Empirical generating partitions of driven oscillators using optimized symbolic shadowing

Navendu S. Patil* and Joseph P. Cusumano†
Department of Engineering Science and Mechanics, Pennsylvania State University, University Park, Pennsylvania 16802, USA

(Received 26 July 2016; revised manuscript received 27 June 2018; published 17 September 2018)

We present an optimized version of the symbolic shadowing algorithm for coarse graining a continuous state dynamical system, originally due to Hirata and co-workers [Phys. Rev. E **70**, 016215 (2004)]. We validate our algorithm by finding generating partitions presented previously in the literature. We show that, unlike the original, the optimized algorithm can approximate generating partitions for periodically driven continuous-time nonlinear oscillators. We recover known generating partitions for the driven Duffing oscillator and compute generating partitions for the driven van der Pol oscillator. We also examine the problem of how algorithms such as ours can be applied "objectively," that is, by starting from arbitrary initial partition guesses. By applying our algorithm to large ensembles of initial random partitions, we show that symbolic shadowing leads to a multiplicity of candidate generating partitions that localize points in phase space to a high degree, thus making it difficult to select the best choice(s). We thus propose using the Lempel-Ziv complexity to identify partitions from this set of candidates that are, in a specific sense, "minimal," i.e., those with contiguous cells, fewer cell boundaries, and a smaller number of cells compared to their rivals. We also show how our methods can be used to indicate the appropriate number of symbols needed to approximate a generating partition.

DOI: 10.1103/PhysRevE.98.032211

I. INTRODUCTION

In many practical situations, chaotic time series data are gathered and used to gain an understanding of the dynamics of a complex physical system. In particular, it is of paramount importance to use such data to provide a model-based and objective quantification of the dynamical complexity of a system, which, together with a measure of its stochasticity, determine fundamental limits to the system's predictability. However, a number of approaches for estimating dynamical complexity are based on information theory [1] and require sequences of *coarse-grained* or, more specifically, discrete observables, which are not usually directly available, and typically have to be generated from continuous-valued discrete time series. In this work, we obtain such discrete observables for deterministic chaotic systems *solely and directly* from their continuous-valued trajectories in the phase space, using a modification of a general numerical algorithm [2]. Furthermore, we devise a selection procedure utilizing the Lempel-Ziv complexity of symbol sequences [3] to pick good observables that have a simple structure in the continuous phase space, and that also remain faithful to the original time series measurements without much loss of the information. Finally, we show how our approach allows us to estimate the number of symbols needed, i.e., the alphabet size of discrete observables.

To construct coarse-grained observables, it is necessary to partition data in the continuous phase space into disjoint subsets, each of which is represented by a unique symbol from a finite alphabet. Using such a partition, the original continuous-valued time series is transformed into a symbol sequence which can be studied using a variety of tools from information theory, formal languages, and automata theory [4]. Some applications for which the experimental analysis of symbolic time series have furnished useful results can be found in the review paper [5] and references therein. Furthermore, symbolic dynamics [6] provides a useful theoretical framework to build models from discrete observables, which is conceptually simple and easier to handle empirically. However, for such models to capture the dynamics in a continuous phase space, one must obtain a "faithful" set of observables for which the resulting symbolic dynamics are, in a specific sense, equivalent to the dynamics in the continuous phase space. *Generating partitions* [7] provide such a description by establishing a one-to-one correspondence between a dynamical trajectory in the continuous phase space and an infinitely long symbol sequence that encodes that trajectory. In general, this one-to-one correspondence does not survive in the presence of process noise [8], although there exist rigorous mathematical formulations for generating partitions of stochastic systems evolving according to deterministic maps drawn at random from an *a priori* fixed collection [9]. However, whether or not the system is purely deterministic, in an empirical context one can only hope to *approximately* reconstruct an equivalence between the dynamics in the continuous phase space and the symbolic dynamics. This is done by identifying increasingly long symbol strings, as allowed by the amount of data available, with progressively smaller sets of points (or sets of trajectories thereof) in the continuous space. In this sense, a reasonable algorithm relying on a good heuristic, if known to yield good approximations of known generating partitions, can be used to obtain empirical partitions of noisy time series. Such empirical partitions should then be useful for the construction of good coarse-grained observables for deterministic dynamical systems, including, possibly, those with sufficiently small stochastic perturbations.

*nsp129@psu.edu
†jpc3@psu.edu

In this paper, we focus on the deterministic case and present an optimized version of the symbolic shadowing algorithm [2] that provides just such coarse-grained observables. We apply our general algorithm to find generating partitions for a number of systems found previously in the literature by various means but show that, unlike the original symbolic shadowing algorithm, our optimized algorithm allows us to obtain generating partitions for periodically driven continuous-time nonlinear oscillators. Specifically, we find generating partitions for Duffing and van der Pol oscillators: while those for the Duffing oscillator were found previously, using entirely different approaches (mentioned below), our results for the van der Pol oscillator, to the best of our knowledge, have not been previously published. We also examine the problem of how algorithms such as ours can be applied to obtain approximate generating partitions "objectively," that is, by starting from arbitrary initial guesses for the partitions. We show, by applying our algorithm to large ensembles of initial random encodings, that symbolic shadowing does not lead to unique solutions, and, indeed, can result in a certain degree of ambiguity regarding the best choice to approximate a system's generating partition. We thus propose the use of the Lempel-Ziv complexity [3], as an additional tool, for identifying partitions from a set of candidates that are, in a specific sense, "minimal," i.e., those with contiguous cells, fewer cell boundaries, and a smaller number of cells compared to their rivals. Finally, we also show how our methods can be used to indicate the appropriate number of symbols needed to approximate a generating partition. In the next two paragraphs, we briefly review approaches based on well-accepted heuristics used to construct generating partitions. These methods use time series data only indirectly, i.e., they require knowledge of various invariant structures in the phase space. We discuss them below because we validate our solutions by comparing to results obtained with these alternative methods, whenever available.

The problem of finding generating partitions, especially from observed time series data alone, is nontrivial. If a scalar time series is well approximated using the output of a one-dimensional map, the critical points of the corresponding return map determine the required generating partition [10]. However, although certain insights picked up from one-dimensional discrete-time systems (such as, e.g., the folding of the attractor at cell boundaries of the generating partition) have proven useful in the two-dimensional setting, the problem of finding generating partitions for multidimensional maps is far from being completely solved. This is primarily because generating partitions are not unique, and the space of partitions is prohibitively large, even for two-dimensional systems, to allow for a systematic search in the absence of additional dynamical information. However, a number of approaches, some heavily relying on the geometry of phase space, have been successfully used to find generating partitions of some well-known two-dimensional maps, including Hénon and Ikeda maps. A two-cell generating partition for the Hénon map has been obtained by joining points of homoclinic tangencies, i.e., points where the stable and unstable manifolds of the map intersect tangentially [11]. This approach has been successfully extended to flows to obtain a three-cell generating partition of the attractor of the two-well, driven, Duffing oscillator, where the primary homoclinic tangency points, at which the sum of curvatures of both manifolds is small, were used to obtain partition boundaries [12,13]. Difficulties with the construction of generating partitions using homoclinic tangencies were discussed in [14] and [15]. Specifically, it was observed that the concept of "primary" tangencies is not well defined: in some situations, boundaries consisting of nonprimary tangencies also yield generating partitions. Furthermore, locating these tangency points is especially challenging with noisy time series data, as can be expected in physical experiments and even, to some extent, numerical simulations.

Another important approach to finding generating partitions relies on the requirement that every periodic orbit, stable or unstable, be encoded by a unique symbol sequence [16]. The points on an aperiodic trajectory are then encoded by interpolating from periodic points with known symbol labels, which act as reference points. An algorithm to find generating partitions of a strange attractor, which is dense with unstable periodic orbits (UPOs), uses "proximity functions" in the phase space to hierarchically encode orbits of increasing period starting from an initially chosen encoding of short UPOs [17]. In other work, extensive topological analysis of UPOs embedded in the attractor is carried out to obtain their knot invariants and a branched manifold (or "template") holding them, consistent with the topological structure of the attractor [18]. An interpolation algorithm is then used to combine possible symbolic names of UPOs, consistent with their topological invariants, along with their spatial locations on the attractor, to obtain an approximation of a generating partition [19]. Unlike other approaches, this method gives the required number of cells of a generating partition, which equals the number of branches of the template. It has also been employed to independently recover the same three-cell generating partition for the two-well Duffing oscillator, previously obtained using the method of homoclinic tangencies, as discussed above. However, finding a sufficient number of UPOs to reasonable accuracy, as required by these methods, can be a challenging task, especially for limited, noisy, time series data.

In light of difficulties associated with approaches relying heavily on the geometry and/or topology of phase space, we are interested in general-purpose algorithms that work directly on time series data to obtain good approximations of generating partitions. From this more empirical perspective, two important pieces of work are available. First, Kennel and Buhl [20,21] present algorithms to find good partitions that try to avoid *topological degeneracies*, i.e., situations where two "similar" symbol strings map to points on the attractor which are quite far apart. This is done using stochastic optimization techniques to iteratively minimize the number of symbolic false nearest neighbors (SFNNs), i.e., points that are neighbors in the space of symbol sequences but not in the phase space. Using the SFNN approach, the authors were able to obtain generating partitions for discrete-time dynamical systems, the Ikeda and Hénon maps, including in the presence of observational noise. They also applied their approach to some engine test data to obtain a partition with low values of their SFNN-based statistic. However, because of the stochastic nature of these algorithms, for the same initial guess of

the partition, in different runs one can obtain quite different final partitions. Furthermore, to obtain consistent, meaningful results near global minima of the statistic, it is required to have a good set of tuning parameters for each system, which is not known *a priori*. To our knowledge, these algorithms have not been demonstrated to recover the known three-symbol generating partition of the driven Duffing oscillator.

Here, we consider another general numerical approach due to Hirata *et al.* [2], viz., the *symbolic shadowing algorithm*, which is deterministic in the sense that the same input (that is, the same initial partition) leads to same output for a given set of algorithm parameters and time series data. In this approach, each partition is specified by a set of *representatives*, i.e., points in the phase space in one-to-one correspondence with all substrings, of given length, as observed in a symbolic encoding of the time series. Each representative is located at the mass center of the set of points in the phase space it *approximates* or "represents." The goal of the symbolic shadowing algorithm is then to minimize the discrepancy between the shadowing trajectory, consisting of representatives, and the given time series in the phase space. Our primary interest in this algorithm is due to its simplicity, ease of implementation, computational efficiency, and the fact that it does not require the extensive tuning of algorithmic parameters. Symbolic shadowing has been successfully used to estimate generating partitions of one and two-dimensional maps with observation and process noise, including two-cell partitions of Hénon and Ikeda maps [2,22].

However, the symbolic shadowing algorithm has not yet been shown to empirically converge (i.e., to converge using methods that work directly on data) to known generating partitions of driven nonlinear oscillators such as the two-well, driven Duffing oscillator [12,19]. In fact, in preliminary studies we found that using a large ensemble of randomly chosen initial partitions, or even starting with the partition uniquely encoding the first few of the low period UPOs observed in time series, the standard symbolic shadowing algorithm could not find the aforementioned three-cell partition for the two-well Duffing oscillator. In this work, we propose a modification of the symbolic shadowing algorithm which achieves the required convergence for many different initial partitions. For a *specified* symbolic encoding (or, partition), this modification yields shadowing trajectory of representatives which approximates the original time series data in the phase space optimally. Here, "optimal" means that the discrepancy, quantified as the squared Euclidean distance, between the time series data and its shadowing trajectory is minimized.

Generating partitions are not unique, and this nonuniqueness affects the convergence of algorithms in the sense that different randomly chosen initial partitions can yield different final results. It is thus desirable to objectively assess if some partitions are "more generating" than others so that one can pick the "best" estimate of a generating partition out of distinct partitions obtained. Here, "more generating" means the degree to which some partitions yield better correspondence between progressively smaller sets of points in the phase space and increasingly longer symbol strings. However, with a limited amount of data it is conceivable that some candidate generating partitions might not be easily distinguishable in this fashion. Empirically speaking, the problem of selecting a good approximation to a generating partition could be looked at as analogous to model selection with the usual "model error" versus "model size" trade-off: whenever possible, the goal is to find the generating partition with "simplest" boundary structure and the minimal number of cells in the phase space. Here we propose to use the "minimality" of a partition *in addition* to its "generating property" to further distinguish between different candidate partitions. We hypothesize that the Lempel-Ziv complexity [3] calculated for individual symbol sequences provides a measure of "minimality" of a partition in the sense that partitions with contiguous cells, fewer cell boundaries, and fewer cells have lower Lempel-Ziv complexity (are "more minimal") than their rivals.

Finally, it is to be noted that, typically, the number of cells k required for a minimal generating partition is not known *a priori*. A lower bound can be obtained if we know the topological entropy of an attractor under consideration. In particular, $k \geqslant \lceil 2^{h_0} \rceil$, where the topological entropy, h_0 (in bits per iteration), quantifies the exponential rate of growth of the number of observed "orbits" (symbol sequences) of increasing length [7,23]. Alternatively, the number of low period UPOs provides a lower bound on k which can sometimes be tighter than that found using topological entropy. For example, it is not possible to uniquely encode an attractor which has three period-3 UPOs embedded in it using a binary alphabet, even if the topological entropy is less than 1 bit per iteration. Although we do not resolve the issue of exactly estimating k, in this work, we take an indirect empirical approach. We use ensembles of symbol sequences over differently sized alphabets ($k \geqslant 2$) as inputs to symbolic shadowing algorithms and assess the distinct solutions that result for their potential to serve as good candidates for minimal generating partitions. Our empirical results suggest that the number of distinct candidate partitions increases considerably for alphabets with more symbols than needed for the minimal generating partition. On the other hand, if we use fewer symbols than required, we obtain very few distinct solutions, which also inadequately approximate generating partitions.

The remainder of this paper is organized as follows. In Sec. II, we discuss the role of coarse-grained observables in the modeling process from a dynamical systems perspective. We also define generating partitions, which yield the "best" observables, along with a criterion to empirically assess their "generating property." In Sec. III, we present the basic notion of symbolic shadowing and lay out the original symbolic shadowing algorithm step by step. We propose two different methods to randomly generate large ensembles of initial partitions and discuss the results after applying the original algorithm to the two-well Duffing oscillator. In Sec. IV, we present an optimized symbolic shadowing algorithm in detail and compare and contrast it with the original algorithm on various fronts, such as convergence, performance, and computational efficiency. We also provide an empirically grounded explanation as to why the optimized algorithm significantly improves convergence to generating partitions for systems like the driven Duffing oscillator, as opposed to Hénon and Ikeda maps, which are not derived from flows and for which the original algorithm is adequate. In Sec. V, we explain the role that the "minimality" of a partition plays in helping

select candidate generating partitions, and provide a heuristic argument, supported by some empirical results, to advocate the use of Lempel-Ziv complexity in its quantification. Using these ideas, we analyze the results of applying the original and optimized symbolic shadowing algorithms to a large ensemble of random initial partitions for the driven Duffing oscillator. We conclude by presenting approximate three-cell generating partitions, found using optimized symbolic shadowing for the driven van der Pol oscillator.

II. COARSE-GRAINED MODELS OF NONLINEAR OSCILLATORS

We are interested in physical systems with an underlying mechanism of deterministic chaos, and which thus show complicated steady-state behavior as they evolve in time. Ultimately, we aim to characterize such systems by building coarse-grained probabilistic models of their observed temporal behavior. This paper focuses on the first step in creating such models: namely, the identification and selection of suitable coarse-grained variables solely and directly from a trajectory in the phase space. Here, we further restrict our attention to periodically driven continuous-time nonlinear oscillators whose phase space is a three-dimensional manifold, $\mathbb{R}^2 \times S^1$, where the circular component S^1 results from the periodicity of their vector field [6]. We study the discrete-time dynamics of these nonlinear oscillators by means of an appropriately defined two-dimensional Poincaré section $\Gamma \subset \mathbb{R}^2 \times S^1$: here we fix an element in S^1 (indicating a specified phase of the forcing) so that $\Gamma = \mathbb{R}^2$. For dissipative systems, we have a compact invariant subset $\Lambda \subset \Gamma$, i.e., an attractor, on which the observed steady-state dynamics takes place. The corresponding Poincaré map $\boldsymbol{F} : \Lambda \mapsto \Lambda$ is given by

$$\boldsymbol{x}_{i+1} = \boldsymbol{F}(\boldsymbol{x}_i), \tag{1}$$

where $\boldsymbol{x}_i \in \Gamma$ denotes the state of a physical system at discrete time $i \in \mathbb{Z}$.

In what follows, we assume that we only have access to time series data $\{\boldsymbol{x}_i : \boldsymbol{x}_i \in \Lambda \subset \Gamma, i \in \mathbb{Z}\}$ from an unknown, yet stationary, ergodic, dissipative, discrete-time map $\boldsymbol{F}$ as in Eq. (1). The ergodicity of $\boldsymbol{F}$ implies that typical trajectories visit neighborhoods of almost all points on the attractor Λ in the Poincaré section Γ, and that observed trajectories sample the natural invariant measure (or physical measure) μ [24].

A. Coarse-grained observations

By analogy with statistical mechanics, in experimental contexts we can think of each $\boldsymbol{x} \in \Gamma$ as a "microstate," which in most practical situations is "hidden" in the sense that we only have access to noisy measurements of real-valued vectors or some of their scalar components. As a consequence, a single experimental measurement at each time can be attributed to many conceivable microstates from Γ. This motivates the concept of *coarse graining*, which means the lumping together of many microstates, or experimental measurements thereof, and identifying them with a unique symbol, or "macrostate." One then aims to develop a finite collection of such symbols in an *alphabet* $\mathcal{A}$, with each symbol representing a "cell" or region of Γ. Taken together, these cells form a partition covering every microstate from the invariant set $\Lambda \subset \Gamma$.

Formally, we choose a time-invariant, noninvertible function $\boldsymbol{G} : \Gamma \mapsto \mathcal{A}$ of microstates, which gives *coarse-grained observations* or, more simply, *observables*, $X = \boldsymbol{G}(\boldsymbol{x}) \in \mathcal{A}$, where, without loss of generality, we let $\mathcal{A} \equiv \{0, 1, \ldots, k-1\}$. We refer to $k = |\mathcal{A}|$ as the *size* of $\mathcal{A}$. The function $\boldsymbol{G}$ divides the entire phase space Γ into k disjoint cells to obtain a partition $\mathcal{P} = \{P_0, P_1, \ldots, P_{k-1}\}$ of Γ with each cell given by $P_u = \{\boldsymbol{x} \in \Gamma : X = \boldsymbol{G}(\boldsymbol{x}) = u \in \mathcal{A}\}$. By successively applying $\boldsymbol{G}$ at each time, we get a stationary symbolic encoding, $\{X_i, i \in \mathbb{Z}\}$, of the dynamical trajectory $\{\boldsymbol{x}_i\}$. A key issue is that the partition $\mathcal{P}$ must be chosen carefully so that there is no loss of relevant dynamical information. This leads directly to the concept of generating partitions, reviewed in the next section, which yield the "best" coarse-grained observables, that is, that preserve all of the relevant dynamical information, and that are central objects of study in this paper.

B. Generating partitions

Let us consider a dynamical trajectory that visits cells P_u and P_v of partition $\mathcal{P}$ of Γ at times i and $i+1$ respectively: i.e., $\boldsymbol{x}_i \in P_u$ and $\boldsymbol{x}_{i+1} \in P_v$. Thus, the corresponding observed process $\{X_i\}$ is such that $X_i = u$ and $X_{i+1} = v$ for $u, v \in \mathcal{A}$. Since $X_{i+1} = v$, $\boldsymbol{x}_i$ must also lie in the first preimage of P_v under $\boldsymbol{F}$, which is, $\boldsymbol{F}^{-1}(P_v) = \{\boldsymbol{x} \in \Gamma : \boldsymbol{F}(\boldsymbol{x}) \in P_v\}$. We note that, for all $v \in \mathcal{A}$, the sets $\boldsymbol{F}^{-1}(P_v)$ themselves form a partition of Γ given by $\boldsymbol{F}^{-1}\mathcal{P} = \{\boldsymbol{F}^{-1}(P_0), \boldsymbol{F}^{-1}(P_1), \ldots, \boldsymbol{F}^{-1}(P_{k-1})\}$. Furthermore, the sets $\{P_u \cap \boldsymbol{F}^{-1}(P_v) : u, v \in \mathcal{A}\}$ form a partition, denoted by $\mathcal{P} \vee \boldsymbol{F}^{-1}\mathcal{P}$, called the *join* of the partitions $\mathcal{P}$ and $\boldsymbol{F}^{-1}\mathcal{P}$. It is evident that for a given choice of partition $\mathcal{P}$ the uncertainty in the true value of $\boldsymbol{x}_i$, still remaining after recording $X_i = u$, will typically reduce as a result of an additional coarse-grained measurement at time $i+1$ (i.e., the value of X_{i+1}), *provided* that $P_u \cap \boldsymbol{F}^{-1}(P_v) \subset P_u$.

When true, the joined partition $\mathcal{P} \vee \boldsymbol{F}^{-1}\mathcal{P}$ is the first *dynamical refinement* of the cells in $\mathcal{P}$ under $\boldsymbol{F}$ [7]. Note that if $\boldsymbol{F}$ is invertible, then $\boldsymbol{F}\mathcal{P}$ consisting of cells, $\boldsymbol{F}(P_v) = \{\boldsymbol{x} \in \Gamma : \boldsymbol{F}^{-1}(\boldsymbol{x}) \in P_v\}$ for every $v \in \mathcal{A}$, is also a legitimate partition. Thus, if $\boldsymbol{F}\mathcal{P} \vee \mathcal{P}$ is a (dynamical) refinement of $\mathcal{P}$ under the image of $\boldsymbol{F}$ then the knowledge of X_{i-1} will also reduce the uncertainty about $\boldsymbol{x}_i$. Of course, this process of refinement of $\mathcal{P}$ under $\boldsymbol{F}$ can continue as subsequent coarse-grained measurements, $X_{i+2}, X_{i+3}, \ldots$ (or, from the past, $X_{i-2}, X_{i-3}, \ldots$) become available, yielding a further reduction in the uncertainty about $\boldsymbol{x}_i$.

A partition $\mathcal{P}$ for which this process continues *indefinitely*, i.e., for infinitely long symbol sequences, is known as a *generating* partition. In general, we can construct a tree of partitions with cells of partition $\mathcal{P}$ occupying k nodes at the first level below the root and cells of its successive dynamical refinements occupying nodes of subsequent levels. At level l, we have k^l nodes consisting of cells of the partition, $\mathcal{Q}_l \equiv \bigvee_{q=0}^{q=l-1} \boldsymbol{F}^{-q}\mathcal{P} = \mathcal{P} \vee \boldsymbol{F}^{-1}\mathcal{P} \vee \cdots \vee \boldsymbol{F}^{-l+1}\mathcal{P}$ (or, k^{2l-1} nodes consisting of cells of $\mathcal{Q}_l \equiv \bigvee_{q=-l+1}^{q=l-1} \boldsymbol{F}^{-q}\mathcal{P}$, if $\boldsymbol{F}$ is invertible), each of which is uniquely labeled by a length-l [length-$(2l-1)$,

if $\boldsymbol{F}$ is invertible] symbol sequence over $\mathcal{A}$. Measure theoretically, if a partition $\mathcal{P}$ *generates* under the action of $\boldsymbol{F}$, then the full σ-algebra $\mathcal{F}$ (corresponding to the dynamical system $\boldsymbol{F}$ as a measure-preserving transformation) equals the σ-algebra generated by $\bigvee_{l=1}^{\infty} \mathcal{Q}_l$: that is, by taking countable unions and intersections of cells of $\mathcal{Q}_l$, for all l, it is possible to generate every set (including singleton sets) of $\mathcal{F}$ [25]. Thus, for a generating partition $\mathcal{P}$ it is possible to *separate* almost any two distinct points $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$ of the phase space Γ into two cells of $\mathcal{Q}_l$ for a sufficiently large l. As a consequence, almost every point (or its dynamical trajectory) in the phase space is assigned a unique infinitely long symbol sequence. Furthermore, the cells of $\mathcal{Q}_l$ "shrink" to singleton sets *almost* everywhere as $l \to \infty$: i.e., $\max_i \mu(\mathcal{Q}_{l,i}) \to 0$, where $\mathcal{Q}_{l,i}$ is the i^{th} cell of partition $\mathcal{Q}_l$ and $\mu(\cdot)$ denotes its invariant probability measure. As a result, to quantify the "generating property" of a given partition, in the next section we will utilize the decreasing "size" of cells in the phase space with increasing l as an easier-to-calculate proxy for the measure μ. While there could be regions in the phase space in which points cannot be coded uniquely, even using generating partitions, as stated above such regions will have vanishingly small measure μ and so will not be observed in typical time-series data.

The separation property of a generating partition is key to preserving an equivalence between the map $\boldsymbol{F}$ on Γ and the dynamics on the space of symbol sequences, the study of which is known as *symbolic dynamics* [6]. Because of this equivalence, generating partitions are extremely appealing from a conceptual viewpoint because the dynamics on the symbol sequences is always governed by the *shift map*, as described below, irrespective of the actual functional form of the map $\boldsymbol{F}$, which frequently is not even known in explicit form. In the following, we assume that $\boldsymbol{F}$ is invertible.

Let $\mathcal{A}^{\mathbb{Z}}$ be the space of all symbol sequences for a given alphabet, i.e., the set of all possible bi-infinite sequences $S \equiv (\ldots s_{-2}s_{-1} \centerdot s_0 s_1 s_2 \ldots)$, with each coordinate $s_i : i \in \mathbb{Z}$, drawn from $\mathcal{A}$, and with the coordinate s_0, to the immediate right of the "center" (denoted by a dot "$\centerdot$"), being the zeroth or time-0 coordinate of a sequence S. Then, the dynamics on the space of symbol sequences is governed by the shift map, $\sigma : \mathcal{A}^{\mathbb{Z}} \mapsto \mathcal{A}^{\mathbb{Z}}$, defined as

$$\sigma(\ldots s_{-2}s_{-1} \centerdot s_0 s_1 s_2 \ldots) = (\ldots s_{-2}s_{-1}s_0 \centerdot s_1 s_2 \ldots), \quad (2)$$

which simply shifts each input bi-infinite sequence S by one coordinate to the left to obtain an output bi-infinite sequence $\sigma(S)$, so that $\sigma(S)_0$ [i.e., time-0 coordinate of $\sigma(S)$] is s_1. Any partition induces a map, $\phi : \Gamma \mapsto \mathcal{A}^{\mathbb{Z}}$, which provides a relationship between points in the phase space and symbol sequences. For an invariant set $\Lambda \subset \Gamma$ (e.g., an attractor) in the phase space, the image $\phi(\Lambda)$ constitutes a shift space which is closed under the shift map σ. For a physical system, typically, $\phi(\Lambda)$ is a subset of $\mathcal{A}^{\mathbb{Z}}$ since many of the finite sequences of transitions between partition cells do not occur. Most importantly, for a generating partition $\mathcal{P}$, the map ϕ is *one to one* and *continuous*. Also, for the ϕ induced from a generating partition, the shift map σ acting on $\phi(\Lambda)$ is conjugate to $\boldsymbol{F}$ acting on Λ: that is, $\phi(\boldsymbol{F}(\boldsymbol{x})) = \sigma(\phi(\boldsymbol{x}))$.

Heuristically speaking, we might expect a reasonable partition with a large number of cells, e.g., a coordinate grid over Γ with many "cubic" boxes, and which is technically not generating, to refine dynamically for fairly long symbol sequences, meaning that newly measured observables bring in new information about the system's initial condition. However, generating partitions are advantageous in that they typically have only a few cells. This property is crucial in practice, as the length of the time series required to build useful probabilistic models grows with the number of cells of the partition. In particular, for large $\mathcal{A}$ the number of distinct strings of length l observed in the symbolic time series grows rapidly with l, so that their true probabilities become very small, which in turn means that we need ever longer time series for adequate statistical estimates. Nevertheless, it is important to compare different reasonable partitions, including generating ones, using some objective criterion applied to finite amounts of data. This is the subject of the next subsection.

C. How "generating" are partitions?

For a general empirical approach to assess generating partitions, as we adopt in our work, it is important to test their defining properties as directly as possible, especially given their abstract nature. However, the "guiding principle" used to construct generating partitions [11], that two points of any homoclinic intersection be coded with distinct symbol sequences, is difficult to test using merely time-series data. In other work [26], the relabeling or misrepresentation of the graph associated with the full shift of the tent map was used to explain the topological error due to misplaced partitions. However, most dynamical systems (including the nonlinear oscillators to be studied here) are not known to admit such *exact* (i.e., finite graph) representations, making it difficult to accurately quantify topological errors arising from nongenerating partitions.

In this work, given a partition, we take as an *empirical* measure of its "generating property" its ability to localize points in the phase space for increasingly long symbol strings [2,24]. For any partition, each observed delimited string

$$s_{-m}s_{-m+1}\ldots s_{-1} \centerdot s_0 s_1 \ldots s_{n-1}s_n,$$

of length $l = m + n + 1$, denoted henceforth by s_{-m}^{n}, corresponds to a set of points in the phase space,

$$\mathcal{C}\left(s_{-m}^{n}\right) \equiv \{\phi^{-1}(S) : S_i = s_i, \\ -m \leqslant i \leqslant n = l - m - 1, \forall\, S \in \mathcal{A}^{\mathbb{Z}}\}, \quad (3)$$

also known as a *cylinder*. The preimage ϕ^{-1} is a continuous map if and only if the partition is generating [24,27]. Each of the cylinders can be empirically estimated from a long time series $\{X_i, i = 1, \ldots, N\}$ by simply sliding a window of length l by one symbol from left to right to get

$$\mathcal{C}\left(s_{-m}^{n}\right) = \left\{\boldsymbol{x}_i : X_{i-m}^{i+n} = s_{-m}^{n}, m+1 \leqslant i \leqslant N-n\right\}, \quad (4)$$

where X_{i-m}^{i+n} denotes a delimited string, $X_{i-m}X_{i-m+1}\ldots X_{i-1} \centerdot X_i X_{i+1}\ldots X_{i+n-1}X_{i+n}$. Continuity of ϕ^{-1} implies that points belonging to the same cylinder are closer together so that the decrease in the size of cylinders $\mathcal{C}(s_{-m}^{n})$ for increasing string length l provides a natural way to quantify the

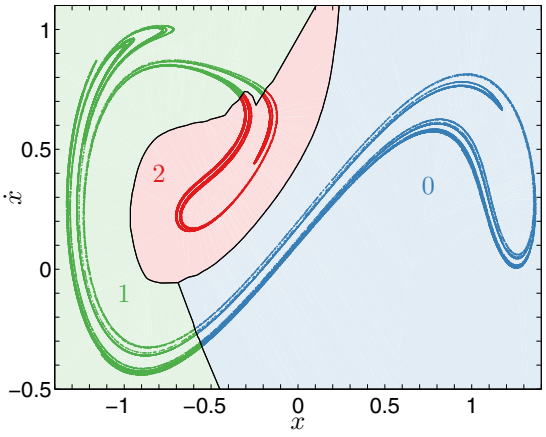

FIG. 1. Generating partition consisting of three cells labeled using symbols {0, 1, 2} for the strange attractor of the stochastically unperturbed, two-well, driven Duffing oscillator given by a second-order ordinary differential equation (ODE): $\ddot{x} + 0.25\,\dot{x} - x + x^3 = 0.4\cos t$. Also shown in the phase space $\mathbb{R}^2$ are iterates of the Poincaré map, each obtained by numerical integration for half the forcing period (i.e., from 0 to π), followed by inversion of coordinates: $x \to -x$ and $\dot{x} \to -\dot{x}$. Although the partition has been obtained for the strange attractor, the cell boundaries have been extrapolated outside the attractor for clarity. This partition, obtained from time series data using methods discussed later in the paper, matches well with the previously published generating partition found using independent methods [12,19].

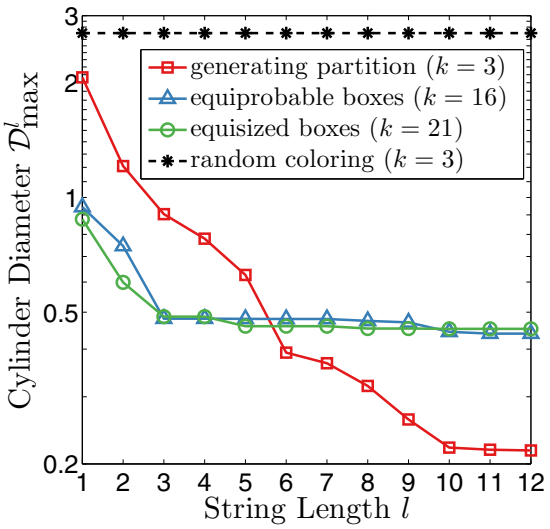

FIG. 2. Comparison of cylinder diameter $\mathcal{D}^l_{\max}$ [Eq. (6)] vs string length l for four partitions (with different alphabet sizes k) of the Duffing attractor, empirically estimated from a time series of four million points. Note the base-10 logarithmic scale on the y axis. For the $k = 3$ (three symbol) generating partition of Fig. 1, $\mathcal{D}^l_{\max}$ shows a sustained decrease until about $l = 10$. For the $k = 16$ "maximum entropy" (equiprobable) and $k = 21$ equisized (same area) partitions, $\mathcal{D}^l_{\max}$ decreases by a factor of one-half until $l = 3$, without further substantial decrease for longer strings. For a $k = 3$ random coloring every data point is uniformly randomly assigned a symbol from $\mathcal{A} = \{0, 1, 2\}$. Symbol sequences from this partition correspond to sets of points spread over all of the attractor so that $\mathcal{D}^l_{\max}$ does not decrease with l and equals the diameter of the attractor itself. As expected, unlike other partitions, the generating partition consistently increases the degree of localization of points on the Duffing attractor, even at string lengths larger than $l = 5$.

generating property of a given partition. Indeed, for a fixed l, cylinders are synonymous with the cells of partition $\mathcal{Q}_l$ belonging to the tree of partitions discussed in Sec. II B. In particular, along the line of [19], we empirically estimate the maximum cylinder diameter $\mathcal{D}^l_{\max}$, corresponding to all strings of length l observed in a long time series, as follows:

$$\mathcal{D}^{l,m}_{\max} \equiv \max_{\mathcal{C}(s^{l-m-1}_{-m}) \in \mathcal{C}^{l,m}} \operatorname{diam}\left[\mathcal{C}\left(s^{l-m-1}_{-m}\right)\right], \tag{5}$$

$$\mathcal{D}^{l}_{\max} \equiv \min_{0 \leqslant m \leqslant l-1} \mathcal{D}^{l,m}_{\max}, \tag{6}$$

where $\operatorname{diam}[\mathcal{C}(s^{l-m-1}_{-m})]$ is a maximum Euclidean distance between all pairs of points belonging to the cylinder $\mathcal{C}(s^{l-m-1}_{-m})$; and $\mathcal{C}^{l,m}$ is the collection of all cylinders [estimated empirically as per Eq. (4)] corresponding to all observed delimited strings of length l with exactly m symbols to the left of their center.

To illustrate the significance of these ideas, consider the three-symbol generating partition for the driven Duffing oscillator in Fig. 1, obtained previously [12,19], and to be found from time-series data using methods developed later in this paper. As can be seen in Fig. 2, this generating partition substantially decreases the maximum cylinder diameter $\mathcal{D}^l_{\max}$, even at string lengths greater than $l = 5$, indicating that it is expected to localize points in the phase space well. In contrast, other reasonable partitions, including those consisting of equiprobable boxes ($k = 16$) or equisized boxes ($k = 21$), do not show substantial decrease in $\mathcal{D}^l_{\max}$ beyond $l = 3$ and have, in fact, higher values of $\mathcal{D}^l_{\max}$ for $l > 5$ than those for the generating partition ($k = 3$). This suggests that partitions which retain greater "static" information by using a greater number of cells (and, therefore, more symbols) are not necessarily good at retaining dynamically relevant information.

As we will see, the minimization of diameter in Eq. (6) proves crucial for estimating generating partitions empirically. It is easy to see that, from any symbolic itinerary $\{X_i, X_{i+1}, \ldots, X_{i+l-1}; 1 \leqslant i \leqslant N - l + 1\}$ of length $l = m + n + 1$, we get $l - 1$ delimited strings s^n_{-m} of length l (and their cylinders) depending on the somewhat arbitrary choice of the location of the string center, that is, of the value of m between 0 and $l - 1$. However, as can be seen in Fig. 3, the maximum diameter $\mathcal{D}^{l,m}_{\max}$ [Eq. (5)], estimated from the three-symbol generating partition of Fig. 1, strongly varies with m for fixed l. In particular, for each l there is a value of $m = m_{\text{opt}}$ that yields the smallest cylinder (i.e., that localizes points within it most strongly) of diameter $\mathcal{D}^l_{\max}$, consistent with the minimization of Eq. (6).

This can be explained as follows. For any given delimited string s^n_{-m} the set of points $\mathcal{C}(s^n_0)$ align along a local stable manifold as they share a common future, while points in the cylinder $\mathcal{C}(s^0_{-m})$ fall along a local unstable manifold since they share a common past (i.e., the same backward sequence of symbols). Also, we clearly have $\mathcal{C}(s^n_{-m}) = \mathcal{C}(s^0_{-m}) \cap \mathcal{C}(s^n_0)$. For dissipative chaotic systems, such as the Duffing oscillator, the rate of convergence of trajectories along a local stable

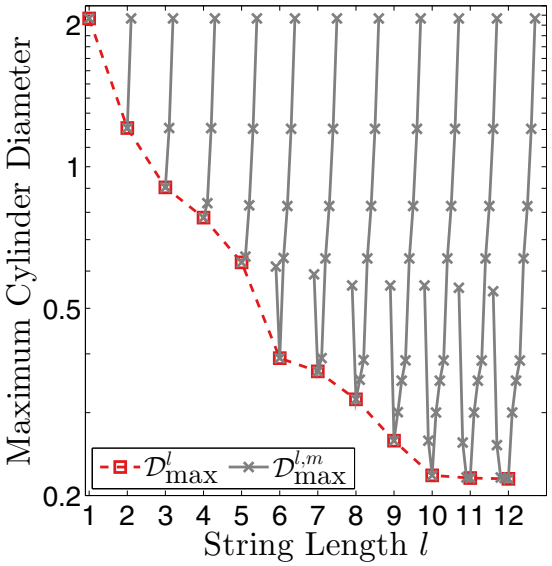

FIG. 3. Variation of maximum cylinder diameter $\mathcal{D}^{l,m}_{\max}$ [Eq. (5)] for three-symbol generating partition of the Duffing attractor (Fig. 1). For each l, the l largest diameters $\mathcal{D}^{l,m}_{\max}$ are shown by $\times$ and solid line, corresponding to values of m from 0 to $l-1$ from left to right within each curve. We see $\mathcal{D}^{l,m}_{\max}$ varies significantly with m, with minimum $\mathcal{D}^{l}_{\max}$ shown by $\square$ and dashed line: these minima are the same values as in Fig. 2, with m_{opt} as in Eq. (7).

manifold is higher than the rate of divergence of nearby points along a local unstable manifold, when averaged over the entire attractor. Thus, on average, the forward sequence of symbols is more "efficient" than the backward sequence in localizing points of a given partition cell. This means, for *fixed* l, the optimal choice of m satisfies

$$m_{\text{opt}} = \underset{0\leqslant m\leqslant l-1}{\operatorname{argmin}}\, \mathcal{D}^{l,m}_{\max} \leqslant \lfloor l/2 \rfloor. \tag{7}$$

This observation is critical for the estimation of generating partitions from time series data. In the next section, we consider the symbolic shadowing algorithm [2] for estimating generating partitions and its modification based on the above observation, which significantly improves its convergence.

III. SYMBOLIC SHADOWING

As we have seen, cylinder sets of a generating partition become smaller with increasing string lengths. This property of generating partitions can be used to search for them using the *symbolic shadowing* algorithm. In the symbolic shadowing algorithm, each cylinder set $\mathcal{C}(s^n_{-m})$ corresponding to a partition of the attractor is associated with a unique point, $\boldsymbol{r}(s^n_{-m}) \in \mathbb{R}^d$, called its *representative*. As cylinders themselves form a partition of the attractor, the set of representatives can be viewed as a piecewise constant approximation of points in them. Thus, each data point of the time series $\boldsymbol{x}_i$ is *approximated* by a unique representative $\boldsymbol{r}(X^{i+n}_{i-m})$ corresponding to a symbolic encoding $\{X_i, i = 1, \ldots, N\}$ of the original time series obtained with a given estimate of the generating partition. In the symbolic shadowing algorithm, the estimated symbolic encoding and the corresponding set of representatives are iteratively updated so that the error of approximation between the given trajectory in the phase space and this *shadowing* trajectory formed by the representatives is systematically reduced.

More precisely, the problem of estimating a generating partition is formulated as the following minimization problem at each string length $l = m + n + 1$ [2]:

$$\min_{\{X_i\};\{\boldsymbol{r}(X^{i+n}_{i-m})\}} \sum_{i=m+1}^{N-n} \left\|\boldsymbol{x}_i - \boldsymbol{r}\left(X^{i+n}_{i-m}\right)\right\|^2, \tag{8}$$

where the minimization takes place over both the encoded sequence of visited cells and the set of representatives, as defined below. The cost function in the above is referred to as the *discrepancy* between a given time series and its shadowing. Considering any estimated partition, for fixed l and m, representatives are located at the center of mass of the corresponding cylinders, i.e.,

$$\boldsymbol{r}\left(s^n_{-m}\right) = \frac{1}{\left|\mathcal{C}\left(s^n_{-m}\right)\right|} \sum_{\boldsymbol{x}\in\mathcal{C}(s^n_{-m})} \boldsymbol{x}, \tag{9}$$

where $|\mathcal{C}(s^n_{-m})|$ is the number of points in a cylinder $\mathcal{C}(s^n_{-m})$. In fact, given a symbolic encoding of an ergodic system with invariant measure μ, the above choice of representatives minimizes the *expected squared error* $\mathcal{H}(l, m)$ given by

$$\mathcal{H}(l,m) \equiv \int \left\|\boldsymbol{x} - \boldsymbol{r}\left(X^n_{-m}\right)\right\|^2 d\mu(\boldsymbol{x}) \tag{10}$$

$$= \lim_{N\to\infty} \mathcal{H}_N(l,m), \tag{11}$$

where

$$\mathcal{H}_N(l,m) = \frac{1}{N-m-n} \sum_{i=m+1}^{N-n} \left\|\boldsymbol{x}_i - \boldsymbol{r}\left(X^{i+n}_{i-m}\right)\right\|^2. \tag{12}$$

Furthermore, if as $m \to \infty$ (which implies $l \to \infty$) $\mathcal{H}(l, m)$ approaches zero, then the estimated symbolic encoding is a generating partition. This justifies the iterative minimization of Eq. (8) carried out by the symbolic shadowing algorithm.

Since the minimization takes place over both the encoded sequence and the set of representatives, it is done in two steps. First, starting with some *initial* partition $\mathcal{P}^j$ (at say, iteration j), representatives at fixed string length l and $m = \lfloor l/2 \rfloor$ are found using Eq. (9) so as to minimize $\mathcal{H}(l, m)$ estimated from a long time series, i.e., $\mathcal{H}_N(l, m)$ with large N using Eq. (12). In the second step, given a set of representatives for fixed l and m, a Voronoi tessellation of the phase space is constructed to obtain a new partition (or, equivalently, a new symbolic encoding) of data as follows. For a given set of *generator points*, a Voronoi tessellation of the space consists of as many cells as there are generator points. A Voronoi cell corresponding to a particular generator point collects all points that are closer to *that* generator point than to any other. For our purposes, we consider the representatives at given l and m as the generator points. Then, a Voronoi cell $\mathcal{T}(s^n_{-m})$, for each string s^n_{-m} obtained from the previous encoding sequence, is defined as a set of points that are closest (considering Euclidean distance) to its representative $\boldsymbol{r}(s^n_{-m})$, i.e.,

$$\mathcal{T}\left(s^n_{-m}\right) \equiv \left\{\boldsymbol{x} \in \Gamma : \left\|\boldsymbol{x} - \boldsymbol{r}\left(s^n_{-m}\right)\right\| < \left\|\boldsymbol{x} - \boldsymbol{r}\left(\tilde{s}^n_{-m}\right)\right\|, \right.$$
$$\left. \forall \tilde{s}^n_{-m} \neq s^n_{-m}\right\}, \tag{13}$$

which we empirically estimate from the time series as

$$\mathcal{T}(s_{-m}^{n}) = \{\boldsymbol{x}_i : \|\boldsymbol{x}_i - \boldsymbol{r}(s_{-m}^{n})\| < \|\boldsymbol{x}_i - \boldsymbol{r}(X_{i-m}^{i+n})\|, \; \forall\, X_{i-m}^{i+n} \neq s_{-m}^{n}, i = m+1, \ldots, N-n\}. \quad (14)$$

Finally, these Voronoi cells are combined to obtain a new estimate of the generating partition $\mathcal{P}^{j+1} = \{P_u^{j+1} : \forall\, u \in \mathcal{A}\}$ with its cells given by

$$P_u^{j+1} = \cup\{\mathcal{T}(s_{-m}^{n}) \,\forall\, s_{-m}^{n} : s_0 = u, u \in \mathcal{A}\}. \quad (15)$$

It is to be noted that if the original partition $\mathcal{P}^j$ is generating then its approximation $\mathcal{P}^{j+1}$ obtained by Voronoi tessellation will converge to a generating partition as string length $l \to \infty$ [2].

We are now able to state the original symbolic shadowing algorithm step by step:

(1) *Given*: time series in the continuous phase space, $\{\boldsymbol{x}_i, i = 1, 2, \ldots, N\}$; alphabet size, k; initial symbol sequence, $y_1^N \equiv \{y_1, y_2, \ldots, y_N\}$ over alphabet $\mathcal{A} = \{0, 1, \ldots, k-1\}$; initial string length, l_{initial}; maximum string length, $l_{\max}$; maximum number of iterations, `MaxIter`.
(2) *Initialize*: $\{X_i\} \leftarrow y_1^N$, $l \leftarrow l_{\text{initial}}$.
(3) Set $m = \lfloor l/2 \rfloor$ and $n = \lfloor (l-1)/2 \rfloor$. For each string s_{-m}^{n} observed in $\{X_i\}$, estimate cylinder $\mathcal{C}(s_{-m}^{n})$ and its representative $\boldsymbol{r}(s_{-m}^{n})$ according to Eqs. (4) and (9), respectively.
(4) For i from 1 to N, find representative $\boldsymbol{r}(s_{-m}^{n})$ which is closest to $\boldsymbol{x}_i$, and assign symbol $s_0 \in \mathcal{A}$ appearing at the center of the string s_{-m}^{n} to X_i.
(5) If the newly obtained symbol sequence $\{X_i\}$ at the end of step (4) is different from the previous one, and if the number of iterations is less than `MaxIter`, then go back to step (3).
(6) $l \leftarrow l + 1$. Go to step (3) if $l \leqslant l_{\max}$.

One could also stop the algorithm when either the mean squared error, $\mathcal{H}_N(l, m)$ [Eq. (12)], or the *maximum* squared error, $\max_i \|\boldsymbol{x}_i - \boldsymbol{r}(X_{i-m}^{i+n})\|^2$, drops below a certain prescribed threshold value. However, the correct choice for the threshold value of error is not known *a priori* and is usually set to some small value. In this work, we instead choose $l_{\max}$ and `MaxIter`, so that the algorithm finishes in a reasonable amount of time. As reported in [2], we have also seen periodic behavior in a few cases, where the algorithm alternates between different "solutions" (i.e., the set of representatives and the corresponding symbol sequence), and would do so indefinitely unless stopped (at `MaxIter`). In such cases, we choose a solution which has the least mean error among the other iterates.

A. Choice of initial partitions

In step (2) of the algorithm presented above, the symbolic shadowing algorithm is initialized by providing a symbol sequence y_1^N over the chosen alphabet $\mathcal{A}$. This can be done in several ways. One way is to use surrogates of low-period UPOs found from the time-series data. Each orbit point is assigned a symbol so that every distinct periodic orbit is coded with a unique symbol sequence. Then, each of the remaining points of the time series are assigned a symbol of the nearest encoded periodic point in the phase space, thus creating a partition of the data. However, surrogates of many of the most prominent lowest-period orbits in the system may not be found from the time-series data, as for many systems such orbits may lie outside the attractor. In such cases, one must look for relatively high period orbits, which becomes a difficult task, especially in the presence of noise in the data. Moreover, typically there are many possible initial partitions which could uniquely encode a given set of low-period UPOs (i.e., assign unique symbolic names to each orbit), some of which may not converge to good approximations of generating partitions when fed to the symbolic shadowing algorithm [2]. Thus, this process inherently requires some manual intervention, and the number of admissible choices of initial partitions can grow very quickly with the period of available UPOs.

In this work, we adopt a different approach by choosing a large ensemble of random initial partitions from the space of partitions. The main idea here is that we want to explore the space of partitions as much as possible (within some reasonable computational time limit) so that we better understand the nature of the algorithmic convergence, and can examine multiple approximations of the many generating partitions that are believed to exist. We use two methods, *random representatives* (RRs) and *random symbols* (RSs) to construct initial symbol sequences. In the RR method, we select a random sample of $k^{l_{\text{initial}}}$ points from the time-series data without replacement, and use its elements to form a set of representatives of string length l_{initial}. Each of these representatives is then assigned a symbol from an alphabet $\mathcal{A}$ (of size k), uniformly randomly, so that exactly $k^{l_{\text{initial}}-1}$ representatives have the same symbol. Finally, each data point of the time series is assigned the symbol of its nearest representative to obtain an initial encoded sequence y_1^N. In the RS method, each data point of the time series is uniformly randomly assigned a symbol from $\mathcal{A}$ to directly obtain y_1^N. As can be seen, these methods are objective in that no particular knowledge of the system or its dynamics in phase space are used while forming the initial symbol sequences. In what follows, we used both the RR and RS methods to generate an ensemble of 1008 initial symbol sequences over different alphabet sizes, which were made to evolve under the symbolic shadowing algorithm. We next discuss the results of applying the symbolic shadowing algorithm with these initial sequences for the driven, two-well Duffing oscillator.

B. Driven Duffing oscillator

Let us consider the two-well, periodically driven Duffing oscillator,

$$\ddot{x} + \delta\dot{x} - x + x^3 = \gamma\cos(\omega t), \quad (16)$$

where overdots denote time derivatives. We fix $\gamma = 0.4$, $\omega = 1$, and $\delta = 0.25$. Moon and Holmes ([6], pp. 82–91) have shown that this ODE serves as a good model to describe periodically forced steady state vibrations of a cantilever beam in its first mode when placed in a nonuniform field created by two permanent magnets near its free end. For our calculations, we exploit the fact that the Duffing equation is invariant under the transformation $(x, \dot{x}, t) \to (-x, -\dot{x}, t + \pi/\omega)$. Thus, points

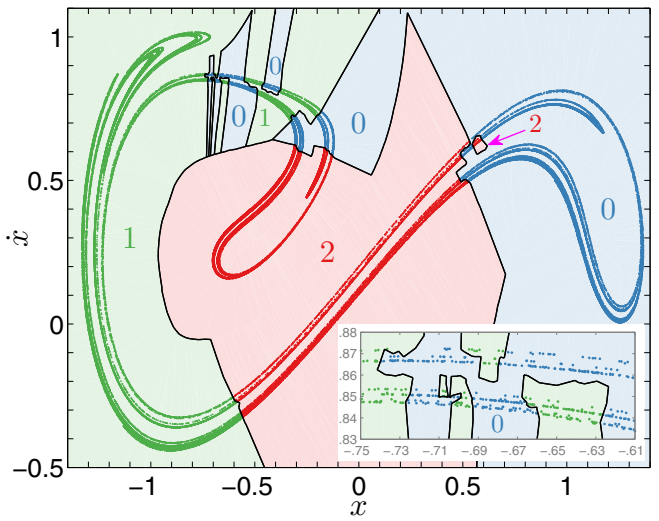

FIG. 4. The most frequently obtained three-cell partition of the Duffing attractor, generated by application of the symbolic shadowing algorithm to thousands of random initial symbol sequences. Note that this *does not* match the generating partition of Fig. 1, a limitation addressed by our modification to the original algorithm. The cells of the partition are a union of multiple, apparently disjoint regions in the phase space. Inset: An enlarged view of one area in the phase space where the cells for symbols {0} and {1} are intermingled in a complicated way.

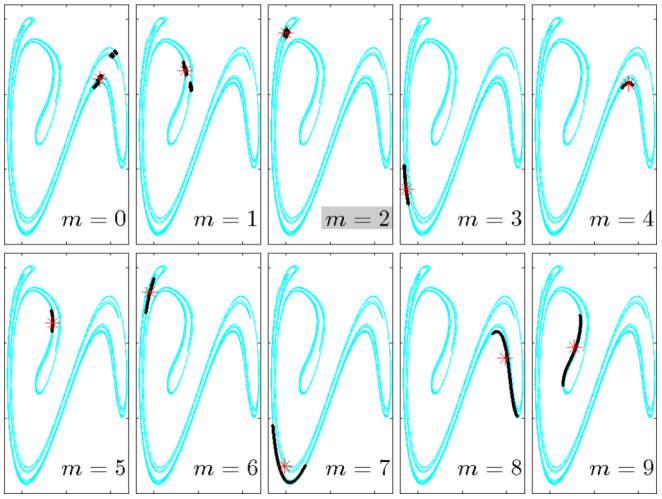

FIG. 5. Cylinder sets $\mathcal{C}(s_{-m}^{l-m-1})$ for $m = 0, \ldots, 9$ and $l = 10$, corresponding to the symbolic itinerary $\{0, 2, 1, 1, 0, 2, 1, 1, 0, 2\}$, the most probable itinerary (probability ≈ 0.0209) among all observed length-10 strings obtained from the generating partition (Fig. 1), overlaid on the Duffing attractor. Sets were estimated from a time series of four million points. As the number of backward symbols m change from 0 to 9, we get cylinders (*heavy dots*) of different shapes and sizes due to stretching and folding on the Duffing attractor. Also indicated, by the symbol ∗, is each cylinder's representative $\boldsymbol{r}(s_{-m}^{l-m-1})$ [Eq. (9)]. Observe that the cylinder size is smallest, and is best approximated by its representative, when $m = 2$.

$\boldsymbol{x} = (x, \dot{x})^\intercal$ in the continuous Poincaré section Γ are acted upon by the map $\boldsymbol{F}_\pi$, constructed by numerically integrating Eq. (16) for half the forcing period, i.e., over the interval $[0, \pi/\omega]$, followed by an inversion of coordinates: $x \to -x$ and $\dot{x} \to -\dot{x}$. Thus, two iterates of $\boldsymbol{F}_\pi$ are required for each full forcing cycle, so that $\boldsymbol{F}_\pi$ is the "square root" of the Poincarè map $\boldsymbol{F} : \Gamma \mapsto \Gamma$ [Eq. (1)], which is therefore obtained by composition as $\boldsymbol{F} \equiv \boldsymbol{F}_\pi \circ \boldsymbol{F}_\pi$ [28]. This system at the chosen parameter values is known to be chaotic and admits the three-symbol generating partition of its strange attractor shown in Fig. 1.

Using the above approach, we generated a chaotic time series of length $N = 10^5$ for the Duffing oscillator. For the numerical integration, we employed a fifth-order solution of the embedded Runge-Kutta method with Cash-Karp parameters [29,30]. We found, after a comparison of fourth and fifth order solutions at several thousand points on the attractor, that a fixed step-size implementation with 2000 time steps in each forcing period yielded a solution with absolute and relative errors below 10^{-8}. We then applied the symbolic shadowing algorithm to the time series, with the following parameters: $k = 3$; $l_{\text{initial}} = 2$; $l_{\max} = 10$; and `MaxIter` $= 100$. After running the algorithm on 1008 distinct initial symbol sequences generated using the RR method, we found only 46 *distinct* encoded sequences. Out of these encodings, the partition shown in Fig. 4 appeared most frequently (314 times out of 1008) and had satisfactory convergence so that the mean and maximum squared errors were 7.9951×10^{-4} and 7.2883×10^{-2} respectively. This partition was also the most frequent of the 37 *distinct* encodings obtained from the initial symbol sequences generated using the RS method, with 497 out of 1008 sequences converging to it. As can be seen, the partition cells consist of a union of multiple disjoint regions in the phase space. Also, this partition has a much more complicated boundary structure compared to the generating partition of Fig. 1, which has fewer and simpler cell boundaries. Furthermore, we found that *none* of more than two thousand initial symbol sequences, generated using both the RR and RS methods, converged to the generating partition of Fig. 1. One might attribute this failure to the choice of initial symbol sequences, in the sense that we might accidentally not have selected initial encodings that happen to converge to the generating partition under the action of the algorithm. However, after considerable experimentation, we discovered that the standard choice of m made in the original symbolic shadowing algorithm [see step (3)] has a strong effect on the algorithm's ability to find a generating partition of the Duffing attractor.

As we have seen in the discussion surrounding Fig. 3, the diameter of cylinder sets at each l corresponding to the generating partition of the Duffing attractor varies strongly with the choice of m. As an illustration, consider the cylinder sets corresponding to the length $l = 10$ symbolic itinerary $\{0, 2, 1, 1, 0, 2, 1, 1, 0, 2\}$, shown in Fig. 5 along with the cylinder set representatives [Eq. (9)]. For the delimited string, `02.110211002`, which corresponds to $m = 2$ we get the "smallest" cylinder so that all points within this cylinder are clustered together better than in cylinders for other m values in [0,9]. In fact, both mean and maximum squared errors as well as the cylinder diameter are minimized at $m = 2$ (see Fig. 6). However, the standard symbolic shadowing algorithm uses the cylinder corresponding to $m = \lfloor \frac{l}{2} \rfloor = 5$, which is clearly suboptimal. Since the main goal of the symbolic shadowing

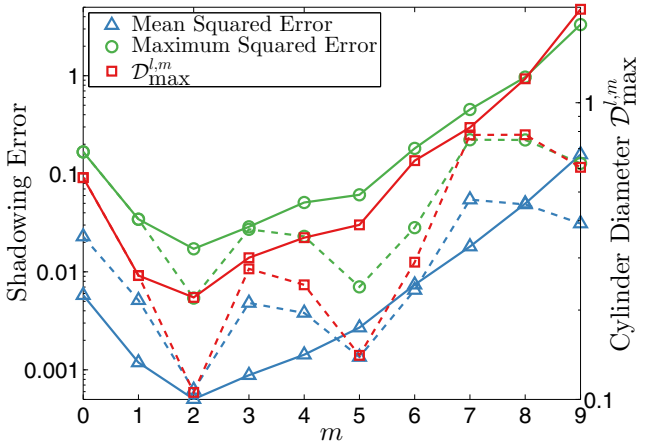

FIG. 6. Comparison of shadowing errors (i.e., mean and maximum squared errors) and the maximum cylinder diameter vs number of backward symbols m in delimited strings of length $l = 10$ for the generating partition (Fig. 1) of the Duffing attractor. The *solid lines* show values estimated from *all* of the cylinder sets over the attractor, while the *dotted lines* correspond to the cylinder sets (Fig. 5) of the symbolic itinerary {0, 2, 1, 1, 0, 2, 1, 1, 0, 2}. Both shadowing errors as well as the cylinder diameter are minimized at $m = 2$ suggesting an optimal choice for the formation of cylinder sets.

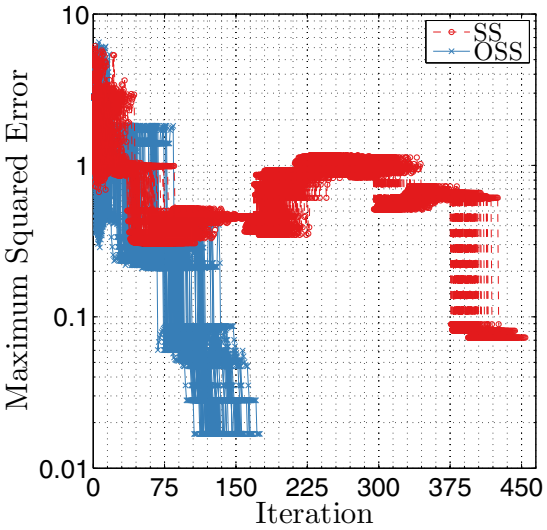

FIG. 7. Typical convergence traces of the maximum squared error for symbolic shadowing (SS) and optimized symbolic shadowing (OSS) algorithms starting from 175 initial symbol sequences generated using the RS method. For all these sequences, the optimized algorithm converged to the generating partition of Fig. 1 for which mean and maximum squared errors were 5.0012×10^{-4} and 1.6851×10^{-2} respectively, while the original algorithm converged to the partition of Fig. 4 with mean and maximum squared errors equal to 7.9951×10^{-4} and 7.2883×10^{-2} respectively. Also, the original algorithm took about thrice as many iterations as required by the optimized algorithm before it stopped.

algorithm is to minimize the discrepancy [Eq. (8)] between the phase space trajectory and its shadowing orbit passing through the cylinder representatives, it makes sense to choose a value of m at which cylinders are *best* approximated by their representatives. This observation motivates a modified symbolic shadowing algorithm, as discussed in the next section.

IV. OPTIMIZED SYMBOLIC SHADOWING

In light of the above discussion, we here present a modification to the original symbolic shadowing algorithm so that cylinder sets are formed in a way that renders their approximation by representatives optimal. Thus, instead of fixing the value of m at $\lfloor \frac{l}{2} \rfloor$ we choose it so that the mean squared error $\mathcal{H}_N(l, m)$ [Eq. (12)] is minimized at each l. We expect to yield the same value of m as m_{opt} in Eq. (7) since the mean squared error and the cylinder diameter behave similarly as m changes (see Fig. 6). The resulting optimized symbolic shadowing algorithm only requires the modification of step (3) of the original algorithm, as presented below: the rest of the algorithm remains the same and is thus omitted to highlight the modification.

(3) *Optimized step*:

(a) Set $m = 0$ so that $n = l - 1$.

(b) For each string s^n_{-m} observed in $\{X_i\}$, estimate the cylinder $\mathcal{C}(s^n_{-m})$ and its representative $\boldsymbol{r}(s^n_{-m})$ according to Eqs. (4) and (9), respectively.

(c) Calculate the mean squared error, $\mathcal{H}_N(l, m)$, according to Eq. (12).

(d) $m \leftarrow m + 1$ so that $n \leftarrow n - 1$. Go to step (3b) if $m \leqslant l - 1$.

(e) Set $m = m_{\text{opt}}$, where

$$m_{\text{opt}} \equiv \underset{0\leqslant m\leqslant l-1}{\text{argmin}}\ \mathcal{H}_N(l, m). \tag{17}$$

To test the optimized algorithm, we used the same time series (of length $N = 10^5$) of the Duffing oscillator as was used previously, along with the same algorithmic parameters: $k = 3$; $l_{\text{initial}} = 2$; $l_{\max} = 10$; and `MaxIter` $= 100$. After running the algorithm on 1008 distinct initial symbol sequences generated using the RR method, we found only 17 distinct encoded sequences. Out of these encodings, the partition shown in Fig. 1 was the most frequent (424 out of 1008) with mean and maximum squared errors equal to 5.0012×10^{-4} and 1.6851×10^{-2} respectively. This partition was also the most frequent among the 21 distinct encodings obtained from the initial symbol sequences generated using the RS method, with 426 out of 1008 sequences converging to it. To directly compare the numerical performance between the original and optimized symbolic shadowing algorithms, we used the same set of initial sequences generated using the RS method for both: Fig. 7 shows the convergence traces for 175 of these common sequences. On average, the optimized algorithm required a third of the iterations for convergence and achieved a lower maximum squared error as compared to the original algorithm.

Each iteration of the optimized symbolic shadowing algorithm finds a symbol sequence $\{X_i\}$ corresponding to each of the l values of m from 0 to $l - 1$, to identify the value of $m = m_{\text{opt}}$ that yields the best encoding. In contrast, for the original algorithm, encoding is found only once per iteration with $m = \lfloor \frac{l}{2} \rfloor$. Thus, it may initially seem that the execution time for the optimized algorithm is l times slower than that of the

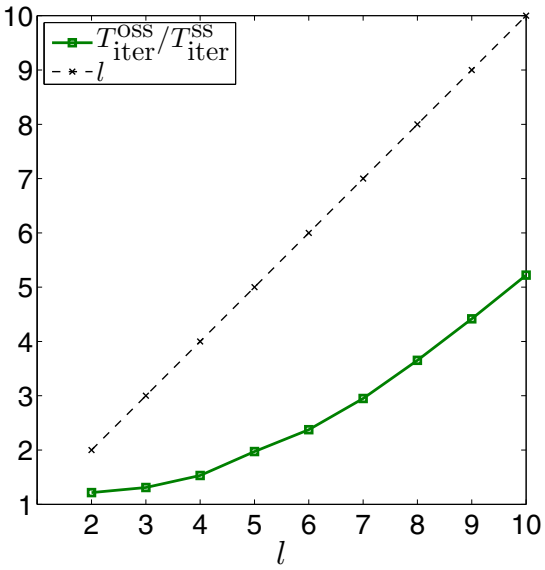

FIG. 8. Ratio of average execution time per iteration for the optimized symbolic shadowing algorithm (OSS) and the original algorithm (SS) with increasing l, estimated using convergence traces of Fig. 7. In the OSS algorithm, although l symbol sequences are found at each iteration, it does not run l times slower as compared to the SS algorithm. For the Duffing oscillator, at $6 < l \leqslant 10$, we have $T^{\rm OSS}_{\rm iter}/T^{\rm SS}_{\rm iter} \approx \epsilon l$, where $\epsilon \in (0.42, 0.53)$. A simple extrapolation, assuming ϵ increases linearly with l, leads to a prediction: $\epsilon_{\rm max} \approx 0.61$ at $l = 15$.

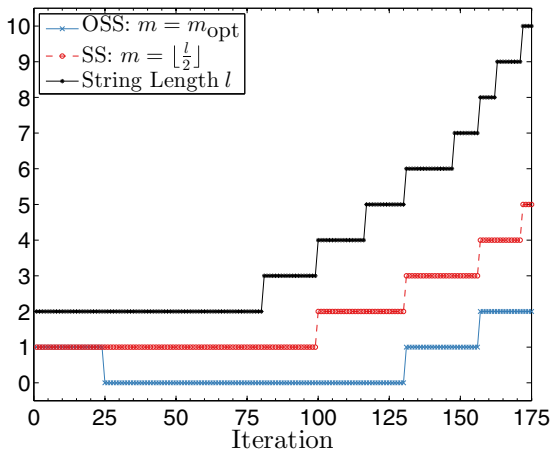

FIG. 9. Typical variation of values of $m_{\rm opt}$, estimated at each iteration, and l for one of the traces in Fig. 7 for the optimized algorithm (OSS). For reference, values of $m = \lfloor l/2 \rfloor$ for the original algorithm (SS) are also shown. As can be seen, the estimated $m_{\rm opt}$ is less than or equal to $\lfloor l/2 \rfloor$, dependent on l, and seldom changes from iteration to iteration at fixed l.

original. However, this is not the case since the determination of cylinder sets via Eq. (4) for different values of m [step (3b), above] at each iteration can be carried out quite efficiently using the shift map. In particular, since

$$\sigma\left(s^n_{-m}\right) = s_{-m} \ldots s_{-1} s_0 \cdot s_1 \ldots s_n,$$

we have

$$\mathcal{C}(s_{-m} \ldots s_{-1} s_0 \cdot s_1 \ldots s_n) = \left\{ \boldsymbol{x}_{i+1} : \boldsymbol{x}_i \in \mathcal{C}\left(s^n_{-m}\right), m+1 \leqslant i \leqslant N-n \right\}. \quad (18)$$

Indeed, for every given symbol string, we can obtain its cylinder at a particular value of m, say $\tilde{m}$, with $0 < \tilde{m} \leqslant l-1$, recursively as an *image*, under the action of the Poincaré map $\boldsymbol{F}$, of its cylinder at $m = \tilde{m} - 1$. We have found empirically that, for the values of l used here, each iteration of the optimized algorithm runs much faster than the "worst case scenario" of being l times slower than the original algorithm (see Fig. 8).

For the sequences considered in Fig. 7, on average the total execution time for each initial sequence was about 330 s for the optimized algorithm (running on a single Intel Xeon X5675 3.07 GHz processor), as compared to 293 s for the original algorithm. Furthermore, it is interesting to note that, as shown in Fig. 9, the values of $m_{\rm opt}$ found at any given l seldom change from iteration to iteration. As a consequence, with reference to Fig. 8, it seems evident that we could further save significant computational time for the optimized algorithm, especially at higher values of l, by choosing to optimize m at only the first few iterations and using that $m_{\rm opt}$ for subsequent iterations.

To further test our optimized symbolic shadowing algorithm, we used it to recover previously known two-cell generating partitions of the Hénon and Ikeda maps [2,11,17,27]. Computations were done using our implementations of both the original and optimized algorithms to allow for direct comparisons. For these discrete maps, there is a good match between partition boundaries as found by original and optimized algorithms [see Fig. 10 left, (a)–(c)]. Thus, it may seem puzzling that the original symbolic shadowing algorithm, unlike the optimized version, could not recover the three-cell generating partition (Fig. 1) of the Duffing oscillator. Moreover, we have also found a previously published [31] two-cell generating partition using the optimized algorithm for a chaotic Duffing oscillator at parameter values, $\delta = 0.65$, $\gamma = 1$ in Eq. (16). This partition [see Fig. 10 left, (d)] of the Duffing attractor in its Poincaré section (prepared as described in Sec. III B except with the forcing phase of the Poincaré section set to $\pi/2$) shows good agreement with the partition obtained by joining homoclinic tangencies (see Fig. 4 of [31]). In contrast, the original symbolic shadowing algorithm could not find this partition even for this simpler two-symbol case. However, using additional numerical results, we can provide an explanation for this apparently "system-specific convergence" issue of the original symbolic shadowing algorithm, one that is seemingly resolved by the optimized algorithm.

In Fig. 10 right, we compare the variation of maximum cylinder diameter $\mathcal{D}^{l,m}_{\rm max}$ as a fraction of the diameter of the attractor with the number m of the backward symbols in delimited length-15 strings using the aforementioned two-cell generating partitions. For the Hénon map, the cylinder diameter $\mathcal{D}^{l,m}_{\rm max}$ is minimized at multiple values of m, with $m_{\rm opt} \in \{2, 3, 4, 5\}$ for partition I [Fig. 10 left, (a)] and $m_{\rm opt} \in \{3, 4, 5\}$ for partition II [Fig. 10 left, (b)]. Note that the $\mathcal{D}^{l,m}_{\rm max}$ "curve" for partition II can be easily obtained from that of partition I by shifting the latter plot to the right by one symbol, which is not surprising given that partition II is a preimage of partition

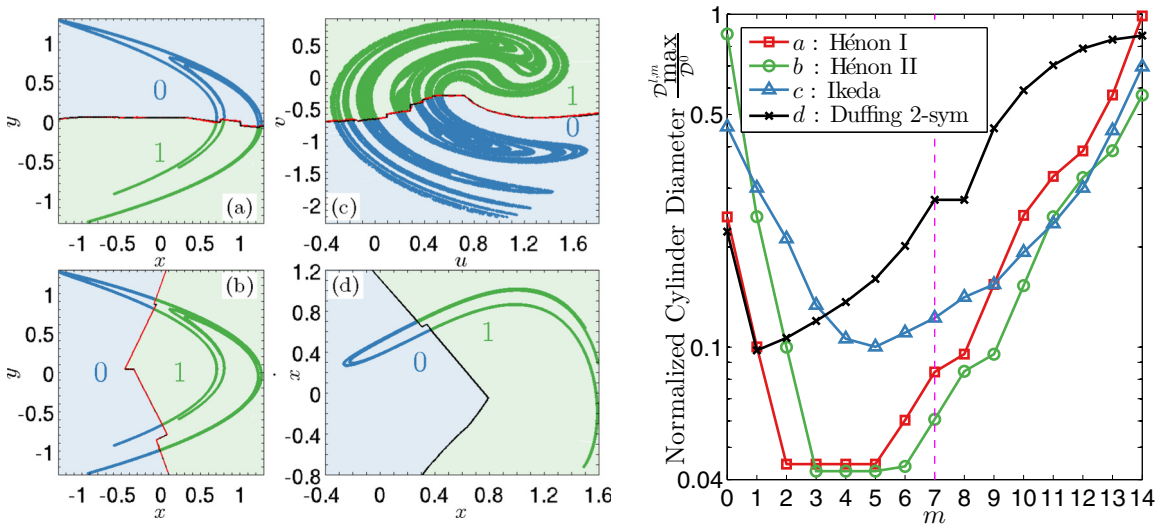

FIG. 10. Left: Two-cell generating partitions of the Hénon map (a), (b), Ikeda map (c), and the driven, two-well Duffing oscillator at an alternate set of parameters (d), respectively, found with optimized symbolic shadowing (cell boundaries shown with solid lines). For Heńon and Ikeda maps, cell boundaries for partitions obtained from the original algorithm (broken lines) lie almost exactly on top of those found using the optimized algorithm. As a result, the cell boundaries of these maps obtained using both the algorithms are virtually indistinguishable from each other in these figures. Right: Maximum cylinder diameter $\mathcal{D}^{l,m}_{\max}$ normalized by attractor diameter $\mathcal{D}^0$ vs number of backward symbols m in delimited strings of length $l = 15$, compared for all systems on left. The sampling of strings was done using $N = 10^6$ iterates on the attractor for each system. The vertical broken line corresponds to $m = \lfloor l/2 \rfloor = 7$. For the Hénon and Ikeda maps, there is only about 2–4% of the attractor diameter increase in $\mathcal{D}^{l,m}_{\max}$ for m values between $m = m_{\text{opt}}$ and $m = \lfloor l/2 \rfloor$, as opposed to an 18% increase for the Duffing oscillator.

I under the action of the Hènon map. More importantly, values of normalized $\mathcal{D}^{l,m}_{\max}$ at $m = \lfloor l/2 \rfloor = 7$ are only about 2–4% higher than the optimized values. This also holds for the Ikeda map, as the normalized $\mathcal{D}^{l,m}_{\max}$ increases from about 10% at $m_{\text{opt}} = 5$ to $\approx 12\%$ at $m = 7$. This indicates that, for the Hénon and Ikeda maps, the localization of points in the phase space using short delimited strings is not too sensitive to the specific location of the "string center," as long as the center is chosen to be not much greater than the middle of the string, which is precisely what occurs with the original symbolic shadowing algorithm. Thus, good approximations of the generating partitions obtained from the original and the optimized algorithm are expected to be very similar to each other for these maps. In contrast, for the two-cell generating partition of the Duffing attractor in Fig. 10 left, (d), the normalized increase in $\mathcal{D}^{l,m}_{\max}$ is significant, from about 10% at $m_{\text{opt}} = 1$ to $\approx 28\%$ at $m = 7$. This suggests that, in contrast with the Hénon and Ikeda maps, and as we found in our calculations, the optimization of m can significantly improve the convergence to the correct generating partition.

V. "MINIMAL" PARTITIONS AND LEMPEL-ZIV COMPLEXITY

As seen above, both the original symbolic shadowing algorithm and its optimized version can converge to multiple distinct partitions, depending on the initial partitions fed into them. Although the partitions of Figs. 1 and 4 appeared most frequently, using optimized and standard symbolic shadowing, respectively, when starting from a large ensemble of initial partitions, other partitions were obtained after satisfactory convergence to lower values of mean and maximum squared errors. As discussed in Sec. II C above, one can rank partitions based on their "generating property," i.e., their ability to consistently localize points in the phase space or decrease cylinder diameter $\mathcal{D}^l_{\max}$ with increasing l (Fig. 2). However, with a limited amount of data it could be misleading to pick the single "best" partition based on cylinder sizes alone due to inherent sampling fluctuations, especially at large l. In the limit of infinitely long symbol sequences, a true generating partition separates every given point in the phase space from all other points by splitting cylinder sets into sets of ever smaller size. Thus, it is at least plausible that the process of separation can be facilitated better for some rather convoluted partitions, that is, which have many more cell boundaries, among all candidate k-cell partitions. This is because the boundaries of cylinders are formed from the preimages and images of the boundaries of cells. Indeed, while it was the most probable, we found that the generating partition of Fig. 1, which has three, rather simple, contiguous cells, was not the one with the smallest value of $\mathcal{D}^l_{\max}$ among different partitions obtained using the optimized symbolic shadowing algorithm. Thus, there is a need for an additional reasonable criterion to further distinguish between different candidate generating partitions.

In our opinion, such a criterion should reflect what we here refer to as the "minimality" of a partition: among a candidate set of partitions which localize points in the phase space equally well, those with contiguous cells, fewer cell boundaries, and a smaller number of cells are preferred over their rivals. Apart from the obvious appeal of simplicity, such partitions seem advantageous in the presence of observation noise. With bigger, contiguous cells there are likely to be fewer instances of the misencoding of a noisy dynamical trajectory when the true state of the system lies near cell bound-

aries. Furthermore, such minimal partitions are "efficient" in terms of matching the information generation rate (i.e., the entropy rate) in the continuous phase space using fewer cylinders. As a consequence, in practice, using a finitely long encoded sequence we can estimate probabilities of cylinders or corresponding strings more accurately.

Unfortunately, however, it is a nontrivial issue to express this notion of partition minimality using some calculable measure. In the context of the symbolic shadowing algorithm, since at each string length l we get different partitions iteratively, choosing the iterate at the "best" l has been looked at as a model selection problem [22,27], especially for short, noisy time series. In this previous work, l has been selected using the Akaike or Bayesian information criteria, as well as by minimizing the description length, with the set of cylinder representatives treated as "model parameters." It also seems reasonable to extend these criteria to select among candidate k-cell partitions obtained at some large l (fixed *a priori*), after iterating distinct initial symbol sequences. Each of these criteria differ in terms of the relative importance placed on the model size and the prediction error, and can lead to different model choices.

Here, we propose to instead use the Lempel-Ziv (LZ) complexity [3] of encoded sequences, which we hypothesize indirectly quantifies the "minimality" of the corresponding partition, as characterized above. The LZ complexity equals the number of new subsequences encountered as the symbolic time series is parsed from left to right. It is high for "random" sequences, which is desirable here since the initial symbol sequences we used for the symbolic shadowing algorithms were randomly generated, i.e., with no consideration to the dynamics in the phase space. Here, we consider the *exhaustive* parsing of the sequence, $y_1^N \equiv y_1 y_2 \ldots y_{N-1} y_N$, into p phrases recursively, $y_{w_{(j-1)}+1}^{w_{(j)}}$ $(j = 1, 2, \ldots, p)$, where $w_{(0)} \equiv 0$, $w_{(1)} = 1$ and $w_{(p)} = N$:

$$y_1^{w_{(1)}} y_{w_{(1)}+1}^{w_{(2)}} \cdots y_{w_{(p-1)}+1}^{w_{(p)}},$$

such that $y_{w_{(j-1)}+1}^{w_{(j)}}$ $(1 < j < p)$ is the *shortest* such phrase which has not appeared anywhere in the subsequence $y_1^{w_{(j)}-1}$. The last phrase, $y_{w_{(p-1)}+1}^{N}$, may or may not be part of the subsequence y_1^{N-1}. The Lempel-Ziv complexity of a sequence, denoted henceforth as c_E, equals the number of phrases p in its uniquely defined exhaustive parsing. For example, a sequence $y_1^{20} = 10\,212\,102\,212\,221\,022\,102$, over the alphabet $\{0, 1, 2\}$, has the exhaustive parsing consisting of $p = 8$ phrases:

$$1 \mid 0 \mid 2 \mid 12 \mid 1022 \mid 122 \mid 2\,102\,210 \mid 2,$$

where the vertical lines separate individual phrases. Thus, we have $c_E(y_1^N) = 8$. Clearly, it is possible to employ various alternative schemes to parse a given sequence in many different ways. However, among all "iterative self-delimiting vocabulary-building processes" [3] which could be considered to produce a given sequence step by step, the exhaustive parsing scheme is minimal as it results in the least number of phrases. Moreover, for a stationary ergodic source from which a given sequence is assumed to arise, the appropriately normalized LZ complexity approaches the entropy rate [1,32] as the length of phrases and the sequence length N grows without bound, i.e.,

$$\limsup_{N\to\infty} \frac{c_E\left(y_1^N\right)}{N/\log_2 N} \leqslant h \quad \text{(almost surely)}, \tag{19}$$

where the entropy rate of the source h is expressed in bits. We note that, while the entropy rate is a property of the source itself, the LZ complexity characterizes individual sequences, i.e., particular realizations from a source. In fact, assuming an underlying unknown "finite-state" probabilistic source model, the individual sequences could be ranked, as $N \to \infty$, essentially according to their LZ complexity so that the most probable sequence emitted by the source would have the least complexity [32,33]. Furthermore, LZ complexity is relatively easy to calculate, and has been shown to be useful to characterize spatiotemporal patterns [34].

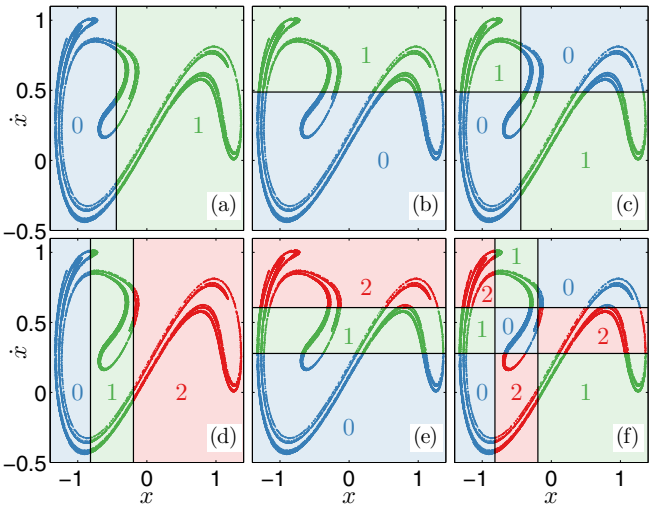

FIG. 11. Six examples of partitions of the Duffing attractor, with two cells [partitions (a)–(c)] and three cells [partitions (d)–(f)]. Symbols {0, 1, 2} label partition cells. Cell boundaries for partitions (a), (b) and (d), (e) were chosen so that each cell for a given partition had an equal number of data points falling in it. Furthermore, by construction, the boundaries of partition (c) are the union of those of partitions (a) and (b). Similarly, cell boundaries of (f) consist of boundaries of (d) and (e). Partitions (c) and (f) have noncontiguous cells which are disjoint unions of smaller boxes.

We now give some empirical evidence supporting our heuristic argument regarding the usefulness of LZ complexity for characterizing minimality of a partition. For this, we have considered six partitions, as shown in Fig. 11. By construction, four partitions, viz., (a), (b), (d), (e), consist of equiprobable boxes, i.e., of cells having the same number of data points falling in them. Partitions (a), (b), (c) have two cells labeled 0 and 1 while partitions (d), (e), (f) have three cells labeled 0, 1, and 2. Partitions (a), (b) and (d), (e) have contiguous cells defined by one and two boundaries respectively. In this case, by construction, we can specify these boundaries using simple inequalities with thresholds, say, $x < x_{\text{th}}$ or $\dot{x} > \dot{x}_{\text{th}}$. On the other hand, partitions (c) and (f), which have noncontiguous cells, require that their boundaries be specified by more than one condition of the form, say, $x_{\text{left}} < x < x_{\text{right}}$ or $\dot{x}_{\text{low}} < \dot{x} < \dot{x}_{\text{high}}$. Thus, from a standpoint of specification, we can argue that partitions

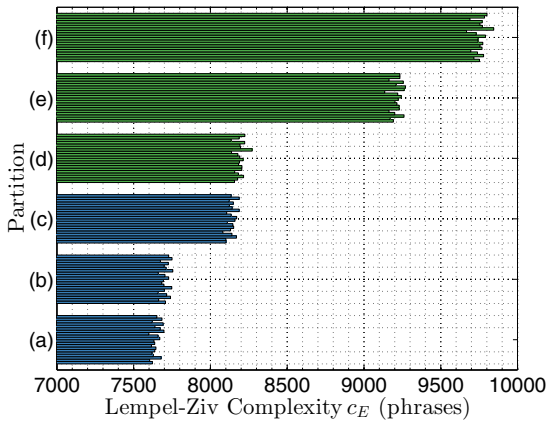

FIG. 12. Comparison of Lempel-Ziv (LZ) complexity c_E of partitions of Fig. 11. A time series of four million points of the Duffing oscillator was divided into 20 consecutive segments of length $N = 2\times10^5$ each. For each partition, these segments were encoded and their c_E values are as shown in the bar plot. Statistical fluctuations in c_E estimates are about 1–2%. Evidently, partitions (c) and (f), which have noncontiguous cells but use same number of symbols as partitions (a), (b) and (d), (e) respectively, are more complex in the LZ sense. Furthermore, three-cell partitions (d)–(f) are also more complex than two-cell partitions (a)–(c).

(c) and (f) are more "complex" (i.e., less "minimal") than partitions (a), (b) and (d), (e), respectively. Moreover, with additional cell boundaries we also yield new temporal patterns (i.e., symbol sequences) due to the "renaming" of some of the dynamical transitions. For instance, a dynamical transition, "0 → 1" (i.e., from cell "0" to cell "1") in partition (a), is "split" into four transitions, viz., "1 → 0," "1 → 1," "0 → 0," and "0 → 1"' in partition (c). As the LZ scheme parses the symbol sequence into distinct phrases, it is expected to build a larger vocabulary for sequences encoded using partition (c), by incorporating these new names of dynamical transitions, than it would with partitions (a) or (b).

Indeed, as shown in Fig. 12, the LZ complexity c_E of partitions (c) and (f) is higher than those of partitions (a), (b) and (d), (e) respectively. Also, as expected, two-cell partitions (a)–(c) have lower complexity than three-cell partitions (d)–(f). Clearly, in this discussion, based on their appearance in the phase space, we are treating partitions (a) and (b) or (d) and (e) as equally "complex" in a *specification* sense because they have the same number and type of cells (contiguous, equiprobable).

Now, it is clear by their very construction that the above partitions do not pay particular attention to the actual dynamics on the attractor, so that they can be expected to yield symbolic dynamics quite different from partitions with better "generating behavior." Likewise, because of this rather arbitrary relationship with the intrinsic dynamics of the system, they can have widely varying values of c_E, as is observed for the two seemingly similar three-symbol partitions (d) and (e). However, for our intended purpose, this will not be an issue because we will select partitions with minimal LZ complexity from a group of partitions which have already been determined to localize points in the phase space equally well, i.e., from a group of partitions with similar values of $\mathcal{D}^l_{\max}$ at a given l. That is, we propose to use c_E in *conjunction* with $\mathcal{D}^l_{\max}$ to assess partitions in a way that is analogous to a more conventional model selection approach: specifically, c_E can be interpreted as playing the role of "model size," while $\mathcal{D}^l_{\max}$ plays the role of "model error" since, as we have seen, it behaves similarly to the mean and maximum squared errors in the symbolic shadowing algorithms (see Fig. 6).

A. Revisiting the Duffing oscillator

We now compare distinct partitions obtained from the original and the optimized symbolic shadowing algorithms, in which we consider their generating property, characterized by the cylinder diameter $\mathcal{D}^l_{\max}$, in conjunction with their complexity (here taken as the opposite of "minimality") as revealed by the Lempel-Ziv complexity c_E. A time series of four million points of the Duffing oscillator was divided into 20 consecutive segments of length $N = 2\times10^5$ each. After encoding each of these segments using a given partition, values of c_E and $\mathcal{D}^l_{\max}$ (at $l = 12$) were calculated. Furthermore, we took the minimum of c_E and maximum of $\mathcal{D}^l_{\max}$ of these 20 values as our best estimates for each partition. The estimate of the cylinder diameter $\mathcal{D}^l_{\max}$ was then normalized using the diameter of the Duffing attractor, $\mathcal{D}^0 \approx 2.6995$, while the LZ complexity c_E was normalized using its asymptotic value for binary random sequences [3], which is $N/\log_2 N \approx 11\,357$ phrases. Note that using the *maximum* of the estimates of $\mathcal{D}^l_{\max}$ is appropriate here because, for fixed l, with increasing N the cylinder sizes cannot decrease. On the other hand, since the normalized LZ complexity approaches its asymptotic value from above as $N \to \infty$, taking the *minimum* of estimates of c_E makes more sense.

In Fig. 13, we show the results in the "generating-complexity plane" (normalized $\mathcal{D}^l_{\max}$ vs normalized c_E), for all three-cell partitions of the Duffing attractor that were obtained after running symbolic shadowing algorithms with both the RR and RS initialization methods. As can be seen, the distinct partitions output by the computations primarily fall into two groups: those that did not converge satisfactorily after running the symbolic shadowing algorithms, and those that did, with the latter indicated by the fact that they achieved $\mathcal{D}^l_{\max}/\mathcal{D}^0 < 0.1$, indicating that they are good approximations of generating partitions. To aid in interpretation, we have also constructed piecewise linear, nonincreasing, greatest lower envelopes for each of the four sets of solutions. Any partition on the envelope is "more generating" (i.e., has the smallest $\mathcal{D}^l_{\max}$) than other partitions which are "equally complex" (i.e., with same c_E). Similarly, any partition on the envelope is more "minimal" (i.e., has a lower LZ complexity c_E) than its rivals which are "equally generating" (i.e., having same $\mathcal{D}^l_{\max}$). In this sense, from a *given* set of solutions, a partition lying on the envelope is preferred over comparable rivals lying above and/or to its right in the normalized c_E - $\mathcal{D}^l_{\max}$ plane. The generating partition previously found in the literature, shown in Fig. 1, indicated by the up arrow in Fig. 13, lies on the envelopes for partitions obtained with optimized symbolic shadowing, while the partition of Fig. 4 (indicated by the left arrow) does not lie on any of the envelopes (although it is not too far away) because of the relatively high LZ complexity of its encoded sequences. Thus the use of these envelopes

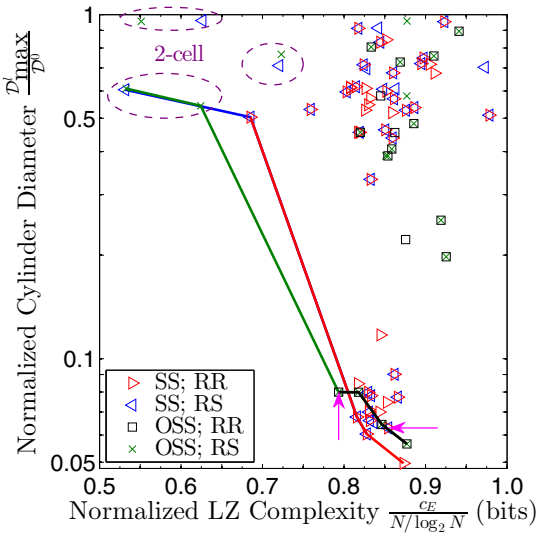

FIG. 13. For Duffing oscillator with parameters as for Figs. 1 and 4: comparison of three-cell partitions obtained from symbolic shadowing, presented in normalized $\mathcal{D}^l_{\max}$ vs c_E plane. All distinct solutions from four data sets obtained after applying both original (SS) and optimized symbolic shadowing (OSS). Initial symbol sequences generated using both RR and RS methods. Heavy lines indicate piecewise linear greatest lower envelope for each of the four sets of solutions. Up arrow and left arrow point to most frequently appearing partitions, as in Figs. 1 and 4, respectively.

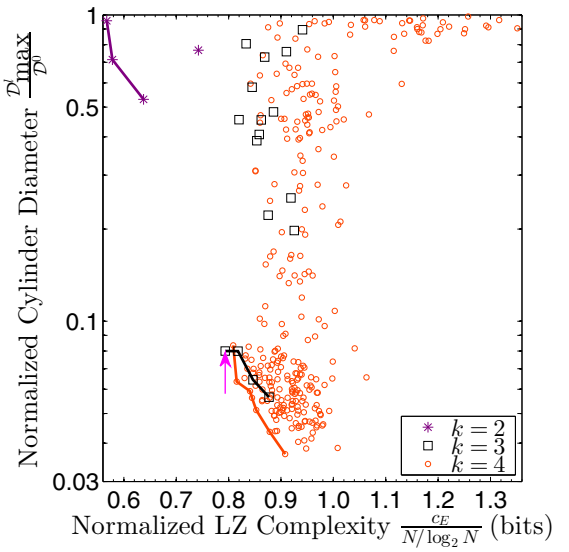

FIG. 14. For Duffing oscillator with parameters as in Fig. 13: comparison of partitions in the normalized c_E - $\mathcal{D}^l_{\max}$ plane, obtained with optimized symbolic shadowing applied to initial symbol sequences prepared using the RR method. All distinct solutions from three sets corresponding to alphabet sizes $k \in \{2, 3, 4\}$ of the initial sequences are shown. The up arrow points to the generating partition of Fig. 1.

in the normalized c_E - $\mathcal{D}^l_{\max}$ plane provides a systematic way to consistently select good partitions (that are minimal and generating) among a candidate set of generating partitions.

The symbolic shadowing algorithms do not in themselves tell us the number of symbols, k, that should be used to encode points in the phase space. It is thus desirable to have an objective method that allows researchers to examine the effect of alternative values of k. We therefore computed partitions using initial symbol sequences over alphabets with sizes $k = 2$ and $k = 4$ for the Duffing oscillator with parameters as in Fig. 13. We set $l_{\max} = 15$ for $k = 2$ and $l_{\max} = 8$ for $k = 4$: all other algorithmic parameters were identical to those used previously. After running the optimized symbolic shadowing algorithm on an ensemble of 1008 distinct initial symbol sequences generated using the RR method, we found only four distinct two-cell partitions as opposed to 287 distinct four-cell partitions. In Fig. 14, we have compared these partitions with the three-cell partitions in the normalized c_E - $\mathcal{D}^l_{\max}$ plane. As can be seen, the two-cell partitions fail to satisfactorily localize points in the phase space since the maximum cylinder diameter at $l = 12$ is comparable to the size of the Duffing attractor itself. On the other hand, there are many four-cell partitions which satisfactorily localize points in the phase space, similar to their three-cell counterparts, but that have marginally higher values of normalized LZ complexity. Although we do expect the number of distinct partitions found by symbolic shadowing applied on an ensemble of initial symbol sequences to increase with k, the scarcity of two-cell partitions and the abundance of four-cell partitions with low values of $\mathcal{D}^l_{\max}$ provides a convenient indication that the minimum number of symbols needed to find generating partitions for the Duffing attractor is 3.

B. Generating partitions for the van der Pol oscillator

We conclude our study by using symbolic shadowing algorithms to look for generating partitions for the periodically driven van der Pol oscillator studied by Shaw and others [23,35], which is governed by the following set of ODEs:

$$\dot{x} = 0.7y + 10x(0.1 - y^2), \quad \dot{y} = -x + 0.25\sin(\omega t), \quad (20)$$

where the forcing frequency ω is set to $\pi/2$. This system also exhibits a half-period symmetry similar to the Duffing oscillator since the above set of equations are invariant under the change of coordinates: $(x, y, t) \to (-x, -y, t + \pi/\omega)$. Thus, each iterate of the map, $\boldsymbol{F}_\pi$, in the continuous phase space is obtained by numerically integrating Eq. (20) for half the forcing period, i.e., over the interval $[0, \pi/\omega]$, followed by inversion of coordinates. As before, the Poincaré map is then given by $\boldsymbol{F} = \boldsymbol{F}_\pi \circ \boldsymbol{F}_\pi$. With these parameter values, the system admits a strange attractor and the corresponding dynamics is equivalent to a return map with three branches over an annulus [23]. As a result, its generating partition is expected to consist of three cells.

We employed the optimized symbolic shadowing algorithm developed in this paper to find candidates for a minimal generating partition, and compared our results to those obtained with the original, nonoptimized algorithm. For this purpose, we generated a time series of 2×10^5 iterations of the Poincaré map in the phase space using the same numerical method as for the Duffing oscillator, albeit with 1000 time steps in each external forcing period, as was adequate to keep absolute and relative one-step errors below 10^{-8}. After applying the symbolic shadowing algorithms to ensembles of

TABLE I. Application of original (SS) and optimized (OSS) symbolic shadowing algorithms to the van der Pol oscillator [Eq. (20)]: parameters and solution sets.

Solution set no.	1	2	3	4	5	6
Alphabet size k	3	3	3	3	2	4
Algorithm	SS	SS	OSS	OSS	OSS	OSS
Initialization method	RR	RS	RR	RS	RR	RR
Initial string length l_{initial}	2	2	2	2	2	2
Maximum string length $l_{\max}$	15	15	15	15	15	12
Maximum iterations at each l, `MaxIter`	500	500	500	500	500	500
Number of distinct solutions (from 1008 initial sequences)	43	24	62	38	1	572

initial symbol sequences we obtained six solution sets of distinct candidate partitions. The number of distinct partitions in each of the solution sets and the parameters of the algorithms used to generate them are given in Table I.

In Fig. 15, we show the partition selection statistics from the first four solution sets in Table I, plotted in the normalized c_E - $\mathcal{D}^l_{\max}$ plane. For the van der Pol oscillator, the greatest

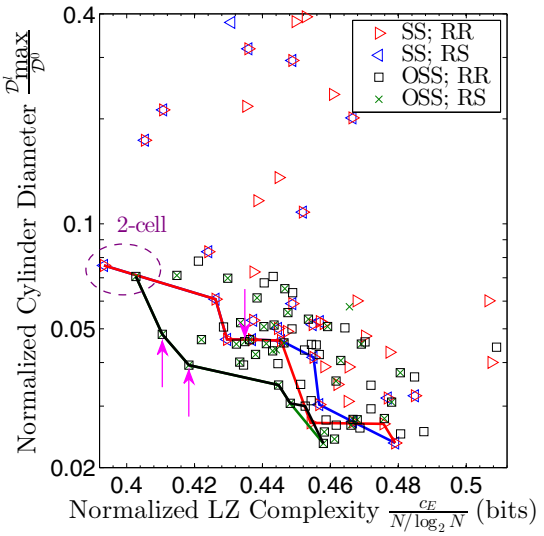

FIG. 15. van der Pol Oscillator: comparison of partitions obtained from symbolic shadowing algorithms in the normalized c_E-$\mathcal{D}^l_{\max}$ plane. All distinct solutions from set 1 to 4 (see Table I) are shown. All are three-cell partitions except for the encircled ones which have two cells. A time series of two million points of the van der Pol oscillator was divided into ten consecutive segments of length $N = 2\times10^5$ each. After encoding each of these segments using a given partition, values of c_E and $\mathcal{D}^l_{\max}$ (at $l = 20$) were calculated. The points shown correspond to minimum of c_E and maximum of $\mathcal{D}^l_{\max}$ over these ten values for each partition. The cylinder diameter $\mathcal{D}^l_{\max}$ was normalized using the diameter of the van der Pol attractor, $\mathcal{D}^0 = 1.6026$, while the LZ complexity c_E was normalized using its asymptotic value for binary random sequences, i.e., $N/\log_2 N \approx 11\,357$ phrases. Heavy lines constitute piecewise linear, nonincreasing, greatest lower envelope for each of the four sets of solutions. Up arrows, with increasing values of LZ complexity, point to partitions (a) and (b) respectively of Fig. 17, while the down arrow points to the partition (i) of Fig. 18.

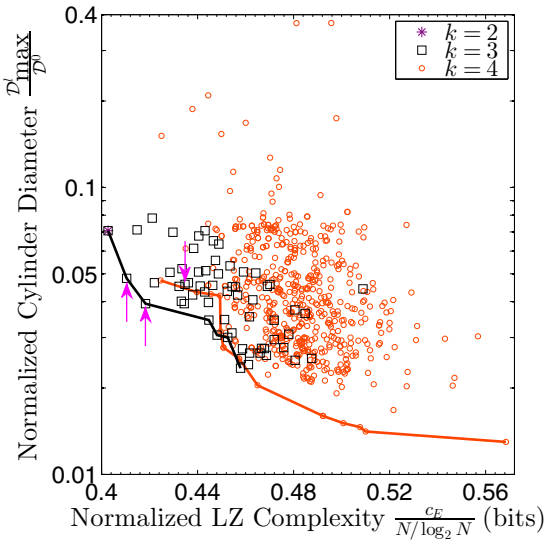

FIG. 16. van der Pol Oscillator: comparison of partitions in the normalized c_E-$\mathcal{D}^l_{\max}$ plane (similar to Fig. 15) obtained after applying the optimized symbolic shadowing algorithm. All distinct solutions from sets 3, 5, and 6 (see Table I) are shown. See caption of Fig. 15 for details on the calculation of c_E and $\mathcal{D}^l_{\max}$ values.

lower envelopes of solution sets obtained from optimized symbolic shadowing algorithm lie below their counterparts obtained with the original algorithm, again indicating the usefulness of the optimization in m proposed in this paper. As done for the Duffing oscillator, we also found partitions of the van der Pol attractor by initializing the algorithm using sequences over two and four symbols (Table I, solution sets 5 and 6). We found only one two-cell partition as opposed to 572 distinct four-cell partitions. This large jump in the number of partitions indicates that $k = 3$ gives an appropriate alphabet size for the system. In Fig. 16, we have compared the two and four-cell partitions with the three-cell partitions in the normalized c_E - $\mathcal{D}^l_{\max}$ plane. It is interesting to note that the two-cell partition has cells $\{P_0\}$ and $\{P_1 \cup P_2\}$, where $\{P_0, P_1, P_2\}$ are those of the three-cell partition (a) of Fig. 17. Unlike the Duffing attractor, the two-cell partition here is able to localize points down to $\approx 7\%$ of the size of van der Pol attractor, in comparison with 4–5% for the three-cell candidate generating partition shown in Fig. 17. This suggests that the two-cell partition under consideration can almost adequately capture the symbolic dynamics of the van der Pol oscillator. Furthermore, we have found hundreds of four-cell partitions, which are very good at localizing points in the phase space at the expense of higher values of LZ complexity when compared to their three-cell counterparts.

We pick out two partitions having three contiguous cells lying on the envelopes of the solution sets of the optimized algorithm as pointed out by up arrows in Figs. 15 and 16. These partitions, shown in Fig. 17, are very good candidates for the minimal generating partition of the van der Pol attractor because of their low complexity ("minimality") and small shadowing errors (see Table II). We emphasize that none of these partitions were found with the original symbolic shadowing algorithm: the value of m_{opt} for these partitions is much smaller than $\lfloor l/2 \rfloor$, signifying the role of

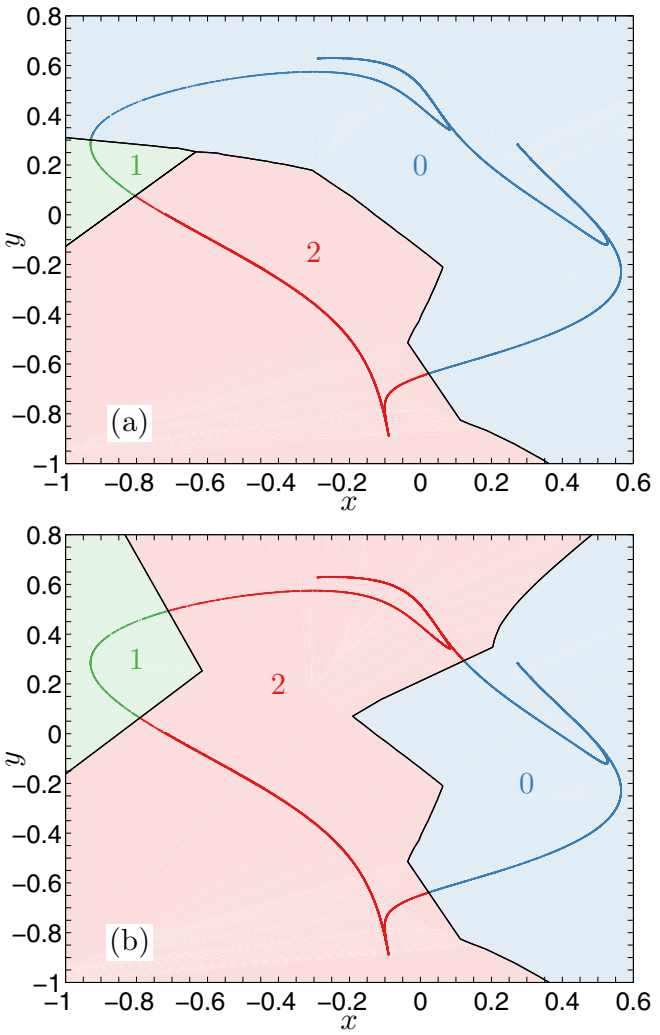

FIG. 17. Partitions of the van der Pol attractor obtained using optimized symbolic shadowing algorithm consisting of three contiguous cells labeled {0, 1, 2}.

the optimization subroutine. Given that we used a limited amount of data, the differences in complexity and cylinder diameters for these partitions are negligible, given that they are comparable to sample fluctuations. Thus, in the absence of any further information, it is difficult to select one of these partitions as being the "most" generating.

TABLE II. Postconvergence summary of the three-cell candidate generating partitions of the van der Pol attractor found using optimized symbolic shadowing algorithm.

Partitions (Figs. 17 and 18)	(a)	(b)	(i)
Mean squared error (10^{-5})	11.02	7.39	5.44
Maximum squared error (10^{-3})	10.15	5.52	9.40
m_{opt} at $l = 15$	2	2	1
Number of occurrences	49 (RR)	26 (RR)	70 (RR)
(out of 1008 initial sequences)	128 (RS)	46 (RS)	174 (RS)

C. Robustness of generating partition estimates for the van der Pol oscillator

We conclude this section by making few observations about the robustness of our partition estimates, particularly as it relates to the random partition initialization approach used for this paper and the subtle way it can interact with the precision of the numerical integration scheme. The end result of this discussion is another candidate generating partition for the van der Pol equation. The code used for this paper was written in MATLAB, with C code generated and used to create MATLAB-callable functions as required for speed and accuracy. All of our Duffing oscillator results were found to be highly robust: they were all run many times with different random initial partition ensembles, and it was found that for sufficiently small values of the integrator precision the statistics obtained, and the "best" partitions chosen according to our criteria, did not change. In addition, we found that changing the particular integration scheme (i.e., using our own solver vs a built-in MATLAB solver) had no significant effect. However, the van der Pol oscillator exhibited a more subtle behavior.

Consider partitions (i) and (ii) of Fig. 18, both found using the optimized symbolic shadowing algorithm, and which at first glance appear to be small perturbations of each other. Partition (i) is not "minimal": it has higher LZ complexity than partition (ii), due to the former's noncontiguous cells, as well as than either of the partitions in Fig. 17. Partition (i) was found using the same custom solver used for all of our previously presented driven oscillator results, as discussed in Sec. V B. In contrast, partition (ii) was obtained using time series generated using MATLAB's `ode45` solver with absolute and relative tolerances set at 10^{-8} and 10^{-6} respectively. However, in our early investigations of numerical robustness, we found that the MATLAB solver could not recover this partition when the absolute and relative tolerances were reduced to 10^{-9}, in which case the solution was very close in appearance to partition (i). Thus, there was some question as to whether or not partition (ii) in Fig. 18 should be considered a valid candidate. Indeed, it was this peculiar behavior, which we emphasize was only observed with the van der Pol equation, that motivated us to write our own solver. However, partition (ii), due to its simple structure, is appealing and if plotted in the normalized c_E-$\mathcal{D}^l_{\text{max}}$ plane corresponds to a point with normalized LZ complexity of 0.4116 bits and a normalized maximum cylinder diameter equal to 4.68%. As a consequence, it would fall *on the envelope* in Fig. 15, just to the right of the first up arrow corresponding to partition (a) of Fig. 17. Furthermore, we discovered that if we initialize the optimized algorithm at a partition with contiguous cells *close* to partition (ii), while starting at $l_{\text{initial}} = 8$ instead of 2 as done for all of our other computations, the algorithm would converge to partition (ii) by $l = 15$. Indeed, this is the very solution that is depicted in part (ii) of Fig. 18. This, then, suggests that partition (ii) is not a numerical artifact but is at least a "local" solution obtainable from the optimized algorithm (that is, for initial partitions that are sufficiently close). It is therefore a worthy generating partition candidate.

The above discussion suggests a potential limitation of our random partition initialization schemes. Such an approach

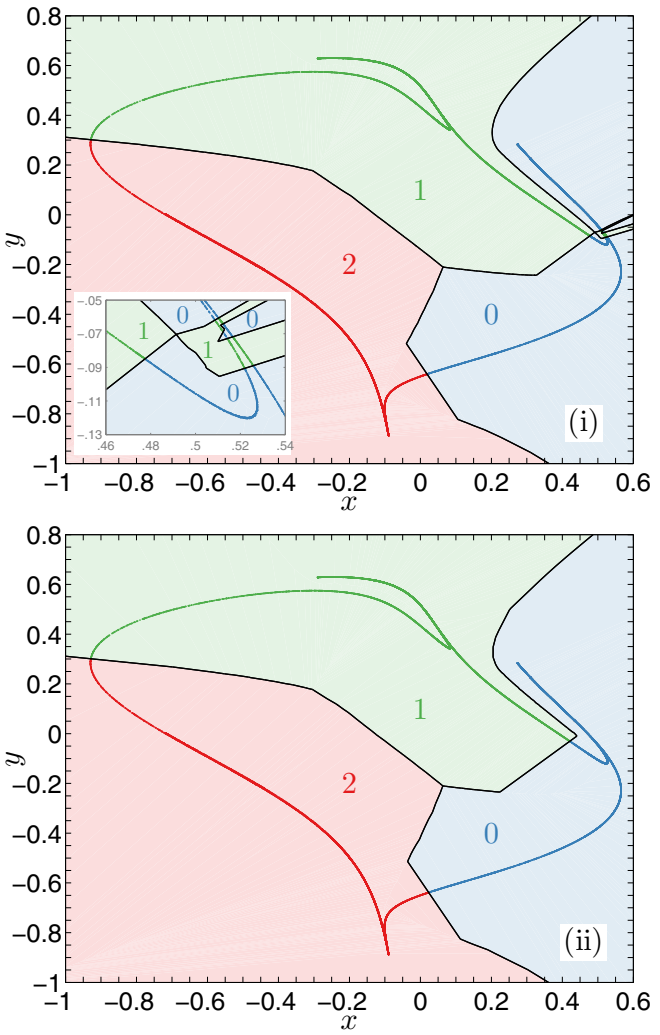

FIG. 18. Three-cell partitions of the van der Pol attractor found using optimized symbolic shadowing algorithm. Unlike partitions of Fig. 17, these solutions are sensitive to parameters of the numerical integration algorithm used to prepare the data set and also to some extent the initial string length l_{initial}. Partitions (i) and (ii) are very similar except in a small area around the boundary between cells {0} and {1}. An enlarged view of this area for partition (i) is shown in its inset.

helps to expand the search range of the algorithm, allows us to identify a multiplicity of possible algorithm outputs, and is useful for identifying an appropriate alphabet size k. However, we cannot at this stage conclude that such random initializations will exhaustively explore the partition space or find all of the "best" approximate generating partitions. Furthermore, the need to both start close to the partition of Fig. 18 (ii) and with longer initial string length suggests extremely random initial partitions *and* low initial string lengths can both inhibit convergence to minimal generating partitions (i.e., that localize states *and* have simple, contiguous cells). It would appear that in this case the small changes in numerical precision related to the different numerical solvers acted in a manner analogous to noise, in that the small differences "nudged" the algorithm toward a valid local solution.

VI. CONCLUSIONS

We have presented a modification of the symbolic shadowing algorithm [2] that can be used to find generating partitions in phase space from time series data alone. Our modified algorithm optimally approximates the cylinder sets corresponding to symbol strings without significantly compromising the simplicity of the algorithm. As with the original algorithm, every cylinder set is approximated by a representative point at its mass center in the phase space. However in our optimized algorithm the number of backward symbols used to construct symbol strings is adjusted to find the cylinder set with the smallest image in the phase space. This subtle modification results in partitions which are much better at localizing points in the phase space with increasingly long symbol sequences and hence are "more generating" than their counterparts found using the original algorithm. This allows us to find generating partitions for driven nonlinear oscillators, something not possible with the original algorithm. In fact, the previously found 3-cell generating partition of the two-well Duffing attractor [12,19] has been obtained in this work from time series data alone using our optimized symbolic shadowing algorithm.

A central issue in data-driven methods for approximating generating partitions is the "objectivity" of the process with respect to the initial guess input to any algorithm. The fact that, in theory, many generating partitions exist for a given chaotic attractor further complicates the problem of selecting the best choice from candidate partitions obtained via symbolic shadowing. To address this issue, we used a large ensemble of random symbol sequences as inputs to the symbolic shadowing algorithms to empirically explore the space of partitions. This resulted in many distinct partitions, some of which are good approximations of generating partitions in that they consistently yield smaller cylinders with increasing symbol string length. In addition, it is desirable to have partitions of minimal complexity, that is, that possess a simple arrangement of contiguous cells, and this provides another criterion that can assist in selecting among candidate partitions. To address this issue in a manner that can be applied objectively and consistently, we proposed the use of the Lempel-Ziv (LZ) complexity [3]. The LZ complexity represents the "growth of vocabulary" from left to right within symbol strings, and we argue heuristically that it indirectly characterizes minimal partitions by preferring partitions with contiguous cells, fewer cell boundaries and a fewer number of cells over their rivals. Using cylinder size in conjunction with Lempel-Ziv complexity we have been able to select the previously known, three-cell generating partition of the Duffing attractor from many potential candidates. The minimal generating partitions so selected yield small cylinders with increasing string lengths, and also have small values of Lempel-Ziv complexity. We argue that this represents a good solution to the "model error" versus "model size" trade-off. By varying the alphabet sizes of the input symbol sequences for the Duffing oscillator and analyzing the distinct solution sets we are able to show that two-cell (two-symbol) partitions are inadequate to capture the symbolic dynamics, while the very large number of four-cell partitions satisfactorily localize points in the phase space at the expense of higher LZ complexity. This provides an indication that a three-symbol alphabet is needed to find a minimal approximate generating partition for the Duffing oscillator.

This observation is useful given that the symbolic shadowing algorithm requires the user to input the alphabet size.

Last, we have applied the optimized symbolic shadowing algorithm to the driven van der Pol oscillator to estimate its generating partition. Following the methodology developed for the Duffing oscillator, we identified two partitions (Fig. 17) consisting of three cells as good approximations to generating partitions of the van der Pol attractor [35]. Furthermore, we were also able to obtain a three-cell partition with simple structure which appears as a "local" solution of the optimized algorithm and is a very good candidate for a generating partition.

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications (Wiley-Interscience, New York, 1991).

[2] Y. Hirata, K. Judd, and D. Kilminster, Estimating a generating partition from observed time series: Symbolic shadowing, Phys. Rev. E **70**, 016215 (2004).

[3] A. Lempel and J. Ziv, On the complexity of finite sequences, IEEE Trans. Inf. Theory **22**, 75 (1976).

[4] H. R. Lewis and C. H. Papadimitriou, *Elements of the Theory of Computation*, 2nd ed. (Prentice-Hall, Upper Saddle River, NJ, 1997).

[5] C. S. Daw, C. E. A. Finney, and E. R. Tracy, A review of symbolic analysis of experimental data, Rev. Sci. Instrum. **74**, 915 (2003).

[6] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Applied Mathematical Sciences (Springer-Verlag, New York, 1983), Vol. 42.

[7] R. Badii and A. Politi, *Complexity: Hierarchical Structures and Scaling in Physics*, Cambridge Nonlinear Science Series 6 (Cambridge University Press, Cambridge, UK, 1997).

[8] J. P. Crutchfield and N. H. Packard, Symbolic dynamics of noisy chaos, Physica D **7**, 201 (1983).

[9] Y. Kifer and B. Weiss, Generating partitions for random transformations, Ergodic Theory Dyn. Syst. **22**, 1813 (2002).

[10] P. Collet and J.-P. Eckmann, *Iterated Maps on the Interval as Dynamical Systems* (Birkhäuser Boston, Cambridge, 1980).

[11] P. Grassberger and H. Kantz, Generating partitions for the dissipative Hénon map, Phys. Lett. A **113**, 235 (1985).

[12] F. Giovannini and A. Politi, Homoclinic tangencies, generating partitions and curvature of invariant manifolds, J. Phys. A: Math. Gen. **24**, 1837 (1991).

[13] A. Politi, Symbolic encoding in dynamical systems, in *From Statistical Physics to Statistical Inference and Back*, edited by P. Grassberger and J.-P. Nadal, NATO ASI Series (Springer, Dordrecht, 1994), Vol. 428, pp. 293–309.

[14] F. Giovannini and A. Politi, Generating partitions in Hénon-type maps, Phys. Lett. A **161**, 332 (1992).

[15] L. Jaeger and H. Kantz, Structure of generating partitions for two-dimensional maps, J. Phys. A: Math. Gen. **30**, L567 (1997).

[16] R. Badii, E. Brun, M. Finardi, L. Flepp, R. Holzner, J. Parisi, C. Reyl, and J. Simonet, Progress in the analysis of experimental chaos through periodic orbits, Rev. Mod. Phys. **66**, 1389 (1994).

[17] R. L. Davidchack, Y.-C. Lai, E. M. Bollt, and M. Dhamala, Estimating generating partitions of chaotic systems by unstable periodic orbits, Phys. Rev. E **61**, 1353 (2000).

[18] J. Plumecoq and M. Lefranc, From template analysis to generating partitions I: Periodic orbits, knots and symbolic encodings, Physica D **144**, 231 (2000).

[19] J. Plumecoq and M. Lefranc, From template analysis to generating partitions II: Characterization of the symbolic encodings, Physica D **144**, 259 (2000).

[20] M. B. Kennel and M. Buhl, Estimating Good Discrete Partitions from Observed Data: Symbolic False Nearest Neighbors, Phys. Rev. Lett. **91**, 084102 (2003).

[21] M. Buhl and M. B. Kennel, Statistically relaxing to generating partitions for observed time-series data, Phys. Rev. E **71**, 046213 (2005).

[22] Y. Hirata and K. Aihara, Estimating optimal partitions for stochastic complex systems, Eur. Phys. J.: Spec. Top. **222**, 303 (2013).

[23] R. Gilmore and M. Lefranc, *The Topology of Chaos: Alice in Stretch and Squeezeland*, 2nd ed. (Wiley-VCH, Weinheim, 2012).

[24] J.-P. Eckmann and D. Ruelle, Ergodic theory of chaos and strange attractors, Rev. Mod. Phys. **57**, 617 (1985).

[25] D. J. Rudolph, *Fundamentals of Measurable Dynamics: Ergodic Theory on Lebesgue Spaces* (Oxford University Press, New York, 1990).

[26] E. M. Bollt, T. Stanford, Y.-C. Lai, and K. Życzkowski, What symbolic dynamics do we get with a misplaced partition?: On the validity of threshold crossings analysis of chaotic time-series, Physica D **154**, 259 (2001); Validity of Threshold-Crossing Analysis of Symbolic Dynamics from Chaotic Time Series, Phys. Rev. Lett. **85**, 3524 (2000).

[27] Y. Hirata, Symbolic Reconstruction of Dynamical Systems, Ph.D. Dissertation, School of Mathematics and Statistics, The University of Western Australia, Crawley, Australia, 2003.

[28] H. G. Solari and R. Gilmore, Organization of periodic orbits in the driven Duffing oscillator, Phys. Rev. A **38**, 1566 (1988).

[29] J. R. Cash and A. H. Karp, A variable order Runge-Kutta method for initial value problems with rapidly varying right-hand sides, ACM Trans. Math. Software **16**, 201 (1990).

[30] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, New York, 1992).

[31] F.-G. Xie, W.-M. Zheng, and B.-L. Hao, Symbolic dynamics of the two-well Duffing equation, Commun. Theor. Phys. **24**, 43 (1995).

[32] J. Ziv and A. Lempel, Compression of individual sequences via variable-rate coding, IEEE Trans. Inf. Theory **24**, 530 (1978).

[33] M. J. Weinberger, J. Ziv, and A. Lempel, On the optimal asymptotic performance of universal ordering and of discrimination of individual sequences, IEEE Trans. Inf. Theory **38**, 380 (1992).

[34] F. Kaspar and H. G. Schuster, Easily calculable measure for the complexity of spatiotemporal patterns, Phys. Rev. A **36**, 842 (1987).

[35] R. Shaw, Strange attractors, chaotic behavior, and information flow, Z. Naturforsch. A **36**, 80 (1981).