Max-Planck-Institut für Mathematik in den Naturwissenschaften Leipzig

Non-generating Partitions of Unimodular Maps

by

Oliver Pfante and Jürgen Jost

Preprint no.: 3 2015

Non-generating Partitions of Unimodular Maps

Oliver Pfante[1] and Jürgen Jost[1,2]
[1)] *Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, 04103 Leipzig, Germany*[a)]
[2)] *Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA*

(Dated: 26 January 2015)

We quantify the relationship between the dynamics of a time-discrete dynamical system, driven by a unimodular map $T : [0,1] \to [0,1]$ on the unit interval and its iterations T^m, and the induced dynamics at a symbolic level in information theoretical terms. The symbolic dynamics are obtained by a threshold crossing technique. A binary string s of length m is obtained by choosing a partition point $\alpha \in [0,1]$ and putting $s^i = 1$ or 0 depending on whether $T^i(x)$ is larger or smaller than α.
First, we investigate how the choice of the partition point α determines which symbolic sequences are forbidden, that is, cannot occur in the symbolic dynamics. The periodic points of T mark the choices of α where the set of those forbidden sequences changes. Second, we interpret the original dynamics and the symbolic ones as different levels of a complex system. This allows us to quantitatively evaluate a closure measure that has been proposed for identifying emergent macro-levels of a dynamical system. In particular, we see that this measure necessarily has its local minima at those choices of α where also the set of forbidden sequences changes. Third, we study the limit case of infinite binary strings and interpret them as a series of coin tosses. These coin tosses are not i.i.d. but exhibit memory effects which depend on α and can be quantified in terms of the closure measure.

Keywords: Markov Chains, Information Theory, Chaotic Systems, Aggregation, Time-Discrete Dynamical Systems

In order to deal with chaotic dynamical systems induced by unimodular maps one often converts an analog time series into a binary sequence by introducing a threshold. The choice of the maximum of the unimodular map provides a generating partition, i.e., the one with maximal topological entropy, and the powerful techniques of kneading theory are available. Nearly nothing is known if the choice of the threshold differs from the maximum. The induced partition have been simply considered as "misplaced".
Our approach is entirely different because we do not consider the symbolic sequence merely as a vehicle to get information about the original, analog time series where choices of the threshold different from the maximum are only considered as those which lead to unnecessary information losses. We think of the symbolic dynamics as new, macroscopic ones in their own right which are derived from a microscopic time series – the initial analog signal – by means of an aggregation which assigns the same symbol to all elements in a cell of the partition.
Due to this new point of view choices of the threshold different from the maximum lead to interesting macroscopic dynamics in their own right. They exhibit a rich structure which can be captured by information theoretical measures proposed by the authors and others.

[a)] Electronic mail: pfante@mis.mpg.de

I. INTRODUCTION

Studying one-dimensional dynamical systems, that is, time-discrete $x_n \to x_{n+1} = T(x_n)$ dynamics with a map $T : [0,1] \to [0,1]$, has a long tradition in mathematics –[8] is the standard textbook or[24] for a short exposure. Among one-dimensional maps, the unimodular maps are the best studied ones. A map $T : [0,1] \to [0,1]$ is called unimodular iff it is continuous, $T(0) = T(1) = 0$ and has a unique maximum c s.t. that T is monotonically increasing on $[0,c]$ and decreasing on $[c,1]$.

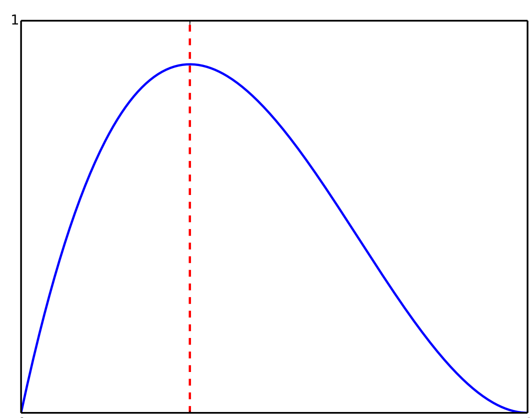

FIG. 1. The graph of a unimodular map on the unit interval $[0,1]$.

The focus on unimodular maps is due to the powerful results provided by kneading theory, see for instance[9] for a short exposure, to analyse the properties of time-discrete, one-dimensional dynamical systems driven by those maps. Essentially, kneading theory suggests an investigation of a symbol dynamics which are derived from a threshold crossing technique, i.e., one simply replaces the real valued data with symbolic data: 1 if the value is larger than the threshold α and 0 otherwise. This provides a map $\phi_\alpha^{\mathbb{N}} : [0,1] \to \{0,1\}^{\mathbb{N}}; \phi_\alpha^{\mathbb{N}}(x) = (s_0, s_1, s_2, \ldots)$ from the unit interval into the set of binary sequences of infinite length with $s_n = 1$ if $T^n(x) > \alpha$ and 0 otherwise. In order to attain a one-to-one correspondence between the initial real valued dynamics and the symbolic ones, research was in particular focused on the problem finding *generating partitions*, i.e., partitions whose corresponding topological entropy is maximal. For a unimodular one-dimensional map, the threshold encoding provides a generating partition if the maximum c of the map is used as the threshold[5]. The infinite symbolic sequence $\phi_c^{\mathbb{N}}(c)$ of c is called the kneading sequence of the map T. It allows for a determination of the support $\phi_c^{\mathbb{N}}([0,1]) \subset \{0,1\}^{\mathbb{N}}$ of the encoding. These supports may exhibit a rather difficult structure as shown by E. Friedman[11] who proves that the supports of a one-parameter family of logistic maps with different heights are even no longer Turing computable.

In informal terms, a generating partitions allows for an entire reconstruction of the real valued, initial dynamics from the symbolic ones. But in more than one dimension generating partitions are difficult to find (see e.g.[4]) and the threshold crossing technique turned out to be a widely used method to detect partitions whose corresponding symbolic sequence allows for all possible strings of a certain length and the occurrence of all these strings is equally distributed. These partitions to which one refers in the literature often as *statistical partitions* were introduced in[21] and[22] to get best possible approximations of a generating partition. These methods became popular in the physics literature and were applied to the quantitative analysis of the death rate of patients with coronary disease[13], describing non-linear dynamics in an internal combustion engine[7], and in cognitive sciences to study movement control[10].

Despite these achievements attained by the threshold technique in the case of unimodular maps, the symbol dynamics derived from thresholds α which are different from the maximum c were always considered as "misplaced"[2] or as misrepresentations of the original dynamics. There is a thorough analysis in[2] to capture quantitatively the information loss resulting from choosing a non-generating partition and to which extent the derived symbol dynamics reflects the dynamics of the original system. The authors of[2] voted for the tent map as a test ground to perform their explicit computations. As a reaction to[2] the author of[14] presented enhanced methods to improve the quality of the non-generating partitions obtained from the threshold technique.

Here, we develop a completely different point of view on the problem of non-generating partitions. We do not try to find generating partitions of unimodular dynamical system, or to improve already existing "imperfect" ones. Rather than trying to reconstruct the original real valued dynamics from the symbolic ones, we consider the symbol dynamic in its own right as a macroscopic level derived from a microscopic one whose time-discrete dynamics are determined by iterating the underlying dynamical system. From this point of view, the threshold crossing technique takes over the role of a coarse-graining which links the microscopic level with the macroscopic, symbolic one, whereas in the previous papers this method was more or less considered as an encoding. The choice of a generating partition is then justified by the fact that it yields the symbolic dynamics of highest entropy, that is, the most random looking symbolic dynamics. Therefore, observing it provides most information about the initial value of the microscopic dynamics. In contrast, for non-generating partitions, dependencies show up in the sense that certain symbolic sequences do not occur, i.e., are “forbidden”[1]. In particular, such symbolic dynamics then indicate that the underlying microscopic process is not completely random itself. Our considerations will elucidate this phenomenon.

In statistical mechanics coarse-graining is one method to overcome the contradiction between the reversibility of the Hamiltonian dynamics and the second law of thermodynamics. For instance, according to Poincaré's recurrence theorem a trajectory in the phase space will return arbitrarily close to its initial state after sufficiently long time. Hence, any smooth state function including the entropy cannot be a strictly monotonic function in time, as proposed by the second law of thermodynamics. But if one introduces a certain kind of coarse-graining of the phase space, which can be considered to be a consequence of the unavoidable inaccuracy in the measurement, an increase of the entropy of the coarse-grained process can be observed, see[3] for more details. Beside their importance in the foundations of statistical physics, coarse-graining techniques play also an important role in biology, chemistry and material science to bridge the gap between the temporal and spatial scales that can be achieved by atomistic molecular dynamics simulations and the scales that are relevant for the macroscopic properties of the studied systems, see[18] for a recent review.

But if one considers the threshold crossing method as a coarse-graining the question arises how to find emergent levels whose dynamics allow for a self-sufficient description.

In[17] we tackled this problem by proposing and investigating several information theoretical measures to detect those emergent levels. More precisely, we considered a dynamical system $T : X \to X$ on a probability space X with measure μ where T can be a measurable map, a Markov kernel, etc. Suppose we have an operator $\phi_\alpha : X \to \hat{X}$ – for instance a coarse-graining, aggregation, averaging, etc. – of the lower, microscopic level X onto an upper, macroscopic level $\hat{X}$. As the dynamics evolve on the lower level, induced dynamics can be observed on the upper state space $\hat{X}$. We say that the upper level is *closed* if it can be also represented by a dynamical system: there is a measurable map, a Markov kernel, etc. $\hat{T} : \hat{X} \to \hat{X}$ such that $\phi_\alpha \circ T = \hat{T} \circ \phi_\alpha$. $\alpha \geq 0$

$$\begin{tikzcd} \hat{X} \arrow[r, dashed, "\hat{T}"] & \hat{X} \\ X \arrow[u, "\phi_\alpha"] \arrow[r, "T"] & X \arrow[u, "\phi_\alpha"] \end{tikzcd}$$

FIG. 2. Basic setup of multilevel dynamical system.

may stand for a scalar parameter that distinguishes the different *scales* where the coarse-graining, etc. is carried out. We characterize a relevant scale as one where special structural or dynamical regularities can be detected. Closure measures provide a link between the two concepts of "levels" and "scales" because they should allow us to identify emergent levels, i.e., scales for which a(n approximately) closed description exists, by means of quantifying to which extent the induced system deviates from being closed. The following closure measures have been proposed so far: in[20] Shalizi and Moore, and in[12] Görnerup and Jacobi proposed Markovianity of the upper process $s_t \to s_{t+1}$ as a property of an emergent level; in his PhD thesis Shalizi[19] proposed to measure the "predictive efficiency" to identify emergent levels. Here, we shall evaluate the measure in[17] which tests for *informational closure*. We called the upper process to be informationally closed if there is no information flow from the lower to the higher level. In that case, knowledge of the microstate will not improve predictions of the macrostate, i.e., for $s_t = \phi_\alpha(x_t)$ we have

$$I(s_{t+1} : x_t | s_t) = H(s_{t+1}|s_t) - H(s_{t+1}|s_t, x_t) = 0 \,, \quad \text{(I.1)}$$

where I denotes the conditional mutual information, and H the entropy.

The entropy of a random variable $Y : X \to \mathbb{R}$ on a probability space X with measure μ is defined by

$$H(Y) = -\sum_y p(y) \log(p(y))$$

where $p(y) = \mu(Y = y)$ denotes the distribution on $\mathbb{R}$ induced by Y – the probability mass function of Y. We use logarithms with base 2 so that the entropy will be measured in bits. The entropy is a measure of the average uncertainty in the random variable.

The conditional entropy $H(Z|Y)$ for two random variables Z and Y with conditional distribution $p(z|y)$ is defined as

$$H(Z|Y) = -\sum_y p(y) \sum_z p(z|y) \log(p(z|y)) \,,$$

which is the average uncertainty of the random variable Z conditional on the knowledge of the other random variable Y.

The reduction in uncertainty due to another random variable is called the mutual information

$$I(Z : Y) = H(Z) - H(Z|Y) \,.$$

The mutual information $I(Z : Y)$ is a measure of the dependence between the two random variables. It is symmetric in Z and Y and always non negative and is equal to zero if and only if Z and Y are independent, see[6].

In[16] we have already studied the closure measures of[17] for a concrete example: $X = [0, 1]$ is the unit interval, $\mu = \lambda$ the Lebesgue measure, and $T : [0, 1] \to [0, 1]$ is the full tent map, that is, $T(x) = 2x$ if $x \in [0, 1/2]$ and $T(x) = 2x - 2$ else. For an integer $m \geq 0$ we got a one-parameter family of *m-th order coarse-grainings*

$$\begin{aligned} \phi_\alpha^m : [0,1] &\to \{0,1\}^{m+1} \\ x &\mapsto (s_m, \ldots, s_0) \\ \text{with } s_k &= \begin{cases} 1 & \text{if } T^k(x) > \alpha \\ 0 & \text{else} \end{cases} \,, \end{aligned} \quad \text{(I.2)}$$

with $\alpha \in [0, 1]$. This concept of a refinement of a given coarse-graining – in the present case the one derived from the partition threshold technique – is crucial in[16] and in the present paper because it links the concept of relevant scales with the one that certain sequences of length $m + 1$ do not occur any longer. That is, the support $\phi_\alpha^m([0, 1])$ of the macro-dynamics is different from $\{0, 1\}^{m+1}$, the *extended state space*. Beside the results on the closure measures introduced in[17], there were also beautiful findings on the chaotic dynamics driven by the tent map itself: we found all Markovian partitions with a full support, even Markovian symbol dynamics whose underlying partition is not Markovian at all, and conditions on the thresholds α s.t. all sequences of length $m + 1$ occur, i.e., we could determine a subset I of $[0, 1]$ s.t. for all $\alpha \in I$ we have $\phi_\alpha^m = \{0, 1\}^{m+1}$.

Even though the results in[16] justified the choice of the tent map as a nice toy model and test ground for our closure measures they were all restricted to the case when the one-dimensional map T is the full tent map. We begin the present paper with abandoning this restriction and showing that most of the results in[16] hold even true in the case of arbitrary unimodular maps. We then move on and discover a tight relationship between the occurrence of forbidden sequences, i.e., the fact that $\phi_\alpha^m \neq \{0, 1\}^{m+1}$, and the periodicity of the threshold α w.r.t. the unimodular map T. Furthermore, we provide a link between these results and the local minima of one of the information theoretical measures defined in[17]: the measure Eq. (I.1) quantifying the informational flow. Finally, we apply these insights to derive results on the infinite symbol dynamics, that is, $m = \infty$, as well.

[17] dealt with different notions of closure and their interdependencies in general, and[16] with the investigation of these measures on a toy model where the microdynamics is defined by the tent map $T : [0,1] \to [0,1]$. In this paper we consider unimodular maps T in general and the properties of their symbolic sequences derived from the partition threshold technique. We investigate the image of the m-th order coarse grainings $\phi^m_\alpha([0,1])$ of Eq. (I.2), which is the set of all sequences of length $m+1$ which may occur and which is a subset of the extended state space $\{0,1\}^{m+1}$. More precisely, we prove that all points α where a shift of the support $\phi^m_\alpha([0,1])$ occur – i.e., for $\alpha_0 < \alpha < \alpha_1$ we have $\phi_{\alpha_0}([0,1]) \neq \phi_{\alpha_1}([0,1])$ – are periodic points of the map T, local maxima, or local minima of the maps T^k, for $k = 1,\dots,m$.
In the third section, we relate these results to the measure Eq. (I.1) testing for the informational closure. We prove that the local minima of the informational flow occur at points α where the support $\phi^m_\alpha([0,1])$ of the corresponding symbolic sequences of length $m+1$ shifts, and that such a shift is a necessary but not a sufficient condition for the occurrence of a local minimum. Initially, the proof is done only for coarse-grainings which provide a piecewise linear dependency of the distribution induced on the extended state space $\{0,1\}^{m+1}$. Let

$$A(s,\alpha) = \{x \in [0,1] : \phi_\alpha \circ T^k(x) = s_k;\, k = 0,\dots,m\} \setminus \mathcal{T}^m(\alpha) \quad \text{(I.3)}$$

be the support of a symbolic sequence $s = (s_m, s_{m-1},\dots,s_0) \in \{0,1\}^{m+1}$ of length $m+1$ for the coarse-graining ϕ^m_α. Every sequence $s \in \{0,1\}^{m+1}$ provides a non-negative function $\alpha \mapsto p(s,\alpha) = \lambda(A(s,\alpha))$ on $[0,1]$ where λ denotes the uniform Lebesgue measure on the unit interval. By piecewise linear we mean that for all $s \in \{0,1\}$ the mappings $\alpha \mapsto p(s,\alpha)$ are piecewise linear on $[0,1]$. The definition of these non-negative functions has a dual aspect: if one keeps the threshold parameter $\alpha \in [0,1]$ fixed, then the mapping $s \mapsto p(s,\alpha)$ defines a distribution on the extended state space $\{0,1\}^{m+1}$. This duality is finally one of two reasons why the occurrence of forbidden sequences is so closely related to the minima of the informational flow. The second one is the close link between the information flow $I(s^m_{n+1} : x_n|s^m_n)$ of an $m-th$ order coarse-graining ϕ^m_α, with $s^m_n = (s_{m+n}, s_{m+n-1},\dots,s_n) = \phi^m_\alpha(x)$, and the entropy rate $h(T,\alpha)$ of the unimodular map for the partition $\{[0,\alpha],(\alpha,1]\}$ of the unit interval induced by the threshold α – see[23] for a precise definition. More precisely, we prove that $I(s^m_{n+1} : x_n|s^m_n) = H(s_0|s_1,\dots,s_{m+1})$, i.e., the information flow of the m-th order coarse-graining is the m-th order proxy of the entropy rate $h(T,\alpha)$. Since the proof on the characterization of the local minima of the informational flow makes only use of the concept of entropy, the results generalise to one-dimensional dynamical systems which are conjugate to ones whose coarse-graining provide piecewise linear functions $\alpha \mapsto p(s,\alpha)$ for all sequences $s \in \{0,1\}^{m+1}$. It turns out that due to this invariance under conjugacy all well known chaotic maps on the unit interval are covered by this result: the tent map, the logistic map, and even the non-unimodular Bernoulli shift.
Finally, we unify the assembled results on m-th order coarse-grainings to interpret the information theoretical behaviour of the infinite symbolic sequence $s_{\mathbb{N}} = (s_0, s_1, s_2,\dots) = \phi^{\mathbb{N}}_\alpha(x)$ discussed at the beginning of the introduction where $s_k = 0$ iff $T^k(x) \leq \alpha$. We interpret this symbolic sequence as a coin toss with a certain memory. It turns out that this memory is given by the mutual information $I(s_0 : s_1, s_2,\dots)$ between the initial state s_0 and its future trajectory $s_{\mathbb{N}_{\geq 1}} = (s_1, s_2, s_3,\dots)$. Furthermore, we have $I(s_0 : s_1, s_2,\dots) = H(s_0) - h(T,\alpha)$, that is, the mutual information is the difference between the entropy of the single symbol space and the entropy rate which is – as we have seen in the previous section – in the limit identical with the informational flow from the micro to the macro level $\{0,1\}^{\mathbb{N}}$. We conclude this section by numerics performed in the case that the unimodular map T is the tent map. The numerical results indicate that for the threshold $\alpha = 2/3$ the memory of the sequences $s_{\mathbb{N}}$ turns out to be maximal. $\alpha = 2/3$ is the unique periodic point of the tent map T with period length 1 and the induced single step macro-dynamics $s_0 = \phi^0_{2/3}(x) \to s_{n+1} = \phi^0_{2/3} \circ T(x)$, with the notation of Eq. (I.2), is Markovian and considered as a right candidate for an emergent level – see[17].
Most of the technical proofs are relegated to the appendix, except for those that are needed to follow the main line of reasoning in the text.

II. FORBIDDEN SEQUENCES

We consider a continous unimodal map T on the unit interval $[0,1]$. That is, $T(0) = T(1) = 0$ and there is a unique maximum at $c \in [0,1]$ s.t. T is strictly increasing on $[0,c]$ and strictly decreasing on $[c,1]$. The space unit interval is a measurable space endowed with the Borel σ-algebra and the Lebesgue measure λ.
Every choice of a partition threshold $\alpha \in [0,1]$ induces a coarse-graining $\phi_\alpha = \chi_{(\alpha,1]}$, the characteristic function on the interval $(\alpha,1]$, of the dynamics $T : [0,1] \to [0,1]$. From the sequence $x_n = T^n(x)$, for an initial value $x \in X$, one obtains derived symbol dynamics $s_n = \phi_\alpha(x_n) \in \{0,1\}$, and the probability of finding s_n in the state 0 or 1 is given by the probability that x_n lies in the interval $[0,\alpha]$ or $(\alpha,1]$, respectively. One can go further and compute the symbol dynamics derived from more than one consecutive time step, i.e., one can consider the dynamics of the extended state $(s_{n+m}, s_{n+m-1},\dots,s_n)$ and its support, that is, all symbol sequences $s = (s_m, s_{m-1},\dots,s_0) \in \{0,1\}^{m+1}$ of length $m+1$ s.t. $\lambda(A(s,\alpha)) > 0$ with

$$A(s,\alpha) = \{x \in [0,1] : \phi_\alpha \circ T^k(x) = s_k;\, k = 0,\dots,m\} \setminus \mathcal{T}^m(\alpha) \quad \text{(II.1)}$$

where the set $\mathcal{T}^m(\alpha)$ is defined as follows:

Definition II.1. Let $\alpha \in [0,1]$ and $m \in \mathbb{N}_0$. We define

$$\mathcal{T}^m(\alpha) = \bigcup_{k=0}^{m} T^{-k}(\alpha) \cup \{0,1\}\,. \tag{II.2}$$

$\mathcal{T}^m(\alpha)$ is the union of all preimages of α under the iterated maps T^k, with $k = 0, \ldots, m$, including $\{0,1\}$, where $T^0 = \mathrm{id}_{[0,1]}$ denotes the identity map on $[0,1]$.

The following lemma gives an impression of the form of the supports Eq. (II.1).

Lemma II.1. *The sets $A(s,\alpha)$ for $s \in \{0,1\}^{m+1}$ are a (possibly empty) union of intervals whose boundary points are adjacent elements of $\mathcal{T}^m(\alpha)$.*

Definition II.2. We define the map

$$\begin{aligned} p : \{0,1\}^{m+1} \times [0,1] &\to [0,1] \\ (s,\alpha) &\mapsto \lambda(A(s,\alpha))\,. \end{aligned}$$

$p(\cdot,\alpha)$ is a probability measure on $\{0,1\}^{m+1}$ for all $\alpha \in [0,1]$, whereas $p(s,\cdot)$ defines a non-negative function on $[0,1]$ for all $s \in \{0,1\}^{m+1}$ which is even continous.

Lemma II.2. *The map $\alpha \mapsto p(s,\alpha)$ is continous for all $s \in \{0,1\}^{m+1}$.*

For $\alpha, \beta \in [0,1]$ we define $\alpha \sim \beta$ iff $p(s,\alpha) > 0$ for a sequence $s \in \{0,1\}^m$ if and only if $p(s,\beta) > 0$. One checks immedately that $\sim$ defines an equivalence relation on $[0,1]$. We write $[\alpha]$ for the equivalence class of α. Furthermore, in the sequel, for $U \subset [0,1]$ let $\mathring{U}$ denote the largest open subset of U.

Definition II.3. For all $m \geq 1$ and $\alpha \in [0,1]$ let $\mathcal{S}^m$ denote the set containing all periodic points T with period less or equal to m, $\{0,1\}$ and all local maxima and minima of T^k for $k = 1, \ldots, m$.

The following theorem provides a characterization of the equivalence classes $\{[\alpha] : \alpha \in [0,1]\}$.

Theorem II.3. *Let $\alpha \in [0,1]$. The set $[\mathring{\alpha}]$ consists of a union of intervals whose boundary points are in $\mathcal{S}^m$.*

From this theorem, we can easily derive the main result of[16] where it had required a much longer proof.

Corollary II.4. *Let $T : [0,1] \to [0,1]$ be the tent map, i.e., $T(x) = 2x$ if $x \in [0,1/2]$ and $T(x) = 2 - 2x$ else. Then for*

$$\alpha \in \left(\frac{2^{m-1}}{2^m + 1}, \frac{2^{m-1}}{2^m - 1} \right).$$

all sequences of length $m+1$ occur, i.e., we have $p(s,\alpha) > 0$ for all $s \in \{0,1\}^{m+1}$.

Proof. Since the choice $\alpha = 1/2$ provides the generating partition, all sequences of length $m+1$ occur. Hence, the equivalence class $[1/2]$ contains all $\alpha \in [0,1]$ where all sequences of length $m+1$ occur. One checks immediately that the points $2^{m-1}/2^m + 1$ and $2^{m-1}/2^m - 1$ are not only m-periodic points but also that there are no further periodic points with period less or equal to m contained in $(2^{m-1}/2^m + 1, 2^{m-1}/2^m - 1)$. Combining this with theorem II.3 yields the corollary. □

III. INFORMATIONAL FLOW

Since the results of this section hold true in a more general context than that of unimodular maps on unit intervals and their binary encoding, we formulate them beyond the scope of the introduction and the second section in terms of Fig. (2).

Consider a dynamical system $T : X \to X$ on a probability space X with a measure μ where T is a measurable and measure preserving map. Let us consider a family of deterministic coarse grainings $\phi_\alpha : X \to \hat{X}$ parametrized by $\alpha \in [0,1]$ and the space $\hat{X} = \{1, \ldots, r-1\}$ is assumed to be finite.

From the sequence $x_n = T^n(x)$, for an initial value $x \in X$, one obtains derived symbol dynamics $s_n = \phi_\alpha(x_n) \in \{0, 1, \ldots, r-1\}$, and the probability of finding s_n in the state i is given by the probability $p_i = \mu(\phi_\alpha^{-1}(i))$.

Given this initial coarse-graining family ϕ_α one can define a refined coarse-graining in the spirit of Eq. (I.2).

Definition III.1. Let $\phi_\alpha : X \to \hat{X}$ be a family of coarse-grainings. For every m we obtain a a family of coarse-grainings $\phi_\alpha^m : X \to \hat{}X^m$ by

$$x \mapsto (\phi_\alpha \circ T^m(x), \phi_\alpha \circ T^{m-1}(x), \ldots, \phi_\alpha(x))$$

for all $x \in X$, which we call the *m-th order coarse-graining*.

In the sequel we want to characterize the local minima of the information flow, see Eq. (I.1),

$$I(s_{n+1}^m : x_n | s_n^m)(\alpha) = H(s_{n+1}^m | s_n^m)(\alpha) - H(s_{n+1}^m | s_n^m, x_n)(\alpha)\,.$$

of the coarse-graining ϕ_α^m of order m which measures the information contained in the macrostate $s_{n+1}^m = (s_{m+n+1}, \ldots, s_{n+1}) = \phi_\alpha^m(T(x_n))$ about the microstate x_n if s_n^m is known. Since s_{n+1}^m is fully determined by x_n, the second term vanishes and we have

$$\begin{aligned} &I(s_{n+1}^m : x_n | s_n^m)(\alpha) = H(s_{n+1}^m | s_n^m)(\alpha) \\ =&H(s_{n+1}^m, s_n^m)(\alpha) - H(s_n^m)(\alpha) \\ =&H(s_{n+1}^m, s_n^m)(\alpha) - H(s_{n+1}^m)(\alpha) \\ =&H(s_{n+m+1}, \ldots, s_n)(\alpha) - H(s_{n+m}, \ldots, s_n)(\alpha) \\ =&H(s_{m+1}, \ldots, s_0)(\alpha) - H(s_{m+1}, \ldots, s_1)(\alpha) \\ =&H(s_0 | s_1, \ldots, s_{m+1})(\alpha) \end{aligned}$$

where the next-to-last identity follows from the stationarity of the process. Hence, the information flow is equal to the $m+1$th order proxy of the entropy rate of the stationary process $T : X \to X$ – see[23] for a definition. These proxies constitute a decreasing sequence. Hence, with every additional order of the coarse-graining of the extended states s_n^m, the information flow decreases.

Lemma III.1. *The information flow $I(s_{n+1}^m : x_n|s_n^m)(\alpha)$ of the extended state dynamics $s_n^m \to s_{n+1}^m$ is $H(s_0|s_1,\dots,s_{m+1})(\alpha)$, i.e., the $m+1$-th order proxy of the entropy rate $h(T,\mathcal{A})$ where $\mathcal{A} = \{\{\phi_\alpha^{-1}(i) : i = 0,\dots,r-1\}$ is the partition of the microstate X induced by the coarse-graining $\phi_\alpha : X \to \hat{X}$.*

The domains of the symbol sequences $s = (s_m,\dots,s_1,s_0)$, with $s_i \in \{0,\dots,r-1\}$ is

$$A(s,\alpha) = \{x \in X : \phi_\alpha \circ T^k(x) = s_k;\, k = 0,\dots,m\}\ ,$$

We define, as in the previous section, $p(s,\alpha) = \mu(A_a)$ for $s \in \{0,\dots,r-1\}^m$, i.e., the probability that the sequence s occurs for a given $\alpha \in [0,1]$. For $\alpha,\beta \in [0,1]$ we define the equivalence relation $\alpha \sim \beta$ iff $p(s,\alpha) > 0$ for a sequence $s \in \{0,\dots,r-1\}^m$ implies $p(s,\beta) > 0$ and vice versa.

Definition III.2. We call the coarse-graining ϕ_α linear iff for all sequences $s \in \{0,\dots,r-1\}^m$ the function $\alpha \mapsto p(s,\alpha)$ depends piecewise linearly and continuously on α.

Example III.1. Let $T : [0,1] \to [0.1]$ be a unimodular map with linear slopes and $\phi_\alpha : [0,1] \to \{0,1\}$ the coarse graining induced by the choice of a partition threshold $\alpha \in [0,1]$. Consider the support $A(s,\alpha)$ of a symbol sequence $s \in \{0,1\}^{m+1}$. Lemma II.1 provides that $A(s,\alpha) = \bigcup_n (b_n,c_n)$ is a finite union of disjoint open intervals where b_n and c_n are in $\mathcal{T}^m(\alpha)$. Since the slopes of T are linear, the preimages $T^{-m}(\alpha)$ depend also linearly on α. Therefore, $\lambda((b_n,c_n))$ depends linearly on α for all n, and so does the sum $\sum \lambda((b_n,c_n)) = \lambda(A(s,\alpha)) = p(s,\alpha)$ for all $s \in \{0,1\}^m$. Furthermore, due to lemma II.2, the functions $\alpha \mapsto p(s,\alpha)$ are continuous.

The following theorem links the local minima of the information flow, i.e., possible candidates for emergent levels induced by the m-th order coarse-graining ϕ_α^m, with the results of the previous section. It tells us that in the case of a piecewise linear coarse-graining III.2 the information flow, i.e., the $m+1$-th order proxy of the entropy rate $H(s_0|s_1,\dots,s_{m+1})(\alpha)$, see III.1, is either strictly convex in the parameter α or linear as long as it is differentiable. But the only points where the conditional entropy $\alpha \mapsto H(s_0|s_1,\dots,s_{m+1})(\alpha)$ is not differentiable are those where a shift in the support $\phi_\alpha^m(X) \subset \hat{X}^{m+1}$ of the macrostate dynamics appears because then one of the probabilities which occur in the logarithm of the conditional entropy vanishes.

Theorem III.2. *Let $\alpha \in [0,1]$ be s.t. $\mathring{[\alpha]} \neq \emptyset$ and the coarse-graining ϕ_α is assumed to be linear. The conditional entropy $\beta \mapsto H(\beta) = H(s_0|s_1,\dots,s_{m+1})(\beta)$ is differentiable in $\mathring{[\alpha]}$ and we have*

$$\begin{aligned} H'(\beta) &= -\sum p'(s)\log p(s|\hat{s}) \\ H''(\beta) &\le 0 \end{aligned}$$

with $s = (s_{m+1},\dots,s_0)$ and $\hat{s} = (s_m,\dots,s_0)$. Furthermore, there is a point $\alpha_0 \in \mathring{[\alpha]}$ with $H''(\alpha_0) = 0$ if and only if $H(\beta)$ is a linear function on $[\alpha]$.

Let $U \subset \mathbb{R}$ be a subset of the real numbers. We call an element of the set $\overline{U}\setminus\mathring{U}$, i.e., the difference of the smallest closed subset of $\mathbb{R}$ which contains U and the biggest open subset contained in U, a *boundary point* of the set U.

Corollary III.3. *Assume that the coarse-graining ϕ_α is linear. The information flow has a local minimum in α_0 if it is a boundary point or $\alpha \mapsto H(\alpha)$ is constant on $[\alpha]$.*

Proof. Let $\alpha \in [0,1]$ be a local minimum of $\beta \mapsto H(\beta)$. Assume that $\alpha \in \mathring{[\alpha]}$. The function $\beta \mapsto H(\beta)$ is differentiable in $\mathring{[\alpha]}$ and strictly concave or linear. Since there is a local minimum in $[\alpha]$, the first case cannot hold. Furthermore, we have $H'(\alpha_0) = 0$. Hence, $\beta \mapsto H(\beta)$ is not only linear on $[\alpha]$ but even constant. ☐

Lemma III.1 provides an identity between the information flow of the m-th order coarse-graining and the $m+1$-th order proxy $H(s_0|s_1,\dots,s_{m+1})(\alpha)$ of the entropy rate of the stationary, deterministic dynamical system $T : X \to X$. But this proxy is invariant under conjugacy. Recall, a dynamical system $S : Y \to Y$ on a measurable space Y with measure λ is called conjugate to the dynamical system $T : X \to X$ if there is a measurable and measure preserving map $\Phi : Y \to X$ which has a measurable inverse Φ^{-1} s.t. $\Phi \circ S = T \circ \Phi$.
The family $\phi_\alpha : X \to \hat{X}$ of coarse-grainings defines a family $\psi_\alpha : Y \to \hat{X}$ of those for the dynamical system $S : Y \to Y$ via $\psi_\alpha = \phi_\alpha \circ \Phi$ for all $\alpha \in [0,1]$. We define the m-th order coarse-graining $\psi_\alpha^m(y) = (\psi_\alpha \circ S^m(y),\dots,\psi_\alpha(y)) = (t_m,\dots,t_0)$ for $y \in Y$. One can prove[23] that ϕ_α^m and ψ_α^m induce partitions on X with the same entropy. Hence, $H(t_0|t_1,\dots,t_m)(\alpha) = H(s_0|s_1,\dots,s_m)(\alpha)$ for all m. Since the proof of corollary III.3 incorporates only computations of the conditional entropy, it holds also for systems which are only conjugate to a linear system.

Corollary III.4. *Suppose that we have stationary, deterministic dynamics $S : Y \to Y$ with a family of coarse graining $\psi_\alpha : Y \to \hat{Y}$ for $\alpha \in [0,1]$, and that there is a conjugate system $T : X \to X$ with conjugacy map Φ s.t. $\psi_\alpha \circ \Phi$ which is a linear coarse graining. Then, the information flow of the m-th order coarse graining ψ_α^m has a local minimum in α if it is a boundary point or $\alpha \mapsto H(\alpha)$ is constant on $[\alpha]$.*

If we combine these results with II.3 we obtain a nice description of the local minima of the information flow if the dynamics are induced by a unimodular map and the family of coarse-grainings comes from the partition threshold technique.

Corollary III.5. *Let $T : [0,1] \to [0,1]$ be a unimodular map conjugate to a unimodular one with linear slopes and $\phi_\alpha : [0,1] \to \{0,1\}$ the coarse-grainings obtained from the threshold technique. The local minima of the information flow of the m-th order coarse grainings ϕ_α^m are elements of the set $\mathcal{S}^m$, i.e., the set of all periodic points with period less or equal to m or local extrema of the maps T^s with $s \leq m$.*

The previous corollary provides a characterization of the local minima of the information flow if the multilevel dynamical system comes from a unimodular map and the coarse-grainings are obtained from the partition threshold technique and its m-th orders, respectively. Unfortunately, being an element of the set $\mathcal{S}^m$ is only a necessary condition for the information flow to be minimal in this point but not sufficient.

Example III.2. Let $T_c : [0,1] \to [0,1]$ be a skewed version of the full tent map, that is, we define

$$T_c(x) = \begin{cases} \frac{1}{c}x & \text{if } x \in [0,c] \\ 1 - \frac{1}{1-c}(x-c) & \text{else} \end{cases}$$

for a parameter $c \in (0,1)$. The choice $c = 1/2$ yields the tent map. We compute the information flow of the 1-st order coarse-graining, that is, the 1-st order proxy $H(s_0|s_1)(\alpha)$ of the entropy rate. The value of the joint distribution $p((s_1,s_0),\alpha)$ of the different symbol sequences $(s_1,s_0) \in \{0,1\}^2$ for different partition thresholds $\alpha \in [0,1]$ is listed in the following table.

	$\alpha \leq 1/(2-c)$	$\alpha > 1/(2-c)$
$p(0,0)$	αc	$2\alpha - 1$
$p(1,0)$	$\alpha(1-c)$	$1-\alpha$
$p(0,1)$	$\alpha(1-c)$	$1-\alpha$
$p(1,1)$	$1-\alpha(2-c)$	0

From these joint distributions one obtains the conditional information $H(s_0|s_1)(\alpha)$. Since the skewed tent maps are full, that is, $T_c(c) = 1$, the only minima and maxima of T_c^s are 0 and 1. Hence, the only possible candidate for a local minimum of the information flow is the fixed point of T_c which is $c_0 = 1/(2-c)$. One can check analytically that c_0 provides a local minimum of the information flow if and only if $c \leq 1/2 + 1/2\sqrt{1/5}$. The conditional entropy $H(s_0|s_1)(\alpha)$ whose properties were thoroughly investigated in[1] is plotted for different values of c in Fig. (3)

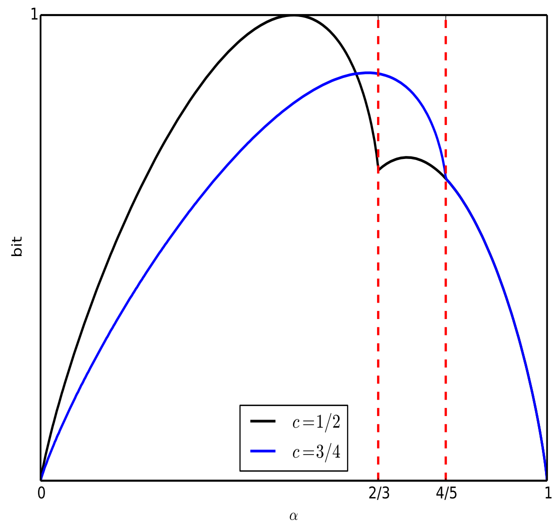

FIG. 3. The conditional entropy $H(s_0|s_1)(\alpha)$ for two different values of the parameter c. The choice $c = 1/2$ yields a local minimum of at the fix point $2/3$ whereas for $c = 3/4$ we do not have a local minimum at the fix point $4/5$.

IV. CHAOTIC COINS

From the general setting of the previous section we return to the case where the micro-dynamics are induced by a unimodular map $T : [0,1] \to [0,1]$. So far we have considered coarse-grainings obtained from symbol sequences $s \in \{0,1\}^m$ of finite length only. This section tackles symbol sequences of infinite length, that is, for a given partition threshold $\alpha \in [0,1]$ we define the coarse-graining $\phi_\alpha^{\mathbb{N}} : [0,1] \to \Sigma_2 = \{0,1\}^{\mathbb{N}}$ via

$$\begin{aligned}\phi_\alpha^{\mathbb{N}}(x) &= (s_0, s_1, \ldots, s_m, \ldots) \\ &= (\phi_\alpha(x), T \circ \phi_\alpha(x), \ldots, T^m \circ \phi_\alpha(x), \ldots).\end{aligned}$$

The dynamics on Σ_2 induced by $T : [0,1] \to [0,1]$ is the shift operator which maps $s_{\mathbb{N}} \in \{0,1\}^{\mathbb{N}}$ onto $s'_{\mathbb{N}}$ s.t. $s'_n = s_{n+1}$ for all $n \geq 0$.
In order to derive insights for this limit case of the scenaria considered in the previous sections, it is intriguing to think of an infinite symbol sequence $\phi_\alpha^{\mathbb{N}}(x)$, induced by a point $x \in [0,1]$ and a choice $\alpha \in [0,1]$, as an infinite sequence of coin tosses, where head indicates that the point x is mapped to the right of α and tail to the left of it. The results of the second subsection then mean that for certain choices of the partition threshold α some sequences of heads and tails may not appear at all. Assuming ergodicity of the unimodular map allows for estimating the frequency of 1s and 0s in such an infinite sequence.

Proposition IV.1. *Let $T : [0,1] \to [0,1]$ be ergodic. For almost all $x \in [0,1]$ the image $\phi_\alpha(x)$ is a sequence where 0s occurs with frequency α.*

Proof. From Birkhoff's ergodicity theorem it follows that

$$\frac{1}{n}\sum_{n=0}^{\infty} \phi_\alpha \circ T^n(x) = \int \phi_a(y)\, dy = 1 - \alpha$$

and the sum on the left hand side counts precisely the frequency of 1s in the image $\phi_\alpha^{\mathbb{N}}(x)$. □

The result might indicate that the infinite symbol sequence results from a coin toss with a biased coin. This actually does not hold true because the sequence $(s_n)_n$ is not independent in general.
With the notation of definition II.2, the entropy rate $h(T,\alpha)$ for the partition induced by $\alpha \in [0,1]$ is the limit

$$\lim_{m\to\infty} H(s_0|s_1,\dots,s_{m+1})(\alpha) \tag{IV.1}$$

of the conditional entropies $H(s_0|s_1,\dots,s_{m+1})(\alpha)$, that is, the uncertainty left about the the initial state s_0 if the future trajectory $s_1,\dots,s_{m+1}$ of length $m+1$ is known.
If the symbol dynamics $s_n \to s_{n+1}$ result from an i.i.d. process – like for a coin toss, the conditional entropy $H(s_0|s_1,\dots,s_{m+1})(\alpha)$ is $H(s_0)(\alpha)$ because no information about the initial state s_0 can be derived from the subsequent outcomes $s_1,\dots,s_{n-1}$. Hence, the entropy rate $h(T,\alpha)$is $H(s_0)(\alpha)$ as well. This does not hold in general as one can read off Fig. (4) where a proxy of the entropy rate $h(T,\alpha)$ and the difference $H(s_0)(\alpha) - h(T,\alpha)$ is plotted when $T : [0,1] \to [0,1]$ is the full tent map, i.e., $T(x) = 2x$ if $x \le 1/2$ and $T(x) = 2-2x$ else.
The difference $H(s_0)(\alpha) - h(T,\alpha) = I(s_0 : s_1, s_2, \dots)$ is the mutual information between the infinite future trajectory $s_1, s_2, \dots$ of the symbol sequence $s^{\mathbb{N}} = (s_0, s_1, s_2, \dots) \in \Sigma_2$ and its initial state s_0, i.e., the memory of the sequence about its original state s_0. Since this difference is greater than 0 except for $\alpha = 0, 1/2, 1$, the sequences $s^{\mathbb{N}} \in \phi_\alpha^{\mathbb{N}}([0,1])$ in the image of the symbol map have a memory in contrast to a sequence that results from simply tossing a biased coin.
From lemma III.3 we know that the information flow $I(s_{n+1}^m : x_n|s_n^m)(\alpha)$ of the m-th order coarse-graining is $H(s_0|s_1,\dots,s_{m+1})(\alpha)$, the $m+1$-st order proxy of the entropy rate $h(T,\alpha)$. Therefore, we can interpret $h(T,\alpha) = \lim_{m\to\infty} I(s_{n+1}^m : x_n|s_n^m)$ as the information flow of the coarse-graining $\phi_\alpha^{\mathbb{N}} : [0,1] \to \Sigma_2$. Hence, the memory of the sequences $s \in \phi_\alpha([0,1])$ in the macro space is the information left about the original micro dynamics $T : [0,1] \to [0,1]$.

If the map T is the full tent map, Fig. (4) indicates that at $\alpha = 2/3$ the memory of the infinite sequence $s_{\mathbb{N}} = \phi_\alpha^{\mathbb{N}}(x)$ about its microstate x is largest among all possible choices of the partition threshold $\alpha \in [0,1]$. We know from[16] that at $\alpha = 2/3$ the single step symbol dynamics $s_n \to s_{n+1}$ are Markovian. This implies $h(T, 2/3) = H(s_0|s_1)$. From the explicit computation in[16] we obtain $H(s_0) - h(T, 2/3) = H(s_0) - H(s_1|s_0) = I(s_0 : s_1) = log(3) - 4/3 \approx 0.251$ bit.

The entropy rate $\alpha \mapsto h(T,\alpha)$ plays a crucial role if we want to quantify the quality of the corresponding coarse-graining $x \mapsto \phi_\alpha^{\mathbb{N}}(x)$ because it is its information flow. An explicit computation of the entropy rate is quite hard. There are results available for the skewed tent maps of example III.2, see[15]. Even though we are not able to provide an explicit expression for the entropy $h(T,\cdot)$ depending on α, we can prove that the mapping $\alpha \mapsto h(T,\alpha)$ is continuous (see the appendix).
Furthermore, from the perspective of the second section, which discussed thoroughly the phenomenon of vanishing sequences, it is an interesting question which infinite sequences $s_{\mathbb{N}} \in \Sigma_2$ are attainable, that is, which sequences are in the image $\phi_\alpha^{\mathbb{N}}([0,1]) \subset \Sigma_2$. We speculate that the support $\phi_\alpha^{\mathbb{N}}([0,1])$ of the macro-dynamics are even not Turing computable: there are $\alpha \in [0,1]$ for which one cannot decide whether an infinite sequence $s \in \Sigma_2$ is an element of $\phi_\alpha^{\mathbb{N}}([0,1])$ or not. A similar result is already known for the logistic maps of different heights, see[11]. Even though we have not been able to prove this result we could at least prove (see the appendix for details) that the support $\phi_\alpha^{\mathbb{N}}([0,1])$ is at least uncountable when the map T is not only assumed to be ergodic but also weakly mixing – see[23] for a precise definition.

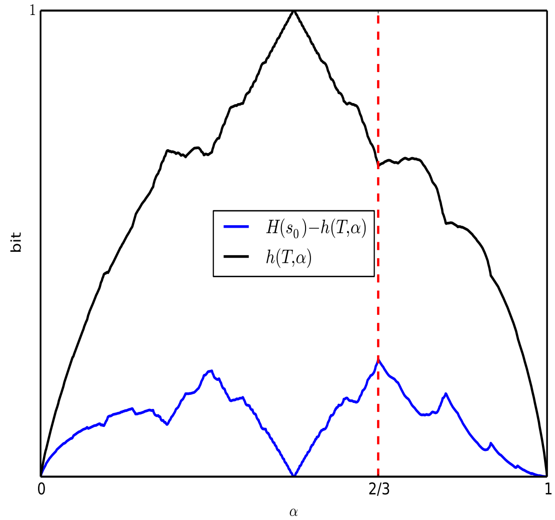

FIG. 4. The entropy rate $h(T,\alpha)$ computed as 7-th order proxy $H(s_0|s_1,\dots,s_7)(\alpha)$ and the memory $H(s_0)(\alpha) - h(T,\alpha)$ of a sequence where $T : [0,1] \to [0,1]$ is the tent map.

V. CONCLUSIONS

First, as far as we know, the present paper provides the first general treatment of non-generating partitions of unimodular maps. Previous papers – see[14],[2] – provide only numerical computations for the full tent map. Second, we uncovered a close relationship between the phenomenon of forbidden sequences and the local minima of the information flow, a closure measure proposed in[17] in order to detect emergent levels in multilevel dynamical systems. Third, we considered also infinite binary sequences which are popular among mathematicians in order to study the underlying chaotic dynamics driven by a unimodular map (see[9]). We could interpret the entropy rate as the information flow from the lower level, the time-discrete dynamics driven by the unimodular map, to the macro level, the infinite binary sequences and their dynamics defined via the shift map.

But there is still much work left. First, we believe that theorem II.3 can be improved: the equivalence classes $[\alpha]$ are not only unions of intervals whose boundaries are in the set $\mathcal{S}^m$ but consist only of a single interval whose boundaries are adjacent elements in the set $\mathcal{S}^m$. If we could prove this, we would be able to show that in the limit $m \to \infty$, that is, in the case of infinite symbolic dynamics on the space $\{0,1\}^{\mathbb{N}} = \Sigma_2$, we have $\phi_\alpha^{\mathbb{N}}([0,1]) \neq \phi_\beta^{\mathbb{N}}([0,1])$ for all $\alpha \neq \beta$. This would imply that the family of supports $(\phi_\alpha^{\mathbb{N}}([0,1]))_{\alpha\in[0,1]}$ is uncountable and due to similar reasons as in[11] no longer Turing computable for all $\alpha \in [0,1]$. Hence, the distribution induced by $\phi_\alpha^{\mathbb{N}}$ on Σ_2 is not Turing computable either and therefore neither is the entropy rate $h(T,\alpha)$ for all $\alpha \in [0,1]$. Second, these results would provide insights into the predictive efficiency[19] of the coarse-grainings $(\phi_\alpha)_{\alpha\in[0,1]}$, because they suggest that this information measure is already incomputable in the simple case that the multilevel dynamical systems is defined by a unimodular map and coarse-grainings induced by the partition threshold technique.

VI. ACKNOWLEDGEMENT

It is a pleasure for the authors to appreciate the fruitful discussions with Nils Bertschinger.

Appendix A: Proofs of the results

Proof. **Lemma II.1** For all $s = (s_m, \dots, s_0) \in \{0,1\}^{m+1}$ we have

$$A(s,\alpha) = \bigcap_{k=0}^{m} (\phi_a \circ T^k)^{-1}(s_k)\,.$$

But we have either $(\phi_a \circ T^k)^{-1}(s_k) = T^{-k}([0,\alpha])$ or $(\phi_\alpha \circ T^k)^{-1}(s_k) = T^{-k}(\alpha,1])$. Let $\gamma_0 < \dots < \gamma_r$ be the elements of the set $T^{-k}(\alpha) \cup \{0,1\}$. Then we have

$$\left(\phi_a \mathring{\circ} T^k\right)^{-1}(s_k) = \begin{cases} \bigcup_{j=0}(\gamma_{2j}, \gamma_{2j+1}) & \text{if } s_k = \epsilon \\ \bigcup_{j=0}(\gamma_{2j+1}, \gamma_{2j+2}) & \text{if } s_k = \\ & \epsilon + 1 \mod 2 \end{cases}$$

where $\epsilon \in \{0,1\}$ has to be chosen adequately. Hence, $(\phi_a \circ T^k)^{-1}(s_k)$ is a disjoint union of intervals whose boundary points are elements of $\mathcal{T}^m(\alpha)$. Since the intersection of intervals yields a (possibly empty) interval again, the same holds true for the support $A(s,\alpha)$.

Assume further that there is an interval $(a,b) \subset A(s,\alpha)$ which contains an element of $\gamma \in \mathcal{T}^m(\alpha)$. There is a $k \le m$ s.t. $\gamma \in T^{-k}(\alpha)$. This implies not only $\lambda\left(T^{-k}([0,\alpha]) \cap (a,b)\right) > 0$ but also $\lambda\left(T^{-k}((\alpha,1]) \cap (a,b)\right) > 0$. But we have $\phi_a \circ T^k(a,b) = s_k$ which is either 0 or 1 – a contradiction. □

Proof. **Lemma II.2** Let us assume the opposite. Then there is an $s = (s_m, \dots, s_0) \in \{0,1\}^{m+1}$ and an $\alpha_0 \in [0,1]$ s.t. $\alpha \mapsto p(s,\alpha)$ is not left- or right-continuous at α_0. We assume the first case, the second can be disproved analogously. Then there is a monotonically increasing sequence $(\alpha_n)_{\mathbb{N}}$ converging to α_0 s.t. $(p(s,\alpha_n))_{n\in\mathbb{N}}$ does not converge to $p(s,\alpha_0)$.

Define the monotonically increasing sequence of functions $f_n^{0,k} = \phi_{\alpha_n} \circ T^k$. From the monotone convergence theorem, we obtain the existence of a measurable function $f \le 1$ s.t. $f_n^{0,k} \to f^{0,k}$ *a.s.* and

$$\int f_n^{0,k}\, d\lambda \to \int f^{0,k}\, d\lambda \tag{A.1}$$

for all $k \ge 0$. We define $f_n^{1,k} = 1 - f_n^{0,k}$ and $f^{1,k} = 1 - f^{0,k}$ for all k and n. Define $A_n^k = (f_n^{s_k,k})^{-1}(s_k)$, $A^k = (f^{s_k,k})^{-1}(s_k)$ and $B^k = (\phi_\alpha \circ T^k)^{-1}(s_k)$. $(A_n^k)_{n\in\mathbb{N}}$ is an increasing or decreasing sequence of sets if $s_k = 0$ or $s_k = 1$, respectively. Eq. (A.1) provides $\bigcup_n A_n^k = A^k$ or $\bigcap_n A_n^k = A^k$ a.s. if $s_k = 0$ or $s_k = 1$, respectively. Assume that for all k the symmetric differences $A^k \Delta B^k$ are sets of measure zero. This implies $\lambda(A_n^k \Delta B^k) \to 0$ for $n \to \infty$. Since $\lambda\left(\bigcap_k B^k\right) = p(s,\alpha_0)$ and $\lambda\left(\bigcap_k A_n^k\right) = p(s,\alpha_n)$ this implies $p(s,\alpha_n) \to p(s,\alpha_0)$ – a contradiction. Hence, there is a k_0 s.t. $\lambda(A^{k_0} \Delta B^{k_0}) > 0$. We assume that $s_{k_0} = 0$. Then $A^{k_0} \subset B^{k_0}$ and $\phi_{\alpha_n} \circ T^{k_0} = f_n^{0,k}$ does not converge a.s. to $\phi_\alpha \circ T^{k_0}$. This implies $\lambda(T^{-k_0}(\alpha)) > 0$ which is not possible because T is a unimodal map which implies that $T^{-k_0}(\alpha)$ consists of at most 2^{k_0} points. □

Proof. **Theorem II.3** If $[\mathring{\alpha}] = \emptyset$, there is nothing to prove. Hence, we assume $[\mathring{\alpha}] \neq \emptyset$. This implies in particular that $\alpha \neq \epsilon$ for $\epsilon \in \{0,1\}$ because in both cases we get $[\epsilon] = \{\epsilon\}$. Since $[\mathring{\alpha}]$ is an open subset of $[0,1]$ there is a sequence of open intervals $(I_m = (a_m, b_m))_m$ s.t. $[\mathring{\alpha}] = \bigcup_m I_m$. Consider an interval $I_m = (a_m, b_m)$ s.t. $b_m < 1$. For all $n \in \mathbb{N}$ there is a $\beta_n \in [0,1] \setminus [\alpha]$ s.t. $0 < \beta_n - b_m < 1/n$. For every $n \in \mathbb{N}$, there exists a sequence $a_n \in \{0,1\}^m$ s.t. $p(a_n,\beta) > 0$ for all $\beta \in I_m$ and $p(a_n,\beta_n) = 0$ or $p(a_n,\beta_n) > 0$ and $p(a_n, b_m) = 0$, respectively. Since there are only finitely many sequences we can assume – choosing a proper subsequence – that all sequences a_n are the same sequence a. In both cases we obtain a sequence $(\alpha_n)_{n\in\mathbb{N}}$ converging to b_m s.t. $p(a,\alpha_n) > 0$ for all $n \in \mathbb{N}$ and $p(a, b_m) = 0$. In the first case, this sequence is taken from the interval I_m, in the second case, we simply set $\alpha_n = \beta_n$ for all n. From II.1 we know that the support $A(a,\alpha_n)$ of the sequence a is a non-empty disjoint union of intervals with endpoints contained in $\mathcal{T}^m(\alpha_n)$. Hence for all $n \geq 1$ there is a pair (r_n, s_n), with $r_n, s_n \in \{0, \ldots, m\}$ s.t. $(\gamma_n^0, \gamma_n^1) \subset A(a, \alpha_n)$ with $\gamma_n^0 \in T^{-r_n}(\alpha_n)$ and $\gamma_n^1 \in T^{-s_n}(\alpha_n)$. Since $\{0, \ldots, m\}^2$ is a finite set, we can choose a subsequence of $(\alpha_n)_n$, which we again denote with $(\alpha_n)_n$, s.t. $(r_n, s_n) = (r, s)$ for all n and fixed r and s. W.l.o.g. we assume $r \leq s$ which implies $T^{s-r}(\alpha_n) = T^s(\gamma_n^0), \alpha_n = T^s(\gamma_n^1) \in I$ for all n. But $\lambda([\gamma_n^0, \gamma_n^1]) \leq p(a, \alpha_n) \to 0$ implies $\gamma_n^1 - \gamma_n^0 \to 0$ for $n \to \infty$. Since $[0,1]$ is compact there is a convergent subsequence of $(\gamma_n^1)_n$, which we again denote by $(\gamma_n^1)_n$, with limit γ. Hence, also γ_n^0 converges to γ. This implies

$$\begin{aligned} T^{s-r}(b_m) &= \lim_{n\to\infty} T^{s-r}(\alpha_n) = \lim_{n\to\infty} T^s(\gamma_n^0) \\ &= T^s(\gamma) = \lim_{n\to\infty} T^s(\gamma_n^1) = \lim_{n\to\infty} \alpha_n = b_m \,. \end{aligned} \tag{A.2}$$

If $s > r$, this implies that b_m a periodic point with period less or equal to $s - r$. Suppose $r = s$. Then $T^s(\gamma_n^0) = T^s(\gamma_n^1) = \alpha_n$. From lemma II.1, it follows that the preimages γ_n^0, γ_n^1 are adjacent in $\mathcal{T}^m(\alpha_n)$. This implies that $[\gamma_n^0, \gamma_n^1]$ contains either a local minimum or a maximum of T^s for all n. Since there are only finitely many of them we can assume, by choosing a subsequence, that it is the same local extremum for all n. But Eq. (A.2) implies in this case that b_m is precisely this local extremum of T^s.
The proof that the lower bound a_m of the interval I_m is an element of $\mathcal{S}^m$ works analogously. □

Proof. **Theorem III.2** Suppose $\alpha_0 \in [\mathring{\alpha}]$. Then there is an $\epsilon > 0$ s.t. $p(s, \alpha_0) > 0$ for a sequence $s \in \{1, \ldots, r-1\}^{m+1}$ implies $p(s, \beta) > 0$ for all $\beta \in I = [\alpha_0 - \epsilon, \alpha_0 + \epsilon]$. Let $\mathcal{A} \subset \{0, \ldots, r-1\}^{m+1}$ be the collection of all those sequences. If $a = (a_m, \ldots, a_0)$, we define $\hat{s} = (s_{m-1}, \ldots, s_0)$. The functions $\beta \to p(s, \beta)$ and $\beta \to p(\hat{s}, \beta)$ depend linearly on β and have for I values different from zero if and only if $s \in \mathcal{A}$. Since I is compact there is a $\gamma > 0$ s.t. $p(s, \beta), p(\hat{s}, \beta) > \gamma$ for all $\beta \in I$ and $s \in \mathcal{A}$. This implies that the conditional entropy $\beta \mapsto H(\beta)$ is smooth on $\mathring{I}$ and therefore differentiable in α_0. If we take the derivative we obtain

$$\begin{aligned} H'(\beta) &= -\sum p'(s) \log p(s|\hat{s}) + \frac{p(s)}{p(s|\hat{s})} p'(s|\hat{s}) \\ &= -\sum p'(s) \log p(s|\hat{s}) + p(\hat{s}) p'(s|\hat{s}) \\ &= -\sum p'(s) \log p(s|\hat{s}) - \sum_{\hat{s}} p(\hat{s}) \sum_{i=0}^{r-1} p'(s_0 = i|\hat{s}) \,. \end{aligned}$$

But $\sum_{i=0}^{r-1} p(s_0 = i|\hat{s}) = 1$ for all β, hence $\sum_{i=0}^{r-1} p'(s_0 = i|\hat{s}) = 0$ and the second term vanishes. Computing the second derivative, we firstly observe that the linearity of $p(s)$ yields $p''(s) = 0$ and therefore

$$\begin{aligned} H''(\beta) &= -\sum p'(s) \frac{p'(s|\hat{s})}{p(s|\hat{s})} \\ &= -\sum \frac{p'(s)p(\hat{s})}{p(s)} \left(\frac{p(s)}{p(\hat{s})}\right)' \\ &= -\sum \frac{p'(s)p(\hat{s})}{p(s)} \frac{p'(s)p(\hat{s}) - p(s)p'(\hat{s})}{p^2(\hat{s})} \\ &= -\sum \frac{p'^2(s)}{p(s)} - \frac{p'(s)p'(\hat{s})}{p(\hat{s})} \\ &= -\sum_{\hat{s}} \sum_{i=0}^{r-1} \frac{p'^2(\hat{s}, s_0 = i)}{p(\hat{s}, s_0 = i)} \\ &\quad - \frac{\sum_{i=0}^{r-1} p'(\hat{s}, s_0 = i) \sum_{i=0}^{r-1} p(\hat{s}, s_0 = i)}{p(\hat{s})} \\ &= -\sum_{\hat{s}} \sum_{i=0}^{r-1} p'^2(\hat{s}, s_0 = i) \left(\frac{1}{p(\hat{s}, s_0 = i)} - \frac{1}{p(\hat{s})}\right) \\ &\quad - \sum_{i \neq j} \frac{p'(\hat{s}, s_0 = i)p'(\hat{s}, s_0 = j)}{p(\hat{s})} \\ &= -\sum_{\hat{s}} \frac{1}{p(\hat{s})} \sum_{i \neq j}^{r-1} \frac{p'^2(\hat{s}, s_0 = i)p(\hat{s}, s_0 = j)}{p(\hat{s}, s_0 = i)} \\ &\quad - \frac{1}{p(\hat{s})} \sum_{i \neq j} p'(\hat{s}, s_0 = i)p'(\hat{s}, s_0 = j) \\ &= -\sum_{\hat{s}} \frac{1}{p(\hat{s})} \sum_{i \neq j}^{r-1} \Big(p'^2(\hat{s}, s_0 = i)p^2(\hat{s}, s_0 = j) \\ &\quad - p'(\hat{s}, s_0 = i)p'(\hat{s}, s_0 = j)p(\hat{s}, s_0 = i)p(\hat{s}, s_0 = j)\Big) \\ &\quad \Big/ \, p(\hat{s}, s_0 = i)p(\hat{s}, s_0 = j) \\ &= -\sum_{\hat{s}} \frac{1}{p(\hat{s})} \sum_{i < j}^{r-1} \Big(p'(\hat{s}, s_0 = i)p(\hat{s}, s_0 = j) \\ &\quad - p'(\hat{s}, s_0 = j)p(\hat{s}, s_0 = i)\Big)^2 \Big/ \, p(\hat{s}, s_0 = i)p(\hat{s}, s_0 = j) \end{aligned}$$

and the last term is zero if

$$\frac{p'(\hat{s}, s_0 = i)}{p(\hat{s}, s_0 = i)} = \frac{p'(\hat{s}, s_0 = j)}{p(\hat{s}, s_0 = j)} \qquad \text{for all } i, j = 0, \ldots, r-1 \,.$$

This implies

$$\begin{aligned} p'(\hat{s}) &= \sum_{i=0}^{r-1} p'(\hat{s}, s_0 = i) = \sum_{i=0}^{r-1} \frac{p'(\hat{s}, s_0 = i)}{p(\hat{s}, s_0 = i)} p(\hat{s}, s_0 = i) \\ &= \frac{p'(\hat{s}, s_0 = j)}{p(\hat{s}, s_0 = j)} \sum_{i=0}^{r-1} p(\hat{s}, s_0 = i) = \frac{p'(\hat{s}, s_0 = j)}{p(\hat{s}, s_0 = j)} p(\hat{s}) \,. \end{aligned}$$

for all $j = 0, \dots, r-1$ and $\hat{s}$. Hence, we obtain

$$p'(s|\hat{s}) = \left(\frac{p(s)}{p(\hat{s})}\right)' = \frac{p'(\hat{s}, s_0 = j)p(\hat{s}) - p(\hat{s}, s_0 = j)p'(\hat{s})}{p^2(\hat{s})} = 0\,.$$

for all s. But we have $p(s|\hat{s}) = f(\beta)/g(\beta)$ for two functions f and g which are linear in β. This implies

$$p'(s|\hat{s}) = \frac{f'(\beta)g(\beta) - g'(\beta)f(\beta)}{g^2(\beta)}$$

where the numerator does no longer depend on β. Hence, the existence of a single point α_0 where the derivative $p'(s|\hat{s})$ is zero implies that $f'(\beta)g(\beta) = g'(\beta)f(\beta)$ for all $\beta \in I$. This proves that $p(s|\hat{s}) = \frac{f'(\beta)}{g'(\beta)} \in \mathbb{R}$ is constant for all $\beta \in [\mathring{\alpha}]$ and sequences s. Hence, the first derivative $H'(\beta)$ of the conditional entropy is constant and the entropy $H(\beta)$ itself is a linear function in β. □

Proof. **Continuity of the entropy rate in α:** Let $\alpha < \beta \in [0,1]$. We denote by $\mathcal{A}_\alpha = \{[0,\alpha],(\alpha,1]\}$ and $\mathcal{A}_\beta = \{[0,\beta],(\beta,1]\}$ the partitions induced from the thresholds α and β. Due to corollary 4.12.1 in[23] we have $|h(T,\alpha) - h(T,\beta)| \leq H(\mathcal{A}_\alpha|\mathcal{A}_\beta) + H(\mathcal{A}_\beta|\mathcal{A}_\alpha)$ where $H(\mathcal{A}_\alpha|\mathcal{A}_\beta)$ denotes the entropy of the partition $\mathcal{A}_\alpha$ conditioned on the partition $\mathcal{A}_\beta$ and analogously $H(\mathcal{A}_\beta|\mathcal{A}_\alpha)$. This implies

$$\begin{aligned} &H(\mathcal{A}_\alpha|\mathcal{A}_\beta) + H(\mathcal{A}_\beta|\mathcal{A}_\alpha) \\ &= -\,\alpha \log\left(\frac{\alpha}{\beta}\right) - (\beta-\alpha)\log\left(\frac{\beta-\alpha}{\beta}\right) \\ &\quad - (\beta-\alpha)\log\left(\frac{\beta-\alpha}{1-\alpha}\right) - (1-\beta)\log\left(\frac{1-\beta}{1-\alpha}\right) \end{aligned}$$

which converges to 0 if $\alpha \to \beta$. □

Proof. **The image $\phi_\alpha^{\mathbb{N}}([0,1])$ is uncountable:** Since T is weakly mixing we obtain from theorem 1.24 in[23] that $T \times T$ is ergodic. From theorem 1.7 in[23] we obtain the existence of a set $A \subset [0,1]\times[0,1]$ s.t. $\lambda(A) = 1$ and for all $(x,y) \in A$ the orbit $(T^n(x), T^n(y))_{n\in\mathbb{N}}$ is dense in $[0,1]\times[0,1]$.

Suppose that the image $\phi_\alpha([0,1]) = \{\mathfrak{s}_n : n \in \mathbb{N}\} \subset \{0,1\}^{\mathbb{N}}$ is countable for an $\alpha \in (0,1)$. We denote by $A_n = \phi_\alpha^{-1}(\mathfrak{s}_n)$ the domain of the symbol $\mathfrak{s}_n$ for all $n \in \mathbb{N}$. The set $\{A_n : n \in \mathbb{N}\}$ is a partition of the set $[0,1]$ and from σ-additivity, it follows that

$$\sum_{n\in\mathbb{N}} \lambda(A_n) = \lambda([0,1]) = 1$$

which implies the existence of an integer $i \in \mathbb{N}$ s.t. $\lambda(A_i) > 0$. This implies $\lambda(A \cap A_i \times A_i) = \lambda(A_i)^2 > 0$. This implies that the intersection $A \cap A_i \times A_i$ is not empty. Let $(x,y) \in A \cap A_i \times A_i$. Since the orbit of (x,y) is dense in $[0,1]\times[0,1]$ there is an $n \in \mathbb{N}$ s.t. $(T^n(x), T^n(y)) \in (0,\alpha)\times(\alpha,1)$. Hence, $\phi_\alpha \circ T^n(x) = 0$ and $\phi_\alpha \circ T^n(y) = 1$ and $\phi_\alpha(x) \neq \phi_\alpha(y)$ which contradicts the fact that $x, y \in A_i$, that is, $\phi_\alpha(x) = \phi_\alpha(y)$. □

[1] F. M. Atay, S. Jalan, and J. Jost, *Randomness, Chaos, and Structure*, Complexity **15** (2009), no. 1, 29–35.

[2] E. M. Bollt, T. Stanford, Y. Cheng-Lai, and K. Życzkowski, *What symbolic dynamics do we get with a misplaced partition? On the validity of threshold crossing analysis of chaotic time-series*, Physica D **154** (2001), 259–286.

[3] Patrizia Castiglione, Massimo Falcioni, Annick Lesne, and Angelo Vulpiani, *CHAOS AND COARSE GRAINING IN STATISTICAL MECHANICS*, Cambridge University Press, 2008.

[4] F. Christiansen and A. Politi, *Guidelines for the construction of a generating partition in the standard map*, Physica D **109** (1997), 32–41.

[5] P. Collett and J.-P.-Eckmann, *Iterated Maps on the Interval as Dynamical Systmes*, Birkhäuser, 1980.

[6] Thomas M. Cover and Joy A. Thomas, *Elements of information theory*, John Wiley & Sons, 2006.

[7] C. S. Daw, M. B. Kennel, C. E. A. Finney, and F. T. Connolly, *Observing and modeling nonlinear dynamics in an internal combustion engine*, Phys. Rev. E **57** (1998), no. 3.

[8] W. de Melo and S. van Strien, *One-Dimensional Dynamics*, Springer, 1993.

[9] Robert L. Devaney, *An Introduction to Chaotic Dynamical Systems*, Addison-Wesley, 1989.

[10] R. Engbert, C. Scheffczyk, R. T. Krampe, J. Kurths, and R. Kliegl, *Symbolic dynamics of bimanual production of polyrythms*, pp. 271–282, Springer, 1998.

[11] Eric J. Friedman, *Structure and uncomputability in one dimensional maps*, Complex Systems **5** (1991), 335–349.

[12] Olof Görnerup and Martin Nilsson Jacobi, *A method for inferring hierarchical dynamics in stochastic processes*, Adv.Complex Syst. **11** (2008), 1–16.

[13] J. Kurths, A. Voss, P. Saparin, A. Witt, H. J. Kleiner, and N. Wessel, *Quantitative analysis of heart rate variability*, Chaos **5** (1995), 88–94.

[14] C. Letellier, *Symbolic seqeunce analysis using approsimated partition*, Chaos, Solitons Fractals **36** (2008), 32–41.

[15] M. Misiurewicz and E. Visinescu, *Kneading sequences of skew tent maps*, Annales de l'I.H.P. **27** (1991), no. 1, 125–140.

[16] O. Pfante, E. Olbrich, N. Bertschinger, N. Ay, and J. Jost, *Closure measures and the tent map*, Chaos **24** (2014).

[17] ______, *Comparison between different methods of level identification*, Adv.Complex Syst. **17** (2014), no. 2, 215–235.

[18] M. G. Saunders and G. A. Voth, *Coarse-Graining Methods for Computational Biology*, Annu. Rev. Biophys. **42** (2013), 73–93.

[19] Cosma Rohilla Shalizi, *Causal Architecture, Complexity and Self-Organization in Time Series and Cellular Automata*, Ph.D. thesis, University of Wisconsin at Madison, Physics Department, 2001.

[20] Cosma Rohilla Shalizi and Christopher Moore, *What is a Macrostate? Subjective Observations and Objective Dynamics*, Tech. report, 2003, arXiv:cond-mat/0303625v1.

[21] X. Z. Tang, E. R. Tracy, and R. Brown, *Symbol sequence statistics in noisy chaotic signal reconstruction*, Phys. Rev. E **51** (1997), no. 3, 253–261.

[22] ______, *Symbol statistics and spatio-temporal systems*, Physica D **102** (1997), 253–261.

[23] Peter Walters, *An Introduction to Ergodic Theory*, Springer-Verlag, 1981.

[24] David Whitley, *Discrete dynamical systems in dimensions one and two*, Bull. Lond. Math. Soc. (1983), 177–217.