# Space-time nature of causality

Ezequiel Bianco-Martinez, and Murilo S. Baptista

# Space-time nature of causality

Ezequiel Bianco-Martinez[1,2] and Murilo S. Baptista[1,a]

[1]*Institute of Complex Sciences and Mathematical Biology, University of Aberdeen, SUPA, AB24 3UE Aberdeen, United Kingdom*
[2]*ZorgDomein, Straatweg 68, 3621BR Breukelen, The Netherlands*

In a causal world the direction of the time arrow dictates how past causal events in a variable $X$ produce future effects in $Y$. $X$ is said to cause an effect in $Y$, if the predictability (uncertainty) about the future states of $Y$ increases (decreases) as its own past and the past of $X$ are taken into consideration. Causality is thus intrinsic dependent on the observation of the past events of both variables involved, to the prediction (or uncertainty reduction) of future event of the other variable. We will show that this temporal notion of causality leads to another natural spatiotemporal definition for it, and that can be exploited to detect the arrow of influence from $X$ to $Y$, either by considering shorter time-series of $X$ and longer time-series of $Y$ (an approach that explores the time nature of causality) or lower precision measured time-series in $X$ and higher precision measured time-series in $Y$ (an approach that explores the spatial nature of causality). Causality has thus space and time signatures, causing a break of symmetry in the topology of the probabilistic space, or causing a break of symmetry in the length of the measured time-series, a consequence of the fact that information flows from $X$ to $Y$. *Published by AIP Publishing.* https://doi.org/10.1063/1.5019917

In a causal world the direction of the time arrow dictates how past causal events produce future effects. The determination of the direction and the intensity of the arrow of influence, causality, is one of the first questions one tries to answer in order to model a system. In ecology, it is fundamental to understand whether zooplankton concentration drives fish population. In meteorology, one wishes to determine whether and how surface sea temperature affects atmospheric temperature in different parts of the globe, or how green house gases drive global temperature. In finance, tax and expenditure correlates with saving and growth. In geology, one wants to access the direction of the flows of underground water from some measurements of water reservoir levels. In urbanism, one wants to understand how electricity consumption drives (or is driven by) urbanism or how building environment leads to obesity. Given the relevance of the topic, several methods have been developed in the last decades to study causality. Among them, there are the approaches that access causality based on informational quantities. They are sustained by the fundamental idea that if $X$ causes an effect in $Y$, then uncertainty about future states of $Y$ is reduced by considering the past of $Y$ and the past of $X$, a hypothesis that implicitly adopts the Granger causal idea that observations in the past of both $X$ (causing system) and $Y$ (where the effect is produced) can be used to predict the future state of $Y$. This work aims at unifying the Granger definition of causality defined in terms of predictability with those based on information quantities by studying the spatiotemporal dynamics of causality. We will show that if a system $X$ causes an effect in a system $Y$, then not only causal information from $X$ to $Y$ is positive but also longer-time or higher-resolution observations in $Y$ can be used to predict the past states of the system $X$, an observation that

will lead us to propose a new informational theoretic quantity that we name Causal Mutual Information (CaMI), and that can assist us in easily quantifying the direction of the flow of information. This work will show that causality has space and time signatures, and each signature can be advantageously exploited to study the direction of influence in different systems. Moreover, we will show that our quantity allows for a simple, experimental appealing and less computational demanding approach, but rigorous, quantification of causality.

---

## I. INTRODUCTION

The determination that a past event in a system has caused a present effect on another system provides a straight measurement of the direction of influence in these systems. Causal relationships between two events happening in two different systems $X$ and $Y$ can be established by verifying whether past events in $X$ and $Y$ influence future events in $Y$. Such understanding is fundamental to characterise, model, and predict behaviour in natural, social, and technological systems. The study of the cause-effect relationships is defined as causality. Causality is a concept that involves the temporal relationship among past, present, and future events of variables. Our studies show, however, that the temporal nature of causality from $X$ to $Y$ can not only be redefined in terms of the reduction of uncertainty from the variable $X$ solely based on observations in the past, present, and future, of $Y$ [a property that can be understood from the way Transfer Entropy (TE)[1] is defined, and also intuitively derived from Granger causality defined by the way past dynamical states directly and linearly influences dynamical future states], but also that causality can be defined in terms of the topological feature of the probabilistic space. In a deterministic system, two temporally related events defined

a) murilo.baptista@abdn.ac.uk

by two particular states of the system are also intrinsically related in space. This space-time ergodic duality in a deterministic system indicates that time-causality should lead to space-causality, a property that we will explore in this work to create novel ways to quantify causality. If there is a direct flow of information from $X$ to $Y$, this is physically interpreted as to that the uncertainty about past of $X$ is reduced by observations of future states of $Y$. This physical interpretation of causal information, and that is the core behind the definition of Granger causality, can be demonstrated by analysing the topology of the probabilistic space—exploring the spatiocharacter of causality, without the need to construct a model of the observed data, as it would be the usual procedure from the Granger approach.

When a perturbation affects a system, its influence is transmitted from the perturbation's source to the other variables of the system. The path the perturbation takes to propagate within the system can be predicted by analysing the causality of events in the system. A smart way to study causality is through a controlled experiment where perturbations can be designed to extract the causal structure of the system. However, the desired experiment could be too expensive, technically impossible to perform, or too invasive. Therefore, it is important to develop methods to identify the causal structure of a system only from observational data, without employing any perturbative technique. The identification of the causing and affected systems from observational data has been of great interest for many scientists. Consequently, and also because the identification of causality is fundamental to the effective observation, modelling, and controlling of any complex system, several techniques to infer and quantify causality have been recently proposed.[1–10]

Granger[4,11,12] considered that if a variable $x(t)$ causes an effect in $y(t)$, then predictions of $y(t)$ are improved by considering its own past complemented by the past of $x(t)$. Based on this assumption, he constructed statistical tests to validate this hypothesis. To adopt this hypothesis to study causality from data by constructing linear models, Granger causality introduces measures mostly used in correlation-based approaches (directed partial coherence):[13] Directed Coherence,[14] Partial Direct Coherence,[15] and direct Directed Transfer Function,[16] which are capable of identifying interactions in linear systems but are not suitable to detect causality among subsystems composing a non-linear system.[7] Due to this, methods based on Granger causality appropriated to detect causality in a nonlinear system were developed.[8,9] Granger causality can also be adopted by informational theoretical quantities, of special interest to us, such as the TE, Directed Information Theory,[13,17] Conditional Mutual Information,[5] Partial Transfer Entropy,[18] and Mutual Information from Mixed Embeddings (MIME),[19] which explore the intuitive notion that if system $X$ causes an effect in system $Y$, then, as specifically defined by transfer entropy, the amount of uncertainty in future values of Y is reduced by knowing the past values of X given past values of Y.

In this work, along the lines of the work in Ref. 13, we intend to unify the concept of causality based on the predictability of dynamical states introduced by Granger with the concept of causality defined in terms of transfer entropy, by considering the spatiotemporal character of causality. We will also introduce an unnoticed informational quantity that fully explores the space-time properties of causality, and that we call Causal Mutual Information (CaMI). $\text{CaMI}_{X\to Y}$ measures the total amount of information being transmitted from $X$ to $Y$, including both the information shared between both variables and that can be used to predict the present state of $X$ by observations in $Y$ (i.e., Mutual Information), and the causal information transmitted from $X$ to $Y$ and that can be used to predict the past states of $X$ by observations of the past and future states of $Y$ (i.e., Transfer Entropy).

Our proposed definition of causality, quantified by the quantity CaMI, is based on the physical notion that if $X$ causes effect in $Y$, then longer observations of the variable $Y$ (or alternatively higher resolution observations) than the one considered for $X$ can be used to predict the past of $X$. CaMI is calculated only by the probabilities of joint events, without the need of conditional events. This allows one to do a reduction in the dimensionality of the probability space used to quantify causality, resulting in a method that demands low computational power, and therefore allows for a quick assessment of causality. To illustrate why this is of any interest, notice that when calculating TE one needs to calculate conditional probabilities. However, conditional probabilities require the calculation of joint probabilities. CaMI only considers joint probabilities, sparing one from the need to further calculate conditional probabilities, but nevertheless a quantity that trivially provides the directionality index, the net transfer entropy between two systems, thus permitting the study of causality with less computational resources. Then, we show that the topology of the probabilistic space of joint probabilities determined by the shapes and forms of the partitions being generated by a dynamical process can also be used to not only visualize the spatiotemporal character of causality but also quantify causality, through the here defined CaMI. The emergence of the spatiotemporal nature of causality can be cumbersome or impossible to obtain with the usual higher-dimensional space of conditional probabilities, from which TE is calculated. A direct application of the topological properties of the joint probabilistic space is that it is not only possible to state about the direction of information between the measured variables but also to determine regions, the here called causal bubbles, that define ranges for the variables that are responsible for most of the information transmitted between two systems. The topology of this space can also be used to demonstrate that if there is a flow of information from the variable $X$ to $Y$, then it is also true that observations in $Y$ allows for an accurate prediction of location of the past dynamical states of variable $X$. As we shall see, there are preferential places and preferential times to measure the information being transmitted. Another advantage of our approach is that our detection of causality is oriented to treat experimental systems since our probabilistic space is based on partitions, which can be constructed based on the available experimental resolution of the data, or on the sampling rate of the measurements, allowing one to work with longer or short time-series. The usual causal analysis based on the value of TE calculated over equal-sized cells or from probabilities estimated by kernels would only provide a scalar number, with no information

about the dynamics behind the process or the topology of the probabilistic space.

## II. COUPLED MAPS

For this study of causality, we consider discrete coupled maps, whose connected nodes are described by

$$x_{n+1}^i = f(x_n^i, r)(1 - \alpha) + \frac{\alpha}{k_i} \sum_{j=1}^{M} A_{ij} f(x_n^j, r), \qquad (1)$$

where $x_n^i$ is the trajectory of map $i$, $n$ is the time index of the variable of the dynamics, $\alpha \in [0, 1]$ is the coupling parameter, $A_{ij}$ is the adjacency matrix (with entries of 1 or 0 depending on the existence of a connection between two nodes or not, respectively), $r$ is the fixed parameter of each map, $k_i$ is the degree of node $i$ [$k_i = \sum(A_{ij})$], and $f(x_n, r)$ is the map governing the dynamics that can be described by the Logistic map $f_1(x_n, r) = rx_n(1 - x_n)$.[20] A disconnected node ($k_i = 0$) is described by $x_{n+1}^i = f(x_n^i, r)$. We assume that there are $N_d$ maps forming the network.

Giving two variables $X$ and $Y$, we are interested in determining the direction of influence that one variable imposes over the other one. If $X$ influences $Y$, we represent this interaction by $X \rightarrow Y$.

## III. PARTITIONS, STATE AND PROBABILISTIC SPACE, AND SYMBOLIC TRAJECTORIES

In what follows, we consider that marginal observations are being made in the relevant variables, defining events in one variable by the falling of a trajectory point within an interval. This interval represents the resolution of the observer. These marginal observations and their probabilities will be used to calculate the probabilities to quantify causality. For simplicity in our analysis, we encode these partitions into symbols and treat the trajectories as symbolic trajectories.

### A. Order-*m* partitions, symbolic representation, and dynamics on it

Consider two discrete scalar time-series $\mathbf{X} = \{x_0, x_1, x_2, \ldots, x_{n-1}\}$ and $\mathbf{Y} = \{y_0, y_1, y_2, \ldots, y_{n-1}\}$ with $n$ elements, and $\Phi = \{\mathbf{X}, \mathbf{Y}\}$ defines a pair of variables taken from two subsystems in a complex network or coupled system. Therefore, a point in a 2-dimensional state space $\Omega^{XY}$ with coordinates $[\mathbf{X} \times \mathbf{Y}]$ representing the states of the subsystems $\Phi$ at time $t$ has coordinates $\{x_t, y_t\}$.

We define a marginal partition of order-$m$ of the coordinate $X$ as $\mathcal{C}^X(m)$, defined by the boundary curves $\mathcal{L}_X(m) = \{l_1^X(m), \ldots, l_r^X(m)\}$, which in this work are assumed to be straight lines, orthogonal to the direction of $X$. Then, this partition is composed of columns $\mathfrak{c}_i^X(m)$ where each one is separated from any other by one and only one curve $l_i^X(m) \in \mathcal{L}_X(m)$. Similarly, for the coordinate $\mathbf{Y}$, we can define a marginal partition $\mathcal{C}^Y(m)$, formed by rows $\mathfrak{c}_i^Y(m)$, enclosed by the set of boundary curves $\mathcal{L}_Y(m) = \{l_1^Y(m), \ldots, l_r^Y(m)\}$, which in this work are assumed to be straight lines, orthogonal to the direction of $Y$. Since we have a 2D time-series, we can construct a space partition $\mathcal{C}^{XY}(m)$ as a splitting of the space

$\Omega^{XY}$ formed by the union of the lines in $\mathcal{L}_X(m)$ and $\mathcal{L}_Y(m)$, so

$$\mathcal{C}^{XY}(m) = \mathcal{C}(m)^X \cup \mathcal{C}(m)^Y. \qquad (2)$$

Areas enclosed by the straight lines of $\mathcal{C}^{XY}(m)$ form the cells $\mathfrak{c}_i^{XY}(m)$ of the partition $\mathcal{C}^{XY}(m)$ that are encoded by the symbols $\mathfrak{s}_i^{XY}(m)$. $S^{XY}(m)$ represents all the possible symbols encoding cells in the partition of order-$m$.

The dynamics of points in this partition are represented by the transformation $\mathcal{U}_t$: $(x_{i+1}, y_{i+1}) = \mathcal{U}_t(x_i, y_j)$ and $\mathcal{U}_t^p(x_t, y_t) = (x_{t+p}, y_{t+p})$. The symbolic dynamics of points in this partition are regulated by the transformation $\mathcal{T}$, a surjective mapping of the states of variables in $\Omega^{XY}$ to a specific symbol in $S^{XY}(m)$. $\mathcal{T}$ provides a symbolic sequence $\Phi$ in the partition $\mathcal{C}^{XY}(m)$. From Eq. (1), $\mathcal{T}$ is the transformation that maps points from $\Omega^{XY}$ into itself, a 2D projection of the whole $N_d$-dimensional network.

Given the partition $\mathcal{C}^{XY}(m)$, we define a transition matrix $\Pi(m)$ where the element $\Pi(m)_{ij} = 1$ if the cell $\mathfrak{c}_i^{XY}(m)$ is the pre-image of the cell $\mathfrak{c}_j^{XY}(m)$ (i.e., there is a dynamical evolution from cell $\mathfrak{c}_i^{XY}$ to cell $\mathfrak{c}_j^{XY}$).

We define a *transition matrix of order-m [$\Pi\wp(m)$]* as follows:

$$\Pi\wp_{ij}(m) = \begin{cases} 1 & \text{if } F^m(\mathfrak{c}_i) \cap \mathfrak{c}_j \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases} \qquad (3)$$

A partition is defined as an order-$m$ if it generates a transition matrix of order $m$.

We adopt a partition defined by marginal probabilities because we want to define informational measures that quantify the predictability one has to predict the state of one variable by measuring only the state of the other variable, assuming that variables are being measured by a physical process, i.e., there is a measurement resolution.

### B. Probabilistic space and symbolic trajectory

Now, let us define a $L$ time-delay and time-forward coordinate system from which probabilities are calculated. The time-delay trajectory $\Phi_{-L}(t) = \{X_{-L}(t), Y_{-L}(t)\} = \{x_{t-L}, \ldots, x_{t-1}, y_{t-L}, \ldots, y_{t-1}\}$ represents a short segment with length $L$ (e.g., $L$ points) of the time-series $\Phi(t)$ taken for a time span between the integer time $t - L$ until the time $t - 1$, the time $t$ representing the time moment from where past and future are defined. By applying the transformation $\mathcal{T}$ to a segment of length $L$ of the time-series $\Phi_{-L}(t)$, we generate a sequence of symbols that represent the itinerary followed by the past length-$L$ trajectory. The trajectory points $\Phi_{-L}^\sigma(t)$ follow an itinerary along the partitions $\mathcal{C}^X(m)$ ($\sigma \equiv X$), $\mathcal{C}^Y(m)$ ($\sigma \equiv Y$), or $\mathcal{C}^{XY}(m)(\sigma \equiv XY)$, which are given by $\{\mathfrak{c}_{i_{t-L}}^\sigma(m), \mathfrak{c}_{i_{t-L+1}}^\sigma(m), \ldots, \mathfrak{c}_{i_{t-1}}^\sigma(m)\}$. If $\sigma = XY$, then $(x_{t-L}, y_{t-L}) \in \mathfrak{c}_{i_{t-L}}^{XY}(m)$. If $\sigma = X$, then $x_{t-L} \in \mathfrak{c}_{i_{t-L}}^X(m)$. The itinerary $\mathcal{C}_{-L}^\sigma(t, m) = \{\mathfrak{c}_{i_{t-L}}^\sigma(m), \mathfrak{c}_{i_{t-L+1}}^\sigma(m), \ldots, \mathfrak{c}_{i_{t-1}}^\sigma(m)\}$ is encoded by a symbolic sequence $S_{-L}^\sigma(t, m) = \{\mathfrak{s}_{i_{t-L}}^\sigma(m), \mathfrak{s}_{i_{t-L+1}}^\sigma(m), \ldots, \mathfrak{s}_{i_{t-1}}^\sigma(m)\}$, from which probabilities can be calculated. Similarly, the forward-time trajectory $\Phi_L^\sigma(t) = \{X_L(t), Y_L(t)\}$ follows an itinerary (or visits the sequence of cells) $\mathcal{C}_L^\sigma(t, m) = \{\mathfrak{c}_{i_{t+1}}^\sigma(m), \mathfrak{c}_{i_{t+2}}^\sigma(m), \ldots, \mathfrak{c}_{i_{t+L-1}}^\sigma(m)\}$ that is encoded by the symbolic sequence $S_L^\sigma(t, m) = \{\mathfrak{s}_{i_{t+1}}^\sigma(m), \mathfrak{s}_{i_{t+2}}^\sigma(m), \ldots, \mathfrak{s}_{i_{t+L-1}}^\sigma(m)\}$.

A $(k + m)$-order partition $\mathcal{C}^\sigma(k + m)$ is generated by the $k$-pre-iteration of the boundary curves composing the $m$-order partition $\mathcal{C}^\sigma(m)$. The pre-iteration is given by the evolution operator $\mathcal{U}^{-k}$. This order-$(k + m)$ partition is formed by the cells $\mathfrak{c}_i^\sigma(k + m)$. Notice that a cell $\mathfrak{c}_i^{XY}(k + m) \in \mathcal{C}^{XY}(m + k)$, with $\mathcal{C}^{XY}(m + k) \equiv \mathcal{U}^{-k}[\mathcal{C}^{XY}(m)]$ in an order-$(k + m)$ partition, represents points that follow a particular length-$L = k + m$ symbolic itinerary (or length-$L$ trajectory) in the order-1 partition. The probability measure of a length-$L$ itinerary $\mu[\{\mathfrak{c}_{i_{t-L}}^\sigma(1), \mathfrak{c}_{i_{t-L+1}}^\sigma(1), \ldots, \mathfrak{c}_{i_{t-1}}^\sigma(1)\}]$ is assumed to be equal to the probability measure of points in a cell of an order-$L$ partition and given by $\mu[\{\mathfrak{c}_i^\sigma(L)\}]$, with $x_{t-L} \in \mathfrak{c}_i^\sigma(L)$. Many length-$L$ trajectories can follow the same itinerary. The probability $P$ calculated over the symbolic sequence of a length-$L$ itinerary along the order-$m$ partition is represented by $P[\{\mathfrak{s}_{i_{t-L}}^\sigma(1), \mathfrak{s}_{i_{t-L+1}}^\sigma(1), \ldots, \mathfrak{s}_{i_{t-1}}^\sigma(1)\}]$. If the partition is generating, then this probability is also equal to the probability of points to belong to a cell $\mathfrak{s}_i(L)$ of an order $L$ partition and given by $P[\mathfrak{s}_i(L)]$, with $i$ such that $\mathcal{T}(X_{-L}) = \mathfrak{s}_i(L)$, where $\mathfrak{s}_i(L)$ is a length-$L$ symbolic sequence that gives the name of a cell in an order-$L$ partition.

Thus, there are two ways of calculating probabilities. One based on the probability of the trajectory itineraries, which produce the probability measures $\mu$. The other based on the symbolic itinerary, which produces the probability measures $P$. There is, however, a fundamental difference between both probabilities. Whereas $\mu$ is calculated over a higher $(m + k)$-order partition with non-overlapping well defined cells, and therefore, it requires the use of a generating partition, $P$ refers to the probability of a symbolic sequence defined by the marginal lower $m$-order original partition. It therefore does not require that the higher-order partition is generating. In our practical numerical calculations, we adopt the probabilities $P$ to calculate our informational quantities.

We assume in the following that the initial partition is order-1 ($m = 1$); therefore, there is only one straight line in $\mathcal{L}_X(1)$ and one straight line in $\mathcal{L}_Y(1)$. Each of these symbolic itineraries along the order-1 original partition can be encoded by the symbolic name of a cell in an order-$L$ partition. An event in $\mathcal{C}^{XY}(1)$ is defined by trajectory points falling in $\mathfrak{c}_i^{XY}(1) \in \mathcal{C}^{XY}(1)$.

## C. A generating partition

An order-$m + k$ partition $\mathcal{C}^\sigma(m + k)$ is generated from an order-m $\mathcal{C}^\sigma(m)$ by

$$\mathcal{C}^\sigma(m + k) = \mathcal{U}^{-k}[C^\sigma(m)]. \tag{4}$$

The partition $\mathcal{C}^\sigma(m)$ is a "generating" partition if all the cells in $\mathcal{C}^\sigma(m + k)$ are non-overlapping and the union of the partition boundary curves in an order-$(m + 1)$ partition restores the boundary curves of an order-$(m)$ partition. The higher order partitions shown in this work satisfy Eq. (4), however, higher-order cells do overlap. This overlapping is, however, minor to the orders treated here; although this overlapping will affect the topological properties of the probabilistic space, this is not significant to all the results presented in this work, in particular to the nominal value of the CaMI quantity. This naturally will prevent the observed partitions to

be generating. Consequently, this would have an impact if one would want to calculate CaMI for longer time-delays, in terms of the order-1 partition. Not the case for the present work.

In practice, when dealing with experimental (with noise) or simulated time-series, as we will proceed in this work, the boundaries and the cells of higher order-$L$ partitions are determined from symbolic sequences of length $L$ created from an order-1 partition. Initial conditions generating a symbolic sequences of length $L$ in the $X_{-L}$ coordinate created from an order-1 partition (a binary partition) will belong to an order-$L$ marginal partition along this coordinate. Marginal and joint probabilities are also estimated from the probabilities of appearances of symbolic sequences by Eqs. (6) and (8), respectively. More details on how this is done including the composed subspace $Y_{-L}Y_L$ where joint events in the variable $Y$ are calculated and where marginal partitions will have an order of $2L$ can be seen in Sec. IV. See also Eq. (12) and following explanations to understand how to calculate CaMI from the higher-order partitions, from symbolic sequences.

## D. An example

As an example of how trajectory points visit the partition with different orders and how this trajectory is symbolic encoded and probabilities are calculated, we consider a dynamic process along a 1D binary partition. Assume $x_t \in \mathcal{C}^X(1)$. Then, $\Phi_{-L}^X(t) = \{x_{t-L}, x_{t-L+1}, \ldots, x_{t-1}\} = X_{-L}(t)$ and $\Phi_L^X(t) = \{x_t, x_{t+1}, \ldots, x_{t+L-1}\} = X_L(t)$, and $S_{-L}^X(t, 1) = \{\mathfrak{s}_{i_{t-L}}^X(1), \mathfrak{s}_{i_{t-L+1}}^X(1), \cdots, \mathfrak{s}_{i_{t-1}}^X(1)\}$ and $S_L^X(t, 1) = \{\mathfrak{s}_{i_t}^X(1), \mathfrak{s}_{i_{t+1}}^X(1), \ldots, \mathfrak{s}_{i_{t+L-1}}^X(1)\}$. Observing a trajectory composed by $\{X_{-L}(t), X_L(t)\}$ considering the partition $\mathcal{C}^X(1)$ allows us to conclude that the trajectory of the system has visited a sequence of cells described by a sequence of symbols $\{S_{-L}^{XY}(t, 1), S_L^{XY}(t, 1)\}$.

Assuming $L = 2$, there exists an order-2 partition generated by $\mathcal{U}^{-1}[\mathcal{C}^X(1)]$ whose cells represent intervals where points within generate $S_{-2}^X(t, 1)$ and $S_2^X(t, 1)$. Moreover, a cell in an order-2 partition is encoded by a symbol that represents the whole symbolic sequence of length-2 trajectories along the partition $\mathcal{C}^X(1)$, and therefore there exist $i$ such that $\mathfrak{s}_i^X(2) = S_{-2}(t, 1)^X$ and there exist $j$ such that $\mathfrak{s}_j^X(2) = S_2(t, 1)^X$. Consequently, $P[\mathfrak{s}_i^X(2)] = P[S_{-2}(t, 1)^X]$ and $P[\mathfrak{s}_j^X(2)] = P[S_2(t, 1)^X]$.

In Fig. 1, we can see the relationship between a trajectory of length-4 in an order-1 and an order-2 partition. This figure also illustrates how an order-2 partition is generated from an order-1 partition.

## E. Probabilistic spaces and informational quantities

We now define some notations for the probabilities and informational quantities to simplify the exposition of our next derivations.

The notations $P(X_L)$ or $P(Y_L)$ represent $P[\mathfrak{c}_i^{X_L}(L)]$ or $P[\mathfrak{c}_i^{Y_L}(L)]$. So, $P(X_L)$ represents the probabilities of finding points in the cells $\mathfrak{c}_i^{X_L}(L)$ or similarly $P(X_L) = P[\mathfrak{s}_i^{X_L}(L)]$. Therefore, the Shannon's entropy of length-$L$ symbolic
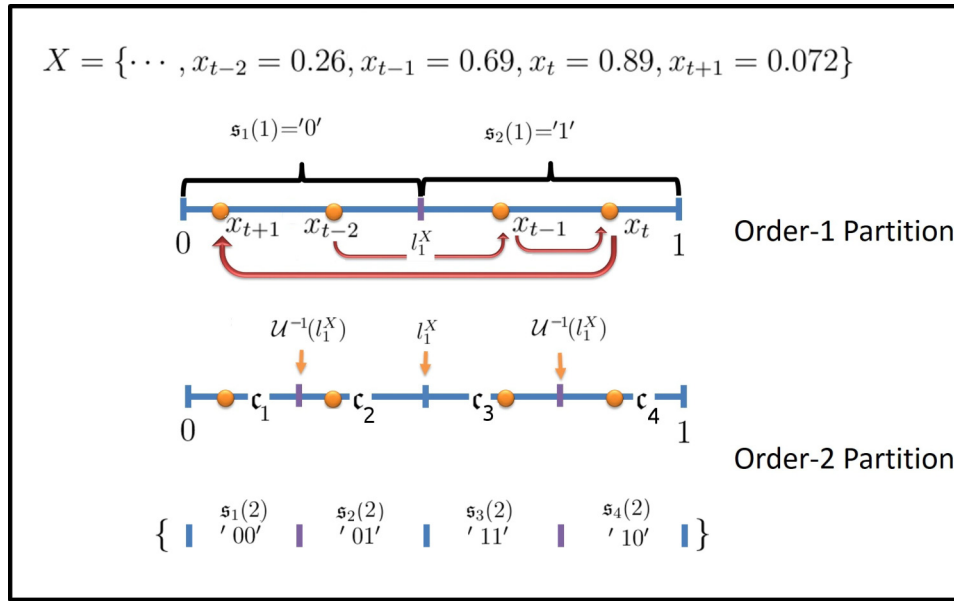
FIG. 1. In this diagram a length-4 trajectory is plotted in the state space and two partitions are shown, one of order-1 defined by the line $l_1^X$ (and the borders of the state space) and another of order-2 defined by the union of $l_1^X$ with $\mathcal{U}^{-1}(l_1^X)$. The trajectory along the order-1 partition generates the symbolic itinerary $\{S_{-L}^X(1), S_L^X(1)\} = \{\mathfrak{s}_{t-2} = \text{'0'}, \mathfrak{s}_{t-1} = \text{'1'}, \mathfrak{s}_t = \text{'1'}, \mathfrak{s}_{t+1} = \text{'0'}\}$ and along the order-2 partition generates the symbolic itinerary $\{S_{-L}^X(t,2), S_L^X(t,2)\} = \{\mathfrak{s}_{t-2}(2) = \text{'01'}, \mathfrak{s}_{t-1}(2) = \text{'11'}, \mathfrak{s}_t(2) = \text{'10'}, \mathfrak{s}_{t+1}(2) = \text{'00'}\}$. Notice, however, that the first symbol in $\mathfrak{s}_t(2) = \text{'10'}$ represents the present location in an order-1 partition and the second symbol represents the location of the first iteration. Therefore, the second symbol in $\mathfrak{s}_t(2) = \text{'10'}$, represents the same first symbol in $\mathfrak{s}_{t+1}(2) = \text{'00'}$.

sequences, represented by $H(X_L)$ or $H(X_{-L})$, is calculated by

$$H(X_L) = -\sum_i P[\mathfrak{c}_i^{X_L}(L)] \log\{P[\mathfrak{c}_i^{X_L}(L)]\}, \qquad (5)$$

So, if the "generating" property of the partition holds, entropies of length-$L$ trajectories along an order-1 partition space can be calculated by the measure of the cells encoding symbolic sequences of length 1 appearing in the higher order-$L$ partitions. This approach is especially oriented for analytical derivations based on the study of networks of coupled dynamical systems. Otherwise, when dealing with time-series coming from experiments or numerical simulations, as it is the case of the present work, we calculate entropies considering the probabilities of length-$L$ symbolic sequences observed along the order-1 partition. So,

$$H(X_L) = -\sum_i P[\mathfrak{s}_i^{X_L}(L)] \log\{P[\mathfrak{s}_i^{X_L}(L)]\}, \qquad (6)$$

where $\mathfrak{s}_i^{X_L}(L)$ represents a length-$L$ symbolic sequence.

Let us now define a 2-dimensional state space $\Omega^{X_{-L}Y_{-L}} \equiv [X_{-L} \times Y_{-L}]$ and with a trajectory $\Phi^{X_{-L}Y_{-L}}$ on it.

The joint entropy of a composed space $H(X_L, Y_L)$ is then defined by

$$H(X_L, Y_L) = -\sum_i P[\mathfrak{c}_i^{X_L Y_L}(L)] \log\{P[\mathfrak{c}_i^{X_L Y_L}(L)]\}. \qquad (7)$$

Assuming that the generating property of the partitions do not hold, then the quantity in Eq. (7) is estimated by

$$H(X_L, Y_L) = -\sum_i P[\mathfrak{s}_i^{X_L Y_L}(L)] \log\{P[\mathfrak{s}_i^{X_L Y_L}(L)]\}. \qquad (8)$$

By $P(X_L, Y_L)$, we represent the probability of a joint event calculated by $P(\mathfrak{c}_i^{X_L(L)Y_L(L)})$, and by $P(X_L|Y_L)$, we represent a conditional probability event representing the probability density of a point falling in the row $\mathfrak{c}_i^{Y_L(L)}$ and then being iterated to the column $\mathfrak{c}_i^{X_L(L)}$.

Extended 3-dimensional spaces can be constructed by the composition of $\Omega^{X_{-L}Y_{-L}}$ with the 1-dimensional space representing the presence of variable $Y$ observed in an order-$L$ partition, or a space constructed from the time-forward trajectory that visits an itinerary of length-$L$ in an order-1 partition. We represent this space by $(\Omega^{Y_L})$ or $X$ $(\Omega^{X_L})$. Notice that a point belonging to a partition in $\mathfrak{c}_i^{Y_L}(L)$ will produce an length-$L$ itinerary along the order-1 partition.

We are interested in the spaces $\Omega^{X_{-L}Y_{-L}Y_L}$ or $\Omega^{Y_{-L}X_{-L}X_L}$ composed by the variables $\{X_{-L}(t), Y_{-L}(t), Y_L(t)\}$ or $\{Y_{-L}(t), X_{-L}(t), X_L(t)\}$. It will be of further interest the spaces $\Omega^{Y_{-L}Y_L}$ and $\Omega^{X_{-L}X_L}$

Figure 2(a) shows the space $\Omega^{X_{-L}Y_{-L}Y_L}$ formed by the time-delay coordinates $X_{-L}$, $Y_{-L}$ and time-forward coordinate $Y_L$, with $L = 1$, and an order-1 partition in all subspaces defining our probabilistic space. The order-1 partition for the 2D space formed by $(Y_{-1} \times Y_1)$ shows the symbolic names of columns, rows, and the composed cells.

Notice that the 2D space $\{Y_{-1}, Y_1\}$ with an order-1 partition, where probabilities are calculated [Fig. 2(a)], can be reduced to a 1D space $W_2^{Y_1 Y_{-1}}$ with an order-2 partition [Fig. 2(b)]. A partition cell in the space $W_2^{Y_1 Y_{-1}}$ represents point that are in $\mathfrak{c}_i^{Y_{-1}}(l)$ and move to $\mathfrak{c}_j(1) \in Y_1$, and therefore produce probabilities of the joint events $P(Y_{-1}, Y_1) = P(Y_{-1})P(Y_1|Y_{-1})$. In a general situation, for an arbitrary $L$, probabilities in the space $\{X_{-L}, Y_{-L}\}$ and $\{Y_{-L}, Y_L\}$ could be calculated over an order-$L$ partition on each subspace. The reduced probabilistic space would be composed by a coordinate $X_{-L}$ where probabilities are calculated over an order-$L$
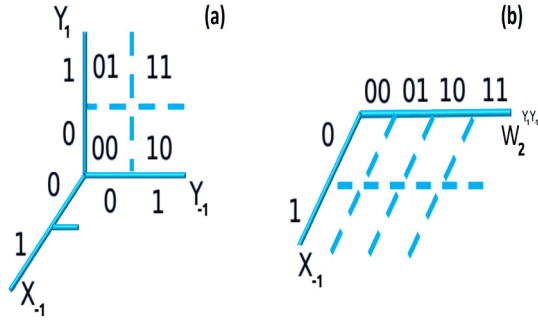
FIG. 2. (a) Visualization of the space $\Omega^{X_{-L}Y_{-L}Y_L}$ formed by the coordinates $(X_{-1}, Y_{-1}, Y_1)$ and an order-1 partition in each subspace formed by pair of coordinates. (b) Visualization of the reduced-dimensional space with coordinates $(X_{-1}, W_2^{Y_{-1}Y_1})$ and its asymmetric partition. Along $X_{-1}$ an order-1 partition is considered. Along $W_2^{Y_{-1}Y_1}$ an order-2 partition is considered.

partition and the coordinate $W_{2L}^{Y_{-L}Y_L}$ where probabilities would be calculated over an order-$2L$ partition. A cell in $W_{2L}^{Y_{-L}Y_L}$ would represent joint events $P(Y_{-L}, Y_L) = P(Y_{-L})P(Y_L|Y_{-L})$.

Notice that one can consider subspaces $X_{-L}$ and $Y_{-L}$ with an order-$L$ partition each, and the subspace $Y_J$ with an order-$J$ partition, with $J \neq L$, composing the space $\Omega^{X_{-L}Y_{-L}Y_J}$. Then, the reduced space $W$ would have a probabilistic space formed by a partition of order $(J + L)$.

## IV. THE TOPOLOGY OF CAUSALITY

### A. Generating higher-order partitions

We consider two non-coupled Logistic maps [$\alpha = 0$ in Eq. (1)], represented by $X$ and $Y$, to illustrate how we construct our partitions. Setting $l_1^X = 0.5$, if $x_i \leq l_1^X$, then $\mathcal{T}(x_i) = 0$ [and $\mathfrak{s}_i^X(1) = $ "0"], and if $x_i > l_1^X$, then $\mathcal{T}(x_i) = 1$ [and $\mathfrak{s}_i^X(1) = $ "1"]. Applying these rules for a trajectory of this uncoupled system, we generate Fig. 3(a). In Fig. 3(b), we show in green two columns of the order-2 partition obtained by $\mathcal{U}^{-1}(l_1^X)$. Setting $l_1^Y = 0.5$, therefore $\mathcal{L}^{XY} = \{l_1^X, l_1^Y\}$, we generate Fig. 3(c). The same coloured regions in this figure represent cells in an order-3 partition created by $\mathcal{U}^{-2}[\mathcal{C}^{XY}(1)]$. The order-3 partition has columns and rows enclosed by straight lines $\mathcal{L}_X(3) = \mathcal{U}^{-2}[l_1^X(1)]$ and $\mathcal{L}_Y(3) = \mathcal{U}^{-2}[l_1^Y(1)]$, respectively, forming the partition $\mathcal{C}^{XY}(3) = \mathcal{U}^{-2}[C^{XY}(1)]$. In this case, each column and row have boundaries that describe a generating partition of one Logistic map.

In practice, we do not make pre-iterations of the partition lines to determine the higher-order partitions. Once we choose $\mathcal{L}_X$ and $\mathcal{L}_Y$, the rows, columns, and cells of higher-order partitions are visualised by the colours of points that encode a particular symbolic sequence, using the following algorithm. Given a trajectory, we construct the length-$L$ segments of it $\Phi_L^{XY}(n) = \{(x_n^1, x_n^2), \ldots, (x_{n+L}^1, x_{n_L}^2)\}$, and whose symbolic sequence is represented by $S_L^{XY}(n, 1) = \{\mathfrak{s}_n^X(1), \mathfrak{s}_{n+1}^X(1), \ldots, \mathfrak{s}_{n+L-1}^X(1) \bullet \mathfrak{s}_n^Y(1), \mathfrak{s}_{n+1}^Y(1), \ldots, \mathfrak{s}_{n+L-1}^Y(1)\}$. This symbolic sequence is then encoded into an integer number that is used in the palette of colours to set the colour of the point $(x_n^1, x_n^2)$ that will produce the length-$L$ symbolic sequence $S_L^{XY}(n, 1)$. Points will belong to the same column (row) if their symbolic sequence
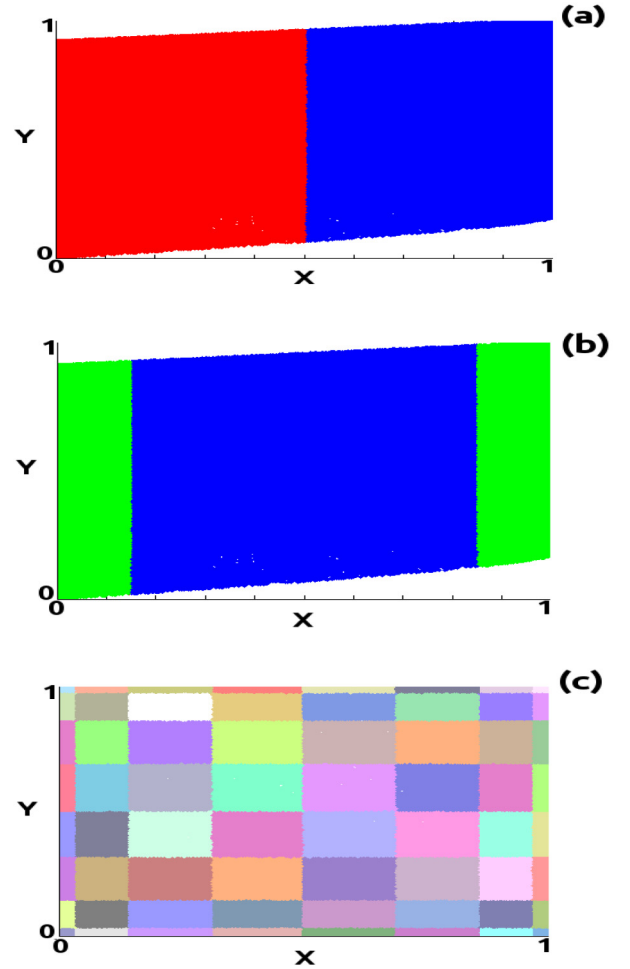


FIG. 3. Panel (a) shows two columns with name $\mathfrak{s}_1(1) = $ "0" in red and $\mathfrak{s}_2(1) = $ "1" in blue, for the order-1 partition in $X$. Panel (b) shows in green two columns from an order-2 partition in $X$ with names $\mathfrak{s}_1(2) = $ "00" and $\mathfrak{s}_4(2) = $ "10". Panel (c) shows, by the regions of the same color, cells $\mathfrak{c}_i^{XY}(3)$ of an order-3 partition.

$S_L^X(n, 1)$ [$S_L^Y(n, 1)$] is the same and will belong to the same cell if their symbolic sequence $S_L^{XY}(n, 1)$ is the same. To set the palette of colours, we produce an integer for the colour of the point $\Phi_L(n) = (x_n^1, x_n^2)$ of an order-$k$ partition generated from an order-1 partition, using the following encoding rule:

$$colour[\Phi_L(n)] = \theta_x(n) * 2^k + \theta_y(n), \quad (9)$$

where $\theta_x(n) = \sum_{i=1}^L \mathfrak{s}_{n+1-i}^X(1)2^{L-i}$ and $\theta_y(n) = \sum_{i=1}^{2L} \mathfrak{s}_{n+1-i}^Y(1)2^{2L-i}$. Then, colours are randomly assigned to each of these integer numbers.

### B. Understanding the arrow of influence

We now show how the topological properties of higher-order partitions change according to the coupling strength $\alpha$ between 2 or more coupled systems as in Eq. (1), and how these topological asymmetries can be used to determine the arrow of influence in these systems.

The symmetry in the structure of the partition in Fig. 3(c) reflects the fact that the two systems are not coupled. Imagine that an observer measure an event in the variable $Y$ at a
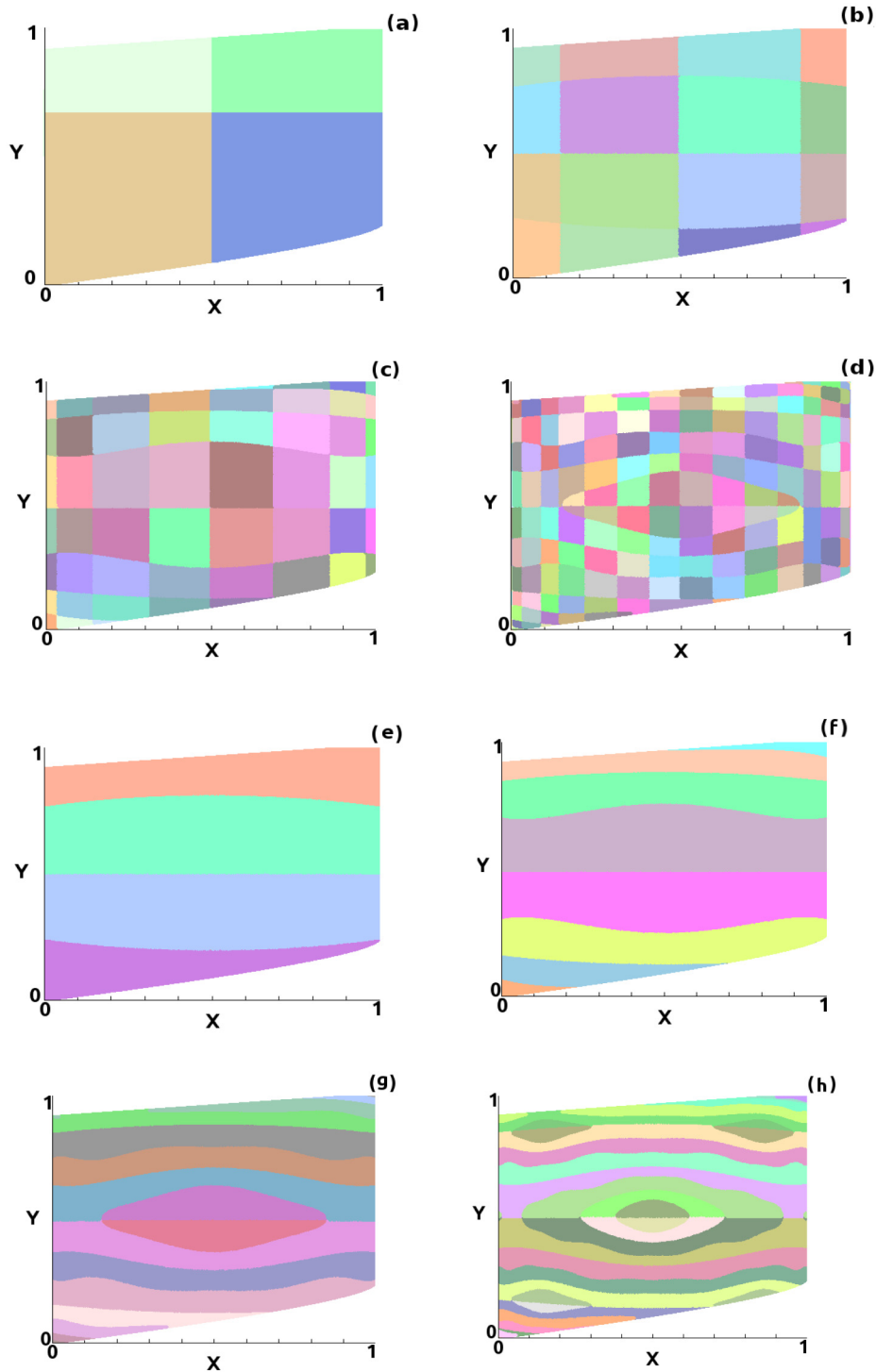
FIG. 4. Panel (a) shows an order-1 partition $\mathcal{C}_1^{XY}$, each coloured region represents a cell. Panel (b) shows the order-2 partition given by $\mathcal{C}_1^{XY} = \mathcal{U}^{-1}[\mathcal{C}^{XY}(1)]$. Panels [(c) and (d)] show the corresponding order-3 and order-4 partitions, generated by $\mathcal{U}^{-2}[\mathcal{C}^{XY}(1)]$ and $\mathcal{U}^{-3}[\mathcal{C}^{XY}(1)]$, respectively, coloured regions represent the cells in the partition. Panels (e-g) show $\mathcal{U}^{-1}[l^Y(1)], \mathcal{U}^{-2}[l^Y(1)]$ and $\mathcal{U}^{-3}[l^Y(1)]$, respectively, and each coloured "horizontal" stripe represents a higher-order row. Finally, panel (h) shows only the higher-order rows of an order-5 partition, enclosed by $\mathcal{U}^{-4}[l^Y(1)]$.

time $n$: $Y(n) = x_n^2 = 0.5$. Applying $\mathcal{U}^{-k}(l_1^Y)$, for any $k$, will create always a vertical line stretching from 0 to 1, meaning that an observation in $Y$ at time $n$ cannot be used to localize the state of the variable $X$ at time $n - k$. The consequence, as we will show next, is that there is no flow of information from $X$ to $Y$. The contrary is also true, i.e., one can also use similar

arguments to conclude that there is no flow of information from $Y$ to $X$.

For a coupled system in a master (node $X$) and a slave (node $Y$) configuration, assuming a coupling strength of $\alpha = 0.09$, we have created partitions of different orders (from order-1 to order-5) and shown in Fig. 4. One can see how the
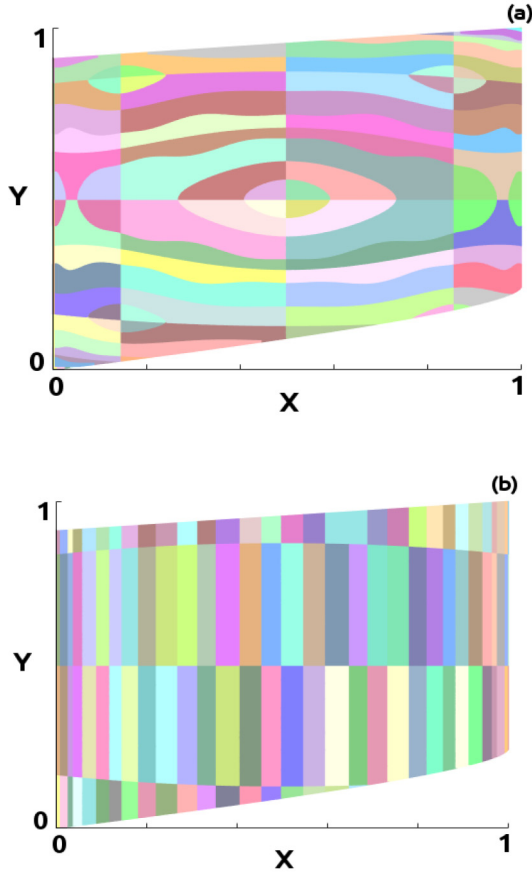
FIG. 5. Panel (a) shows the partition of order 5 in $Y$ and order 2 in $X$. Panel (b) shows the partition of order 2 in $Y$ and order 5 in $X$.
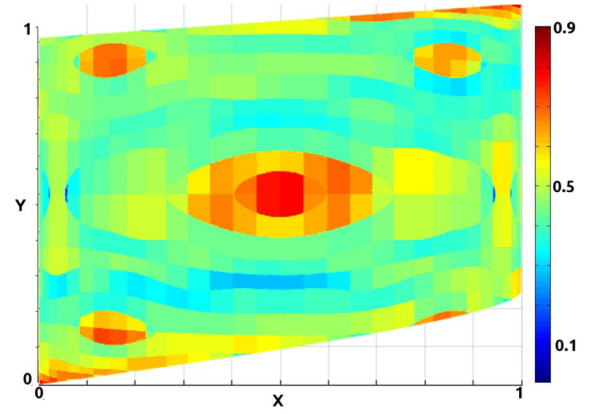


FIG. 6. nPMI for a directed coupled Logistic map with a coupling strength of 0.09. It can be observed how the nPMI is higher inside the bubbles. Minimum and maximal values of nPMI are $-5.12$ and $4.12$ respectively. These values were normalized to fit in the interval [a,b], with $a = 0$ and $b = 1$. The equation for the normalization is $y = (b - a) * (\mathrm{nPMI} - \min\{\mathrm{nMPI}\})/(\max\{\mathrm{nMPI}\} - \min\{\mathrm{nMPI}\}) + a$.

increase of the order increases the topological complexity of the partitions, for example going from Figs. 4(a) to 4(d). Paying attention to the higher order rows, defined by the enclosure of $\mathcal{U}^{-k}(l_1^Y)$ along the $Y$ variable, in Figs. 4(e)–4(h), and the higher-order columns, defined by enclosure of $\mathcal{U}^{-k}(l_1^X)$, we can observe how they are not enclosed any longer by straight lines. This asymmetry is the consequence of $X$ driving $Y$.

We now want to analyse the different topological features of the higher-order partitions when we consider different orders in $X$ and $Y$. For that we produce Fig. 5 obtained from two Logistic maps coupled in a master ($X$) and slave ($Y$) configuration with a coupling strength $\alpha = 0.09$. We have selected two orders for our partitions in $X$ and $Y$: 2 and 5. Figure 5(a) shows the different cells (same colour region) of a partition created by the intersection of an order 2 partition in $X$ and order 5 partition in $Y$. Figure 5(b) shows the different cells of a partition of an order 2 in $Y$ and an order 5 in $X$. The asymmetry in Fig. 5(a) indicates that the system has an arrow of influence $X \rightarrow Y$.

## C. Local Mutual Information

Pointwise Mutual Information (PMI) is a probabilistic measure of the amount of information that two different random variables posses locally between them. Given a particular partition, the PMI only takes into consideration the information computed over a single cell and not over the entire set of cells as the MI. PMI is defined as

$$PMI(i,j) = H_X^i + H_Y^j - H_{XY}^{i,j}, \tag{10}$$

where $H_X^i = -P(\mathfrak{s}_i^X) \log[P(\mathfrak{s}_i^X)]$, $H_Y^j = -P(\mathfrak{s}_j^Y) \log[P(\mathfrak{s}_j^Y)]$, and $H_{XY}^{i,j} = -P(\mathfrak{s}_k^{XY}) \log[P(\mathfrak{s}_k^{XY})]$, with $k$ representing the cell formed by the overlapping of the higher-order row $i$ with the higher-order column $j$. Therefore, MI is just the average of PMI over cells of the partition. In the following, we consider a normalized variant of PMI, named normalized Pointwise Mutual Information (nPMI), introduced in Ref. [21], and defined as follows:

$$nPMI = -\frac{PMI(i,j)}{\log[P(\mathfrak{s}_k^{XY})]}. \tag{11}$$

The advantage of the nPMI over PMI is the reduction of the sensitivity of the measure to short time-series.

Using nPMI in the partitions considered in Fig. 5, we can calculate the amount of information exchanged between variables $X$ and $Y$, the nMPI values for each cell represent the contribution of specific symbolic sequences (time nature of causality), or of particular variables domains (spatial nature of causality) contributing to the information transferred from $X$ to $Y$.

Figure 6 shows the nPMI for two directed coupled Logistic maps for an order-5 partition $\mathcal{C}^{XY}$ as the order of the partition along the variable $Y$ in Fig. 5(a). Rule for the normalization is explained in the figure caption. The surprising fact is that nPMI is larger for a special union of cells enclosed by one of the solutions of $\mathcal{U}^{-3}(l_1^Y)$ that forms closed curves. We call these special union of cells, causal bubbles, from its closed graphical representation. We can see that the bubbles are areas containing trajectory points responsible for a large amount of information exchanged between $X$ and $Y$, and as explained in Sec. IV D, a consequence of the fact that $X \rightarrow Y$.

## D. Causal bubbles

Figure 7 shows an illustration of how these bubbles are created. The partition line $l_1^Y$ represented by the black dashed
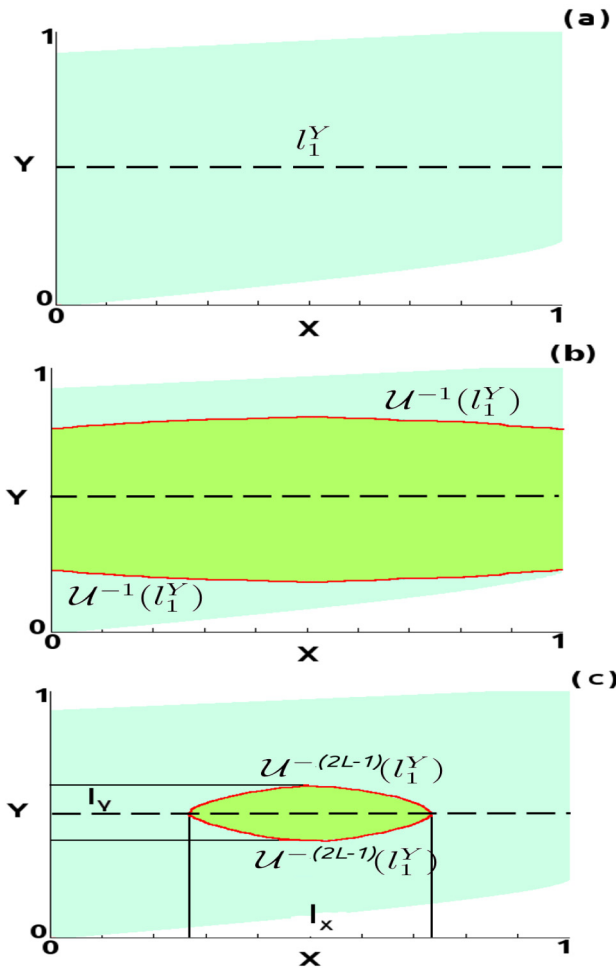
FIG. 7. (a) Dashed black line represents $l_1^Y$. (b) The red curves are obtained by one backward iteration of $l_1^Y$. (c) The closed red curve is obtained by $2L - 1$ backward iterations of $l_1^Y$.

line in panel (a) is iterated once producing the red curves such as the ones showed in panel (b), and eventually after $2L - 1$ backwards iterations these curves form a closed contour as is shown in panel (c) by the closed red curves.

Assume that the future of the observed variable $y_{t+L-1}$, $L$ iterations forward in $y$, has a value that lays exactly at the partition line, i.e., $y_{t+L-1} \in l_1^Y$. The variable to be predicted has an arbitrary value at $t + L - 1$, so $x_{t+L-1}$ can assume any value in the state space. Then, after one backward iteration, $y_{t+L-2} \in \mathcal{U}^{-1}(l_1^Y)$ is located on the red lines in panel (b). After $2L - 1$ backward iterations, $y_{t-L}$ has a position along the red closed curve in panel (c), enclosing a bubble area.

Therefore, the first observation in the variable $y$ at time $t + L - 1$ cannot tell anything about the position of the variable $x_{t+L-1}$. Assuming that the observer has full knowledge of the dynamical equations, it makes one observation at $y$ obtaining the value $y_{t+L-1}$, which in this imaginary example lays at $l_1^Y$. The smart observer uses the knowledge of the dynamics and makes $2L - 1$ backward iteration of $l_1^Y$. The observer will conclude that, if $y_{t+L-1} \in l_1^Y$, then $y_{t-L} \in I_Y$ and $x_{t-L} \in I_X$, where $I_X$ is the $x$-interval enclosed by the red curve, panel(c). So, by making one observation of the future of $y$ and using the knowledge of the dynamics (which can be obtained by

inspection of the time-series or from modelling the system), the observer can better localize the state of the past variable $x$, if $X \rightarrow Y$. Moreover, if more observations are done (from the future to the past) in $y$, more likely the observer is to improve its knowledge about the location of $x_{t-L}$ by doing similar analysis.

If there is no physical connection between system $X$ and system $Y$, and no flow of information from $X$ to $Y$, the bubbles are not formed, and therefore, one cannot localize the position of the variable $X$ by observing $Y$.

It is worth mentioning that the studied system is non-invertible, and therefore, if future of $Y$ depends on the past $X$ (as one would naturally conclude if doing the Granger analysis by constructing a model from the data), it is not necessarily obvious that the past of $X$ can be predicted by the future of $Y$. The causal bubble, however, demonstrates that this is indeed true.

## V. CAUSAL MUTUAL INFORMATION (CAMI)

We are now ready to define a new informational quantity named *Causal Mutual Information* from $X$ to $Y$ (CaMI$_{X \rightarrow Y}$) as the mutual information between joint events in $X_{-L}$ and the set composed by the joint events of $Y_{-L}$ and $Y_L$ as

$$\text{CaMI}_{X \rightarrow Y} = MI(X_{-L}; Y_{-L}, Y_L) = MI(X_{-L}; W_{2L}^{Y_L Y_{-L}}). \quad (12)$$

Analogously, CaMI$_{Y \rightarrow X} = MI(Y_{-L}; W_{2L}^{X_L X_{-L}})$. So, in practice, CaMI$_{X \rightarrow Y}$ is calculated by computing the mutual information between symbolic sequences of length $L$ in the variable $X$ and symbolic sequences of length $2L$ in the variable $Y$.

Notice that CaMI is not permutable since CaMI$_{X \rightarrow Y} \neq$ CaMI$_{Y \rightarrow X}$. As we can see, CaMI$_{X \rightarrow Y}$ is the Mutual information between trajectory points in the subspace $X_L$ and the subspace $W_{2L}^{Y_L Y_{-L}}$ and that measures the amount of information between longer time-series of past, present, and future of $Y$ and shorter time-series of the past of $X$. The fundamental idea behind the reason for us to propose CaMI as a measure of causality is that if there is a flow of information from $X$ to $Y$, then longer observations in the $Y$ variable can be used to predict the past of states of $X$. CaMI is also a quantity that measures the total amount of information extracted from one variable by observing another variable, not only the shared amount (non-causal, measured by the Mutual Information) but also the transmitted amount (causal, measured by the Transfer Entropy).

Considering the coupled system and the same partition used to create Fig. 5(a), the magnitude of the computed CaMI$_{X \rightarrow Y}$ is 0.17. Considering the coupled system and the same partition used to create Fig. 5(b), the magnitude of the computed CaMI$_{Y \rightarrow X}$ is 0.04. The difference in the CaMI's magnitudes, meaning CaMI$_{X \rightarrow Y}$-CaMI$_{Y \rightarrow X} > 0$, and the asymmetry in Fig. 5(a) with the existence of the causal bubbles indicate the presence of a system whose direction of causality is given by $X \rightarrow Y$, and therefore, there is a flow of influence from $X$ to $Y$. As we shall see, for this situation, CaMI$_{Y \rightarrow X}$=$MI(X_{-L}; Y_{-L})$, meaning that the variables share some common, non-causal information.

Notice that the PMI as defined in Eq. (10) is just one of the terms considered in the calculation of CaMI.

## A. Transfer Entropy and Causal Mutual Information

Having two random processes $X$ and $Y$, the amount of information transferred from process $X$ to $Y$ ($X \rightarrow Y$) can be quantified by the Transfer Entropy,[1] defined as follows:

$$T_{X \rightarrow Y} = H(Y_L \mid Y_{-L}) - H(Y_L \mid Y_{-L}, X_{-L}), \quad (13)$$

where $H(X)$ is the Shannon's entropy of $X$, $H(X_L \mid Y_{-L})$ is the knowledge (reduce of entropy) of process $X$ from time $t$ to $t + L - 1$ if the past of process $Y$ from $t - 1$ to $t - L$ is known, and $H(Y_L \mid Y_{-L}, X_{-L})$ represents the knowledge of the process $Y$ from time $t$ to $t + L - 1$, if the past of $X$ and $Y$ in the interval from $t - 1$ to $t - L$ is known. Transfer entropy was shown to be related [22,23] to directed information, a sort of a cumulative version of transfer entropy. The later quantity is being currently considered as an appropriate measure to deal with channels of communication with feed-back, describing channels where the output is fed back to the input.

We can express Eq. (13) as a function of joint entropies and not conditional one using the chain rule for entropy and Bayes theorem:

$$T_{X \rightarrow Y} = H(Y_L, Y_{-L}) - H(Y_{-L})$$
$$- H(Y_{-L}, Y_L, X_{-L}) + H(Y_{-L}, X_{-L}). \quad (14)$$

But notice that CaMI can also be written as

$$\text{CaMI}_{X \rightarrow Y} = MI(X_{-L}; Y_L, Y_{-L})$$
$$= H(X_{-L}) + H(Y_L, Y_{-L})$$
$$- H(X_{-L}, Y_L, Y_{-L}). \quad (15)$$

Finally, comparing Eq. (15) with Eq. (14), we conclude that

$$\text{CaMI}_{X \rightarrow Y} = T_{X \rightarrow Y} + MI(X_{-L}; Y_{-L}), \quad (16)$$

where $MI(X_{-L}; Y_{-L}) = H(X_{-L}) + H(Y_{-L}) - H(X_{-L}, Y_{-L})$ is the mutual information of the system composed by $X$ and $Y$. Both quantities are shown over a Venn diagram in Fig. 8 for their comparison. One can see that CaMI carries more information about the considered variables than transfer entropy. $\text{CaMI}_{X \rightarrow Y}$ represents the amount of information exchanged between $X$ and $Y$ [provided by the term $MI(X_{-L}; Y_{-L})$] and the transfer entropy from $X$ to $Y$. Whereas $MI(X_{-L}; Y_{-L})$ measures how much the observation of a length-$L$ trajectory along the variable $X$ (or $Y$) can be predicted by observations of a length-$L$ trajectory of the variable $Y$ (or $X$), the transfer entropy from $X$ to $Y$ measures how much one can predict from the past state of the $X$ by making observations of the past, present, and future states of the variable $Y$. One important fact to notice is that since $MI(X_{-L}; Y_{-L}) = MI(Y_{-L}; X_{-L})$, the directionality index defined by $T_{X \rightarrow Y} - T_{Y \rightarrow X}$, and therefore representing the net flow of information between both variables, can be calculated by

$$DI = T_{X \rightarrow Y} - T_{Y \rightarrow X} = \text{CaMI}_{X \rightarrow Y} - \text{CaMI}_{Y \rightarrow X}. \quad (17)$$

As another remark, notice that transfer entropy is defined as the conditional mutual information, and therefore, $T(X \rightarrow Y) = MI(X_{-L}; Y_L | Y_{-L})$, see Ref. [5]. Recalling that $\text{CaMI}_{X \rightarrow Y} = MI[X_{-L}; (Y_{-L}, Y_L)]$, it is easy to see that to define CaMI we have replaced the conditional probabilities in the transfer entropy $T_{X \rightarrow Y}$ to joint probability ones in $\text{CaMI}_{X \rightarrow Y}$.
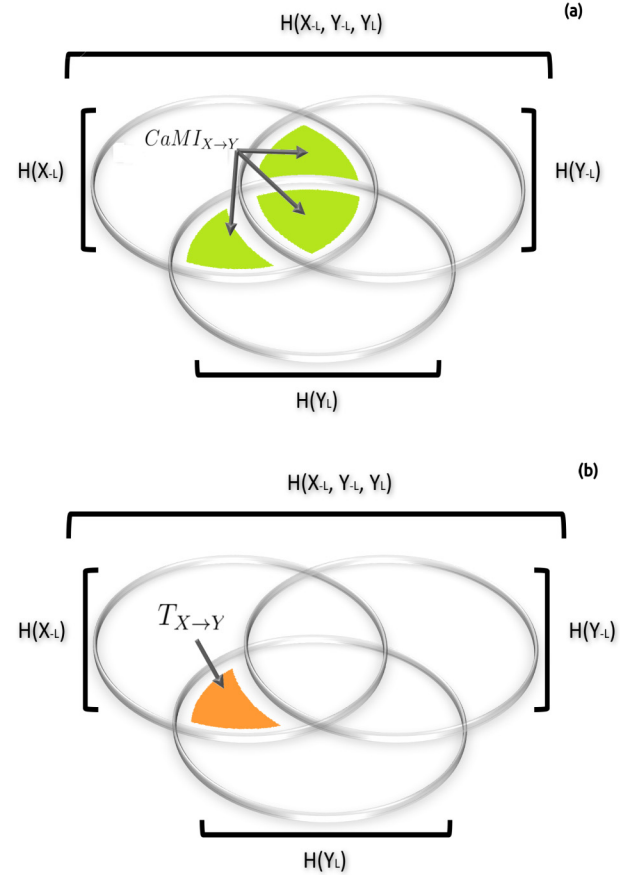


FIG. 8. Representation over a Venn diagram of the CaMI. Panel (a): Causal Mutual Information, $\text{CaMI}_{X \rightarrow Y}$. Panel (b): Transfer entropy, $T_{X \rightarrow Y}$.

## VI. HIGHER-RESOLUTIONS INITIAL MARGINAL PARTITIONS

We have previously seen that the probabilities of length-$L$ symbolic sequences representing trajectory points following an itinerary along order-1 marginal partitions can be used to calculate CaMI, and therefore, be used to detect the causal direction of the flow of information. We have seen that the topology of a higher-order 2D asymmetric partition, where each coordinate has different orders, can be used to determine the arrow of influence between two variables. The higher-order partitions were generated out of order-1 marginal partitions. However, in certain practical situations, for example in stochastic or experimental systems, higher-order partitions generated out of lower-order partitions should be expected to produce no discernible topological set that could orient one to the correct direction of the flow of information. It is thus interesting to verify whether and for which boundary conditions a marginal partition with $N_X$ boundary lines along the variable $X$ [$\mathcal{L}_X(m)$] and a marginal partition with $N_Y$ boundary lines along the variable $Y$ [$\mathcal{L}_Y(2m)$] could be used to estimate a physically consistent $\text{CaMI}_{X \rightarrow Y}$. Simply put, $\text{CaMI}_{X \rightarrow Y}$ calculated in this way would be obtained by measuring the mutual information between lower-resolution observations along the variable $X$ and higher-resolutions observations along the variable $Y$.
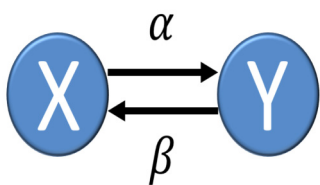
FIG. 9. Two Logistic maps bi-directionally coupled. The coupling strengths $\alpha$ and $\beta$ are related by $\beta = 0.1 - \alpha$.

We study the mutual information of a system (as the one shown in Fig. 9), with a coupling strength $\alpha = \{0, 0.05, 0.1\}$ and $\beta = 0.1 - \alpha$. In Fig. 10, we show by colours the values of the MI calculated considering Eqs. (5) and (7) between variables $X$ and $Y$ for a partition $\mathcal{C}^X$ for $X$ and $\mathcal{C}^Y$ for $Y$, with a different number of columns and rows, respectively. The number of $N_X$ and $N_Y$ are shown in the axis of Fig. 10. Recall that CaMI$_{X \rightarrow Y}$ is just the mutual information between variables $X$ and $Y$ where the partition of $X$ has order $L$ and the partition of $Y$ has order $2L$. Here, we want to test the plausible idea that causality can also be detected when the variables are observed with different spatial resolutions. We, therefore, want to test whether MI is capable of detecting the direction
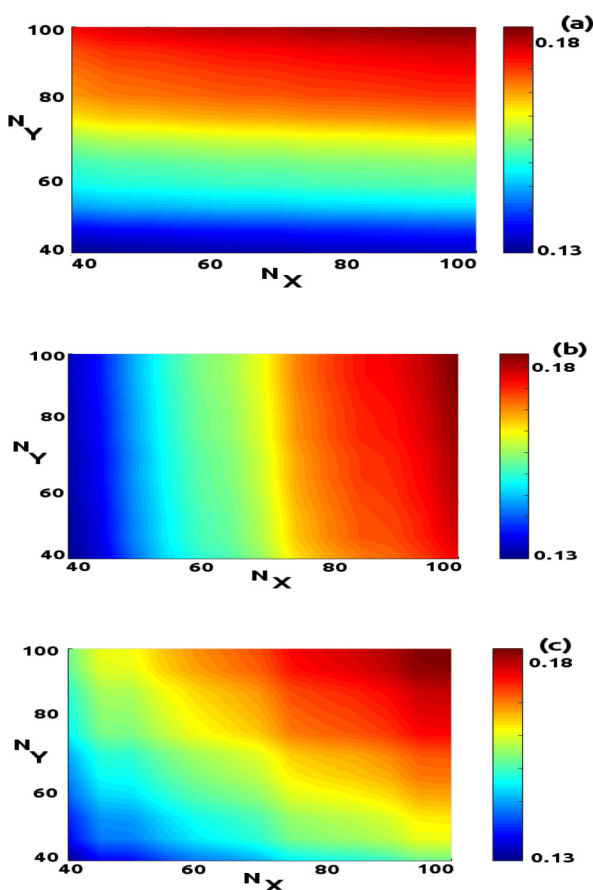


FIG. 10. Mutual information of coupled Logistic maps computed on an asymmetric partition with different number of rows and columns. The horizontal and vertical axis show the different amount of rows and columns of the partition. Panel (a) shows the result for $\alpha = 0.1$ and $\beta = 0$, so information flows from $X$ to $Y$. Panel (b) the results for $\alpha = 0$ and $\beta = 0.1$, so information flows from $X$ to $Y$. Panel (c) shows the result for $\alpha = 0.05$ and $\beta = 0.05$.

for the flow of information when a probabilistic space has an arbitrary number of equal rectangular areas.

In Figs. 10(a)–10(c), if the partition in $X$ has the same number of cells than the partition in $Y$, MI grows with the growing of the number of cells. As expected, one can see that if the flow of information is from $X \rightarrow Y$ [as in panel (a)], then partitions with more cells in $Y$ than in $X$ produce larger MI (or CaMI) than partition with less cells in $Y$ than in $X$.

Surprisingly, Fig. 10(a) shows a novel feature for the MI of coupled systems. If the flow of information goes from $X \rightarrow Y$, then given a partition in $Y$ with a particular number of cells (i.e., resolution), the value of MI obtained is roughly invariant for the chosen resolution in $X$. This implies that the amount of information one can realize from $X$ by making measurements in $Y$ is almost solely dependent on the resolution of the observation in $Y$, and there is a sufficient large amount of number of cells in $Y$. In Fig. 10(b), the flow of information goes from $Y$ to $X$ and therefore for a sufficiently large number of cells in $X$, the information that one can deduce from $Y$ by measuring $X$ almost solely depends on the resolution of $X$. Finally, in Fig. 10(c) for the bidirectionally coupled system, with equal coupling strengths, the values for MI will depend on both resolution of variables $X$ and $Y$ in a complementary way; i.e., if the sum of the number of rows and columns is maintained, MI remains roughly invariant. In fact, the relationship is described by a diagonal hyperbola. This phenomenon can be also used to detect the directionality of the coupling.

## VII. CONCLUSION

In this work, we have investigated the spatiotemporal properties of causality, causality meaning the study of the arrow of influence between two systems. The spatiotemporal nature of causality can be exploited to detect the arrow of influence from $X$ to $Y$, by considering either shorter time-series of $X$ and longer time-series of $Y$, an approach that explores the time nature of causality, or lower precision measured time-series in $X$ and higher precision measured time-series in $Y$, an approach that explores the spatial nature of causality. Thus, this work shows that causality can be detected not only by the analyses of the topological properties of higher-order partitions generated by lower-order marginal partitions (the "space" property of causality) but also by considering the probabilities on these partitions, reflecting the density of trajectories of a given time-length (the "time" property of causality). To apply this abstract notion of causality into a quantitative approach, we have introduced a new informational quantity, the causal mutual information, CaMI$_{X \rightarrow Y}$, that measures the total amount of information being transmitted from $X$ to $Y$, the information shared between both variables and that can be used to predict the present state of $X$ by observations in $Y$ (i.e., the Mutual Information), and the causal information directed transmitted from $X$ to $Y$, which can be used to predict the past states of $X$ by observations of the past and future states of $Y$ (i.e., the Transfer Entropy).

Since CaMI does not require the calculations of conditional probabilities, but rather only joint probabilities, the

probabilistic spaces involved in its calculation can be lower-dimensional, enabling a quick estimation of TE. This property is well wished for causal analysis of large complex systems such as the brain or for technological applications that employ TE, for example, the recently proposed brain-based cryptography (see Ref. 24). Also, less data are required for the determination of causality since the probability space can be constructed according to the available data.

Another important result of this work was to show that measuring a driven variable with finer resolution than that used to observe the driving variable allows us to obtain more information about the driving system but not the other way around. Increasing the resolution of observation of the driving variable brings no additional information about the driven variable. This observation could also be exploited to detect directionality in networked systems.

A potential advantage of our approach is that even though the analysis of causality is bivariate, employing two observables taken from two subsystems $X$ and $Y$ in a larger system, the topological properties of the constructed probabilistic space can potentially discern whether information is being sent physically from $X$ to $Y$, or whether it is being mediated by other subsystems and variables (in this case, there is no physical connection between $X$ and $Y$). This special property of the probabilistic space shown to exist to fully deterministic systems to detect causality allows one to detect direct or mediate effects without the need to calculate multivariate conditional probabilities, from which one can detect direct or mediated influences in stochastic systems,[10,25] an approach suitable for both dynamical and stochastic systems, but that however requires the use of large dimensional probabilistic spaces. In the multivariate approach to detect causality, the conditional probabilities of multivariate variables need to take into consideration the influence of co-founders,[26] entities that mediates the transfer of information from $X$ and $Y$ systems. In fact, our approach was recently tested to infer the topology and the synaptic nature (either chemical or electrical) of complex neural networks. It was shown[27] that CaMI can be used to successfully infer the topology of the neural network with no mistakes, and also discern about the nature of the connections, even when the network is in the presence of both dynamical and observational additive Gaussian noise, and even when only observational time-series based on local averages are available.

## ACKNOWLEDGMENTS

[1] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).
[2] I. Guyon, D. Janzing, and B. Schölkopf, "Causality: Objectives and assessment," in *JMLR Workshop and Conference Proceedings* (2010), Vol. 6, pp. 1–38.
[3] Y. Kano and S. Shimizu, "Causal inference using nonnormality," in *Proceedings of the International Symposium on the Science of Modeling, the 30th Anniversary of the Information Criterion* (2013), pp. 261–270.
[4] C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," Econometrica **37**, 424 (1969).
[5] M. Paluš, Physica D **93**, 64 (1996).
[6] Y. Hirata and K. Aihara, Phys. Rev. E **81**, 016203 (2010).
[7] S. Haufe, V. V. Nikulin, K.-R. Müller, and G. Nolte, NeuroImage **64**, 120 (2013).
[8] Y. Chen, G. Rangarajan, J. Feng, and M. Ding, Phys. Lett. A **324**, 26 (2004).
[9] N. Ancona, D. Marinazzo, and S. Stramaglia, Phys. Rev. E **70**, 056221 (2004).
[10] E. Bollt, P. Gora, A. Ostruszka, and K. Zyczkowski, SIAM J. Appl. Dyn. Syst. **7**, 341 (2008).
[11] C. W. Granger, J. Econ. Dyn. Control **2**, 329 (1980).
[12] C. W. Granger, J. Econom. **39**, 199 (1988).
[13] P.-O. Amblard, and O. J. Michel, J. Comput. Neurosci. **30**, 7 (2011).
[14] M. Takigawa, G. Wang, H. Kawasaki, and H. Fukuzako, Int. J. Psychophysiol. **21**, 65 (1996).
[15] L. A. Baccalá and K. Sameshima, Biol. Cybern. **84**, 463 (2001).
[16] M. Kaminski and K. J. Blinowska, Biol. Cybern. **65**, 203 (1991).
[17] H. Marko, IEEE Trans. Commun. **21**, 1345 (1973).
[18] V. A. Vakorin, O. A. Krakovska, and A. R. McIntosh, J. Neurosci. Methods **184**, 152 (2009).
[19] I. Vlachos and D. Kugiumtzis, Phys. Rev. E **82**, 016207 (2010).
[20] P. Myrberg, J. Math. Pures Appl. (9) **41**, 339 (1962).
[21] G. Bouma, "Normalized (Pointwise) mutual information in collocation extraction," *Proc. Ger. Soc. Comput. Linguist*, (2009), pp. 31–40.
[22] N. J. Newton, arXiv:1604.01969 (2016).
[23] Y. Liu and S. Aviyente, "The relationship between transfer entropy and directed information," in *2012 IEEE Statistical Signal Processing Workshop (SSP)*, (2012), pp. 73–76.
[24] R. Szmoski, F. Ferrari, S. d. S. Pinto, M. Baptista, and R. Viana, Phys. Lett. A **377**, 760 (2013).
[25] J. Runge, Phys. Rev. E **92**, 062829 (2015).
[26] S. Mangiarotti, Chaos Solitons Fractals **81**, 184 (2015).
[27] F. S. Borges *et al.*, Phys. Rev. E **97**, 022303 (2018).