Dated: November 17, 2011)

We present two examples of finite-alphabet, infinite excess entropy processes generated by invariant hidden Markov models (HMMs) with countable state sets. The first, simpler example is not ergodic, but the second is. It appears these are the first constructions of processes of this type. Previous examples of infinite excess entropy processes over finite alphabets admit only invariant HMM presentations with uncountable state sets.

Keywords: stationary stochastic process, hidden Markov model, epsilon-machine, ergodicity, entropy rate, excess entropy, mutual information

PACS numbers: 02.50.-r 89.70.+c 05.45.Tp 02.50.Ey

I. INTRODUCTION

For a stationary process (X_t) the *excess entropy* \mathbf{E} is the mutual information between the infinite past $\overleftarrow{X} = \dots X_{-2}X_{-1}$ and the infinite future $\overrightarrow{X} = X_0X_1\dots$. It has a long history and is widely employed as a measure of correlation and complexity in a variety of fields, from ergodic theory and dynamical systems to neuroscience and linguistics [1–6]; see Ref. [7] and references therein for a review.

An important question in classifying a given process is whether it is *finitary* (finite excess entropy) or *infinitary* (infinite excess entropy). Over a finite alphabet, many of the simple process classes commonly studied are always finitary. These include all i.i.d. processes, Markov chains, and processes with finite-state hidden Markov model (HMM) presentations. There also exist several well known examples of finite-alphabet infinitary processes, though. For instance, the symbolic dynamics at the onset of chaos in the logistic map and similar dynamical systems [7] and the stationary representation of the binary Fibonacci sequence [8] are both infinitary.

These latter processes, however, only admit invariant HMM presentations [*] with uncountable state sets. Indeed, any process generated by an invariant countable-state HMM either has positive entropy rate or consists entirely of periodic sequences, which these do not; see App. B. Versions of the *Santa Fe Process* introduced in Ref. [6] are finite-alphabet infinitary processes with positive entropy rate. However, they were not constructed directly as HMMs, and it seems unlikely that they should have any invariant countable-state presentations. To the best of our knowledge, to date there are no examples of finite-alphabet, infinitary processes with invariant countable-state presentations.

We present two such examples. The first is nonergodic, and the information conveyed from the past to the future essentially consists of the ergodic component along a given realization. This example is straightforward to construct and, though previously unpublished, we suspect that others are aware of this or similar constructions. The second, ergodic example, though, is more involved and we believe that both its structure and properties are novel.

To put these contributions in perspective, note that *any* stationary finite-alphabet process may be trivially represented as an invariant HMM with an uncountable state set, in which each infinite history \overleftarrow{x} corresponds to a single state. Thus, it is clear invariant HMMs with uncountable state sets can generate finite-alphabet infinitary processes. In contrast, for any finite-state HMM \mathbf{E} is always finite—bounded by the logarithm of the number of states. The case of countable-state HMMs lies in between the finite-state and uncountable-state cases, and it was previously not clear whether it is possible to have countable-state invariant HMMs that generate infinitary finite-alphabet processes and, in particular, ergodic ones. Here, we show that infinite excess entropy is indeed possible for processes with

*Electronic address: ntravers@math.ucdavis.edu

†Electronic address: chaos@ucdavis.edu

[*] *Invariant* here means that the state sequence of the underlying Markov chain and, hence, the output sequence generated by the HMM are both stationary processes [9].

countable-state state generators, even when ergodicity is required.

II. BACKGROUND

A. Excess Entropy

Definition 1. For a stationary, finite-alphabet process $(X_t)_{t \in \mathbb{Z}}$ the excess entropy \mathbf{E} is the mutual information between the infinite past $\overleftarrow{X} = \dots X_{-2} X_{-1}$ and the infinite future $\overrightarrow{X} = X_0 X_1 \dots$:

$$\mathbf{E} = I[\overleftarrow{X}; \overrightarrow{X}] = \lim_{t \rightarrow \infty} I[\overleftarrow{X}^t; \overrightarrow{X}^t], \quad (1)$$

where $\overleftarrow{X}^t = X_{-t} \dots X_{-1}$ and $\overrightarrow{X}^t = X_0 \dots X_{t-1}$ are the length- t past and future, respectively.

In Refs. [7, 9] it is shown that \mathbf{E} may also be expressed alternatively as:

$$\mathbf{E} = \lim_{t \rightarrow \infty} \left(H[\overrightarrow{X}^t] - h_\mu t \right), \quad (2)$$

where h_μ is the process entropy rate:

$$h_\mu = \lim_{t \rightarrow \infty} \frac{H[\overrightarrow{X}^t]}{t} = \lim_{t \rightarrow \infty} H[X_t | \overrightarrow{X}^t]. \quad (3)$$

That is, the excess entropy \mathbf{E} is the asymptotic amount of entropy (information) in length- t blocks of random variables beyond that explained by the entropy rate. The excess entropy derives its name from this formulation. We also use this formulation to establish that the process of Sec. III A is infinitary.

Expanding the block entropy $H[\overrightarrow{X}^t]$ in Eq. (2) with the chain rule and recombining terms gives another important formulation:

$$\mathbf{E} = \sum_{t=1}^{\infty} (h_\mu(t) - h_\mu), \quad (4)$$

where $h_\mu(t)$ is the length- t entropy-rate approximation:

$$h_\mu(t) = H[X_{t-1} | \overrightarrow{X}^{t-1}], \quad (5)$$

the conditional entropy in the t_{th} symbol given the previous $t - 1$ symbols. This final formulation will be used to establish that the process of Sec. III B is infinitary.

B. Hidden Markov Models

There are two primary types of hidden Markov models: edge-emitting (or *Mealy*) and state-emitting (or *Moore*). We work with the former edge-emitting type, but the two are equivalent in that any model of one type over a finite alphabet may be converted to a model of the other type without changing the cardinality of the state set by more than a constant factor—the alphabet size. Thus, for our purposes, Mealy HMMs are sufficiently general. We also consider only *invariant HMMs*, as defined in [9], over finite alphabets and with countable state sets.

Definition 2. An invariant, edge-emitting, countable-state, finite-alphabet hidden Markov model (hereafter referred to simply as a countable-state HMM) is a 4-tuple $(\mathcal{S}, \mathcal{X}, \{T^x\}, \pi)$ where:

1. \mathcal{S} is a countable set of states.
2. \mathcal{X} is a finite alphabet of output symbols.
3. $T^{(x)}$, $x \in \mathcal{X}$, are symbol labeled transition matrices. $T_{\sigma\sigma'}^{(x)}$ is the probability that state σ transitions to state σ' on symbol x .
4. π is an invariant or stationary distribution for the underlying Markov chain over states with transition matrix $T = \sum_{x \in \mathcal{X}} T^{(x)}$. That is, π satisfies $\pi = \pi T$.

Remark. “Countable” in Property 1 means either finite or countably infinite. If the state set \mathcal{S} is finite, we also refer to the HMM as finite-state.

A hidden Markov model may be depicted as a directed graph with labeled edges. The vertices are the states $\sigma \in \mathcal{S}$ and, for all $\sigma, \sigma' \in \mathcal{S}$ with $T_{\sigma\sigma'}^{(x)} > 0$, there is a directed edge from state σ to state σ' labeled $p|x$ for the symbol x and transition probability $p = T_{\sigma\sigma'}^{(x)}$. These probabilities are normalized so that the sum of probabilities on all outgoing edges from each state is 1. An example is given in Fig. 1.



FIG. 1: A hidden Markov model (the ϵ -machine) for the Even Process. The support for this process consists of all binary sequences in which blocks of uninterrupted 1s are even in length, bounded by 0s. After each even length is reached, there is a probability p of breaking the block of 1s by inserting a 0. The machine has two internal states $\mathcal{S} = \{\sigma_1, \sigma_2\}$, a two symbol alphabet $\mathcal{X} = \{0, 1\}$, and a single parameter $p \in (0, 1)$ that controls the transition probabilities. The associated Markov chain over states is finite-state and irreducible and, thus, has a unique stationary distribution $\pi = (\pi_1, \pi_2) = (1/(2-p), (1-p)/(2-p))$. The graphical representation of the machine is given on the left, with the corresponding transition matrices on the right. In the graphical representation the **symbols** labeling the transitions have been colored **blue**, for visual contrast, while the transition probabilities are black.

The operation of a HMM may be thought of as a weighted random walk on the associated graph. That is, from the current state σ the next state σ' is determined by following an outgoing edge from σ according to the edges' relative probabilities (or weights). During the transition, the HMM outputs the symbol x labeling this edge.

The state sequence (S_t) determined in such a fashion is simply a Markov chain with transition matrix T . However, we are interested not simply in the state sequence of the HMM, but rather the associated sequence of output symbols (X_t) that are generated by reading the labels off the edges as they are followed. The interpretation is that an observer of the HMM may directly observe this sequence of output symbols, but not the hidden internal states. Alternatively, one may consider the Markov chain over edges (E_t) , of which the observed symbol sequence (X_t) is simply a projection.

In either case, the process (X_t) generated by the HMM $(\mathcal{S}, \mathcal{X}, \{T^x\}, \pi)$ is defined as the output sequence of edge symbols, which results from running the Markov chain over states according to the stationary law with marginals $\mathbb{P}(S_0) = \mathbb{P}(S_t) = \pi$. It is easy to verify that this process is itself stationary, with word probabilities given by:

$$\mathbb{P}(w) = \|\pi T^{(w)}\|_1, \quad (6)$$

where for a given word $w = w_1 \dots w_n \in \mathcal{X}^+$, $T^{(w)}$ is the word transition matrix $T^{(w)} = T^{(w_1)} \dots T^{(w_n)}$. The *process language* is the set of words $\mathcal{L} = \{w : \mathbb{P}(w) > 0\}$.

Remark. Even for a noninvariant HMM $(\mathcal{S}, \mathcal{X}, \{T^x\}, \pi)$, where the state distribution π is not stationary, one may always define a one-sided process $(X_t)_{t \geq 0}$ with marginals given by:

$$\mathbb{P}(\vec{X}^{|w|} = w) = \|\pi T^{(w)}\|_1. \quad (7)$$

Furthermore, though the state sequence $(S_t)_{t \geq 0}$ will not be a stationary process if π is not a stationary distribution for T , the output sequence $(X_t)_{t \geq 0}$ may still be stationary. In fact, Ref. [9, Example 2.9] showed that any one-sided process over a finite alphabet \mathcal{X} , stationary or not, may be represented as a countable-state noninvariant HMM in which the states correspond to finite-length words in \mathcal{X}^+ , of which there are only countably many. By stationarity, a one-sided stationary process generated by such a noninvariant HMM can be uniquely extended to a two-sided stationary process. So, in a sense, any two-sided stationary process $(X_t)_{t \in \mathbb{Z}}$ can be said to be generated by a noninvariant countable-state HMM. Though, this is a slightly unnatural interpretation of process generation in that the two-sided process $(X_t)_{t \in \mathbb{Z}}$ is not directly the process obtained by reading symbols off the edges of the HMM as it runs along transitioning between states in bi-infinite time. In either case, the space of stationary finite-alphabet processes generated by noninvariant countable-state HMMs is too large: it includes all stationary finite-alphabet processes. Due to this, we restrict to the case of invariant HMMs where both the state sequence (S_t) and output sequence (X_t) are stationary. Clearly, if one allows finite-alphabet processes generated by noninvariant countable-state HMMs there are infinitary examples. And so, in the following development HMM will implicitly mean invariant HMM, but this will no longer be stated.

We consider now an important property known as unifilarity. This property is useful in that many quantities are analytically computable only for unifilar HMMs. In particular, for unifilar HMMs the entropy rate h_μ is often directly

computable, unlike the nonunifilar case. Both of the examples constructed in Sec. III are unifilar, as is the Even Process HMM of Fig. 1.

Definition 3. A HMM $(\mathcal{S}, \mathcal{X}, \{T^x\}, \pi)$ is unifilar if for each $\sigma \in \mathcal{S}$ and $x \in \mathcal{X}$ there is at most one outgoing edge from state σ labeled with symbol x in the associated graph G .

It is well known that for any finite-state unifilar HMM the entropy rate in the output process (X_t) is simply the conditional entropy in the next symbol given the current state:

$$h_\mu = H[X_0|S_0] = \sum_{\sigma \in \mathcal{S}} \pi_\sigma h_\sigma, \quad (8)$$

where π_σ is the stationary probability of state σ and $h_\sigma = H[X_0|S_0 = \sigma]$ is the conditional entropy in the next symbol given that the current state is σ .

We are unaware, though, of any proof that this is generally true for countable-state HMMs. If the entropy in the stationary distribution $H[\pi]$ is finite, then a proof along the lines given in Ref. [10] carries through to the countable-state case and Eq. (8) still holds. However, countable-state HMMs may sometimes have $H[\pi] = \infty$. Furthermore, it can be shown [9] that the excess entropy \mathbf{E} is always bounded above by $H[\pi]$. So, for the infinitary process of Sec. IIIB we need slightly more than unifilarity to establish the value of h_μ . To this end, we consider a property known as *exactness* [11].

Definition 4. A HMM is said to be exact if for a.e. infinite future $\vec{x} = x_0x_1\dots$ generated by the HMM an observer synchronizes to the internal state after a finite time. That is, for a.e. \vec{x} there exists $t \in \mathbb{N}$ such that $H[S_t|\vec{X}^t = \vec{x}^t] = 0$, where $\vec{x}^t = x_0x_1\dots x_{t-1}$ denotes the first t symbols of a given \vec{x} .

In App. A we prove the following proposition.

Proposition 1. For any countable-state, exact, unifilar HMM, the entropy rate is given by the standard formula of Eq. (8).

The HMM constructed in Sec. IIIB is both exact and unifilar, so Prop. 1 applies. Using this explicit formula for h_μ , we will show that $\mathbf{E} = \sum_{t=1}^{\infty} (h_\mu(t) - h_\mu)$ is infinite.

III. CONSTRUCTIONS

We present two constructions of (invariant) countable-state HMMs that generate infinitary processes. In the first example the output process is not ergodic, but in the second it is.

A. Heavy-Tailed Periodic Mixture: An infinitary nonergodic process with a countable-state presentation

Figure 2 depicts a countable-state HMM M , for a nonergodic infinitary process \mathcal{P} . The machine M consists of a countable collection of disjoint strongly connected subcomponents M_i , $i \geq 2$. For each i , the component M_i generates the periodic processes \mathcal{P}_i consisting of $i - 1$ 1s followed by a 0. The weighting over components is taken as a heavy-tailed distribution with infinite entropy. For this reason, we refer to the process M generates as the *Heavy-Tailed Periodic Mixture* (HPM) process.

Intuitively, the information transmitted from the past to the future for the HPM Process is the ergodic component i along with the phase of the period- i process \mathcal{P}_i in this component. This is more information than simply the ergodic component i , which is itself an infinite amount of information: $H[(\mu_2, \mu_3, \dots)] = \infty$. Hence, \mathbf{E} should be infinite. This intuition can be made precise using the ergodic decomposition theorem of Debowski [12], but we present a more direct proof here.

Proposition 2. The HPM Process has infinite excess entropy.

Proof. For the HPM Process \mathcal{P} we will show that (i) $\lim_{t \rightarrow \infty} H[\vec{X}^t] = \infty$ and (ii) $h_\mu = 0$. The conclusion then follows immediately from Eq. (2). To this end, we define sets:

$$\begin{aligned} W_{i,t} &= \{w : |w| = t \text{ and } w \text{ is in the support of process } \mathcal{P}_i\}, \\ U_t &= \bigcup_{2 \leq i \leq t/2} W_{i,t}, \text{ and} \\ V_t &= \bigcup_{i > t/2} W_{i,t}. \end{aligned}$$

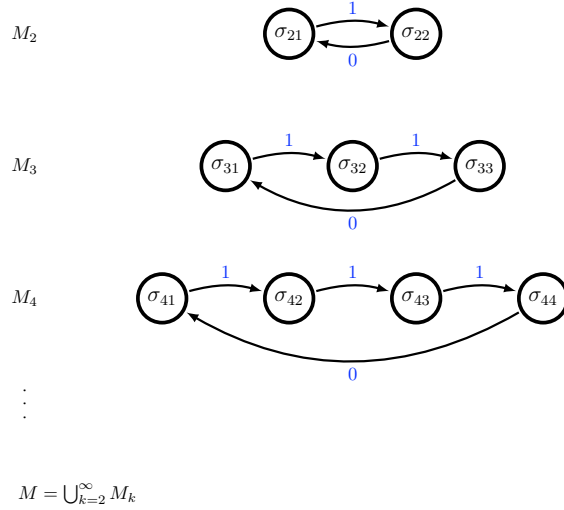


FIG. 2: A countable-state HMM for the Heavy-Tailed Periodic Mixture Process. The machine M is the union of the machines $M_i, i \geq 2$, generating the period- i processes of $i - 1$ 1s followed by a 0. All topologically allowed transitions have probability 1. So, for visual clarity these probabilities are omitted from the edge labels and only the symbols labeling the transitions are given. The stationary distribution π is chosen such that the combined probability μ_i of all states in the the i th component is $\mu_i = C/(i \log^2 i)$, where $C = 1/(\sum_{i=2}^{\infty} 1/(i \log^2 i))$ is a normalizing constant. Formally, the HMM $M = (\mathcal{S}, \mathcal{X}, \{T^{(x)}\}, \pi)$ has alphabet $\mathcal{X} = \{0, 1\}$, state set $\mathcal{S} = \{\sigma_{ij} : i \geq 2, 1 \leq j \leq i\}$, stationary distribution π defined by $\pi_{ij} = C/(i^2 \log^2 i)$, and transition probabilities $T_{ij, i(j+1)}^{(1)} = 1$ for $i \geq 2$ and $1 \leq j < i$, $T_{ii, i1}^{(0)} = 1$ for $i \geq 2$, and all other transitions probabilities 0. Note that all logs here (and throughout) are taken base 2, as is typical when using information-theoretic quantities.

Note that any word $w \in W_{i,t}$ with $i \leq t/2$ contains at least two 0s. Therefore:

1. No two distinct states σ_{ij} and $\sigma_{i'j'}$ with $i \leq t/2$ generate the same length t word.
2. The sets $W_{i,t}, i \leq t/2$, are disjoint from both each other and V_t .

It follows that each word $w \in W_{i,t}$, with $i \leq t/2$, can only be generated from a single state σ_{ij} of the HMM and has probability:

$$\begin{aligned}
 \mathbb{P}(w) &= \mathbb{P}(\vec{X}^t = w) \\
 &= \mathbb{P}(S_0 = \sigma_{ij}) \cdot \mathbb{P}(\vec{X}^t = w | S_0 = \sigma_{ij}) \\
 &= \pi_{ij} \cdot 1 \\
 &= C/(i^2 \log^2 i).
 \end{aligned} \tag{9}$$

Hence, for any fixed t :

$$\begin{aligned}
 H[\vec{X}^t] &= \sum_{|w|=t} \mathbb{P}(w) \log \left(\frac{1}{\mathbb{P}(w)} \right) \\
 &\geq \sum_{i=2}^{\lfloor t/2 \rfloor} \sum_{w \in W_{i,t}} \frac{C}{i^2 \log^2(i)} \log \left(\frac{i^2 \log^2(i)}{C} \right) \\
 &= \sum_{i=2}^{\lfloor t/2 \rfloor} \frac{C}{i \log^2(i)} \log \left(\frac{i^2 \log^2(i)}{C} \right),
 \end{aligned}$$

so:

$$\lim_{t \rightarrow \infty} H[\vec{X}^t] \geq \sum_{i=2}^{\infty} \frac{C}{i \log^2(i)} \log \left(\frac{i^2 \log^2(i)}{C} \right) = \infty, \tag{10}$$

which proves Claim (i). Now, to prove Claim (ii) consider the quantity:

$$\begin{aligned} h_\mu(t+1) &= H[X_t|\vec{X}^t] \\ &= \sum_{w \in U_t} \mathbb{P}(w) \cdot H[X_t|\vec{X}^t = w] + \sum_{w \in V_t} \mathbb{P}(w) \cdot H[X_t|\vec{X}^t = w]. \end{aligned} \quad (11)$$

On the one hand, for $w \in U_t$, $H[X_t|\vec{X}^t = w] = 0$ since the current state and, hence, entire future are completely determined by any word $w \in U_t$. On the other hand, for $w \in V_t$, $H[X_t|\vec{X}^t = w] \leq 1$ since the alphabet is binary. Moreover, the combined probability of all words in the set V_t is simply the probability of starting in some component M_i with $i > t/2$: $\mathbb{P}(V_t) = \sum_{i > t/2} \mu_i$. Thus, by Eq. (11), $h_\mu(t+1) \leq \sum_{i > t/2} \mu_i$. Since $\sum_i \mu_i$ converges, it follows that $h_\mu(t) \searrow 0$, which verifies Claim (ii). \square

B. Branching Copy Process: An infinitary ergodic process with a countable-state presentation

Figure 3 depicts a countable-state HMM M for the ergodic, infinitary *Branching Copy Process*. Essentially, the machine M consists of a binary tree with loop backs and a self-loop on the root node. From the root node a path is chosen down the tree with each left-right (or 0-1) choice equally likely. But, at each step there is also a chance of turning back towards the root. The path back is not a single step, however. It has length equal to the number of steps taken down the tree before returning back, and copies the path taken down symbol-wise with 0s replaced by 2s and 1s replaced by 3s. There is also a high self-loop probability at the root node on symbol 4, so some number of 4s will normally be generated after returning to the root node before preceding again down the tree. The process generated by this machine is referred to as the Branching Copy (BC) Process, because the branch taken down the tree is copied on the loop back to the root.

By inspection we see that the machine is unifilar with *synchronizing word* $w = 4$, i.e. $H[S_1|X_0 = 4] = 0$. Since the underlying Markov chain over states (S_t) is positive recurrent, the state sequence (S_t) and symbol sequence (X_t) are both ergodic. Thus, a.e. infinite future \vec{x} contains a 4, so the machine is exact. Therefore, Prop. 1 may be applied, and we know the entropy rate h_μ is given by the standard formula of Eq. (8): $h_\mu = \sum_\sigma \pi_\sigma h_\sigma$. Since $\mathbb{P}(S_t = \sigma) = \pi_\sigma$ for any $t \in \mathbb{N}$, we may alternatively represent this entropy rate as:

$$\begin{aligned} h_\mu &= \sum_\sigma \left(\sum_{w \in \mathcal{L}_t} \mathbb{P}(w) \phi(w)_\sigma \right) h_\sigma \\ &= \sum_{w \in \mathcal{L}_t} \mathbb{P}(w) \left(\sum_\sigma \phi(w)_\sigma h_\sigma \right) \\ &= \sum_{w \in \mathcal{L}_t} \mathbb{P}(w) \tilde{h}_w, \end{aligned} \quad (12)$$

where $\mathcal{L}_t = \{w : |w| = t, \mathbb{P}(w) > 0\}$ is the set of length t words in the process language \mathcal{L} , $\phi(w)$ is the conditional state distribution induced by the word w (i.e., $\phi(w)_\sigma = \mathbb{P}(S_t = \sigma|\vec{X}^t = w)$), and $\tilde{h}_w = \sum_\sigma \phi(w)_\sigma h_\sigma$ is the $\phi(w)$ -weighted average entropy in the next symbol given knowledge of the current state σ .

Similarly, for any $t \in \mathbb{N}$ the entropy-rate approximation $h_\mu(t+1)$ may be expressed as:

$$h_\mu(t+1) = H[X_t|\vec{X}^t] = \sum_{w \in \mathcal{L}_t} \mathbb{P}(w) h_w, \quad (13)$$

where $h_w = H[X_t|\vec{X}^t = w] = H[X_0|S_0 \sim \phi(w)]$ is the entropy in the next symbol given the word w . Combining Eqs. (12) and (13) we have for any $t \in \mathbb{N}$:

$$h_\mu(t+1) - h_\mu = \sum_{w \in \mathcal{L}_t} \mathbb{P}(w) (h_w - \tilde{h}_w). \quad (14)$$

By concavity of the entropy function, the quantity $h_w - \tilde{h}_w$ is always nonnegative. Furthermore, in Claim 5 we show that $h_w - \tilde{h}_w$ is always bounded below by some fixed positive constant for any word w consisting entirely of 2s and 3s. Also, in Claim 3 we show that $\mathbb{P}(W_t)$ scales as $1/t$, where W_t is the set of length- t words consisting entirely of 2s

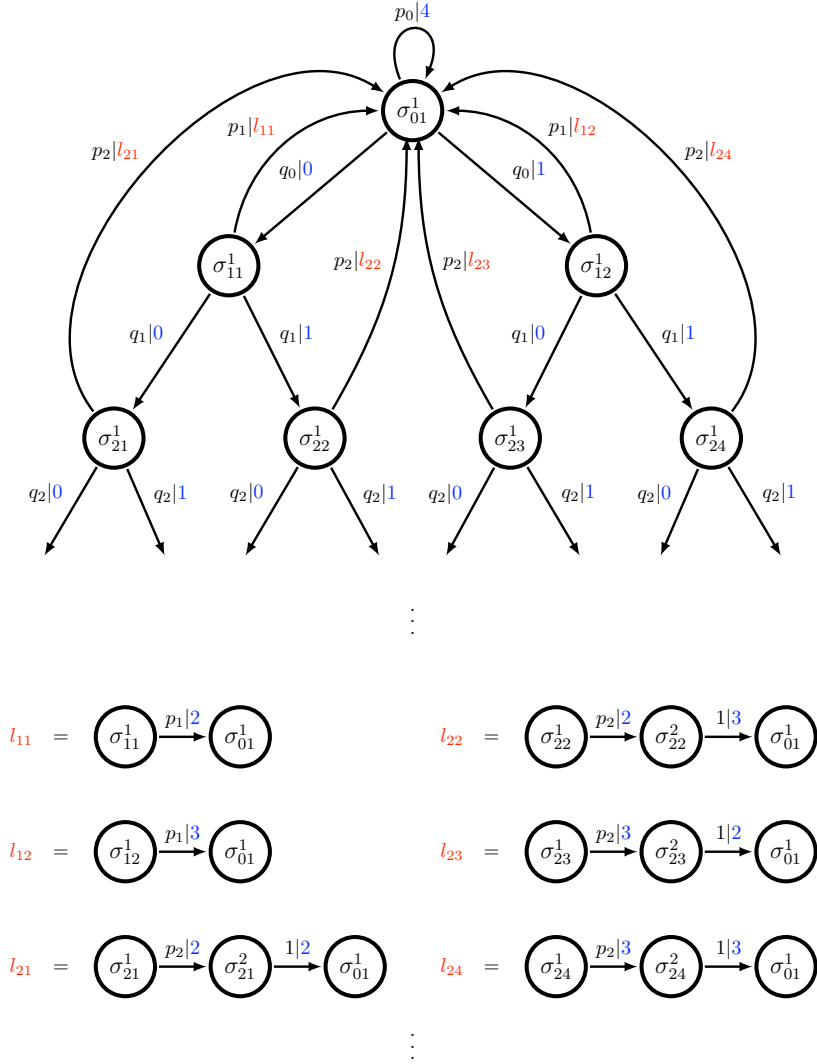


FIG. 3: A countable-state HMM for the Branching Copy Process. The machine M is essentially a binary tree with loop-back paths from each node in the tree to the root node and a self-loop on the root. At each node σ_{ij}^k in the tree there is a probability $2q_i$ of continuing down the tree and a probability $p_i = 1 - 2q_i$ of turning back towards the root σ_{01}^1 on path $l_{ij}^k \sim \sigma_{ij}^k \rightarrow \sigma_{ij}^{k+1} \rightarrow \sigma_{ij}^{k+2} \dots \rightarrow \sigma_{ij}^1 \rightarrow \sigma_{01}^1$. If the choice is made to head back, the next $i - 1$ transitions are deterministic. The path of 0s and 1s taken to get from σ_{01}^1 to σ_{ij}^k is copied on the return with 0s replaced by 2s and 1s replaced by 3s. Formally, the alphabet is $\mathcal{X} = \{0, 1, 2, 3, 4\}$ and the state set is $\mathcal{S} = \{\sigma_{ij}^k : i \geq 0, 1 \leq j \leq 2^i, 1 \leq k \leq \max\{i, 1\}\}$. The nonzero transition probabilities are as depicted graphically with $p_i = 1 - 2q_i$ for all $i \geq 0$, $q_i = i^2/[2(i+1)^2]$ for all $i \geq 1$, and $q_0 > 0$ taken sufficiently small so that $H[(p_0, q_0, q_0)] \leq 1/300$. The graph is strongly connected so the Markov chain over states is irreducible. Claim 1 shows that the Markov chain is also positive recurrent and, hence, has a unique stationary distribution π . Claim 2 gives the form of π .

and 3s. Combining these results it follows that $h_\mu(t+1) - h_\mu \gtrsim 1/t$ and, hence, the sum $\mathbf{E} = \sum_{t=1}^{\infty} (h_\mu(t) - h_\mu)$ is infinite.

A more detailed analysis with the claims and their proofs is given below. In this we will use the following notation:

- $\mathbb{P}_\sigma(\cdot) = \mathbb{P}(\cdot | S_0 = \sigma)$,
- $V_t = \{w \in \mathcal{L}_t : w \text{ contains only 0s and 1s}\}$ and $W_t = \{w \in \mathcal{L}_t : w \text{ contains only 2s and 3s}\}$,
- $\pi_{ij}^k = \mathbb{P}(\sigma_{ij}^k)$ is the stationary probability of state σ_{ij}^k ,
- $R_{ij} = \{\sigma_{ij}^1, \sigma_{ij}^2, \dots, \sigma_{ij}^i\}$, and
- $\pi_{ij} = \sum_{k=1}^i \pi_{ij}^k$ and $\pi_i^1 = \sum_{j=1}^{2^i} \pi_{ij}^1$.

Note that:

$$\mathbb{P}_{\sigma_{01}^1}(\vec{X}^t \in V_t) = \frac{1-p_0}{t^2}, \text{ for all } t \geq 1, \quad (15)$$

and:

$$p_i = \frac{2i+1}{(i+1)^2} \leq \frac{2}{i}, \text{ for all } i \geq 1. \quad (16)$$

These facts will be used in the proof of the Claim 1.

Claim 1. *The underlying Markov chain over states for the HMM is positive recurrent.*

Proof. Let $\tau_{\sigma_{01}^1} = \min\{t > 0 : S_t = \sigma_{01}^1\}$ be the first return time to state σ_{01}^1 . Then, by continuity:

$$\begin{aligned} \mathbb{P}_{\sigma_{01}^1}(\tau_{\sigma_{01}^1} = \infty) &= \lim_{t \rightarrow \infty} \mathbb{P}_{\sigma_{01}^1}(\tau_{\sigma_{01}^1} > 2t) \\ &= \lim_{t \rightarrow \infty} \mathbb{P}_{\sigma_{01}^1}(\vec{X}^{2t+1} \in V_{2t+1}) \\ &= \lim_{t \rightarrow \infty} \frac{1-p_0}{(t+1)^2} \\ &= 0. \end{aligned}$$

Hence, the Markov chain is recurrent and we have:

$$\begin{aligned} \mathbb{E}_{\sigma_{01}^1}(\tau_{\sigma_{01}^1}) &= \sum_{t=1}^{\infty} \mathbb{P}_{\sigma_{01}^1}(\tau_{\sigma_{01}^1} = t) \cdot t \\ &= p_0 \cdot 1 + \sum_{t=1}^{\infty} \mathbb{P}_{\sigma_{01}^1}(\tau_{\sigma_{01}^1} = 2t) \cdot 2t \\ &= p_0 + \sum_{t=1}^{\infty} \mathbb{P}_{\sigma_{01}^1}(\vec{X}^t \in V_t) \cdot p_t \cdot 2t \\ &\leq p_0 + \sum_{t=1}^{\infty} \frac{1-p_0}{t^2} \cdot \frac{2}{t} \cdot 2t \\ &< \infty, \end{aligned}$$

from which it follows that the chain is also positive recurrent. Note that the topology of the chain implies the first return time may not be an odd integer greater than 1. \square

Claim 2. *The stationary distribution π has:*

$$\pi_{ij}^1 = \frac{C}{i^2 \cdot 2^i}, \quad i \geq 1, 1 \leq j \leq 2^i, \quad (17)$$

$$\pi_{ij}^k = \frac{C}{i^2 \cdot 2^i} \cdot \frac{2i+1}{(i+1)^2}, \quad i \geq 1, 1 \leq j \leq 2^i, 2 \leq k \leq i, \quad (18)$$

where $C = \pi_{01}^1(1-p_0)$.

Proof. Existence of a unique stationary distribution π is guaranteed by Claim 1. Given this, clearly $\pi_1^1 = \pi_{01}^1(1-p_0)$. Similarly, for $i \geq 1$, $\pi_{i+1}^1 = \pi_i^1(1-p_i) = \pi_i^1 \frac{i^2}{(i+1)^2}$, from which it follows by induction that $\pi_i^1 = \pi_{01}^1(1-p_0)/i^2$, for all $i \geq 1$. By symmetry $\pi_{ij}^1 = \pi_i^1/2^i$ for each $i \in \mathbb{N}$ and $1 \leq j \leq 2^i$. Therefore, for each $i \in \mathbb{N}$, $1 \leq j \leq 2^i$ we have $\pi_{ij}^1 = \pi_{01}^1(1-p_0)/(i^2 \cdot 2^i) = C/(i^2 \cdot 2^i)$ as was claimed. Moreover, $\pi_{ij}^2 = \pi_{ij}^1 \cdot p_i = \pi_{ij}^1 \cdot \frac{2i+1}{(i+1)^2}$. Combining with the expression for π_{ij}^1 gives $\pi_{ij}^2 = \frac{C}{i^2 \cdot 2^i} \cdot \frac{2i+1}{(i+1)^2}$. By induction, $\pi_{ij}^2 = \pi_{ij}^3 = \dots = \pi_{ij}^i$, so this completes the proof. \square

Note that for all $i \geq 1$ and $1 \leq j \leq 2^i$:

$$\pi_{ij} = \frac{C}{2^i \cdot i^2} + (i-1) \frac{C}{2^i \cdot i^2} \cdot \frac{2i+1}{(i+1)^2} \geq \frac{C}{2^i \cdot i^2}, \text{ and} \quad (19)$$

$$\pi_{ij} = \frac{C}{2^i \cdot i^2} + (i-1) \frac{C}{2^i \cdot i^2} \cdot \frac{2i+1}{(i+1)^2} \leq \frac{3C}{2^i \cdot i^2}. \quad (20)$$

Also note that for any $t \in \mathbb{N}$ and $i \geq 2t$ we have for each $1 \leq j \leq 2^i$:

1. $\mathbb{P}(\vec{X}^t \in W_t | S_0 = \sigma_{ij}^k) = 1$, for $2 \leq k \leq \lceil i/2 \rceil + 1$.
2. $(\sum_{k=2}^i \pi_{ij}^k) / \pi_{ij} \geq 1/3$ and $|\{k : 2 \leq k \leq \lceil i/2 \rceil + 1\}| \geq \frac{1}{2} \cdot |\{k : 2 \leq k \leq i\}|$. Hence, $(\sum_{k=2}^{\lceil i/2 \rceil + 1} \pi_{ij}^k) / \pi_{ij} \geq 1/6$.

Therefore, for each $t \in \mathbb{N}$:

$$\mathbb{P}(\vec{X}^t \in W_t | S_0 \in R_{ij}) \geq 1/6, \text{ for all } i \geq 2t \text{ and } 1 \leq j \leq 2^i. \quad (21)$$

Equations (19), (20), and (21) will be used in the proof of Claim 3 below, along with the following simple lemma.

Lemma 1 (Integral Test). *Let $n \in \mathbb{N}$ and let $f : [n, \infty) \rightarrow \mathbb{R}$ be a positive, continuous, monotone-decreasing function, then:*

$$\int_n^\infty f(x) dx \leq \sum_{k=n}^\infty f(k) \leq f(n) + \int_n^\infty f(x) dx.$$

Claim 3. $\mathbb{P}(W_t)$ decays roughly as $1/t$. More exactly, $C/12t \leq \mathbb{P}(W_t) \leq 6C/t$ for all $t \in \mathbb{N}$.

Proof. For any state σ_{ij}^k with $i < t$, $\mathbb{P}(\vec{X}^t \in W_t | S_0 = \sigma_{ij}^k) = 0$. Thus, we have:

$$\begin{aligned} \mathbb{P}(W_t) &= \mathbb{P}(\vec{X}^t \in W_t) \\ &= \sum_{i=t}^\infty \sum_{j=1}^{2^i} \mathbb{P}(S_0 \in R_{ij}) \cdot \mathbb{P}(\vec{X}^t \in W_t | S_0 \in R_{ij}) \\ &= \sum_{i=t}^\infty 2^i \cdot \mathbb{P}(S_0 \in R_{i1}) \cdot \mathbb{P}(\vec{X}^t \in W_t | S_0 \in R_{i1}), \end{aligned} \quad (22)$$

where the second equality follows from symmetry. We prove the bounds from above and below on $\mathbb{P}(W_t)$ separately using Eq. (22).

- *Bound from below:*

$$\begin{aligned} \mathbb{P}(W_t) &= \sum_{i=t}^\infty 2^i \cdot \mathbb{P}(S_0 \in R_{i1}) \cdot \mathbb{P}(\vec{X}^t \in W_t | S_0 \in R_{i1}) \\ &\geq \sum_{i=2t}^\infty 2^i \cdot \mathbb{P}(S_0 \in R_{i1}) \cdot \mathbb{P}(\vec{X}^t \in W_t | S_0 \in R_{i1}) \\ &\stackrel{(a)}{\geq} \sum_{i=2t}^\infty 2^i \cdot \frac{C}{2^i \cdot i^2} \cdot \frac{1}{6} \\ &= \frac{C}{6} \sum_{i=2t}^\infty \frac{1}{i^2} \\ &\stackrel{(b)}{\geq} \frac{C}{6} \int_{2t}^\infty \frac{1}{x^2} dx \\ &= \frac{C}{12t}. \end{aligned} \quad (23)$$

Here, (a) follows from Eqs. (19) and (21) and (b) from Lemma 1.

- *Bound from above:*

$$\begin{aligned}
\mathbb{P}(W_t) &= \sum_{i=t}^{\infty} 2^i \cdot \mathbb{P}(S_0 \in R_{i1}) \cdot \mathbb{P}(\vec{X}^t \in W_t | S_0 \in R_{i1}) \\
&\stackrel{(a)}{\leq} \sum_{i=t}^{\infty} 2^i \cdot \frac{3C}{2^i \cdot i^2} \cdot 1 \\
&= 3C \sum_{i=t}^{\infty} \frac{1}{i^2} \\
&\stackrel{(b)}{\leq} 3C \left(\frac{1}{t^2} + \int_t^{\infty} \frac{1}{x^2} dx \right) \\
&= 3C \cdot \left(\frac{1}{t^2} + \frac{1}{t} \right) \\
&\leq \frac{6C}{t}.
\end{aligned} \tag{24}$$

Here, (a) follows from Eq. (20) and (b) from Lemma 1. □

Claim 4. $\mathbb{P}(X_t \in \{2, 3\} | \vec{X}^t = w) \geq 1/150$, for all $t \in \mathbb{N}$ and $w \in W_t$.

Proof. Applying Claim 3 we have for any $t \in \mathbb{N}$:

$$\begin{aligned}
\mathbb{P}(X_t \in \{2, 3\} | \vec{X}^t \in W_t) &= \mathbb{P}(\vec{X}^{t+1} \in W_{t+1} | \vec{X}^t \in W_t) \\
&= \mathbb{P}(\vec{X}^{t+1} \in W_{t+1}, \vec{X}^t \in W_t) / \mathbb{P}(\vec{X}^t \in W_t) \\
&= \mathbb{P}(\vec{X}^{t+1} \in W_{t+1}) / \mathbb{P}(\vec{X}^t \in W_t) \\
&\geq \frac{C/12(t+1)}{6C/t} \\
&= \frac{1}{72} \cdot \frac{t}{t+1} \\
&\geq \frac{1}{150}.
\end{aligned}$$

By symmetry, $\mathbb{P}(X_t \in \{2, 3\} | \vec{X}^t = w)$ is the same for each $w \in W_t$. Thus, the same bound must also hold for each $w \in W_t$ individually: $\mathbb{P}(X_t \in \{2, 3\} | \vec{X}^t = w) \geq 1/150$ for all $w \in W_t$. □

Claim 5. For each $t \in \mathbb{N}$ and $w \in W_t$, (i) $\tilde{h}_w \leq 1/300$ and (ii) $h_w \geq 1/150$. Hence, $h_w - \tilde{h}_w \geq 1/300$.

Proof of (i). $h_{\sigma_{ij}^k} = 0$, for all $i \geq 1$, $1 \leq j \leq 2^i$, and $k \geq 2$. And, for each $w \in W_t$, $\phi(w)_{\sigma_{ij}^1} = 0$, for all $i \geq 1$ and $1 \leq j \leq 2^i$. Hence, for each $w \in W_t$, $\tilde{h}_w = \sum_{\sigma \in \mathcal{S}} \phi(w)_{\sigma} h_{\sigma} = \phi(w)_{\sigma_{01}^1} h_{\sigma_{01}^1}$. By construction of the machine $h_{\sigma_{01}^1} \leq 1/300$ and, clearly, $\phi(w)_{\sigma_{01}^1}$ can never exceed 1. Thus, $\tilde{h}_w \leq 1/300$ for all $w \in W_t$. □

Proof of (ii). Let the random variable Z_t be defined by: $Z_t = 1$ if $X_t \in \{2, 3\}$ and $Z_t = 0$ if $X_t \notin \{2, 3\}$. By Claim 4, $\mathbb{P}(Z_t = 1 | \vec{X}^t = w) \geq 1/150$ for any $w \in W_t$ and, by symmetry, the probabilities of a 2 or a 3 following any word $w \in W_t$ are equal, so $\mathbb{P}(X_t = 2 | \vec{X}^t = w, Z_t = 1) = \mathbb{P}(X_t = 3 | \vec{X}^t = w, Z_t = 1) = 1/2$. Therefore, for any $w \in W_t$:

$$\begin{aligned}
h_w &= H[X_t | \vec{X}^t = w] \\
&\geq H[X_t | \vec{X}^t = w, Z_t] \\
&\geq \mathbb{P}(Z_t = 1 | \vec{X}^t = w) \cdot H[X_t | \vec{X}^t = w, Z_t = 1] \\
&\geq 1/150 \cdot 1.
\end{aligned}$$

□

Claim 6. The quantity $h_{\mu}(t) - h_{\mu}$ decays at a rate no faster than $1/t$. More exactly, $h_{\mu}(t+1) - h_{\mu} \geq \frac{C}{3600t}$, for all $t \in \mathbb{N}$.

Proof. As noted above, since the machine satisfies the conditions of Prop. 1, the entropy rate is given by Eq. (8) and the difference $h_\mu(t+1) - h_\mu$ is given by Eq. (14). Therefore, applying Claims 3 and 5 we may bound the quantity $h_\mu(t+1) - h_\mu$ as follows:

$$\begin{aligned} h_\mu(t+1) - h_\mu &= \sum_{w \in \mathcal{L}_t} \mathbb{P}(w)(h_w - \tilde{h}_w) \\ &\geq \sum_{w \in W_t} \mathbb{P}(w)(h_w - \tilde{h}_w) \\ &\geq \mathbb{P}(W_t) \cdot \frac{1}{300} \\ &\geq \frac{C}{3600t} . \end{aligned}$$

□

With the above decay on $h_\mu(t)$ established we easily see the Branching Copy Process must have infinite excess entropy.

Proposition 3. *The excess entropy \mathbf{E} for the BC Process is infinite.*

Proof. $\mathbf{E} = \sum_{t=1}^{\infty} (h_\mu(t) - h_\mu)$. By Claim 6, this sum must diverge. □

IV. CONCLUSION

Any stationary, finite-alphabet process may be represented as an invariant HMM with an uncountable state set. Thus, there exist invariant HMMs with uncountable state sets capable of generating infinitary processes over finite alphabets. It is impossible, however, to have a finite-state invariant HMM that generates an infinitary process. The excess entropy \mathbf{E} is always bounded by the entropy in the stationary distribution $H[\pi]$, which is finite for any finite-state HMM. Countable-state HMMs are intermediate between the finite and uncountable cases, and it was previously unknown whether infinite excess entropy was possible in this case. We have demonstrated that it is indeed possible, by giving two explicit constructions of finite-alphabet infinitary processes generated by invariant HMMs with countable state sets.

The second example, the Branching Copy Process, is also ergodic—a strong restriction. It is a priori quite plausible that infinite \mathbf{E} might only occur in the countable-state case for nonergodic processes. Moreover, both HMMs we constructed are unifilar, so the ϵ -machines [9, 13] of the processes have countable state sets as well. Again, unifilarity is a strong restriction to impose, and it is a priori conceivable that infinite \mathbf{E} might only occur in the countable-state case for nonunifilar HMMs. Our examples have shown, though, that infinite \mathbf{E} is possible for countable-state HMMs, even if one requires both ergodicity and unifilarity.

Appendix A

We prove Prop. 1 from Sec. IIB, which states that the entropy rate of any countable-state, exact, unifilar HMM is given by the standard formula:

$$h_\mu = H[X_0|S_0] = \sum_{\sigma \in \mathcal{S}} \pi_\sigma h_\sigma . \quad (\text{A1})$$

Proof. Let $\mathcal{L}_t = \{w : |w| = t, \mathbb{P}(w) > 0\}$ be the set of length t words in the process language \mathcal{L} , and let $\phi(w)$ be the conditional state distribution induced by a word $w \in \mathcal{L}$: i.e., $\phi(w)_\sigma = \mathbb{P}(S_t = \sigma | \vec{X}^t = w)$. Furthermore, let $\tilde{h}_w = \sum_\sigma \phi(w)_\sigma h_\sigma$ be the $\phi(w)$ -weighted average entropy in the next symbol given knowledge of the current state σ . And let $h_w = H[X_t | \vec{X}^t = w] = H[X_0 | S_0 \sim \phi(w)]$ be the entropy in the next symbol given the word w . Note that:

1. $h_\mu(t+1) = H[X_t | \vec{X}^t] = \sum_{w \in \mathcal{L}_t} \mathbb{P}(w) h_w$, and
2. $\sum_\sigma \pi_\sigma h_\sigma = \sum_\sigma (\sum_{w \in \mathcal{L}_t} \mathbb{P}(w) \phi(w)_\sigma) h_\sigma = \sum_{w \in \mathcal{L}_t} \mathbb{P}(w) (\sum_\sigma \phi(w)_\sigma h_\sigma) = \sum_{w \in \mathcal{L}_t} \mathbb{P}(w) \tilde{h}_w$.

Since we know $h_\mu(t)$ limits to h_μ , it suffices to show that:

$$\lim_{t \rightarrow \infty} \sum_{w \in \mathcal{L}_t} \mathbb{P}(w)(h_w - \tilde{h}_w) = 0 . \quad (\text{A2})$$

By concavity of the entropy function, $h_w - \tilde{h}_w \geq 0$ for any w . However, for a *synchronizing word* $w = w_1 \dots w_t$ with $H[S_t | \vec{X}^t = w] = 0$, $h_w - \tilde{h}_w$ is always 0, since the distribution $\phi(w)$ is concentrated only on a single state. Furthermore, for any w , $h_w - \tilde{h}_w \leq h_w \leq \log |\mathcal{X}|$. Thus:

$$\sum_{w \in \mathcal{L}_t} \mathbb{P}(w)(h_w - \tilde{h}_w) \leq \log |\mathcal{X}| \cdot \mathbb{P}(NS_t) , \quad (\text{A3})$$

where NS_t is the set of length- t words that are nonsynchronizing and $\mathbb{P}(NS_t)$ is the combined probability of all words in this set. Since the HMM is exact, we know that for a.e. infinite future \vec{x} an observer will synchronize exactly at some finite time $t = t(\vec{x})$. And, since it is unifilar, the observer will remain synchronized for all $t' \geq t$. It follows that $\mathbb{P}(NS_t)$ must be monotonically decreasing and limit to 0:

$$\lim_{t \rightarrow \infty} \mathbb{P}(NS_t) = 0 . \quad (\text{A4})$$

Combining Eq. (A3) with Eq. (A4) shows that Eq. (A2) does in fact hold, which completes the proof. \square

Appendix B

We prove the following proposition for the entropy rate of countable-state HMMs.

Proposition 4. *Let M be a countable-state HMM and let $\mathcal{P} = (X_t)$ be the process generated by M . If \mathcal{P} does not consist entirely of periodic sequences, then its entropy rate $h_\mu \mathcal{P}$ is strictly positive.*

Proof. For any countable-state HMM M , the future output sequence and past output sequence are conditionally independent given the current state. Thus, for all $t \in \mathbb{N}$, $H[X_t | \vec{X}^t, S_t] = H[X_t | S_t]$. Also, by stationarity $H[X_t | S_t] = H[X_0 | S_0] = \sum_{\sigma} \pi_{\sigma} h_{\sigma}$, for all t . Combining these facts shows that entropy rate is always bounded below by the standard unifilar formula of Eq. (8):

$$\begin{aligned} h_\mu &= \lim_{t \rightarrow \infty} H[X_t | \vec{X}^t] \\ &\geq \lim_{t \rightarrow \infty} H[X_t | S_t, \vec{X}^t] \\ &= \lim_{t \rightarrow \infty} H[X_t | S_t] \\ &= \sum_{\sigma \in \mathcal{S}} \pi_{\sigma} h_{\sigma} . \end{aligned} \quad (\text{B1})$$

Therefore, the entropy rate is positive if $h_{\sigma} > 0$ for any state σ with nonzero probability π_{σ} or, equivalently, if there are at least two outgoing edges in the associated graph from state σ .

Now, assume there is no such state. Consider the restricted state set $\tilde{\mathcal{S}}$ consisting of states σ with positive probability ($\pi_{\sigma} > 0$) and the restricted graph \tilde{G} associated to this state set. Clearly, the HMM \tilde{M} defined by this graph with stationary distribution π generates the same process \mathcal{P} as the original HMM. And, it is also easily seen that in order to keep the distribution π stationary, the graph \tilde{G} must consist entirely of disjoint strongly connected components. That is, each connected component of \tilde{G} must be strongly connected. Take any strongly connected component C_i in \tilde{G} . Since each state σ in C_i has only a single outgoing edge and C_i is strongly connected, it follows that C_i must be a deterministic loop of some finite length l_i . Since this holds for each strongly connected component C_i in \tilde{G} and the HMM \tilde{M} is always run from one of the C_i s, it follows that all sequences $\vec{x} = \dots x_{-1} x_0 x_1 \dots$ generated by \tilde{M} are periodic. Or, equivalently, all sequences generated by M are periodic. \square

[1] A. del Junco and M. Rahe. Finitary codings and weak Bernoulli partitions. *Proc. AMS*, 75:259, 1979.

- [2] J. P. Crutchfield and N. H. Packard. Symbolic dynamics of one-dimensional maps: Entropies, finite precision, and noise. *Intl. J. Theo. Phys.*, 21:433, 1982.
- [3] P. Grassberger. Toward a quantitative theory of self-generated complexity. *Intl. J. Theo. Phys.*, 25:907, 1986.
- [4] K. Lindgren and M.G. Norhdal. Complexity measures and cellular automata. *Complex Systems*, 2(4), 1988.
- [5] W. Bialek, I. Nemenman, and N. Tishby. Predictability, complexity, and learning. *Neural Computation*, 13(11):2409–2463, 2001.
- [6] L. Debowski. Excess entropy in natural language: Present state and perspectives. *Chaos*, 21(3):037105, 2011.
- [7] J. P. Crutchfield and D. P. Feldman. Regularities unseen, randomness observed: Levels of entropy convergence. *CHAOS*, 13(1):25–54, 2003.
- [8] W. Ebeling. Prediction and entropy of nonlinear dynamical systems and symbolic sequences with Iro. *Physica D: Nonlinear Phenomena*, 109(1-2):42 – 52, 1997.
- [9] W. Löhr. *Models of Discrete Time Stochastic Processes and Associated Complexity Measures*. PhD thesis, Max Planck Institute for Mathematics in the Sciences, Leipzig, 2010.
- [10] N. Travers and J. P. Crutchfield. Asymptotic synchronization for finite-state sources. *J. Stat. Phys.*, 2011. (in press).
- [11] N. Travers and J. P. Crutchfield. Exact synchronization for finite-state sources. *J. Stat. Phys.*, 2011. (in press).
- [12] L. Debowski. A general definition of conditional information and its application to ergodic decomposition. *Stat. Probab. Lett.*, 79(9):1260–1268, 2009.
- [13] J. P. Crutchfield and K. Young. Inferring statistical complexity. *Phys. Rev. Lett.*, 63:105–108, 1989.