 doi:10.20944/preprints201710.0067.v1

Article

Information Landscape and Flux, Mutual Information Rate Decomposition and Entropy Production

Qian Zeng [1,†] and Jin Wang [1,2,†,*]

1 State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Changchun, Jilin 130022, China; qzeng@ciac.ac.cn
2 Department of Chemistry and Physics, State University of New York, Stony Brook, NY 11794 USA; jin.wang.1@stonybrook.edu
* Correspondence: jin.wang.1@stonybrook.edu; Tel.: +1-631-632-1185
† These authors contributed equally to this work.

Version October 11, 2017 submitted to Entropy

Abstract: We explore the dynamics of information systems. We show that the driving force for information dynamics is determined by both the information landscape and information flux which determines the equilibrium time reversible and the nonequilibrium time-irreversible behaviours of the system respectively. We further demonstrate that the mutual information rate between the two subsystems can be decomposed into the time-reversible and time-irreversible parts respectively, analogous to the information landscape-flux decomposition for dynamics. Finally, we uncover the intimate relation between the nonequilibrium thermodynamics in terms of the entropy production rates and the time-irreversible part of the mutual information rate. We demonstrate the above features by the dynamics of a bivariate Markov chain.

Keywords: nonequilibrium thermodynamics; landscape-flux decomposition; mutual information rate; entropy production rate

---

1. Introduction

There are growing interests in studying the information systems in the fields of control theory, information theory, communication theory, and biophysics [1–6]. Significant progresses have been made recently towards the understanding of the information system in terms of information thermodynamics [10–13]. However, the identification of the global driving force for the information system dynamics is still challenging. Here we would like to fill the gap by quantifying the driving forces for the information system dynamics. Inspired by the recent development of landscape and flux theory for the non-equilibrium systems [14–16], we will show that the driving force for information dynamics is determined by both the information landscape and information flux. The information flux is a measure of the degree of nonequilibirumness or time irreversibility. Mutual information represents the correlation between two information subsystems. We uncovered that the mutual information rate between the two subsystems can be decomposed into the time-reversible and time-irreversible parts respectively. This is originated from the information landscape-flux decomposition for dynamics. An important signature of nonequilibriumness is the entropy production or energy cost. We also uncover the intimate relation between the entropy production rates and the time-irreversible part of the mutual information rate. We demonstrate the above features by the dynamics of a bivariate Markov chain.

2. Bivariate Markov Chains

Markov chains have been often assumed for the underlying information dynamics of the total system in random environments. That is, the two subsystems together forms a Markov chain in continuous or discrete times, which is the so-called *Bivariate Markov Chain*(BMC). The processes of the two subsystems are correspondingly said to be marginal processes or marginal chain. The BMC was used to model ion channel currents [2]. It was also used to model delays and congestion in a

 doi:10.20944/preprints201710.0067.v1

computer network [3]. Recently, different models of BMC appeared in non-equilibrium statistical physics for capturing or implementing the Maxwell's demon [4–6], which can be seen as one marginal chain in the BMC playing feedback control to the other marginal chain. Although the BMC has been studied for decades, there are still challenges on quantifying the dynamics of the whole as well as the two subsystems. This is because neither of them needs to be Markovian chain in general [7], and the quantifications of the probabilities (densities) for the trajectories of the two subsystems involve complex random matrices manipulations [8]. This leads to the problem not exactly analytically solvable. The corresponding numerical solutions often lack direct mathematical and physical interpretations.

The conventional analysis of the BMC focuses on the mutual information [9] of the two subsystems for quantifying the underlying information correlations. There are three main representations on this. The first one was proposed by Sagawa [10,11] for explaining the mechanism of Maxwell's demon in Szilard's engine. In this representation, the mutual information between the demon and controlled system characterizes the observation and the feedback of the demon. This leads to an elegant way which includes the increment of the mutual information into a unified fluctuation relation. The second representation was proposed by Esposito [12] in an attempt to explain the violation of the second law in a specified BMC, the bipartite model, where the mutual information is divided into two parts corresponding to the two subsystems respectively, which were said to be the information flows. This representation tries to explain the mechanism of the demon because one can see that the information flows do contribute to the entropy production to both demon and controlled system. The first two representations are based on the ensembles of the subsystem states. This means that the mutual information is defined only on the time-sliced distributions of the system states, which somehow lacks the information of subsystem dynamics: the time-correlations of the observation and feedback of the demon. The last representation was seen in the work of Seifert [13] where he used a more general definition of mutual information in information theory, which is defined on the trajectories of the two subsystem. More exactly, this is the so-called *Mutual Information Rate* (MIR) which quantifies the correlation between the two subsystem dynamics. However, due to the difficulties from the possible underlying non-Markovian property of the marginal chains, exactly solvable models and comprehensive conclusions are still challenging from this representation.

In this study, we study the discrete-time BMC in both stochastic dynamics. To avoid the technical difficulty caused by non-Markovian dynamics, we first assume that the two marginal chains follow the Markovian dynamics. The non-Markovian case will be discussed elsewhere. We explore the time-irreversibility of BMC and marginal processes in steady state. Then we decompose driving force for the underlying information dynamics as the information landscape and information flux [14–16] representing the time-reversible parts and time-irreversible parts respectively. We also prove that the non-vanishing flux fully describes the time-irreversibility of BMC and marginal processes.

We focus on the mutual information rate between the two marginal chains in information dynamics. Since the two marginal chains are assumed to be Markov chains here, the mutual information rate is exactly analytically solvable, which can be seen as the averaged conditional correlation between the two subsystem states. Here the conditional correlations reveal the time correlations between the past states and the future states.

Corresponding to the landscape-flux decomposition in stochastic dynamics, we decompose the MIR into two parts: the time-reversible and time-irreversible parts respectively. The time-reversible part measures the part of the correlations between the two marginal chains in both forward and backward processes of BMC. The time-irreversible part measures the difference between the correlations in forward and backward processes of BMC respectively. We can see that a non-vanishing time-irreversible part of the MIR must be driven by a non-vanishing flux in steady state, and can be seen as the sufficient condition for a BMC to be time-irreversible.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2017 doi:10.20944/preprints201710.0067.v1

We also reveal the important fact that the time-irreversible parts of MIR contributes to the nonequilibrium *Entropy Production Rate* (EPR) of the BMC by the simple equality:

$$\textit{EPR of BMC} = \textit{EPR of 1st marginal chain} + \textit{EPR of 2nd marginal chain} + 2 \times \textit{time-irreversible part of MIR}.$$

And this relation may help to develop general theory on nonequilibrium interacting information system dynamics.

3. Information Landscape and Information Flux for Determining the Information Dynamics, Time-Irreversibility

Consider a finite-state, discrete-time, ergodic, and irreducible bivariate Markov chain

$$Z = (X, S) = \{(X(t), S(t)), t \geq 1\}. \tag{1}$$

We assume that the state space of X is given by $\mathcal{X} = \{1, ..., d\}$ and the state space of S is given by $\mathcal{S} = \{1, ..., l\}$. The state space of Z is then given by $\mathcal{Z} = \mathcal{X} \times \mathcal{S}$. The time evolution of distribution of Z is characterized by the following master equation in discrete time,

$$p_z(z; t+1) = \sum_{z'} q_z(z|z') p_z(z'; t), \text{ for } t \geq 1, \text{ and } z \in \mathcal{Z} \tag{2}$$

where $p_z(z; t) = p_z(x, s; t)$ is the probability of observing state z (or joint probability of $X = x$ and $S = s$) at time t; $q_z(z|z') = q_z(x, s|x', s') \geq 0$ are the transition probabilities from $z' = (x', s')$ to $z = (x, s)$ respectively and are with $\sum_z q_z(z|z') = 1$.

We assume that there exists a unique stationary distribution π_z such that $\pi_z(z) = \sum_{z'} q_z(z|z')\pi_z(z')$. Then given arbitrary initial distribution, the distribution goes to π_z exponentially fast in time. If the initial distribution is π_z, we say that Z is in *Steady State* (SS) and our discussion is based on this SS.

The marginal chains of Z, i.e., X and S, do not need to be Markov chains in general. For simplicity of analysis, we assume that both marginal chains are Markov chains and the corresponding transition probabilities are given by $q_x(x|x')$ and $q_s(s|s')$ (for $x, x' \in \mathcal{X}$ and $s, s' \in \mathcal{S}$) respectively. Then we have the following master equations for X and S,

$$p_x(x; t+1) = \sum_{x'} q_x(x|x') p_x(x'; t), \tag{3}$$

and

$$p_s(s; t+1) = \sum_{s'} q_s(s|s') p_s(s'; t), \tag{4}$$

where $p_x(x; t)$ and $p_s(s; t)$ are the probabilities of observing $X = x$ and $S = s$ at time t respectively.

We consider that both Eqs.(3,4) have unique stationary solutions π_x and π_s which satisfy $\pi_x(x) = \sum_{x'} q_x(x|x')\pi_x(x')$ and $\pi_s(s) = \sum_{s'} q_s(s|s')\pi_s(s')$ respectively. Also, we assume that when Z is in SS, π_x and π_s are also achieved. The relations between π_x, π_s and π_z read,

$$\begin{cases} \pi_x(x) = \sum_s \pi_z(x, s), \\ \pi_s(s) = \sum_x \pi_z(x, s). \end{cases} \tag{5}$$

In the rest of this paper, we let $X^T = \{X(1), X(2), ..., X(T)\}$, $S^T = \{S(1), S(2), ..., S(T)\}$, and $Z^T = \{Z(1), Z(2), ..., Z(T)\} = (X^T, S^T)$ denote the time sequences of X, S, and Z in time T respectively.

To characterize the time-irreversibility of the Markov chain C in stochastic dynamics in SS, we introduce the concept of probability flux. Here we let C denote arbitrary Markov chain in $\{Z, X, S\}$,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2017 doi:10.20944/preprints201710.0067.v1

and let c, π_c, q_c, and C^T denote arbitrary state of C, the stationary distribution of C, the transition probabilities of C, and a time sequence of C in time T and in SS, respectively.

The averaged number transitions from the state c' to state c, denoted by $N(c' \to c)$, in unit time in SS can be obtained as

$$N(c' \to c) = \pi_c(c')q_c(c|c').$$

This is also the probability of the time sequence $C^T = \{C(1) = c', C(2) = c\}, (T = 2)$. Correspondingly, the averaged number of reverse transitions, denoted by $N(c \to c')$, reads

$$N(c \to c') = \pi_c(c)q_c(c'|c).$$

This is also the the probability of the time-reverse sequence $\widetilde{C}^T = \{C(1) = c, C(2) = c'\}, (T = 2)$. The difference between these two transition numbers measures the time-reversibility of the forward sequence C^T in SS,

$$\begin{aligned} J_c(c' \to c) &= N(c' \to c) - N(c \to c') \\ &= P(C^T) - P(\widetilde{C}^T) \\ &= \pi_c(c')q_c(c|c') - \pi_c(c)q_c(c'|c), \text{ for } C = X, S, \text{ or } Z. \end{aligned} \tag{6}$$

Then, $J_c(c' \to c)$ is said to be the probability flux from c' to c in SS. If $J_c(c' \to c) = 0$ for arbitrary c' and c, then C^T $(T = 2)$ is time-reversible; otherwise when $J_c(c' \to c) \neq 0$, C^T is time-irreversible. Clearly, we have from Eq. (6) that

$$J_c(c' \to c) = -J_c(c \to c'). \tag{7}$$

The transition probability determines the evolution dynamics of the information system. We can decompose the transition probabilities $q_c(c|c')$ into two parts: the time-reversible part D_c and time-irreversible part B_c, which read

$$q_c(c|c') = D_c(c' \to c) + B_c(c' \to c), \text{ with} \tag{8}$$
$$\begin{cases} D_c(c' \to c) = \frac{1}{2\pi_c(c')}(\pi_c(c')q_c(c|c') + \pi_c(c)q_c(c'|c)), \\ B_c(c' \to c) = \frac{1}{2\pi_c(c')}J_c(c' \to c). \end{cases}$$

From this decomposition, we can see that the information dynamics is determined by two driving forces. One of the driving force is determined by the steady state probability distribution and is time reversible. The other driving force for the information system dynamics is the steady state probability flux which breaks the detailed balance and quantify the time irreversibility. Since the steady state probability measures the weight of the information state, therefore it quantifies the information landscape. If we define the potential landscape for the information system as $\phi = -\log \pi$, then the $D_c(c' \to c) = \frac{1}{2}(q_c(c|c') + \frac{\pi_c(c)}{\pi_c(c')}q_c(c'|c)) = \frac{1}{2}(q_c(c|c') + \exp[-(\phi_c(c) - \phi_c(c')]q_c(c'|c))$ becomes the difference or "gradient" in the potential landscape. Therefore, this reversible part of the information dynamics is determined by the "gradient" of the information landscape. The steady state probability flux measures the information flow in the dynamics and therefore can be termed as the information flux. It is a direct measure of the nonequilibriumness in terms of time irreversibility.

By Eqs.(7,8), we have the following relations

$$\begin{cases} \pi_c(c')D_c(c' \to c) = \pi_c(c)D_c(c \to c'), \\ \pi_c(c')B_c(c' \to c) = -\pi_c(c)B_c(c \to c'). \end{cases} \tag{9}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2017 doi:10.20944/preprints201710.0067.v1

As we can see in next section, D_c and B_c are useful for us to quantify time-reversible and time-irreversible observables of C respectively.

We give the interpretation that the non-vanishing probability flux J_c fully measures the time-irreversibility of the chain C in time T for $T \geq 2$. Let C^T be arbitrary sequence of C in SS, and with no loss of generality we let $T = 3$. Similar to Eq. (6), the measure of time-irreversibility of C^T can be given by the difference between the probability of $C^T = \{C(1), C(2), C(3)\}$ and that of its time-reversal $\widetilde{C}^T = \{C(3), C(2), C(1)\}$, such as

$$\begin{aligned}
&P(C^T) - P(\widetilde{C}^T) \\
&= \pi_c(C(1))q_c(C(2)|C(1))q_c(C(3)|C(2)) - \pi_c(C(3))q_c(C(2)|C(3))q_c(C(1)|C(2)) \\
&= \pi_c(C(1))\left(D_c(C(1) \to C(2)) + B_c(C(1) \to C(2))\right)\left(D_c(C(2) \to C(3)) + B_c(C(2) \to C(3))\right) - \\
&\pi_c(C(3))\left(D_c(C(3) \to C(2)) + B_c(C(3) \to C(2))\right)\left(D_c(C(2) \to C(1)) + B_c(C(2) \to C(1))\right), \\
&\text{for } C = X, S \text{ or } Z.
\end{aligned}$$

Then by the relations given in Eq.(9), we have $P(C^T) - P(\widetilde{C}^T) = 0$ holds for arbitrary C^T if and only if $B_c(C(1) \to C(2)) = B_c(C(2) \to C(3)) = 0$ or equivalently $J_c(C(1) \to C(2)) = J_c(C(2) \to C(3)) = 0$. This conclusion can be made for arbitrary $T > 3$. Thus, non-vanishing J_c can fully describe the time-irreversibility of C for $C = X, S$, or Z.

We show the relations between the fluxes of the whole system J_z and of the subsystem J_x as following:

$$\begin{aligned}
J_x(x' \to x) &= \pi_x(x')q_x(x|x') - \pi_x(x)q_x(x'|x) \\
&= P(\{x', x\}) - P(\{x, x'\}) \\
&= \sum_{s,s'} \left(P(\{(x', s'), (x, s)\}) - P(\{(x, s), (x', s')\})\right) \\
&= \sum_{s,s'} \left(\pi_z(x', s')q_z(x, s|x', s') - \pi_z(x, s)q_z(x', s'|x, s)\right) \\
&= \sum_{s,s'} J_z((x', s') \to (x, s)). \qquad (10)
\end{aligned}$$

Similarly, we have

$$J_s(s' \to s) = \sum_{x,x'} J_z((x', s') \to (x, s)). \qquad (11)$$

These relations indicate that the subsystem fluxes J_x and J_s can be seen as the coarse-grained levels of total system flux J_z by averaging over the other part of the system S and X respectively. We should emphasize that, Non-vanishing J_z does not mean X or S is time-irreversible and vice versa.

4. Mutual Information Decomposition to Time-Reversible and Time-Irreversible Parts

According to the information theory, the two interacting information systems represented by bivariate Markov chain Z can be characterized by the *Mutual Information Rate* (MIR) between the marginal chains X and S in SS. The mutual information rates represents correlation between two interacting infomration systems. The MIR is defined on the probabilities of all possible time sequences, $P(Z^T)$, $P(X^T)$, and $P(S^T)$, and is given by

$$I(X, S) = \lim_{T \to \infty} \frac{1}{n} \sum_{Z^T} P(Z^T) \log \frac{P(Z^T)}{P(X^T)P(S^T)}. \qquad (12)$$

It measures the correlation between X and S in unit time, or say, the efficient bits of information that X and S exchange with each other in unit time. The MIR must be non-negative, and a vanishing $I(X, S)$

indicates that X and S are independent of each other. More explicitly, the corresponding probabilities of these sequences can be evaluated by using Eqs.(2,3,4), we have

$$\begin{cases} P(X^T) = \pi_x(X(1)) \prod_{t=1}^{T-1} q_x(X(t+1)|X(t)), \\ P(S^T) = \pi_s(S(1)) \prod_{t=1}^{T-1} q_s(S(t+1)|S(t)), \\ P(Z^T) = \pi_z(Z(1)) \prod_{t=1}^{T-1} q_z(Z(t+1)|Z(t)). \end{cases}$$

By substituting these probabilities into Eq.(12) (see Appendix), we have the exact expression of MIR as

$$\begin{aligned} I(X,S) &= \sum_{z,z'} \pi_z(z') q_z(z|z') \log \frac{q_z(z|z')}{q_x(x|x')q_s(s|s')} \\ &= \langle i(z|z') \rangle_{z',z} \geq 0, \text{ for } z = (x,s), \text{ and } z' = (x',s'). \end{aligned} \tag{13}$$

where $i(z|z') = \log \frac{q_z(z|z')}{q_x(x|x')q_s(s|s')}$ is the conditional (Markovian) correlation between the states x and s when the transition $z' = (x',s') \to z = (x,s)$ occurs. This indicates that when the two marginal processes are both Markovian, the MIR is the average of the conditional (Markovian) correlations. These correlations are measurable when transitions occur and can be seen from the observables of Z.

By noting the decomposition of transition probabilities in Eq. (8), we have a corresponding decomposition of $I(X,S)$ such as

$$I(X,S) = I_D(X,S) + I_B(X,S), \text{ with} \tag{14}$$
$$\begin{cases} I_D(X,S) = \sum_{z,z'} \pi_z(z') D_z(z|z') i(z|z') = \frac{1}{2} \sum_{z,z'} (\pi_z(z') q_z(z|z') + \pi_z(z) q_z(z'|z)) i(z|z'), \\ I_B(X,S) = \sum_{z,z'} \pi_z(z') B_z(z|z') i(z|z') = \frac{1}{2} \sum_{z,z'} J_z(z|z') i(z|z') = \frac{1}{4} \sum_{z,z'} J_z(z|z') (i(z|z') - i(z'|z)). \end{cases}$$

This means that the mutual information representing the correlations between the two interacting systems can be decomposed into time reversible equilibrium part and time irreversible nonequilibrium part. The origin of this is from the fact the underlying information dynamics is determined by both the time reversible information landscape and time irreversible information flux. These equations are very important to establish the link to the time-irreversibility. We now give further interpretation for $I_D(X,S)$ and $I_B(X,S)$:

Consider a bivariate Markov chain Z in SS wherein X and S are dependent of each other, i.e., $I(X,S) = I_D(X,S) + I_B(X,S) > 0$. By the ergodicity of Z, we have the MIR which measures the averaged conditional correlation along the time sequences Z^T,

$$\lim_{T\to\infty} \frac{1}{T} \langle i(Z(t+1)|Z(t)) \rangle_{Z^T} = I(X,S), \text{ for } 1 < t < T.$$

Then $I_B(X,S)$ measures the change of averaged conditional correlation between X and S when a sequence of Z turns back in time,

$$\lim_{T\to\infty} \frac{1}{T} \langle i(Z(t+1)|Z(t)) - i(Z(t)|Z(t+1)) \rangle_{Z^T} = 2I_B(X,S).$$

A negative $I_B(X,S)$ shows that the correlation between X and S becomes strong in the time-reversal process of Z; A positive $I_B(X,S)$ shows that the correlation becomes weak in the time-reversal process of Z. Both two cases show that the Z is time-irreversible since we have a non-vanishing J_z. But the case of $I_B(X,S) = 0$ is complicated, since it indicates either a vanishing J_z or a non-vanishing J_z. Anyway, we see that a non-vanishing $I_B(X,S)$ is a sufficient condition for Z to be time-irreversible. On the other hand, $I_D(X,S) = I(X,S) - I_B(X,S)$ measures the correlation remaining in the backward process of Z.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2017 doi:10.20944/preprints201710.0067.v1

5. Relationship Between Mutual Information and Entropy Production

The *Entropy Production Rates* (EPR) or energy dissipation (cost) rate at steady state is a quantitative nonequilibriumness measure which characterizes the time-irreversibility of the underlying processes. The EPRs of the information system described by the bivariate Markov chains here can be given by

$$\begin{cases} R_z = \frac{1}{2}\sum_{z,z'} J_z(z' \to z)\log\frac{q_z(z|z')}{q_z(z'|z)} \geq 0, \\ R_x = \frac{1}{2}\sum_{x,x'} J_x(x' \to x)\log\frac{q_x(x|x')}{q_x(x'|x)} \geq 0, \\ R_s = \frac{1}{2}\sum_{s,s'} J_s(s' \to s)\log\frac{q_s(s|s')}{q_s(s'|s)} \geq 0, \end{cases} \tag{15}$$

where total and subsystem entropy productions R_z, R_x, and R_s correspond to Z, X, and S respectively. Here, R_z usually contains the detailed interaction information of the system (or subsystems) and environments; R_x and R_s provide the coarse-grained information of time-irreversible observables of X and Z respectively. Each non-vanishing EPR indicates that the corresponding Markov chain is time-irreversible. Again, we emphasize that a non-vanishing R_z does not mean X or S is time-irreversible and vice versa.

We are interested in the connection between these EPRs and mutual information. We can associate them with $I_B(X,S)$ by noting Eqs.(10,11,14). We have

$$\begin{aligned} I_B(X,S) &= \frac{1}{4}\sum_{z,z'} J_z(z|z')(i(z|z') - i(z'|z)) \\ &= \frac{1}{4}\sum_{z,z'} J_z(z|z')\log\frac{q_z(z|z')}{q_z(z'|z)} - \frac{1}{4}\sum_{x,x'} J_x(x|x')\log\frac{q_x(x|x')}{q_x(x'|x)} - \frac{1}{4}\sum_{s,s'} J_s(s|s')\log\frac{q_s(s|s')}{q_s(s'|s)} \\ &= \frac{1}{2}(R_z - R_x - R_s). \end{aligned} \tag{16}$$

We note that $I_B(X,S)$ intimated related to the EPRs. This builds up a bridge between these EPRs and irreversible part of the mutual information. Moreover, we also have

$$\begin{cases} R_z = R_x + R_s + 2I_B(X,S) \geq 0, \\ R_x + R_s \geq -2I_B(X,S), \\ R_z \geq 2I_B(X,S). \end{cases} \tag{17}$$

This indicates that the time-irreversible MIR contributes to the detailed EPR. In other words, The differences of entropy production rate of the whole system and subsystems provides the origin of the time irreversible part of the mutual information. This gives the nonequilibrium thermodynamic origin of the irreversible mutual information or correlations. Of course, since the EPR is related to the flux directly as seem from above definitions, the origin of the EPR or nonequilibrium thermodynamics is from the non-vanishing information flux for the nonequilibrium dynamics. On the other hand, irreversible part of the mutual information measures the correlations and contributes to the correlated part of the EPR between the subsystems.

6. A Simple Case: Blind Demon

As a concrete example, we consider a two-state system coupled to two information baths a and b. The states of the system are denoted by $\mathcal{X} = \{x : x = 0, 1\}$ respectively. Each bath sends an instruction to the system. If the system adopts one of them, it then follows the instruction and makes change of the state. The instructions generated from one bath are independently and identically distributed (Bernoulli trials). Both the probability distributions of the instructions corresponding to the baths follow Bernoulli distributions and read $\{\epsilon_a(x) : x \in \mathcal{X}, \epsilon_a(x) \geq 0, \sum_x \epsilon_a(x) = 1\}$ for bath a and $\{\epsilon_b(x) : x \in \mathcal{X}, \epsilon_b(x) \geq 0, \sum_x \epsilon_b(x) = 1\}$ for bath b respectively. Since the system cannot execute two instructions simultaneously, there exists an information demon that makes choices for the system. The demon is blind to care about the system and it makes choices independently and identically distributed. The choices of the demon are denoted by $\mathcal{S} = \{s : s = a, b\}$ respectively. The probability distribution of demon's choices reads $\{P(s) : s \in \mathcal{S}, P(a) = p, P(b) = 1 - p, p \in [0,1]\}$. Still, we use $Z = (X, S)$ with $X \in \mathcal{X}$ and $S \in \mathcal{S}$ to denote the BMC of the system and the demon.

The transition probabilities of the system read

$$q_x(x|x') = p\epsilon_a(x) + (1-p)\epsilon_b(x).$$

The transition probabilities of the demon read

$$q_s(s|s') = P(s).$$

And the transition probabilities of the joint chain read

$$q_z(x, s|x', s') = P(s)\epsilon_{s'}(x).$$

We have the corresponding steady state distributions or the information landscape as,

$$\begin{cases} \pi_x(x) = p\epsilon_a(x) + (1-p)\epsilon_b(x), \\ \pi_s(s) = P(s), \\ \pi_z(x,s) = P(s)\pi_x(x). \end{cases}$$

We obtain the information fluxes as,

$$\begin{cases} J_x(x' \to x) = 0, \text{ for all } x, x' \in \mathcal{X} \\ J_s(s' \to s) = 0, \text{ for all } s, s' \in \mathcal{S} \\ J_z((x', s') \to (x, s)) = P(s)P(s')(\pi_x(x')\epsilon_{s'}(x) - \pi_x(x)\epsilon_s(x')). \end{cases}$$

Here, we use the notations $\epsilon_s(x')$ and $\epsilon_{s'}(x)$ ($s, s' = a$ or b) to denote the probabilities of the instructions x' or x from bath a or b briefly. We obtain the EPRs as

$$\begin{cases} R_x = 0, \\ R_s = 0, \\ R_z = \sum_x p(1-p)(\epsilon_a(x) - \epsilon_b(x))(\log\epsilon_a(x) - \log\epsilon_b(x)). \end{cases}$$

We evaluate the MIR as

$$I(X,S) = -\sum_x \pi_x(x)\log\pi_x(x) + p\sum_x \epsilon_a(x)\log\epsilon_a(x) + (1-p)\sum_x \epsilon_b(x)\log\epsilon_b(x).$$

The time-irreversible part of $I(X,S)$ reads,

$$I_B(X,S) = \frac{1}{2}R_z.$$

7. Conclusion

In this work, we identify the driving forces for the information system dynamics. We show that the information system dynamics is determined by both the information landscape and information flux representing the time reversible and time irreversible part of the information dynamics. We further demonstrate that the mutual information representing the correlations can be decomposed into time reversible part and time irreversible part originated from the landscape and flux decomposition of the information dynamics. Finally we uncover the intimate relationship between the difference of the entropy production of the whole system and the subsystems and the time irreversible part of the mutual information. This will help for understanding the non-equilibrium behaviour of the interacting information system dynamics in random environments. Furthermore, we believe that our conclusion can be made more general for the BMC with non-Markovian marginal chains which we will discuss in a separate work.

Acknowledgments: This work was support in part by National Natural Science Foundation of China (NSFC-91430217) and National Science Foundation[US] (NSF-PHY-76066).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BMC	Bivariate Markov Chain
EPR	Entropy Production Rate
MIR	Mutual Information Rate
SS	Steady State

Appendix

Here, we derive the exact form of Mutual Information Rate (MIR, Eq.(13)) in steady state by using the cumulant-generating function.

We write arbitrary time sequence of Z in time T in the form as following

$$Z^T = \{Z(1), ..., Z(i), ..., Z(T)\}, \text{ for } T \geq 2,$$

where $Z(i)$ (for $i \geq 1$) denotes the state at time i. The corresponding probability of Z^T is in the following form

$$P(Z^T) = \pi_z(Z_1) \left\{ \prod_{i=1}^{T-1} q_z(Z_{i+1}|Z_i) \right\}. \tag{A.1}$$

We let the chain $U = (X, S)$ to denote a process that X and S follow the same Markov dynamics in Z but are independent of each other. Then we have the transition probabilities of U read

$$q_u(u|u') = q(x, s|x', s') = q_x(x|x')q_s(s|s'). \tag{A.2}$$

Then the probability of a time sequence of U, U^T, with the same trajectory of Z^T reads

$$P(U^T) = \pi_u(Z_1) \left\{ \prod_{i=1}^{T-1} q_u(Z_{i+1}|Z_i) \right\}, \tag{A.3}$$

with $\pi_u(x, s) = \pi_x(x)\pi_s(s)$ being the stationary probability of U.

For evaluating the exact form of MIR, we introduce the cumulant-generating function of the random variable $\log \frac{P(Z^T)}{P(U^T)}$,

$$K(m, T) = \log \left\langle \exp \left(m \log \frac{P(Z^T)}{P(U^T)} \right) \right\rangle_{Z^T}. \tag{A.4}$$

 doi:10.20944/preprints201710.0067.v1

We can see that

$$\begin{aligned}&\lim_{T\to\infty}\lim_{m\to 0}\frac{1}{T}\frac{\partial K(m,T)}{\partial m}\\&=\lim_{T\to\infty}\frac{1}{T}\left\langle \log\frac{P(Z^T)}{P(U^T)}\right\rangle_{Z^T}\\&=I(X,S).\end{aligned}\tag{A.5}$$

Thus, our idea is to evaluate $K(m,T)$ at first. We have

$$\begin{aligned}K(m,T)&=\log\left\langle \exp\left(m\log\frac{P(Z^T)}{P(U^T)}\right)\right\rangle_{Z^T}\\&=\log\left\{\sum_{Z^T}\frac{(P(Z^T))^{m+1}}{(P(U^T))^m}\right\}\\&=\log\left\{\sum_{\{Z(0),Z(1),\ldots,Z(T)\}}\frac{(\pi_z^{m+1}(Z_0))}{(\pi_u^m(Z_0))}\prod_{i=0}^{T-1}\frac{q_z^{m+1}(Z_{i+1}|Z_i)}{q_u^m(Z_{i+1}|Z_i)}\right\},\end{aligned}\tag{A.6}$$

256 where we realize that the last equality can be rewritten in the form of matrices multiplication.

We introduce the following matrices and vectors for Eq. (A.6) such that

$$\begin{aligned}\boldsymbol{Q}_z&=\left\{(\boldsymbol{Q}_{\boldsymbol{z}})_{(z,z')}=q_z(z|z'),\ \text{for } z,z'\in\mathcal{Z}\right\},\\\boldsymbol{G}(m)&=\left\{(\boldsymbol{G}(m))_{(z,z')}=\frac{q_z^{m+1}(z|z')}{q_u^m(z|z')},\ \text{for } z,z'\in\mathcal{Z}\right\},\\\boldsymbol{\pi}_z&=\{(\boldsymbol{\pi}_z)_z=\pi_z(z),\ \text{for } z\in\mathcal{Z}\},\\\boldsymbol{v}(m)&=\left\{(\boldsymbol{v}(m))_z=\frac{\pi_z^{m+1}(z)}{\pi_u^m(z)}\right\},\end{aligned}\tag{A.7}$$

where $\boldsymbol{Q}_z$ is the transition matrix of Z; $\boldsymbol{\pi}_z$ is the stationary distribution of Z. It can be also verified that

$$\begin{aligned}&\boldsymbol{Q}_z=\boldsymbol{G}(0),\\&\boldsymbol{\pi}_z=\boldsymbol{v}(0),\\&\boldsymbol{\pi}_z=\boldsymbol{Q}_z\boldsymbol{\pi}_z,\\&\boldsymbol{1}^\dagger\boldsymbol{Q}_z=\boldsymbol{1}^\dagger,\\&\lim_{m\to 0}\frac{d\boldsymbol{G}(m)}{dm}=\left\{\left(\lim_{m\to 0}\frac{d\boldsymbol{G}(m)}{dm}\right)_{(z,z')}=q_z(z|z')\log\frac{q_z(z|z')}{q_u(z|z')},\ \text{for } z,z'\in\mathcal{Z}\right\},\\&\lim_{m\to 0}\frac{d\boldsymbol{v}(m)}{dm}=\left\{\left(\lim_{m\to 0}\frac{d\boldsymbol{v}(m)}{dm}\right)_z=\pi_z(z)\log\frac{\pi_z(z)}{\pi_u(z)},\ \text{for } z\in\mathcal{Z}\right\},\end{aligned}\tag{A.8}$$

257 where $\boldsymbol{1}^\dagger$ is the vector of all 1's with appropriate dimension.

Then $K(m,T)$ can be rewritten in a compact form such that

$$K(m,T)=\log\left\{\boldsymbol{1}^\dagger\boldsymbol{G}^{T-1}(m)\boldsymbol{v}(m)\right\}.\tag{A.9}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2017 doi:10.20944/preprints201710.0067.v1

Then, we substitute Eq. (A.9) into Eq. (A.5) and have

$$\begin{aligned} I(X,S) &= \lim_{T\to\infty}\lim_{m\to 0}\frac{1}{T}\frac{\partial K(m,T)}{\partial m} \\ &= \lim_{T\to\infty}\lim_{m\to 0}\frac{1}{T}\frac{\partial \log\left\{\mathbf{1}^\dagger \boldsymbol{G}^{T-1}(m)\boldsymbol{v}(m)\right\}}{\partial m} \\ &= \lim_{T\to\infty}\lim_{m\to 0}\frac{1}{T}\left\{(T-1)\mathbf{1}^\dagger \boldsymbol{G}^{T-2}(m)\frac{d\boldsymbol{G}(m)}{dm}\boldsymbol{v}(m) + \mathbf{1}^\dagger \boldsymbol{G}^{T-1}(m)\frac{d\boldsymbol{v}(m)}{dm}\right\} \\ &= \lim_{T\to\infty}\frac{1}{T}\left\{(T-1)\mathbf{1}^\dagger \boldsymbol{G}^{T-2}(0)\left(\lim_{m\to 0}\frac{d\boldsymbol{G}(m)}{dm}\right)\boldsymbol{v}(0) + \mathbf{1}^\dagger \boldsymbol{G}^{T-1}(0)\left(\lim_{m\to 0}\frac{d\boldsymbol{v}(m)}{dm}\right)\right\}. \end{aligned} \tag{A.10}$$

By noting Eq. (A.8) and $T \geq 2$, we obtain Eq. (13) from Eq. (A.10) that

$$\begin{aligned} I(X,S) &= \lim_{T\to\infty}\frac{1}{T}\left\{(T-1)\mathbf{1}^\dagger \boldsymbol{G}^{T-2}(0)\left(\lim_{m\to 0}\frac{d\boldsymbol{G}(m)}{dm}\right)\boldsymbol{v}(0) + \mathbf{1}^\dagger \boldsymbol{G}^{T-1}(0)\left(\lim_{m\to 0}\frac{d\boldsymbol{v}(m)}{dm}\right)\right\} \\ &= \lim_{T\to\infty}\left\{\left(1-\frac{1}{T}\right)\mathbf{1}^\dagger\left(\lim_{m\to 0}\frac{d\boldsymbol{G}(m)}{dm}\right)\boldsymbol{\pi}_z + \frac{1}{T}\mathbf{1}^\dagger\left(\lim_{m\to 0}\frac{d\boldsymbol{v}(m)}{dm}\right)\right\} \\ &= \mathbf{1}^\dagger\left(\lim_{m\to 0}\frac{d\boldsymbol{G}(m)}{dm}\right)\boldsymbol{\pi}_z \\ &= \sum_{(x,s),(x',s')}\pi_z(x',s')q_z(x,s|x',s')\log\frac{q_z(x,s|x',s')}{q_x(x|x')q_s(s|s')}. \end{aligned} \tag{A.11}$$

References

1. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* 1948, 27, 379–423, doi:10.1109/9780470544242.ch1.
2. Ball, F.; Yeo, G. F. Lumpability and Marginalisability for Continuous-Time Markov Chains. *J. Appl. Probab.* 1993, 30, 518-528, doi:10.2307/3214762.
3. Wei, W.; Wang, B.; Towsley, D. Continuous-time hidden Markov models for network performance evaluation. Performance Evaluation 2002, 49, 129-146, doi: 10.1016/s0166-5316(02)00122-0.
4. Strasberg, P.; Schaller, G.; Brandes, T.; etc. Thermodynamics of a physical model implementing a Maxwell demon. *Phys. Rev. Lett.* 2013, 110, 040601:1-040601:5, doi:10.1103/physrevlett.110.040601.
5. Koski, J. V.; Kutvonen, A.; Khaymovich, I. M.; etc. On-Chip Maxwell's Demon as an Information-Powered Refrigerator. *Phys. Rev. Lett.* 2015, 115, 260602:1- 260602:5, doi:10.1103/PhysRevLett.115.260602.
6. Mcgrath, T.; Jones, N. S.; Ten Wolde, P. R.; etc. Biochemical Machines for the Interconversion of Mutual Information and Work. *Phys. Rev. Lett.* 2017, 118, 028101:1-028101:5, doi:10.1103/PhysRevLett.118.028101.
7. Mark, B. L.; Ephraim, Y. An EM algorithm for continuous-time bivariate Markov chains. *Comput. Stat. Data. Anal.* 2013, 57, 504-517, doi:10.1016/j.csda.2012.07.017.
8. Ephraim, Y.; Mark, B. L. Bivariate Markov Processes and Their Estimation. *Foundations and Trends in Signal Processing* 2012, 6, 1-95, doi:10.1561/2000000043.
9. Cover, T. M.; Thomas, J. A. *Elements of information theory*, 2nd edn. John Wiley & Sons: Hoboken, New Jersey, USA, 2006; ISBN: 13 978-0-471-24195-9.
10. Parrondo, J. M. R.; Horowitz, J. M.; Sagawa, T. Thermodynamics of information. *Nature Phys.* 2015, 11, 131-139, doi:10.1038/nphys3230.
11. Sagawa, T.; Ueda, M. Fluctuation theorem with information exchange: role of correlations in stochastic thermodynamics. *Phys. Rev. Lett.* 2012, 109, 180602:1-180602:5, doi:10.1103/PhysRevLett.109.180602.
12. Horowitz, J. M.; Esposito, M. Thermodynamics with Continuous Information Flow. *Phys. Rev. X.* 2014, 4, 031015:1-031015:11, doi:10.1103/physrevx.4.031015.
13. Barato, A. C.,; Hartich, D.; Seifert, U. Rate of Mutual Information Between Coarse-Grained Non-Markovian Variables. *J. Stat. Phys.* 2013, 153, 460–478, doi:10.1007/s10955-013-0834-5.

14. Wang, J.; Xu, L.; Wang, E. K. Potential landscape and flux framework of nonequilibrium networks: robustness, dissipation, and coherence of biochemical oscillations. *Proc. Natl. Acad. Sci. USA* 2008, 105, 12271-12276, doi:10.1073/pnas.0800579105.
15. Wang, J. Landscape and flux theory of non-equilibrium dynamical systems with application to biology. *Adv. Phys.* 2015, 64, 1-137, doi:10.1080/00018732.2015.1037068.
16. Li, C. H.; Wang, E. K.; Wang, J. Potential flux landscapes determine the global stability of a Lorenz chaotic attractor under intrinsic fluctuations. *J. Chem. Phys.* 2012, 136, 194108:1-194108:13, doi:10.1063/1.4716466.