

Mutual Information Rate and Bounds for It

Murilo S. Baptista^{1*}, Rero M. Rubinger², Emilson R. Viana³, José C. Sartorelli⁴, Ulrich Parlitz^{5,6}, Celso Grebogi^{1,7}

1 Institute for Complex Systems and Mathematical Biology, Scottish Universities Physics Alliance, University of Aberdeen, Aberdeen, United Kingdom, **2** Institute of Physics and Chemistry, Federal University of Itajubá, Itajubá, Brazil, **3** Instituto de Ciências Exatas, Departamento de Física, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, **4** Institute of Physics, University of São Paulo, São Paulo, Brazil, **5** Biomedical Physics Group, Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany, **6** Institute for Nonlinear Dynamics, Georg-August-Universität Göttingen, Göttingen, Germany, **7** Freiburg Institute for Advanced Studies, University of Freiburg, Freiburg, Germany

Abstract

The amount of information exchanged per unit of time between two nodes in a dynamical network or between two data sets is a powerful concept for analysing complex systems. This quantity, known as the mutual information rate (MIR), is calculated from the mutual information, which is rigorously defined only for random systems. Moreover, the definition of mutual information is based on probabilities of significant events. This work offers a simple alternative way to calculate the MIR in dynamical (deterministic) networks or between two time series (not fully deterministic), and to calculate its upper and lower bounds without having to calculate probabilities, but rather in terms of well known and well defined quantities in dynamical systems. As possible applications of our bounds, we study the relationship between synchronisation and the exchange of information in a system of two coupled maps and in experimental networks of coupled oscillators.

Citation: Baptista MS, Rubinger RM, Viana ER, Sartorelli JC, Parlitz U, Grebogi C. (2012) Mutual Information Rate and Bounds for It. PLoS ONE 7(10): e46745. doi:10.1371/journal.pone.0046745

Editor: Jesus Gomez-Gardenes, Universidad de Zaragoza, Spain

Received: June 20, 2012; **Accepted:** September 4, 2012; **Published:** October 24, 2012

Copyright: © 2012 Baptista et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: MSB was partially supported by the Northern Research Partnership, Alexander von Humboldt foundation, and the Engineering and Physical Sciences Research Council grant Ref. EP/I032606/1. The research leading to the results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement No. HEALTH-F2-2009-241526, EUTrigTreat. Furthermore, support by the Bernstein Center for Computational Neuroscience II Göttingen (BCCN grant 01GQ1005A, project D1) is acknowledged. RR, EV and JCS thanks the Brazilian agencies Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior, Conselho Nacional de Desenvolvimento Científico e Tecnológico, Fundação de Amparo à Pesquisa do estado de Minas Gerais and Fundação de Amparo à Pesquisa do Estado de São Paulo. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: murilo.baptista@abdn.ac.uk

Introduction

Shannon's entropy quantifies information [1]. It measures how much uncertainty an observer has about an event being produced by a random system. Another important concept in the theory of information is the mutual information [1]. It measures how much uncertainty an observer has about an event in a random system \mathbf{X} after observing an event in another random system \mathbf{Y} (or vice-versa).

Mutual information (MI) is an important quantity because it quantifies not only linear and non-linear interdependencies between two systems or data sets, but also is a measure of how much information two systems exchange or two data sets share. Due to these characteristics, it became a fundamental quantity to understand the development and function of the brain [2,3], to characterise [4,5] and model complex systems [6–8] or chaotic systems, and to quantify the information capacity of a communication system [9]. When constructing a model of a complex system, the first step is to understand which are the most relevant variables to describe its behaviour. Mutual information provides a way to identify those variables [10].

However, the calculation of mutual information in dynamical networks or data sets faces three main difficulties [4,11–13]. Mutual information is rigorously defined for random memoryless processes, only. In addition, its calculation involves probabilities of significant events and a suitable space where probability is

calculated. The events need to be significant in the sense that they contain as much information about the system as possible. But, defining significant events, for example the fact that a variable has a value within some particular interval, is a difficult task because the interval that provides significant events is not always known. Finally, data sets have finite size. Probabilities computed from finite data sets are subjected to unavoidable sampling errors. As a consequence, mutual information can often be calculated with a bias, only [4,11–13].

In this work, we show how to calculate the amount of information exchanged per unit of time [Eq. (2)], the so called mutual information rate (MIR), between two arbitrary nodes (or group of nodes) in a dynamical network or between two data sets. Each node represents a d -dimensional dynamical system with d state variables. The trajectory of the network considering all the nodes in the full phase space is denoted by Σ and represents the “attractor”, which in the following calculations is considered to be an asymptotic limiting set. Then, we propose an alternative method, similar to the ones proposed in Refs. [14,15], to calculate significant upper and lower bounds for the MIR in dynamical networks or between two data sets, in terms of Lyapunov exponents, expansion rates, and capacity dimension. These quantities can be calculated without the use of probabilistic measures. As possible applications of our bounds calculation, we describe the relationship between synchronisation and the

exchange of information in small experimental networks of coupled Double-Scroll circuits.

In previous works of Refs. [14,15], we have proposed an upper bound for the MIR in terms of the positive Lyapunov exponents of the synchronisation manifold. As a consequence, this upper bound could only be calculated in special complex networks that allow the existence of complete synchronisation. In the present work, the proposed upper bound can be calculated to any system (complex networks and data sets) that admits the calculation of Lyapunov exponents.

We assume that an observer can measure only one scalar time series for each one of two chosen nodes. These two time series are denoted by X and Y and they form a bidimensional set $\Sigma_{\Omega} = (X, Y)$, a projection of the ‘‘attractor’’ into a bidimensional space denoted by Ω . To calculate the MIR in higher-dimensional projections Ω , see Information S1. To estimate the upper bound of the MIR in terms of Lyapunov exponents obtained from the reconstructed attractor of a scalar time-series see Information S1. Assume that the space Ω is coarse-grained in a square grid of N^2 boxes with equal sides ϵ , so $N = 1/\epsilon$.

Mutual information is defined in the following way [1]. Given two discrete random variables, \mathbf{X} and \mathbf{Y} , each one produces events i and j with probabilities $P_X(i)$ and $P_Y(j)$, respectively, the joint probability between these events is represented by $P_{XY}(i,j)$. Then, mutual information is defined as

$$I_S = H_X + H_Y - H_{XY}. \quad (1)$$

$H_X = -\sum_i P_X(i) \log[P_X(i)]$, $H_Y = -\sum_j P_Y(j) \log[P_Y(j)]$, and $H_{XY} = -\sum_{i,j} P_{XY}(i,j) \log[P_{XY}(i,j)]$. When using Eq. (1) to calculate the mutual information between the dynamical variables X and Y , the probabilities appearing in Eq. (1) are defined such that $P_X(i)$ is the probability of finding points in a column i of the grid, $P_Y(j)$ of finding points in the row j of the grid, and $P_{XY}(i,j)$ the probability of finding points in a box where the column i meets the row j of the grid.

The MIR was firstly introduced by Shannon [1] as a ‘‘rate of actual transmission’’ [16] and later more rigorously redefined in Refs. [17,18]. It represents the mutual information exchanged between two dynamical variables (correlated) per unit of time. To calculate the MIR, the two continuous dynamical variables are transformed into two discrete symbolic sequences X and Y . Then, the MIR is defined by

$$MIR = \lim_{n \rightarrow \infty} \frac{I_S(n)}{n}, \quad (2)$$

where $I_S(n)$ represents the usual mutual information between the two sequences X and Y , calculated by considering words of length n . If $I_S(n)$ is calculated using \log_2 , the MIR in Eq. (2) has units of bits/symbol. If a discrete system is producing the symbols, the units of Eq. (2) are bits/iteration.

The MIR is a fundamental quantity in science. Its maximal value gives the information capacity between any two sources of information (no need for stationarity, statistical stability, memoryless) [19]. Therefore, alternative approaches for its calculation or for the calculation of bounds of it are of vital relevance. Due to the limit to infinity in Eq. (2) and because it is defined from probabilities, the MIR is not easy to be calculated especially if one wants to calculate it from (chaotic) trajectories of a large complex network or data sets. The difficulties faced to estimate the MIR from dynamical systems and networks are similar to the ones

faced in the calculation of the Kolmogorov-Sinai entropy, H_{KS} [20], (Shannon’s entropy per unit of time). Because of these difficulties, the upper bound for H_{KS} proposed by Ruelle [21] in terms of the Lyapunov exponents and valid for smooth dynamical systems ($H_{KS} \leq \sum \lambda_i^+$, where λ_i^+ represent all the i positive Lyapunov exponents) or Pesin’s equality [22] ($H_{KS} = \sum \lambda_i^+$) proved in Ref. [23] to be valid for the large class of systems that possess a SRB measure, became so important in the theory of dynamical systems. Our upper bound [Eq. (5)] is a result similar to the work of Ruelle, but instead we relate mutual information rate with Lyapunov exponents.

Our work is also similar to the work of Wissman-Jones-Binder [24] who have shown that upper and lower bounds for H_{KS} and the sum of the Lyapunov exponents can be calculated in terms of the mutual information, MI, of a trajectory. Their work, like ours, has shown a link between (conditional and joint) probabilities and a dynamical quantity, the Lyapunov exponents. We focus our attention to the relationship between MIR and Lyapunov exponents, Wissman and co-authors focus their attention in the relationship between MI and the Lyapunov exponents.

Results

One of the main results of this work (whose derivation can be seen in Sec. Methods) is to show that, in dynamical networks or data sets with fast decay of correlation, I_S in Eq. (1) represents the amount of mutual information between X and Y produced within a special time interval T , where T represents the time for the dynamical network (or data sets) to lose its memory from the initial state or the correlation to decay to zero. Correlation in this work is not the usual linear correlation, but a non-linear correlation defined in terms of the evolution of probabilities defined by space integrals, the quantity $C(T)$ in Eq. (9). Therefore, the mutual information rate (MIR), between the dynamical variables X and Y (or two data sets) can be estimated by

$$MIR = \frac{I_S}{T} \quad (3)$$

In systems that exhibit sensitivity to initial conditions, e.g. chaotic systems, predictions are only possible for times smaller than this time T . This time has other meanings. It is the expected time necessary for a set of points belonging to an ϵ -square box in Ω to spread over Σ_{Ω} and it is of the order of the shortest Poincaré return time for a point to leave a box and return to it [25,26]. It can be estimated by

$$T \approx \frac{1}{\lambda_1} \log \left[\frac{1}{\epsilon} \right]. \quad (4)$$

where λ_1 is the largest positive Lyapunov exponent measured in Σ_{Ω} . Chaotic systems can exhibit the mixing property (see Methods), and as a consequence the correlation $C(t)$ decays to zero, surely after an infinitely long time. The correlation of chaotic systems can also decay to zero for sufficiently large but finite $t = T$ (see Information S1). T can be interpreted to be the minimum time required for a system to satisfy the conditions to be considered as mixing. Some examples of physical systems that are proved to be mixing and have exponentially fast decay of correlation are nonequilibrium steady-state [27], Lorentz gases (models of diffusive transport of light particles in a network of heavier particles) [28], and billiards [29]. An example of a ‘‘real world’’ physical complex system that presents exponentially fast

decay of correlation is plasma turbulence [30]. We do not expect that data coming from a “real world” complex system is rigorously mixing and has an exponentially fast decay of correlation. But, we expect that the data has a sufficiently fast decay of correlation (e.g. stretched exponential decay or polynomially fast decays), implying that the system has sufficiently high sensitivity to initial conditions and as a consequence $C(t) \cong 0$, for a reasonably small and finite time $t = T$

The other two main results of our work are presented in Eqs. (5) and (7), whose derivations are presented in Sec. Methods. An upper bound for the MIR is given by

$$I_C = \lambda_1 - \lambda_2 = \lambda_1(2 - D), \tag{5}$$

where λ_1 and λ_2 represent the largest and the second largest Lyapunov exponent measured in Σ_Ω , if both exponents are positive. If the i -largest exponent is negative, then we set $\lambda_i = 0$. If the set Σ_Ω represents a periodic orbit, $I_C = 0$, and therefore there is no information being exchanged. The quantity D is defined as

$$D = - \frac{\log(N_C(t=T))}{\log(\epsilon)}, \tag{6}$$

where $N_C(t=T)$ is the number of boxes that would be covered by fictitious points at time T . At time $t=0$, these fictitious points are confined in an ϵ -square box. They expand not only exponentially fast in both directions according to the two positive Lyapunov exponents, but expand forming a compact set, a set with no “holes”. At $t = T$, they spread over Σ_Ω .

A lower bound for the MIR is given by

$$I'_C = \lambda_1(2 - \tilde{D}_0), \tag{7}$$

where \tilde{D}_0 represents the capacity dimension of the set Σ_Ω

$$\tilde{D}_0 = \lim_{\epsilon \rightarrow 0} \left[- \frac{\log(\tilde{N}_C(\epsilon))}{\log(\epsilon)} \right], \tag{8}$$

where \tilde{N}_C represents the number of boxes in Ω that are occupied by points of Σ_Ω .

D is defined in a way similar to the capacity dimension, though it is not the capacity dimension. In fact, $D \leq \tilde{D}_0$, because \tilde{D}_0 measures the change in the number of occupied boxes in Ω as the space resolution varies, whereas D measures the relative number of boxes with a certain fixed resolution ϵ that would be occupied by the fictitious points (in Ω) after being iterated for a time T . As a consequence, the empty space in Ω that is not occupied by Σ_Ω does not contribute to the calculation of \tilde{D}_0 , whereas it contributes to the calculation of the quantity D . In addition, $N_C \geq \tilde{N}_C$ (for any ϵ), because while the fictitious points form a compact set expanding with the same ratio as the one for which the real points expand (ratio provided by the Lyapunov exponents), the real set of points Σ_Ω might not occupy many boxes.

Methods

Mixing, Correlation Decay and Invariant Measures

Denote by $F^T(x)$ a mixing transformation that represents how a point $x \in \Sigma_\Omega$ is mapped after a time T into Σ_Ω , and let $\rho(x)$ to represent the probability of finding a point of Σ_Ω in x (natural invariant density). Let I'_1 represent a region in Ω . Then, $\mu(I'_1) = \int \rho(x) dx$, for $x \in I'_1$ represents the probability measure of

the region I'_1 . Given two square boxes $I'_1 \in \Omega$ and $I'_2 \in \Omega$, if F^T is a mixing transformation, then for a sufficiently large T , we have that the correlation defined as

$$C(T) = \mu[F^{-T}(I'_1) \cap I'_2] - \mu[I'_1] \mu[I'_2], \tag{9}$$

decays to zero, the probability of having a point in I'_1 that is mapped to I'_2 is equal to the probability of being in I'_1 times the probability of being in I'_2 . That is typically what happens in random processes.

Notice that $\mu[F^{-T}(I'_1) \cap I'_2]$ can be interpreted as a joint entropy defined by the probability of being at I'_2 times the conditional probability (that defines elements in a transition matrix) of transferring from the set I'_2 to I'_1 .

If the measure $\mu(\Sigma_\Omega)$ is invariant, then $\mu([F^{-T}(\Sigma_\Omega)]) = \mu(\Sigma_\Omega)$. Mixing and ergodic systems produce measures that are invariant.

Derivation of the Mutual Information Rate (MIR) in Dynamical Networks and Data Sets

We consider that the dynamical networks or data sets to be analysed present either the mixing property or have fast decay of correlations, and their probability measure is time invariant. If a system that is mixing for a time interval T is observed (sampled) once every time interval T , then the probabilities generated by these snapshot observations behave as if they were independent, and the system behaves as if it were a random process. This is so because if a system is mixing for a time interval T , then the correlation $C(T)$ decays to zero for this time interval. For systems that have some decay of correlation, surely the correlation decays to zero after an infinite time interval. But, this time interval can also be finite, as shown in Information S1.

Consider now that we have experimental points and they are sampled once every time interval T . If the system is mixing, then the probability $\tilde{P}_{XY}((i,j),(k,l))$ of the sampled trajectory to be in the box with coordinates (i,j) and then be iterated to the box (k,l) depends exclusively on the probabilities of being at the box (i,j) , represented by $\tilde{P}_{XY}(i,j)$, and being at the box (k,l) , represented by $\tilde{P}_{XY}(k,l)$.

Therefore, for the sampled trajectory, $\tilde{P}_{XY}((i,j),(k,l)) = \tilde{P}_{XY}(i,j) \tilde{P}_{XY}(k,l)$. Analogously, the probability $\tilde{P}_X((i),(k))$ (or $\tilde{P}_Y((j),(l))$) of the sampled trajectory to be in the column i (or row j) of the grid and then be iterated to the column k (or row l) is given by $\tilde{P}_X((i),(k)) = \tilde{P}_X(i) \tilde{P}_X(k)$ (or $\tilde{P}_Y((j),(l)) = \tilde{P}_Y(j) \tilde{P}_Y(l)$).

The MIR of the experimental non-sampled trajectory points can be calculated from the mutual information of the sampled trajectory points \tilde{I}_S that follow itineraries of length n :

$$MIR = \lim_{n \rightarrow \infty} \frac{\tilde{I}_S(n)}{nT}, \tag{10}$$

Due to the absence of correlations of the sampled trajectory points, the mutual information for these points following itineraries of length n can be written as

$$\tilde{I}_S(n) = n[\tilde{H}_X(n=1) + \tilde{H}_Y(n=1) - \tilde{H}_{XY}(n=1)], \tag{11}$$

where $\tilde{H}_X(n=1) = - \sum_i \tilde{P}_X(i) \log[\tilde{P}_X(i)]$, $\tilde{H}_Y(n=1) = - \sum_j \tilde{P}_Y(j) \log[\tilde{P}_Y(j)]$, and $\tilde{H}_{XY}(n=1) = - \sum_{i,j} \tilde{P}_{XY}(i,j) \log[\tilde{P}_{XY}(i,j)]$, and $\tilde{P}_X(i)$, $\tilde{P}_Y(j)$, and $\tilde{P}_{XY}(i,j)$ represent the

probability of the sampled trajectory points to be in the column i of the grid, in the row j of the grid, and in the box (i,j) of the grid, respectively.

Due to the time invariance of the set Σ_Ω assumed to exist, the probability measure of the non-sampled trajectory is equal to the probability measure of the sampled trajectory. If a system that has a time invariant measure is observed (sampled) once every time interval T , the observed set has the same natural invariant density and probability measure of the original set. As a consequence, if Σ_Ω has a time invariant measure, the probabilities $P_X(i)$, $P_Y(j)$, and $P_{XY}(i,j)$ (used to calculate I_S) are equal to $\tilde{P}_X(i)$, $\tilde{P}_Y(j)$, and $\tilde{P}_{XY}(i,j)$.

Consequently, $\tilde{H}_X(n=1)=H_X$, $\tilde{H}_Y(n=1)=H_Y$, and $\tilde{H}_{XY}(n=1)=H_{XY}$, and therefore $\tilde{I}_S(n)=nI_S$. Substituting into Eq. (10), we finally arrive to $MIR = \frac{I_S}{T}$ in Eq. (3), where I_S between two nodes is calculated from Eq. (1).

Therefore, in order to calculate the MIR, we need to estimate the time T for which the correlation of the system approaches zero and the probabilities $P_X(i)$, $P_Y(j)$, $P_{XY}(i,j)$ of the experimental non-sampled experimental points to fall in the column i of the grid, in the row j of the grid, and in the box (i,j) of the grid, respectively.

We demonstrate the validity of Eqs. (10) and (11) by showing that $\tilde{I}_S(n=2)=2\tilde{I}_S(n=1)$, which leads to Eq. (3). For the following demonstration, (i,j) (or (k,l)) represents a box in the subspace Ω placed at coordinates (i,j) , meaning a square of sides ϵ whose lower left corner point is located at $((i-1)\epsilon,(j-1)\epsilon)$. Then, i (or k) represents a column with width ϵ in Ω whose left side is located at $(i-1)\epsilon$ (or $(k-1)\epsilon$) and j (or l) represents a row with width ϵ in Ω whose bottom side is located at $(j-1)\epsilon$ (or $(l-1)\epsilon$).

If the system is mixing for a time T , then the probability of having points in a box (i,j) and going to another box (k,l) , i.e., $P_{XY}(F^{-T}(k,l)\cap(i,j))$ can be calculated by

$$P_{XY}[F^{-T}(k,l)\cap(i,j)] = \tilde{P}_{XY}((i,j),(k,l)) = \tilde{P}_{XY}(i,j)\tilde{P}_{XY}(k,l), \quad (12)$$

Notice that $P_{XY}[F^{-T}(k,l)\cap(i,j)]$ is a joint entropy that is equal to $\tilde{P}_{XY}((i,j),(k,l))$, and could be written as a function of conditional probabilities: $\tilde{P}_{XY}((i,j),(k,l)) = \tilde{P}_{XY}(i,j)\tilde{P}_{XY}((i,j)|(k,l))$, where $\tilde{P}_{XY}((i,j)|(k,l))$ represents the conditional probability of being transferred from the box (i,j) to the box (k,l) .

The same can be done to calculate the probability of having points in a column i that are mapped to another column k , i.e. $\tilde{P}_X(i,(k))$, or of having points in a row j that are mapped to another row l , i.e. $\tilde{P}_Y(j,(l))$. If the system is mixing for a time T , then

$$P_X[F^{-T}(k)\cap(i)] = \tilde{P}_X((i),(k)) = \tilde{P}_X(i)\tilde{P}_X(k) \quad (13)$$

and

$$P_Y[F^{-T}(l)\cap(j)] = \tilde{P}_Y((j),(l)) = \tilde{P}_Y(j)\tilde{P}_Y(l) \quad (14)$$

for the rows. Notice that $P_X(i) = \sum_{j=1}^N P_{XY}(i,j)$ and $P_Y(j) = \sum_{i=1}^N P_{XY}(i,j)$.

The order-2 Mutual information of the sampled points can be calculated by.

$$\tilde{I}_S(n=2) = \sum_{i,j=1}^N \sum_{k,l=1}^N \tilde{P}_{XY}((i,j),(k,l)) \log \left[\frac{\tilde{P}_{XY}((i,j),(k,l))}{\tilde{P}_X((i),(k))\tilde{P}_Y((j),(l))} \right], \quad (15)$$

where $\sum_{i,j=1}^N = \sum_{i=1}^N \sum_{j=1}^N$. $\tilde{I}_S(n=2)$ measures the MI of points that follow an itinerary of one iteration, points that are in a box and are iterated to another box. Substituting Eq. (12) in Eq. (15) we arrive at

$$\begin{aligned} \tilde{I}_S(n=2) &= \sum_{i,j=1}^N \sum_{k,l=1}^N \tilde{P}_{XY}(i,j)\tilde{P}_{XY}(k,l) \\ &\times (\log[\tilde{P}_{XY}(i,j)] + \log[\tilde{P}_{XY}(k,l)] - \\ &\log[\tilde{P}_X((i),(k))] - \log[\tilde{P}_Y((j),(l))]). \end{aligned} \quad (16)$$

Then, substituting (13) and (14) in Eq. (16), and using the fact that $\sum_{i,j} \tilde{P}_{XY}(i,j) = 1$ and $\sum_{k,l} \tilde{P}_{XY}(k,l) = 1$, we arrive at

$$\begin{aligned} \tilde{I}_S(n=2) &= 2 \sum_{i,j=1}^N \tilde{P}_{XY}(i,j) \log[\tilde{P}_{XY}(i,j)] - \\ &\sum_{i,j=1}^N \sum_{k,l=1}^N \tilde{P}_{XY}(i,j)\tilde{P}_{XY}(k,l) (\log[\tilde{P}_X(i)] + \log[\tilde{P}_X(k)] + \\ &\log[\tilde{P}_Y(j)] + \log[\tilde{P}_Y(l)]) \end{aligned} \quad (17)$$

Re-organizing the terms we arrive at

$$\begin{aligned} \tilde{I}_S(n=2) &= 2 \sum_{i,j=1}^N \tilde{P}_{XY}(i,j) \log[\tilde{P}_{XY}(i,j)] \\ &- \sum_{k,l=1}^N \tilde{P}_{XY}(k,l) \sum_{i=1}^N \log[\tilde{P}_X(i)] \sum_{j=1}^N \tilde{P}_{XY}(i,j) + \dots, \end{aligned} \quad (18)$$

where \dots represents other terms that are similar to the term appearing in the last hand-side part of the previous equation. Using the fact that $\sum_{j=1}^N \tilde{P}_{XY}(i,j) = \tilde{P}_X(i)$, we arrive at

$$\begin{aligned} \tilde{I}_S(n=2) &= \\ &2 \sum_{i,j=1}^N \tilde{P}_{XY}(i,j) \log[\tilde{P}_{XY}(i,j)] - \sum_{i=1}^N \log[\tilde{P}_X(i)]\tilde{P}_X(i) + \dots, \end{aligned} \quad (19)$$

which can then be written as

$$\begin{aligned} \tilde{I}_S(n=2) &= 2 \sum_{i,j=1}^N \tilde{P}_{XY}(i,j) \log[\tilde{P}_{XY}(i,j)] \\ &- \sum_{i=1}^N \tilde{P}_X(i) \log[\tilde{P}_X(i)] - \sum_{j=1}^N \tilde{P}_Y(j) \log[\tilde{P}_Y(j)] - \\ &\sum_{k=1}^N \tilde{P}_X(k) \log[\tilde{P}_X(k)] - \sum_{l=1}^N \tilde{P}_Y(l) \log[\tilde{P}_Y(l)]. \end{aligned} \quad (20)$$

Since $\sum_{i=1}^N \tilde{P}_X(i) \log[\tilde{P}_X(i)] = \sum_{k=1}^N \tilde{P}_X(k) \log[\tilde{P}_X(k)]$ and $\sum_{j=1}^N \tilde{P}_Y(j) \log[\tilde{P}_Y(j)] = \sum_{l=1}^N \tilde{P}_Y(l) \log[\tilde{P}_Y(l)]$, we finally arrive at that $\tilde{I}_S(n=2) = 2\tilde{I}_S(n=1)$. Similar calculations can be performed to state that $\tilde{I}_S(n) = n\tilde{I}_S(n=1)$. As previously discussed, $\tilde{I}_S(n=1) = I_S$, which lead us to Eq. (3).

Derivation of an Upper (I_C) and Lower (I_C^l) Bounds for the MIR

Consider that our attractor Σ is generated by a 2d expanding system with constant Jacobian that possesses two positive Lyapunov exponents λ_1 and λ_2 , with $\lambda_1 \geq \lambda_2$. $\Sigma \in \Omega$. Imagine a box whose sides are oriented along the orthogonal basis used to calculate the Lyapunov exponents. Then, points inside the box spread out after a time interval t to $\epsilon\sqrt{2}\exp^{\lambda_1 t}$ along the direction from which λ_1 is calculated. At $t=T$, $\epsilon\sqrt{2}\exp^{\lambda_1 T} = L$, which provides T in Eq. (4), since $L = \sqrt{2}$. These points spread after a time interval t to $\epsilon\sqrt{2}\exp^{\lambda_2 t}$ along the direction from which λ_2 is calculated. After an interval of time $t=T$, these points spread out over the set Σ_Ω . We require that for $t \leq T$, the distance between these points only increases: the system is expanding.

Imagine that at $t=T$, fictitious points initially in a square box occupy an area of $\epsilon\sqrt{2}\exp^{\lambda_2 T} L = 2\epsilon^2 \exp^{(\lambda_2 + \lambda_1)T}$. Then, the number of boxes of sides ϵ that contain fictitious points can be calculated by $N_C = 2\epsilon^2 \exp^{(\lambda_1 + \lambda_2)T} / 2\epsilon^2 = \exp^{(\lambda_1 + \lambda_2)T}$. From Eq. (4), $N = \exp^{\lambda_1 T}$, since $N = 1/\epsilon$.

We denote with a lower-case format, the probabilities $p_X(i)$, $p_Y(j)$, and $p_{XY}(i,j)$ with which fictitious points occupy the grid in Ω . If these fictitious points spread uniformly forming a compact set whose probabilities of finding points in each fictitious box is equal, then $p_X(i) = 1/N$ ($= \frac{1}{N_C} \frac{N_C}{N}$), $p_Y(j) = 1/N$, and $p_{XY}(i,j) = 1/N_C$.

Let us denote the Shannon entropy of the probabilities $p_X(i)$, $p_Y(j)$ and $p_{XY}(i,j)$ as h_X , h_Y , and h_{XY} , respectively. The mutual information of the fictitious trajectories after evolving a time interval T can be calculated by $I_S^u = h_X + h_Y - h_{XY}$. Since, $p_X(i) = p_Y(j) = 1/N$ and $p_{XY}(i,j) = 1/N_C$, then $I_S^u = 2\log(N) - \log(N_C)$. At $t=T$, we have that $N = \exp^{\lambda_1 T}$ and $N_C = \exp^{(\lambda_1 + \lambda_2)T}$, leading us to $I_S^u = (\lambda_1 - \lambda_2)T$. Therefore, defining, $I_C = I_S^u/T$, we arrive at $I_C = \lambda_1 - \lambda_2$.

We define D as

$$D = -\frac{\log(N_C(t=T))}{\log(\epsilon)}, \tag{21}$$

where $N_C(t=T)$ being the number of boxes that would be covered by fictitious points at time T . At time $t=0$, these fictitious points are confined in an ϵ -square box. They expand not only exponentially fast in both directions according to the two positive Lyapunov exponents, but expand forming a compact set, a set with no ‘‘holes’’. At $t=T$, they spread over Σ_Ω .

Using $\epsilon = \exp^{-\lambda_1 T}$ and $N_C = \exp^{(\lambda_1 + \lambda_2)T}$ in Eq. (21), we arrive at $D = 1 + \frac{\lambda_2}{\lambda_1}$, and therefore, we can write that $I_C = \lambda_1 - \lambda_2 = \lambda_1(2 - D)$, as in Eq. (5).

To calculate the maximal possible MIR, of a random independent process, we assume that the expansion of points is uniform only along the columns and rows of the grid defined in the space Ω , i.e., $P_X(i) = P_Y(j) = 1/N$, (which maximises H_X and H_Y), and we allow $P_{XY}(i,j)$ to be not uniform (minimising H_{XY}) for all i and j , then

$$I_S(\epsilon) = -2\log(\epsilon) + \sum_{i,j} P_{XY}(i,j) \log[P_{XY}(i,j)]. \tag{22}$$

Since $T(\epsilon) = -1/\lambda_1 \log(\epsilon)$, dividing $I_S(\epsilon)$ by $T(\epsilon)$, taking the limit of $\epsilon \rightarrow 0$, and reminding that the information dimension of the set Σ_Ω in the space Ω is defined as $\tilde{D}_1 = \lim_{\epsilon \rightarrow 0} \frac{\sum_{i,j} P_{XY}(i,j) \log[P_{XY}(i,j)]}{\log(\epsilon)}$, we obtain that the MIR is given by

$$I_S/T = \lambda_1(2 - \tilde{D}_1). \tag{23}$$

Since $\tilde{D}_1 \leq \tilde{D}_0$ (for any value of ϵ), then $\lambda_1(2 - \tilde{D}_1) \geq \lambda_1(2 - \tilde{D}_0)$, which means that a lower bound for the maximal MIR [provided by Eq. (23)] is given by $I_C^l = \lambda_1(2 - \tilde{D}_0)$, as in Eq. (7). But $D \leq \tilde{D}_0$ (for any value of ϵ), and therefore I_C is an upper bound for I_C^l .

To show why I_C is an upper bound for the maximal possible MIR, assume that the real points Σ_Ω occupy the space Ω uniformly. If $\tilde{N}_C > N$, there are many boxes being occupied. It is to be expected that the probability of finding a point in a column or a row of the grid is $P_X(i) = P_Y(j) \cong 1/N$, and $P_{XY}(i,j) \cong 1/\tilde{N}_C$. In such a case, $MIR \cong I_C^l$, which implies that $I_C \geq MIR$. If $\tilde{N}_C < N$, there are only few boxes being sparsely occupied. The probability of finding a point in a column or a row of the grid is $P_X(i) = P_Y(j) \cong 1/\tilde{N}_C$, and $P_{XY}(i,j) \cong 1/\tilde{N}_C$. There are \tilde{N}_C columns and rows being occupied by points in the grid. In such a case, $I_S \cong 2\log(\tilde{N}_C) - \log(\tilde{N}_C) \cong \log(\tilde{N}_C)$. Comparing with $I_S^u = 2\log(N) - \log(N_C)$, and since $\tilde{N}_C < N$ and $N_C \geq \tilde{N}_C$, then we conclude that $I_S^u \geq I_S$, which implies that $I_C \geq MIR$.

Notice that if $P_{XY}(i,j) = p_{XY}(i,j) = 1/N_C$ and $\tilde{D}_1 = \tilde{D}_0$, then $I_S/T = I_C^l = I_C$.

Expansion Rates

In order to extend our approach for the treatment of data sets coming from networks whose equations of motion are unknown, or for higher-dimensional networks and complex systems which might be neither rigorously chaotic nor fully deterministic, or for experimental data that contains noise and few sampling points, we write our bounds in terms of expansion rates defined in this work by

$$e_k(t) = 1/\tilde{N}_C \sum_{i=1}^{\tilde{N}_C} \frac{1}{t} \log[L_k^i(t)], \tag{24}$$

where we consider $k=1,2$. $L_1^i(t)$ measures the largest growth rate of nearby points. In practice, it is calculated by $L_1^i(t) = \frac{\Delta}{\delta}$, with δ representing the largest distance between pairs of points in an ϵ -square box i and Δ representing the largest distance between pairs of the points that were initially in the ϵ -square box but have spread out for an interval of time t . $L_2^i(t)$ measures how an area enclosing points grows. In practice, it is calculated by $L_2^i(t) = \frac{A}{\epsilon^2}$, with ϵ^2 representing the area occupied by points in an ϵ -square box, and A the area occupied by these points after spreading out for a time interval t . There are \tilde{N}_C boxes occupied by points which are taken into consideration in the calculation of $L_k^i(t)$. An order- k

expansion rate, $e_k(t)$, measures on average how a hypercube of dimension k exponentially grows after an interval of time t . So, e_1 measures the largest growth rate of nearby points, a quantity closely related to the largest finite-time Lyapunov exponent [31]. And e_2 measures how an area enclosing points grows, a quantity closely related to the sum of the two largest positive Lyapunov exponents. In terms of expansion rates, Eqs. (4) and (5) read

$$T = \frac{1}{e_1} \log \left[\frac{1}{\epsilon} \right] \quad \text{and} \quad I_C = e_1(2-D), \quad \text{respectively, and Eqs. (6) and (7) read } D(t) = \frac{e_2(t)}{e_1(t)} \quad \text{and} \quad I_C^l = e_1(2-\tilde{D}_0), \quad \text{respectively.}$$

From the way we have defined expansion rates, we expect that $e_k \leq \sum_{i=1}^k \lambda_i$. Because of the finite time interval and the finite size of the regions of points considered, regions of points that present large derivatives, contributing largely to the Lyapunov exponents, contribute less to the expansion rates. If a system has constant Jacobian, is uniformly hyperbolic, and has a constant natural measure, then $e_k = \sum_{i=1}^k \lambda_i$.

There are many reasons for using expansion rates in the way we have defined them in order to calculate bounds for the MIR. Firstly, because they can be easily experimentally estimated whereas Lyapunov exponents demand more computational efforts. Secondly, because of the macroscopic nature of the expansion rates, they might be more appropriate to treat data coming from complex systems that contain large amounts of noise, data that have points that are not (arbitrarily) close as formally required for a proper calculation of the Lyapunov exponents. Thirdly, expansion rates can be well defined for data sets containing very few data points: the fewer points a data set contains, the larger the regions of size ϵ need to be and the shorter the time T is. Finally, expansion rates are defined in a similar way to finite-time Lyapunov exponents and thus some algorithms used to calculate Lyapunov exponents can be used to calculate our defined expansion rates.

Results and Discussion

MIR and its Bounds in Two Coupled Chaotic Maps

To illustrate the use of our bounds, we consider the following two bidirectionally coupled maps.

$$\begin{aligned} X_{n+1}^{(1)} &= 2X_n^{(1)} + \rho X_n^{(2)^2} + \Sigma(X_n^{(2)} - X_n^{(1)}), \text{ mod } 1 \\ X_{n+1}^{(2)} &= 2X_n^{(2)} + \rho X_n^{(1)^2} + \Sigma(X_n^{(1)} - X_n^{(2)}), \text{ mod } 1 \end{aligned} \quad (25)$$

where $X_n^{(i)} \in [0,1]$. If $\rho=0$, the map is piecewise-linear and quadratic, otherwise. We are interested in measuring the exchange of information between $X^{(1)}$ and $X^{(2)}$. The space Ω is the unit square. The Lyapunov exponents measured in the space Ω are the Lyapunov exponents of the set Σ_Ω that is the chaotic attractor generated by Eqs. (25).

The quantities I_S/T , I_C , and I_C^l are shown in Fig. 1 as we vary Σ for $\rho=0$ (A) and $\rho=0.1$ (B). We calculate I_S using in Eq. (1) the probabilities $P_{XY}(i,j)$ in which points from a trajectory composed of 2,000,000 samples fall in boxes of sides $\epsilon=1/500$ and the probabilities $P_X(i)$ and $P_Y(j)$ that the points visit the intervals $[(i-1)\epsilon, i\epsilon[$ of the variable $X_n^{(1)}$ or $[(j-1)\epsilon, j\epsilon[$ of the variable $X_n^{(2)}$, respectively, for $i,j=1, \dots, N$. When computing I_S/T , the quantity T was estimated by Eq. (4). Indeed for most values of Σ , $I_C \geq I_S/T$ and $I_C^l \leq I_S/T$.

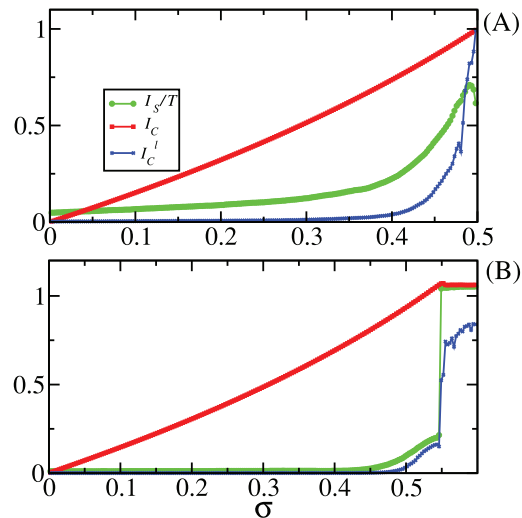


Figure 1. Results for two coupled maps. I_S/T [Eq. (3)] as (green online) filled circles, I_C [Eq. (5)] as the (red online) thick line, and I_C^l [Eq. (7)] as the (blue online) crosses. In (A) $\rho=0$ and in (B) $\rho=0.1$. The units of I_S/T , I_C , and I_C^l are [bits/iteration]. doi:10.1371/journal.pone.0046745.g001

For $\Sigma=0$ there is no coupling, and therefore the two maps are independent from each other. There is no information being exchanged. In fact, $I_C=0$ and $I_C^l \cong 0$ in both figures, since $D=\tilde{D}_0=2$, meaning that the attractor Σ_Ω fully occupies the space Ω . This is a remarkable property of our bounds: to identify that there is no information being exchanged when the two maps are independent. Complete synchronisation is achieved and I_C is maximal, for $\Sigma > 0.5$ (A) and for $\Sigma \geq 0.55$ (B). A consequence of the fact that $D=\tilde{D}_0=1$, and therefore, $I_C=I_C^l=\lambda_1$. The reason is because for this situation this coupled system is simply the shift map, a map with constant natural measure; therefore $P_X(i)=P_Y(j)$ and $P_{XY}(i,j)$ are constant for all i and j . As usually happens when one estimates the mutual information by partitioning the phase space with a grid having a finite resolution and data sets possessing a finite number of points, I_S is typically larger than zero, even when there is no information being exchanged ($\Sigma=0$). Even when there is complete synchronisation, we find non-zero off-diagonal terms in the matrix for the joint probabilities causing I_S to be smaller than it should be. Due to numerical errors, $X^{(1)} \cong X^{(2)}$, and points that should be occupying boxes with two corners exactly along a diagonal line in the subspace Ω end up occupying boxes located off-diagonal and that have at least three corners off-diagonal. Due to such problems, I_S/T is underestimated by an amount of $\lambda_1 \frac{3}{N} \approx 0.18\lambda_1$, resulting in a value of approximately $I_S/T = 0.82\lambda_1$, close to the value of I_S/T shown in Fig. 1(A), for $\Sigma=0.5$. The estimation of the lower bound I_C^l in (B) suffers from the same problems.

Our upper bound I_C is calculated assuming that there is a fictitious dynamics expanding points (and producing probabilities) not only exponentially fast but also uniformly. The “experimental” numerical points from Eqs. (25) expand exponentially fast, but not uniformly. Most of the time the trajectory remains in 4 points: $(0,0)$, $(1,1)$, $(1,0)$, $(0,1)$. That is the main reason of why I_C is much larger than the estimated real value of the MIR, for some coupling strengths. If two nodes in a dynamical network behave in the same

way the fictitious dynamics does, these nodes would be able to exchange the largest possible amount of information.

We would like to point out that one of the main advantages of calculating upper bounds for the MIR (I_S/T) using Eq. (5) instead of actually calculating I_S/T is that we can reproduce the curves for I_C using much less number of points (1000 points) than the ones (2,000,000) used to calculate the curve for I_S/T . If $\rho=0$, $I_C = -\ln(1-\Sigma)$ can be calculated since $\lambda_1 = \ln(2)$ and $\lambda_2 = \ln(2-2\Sigma)$.

MIR and its Bounds in Experimental Networks of Double-Scroll Circuits

We illustrate our approach for the treatment of data sets using a network formed by an inductorless version of the Double-Scroll circuit [32]. We consider four networks of bidirectionally diffusively coupled circuits (see Fig. 2). Topology I in (A) represents two bidirectionally coupled circuits, Topology II in (B), three circuits coupled in an open-ended array, Topology III in (C), four circuits coupled in an open-ended array, and Topology IV in (D), coupled in an closed array. We choose two circuits in the different networks (one connection apart) and collect from each circuit a time-series of 79980 points, with a sampling rate of $\delta=80.000$ samples/s. The measured variable is the voltage across one of the circuit capacitors, which is normalised in order to make the space Ω to be a square of sides 1. Such normalisation does not alter the quantities that we calculate. The following results provide the exchange of information between these two chosen circuits. The values of ϵ and t used to course-grain the space Ω and to calculate e_2 in Eq. (24) are the ones that minimise $|N_C(T, e_2) - \tilde{N}_C(\epsilon)|$ and at the same time satisfy $N_C(T, e_2) \geq \tilde{N}_C(\epsilon)$, where $N_C(T, e_2) = \exp^{T e_2(t)}$ represents the number of fictitious boxes covering the set Σ_Ω in a compact fashion, when $t=T$. This optimisation excludes some non-significant points that make the expansion rate of fictitious points to be much larger than it should be. In other words, we require that e_2 describes well the way most of the points spread. We consider that t used to calculate e_k in Eq. (24) is the time points initially in an ϵ -side box to become at most apart by $0.8L$. That guarantees that nearby points in Σ_Ω are expanding in both directions within the time interval $[0, T]$. Assuming that $t=T$ is calculated by measuring the time points initially in an ϵ -side box to be at most apart by $[0.4L, 0.8L]$ produces already similar results. If $t=T$ is calculated by measuring the time points become at least apart by $0.8L$, the set Σ_Ω might not be only expanding. T might be overestimated.

I_S has been estimated by the method in Ref. [33]. Since we assume that the space Ω where mutual information is being measured is 2D, we will compare our results by considering in the method of Ref. [33] a 2D space formed by the two collected scalar signals. In the method of Ref. [33] the phase space is partitioned in regions that contain 30 points of the continuous trajectory. Since that these regions do not have equal areas (as it is the case for I_C and I_C^l), in order to estimate T we need to imagine a box of sides ϵ_k , such that its area ϵ_k^2 contains in average 30 points. The area occupied by the set Σ_Ω is approximately given by $\epsilon^2 \tilde{N}_C$, where \tilde{N}_C is the number of occupied boxes. Assuming that the 79980 experimental data points occupy the space Ω uniformly, then on average 30 points would occupy an area of $\frac{30}{79980} \epsilon^2 \tilde{N}_C$. The square root of this area is the side of the imaginary box that would occupy 30 points. So, $\epsilon_k = \sqrt{\frac{30}{79980} \tilde{N}_C} \epsilon$. Then, in the following, the ‘‘exact’’ value of the MIR will be considered to be given by

$$I_S/T_k, \text{ where } T_k \text{ is estimated by } T_k = -\frac{1}{e_1} \log(\epsilon_k).$$

The three main characteristics of the curves for the quantities I_S/T_k , I_C , and I_C^l (appearing in Fig. 3) with respect to the coupling strength are that (i) as the coupling resistance becomes smaller, the coupling strength connecting the circuits becomes larger, and the level of synchronisation increases leading to an increase in I_S/T_k , I_C , and I_C^l , (ii) all curves are close, (iii) and as expected, for most of the resistance values, $I_C > I_S/T_k$ and $I_C^l \leq I_S/T_k$. The two main synchronous phenomena appearing in these networks are almost synchronisation (AS) [34], when the circuits are almost completely synchronous, and phase synchronisation (PS) [35]. For the circuits considered in Fig. 3, AS appears for the interval $R \in [0, 3]$ and PS appears for the interval $R \in [3, 3.5]$. Within this region of resistance values the exchange of information between the circuits becomes large. PS was detected by using the technique from Refs. [36,37].

MIR and its Upper Bound in Stochastic Systems

To analytically demonstrate that the quantities I_C and I_S/T can be well calculated in stochastic systems, we consider the following stochastic dynamical toy model illustrated in Fig. 4. In it points within a small box of sides ϵ (represented by the filled square in Fig. 4(A)) located in the centre of the subspace Ω are mapped after one iteration ($n=1, n \in \mathbb{N}$) of the dynamics to 12 other neighbouring boxes. Some points remain in the initial box. The points that leave the initial box go to 4 boxes along the diagonal line and 8 boxes off-diagonal along the transverse direction. Boxes along the diagonal are represented by the filled squares in Fig. 4(B) and off-diagonal boxes by filled circles. At the second iteration ($n=2$), the points occupy other neighbouring boxes, as illustrated in Fig. 4(C), and at a time T ($T \in \mathbb{R}$) the points occupy the attractor Σ and do not spread any longer. For iterations n larger than T , the points are somehow reinjected inside the region of the attractor. We consider that this system is completely stochastic, in the sense that no one can precisely determine the location of where an initial condition will be mapped. The only information is that points inside a smaller region are mapped to a larger region.

At the iteration n , there will be $N_d = 2^{1+n} + 1$ boxes occupied along the diagonal (filled squares in Fig. 4) and $N_t = 2nN_d - C(\tilde{n})$ (filled circles in Fig. 4) boxes occupied off-diagonal (along the transverse direction), where $C(\tilde{n}) = 0$ for $\tilde{n} = 0$, and $C(\tilde{n}) > 0$ for

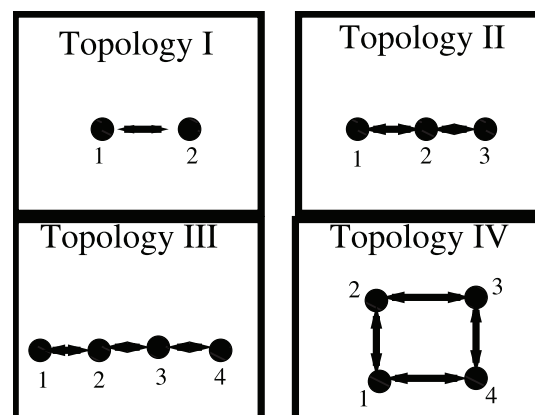


Figure 2. Black filled circles represent a Chua’s circuit and the numbers identify each circuit in the networks. Coupling is diffusive. We consider 4 topologies: 2 coupled Chua’s circuit (A), an array of 3 coupled circuits, an array of 4 coupled circuits, and a ring formed by 4 coupled circuits. doi:10.1371/journal.pone.0046745.g002

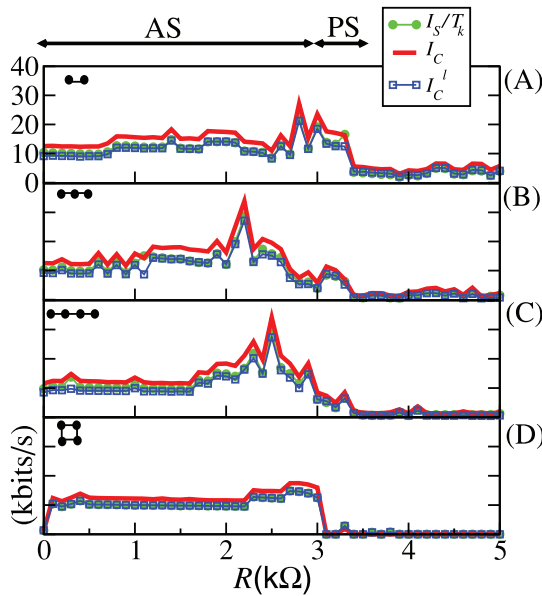


Figure 3. Results for experimental networks of Double-Scroll circuits. On the left-side upper corner pictograms represent how the circuits (filled circles) are bidirectionally coupled. I_S/T_k as (green online) filled circles, I_C as the (red online) thick line, and I_C^l as the (blue online) squares, for a varying coupling resistance R . The unit of these quantities shown in these figures is (kbits/s). (A) Topology I, (B) Topology II, (C) Topology III, and (D) Topology IV. In all figures, \tilde{D}_0 increases smoothly from 1.25 to 1.95 as R varies from 0.1kΩ to 5kΩ. The line on the top of the figure represents the interval of resistance values responsible to induce almost synchronisation (AS) and phase synchronisation (PS).
doi:10.1371/journal.pone.0046745.g003

$\tilde{n} \geq 1$ and $\tilde{n} = n - T - \alpha$. α is a small number of iterations representing the time difference between the time T for the points in the diagonal to reach the boundary of the space Ω and the time for the points in the off-diagonal to reach this boundary. The border effect can be ignored when the expansion along the diagonal direction is much faster than along the transverse direction.

At the iteration n , there will be $N_C = 2^{1+n} + 1 + (2^{1+n} + 1)2n - C(\tilde{n})$ boxes occupied by points. In the following calculations we consider that $N_C \cong 2^{1+n}(1 + 2n)$. We assume that the subspace Ω is a square whose sides have length 1, and that $\Sigma \in \Omega$, so $L = \sqrt{2}$. For $n > T$, the attractor does not grow any longer along the off-diagonal direction.

The largest Lyapunov exponent or the order-1 expansion rate of this stochastic toy model can be calculated by $N_d(n) \exp^{\lambda_1} = N_d(n + 1)$, which takes us to

$$\lambda_1 = \log(2). \tag{26}$$

Therefore, the time T , for the points to spread over the attractor Σ , can be calculated by the time it takes for points to visit all the boxes along the diagonal. It can be calculated by $\epsilon \sqrt{2} \exp^{\lambda_1 T} = \sqrt{2}$, which take us to

$$T = -\frac{\log(\epsilon)}{\lambda_1} = -\frac{\log(\epsilon)}{\log(2)}. \tag{27}$$

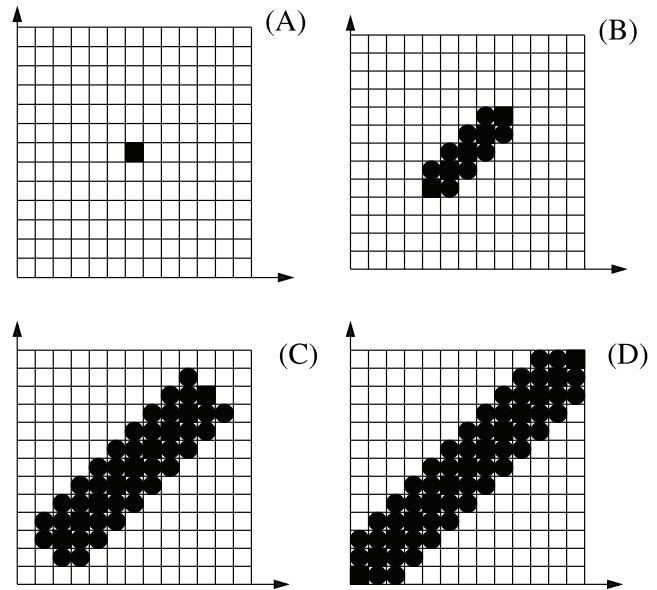


Figure 4. This picture is a hand-made illustration. Squares are filled as to create an image of a stochastic process whose points spread according to the given Lyapunov exponents. (A) A small box representing a set of initial conditions. After one iteration of the system, the points that leave the initial box in (A) go to 4 boxes along the diagonal [filled squares in (B)] and 8 boxes off-diagonal [filled circles in (B)]. At the second iteration, the points occupy other neighbouring boxes as illustrated in (C) and after an interval of time $n = T$ the points do not spread any longer (D).
doi:10.1371/journal.pone.0046745.g004

The quantity D can be calculated by $D = \frac{\log(N_C)}{\log(N)}$, with $n = T$.

Neglecting $C(\tilde{n})$ and the 1 appearing in N_C due to the initial box, we have that $N_C \cong 2^{1+T}[1 + 2^T]$. Substituting in the definition of D , we obtain $D = \frac{(1 + T) \log(2) + \log(1 + 2^T)}{-\log(\epsilon)}$. Using T from Eq. (27), we arrive at

$$D = 1 + r, \tag{28}$$

where

$$r = -\frac{\log(2)}{\log(\epsilon)} - \frac{\log(1 + 2^T)}{\log(\epsilon)} \tag{29}$$

Placing D and λ_1 in $I_C = \lambda_1(2 - D)$, gives us

$$I_C = \log(2)(1 - r). \tag{30}$$

Let us now calculate I_S/T . Ignoring the border effect, and assuming that the expansion of points is uniform, then $P_{XY}(i, j) = 1/N_C$ and $P_X(i) = P_Y(j) = 1/N = \epsilon$. At the iteration $n = T$, we have that $I_S = -2 \log(\epsilon) - \log(N_C)$. Since $N_C \cong 2^{1+T}[1 + 2^T]$, we can write that $I_S = -2 \log(\epsilon) - (1 + T) \log(2) - \log(1 + 2^T)$. Placing T from Eq. (27) into I_S takes us to $I_S = -\log(2) - \log(\epsilon) - \log(1 + 2^T)$. Finally, dividing I_S by T , we arrive that

$$\begin{aligned} \frac{I_S}{T} &= \log(2) \left[1 + \frac{\log(2)}{\log(\epsilon)} + \frac{\log(1+2^T)}{\log(\epsilon)} \right] \\ &= \log(2)(1-r). \end{aligned} \tag{31}$$

As expected from the way we have constructed this model, Eq. (31) and (30) are equal and $I_C = \frac{I_S}{T}$.

Had we included the border effect in the calculation of I_C , denote the value by I_C^b , we would have obtained that $I_C^b \geq I_C$, since λ_2 calculated considering a finite space Ω would be either smaller or equal than the value obtained by neglecting the border effect. Had we included the border effect in the calculation of I_S/T , denote the value by I_S^b/T , typically we would expect that the probabilities $P_{XY}(i,j)$ would not be constant. That is because the points that leave the subspace Ω would be randomly reinjected back to Ω . We would conclude that $I_S^b/T \leq I_S/T$. Therefore, had we included the border effect, we would have obtained that $I_C^b \geq I_S^b/T$.

The way we have constructed this stochastic toy model results in $D \cong 1$. This is because the spreading of points along the diagonal direction is much faster than the spreading of points along the off-diagonal transverse direction. In other words, the second largest Lyapunov exponent, λ_2 , is close to zero. For stochastic toy models which produce larger λ_2 , one could consider that the spreading along the transverse direction is given by $N_t = N_d 2^{2^m} - C(\tilde{n})$, with $\alpha \in [0,1]$.

Expansion Rates for Noisy Data with Few Sampling Points

In terms of the order-1 expansion rate, e_1 , our quantities read $I_C = e_1(2-D)$, $T = \frac{1}{e_1} \log \left[\frac{1}{\epsilon} \right]$, and $I_C^l = e_1(2-\tilde{D}_0)$. In order to show that our expansion rate can be used to calculate these quantities, we consider that the experimental system is being observed in a one-dimensional projection and points in this projection have a constant probability measure. Additive noise is assumed to be bounded with maximal amplitude η , and having constant density.

Our order-1 expansion rate is defined as

$$e_1(t) = 1/\tilde{N}_C \sum_{i=1}^{\tilde{N}_C} \frac{1}{t} \log[L_1^i(t)]. \tag{32}$$

where $L_1^i(t)$ measures the largest growth rate of nearby points. Since all it matters is the largest distance between points, it can be estimated even when the experimental data set has very few data points. Since, in this example, we consider that the experimental noisy points have constant uniform probability distribution, $e_1(t)$ can be calculated by

$$e_1(t) = \frac{1}{t} \log \left[\frac{\Delta + 2\eta}{\delta + 2\eta} \right]. \tag{33}$$

where $\delta + 2\eta$ represents the largest distance between pair of experimental noisy points in an ϵ -square box and $\Delta + 2\eta$ represents the largest distance between pair of the points that were initially in the ϵ -square box but have spread out for an interval of time t . The experimental system (without noise) is responsible to make points that are at most δ apart from each other to spread to at most to Δ apart from each other. These

points spread out exponentially fast according to the largest positive Lyapunov exponent λ_1 by

$$\Delta = \delta \exp^{\lambda_1 t}. \tag{34}$$

Substituting Eq. (34) in (33), and expanding \log to first order, we obtain that $e_1 = \lambda_1$, and therefore, our expansion rate can be used to estimate Lyapunov exponents.

Conclusions

We have shown a procedure to calculate mutual information rate (MIR) between two nodes (or groups of nodes) in dynamical networks and data sets that are either mixing, or exhibit fast decay of correlations, or have sensitivity to initial conditions, and we have proposed significant upper (I_C) and lower (I_C^l) bounds for it, in terms of the Lyapunov exponents, the expansion rates, and the capacity dimension.

Since our upper bound is calculated from Lyapunov exponents or expansion rates, it can be used to estimate the MIR between data sets that have different sampling rates or experimental resolution or between systems possessing a different number of events. For example, suppose one wants to understand how much information is exchanged between two time-series, the heart beat and the level of CO₂ in the body. The heart is monitored by an EEG that collects data with a high-frequency, whereas the monitoring of the CO₂ level happens in a much lower frequency. For every m points collected from an EEG one could collect $n < m$ points in the monitoring of the CO₂ level. Assuming that the higher-frequency variable (the heart beat) is the one that contributes mostly for the sensibility to the initial conditions, then the larger expansion rate (or Lyapunov exponent) can be well estimated only using this variable. The second largest expansion rate (or Lyapunov exponent) can be estimated by the composed subspace formed by these two measurements, but only the measurements taken simultaneously would be considered. Therefore, the estimation of the second largest expansion rate would have to be done using less points than the estimation used to obtain the largest. In the calculation of the second largest expansion rate, it is necessary to know the largest exponent. If the largest is correctly estimated, then the chances we make a good estimation of the second largest increases, even when only a few points are considered. With the two largest expansion rates, one can estimate I_C , the upper bound for the MIR.

Additionally, Lyapunov exponents can be accurately calculated even when data sets are corrupted by noise of large amplitude (observational additive noise) [38,39] or when the system generating the data suffers from parameter alterations (“experimental drift”) [40]. Our bounds link information (the MIR) and the dynamical behaviour of the system being observed with synchronisation, since the more synchronous two nodes are, the smaller λ_2 and D_0 will be. This link can be of great help in establishing whether two nodes in a dynamical network or in a complex system not only exchange information but also have linear or non-linear interdependences, since the approaches to measure the level of synchronisation between two systems are reasonably well known and are been widely used. If variables are synchronous in a time-lag fashion [35], it was shown in Ref. [16] that the MIR is independent of the delay between the two processes. The upper bound for the MIR could be calculated by measuring the Lyapunov exponents of the network (see Information S1), which are also invariant to time-delays between the variables.

If the MIR and its upper bounds are calculated from an “attractor” that is not an asymptotic limiting set but rather a transient trajectory, these values should typically differ from the values obtained when the “attractor” is an asymptotic limiting set. The dynamical quantities calculated, e.g., the Lyapunov exponents, expansion rates, and the fractal dimension should be interpreted as finite time quantities.

In our calculations, we have considered that the correlation of the system decays to approximately zero after a finite time T . If after this time interval the correlation does not decay to zero, we expect that I_S will be overestimated, leading to an overestimated value for the MIR. That is so because the probabilities used to calculate I_S will be considered to have been generated by a random system with uncorrelated variables, which is not true. However, by construction, the upper bound I_C is larger than the overestimated MIR.

References

- Shannon CE (1948) A Mathematical Theory of Communication. *Bell System Technical Journal* 27: 379–423.
- Strong SP, Koberle R, de Ruyter van Steveninck RR, Bialek W (1998) Entropy and Information in Neural Spike Trains. *Phys. Rev. Lett.* 80: 197–200.
- Sporns O, Chialvo DR, Kaiser M, Hilgetag CC (2004) Organization, development and function of complex brain networks. *Trends in Cognitive Sciences.* 8: 418–425.
- Palus M, Komárek V, Procházka T, Hrnčíř Z, Sterbova K (2001) Synchronization and information ow in EEGs of epileptic patients. *IEEE Engineering in Medicine and Biology Sep/Oct*: 65–71.
- Donges JF, Zou Y, Marwan N, Kurths J (2009) Complex networks in climate dynamics. *Eur. Phys. J.* 174: 157–179.
- Fraser AM, Swinney HL (1986) Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* 33: 1134–1140.
- Kantz H, Schreiber T (2004) *Nonlinear Time Series Analysis*. Cambridge: Cambridge University Press.
- Parlitz U (1998) *Nonlinear Time-Series Analysis*, in *Nonlinear Modelling - Advanced Black-Box techniques*. The Netherlands: Kluwer Academic Publishers.
- Haykin S (2001) *Communication Systems*. New York: John Wiley & Sons.
- Rossi F, Lendasse A, François D, Wertz V, Verleysen M (2006) Mutual information for the selection of relevant variables in spectrometric nonlinear modelling. *Chemometrics and Intelligent Laboratory Systems*, 80: 215–226.
- Paninski L (2003) Estimation of entropy and mutual information. *Neural Computation* 15: 1191–1253.
- Steuer R, Kurths J, Daub CO, Weckwerth W (2002) The Mutual Information: Detecting and evaluating dependencies between variables. *Bioinformatics* 18: S231–S240.
- Papana A, Kugiumtzis D, (2009) Evaluation of Mutual Information Estimators for Time Series. *Int. J. Bifurcation and Chaos* 19: 4197–4215.
- Baptista MS, Kurths J (2008) Information transmission in active channels. *Phys. Rev. E* 77: 026205–1–026205–13.
- Baptista MS, de Carvalho JX, Hussein MS (2008) Optimal network topologies for information transmission in active networks. *PLoS ONE* 3: e3479.
- Blanc JL, Pezard L, Lesne A (2011) Delay independence of mutual-information rate of two symbolic sequences. *Phys. Rev. E* 84: 036214–1–036214–9.
- Dobrushin RL (1959) General formulation of Shannon’s main theorem of information theory. *Usp. Mat. Nauk.* 14: 3–104; transl: *Amer. Math. Soc. Translations, series 2* 33: 323–438.
- Gray RM, Kieffer JC (1980) Asymptotically mean stationary measures. *IEEE Transactions on Information Theory* IT-26: 412–421.
- Verdú S, Han TS (1994) A general formula for channel capacity. *IEEE Trans. Information Theory* 40: 1147–1157.
- Kolmogorov AN (1958) A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces. *Dokl. Akad. Nauk SSSR* 119: 861–864;

Supporting Information

Information S1
(PDF)

Acknowledgments

M. S. Baptista would like to thank A. Politi for discussions concerning Lyapunov exponents and N. R. Obrero for discussions concerning MI and MIR.

Author Contributions

Conceived and designed the experiments: MSB RR EV JCS. Performed the experiments: RR EV JCS. Analyzed the data: MSB RR EV JCS UP. Contributed reagents/materials/analysis tools: MSB RR EV JCS UP. Wrote the paper: MSB RR EV JCS UP CG. Computer simulations: MSB. Analytical calculations: MSB. Analysis of analytical calculations: MSB UP.

- (1959) Entropy per unit time as a metric invariant of automorphisms. *Dokl. Akad. Nauk SSSR* 124: 754–755.
- Ruelle D (1978) An inequality for the entropy of differentiable maps. *Bol. Soc. Bras. Mat.* 9: 83–87.
- Pesin YaB (1977) Characteristic Lyapunov exponents and smooth ergodic theory. *Russ. Math. Surveys* 32: 55–114.
- Ledrappier F, Strelcyn JM (1982) A proof of the estimate from below in Pesin’s entropy formula. *Ergod. Theory Dyn. Syst.* 2: 203–219.
- Wissman BD, McKay-Jones LC, Binder PM, (2011) Entropy rate estimates from mutual information. *Phys. Rev. E* 84: 046204–1–046204–5.
- Gao JB (1999) Recurrence Time Statistics for Chaotic Systems and Their Applications. *Phys. Rev. Lett.* 83: 3178–3181.
- Baptista MS, Eulalie N, Pinto PRF, Brito M, Kurths J (2010) Density of first Poincaré returns, periodic orbits, and Kolmogorov-Sinai entropy. *Phys. Lett. A* 374: 1135–1140.
- Eckmann JP (2003) Non-equilibrium steady states. *arXiv:math-ph/0304043*.
- Sinai YaG (1970) Dynamical systems with elastic refections. *Ergodic properties of dispersing billiards.* *Russ. Math. Surv.* 25: 137–189.
- Chernov N, Young LS (2001) Decay of correlations for Lorentz gases and hard balls. *Encycl. Of Math. Sc., Math. Phys.* II 101: 89–120.
- Baptista MS, Caldas IL, Heller MVAP, Ferreira AA (2001) Onset of symmetric plasma turbulence. *Physica A*, 301: 150–162.
- Dawson S, Grebogi C, Sauer T, Yorke JA (1994) Obstructions to Shadowing When a Lyapunov Exponent Fluctuates about Zero. *Phys. Rev. Lett.* 73: 1927–1930.
- Albuquerque HA, Rubinger RM, Rech PC, (2007) Theoretical and experimental time series analysis of an inductorless Chua’s circuit. *Physics D* 233: 66–72.
- Kraskov A, Stogbauer H, Grassberger P (2004) Estimating mutual information. *Phys. Rev. E* 69: 066138-1-066138-16.
- Femat R, Sois-Perales G (1999) On the chaos synchronization phenomena. *Phys. Lett. A* 262: 50–60.
- Pikovsky A, Rosenblum M, Kurths J (2001) *Synchronization: A Universal Concept in Nonlinear Sciences*. Cambridge: Cambridge University Press.
- Baptista MS, Pereira T, Kurths J (2006) Upper bounds in phase synchronous weak coherent chaotic attractors. *Physica D* 216: 260–268.
- Pereira T, Baptista MS, Kurths J (2007) General framework for phase synchronization through localized sets. *Phys. Rev. E* 75: 026216–1–026216–12.
- Mera ME, Morán M (2009) Reduction of noise of large amplitude through adaptive neighborhoods. *Phys. Rev. E* 80: 016207–1–016207–8.
- Gao JB, Hu J, Tung WW, Cao YH, Distinguishing chaos from noise by scale-dependent Lyapunov exponent. *Phys. Rev. E* 74: 066204–1–066204–9.
- Stefański A (2008) Lyapunov exponents of systems with noise and actuating parameters. *Journal of Theoretical and Applied Mechanics* 46: 665–678.