Interactions of information transfer along separable causal paths

Peishi Jiang and Praveen Kumar

Interactions of Information Transfer Along Separable Causal Paths

Peishi Jiang and Praveen Kumar*

Ven Te Chow Hydrosystem Laboratory, Civil and Environmental Engineering, University of Illinois, Urbana, IL 61801, USA

Complex systems arise as a result of inter-dependencies between multiple variables, whose causal interactions can be visualized in a time series graph. Transfer entropy and information partitioning approaches have been used to characterize such dependencies. However, these approaches capture net information transfer occurring through a multitude of pathways involved in the interaction, and as a result mask our ability to discern the causal interaction within a subgraph of interest through specific pathways. We build on recent developments of momentary information transfer along causal paths proposed by Runge [Phys.Rev.E. 92, 062829 (2015)] to develop a framework for quantifying information partitioning along separable causal paths. Momentary information transfer along causal paths captures the amount of information transfer between any two variables lagged at two specific points in time. Our approach expands this concept to characterize the causal interaction in terms of synergistic, unique and redundant information transfer through separable causal paths. Through a graphical model, we analyze the impact of the separable and non-separable causal paths and the causality structure embedded in the graph as well as the noise effect on information partitioning by using synthetic data generated from two coupled logistic equation models. Our approach can provide a valuable reference for an autonomous information partitioning along separable causal paths which form a causal subgraph influencing a target.

I. INTRODUCTION

Complex natural and human systems, such as ecosystems and financial markets, emerge as a result of causal and/or self-organized feedback interactions among multiple variables. Lagged feedback of one variable on another can also be interpreted as a causal outcome at the time scale of the feedback dependence. Causality can also be characterized as the outcome of the interaction of information transfer [1, 2] among the variables in the system. Therefore, the amount and quality of information transfer can be used to quantify the degree and nature of causality. Our goal in this paper is to develop a framework for quantifying causal interaction arising from information transferred along separable paths that affect a target.

In this study, causality is interpreted as *strong Granger causality* [3]. Different from the causality from an interventional perspective [4], Granger causality [5] is anchored on the predictability of a target from one or more sources by measuring the variance of the target given the sources. Strong Granger causality takes a further step by investigating the predictability from sources based on the entire joint distribution of the variables involved [3]. It is extremely useful in detecting the relationships among the variables in a complex system where only observational data is available and intervention in system is hard or impossible as in natural systems which comprise a multitude of interactions. In the rest of the paper, we refer *causality* to indicate strong Granger causality for convenience.

When a source variable influences a target directly, or indirectly through a path comprising of nodes and links, it is called a causal path [6]. For example, consider a three-variable process network [7, 8] shown in Fig. 1a, where the dynamical linkages between these components of the system are assumed to be time-invariant and constructed such that: (1) X drives both Y and Z with a lag of one time step, (2) Y drives Z at a lag of two time steps, and (3) both X and Z have a self-feedback at a lag of one time step. By representing the dynamics of the process network in a time series graph [9], X affects Z lagged at two time steps through a causal path (Fig. 1b), denoted as $C_{X_{t-2} \to Z_t}$, containing two pathways: $X_{t-2} \to X_{t-1} \to Z_t$ and $X_{t-2} \to Z_{t-1} \to Z_t$. Furthermore, considering the contribution of multiple source variables on the target, the target can be affected by different sources each having its own causal path, which can be *separable* or *non-separable*. For the non-separable case, the causal path from one source contains the causal paths of the remaining sources. For instance, the causal path $C_{X_{t-2} \to Z_t}$ includes the causal path from X_{t-1} (i.e., $C_{X_{t-1} \to Z_t}$) (Fig. 1c). $C_{X_{t-2} \to Z_t}$ and $C_{X_{t-1} \to Z_t}$ are non-separable because the node X_{t-1} lies in the causal path $C_{X_{t-2} \to Z_t}$. Also, the causal paths can be separable such that no source is enslaved to the causal path of another. The separability of causal paths can be either totally independent, such as the causal paths $C_{X_{t-1} \to Z_t}$ and $C_{Y_{t-3} \to Z_t}$ (Fig. 1d), or partially independent, such as the causal paths $C_{X_{t-2} \to Z_t}$ and $C_{Y_{t-3} \to Z_t}$ which share the pathway $Z_{t-1} \to Z_t$ (Fig. 1e). We call the causal paths, which transmit the information to the same target from different sources but are not enslaved to the pathways of other sources, as *separable causal paths*. Together, the causal paths from all source nodes of interest
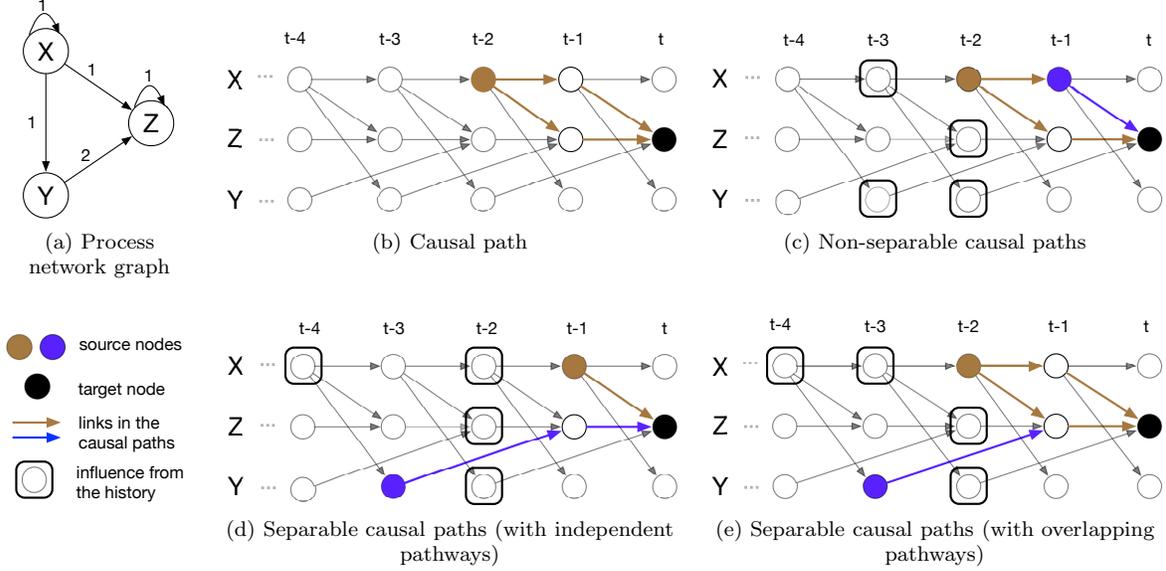
FIG. 1. The illustration of a three-variable dynamical system: (a) the process network graph illustrating the lagged interaction of three variables X, Y and Z, where the numbers on the directed links represent the lagged time step; (b) the time series graph corresponding to the process graph, with the causal path $C_{X_{t-2} \to Z_t}$ whose directed links are highlighted in brown arrows; (c) the non-separable causal paths $C_{X_{t-1} \to Z_t}$ and $C_{X_{t-2} \to Z_t}$; (d) the separable causal paths with independent pathways $C_{X_{t-1} \to Z_t}$ and $C_{Y_{t-3} \to Z_t}$; and (e) the separable causal paths with overlapping pathways $C_{X_{t-2} \to Z_t}$ and $C_{Y_{t-3} \to Z_t}$ (brown and blue nodes: source nodes; black nodes: target nodes; brown and blue links: the corresponding causal paths from the sources to the target; and nodes with black circles: nodes affecting the two separable causal paths and the target node).

that affect a target node comprise a *causal subgraph.*

An emerging approach for characterizing the nature of dependency between multiple interacting variables is based on *partial information decomposition* (PID) [10, 11], also called as *information partitioning*, which is capable of providing insights from large amounts of observational or model datasets that are becoming available [12, 13]. PID provides estimation of the information transfer from the sources to the target variable in terms of: (1) redundant information R, which captures overlapping information content that both sources provide to the target; (2) synergistic information S, which is only available due to the joint interaction of sources as they affect the target; and (3) unique information U that each source shares with the target. In the case of two sources (i.e., $X_{t-\tau_X}$ and $Y_{t-\tau_Y}$) driving Z_t, PID is given as [10, 11]

$$I(X_{t-\tau_X}, Y_{t-\tau_Y}; Z_t) = U_X + U_Y + S + R \quad (1)$$
$$I(X_{t-\tau_X}; Z_t) = U_X + R \quad (2)$$
$$I(Y_{t-\tau_Y}; Z_t) = U_Y + R, \quad (3)$$

where U_X and U_Y are unique information from $X_{t-\tau_X}$ and $Y_{t-\tau_Y}$, respectively; $I(a; b)$ is the mutual information [14] between a and b; and especially, $I(X_{t-\tau_X}, Y_{t-\tau_Y}; Z_t)$ represents the mutual information between Z_t and the union of $X_{t-\tau_X}$ and $Y_{t-\tau_Y}$. Eqs.(1)-(3) further give rise to the expression of S and R in terms

of interaction information II [14] such that

$$II(X_{t-\tau_X}; Y_{t-\tau_Y}; Z_t) = I(X_{t-\tau_X}, Y_{t-\tau_Y}; Z_t) - I(X_{t-\tau_X}; Z_t) - I(Y_{t-\tau_Y}; Z_t)$$
$$= S - R, \quad (4)$$

which quantifies the amount of information bound up in $\{X_{t-\tau_X}, Y_{t-\tau_Y}, Z_t\}$, also called as net synergy [11].

PID captures the net information transfer occurring through a multitude of causal paths involved in the interaction. But due to the high level of dependencies in a complex system, it also masks our ability to discern the causal interaction within a causal subgraph consisting of specific pathways, or possible separable causal paths. For example, as illustrated in Fig. 1d, the information partitioning associated with the interaction of X_{t-1} and Y_{t-3} on Z_t will be influenced by the external factors affecting the intermediate nodes, through their own causal paths, called as *complementary causal subgraph.* Some of the nodes in the complementary causal subgraph, that are outside of the nodes of interest (e.g., $C_{X_{t-1} \to Z_t}$, $C_{Y_{t-3} \to Z_t}$, and Z_t) but have influences on them, are highlighted as nodes with black circles, including $X_{t-4}, X_{t-2}, Y_{t-2}$ and Z_{t-2}.

Therefore, in this paper we quantify the partial information decomposition arising from the information transfer along separable causal paths in a manner that excludes the influence of complementary causal subgraph thereby eliminating the effects of external factors. Also,

we explore (1) how the PID varies between separable and non-separable causal paths; (2) how the structure of the separable causal paths affects the information partitioning; and (3) how the noise in a complex system affects these estimates. We use the recently-proposed *momentary information transfer along causal paths* (MITP) [6], which allows us to isolate the information transfer between two variables along their causal path from that associated with its complementary causal subgraph. We extend this concept to characterize the causal interaction in terms of synergistic, unique and redundant information transfer through separable causal paths by using a PID framework anchored on a rescaled approach for computing the redundancy [12] and also provide formulations associated with other prevalent measures of redundancy. This will be termed as *momentary partial information decomposition* (MPID).

This paper is organized as follows. First, we provide a brief review of the momentary information in Section II. Then, in Section III we develop the details of the mathematical framework for MPID by adopting the rescaled approach for computing redundancy. In Section IV, we show that under some conditions, the proposed MPID is entirely determined by the interactions of the nodes of interest in the causal subgraph, and autonomous of how these nodes are affected by the complementary causal subgraph. This property is called the *coupling strength autonomy* (CSA) [2]. We utilize both a linear and a nonlinear common driver model to verify the coupling strength autonomy property analytically and numerically, respectively. Moreover, in Section V, we define the MPID frameworks based on three alternative redundancy measures, and discuss their properties. In Section VI, based on the rescaled redundancy measure, we analyze MPID and PID under both separable and non-separable causal paths by using synthetic data generated from two models of coupled logistic equations with varying noise strengths. Finally, we provide summary and conclusions in Section VII. Appendix A provides definitions of abbreviations and a list of symbols, and Appendices B to D provide mathematical proofs of several formulations.

II. A REVIEW OF MOMENTARY INFORMATION

In this section, we briefly review the concept of momentary information first proposed by Pompe and Runge [1], which provides the basis for quantifying information partitioning along separable causal paths presented in the next section. This concept originates from the idea that to assess the impact of some variable X on another variable Y at time t with a specific time lag τ, the history of both $X_{t-\tau}$ and Y_t should be accounted for through conditioning so that the response of Y_t to a disturbance on $X_{t-\tau}$ accounts for the causal interaction between the two variables at the specific lag τ [1]. Therefore, the

momentary information quantifies the causal interaction between two lagged nodes of interest, in a time-series graph. To condition at the history of the nodes of interest, a network describing the causal relationships among the variables in the whole system is required [3].

Consider a complex system with a multivariate process $\vec{X} = \{V, X, Y, Z, \dots\}_N$, where N is the number of component processes. We assume that the causal dependence between the component processes are temporally invariant. The process can be represented in a time series graph G, as illustrated in Figs. 1c-1e for a three variable system, of which the basic elements include:

- Node Z_t: a subprocess Z at a specific time t.

- All the nodes at time step t: $\vec{X}_t \equiv \{V_t, X_t, Y_t, Z_t, \dots\}_N$.

- The past or history of Z_t: $\vec{X}_t^- \equiv \{\vec{X}_{t-1}, \vec{X}_{t-2}, \dots\}$.

- Directed link (or causal link) $X_{t-\tau} \to Z_t$: implying $X_{t-\tau}$ ($\tau > 0$ is the time lag) has causal influence on Z_t.

- Parents of Z_t: $P_{Z_t} \equiv \{X_{t-\tau} : X \in \vec{X}, \tau > 0, X_{t-\tau} \to Z_t\}$. A node $X_{t-\tau}$ is a parent of Z_t if and only if there is a directed link/edge from $X_{t-\tau}$ to Z_t (i.e., $X_{t-\tau} \to Z_t$).

- Causal path from $X_{t-\tau}$ to Z_t: $C_{X_{t-\tau} \to Z_t} \equiv \{V_{t-\tau_V} : V \in \vec{X}, \tau_V > 0, X_{t-\tau} \to \cdot \to V_{t-\tau_V} \to \cdot \to Z_{t-\tau_Z}\} \cup \{X_{t-\tau}\}$.

In this study, we assume that all the causal links are identified from the criteria that only the past affects the future. The causality in this study is defined in the context of *strong Granger causality* [3, 5], that is, a pair of nodes $X_{t-\tau} \in G$ ($\tau > 0$) and $Z_t \in G$ are connected by a directed link $X_{t-\tau} \to Z_t$ if and only if:

$$X_{t-\tau} \not\perp Z_t \mid [\vec{X}_t \backslash \{X_{t-\tau}\}], \tag{5}$$

where $\not\perp$, $\mid$ and $\backslash$ are the dependent, conditioning and subtraction symbols, respectively. It is anchored on the idea that $X_{t-\tau}$ *Granger-causes* Z_t if the two are still dependent on each other when conditioned on the remaining process.

Furthermore, because the directionality of the causal relationship between two contemporaneous nodes (e.g., X_t and Y_t) is ambiguous, we do not consider the undirected contemporaneous link (a link connecting two nodes at the same time step, e.g., $X_t - Y_t$) in this study. Thus, it allows the time series graph to be a directed acyclic graph (DAG), where no directed loops exists [15]. To connect the DAG with the joint probability implied by the graph, we assume *causal Markov property* [15] that given the parents , P_{Z_t}, of any $Z_t \in \vec{X}$, Z_t is independent of its non-descendants $\vec{X}_t^- \backslash P_{Z_t}$ in the graph, which are the earlier dynamics excluding the direct causes of Z_t.
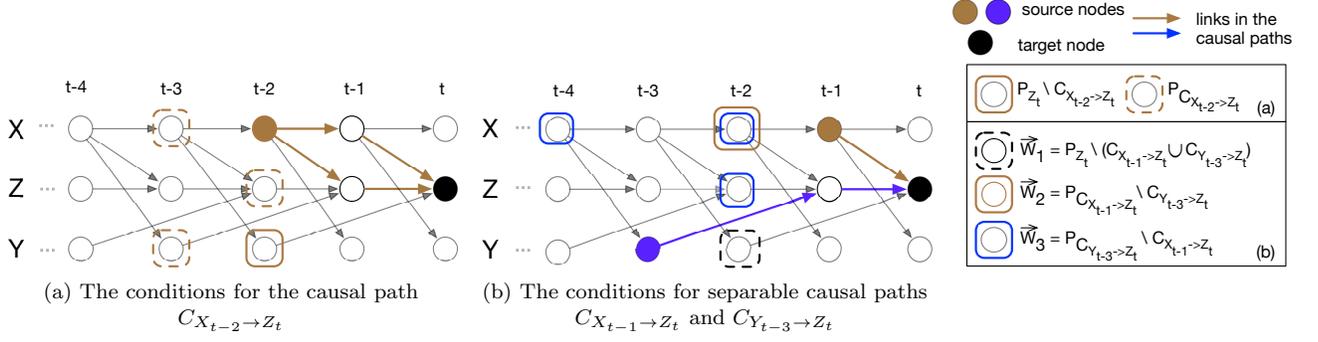
FIG. 2. Illustration of the Markovian conditions for the causal subgraphs comprising one causal path and two separable causal paths shown in Figs. 1b and 1d, respectively. (a) The conditions for the causal path $C_{X_{t-2}\to Z_t}$ (brown solid circles: $P_{Z_t}\backslash C_{X_{t-2}\to Z_t}$ where $\backslash$ represents the subtraction operation for sets; the brown dashed circles: $P_{C_{X_{t-2}\to Z_t}}$). (b) The conditions for the separable causal paths $C_{X_{t-1}\to Z_t}$ and $C_{Y_{t-3}\to Z_t}$ (the black dashed circle: $\vec{W}_1$; the brown solid circle: $\vec{W}_2$; the blue solid circles: $\vec{W}_3$).

By drawing upon the causal Markov property, the potential infinite number of conditions in Eq.(5) from the past can be reduced to a finite number of conditions corresponding to the parents of the nodes of interest [9]. Hence, the condition in Eq.(5) can be reduced into the parents of the two nodes such that

$$X_{t-\tau} \not\perp Z_t \mid [P_{X_{t-\tau}} \cup (P_{Z_t}\backslash\{X_{t-\tau}\})], \qquad (6)$$

where $P_{X_{t-\tau}}$ represents the parents of the source node $X_{t-\tau}$, and $P_{Z_t}\backslash\{X_{t-\tau}\}$ denotes the parents of Z_t excluding the node $X_{t-\tau}$ if $X_{t-\tau} \in P_{Z_t}$.

Based on Eq.(6), momentary information transfer $I^{MIT}_{X_{t-\tau}\to Z_t}$ quantifies the coupling strength between two nodes (e.g., $X_{t-\tau}$ and Z_t), and is estimated as a conditional mutual information as follows [14]:

$$I^{MIT}_{X_{t-\tau}\to Z_t} = I(X_{t-\tau}; Z_t \mid [P_{X_{t-\tau}} \cup (P_{Z_t}\backslash\{X_{t-\tau}\})]). \quad (7)$$

Similarly, to account for the amount of information transferred from $X_{t-\tau}$ to Z_t via their causal path $C_{X_{t-\tau}\to Z_t}$, the momentary information transfer along the causal path (MITP) is expressed as [6]:

$$I^{MITP}_{X_{t-\tau}\to Z_t} = I(X_{t-\tau}; Z_t \mid$$
$$[P_{C_{X_{t-\tau}\to Z_t}} \cup (P_{Z_t}\backslash C_{X_{t-\tau}\to Z_t})]). \qquad (8)$$

$I^{MITP}_{X_{t-\tau}\to Z_t}$ is a conditional mutual information [14], where $P_{C_{X_{t-\tau}\to Z_t}}$ represents the parents of the causal path $C_{X_{t-\tau}\to Z_t}$ and $P_{Z_t}\backslash C_{X_{t-\tau}\to Z_t}$ denotes the parents of Z_t excluding the nodes on the causal path $C_{X_{t-\tau}\to Z_t}$. Consider Fig. 2a as an example. The conditions for $I^{MITP}_{X_{t-2}\to Z_t}$ are $\{X_{t-3}, Z_{t-2}, Y_{t-2}, Y_{t-3}\}$. By conditioning on the parents of the nodes involved (i.e., $\{Z_t\} \cup C_{X_{t-\tau}\to Z_t}$), the calculated interaction information prevents the information of the complementary causal graph from entering into the calculation.

The idea of momentary information is further extended to analyze how one node (i.e., $Y_{t-\tau_Y}$) in the causal path $C_{X_{t-\tau_X}\to Z_t}$ affects the information transfer in $C_{X_{t-\tau_X}\to Z_t}$ through momentary interaction information (MII) [6]. Similar to MITP, by conditioning on the parents of the causal path, MII is given by:

$$\Delta I^M_{X_{t-\tau_X}\to Z_t|Y_{t-\tau_Y}} = II(X_{t-\tau_X}; Y_{t-\tau_Y}; Z_t \mid$$
$$[P_{C_{X_{t-\tau_X}\to Z_t}} \cup (P_{Z_t}\backslash C_{X_{t-\tau_X}\to Z_t})]), \qquad (9)$$

which is a conditional interaction information [14].

The utilization of a time series graph not only allows the visualization of the interactions among several nodes, but also facilitates the quantification of these interactions across different time points through different momentary information measures. Especially, MITP and MII [6] are the first attempts to quantify the information transfer between two nodes through a causal path and the casual interactions among three nodes in a causal path, which provides a starting point to develop the *momentary partial information decomposition* (MPID) described in the next section.

III. MOMENTARY PARTIAL INFORMATION DECOMPOSITION

In general, a target node Z_t is not only influenced by one lagged source node $X_{t-\tau}$ through either a direct link or a causal path, but also driven by multiple lagged source nodes $\vec{V}$ through a multitude of causal paths, which forms a causal subgraph $C_{\vec{V}\Rightarrow Z_t} \equiv \cup_{X_{t-\tau}\in\vec{V}} C_{X_{t-\tau}\to Z_t}$. Also, it is clear that the dynamics within a causal subgraph are affected by the remaining graph preceding the target, $\vec{X}_t^-\backslash C_{\vec{V}\Rightarrow Z_t}$, called as *complementary causal subgraph*. The interaction among multiple nodes in a time series graph poses a question of how to characterize the different contents of information transfer through a causal subgraph consisting of multiple

causal paths, while at the same time, with the influence of its complementary causal subgraph eliminated.

In this section, we describe the mathematical framework of the *momentary partial information decomposition* (MPID) for quantifying and categorizing the information transfer to a target from a preceding causal subgraph starting with two sources with separable causal paths. First, we build on the momentary interaction information, which only considers one causal path, to formulate the *momentary interaction information for separable causal paths* (MII-SCP). Then, we review the recent advancements in PID, including the rescaled redundancy approach for estimating PID [12]. Lastly, based on the chosen PID framework, MPID is developed for partitioning the interaction of information transfer arising from separable causal paths into synergistic, redundant and unique components.

A. Momentary Interaction Information for Separable Causal Paths

From a time series graph perspective, the PID of a three-variable interaction at specific lagged times is induced by the information transfer from a subgraph containing two separable causal paths of the sources towards the target, such as the case shown in Fig. 1e. The separability of the two causal paths emphasizes the fact that neither source lies in the causal path of the other. That is, two causal paths (i.e., $C_{X_{t-\tau_X} \to Z_t}$ and $C_{Y_{t-\tau_Y} \to Z_t}$) are *separable* if $(Y_{t-\tau_Y} \notin C_{X_{t-\tau_X} \to Z_t}) \wedge (X_{t-\tau_X} \notin C_{Y_{t-\tau_Y} \to Z_t})$, where $\wedge$ is the logical AND symbol. However, it should be noted that the two causal paths can also be non-separable if one source belongs to the causal path of the other (such as the case in Fig. 1c), in which case, as shown below, the formulation reduces to that in Eq.(9).

To compute MPID, the calculated interaction information along two separable causal paths in Eq.(4) is required to be isolated from complementary causal subgraph containing the historical information of the dynamics. Nevertheless, MII is not an appropriate option to use as it is formulated for the situation of one causal path as shown in Eq.(9). Hence, the *momentary interaction information for separable causal paths* (MII-SCP) is given as

$$\Delta I^{MSCP}_{\{X_{t-\tau_X}, Y_{t-\tau_Y}\} \to Z_t} = II(X_{t-\tau_X}; Y_{t-\tau_Y}; Z_t \mid \vec{W}), \quad (10)$$

where

$$\vec{W} = \vec{W}_1 \cup \vec{W}_2 \cup \vec{W}_3 \quad (11a)$$

with

$$\vec{W}_1 = P_{Z_t} \backslash (C_{X_{t-\tau_X} \to Z_t} \cup C_{Y_{t-\tau_Y} \to Z_t}) \quad (11b)$$

$$\vec{W}_2 = P_{C_{X_{t-\tau_X} \to Z_t}} \backslash C_{Y_{t-\tau_Y} \to Z_t} \quad (11c)$$

$$\vec{W}_3 = P_{C_{Y_{t-\tau_Y} \to Z_t}} \backslash C_{X_{t-\tau_X} \to Z_t}. \quad (11d)$$

The condition set, $\vec{W}$, for the MII-SCP represents the parents of the union set of both the target Z_t and the causal paths from the two sources to the target (i.e. $C_{X_{t-\tau_X} \to Z_t}, C_{Y_{t-\tau_Y} \to Z_t}$). It consists of three parts: (1) $\vec{W}_1$, the parents of the target node P_{Z_t} excluding those in the two causal paths; (2) $\vec{W}_2$, the parents of the causal path $C_{X_{t-\tau_X} \to Z_t}$ (i.e., $P_{C_{X_{t-\tau_X} \to Z_t}}$) excluding the nodes in $C_{Y_{t-\tau_Y} \to Z_t}$, and (3) $\vec{W}_3$, the parents of $C_{Y_{t-\tau_Y} \to Z_t}$ (i.e., $P_{C_{Y_{t-\tau_Y} \to Z_t}}$) excluding those in $C_{X_{t-\tau_X} \to Z_t}$. It is noted that when the causal paths of the two sources are non-separable, MII-SCP is reduced to MII in Eq.(9).

In the example of the three-variable system in Fig. 2b, the condition $\vec{W}$ for MII-SCP between the target Z_t and two sources X_{t-1} and Y_{t-3} (i.e., $\Delta I^{MII-SCP}_{\{X_{t-\tau_1}, Y_{t-\tau_2}\} \to Z_t}$) includes (1) $\vec{W}_1 = \{Y_{t-2}\}$ (black dashed circle), (2) $\vec{W}_2 = \{X_{t-2}\}$ (brown solid circle) and (3) $\vec{W}_3 = \{X_{t-4}, X_{t-2}, Z_{t-2}\}$ (blue solid circle). Therefore, $\vec{W} = \{X_{t-2}, X_{t-4}, Y_{t-2}, Z_{t-2}\}$.

Furthermore, when two causal paths are non-separable (i.e., one source lies in the causal path of the other source), MII-SCP collapses into MII. Suppose the source $Y_{t-\tau_Y} \in C_{X_{t-\tau_X} \to Z_t}$, then $C_{Y_{t-\tau_Y} \to Z_t} \in C_{X_{t-\tau_X} \to Z_t}$). The first term in the condition $\vec{W}$, i.e., $P_{Z_t} \backslash (C_{X_{t-\tau_X} \to Z_t} \cup C_{Y_{t-\tau_Y} \to Z_t})$, is reduced to $P_{Z_t} \backslash C_{X_{t-\tau_X} \to Z_t}$. Also, the remaining terms in $\vec{W}$, i.e., $\vec{W}_2 \cup \vec{W}_3$, can be simplified to $P_{C_{X_{t-\tau_X} \to Z_t}}$, since the purpose of the two terms is to choose all the parents of the union of two causal paths which is now $C_{X_{t-\tau_X} \to Z_t}$. Therefore, the condition set $\vec{W}$ is reduced to $P_{C_{X_{t-\tau_X} \to Z_t}} \cup (P_{Z_t} \backslash C_{X_{t-\tau_X} \to Z_t})$, the same condition as in MII in (Eq.(9)).

B. Framework for Partial Information Decomposition

Besides MII-SCP, MPID needs to provide a further characterization of the information transfer from $X_{t-\tau_X}$ and $Y_{t-\tau_Y}$ to Z_t through the corresponding causal paths in terms of the synergistic, unique and redundant information. Nevertheless, to compute the four components U_Y, U_X, R and S in the PID from three equations Eqs.(1)-(3), one more condition is needed. To that end, various approaches for quantifying one of the components in PID, that is, redundancy, synergy and unique information, have been proposed. In addition to propose the PID framework, Williams and Beers [10] were also the first to propose a redundancy measure (see Section) based on the minimum specific information for the target at the outcome of a source value. However, this measure overestimates the redundancy [16, 17], because it assumes that the amount of information shared between the two variables is also the same as the impact of the two variables on the target variable. In fact, it represents the

upper bound of the information that can be redundantly shared with the target variable. Despite this drawback, it emphasizes the idea that the redundancy and unique information are dependent on the marginal joint distribution between each source variable and the target, which enables a series of further measures for unique [18], synergistic [19, 20] and redundant [11, 21] information. Barrett [11] defined the redundancy as the minimum mutual information between each source and the target, which is equivalent to some of the early PID frameworks [10, 18, 19] in Gaussian systems. Furthermore, in computing the redundancy, besides anchoring the redundancy upon the marginal joint distributions between the sources and the target, there is also a trend to utilize other aspects in the joint distribution of all the considered variables in computing the redundancy. Harder et al. [16] quantified the redundant information based on the distance between the conditional distribution of the target given each source by using Kullback-Leibler divergence. Ince [17] computed the redundancy as the expected pointwise change of the surprisal of the target given the sources from the local interaction information based on an entropy-maximized joint distribution, thus distinguishing the redundancy-related elements in interaction information. Goodwell and Kumar [12] put forward a rescaled redundancy which considers the mutual information between the sources in accounting for the minimum mutual information-based redundancy.

In this study, to formulate the momentary partial information decomposition (MPID), we adopt the PID framework based on the rescaled redundancy, R_S, [12], which is given by:

$$R_S = R_{min} + I_s(R_{MMI} - R_{min}), \qquad (12a)$$

where

$$R_{MMI} = \min[I(X_{t-\tau_X}; Z_t), I(Y_{t-\tau_Y}; Z_t)] \qquad (12b)$$

$$I_s = \frac{I(X_{t-\tau_X}; Y_{t-\tau_Y})}{\min[H(X_{t-\tau_X}), H(Y_{t-\tau_Y})]} \qquad (12c)$$

$$R_{min} = \begin{cases} 0, & \text{if } II \geq 0 \\ -II, & \text{otherwise.} \end{cases} \qquad (12d)$$

R_{MMI} represents the *minimum mutual information*, originating from the fact that the redundant information should not be larger than the smallest value of the mutual information between any source and target variables [11]. We choose the rescaled approach in computing the redundancy because it is able to reduce the overestimation from R_{MMI} as well as guarantee non-negativity of PID, and it is also empirically computable from time series observational data. To ensure a non-negative PID, R_{MMI} and R_{min} are used for providing the upper and lower bounds, respectively, for the redundant information. It is noted that we assume the non-negativity of PID in this study, which is termed as "local positivity" property in [22], because the non-negative partitioning is essential in illustrating physical phenomenon in a meaningful

way, though there exist considerations for dropping the non-negativity condition for PID [17]. Furthermore, the rescaling coefficient I_s is used to scale R_{MMI} (which overestimates the redundancy) based on the mutual information between the sources. Besides, many proposed PID frameworks are computationally difficult or costly when using observational time series data due to the requirement of an optimization procedure for computing a metric based on marginal distributions [17, 18, 21]. Meanwhile, the rescaled redundancy approach, which has been applied in ecohydrological time series data [12, 13], is empirically computable and thus may have a broader applications associated with empirical analysis across many physical domains. Also, it is noted that the rescaled approach, as formulated in Eq.(12) [12], is applicable to the interaction between two sources similar to other redundancy measures [16, 18, 19].

Nonetheless, recognizing that there is no universal agreement on the additional condition for quantifying PID so far, we also provide the formulations for the MPIDs based on three other redundancy measures in Section V.

C. Framework for Momentary Partial Information Decomposition

Based on the rescaled redundancy in Eq.(12), in a manner similar to MII-SCP, we propose the *momentary partial information decomposition* (MPID) by conditioning all the components in the original PID (Eqs.(1)-(3)) on $\vec{W}$ such that:

$$\Delta I^{MSCP}_{\{X_{t-\tau_X}, Y_{t-\tau_Y}\} \to Z_t} = S_c - R_c \qquad (13a)$$

$$R_c = R_{min,c} + I_{s,c}(R_{MMI,c} - R_{min,c}) \qquad (13b)$$

$$U_{X,c} = I(Z_t; X_{t-\tau_X}|\vec{W}) - R_c \qquad (13c)$$

$$U_{Y,c} = I(Z_t; Y_{t-\tau_Y}|\vec{W}) - R_c, \qquad (13d)$$

where

$$R_{MMI,c} = \min[I(X_{t-\tau_X}; Z_t|\vec{W}), I(Y_{t-\tau_Y}; Z_t|\vec{W})] \quad (13e)$$

$$I_{s,c} = \frac{I(X_{t-\tau_X}; Y_{t-\tau_Y}|\vec{W})}{\min[H(X_{t-\tau_X}|\vec{W}), H(Y_{t-\tau_Y}|\vec{W})]} \qquad (13f)$$

$$R_{min,c} = \begin{cases} 0, & \text{if } \Delta I^{MSCP} \geq 0 \\ -\Delta I^{MSCP}, & \text{otherwise.} \end{cases} \qquad (13g)$$

The subscript c stands for "conditional". We note that the original suggestion of forming conditional redundancies was given by Bertschinger et al. [22] [23].

IV. COUPLING STRENGTH AUTONOMY FOR MOMENTARY PARTIAL INFORMATION DECOMPOSITION

The causal Markov property of the time series graph ensures the momentary information approaches (e.g.,

MII, MITP and MII-SCP) exclude the influence of the complementary causal subgraph. This allows us to approximate the causal interactions among variables of interest. However, these metrics are still dependent on the Markovian conditions, such as $\vec{W}$ in MII-SCP. Nevertheless, Runge [6] showed that under some conditions, MII and MITP are autonomous of how the nodes of interest interact with the nodes in the complementary causal subgraph, a property described as *coupling strength autonomy*. In this section, we generalize this property to MPID. Two conditions (i.e., *additivity* and *linearity*) used for the verification of coupling strength autonomy property in MII are also adopted here, with one more condition, *separability*, which allows the analysis of MPID when the causal paths of the two source nodes are separable. Then, the coupling strength autonomy property of MPID is shown for two common driver models. In the first model, the interaction among the variables of interest is linear, and nonlinear in the second model. For the former, the analytical solution of MPID is derived and compared with the interaction information without conditioning. For the latter, a numerical simulation is conducted to estimate both MPID and PID.

A. Coupling strength autonomy in MPID

Consider a multivariate stationary discrete-time process $\vec{X} = \{X, Y, Z, \dots\}_N$, where X, Y, Z, etc. are sub-processes. The process $\vec{X}$ satisfies the causal Markov property with its corresponding time series graph G as described in Section II. We assume that for $\tau_X, \tau_Y \geq 0$, both the source nodes $X_{t-\tau_X}$ and $Y_{t-\tau_Y}$ are connected with the target node Z_t through two causal paths $C_{X_{t-\tau_X} \to Z_t}$ and $C_{Y_{t-\tau_Y} \to Z_t}$, respectively. Also, the union of the two causal paths and the target node is represented as $\vec{B}$, that is

$$\vec{B} = C_{X_{t-\tau_X} \to Z_t} \cup C_{Y_{t-\tau_Y} \to Z_t} \cup Z_t. \quad (14)$$

For the dependencies of each node $K_{t-\tau} \in \vec{B}$ ($\tau \geq 0$), the following conditions are defined:

(i) *Additivity*: Dependencies of $K_{t-\tau}$ in the following are additive:

- its parents belonging to $\vec{B}$ (denoted as $P^{\vec{B}}_{K_{t-\tau}} \equiv P_{K_{t-\tau}} \cap \vec{B}$),

- its remaining parents (denoted as $P_{K_{t-\tau}} \backslash P^{\vec{B}}_{K_{t-\tau}}$), and

- the noise term representing the modelling uncertainty.

Therefore, any $K_{t-\tau} \in \vec{B}$ ($\tau \geq 0$) can be written as

$$K_{t-\tau} = f_K(P^{\vec{B}}_{K_{t-\tau}}) + g_K(P_{K_{t-\tau}} \backslash P^{\vec{B}}_{K_{t-\tau}}) + \eta^K_{t-\tau}, \quad (15)$$

where f_K, g_K are the arbitrary functions for $K_{t-\tau}$, and $\eta^K_{t-\tau}$ is the noise term which is assumed as independent and identically distributed (i.i.d.). It can be observed that $P^{\vec{B}}_{K_{t-\tau}}$ (i.e., the dependencies in f_K) belongs to $\vec{B}$, while $P_{K_{t-\tau}} \backslash P^{\vec{B}}_{K_{t-\tau}}$ (i.e., the dependencies in g_K) belongs to $\vec{W}$ defined in Eq.(10).

(ii) *Linearity in f_K*: The function f_K for each node $K_{t-\tau} \in \vec{B}$ is linear. It means the dependence of $K_{t-\tau}$ on the part of its parents belonging to $\vec{B}$ (i.e., $P^{\vec{B}}_{K_{t-\tau}}$), is linear. The linearity also implies that the nodes in $\vec{B}$ (Eq.(14)) linearly depend on each other.

(iii) *Separability of the causal paths $C_{X_{t-\tau_X} \to Z_t}$ and $C_{Y_{t-\tau_Y} \to Z_t}$*: Neither of the two sources lies in the causal path from the other source to the target. That is, two causal paths (i.e., $C_{X_{t-\tau_X} \to Z_t}$ and $C_{Y_{t-\tau_Y} \to Z_t}$) are *separable* if $(Y_{t-\tau_Y} \notin C_{X_{t-\tau_X} \to Z_t}) \wedge (X_{t-\tau_X} \notin C_{Y_{t-\tau_Y} \to Z_t})$.

Theorem: Coupling strength autonomy for momentary partial information decomposition. In a stationary discrete-time multivariate process $\vec{X}$, which meets the causal Markov property in its time series graph, the MPID for the contribution from two sources $X_{t-\tau_X}$ and $Y_{t-\tau_Y}$ ($\tau_X, \tau_Y \geq 0$) to the target Z_t have the following properties (proof is given in the Appendices B & C):

(a) If all the three conditions (i.e., *additivity*, *linearity* and *separability*) hold, MPID defined in Eqs.(13a)-(13b) can be written as:

$$\Delta I^{MSCP}_{\{X_{t-\tau_X}, Y_{t-\tau_Y}\} \to Z_t}$$
$$= I(\eta^X_{t-\tau_X}; \eta^Y_{t-\tau_Y} \mid \tilde{f}_Z(\bullet) + \eta^Z_t) \quad (16a)$$
$$R_c = R_{min,c} \quad (16b)$$
$$S_c = \Delta I^{MSCP} + R_c \quad (16c)$$
$$U_{X,c} = I(\eta^X_{t-\tau_X}; \tilde{f}_Z(\bullet) + \eta^Z_t) - R_c \quad (16d)$$
$$U_{Y,c} = I(\eta^Y_{t-\tau_Y}; \tilde{f}_Z(\bullet) + \eta^Z_t) - R_c, \quad (16e)$$

where $R_{min,c}$ is given in Eq.(13b) and $\tilde{f}_Z(\bullet)$ is a linear combination of all the noise terms η_t of the nodes in $\vec{B}$, which are simplified as the symbol $\bullet$. In brief, *additivity* allows the exclusion of the dependencies not in $\vec{B}$ (i.e., $g_K(P_{K_{t-\tau}} \backslash P^{\vec{B}}_{K_{t-\tau}})$) in the calculation of the information partitioning due to the translational invariance property of both entropy and mutual information (see Appendix B for details). Furthermore, *separability* ensures both the minimum redundancy and the zeros of f_X and f_Y, because $X_{t-\tau_X}$ and $Y_{t-\tau_Y}$ now do not depend on any nodes in the two separable causal paths. Also, *linearity* converts the dependencies in the $\vec{B}$ (i.e., $f_K(P^{\vec{B}}_{K_{t-\tau}})$) to be linear functions $\tilde{f}$ of all the noise terms in $\vec{B}$ after the exclusion of the non-linear dependencies $g_K(P_{K_{t-\tau}} \backslash P^{\vec{B}}_{K_{t-\tau}})$. Therefore, the condition set $\vec{W}$ is cancelled out since the calculations are now only determined by the nodes in the causal paths and the target node (*i.e.*, $\vec{B}$) due to the *linearity* and *additivity* conditions.

(b) If only *additivity* and *linearity* hold, both f_X and

f_Y can be converted into linear functions of the noise terms of the nodes in $\vec{B}$ (similar to how f_Z is converted into $\tilde{f}_Z$ in (a)). However, since *separability* does not hold, the redundancy R_c is not the minimum and f_X and f_Y can be nonzero. Hence, we can express the corresponding MPID as

$$\Delta I^{MSCP}_{\{X_{t-\tau_X}, Y_{t-\tau_Y}\}\to Z_t}$$
$$= II(\tilde{f}_X(\bullet) + \eta^X_{t-\tau_X}; \tilde{f}_Y(\bullet) + \eta^Y_{t-\tau_Y}; \tilde{f}_Z(\bullet) + \eta^Z_t) \tag{17a}$$

$$R_c = R_{min,c} + I_{s,c}(R_{MMI,c} - R_{min,c}) \tag{17b}$$

$$S_c = \Delta I^{MSCP} + R_c \tag{17c}$$

$$U_{X,c} = I(\tilde{f}_X(\bullet) + \eta^X_{t-\tau_X}; \tilde{f}_Z(\bullet) + \eta^Z_t) - R_c \tag{17d}$$

$$U_{Y,c} = I(\tilde{f}_Y(\bullet) + \eta^Y_{t-\tau_Y}; \tilde{f}_Z(\bullet) + \eta^Z_t) - R_c, \tag{17e}$$

where both $\tilde{f}_X(\bullet)$ and $\tilde{f}_Y(\bullet)$ are the linear functions of all the noise terms η_t of the nodes in $\vec{B}$ for $X_{t-\tau_X}$ and $Y_{t-\tau_Y}$, respectively, and

$$R_{MMI,c} = \min[I(\tilde{f}_X(\bullet) + \eta^X_{t-\tau_X}; \tilde{f}_Z(\bullet) + \eta^Z_t),$$
$$I(\tilde{f}_Y(\bullet) + \eta^Y_{t-\tau_Y}; \tilde{f}_Z(\bullet) + \eta^Z_t)] \tag{18a}$$

and

$$I_{s,c} = \frac{I(\tilde{f}_X(\bullet) + \eta^X_{t-\tau_X}; \tilde{f}_Y(\bullet) + \eta^Y_{t-\tau_Y})}{\min[H(\tilde{f}_X(\bullet) + \eta^X_{t-\tau_X}), H(\tilde{f}_Y(\bullet) + \eta^Y_{t-\tau_Y})]}. \tag{18b}$$

(c) If only *additivity* and *separability* hold, the redundancy R_c is minimized with f_X and f_Y vanishing. However, because of the nonlinearity in f, the condition set $\vec{W}$ cannot be cancelled out. Thus, the MPID is reduced to

$$\Delta I^{MSCP}_{\{X_{t-\tau_X}, Y_{t-\tau_Y}\}\to Z_t}$$
$$= I(\eta^X_{t-\tau_X}; \eta^Y_{t-\tau_Y} \mid \{f_Z(P^{\vec{B}}_{Z_t}) + \eta^Z_t\} \cup \vec{W}) \tag{19a}$$

$$U_{X,c} = I(\eta^X_{t-\tau_X}; f_Z(P^{\vec{B}}_{Z_t}) + \eta^Z_t \mid \vec{W}) - R_c \tag{19b}$$

$$U_{Y,c} = I(\eta^Y_{t-\tau_Y}; f_Z(P^{\vec{B}}_{Z_t}) + \eta^Z_t \mid \vec{W}) - R_c. \tag{19c}$$

It can be observed that the *additivity* and *linearity* conditions allow MPID to be dependent only on the nodes in the union of the two causal paths and the target node (see Eqs.(16) and (17)).

Furthermore, under the *separability* condition, the redundant information R_c achieves the minimum value $R_{min,c}$ (see Eq.(16b)). It is intuitive that without the influence of other factors (e.g., a common driver), the two *separable* source variables are independent, and therefore, the redundant information is minimized.

B. A Common Driver model

Now we verify the coupling strength autonomy property for MPID in both a linear and a nonlinear model

solved analytically and numerically, respectively. Let's consider a common driver model involving four subprocesses (i.e., V X, Y and Z). V is the common driver of X and Y, both of which cause Z. All the causal relationships are delayed at one time step. Fig. 3 illustrates the process network graph and the time series graph of the model. We show that by adopting MPID, the effect of the common driver V in the PID of X, Y and Z is eliminated completely in the linear model and significantly in the nonlinear model.

1. A linear common driver model

The linear common driver model can be written as

$$V_t = \eta^V_t$$
$$X_t = c_{VX} V_{t-1} + \eta^X_t$$
$$Y_t = c_{VY} V_{t-1} + \eta^Y_t$$
$$Z_t = c_{XZ} X_{t-1} + c_{YZ} Y_{t-1} + \eta^Z_t. \tag{20}$$

The i.i.d. noise terms for each variable are represented as η^V_t, η^X_t, η^Y_t and η^Z_t, following normal distributions $\mathcal{N}(0, \sigma^2_V)$, $\mathcal{N}(0, \sigma^2_X)$, $\mathcal{N}(0, \sigma^2_Y)$ and $\mathcal{N}(0, \sigma^2_Z)$, respectively. The coefficients, c_{VX}, c_{VY}, c_{XZ} and c_{YZ}, are the coupling strengths between variables. Especially, c_{VX} and c_{VY} quantify the effect of the common driver of V_{t-1} on X_t and Y_t, respectively. In the following, the analytical solutions of both MPID and PID for quantifying the information transfer from two sources X_{t-1} and Y_{t-1} and the target Z_t are shown sequentially.

MPID for X_{t-1}, Y_{t-1} and Z_t: According to the definition of the condition in Eq.(10), the condition set $\vec{W}$ for $\Delta I^{MSCP}_{\{X_{t-1}, Y_{t-1}\}\to Z_t}$ is $\{V_{t-2}\}$ as shown in Fig. 3b. Therefore, the analytical solutions of $\vec{W}$ for $\Delta I^{MSCP}_{\{X_{t-1}, Y_{t-1}\}\to Z_t}$, R_c and S_c are given as:

$$\Delta I^{MSCP}_{\{X_{t-1}, Y_{t-1}\}\to Z_t}$$
$$= \frac{1}{2} \ln\left\{1 + \frac{c^2_{XZ} c^2_{YZ} \sigma^2_X \sigma^2_Y}{\sigma^2_Z(\sigma^2_Z + c^2_{XZ} \sigma^2_X + c^2_{YZ} \sigma^2_Y)}\right\} \tag{21a}$$

$$R_c = 0 \tag{21b}$$

$$S_c = \Delta I^{MSCP}_{\{X_{t-1}, Y_{t-1}\}\to Z_t}. \tag{21c}$$

The derivation is available in Appendix D. It is easy to verify that the linear common driver model example fulfills all the three conditions (*linearity*, *additivity* and *separability*), thus the corresponding MPID follows the results in Eqs.(16a)-(16e). Obviously, MPID is autonomous such that it is independent of c_{VX} and c_{VY} – the impact from the common driver V_{t-2}. Moreover, in this case, because $\Delta I^{MSCP}_{\{X_{t-1}, Y_{t-1}\}\to Z_t} > 0$, the redundant information R_c achieves the minimum value $R_{min} = 0$ in MPID. It implies that there is only synergistic information in the contribution of X_{t-1} and Y_{t-1} in generating Z_t.

PID for X_{t-1}, Y_{t-1} and Z_t: The analytical solution of the interaction information $II(X_{t-1}; Y_{t-1}; Z_t)$ is given by

$$II(X_{t-1}; Y_{t-1}; Z_t) =$$
$$\frac{1}{2} \ln \left\{ \frac{(c_{XZ}^2 b + \Gamma_X \sigma_Z^2)(c_{YZ}^2 b + \Gamma_Y \sigma_Z^2)}{\sigma_Z^2 \Gamma_X \Gamma_Y (\sigma_Z^2 + c_{XZ}^2 \Gamma_X^2 + c_{YZ}^2 \Gamma_Y^2 + d)} \right\},$$
$$(22)$$

where $b = c_{VY}^2 \sigma_V^2 \sigma_X^2 + c_{VX}^2 \sigma_V^2 \sigma_Y^2 + \sigma_Y^2 \sigma_X^2$, $d = 2 c_{XZ} c_{YZ} c_{VY} c_{VX} \sigma_V^2$, $\Gamma_X = c_{VX}^2 \sigma_V^2 + \sigma_X^2$, $\Gamma_Y = c_{VY}^2 \sigma_V^2 + \sigma_Y^2$ and $d = c_{YZ}^2 \Gamma_Y^2 + 2 c_{XZ} c_{YZ} c_{VY} c_{VX} \sigma_V^2$. For a full derivation, see Appendix D.

Eq.(22) shows that $II(X_{t-1}; Y_{t-1}; Z_t)$ depends on c_{VX} and c_{VY} through b, d, Γ_X and Γ_Y. The dependence implies that the common driver V plays a role in determining $II(X_{t-1}; Y_{t-1}; Z_t)$, which is in contrast to $\Delta I_{\{X_{t-1}, Y_{t-1}\} \to Z_t}^{MSCP}$ which exhibits the coupling strength autonomy property. It demonstrates that in this linear system, by conditioning on the common driver V_{t-2}, the proposed momentary information measure, $\Delta I_{\{X_{t-1}, Y_{t-1}\} \to Z_t}^{MSCP}$, is able to eliminate the influence from the complementary causal subgraph.

To explore this further, the coupling strength coefficients c_{VX} and c_{VY} are altered to see how V affect S and R. Suppose X and Y are strongly driven by V such that c_{VX} and c_{VY} are much larger than other coefficients in Eq.(20). Also, assume both the coupling coefficients c_{VX} and c_{VY} are in the same order of magnitude, that is:

$$c_{VX} \approx c_{VY} \approx h, \text{ and}$$
$$h \gg c, \forall c \in \{c_{XZ}, c_{YZ}, \sigma_V, \sigma_X, \sigma_Y, \sigma_Z\}.$$

Therefore, $II(X_{t-1}; Y_{t-1}; Z_t)$ in Eq.(22) can be reduced to

$$II(X_{t-1}; Y_{t-1}; Z_t) \approx \frac{1}{2} \ln \frac{1}{h^2}. \qquad (23)$$

Especially, when $|h| > 1$, $II(X_{t-1}; Y_{t-1}; Z_t) < 0$, implying that $S < R$ according to the relationship among II, S and R in Eq.(4). It means a strong coupling strength from the common driver V_{t-2} would result in more redundant information from X_{t-1} and Y_{t-1}, even though the dynamics among X_{t-1}, Y_{t-1} and Z_t alone do not suggest any redundancy from the two sources as shown in Eq.(21c). This is crucial in that the empirically estimated PID without an appropriate conditioning as in ΔI^{MSCP} would probably be influenced by other factors (e.g., the common driver in this example), thus hiding the true dynamics of the variables of interest.

2. A nonlinear common driver model

Next, we examine the coupling strength autonomy of MPID for a nonlinear model. The nonlinear model still follows the same causality structure of the common driver
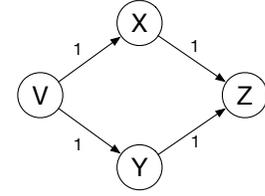
(a) The process network graph
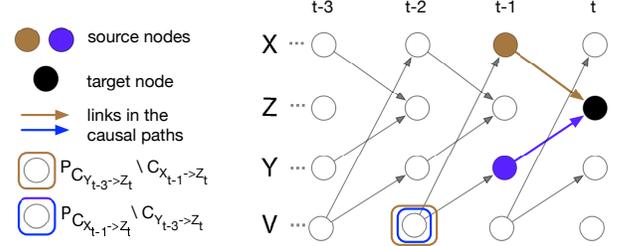
(b) The time series graph

FIG. 3. The common driver model. (a) is the process network graph representation, where the numbers on the directed links represent the delayed time step. (b) is the time series graph representation, where brown and blue circles are the source nodes, (i.e., X_{t-1} and Y_{t-1}), black circle is the target node (i.e., Z_t), brown circle represents the nodes in $P_{C_{X_{t-1} \to Z_t}} \setminus C_{Y_{t-1} \to Z_t}$ and blue circle refers to the nodes in $P_{C_{Y_{t-1} \to Z_t}} \setminus C_{X_{t-1} \to Z_t}$. Also, $P_{Z_t} \setminus (C_{X_{t-1} \to Z_t} \cup C_{Y_{t-1} \to Z_t})$ is empty in this case. Therefore, the condition set $\vec{W}$ is only $\{V_{t-2}\}$.

model in Fig. 3, but with a nonlinear dependency between X and Y such that:

$$V_t = \eta_t^V$$
$$X_t = c_{VX} V_{t-1} + \eta_t^X$$
$$Y_t = c_{VY} V_{t-1} + \eta_t^Y$$
$$Z_t = c_Z X_{t-1} Y_{t-1} + \eta_t^Z. \qquad (24)$$

where c_Z is the coupling coefficient in the function of Z_t.

The MPID and the PID for X_{t-1}, Y_{t-1} and Z_t are estimated numerically for different combinations of c_{VX} and c_{VY}. c_Z and the standard deviations of all the noises (i.e., σ_V, σ_X, σ_Y and σ_Z) are set to 0.5 and 1, respectively. To compute the information-theoretic metrics (e.g., conditional entropy, conditional mutual information), the multivariate probability distribution is required and estimated based on the kernel density estimation method with the multivariate Gaussian kernel [24]. For each combination of c_{VX} and c_{VY}, the sample length is 10,000.

Fig. 4 shows the estimated interaction information and the synergistic, redundant and unique information under both MPID and PID. It can be observed that for MPID, the redundant information, R_c, in the total information, I_{total}, is almost zero for different c_{VX} and c_{VY}, consistent with the conclusion that the redundancy is minimized when *additivity* and *separability* hold. Nevertheless, all
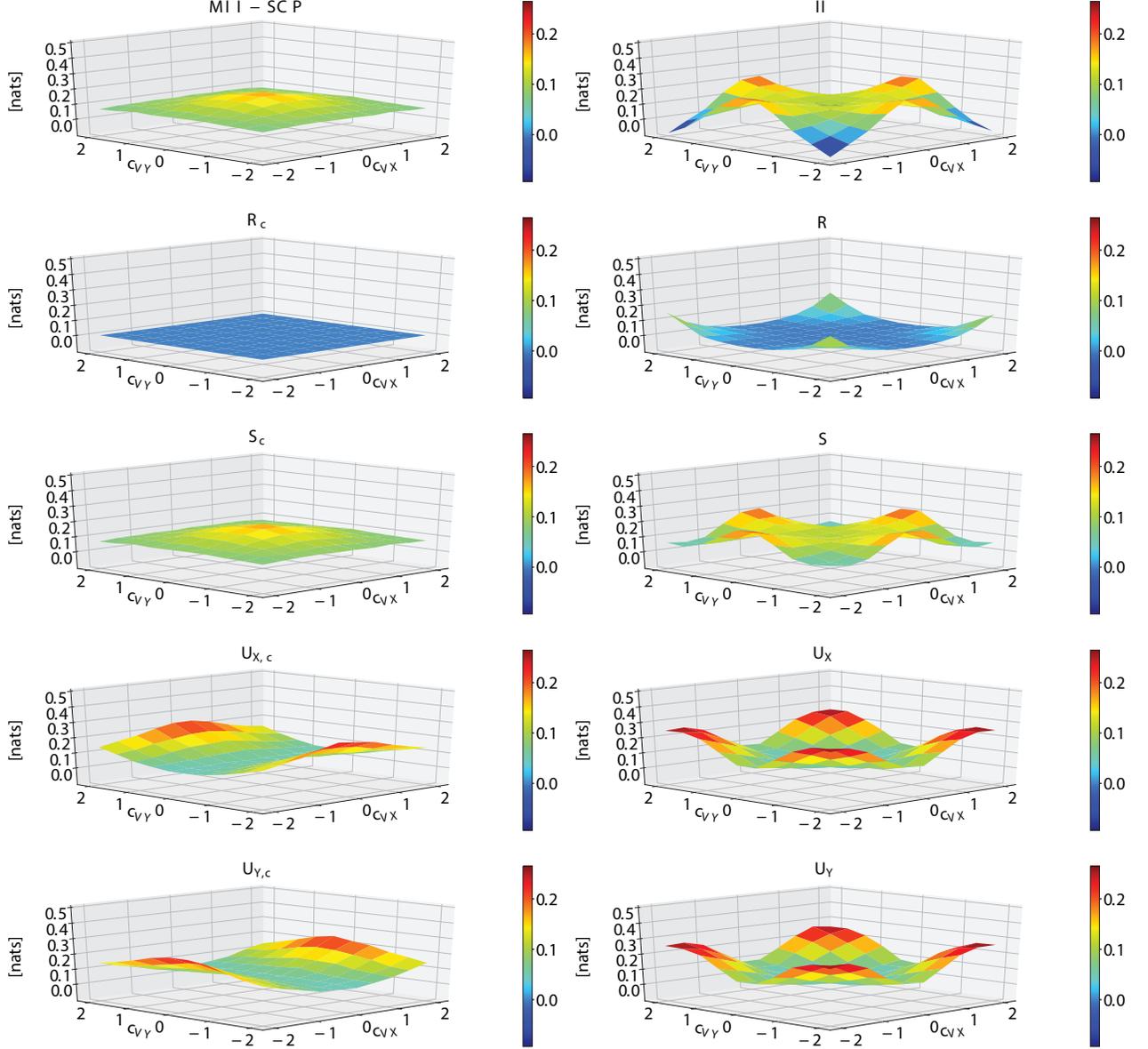
FIG. 4. The estimated interaction information MII-SCP and II as well as SUR (i.e., redundant, synergistic and unique information) from both MPID (the left column) and PID (the right column) for the sources $\{X_{t-1}, Y_{t-1}\}$ and the target Z_t, based on Eq.(24) with a simulation time length 10000 for each combination of c_{VX} and c_{VY}.

the remaining metrics (i.e, MII-SCP, S_c, U_X, U_Y) vary with c_{VX} and c_{VY}. It suggests that the coupling strength autonomy is not entirely valid due to the nonlinearity in Z_t. Especially, both MII-SCP and S_c show higher values when both c_{VX} and c_{VY} are close to zero, illustrating more synergistic information from X_{t-1} and Y_{t-1} with the decrease of the influence from their common driver V_{t-2}.

With regards to PID (shown in the second column in Fig. 4), all the metrics are affected by V through c_{VX} and c_{VY}. The impacts of c_{VX} and c_{VY} on redundancy is illustrated in the increasing redundancy R with the two

coefficients, consistent with the conclusion in the linear case (see Eqs.(22) and (23)). Furthermore, more interaction information II and synergistic information S is also observed for higher values of c_{VX} and c_{VY}.

By comparing the estimated results from MPID and PID in Fig. 4, we can draw the following conclusions. First, the effect from the external driver, V, on the redundancy for MPID (i.e., R_c) is negligible based on the nearly-zero values of R_c. Second, even though MII-SCP, S_c and the two unique information values ($U_{X,c}$ and $U_{Y,c}$) in MPID still depend on c_{VX} and c_{VY}, their variations in terms of the two coefficients (i.e., c_{VX} and c_{VY}) are much

smaller than the corresponding metrics (i.e., II and S) in PID. It suggests that in the nonlinear common driver model, the conditioning in MPID is able to significantly reduce the influences from the common driver V_{t-2} in quantifying the dynamics among X_{t-1}, Y_{t-1} and Z_t.

V. OTHER MOMENTARY PARTIAL INFORMATION DECOMPOSITION FRAMEWORKS

In this study, we adopt the rescaled redundancy measure (Eq.(12)) to define the framework for the momentary partial information decomposition (MPID) in Eq.(13). However, because there is still no universal agreement on an appropriate condition to supplement Eqs.(1)-(3) to provide complete solution for estimating S, R, U_X and U_Y, researchers might prefer to adopt MPID based on alternate approaches, such as those in [10, 11, 16–19]. Therefore, in this section, we provide the frameworks for MPID based upon three other redundancy measures based on: (1) minimum specific information R_{MSI} [10], (2) minimum mutual information R_{MMI} [11], and (3) pointwise common change in surprisal R_{CCS} [17], with the corresponding conditional redundancies represented as $R_{MSI,c}$, $R_{MMI,c}$ and $R_{CCS,c}$ respectively. These conditional redundancy candidates can be used to replace the rescaled redundancy in Eq.(13b) for forming new MPID frameworks. However, except the $R_{MMI,c}$-based MPID, the coupling strength autonomy does not hold for the other two MPID frameworks. This is detailed at the end of this section along with other properties of these alternative MPID frameworks. Despite this, due to the causal Markov property, the proposed MPID frameworks based on various redundancy measurements are still helpful in excluding the influence of the history in quantifying the different information transfer in a causal subgraph, thus revealing the information partitioning from there alternative perspectives.

A. The Minimum Specific Information Approach

Williams and Beer [10] proposed the redundancy measure as the average minimum specific information over the considered input sources. The idea of defining the conditional version $R_{MSI,c}$ is to exclude the influence of the complementary causal subgraph in calculating the specific information for the two source nodes $X_{t-\tau_X}$ and $Y_{t-\tau_Y}$. Therefore, $R_{MSI,c}$ is given by:

$$R_{MSI,c} = \int \min_{A \in \{X_{t-\tau_X}, Y_{t-\tau_Y}\}} \{I(Z_t = z; A \mid \vec{W})\} p(z) dz,$$
(25)

where $I(Z_t = z; A \mid \vec{W})$ is the conditional specific information [10] that $A \in \{X_{t-\tau_X}, Y_{t-\tau_Y}\}$ provides about the outcome $Z_t = z$ conditioned on $\vec{W}$, and can be expressed

as:

$$I(Z = z; A \mid \vec{W}) = \int_a \int_{\vec{w}} p(a, \vec{w}|z) [\ln \frac{p(z|a, \vec{w})}{p(z|\vec{w})}] da \, d\vec{w}.$$
(26)

B. The Minimum Mutual Information Approach

As part of R_c in Eq.(13b), the expression of $R_{MMI,c}$ is given in Eq.(13e).

C. The Pointwise Common Change in Surprisal Approach

Ince [17] characterized the redundancy, R_{CCS}, as the expected pointwise change in surprisal of the target which is common to the sources. The pointwise change in surprisal of the target Z_t is interpreted as the pointwise or local interaction information, whose joint distribution $\tilde{P}$ is estimated from the original joint distribution P based on a game-theoretic approach (see the details in [17]).

R_{CCS} is defined based on the interaction information $II(X_{t-\tau_X}; Y_{t-\tau_Y}; Z_t)$ and its local elements [17]. We develop the corresponding conditional version $R_{CCS,c}$ upon the conditional interaction information $II(X_{t-\tau_X}; Y_{t-\tau_Y}; Z_t \mid \vec{W})$, which is also $\Delta I_{\{X_{t-\tau_X}, Y_{t-\tau_Y}\} \to Z_t}^{MSCP}$ in Eq.(10) given as:

$$II(X_{t-\tau_X}; Y_{t-\tau_Y}; Z_t \mid \vec{W}) = \qquad (27)$$
$$\int_{\vec{w}} \int_x \int_y \int_z p(x, y, z, \vec{w}) i(x; y; z|\vec{w}) dx \, dy \, dz \, d\vec{w},$$

where $i(x; y; z|\vec{w})$ is the local interaction information of $II(X_{t-\tau_X}; Y_{t-\tau_Y}; Z_t \mid \vec{W})$ and can be written as:

$$i(x; y; z|\vec{w}) = \log \frac{p(x, y, z \mid \vec{w}) p(x \mid \vec{w}) p(y \mid \vec{w}) p(z \mid \vec{w})}{p(x, y \mid \vec{w}) p(x, z \mid \vec{w}) p(y, z \mid \vec{w})}$$
(28a)
$$= \Delta_z h(x, y \mid \vec{w}) - \Delta_z h(x \mid \vec{w}) - \Delta_z h(y \mid \vec{w}),$$
(28b)

with the three local individual informations: $\Delta_z h(x, y|\vec{w}) = \log \frac{p(x,y,z|\vec{w})}{p(x,y|\vec{w})p(z|\vec{w})}$, $\Delta_z h(x|\vec{w}) = \log \frac{p(x,z|\vec{w})}{p(x|\vec{w})p(z|\vec{w})}$, and $\Delta_z h(y|\vec{w}) = \log \frac{p(y,z|\vec{w})}{p(y|\vec{w})p(z|\vec{w})}$. Therefore, following the idea that the redundancy is measured with pointwise common change in surprisal, $R_{CCS,c}$ can be defined as the weighted sum of the local conditional interaction information and is given by:

$$R_{CCS,c} = \int_x \int_y \int_z \int_{\vec{w}} \tilde{p}(x, y, z, \vec{w}) \qquad (29)$$
$$\Delta_z h^{com}(x, y \mid \vec{w}) dx \, dy \, dz \, d\vec{w},$$

where

$$\Delta_z h^{com}(x, y \mid \vec{w}) = \qquad\qquad (30)$$
$$\begin{cases} -i_{\tilde{p}}(x; y; z \mid \vec{w}), & \text{if } \operatorname{sgn} \Delta_z h(x, y \mid \vec{w}) = \operatorname{sgn} \Delta_z h(x \mid \vec{w}) \\ & = \operatorname{sgn} \Delta_z h(y \mid \vec{w}) = \operatorname{sgn}[-i_{\tilde{p}}(x; y; z \mid \vec{w})] \\ 0, & \text{otherwise.} \end{cases}$$

Notice that the signs of the three local individual information and the inverse of the local interaction information have to be the same. This is because of the assumption [17] that the positive and negative local information terms are fundamentally different, so that only the local (conditional) interaction information where all three local individual informations have the same signs contributes to the redundancy. Also, $\tilde{p}$ is the element of the maximum entropy distribution $\tilde{P}$ from the original joint distribution P such that:

$$\tilde{P}(X, Y, Z, \vec{W}) = \underset{Q \in \Delta_P}{\arg\max} \int_x \int_y \int_z \int_{\vec{w}} -\tilde{q}(x, y, z, \vec{w})$$
$$\log \tilde{q}(x, y, z \mid \vec{w}) dx \, dy \, dz \, d\vec{w}, \qquad (31)$$

where the set of potential distributions Δ_P is selected from each candidate joint distribution Q, which has the same sources' joint distribution as well as pairwise source-target marginal joint distribution as P, and is given by:

$$\Delta_P = \left\{ Q \in \Delta : \begin{array}{c} Q(X, Z \mid \vec{W}) = P(X, Z \mid \vec{W}), \\ Q(Y, Z \mid \vec{W}) = P(Y, Z \mid \vec{W}), \\ Q(X, Y \mid \vec{W}) = P(X, Y \mid \vec{W}) \end{array} \right\}.$$

The selection of the joint distribution $\tilde{P}$ for $R_{CCS,c}$ follows a similar game-theory approach as R_{CCS} in [17] but is different in that the optimization in Eq.(31) is based on the maximization of the conditional entropy.

D. Properties of the Four Conditional Redundancies

We have provided formulations for MPID based on four different definitions for conditional redundancies as the additional condition for MPID. The four selected redundancies (i.e., R_{MSI}, R_{MMI}, R_{CCS} and R_S) originate from different perspectives. R_{MSI} and R_{MMI} render the redundancy as a function of the marginal distribution between the target and each source, which usually overestimate the redundancy. To overcome that, the rescaled redundancy, R_S, takes the source correlations into account in the redundancy computation. Furthermore, R_{CCS} is developed based on a game-theory perspective and the idea that the redundancy consists of the pointwise common change in surprisal represented as the local interaction information, which essentially utilizes the information of the whole joint distribution.

In terms of the non-negative PID or the corresponding MPID, termed as "local positivity" in [22], only R_{MMI} and R_S allow a non-negative decomposition. The local positive property for PID and MPID is imposed in this study because the non-negative decomposition might be essential in illustrating physical phenomenon in a meaningful way. Despite the fact the other two measures, R_{MSI} and R_{CCS}, cannot guarantee the local positivity of PID and MPID, it should also be noted that the negativity of the PID and MPID induced by R_{CCS} and $R_{CCS,c}$, respectively, is explainable in the context of the definition for R_{CCS} [17].

Furthermore, coupling strength autonomy only holds for the $R_{MMI,c}$ and $R_{S,c}$-based MPID, while the other two conditional redundancy measures do not facilitate this property due to their pointwise or local information based computation. Nevertheless, the definitions for the $R_{MSI,c}$ and $R_{CCS,c}$-based MPIDs are also useful in computing MPIDs with most of the influence from the complementary causal subgraph eliminated owing to causal Markov property of the network.

In addition, when the two causal paths are separable, the conditional redundancy based on the rescaled approach is minimized as $R_{min,c}$ (see Section IV A). The minimization of the redundancy vividly illustrates the separable structure of the two causal paths from the sources, making the rescaled approach (Eq.(12)) suitable in quantifying MPID for separable causal paths based on a causal network.

Despite their pros and cons, the three proposed MPID frameworks in this section would eventually delineate the information transfer along causal paths in a causal network from their own perspectives, and provide for potentially alternative applications.

VI. QUANTIFYING MPID UNDER DIFFERENT CAUSAL PATHS AND CAUSALITY STRUCTURES

Given the four MPID frameworks based on different redundancy measures described above, the MPID using the rescaled redundancy is adopted for further analysis. Specifically, in this section, we aim to investigate how MPID is affected by (1) separable and non-separable causal paths and different causal network structures, and (2) the effect of noise. We use synthetic data generated from two coupled logistic equation models for both separable and non-separable causal paths.

A. Coupled Logistic Equations

The two models, each of which involves three variables, are given as:

(1) A coupled logistic equation model without self-

(a) Case SI – time series graph

(b) Case SI – PID & MPID

(c) Case NI – time series graph

(d) Case NI – PID & MPID
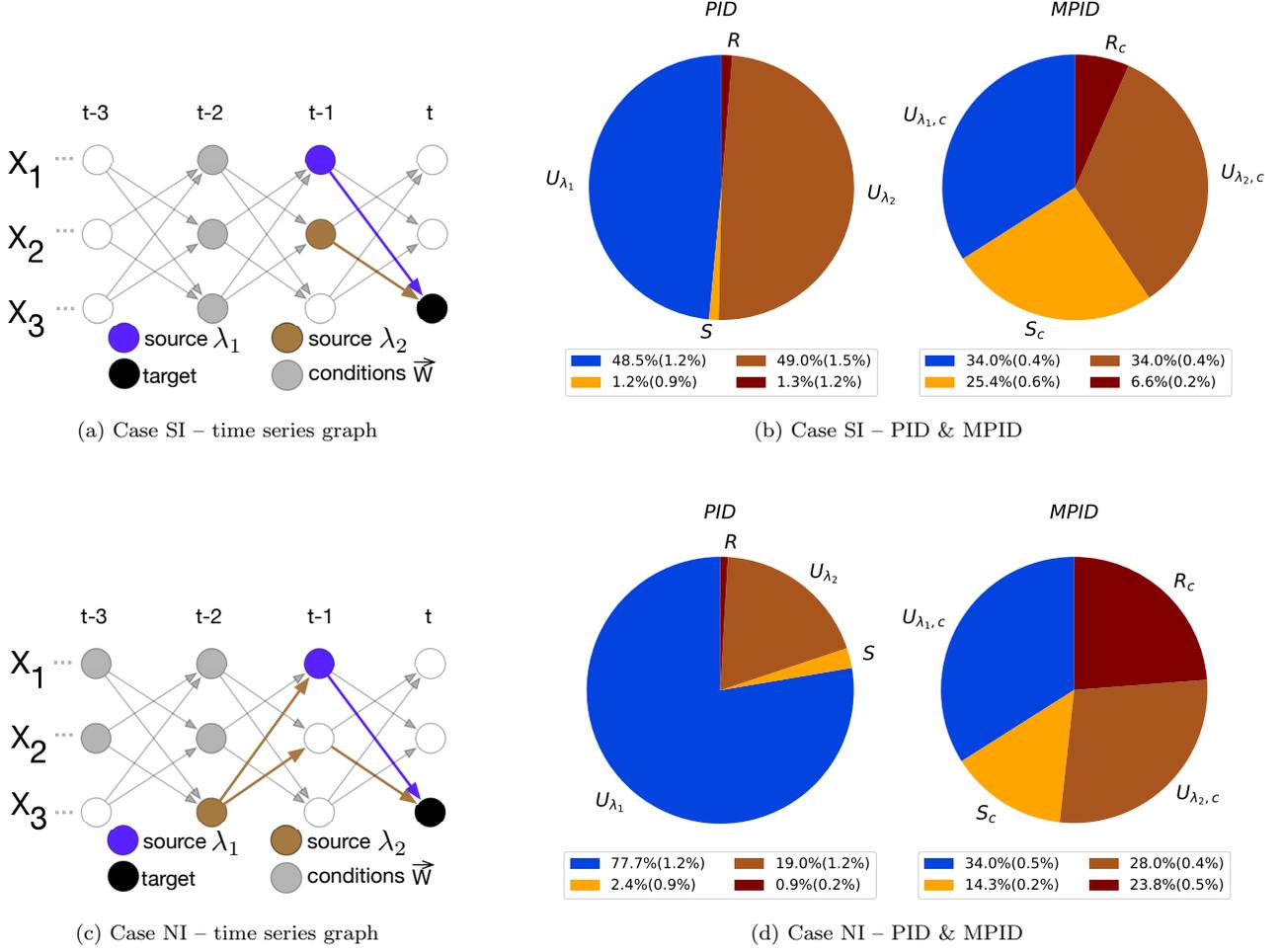
FIG. 5. Illustration of the time series graphs and the averaged information partitioning for the two cases in Eq.(32) (i.e., Case SI and Case NI) when the coupling strength ϵ equals to 0.5. (a) and (c) are the time series graphs of Cases SI and NI respectively (blue and brown nodes: the source nodes; black node: the target node; blue and brown directed links: the links in the causal paths of the two sources; grey solid nodes: the conditioning nodes in $\vec{W}$ for MPID in Eq.(10)). (b) and (d) are the pie charts of the averaged synergistic, redundant and unique information percentages from PID and MPID of Cases SI and NI respectively, with a legend showing the averaged values along with the corresponding standard deviations in the parenthesis (U_{λ_1} and $U_{\lambda_1,c}$: the unique information of $X_{1,t-1}$; and U_{λ_2} and $U_{\lambda_2,c}$: the unique information of $X_{2,t-1}$).

dependency:

$$X_{i,t} = \frac{1-\epsilon}{2} \sum_{\substack{j=0 \\ j \neq i}}^{3} 4 X_{j,t-1}(1 - X_{j,t-1}) + \epsilon \eta_t^{X_i}, \qquad (32)$$

where $\eta_t^{X_i}$ is a standard uniform noise $\sim U(0,1)$, and ϵ is the noise coupling strength ranging from 0 to 1.

(2) A fully coupled logistic equation model given as

$$X_{i,t} = \frac{1-\epsilon}{3} \sum_{j=1}^{3} 4 X_{j,t-1}(1 - X_{j,t-1}) + \epsilon \eta_t^{X_i}. \qquad (33)$$

For each model (Eqs.(32) and (33)), the following two situations of a causal subgraph, comprising the pathways

of two sources affecting a target node, are considered for computing their MPID and PID:

- Scenario 1 (separable causal paths): the sources $\{X_{1,t-1}, X_{2,t-1}\}$ and the target $X_{3,t}$;

- Scenario 2 (non-separable causal paths): the sources $\{X_{1,t-1}, X_{3,t-2}\}$ and the target $X_{3,t}$.

For convenience, we name the four cases as follows:

- Case SI: Scenario 1 (separable) in Eq.(32) (without self-dependency);

- Case NI: Scenario 2 (non-separable) in Eq.(32) (without self-dependency);

- Case SC: Scenario 1 (separable) in Eq.(33) (fully coupled);

(a) Case SC – time series graph

(b) Case SC – PID & MPID

(c) Case NC – time series graph

(d) Case NC – PID & MPID
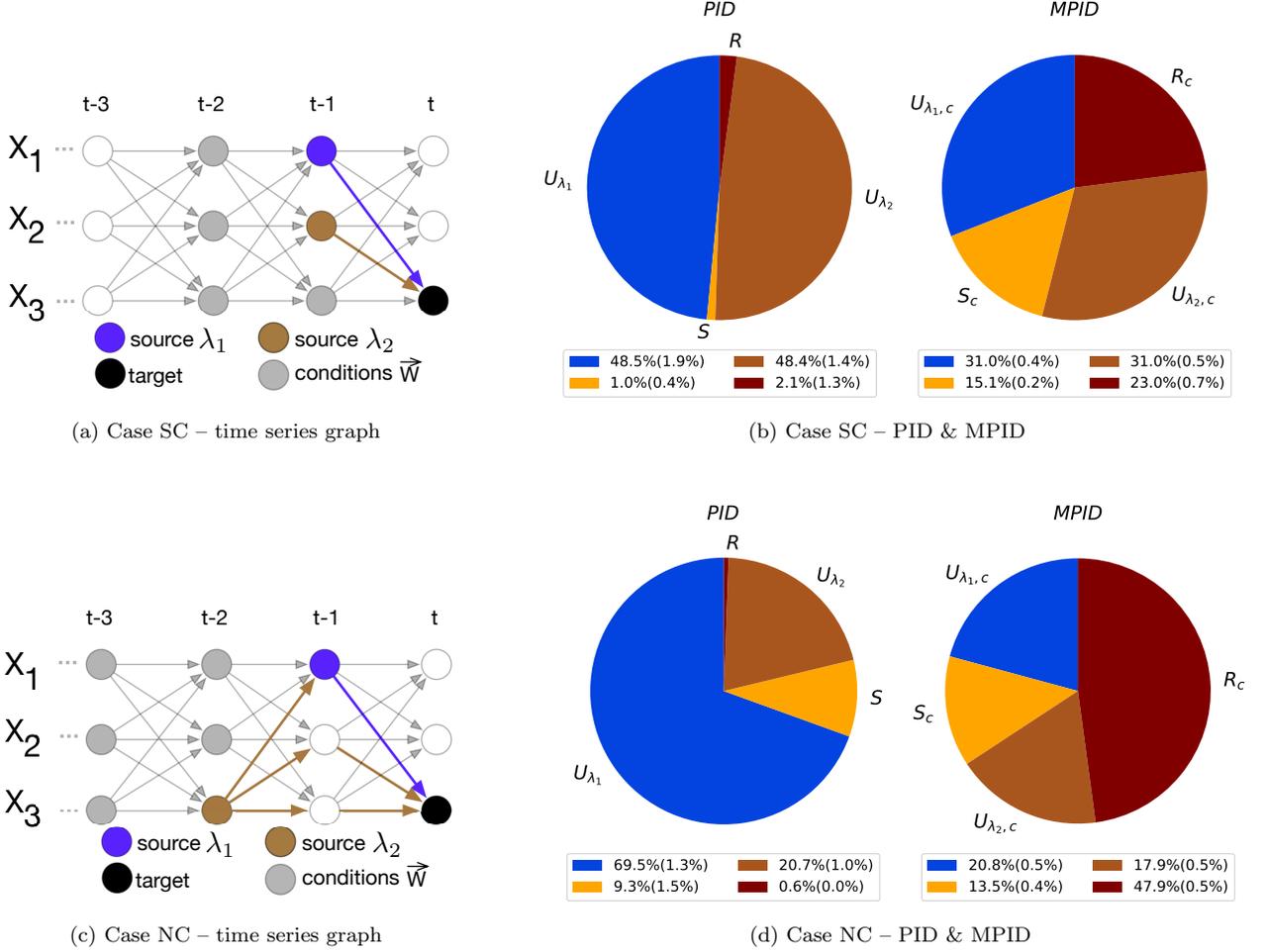
FIG. 6. Illustration of the time series graphs and the averaged information partitioning for the two cases in Eq.(33) (i.e., Case SC and Case NC) when the coupling strength ϵ equals to 0.5. (a) and (c) are the time series graphs of Cases SC and NC respectively (blue and brown nodes: the source nodes; black node: the target node; blue and brown directed links: the links in the causal paths of the two sources; grey solid nodes: the conditioning nodes in $\vec{W}$ for MPID in Eq.(10)). (b) and (d) are the pie charts of the averaged synergistic, redundant and unique information percentages from PID and MPID of Cases SC and NC respectively, with a legend showing the averaged values along with the corresponding standard deviations in the parenthesis (U_{λ_1} and $U_{\lambda_1,c}$: the unique information of $X_{1,t-1}$; and U_{λ_2} and $U_{\lambda_2,c}$: the unique information of $X_{2,t-1}$).

- Case NC: Scenario 2 (non-separable) in Eq.(33) (fully coupled).

Figs. 5a, 5c, 6a and 6c show the time series graphs of the two models and the causal subgraphs for the four cases. For both scenarios, $X_{1,t-1}$ is considered as the first source whose unique information are represented by U_{λ_1} and $U_{\lambda_1,c}$ for PID and MPID, respectively. $X_{2,t-1}$ and $X_{3,t-2}$ are taken as the second source for Scenario 1 and Scenario 2, respectively, with unique information represented as U_{λ_2} and $U_{\lambda_2,c}$ for PID and MPID, respectively.

B. Simulation Setting

We change the noise coupling strength ϵ uniformly in 19 intervals between 0 and 1 to control the influence of the noise on the two logistic models Eqs.(32) and (33) for generating synthetic data for the four cases. For each ϵ, 10,000 data points are generated for computing both MPID and PID in each case. To get an averaged behavior, ensembles of 10 trajectories are conducted for each ϵ. We first investigate the general influences of the separability of the causal paths and the network structure on MPID and PID when the coupling strength ϵ equals to 0.5. Then, we expand the analysis to explore how the noise strength shapes MPID and PID.

C. Influence of the Separability of the Causal Paths and the Network Structure on MPID/PID

The average percentages of the redundancy (i.e., R and R_c), the synergy (i.e., S and S_c) and the unique information (i.e., U_{λ_1}, $U_{\lambda_1,c}$, U_{λ_2} and $U_{\lambda_2,c}$) of both MPID and PID from an ensemble of 10 runs with $\epsilon = 0.5$ are shown in Figs. 5b, 5d, 6b and 6d. The standard deviations of MPID and PID are depicted in the parenthesis in the legends.

Comparison between PID and MPID: For all the four cases, the results of MPID and PID are different. Especially, in Case SI, Fig. 5b shows that there is almost no redundancy R and little synergy S from PID, while more synergy S_c (~25.4%) is observed in MPID. This is explained by the fact that $X_{3,t}$ is entirely determined by $X_{1,t-1}$ and $X_{2,t-1}$ according to Eq.(32), thus in MPID the exclusion of the history, that is, the complementary causal subgraph, through conditioning on $\vec{W}$ results in higher synergistic information. Furthermore, in Cases NI and NC, where the two causal paths are non-separable (Scenario 2), PID reveals near-zero redundancy of the two sources and significant unique information (69% ~78%) from $X_{1,t-1}$ (Figs. 5d and 6d). However, MPID (Cases NI and NC) indicates that unique information percentages of the sources ($U_{\lambda_1,c}$ and $U_{\lambda_2,c}$) are much closer and the proportion of the redundancy, R_c, is larger than that of PID. This is because the non-separable causal paths (Scenario 2) allows the information transfer of the source $X_{1,t-1}$ to be a part of the information transfer of the other source $X_{3,t-2}$, leading to the higher redundancy. Also, almost identical unique information of the two sources in MPID (with $U_{\lambda_1,c}$ slightly larger) for Cases NI and NC results from the balance between the information transfer from the direct driver $X_{1,t-1}$ with one pathway and the indirect driver $X_{3,t-2}$ with multiple pathways. Even though $X_{3,t-2}$ affects the target $X_{3,t}$ indirectly, there are two and three pathways towards the target for Cases NI and NC, respectively, enhancing its unique information contribution to the target.

Different causal networks: Let's first compare the information partitioning results of Scenario 1 – the separable causal paths (i.e. Cases SI and NI shown in Figs. 5b and 6b, respectively). The synergy of the MPID in Case SI (i.e., S_c) is larger than that of Case NI. As discussed in the previous paragraph, the higher percentage of S_c in Case SI arises because the target $X_{3,t}$ is determined entirely by the two sources (i.e., $X_{1,t-1}$ and $X_{2,t-1}$) based on the causal relationship in Eq.(32). Nevertheless, the full coupling in Eq.(33) enables the target $X_{3,t}$ to be dictated by three nodes (i.e., $X_{1,t-1}$, $X_{2,t-1}$ and $X_{3,t-1}$). Therefore, for MPID in Case SC, with the knowledge of only two nodes, it is not sufficient to provide the synergistic information S_c to the target, resulting in a less synergy S_c compared with Case SI. In addition, in the comparison between Cases NI and NC, of which the two causal paths are not separable, the information partitioning patterns of both cases are consistent with

each other. For example, the PIDs of both cases shows a strong unique information U_{λ_1} from X_{t-1}. However, the differences in the causal network quantitatively results in the differences in the information partitioning result. For instance, in terms of MPID, the increased redundant information R_c in Case NC is higher than that of Case SC.

Separable and non-separable causal paths: For each model, we compare the MPID results of the separable and non-separable causal paths. The comparisons (i.e., Case SI versus NI, and Case SC versus NC) reveal the same behavior in the two coupled logistic models that more redundancy in the cases of non-separable causal paths (i.e., Cases NI and NC) is estimated than that of the separable ones (i.e., Cases SI and SC). This is intuitive because in both Case NI and NC, the source $X_{1,t-1}$ lies in the causal path from the other source $X_{3,t-2}$ to the target $X_{3,t}$, thereby a part of the information transfer from $X_{3,t-2}$ to $X_{3,t}$ is overlapped or contributed by the information given by $X_{1,t-1}$, resulting in more redundancy than that of the separable causal paths in Cases SI and SC.

D. Influence of Noise

To understand how additive noise affects the estimation of MPID and PID, we plot these estimates as a function of signal-to-noise ratio (SNR) which is computed as the ratio of the variance of the logistic time series to the variance of the noise terms in Eqs.(32)-(33). Furthermore, it is well known that the coupled logistic equations are prone to synchronize depending on the lags and noise strength [25–29], which impacts the MPID and PID outcomes. To investigate whether the lagged synchronization occurs and, thus, affects MPID and PID, we also plot the similarity functions among the sources and target as well as the percentages of total information given by the two sources in the overall uncertainty for both MPID and PID over SNR in Fig. 7, which are defined as follows:

$$SF_{ij}(\tau) = \left\{ \frac{E[(X_{i,t+\tau} - X_{j,t})^2]}{[E(X_{i,t+\tau}^2)E(X_{j,t}^2)]^{1/2}} \right\}^{0.5} \qquad (34)$$

$$I_{tot,p} = \frac{I(\lambda_1, \lambda_2; X_{tar})}{H(\lambda_1, \lambda_2, X_{tar})} \times 100\% \qquad (35)$$

$$I_{tot,p,c} = \frac{I(\lambda_1, \lambda_2; X_{tar} \mid \vec{W})}{H(\lambda_1, \lambda_2, X_{tar} \mid \vec{W})} \times 100\% \qquad (36)$$

where $SF_{ij}(\tau)$ is the similarity function between $X_{i,t+\tau}$ and X_j, E is the expectation function, $H(\bullet)$ and $H(\bullet \mid \bullet)$ are the Shannon's entropy and the corresponding conditional entropy [14], respectively, $\{\lambda_1, \lambda_2, X_{tar}\}$ represents the two sources and the target variables in the four cases, and $I_{tot,p}$ and $I_{tot,p,c}$ are the percentages of shared information between sources and target for PID and MPID, respectively. The similarity function $SF_{ij}(\tau)$ describes

the degree of synchronization between two variables X_i and X_j with a lag τ. A lower value $SF_{ij}(\tau)$ means a higher similarity between X_i and X_j with lag τ with value 0 representing a complete synchronization. The normalized total information $I_{tot,p}$ and $I_{tot,p,c}$ accounts for the amount of uncertainty reduced by the two sources λ_1 and λ_2 for the target X_{tar} against their overall uncertainty, with and without conditioning on $\vec{W}$, respectively.

Influence of SNR on $I_{tot,p}$ and $I_{tot,p,c}$: The percentages of total information behave differently for PID and MPID. For PID, $I_{tot,p}$ (the red lines in the first column of Fig. 7) increases with SNR. This is because with the decrease in the noise, the sources have a stronger control on the target, resulting in high $I_{tot,p}$. Meanwhile, for MPID, $I_{tot,p,c}$ (the red lines in the second column of Fig. 7) increases gradually until SNR is between 1 to 50. Note that higher SNR is achieved by reducing the noise coupling strength ϵ in Eqs.(32) and (33). With further increase in SNR rapidly, $I_{tot,p,c}$ initially drops down and then flattens (for Case SC) or even increases a bit (for the other cases). The decrease of $I_{tot,p,c}$ for MPID in the range $1 <$ SNR < 50 can be explained by two factors. First, the dynamics tend to be less stochastic in higher SNR and thus the condition set $\vec{W}$ is able to explain more about the target. Note that in a deterministic model, the target variable can be determined entirely by the knowledge of the condition set $\vec{W}$. Therefore, the enhanced contribution of $\vec{W}$ leads to the decline of $I_{tot,p,c}$ under high SNR. Second, the synchronization rate between each source and the target decreases when SNR becomes larger than around 1, shown as the flattened black lines with marked symbols in the second column of Fig. 7. Moreover, with the further growth of SNR ($\gg 1$), the target starts to be desynchronized with each source, reflected as the increase of the similarity function between the two in each case. The change of synchronization rate accounts for not only the decrease of $I_{tot,p,c}$ when SNR is slightly larger than 1, but also the increase of $I_{tot,p,c}$ for MPID when the system is much less stochastic for large SNR.

Influence of lagged synchronization on $I_{tot,p}$ and $I_{tot,p,c}$: The three similarity functions among the two sources and the target are plotted as the three types of black lines in the first and second columns of Fig. 7 for PID and MPID, respectively. It can be observed that when SNR is less than 1 (or the noise coupling strength ϵ is large than around 0.5), the decrease of the three similarity functions captures with the increases of both $I_{tot,p}$ and $I_{tot,p,c}$. It implies the synchronization between each source and the target with the decrease of the noise effect results in an increased proportion of uncertainty reduction in the target given the knowledge of the two sources. Moreover, $I_{tot,p,c}$ decreases significantly once SNR is in the range of 1 to 50. As explained at the end of the previous paragraph, this is due to (1) the increased information provided by the condition set $\vec{W}$ for the target in a less stochastic system, and (2) the desynchronization trend between each source and the target, which

is shown as the growth of the similarity functions when SNR is larger than around 1. However, when SNR is large, that is, the system is weakly stochastic, $I_{tot,p,c}$ flattens for Case SC and goes up again for the other three cases, while at the same time, the increasing rate of the similarity functions between each source and the target declines. The decline in the increasing rate of the similarity functions reflects the slowdown of the desynchronization process between each source and the target, causing the flattening and growth of $I_{tot,p,c}$, for Case SC and the other three cases, respectively. Especially, for Cases NI and NC where the two causal paths are nonseparable, the increase of $I_{tot,p,c}$ in MPID for high SNR (close to 100) is mainly due to the reduction in the rate of growth of the similarity function between the target and the first source, whose unique information, $U_{\lambda_1,c}$, (the blue dashed lines in the fourth column of Fig. 7) contributes to most of $I_{tot,p,c}$.

Influence of SNR on PID: For all the four cases, the synergy S (the orange lines in the third column in Fig. 7) is almost zero with a little increase for high SNR, while the redundancy R (the black lines in the third column in Fig. 7) is close to zero when SNR is less than around 1 but increases significantly for high values of SNR. The near-zero values of both S and R for SNR less than 1 are due to the dominant role of the noise in the system such that the two sources provide little information to the target, which is also manifested as the near-zero values of the corresponding $I_{tot,p}$ (the red lines in the third column of Fig. 7). Meanwhile, the significant increase of R for larger SNR (> 1) is due to the symmetric structure of the coupled logistic equations, whose influence is more significant in larger SNR where the system is less stochastic, resulting in a higher value of redundancy. In terms of the two unique information, for Cases SI and SC, U_{λ_1} (the blue dashed lines) and U_{λ_2} (the sienna dashed lines) are almost identical for both sources which influence the target directly through the same logistic equation. For Cases NI and NC, the two unique information are almost zero for small SNR because of the dominant role of the noise, and start to increase with the growth of SNR with a faster increasing rate for U_{λ_1}. That U_{λ_1} is larger than U_{λ_2} when the system is less stochastic for a higher SNR illustrates the fact that the first source $X_{1,t-1}$ is a direct cause of the target $X_{3,t}$ while the second source $X_{3,t-2}$ influences the target indirectly.

Influence of SNR on MPID: For Case SI, the redundancy R_c (the orange lines in the fourth column in Fig. 7) and the synergy S_c (the black lines in the fourth column in Fig. 7) increases and decreases with SNR, respectively. Especially, S_c is much larger than R_c for SNR < 1, illustrating the fact that the target is entirely determined by the two sources, which is not shown in the corresponding PID plot in Case SI. Meanwhile, with the increase of SNR, R_c increases rapidly while S_c decreases, as a result of the symmetric structure of the coupled logistic equations, which leads to higher redundancy. For the other three cases, both S_c and R_c are much larger than zero
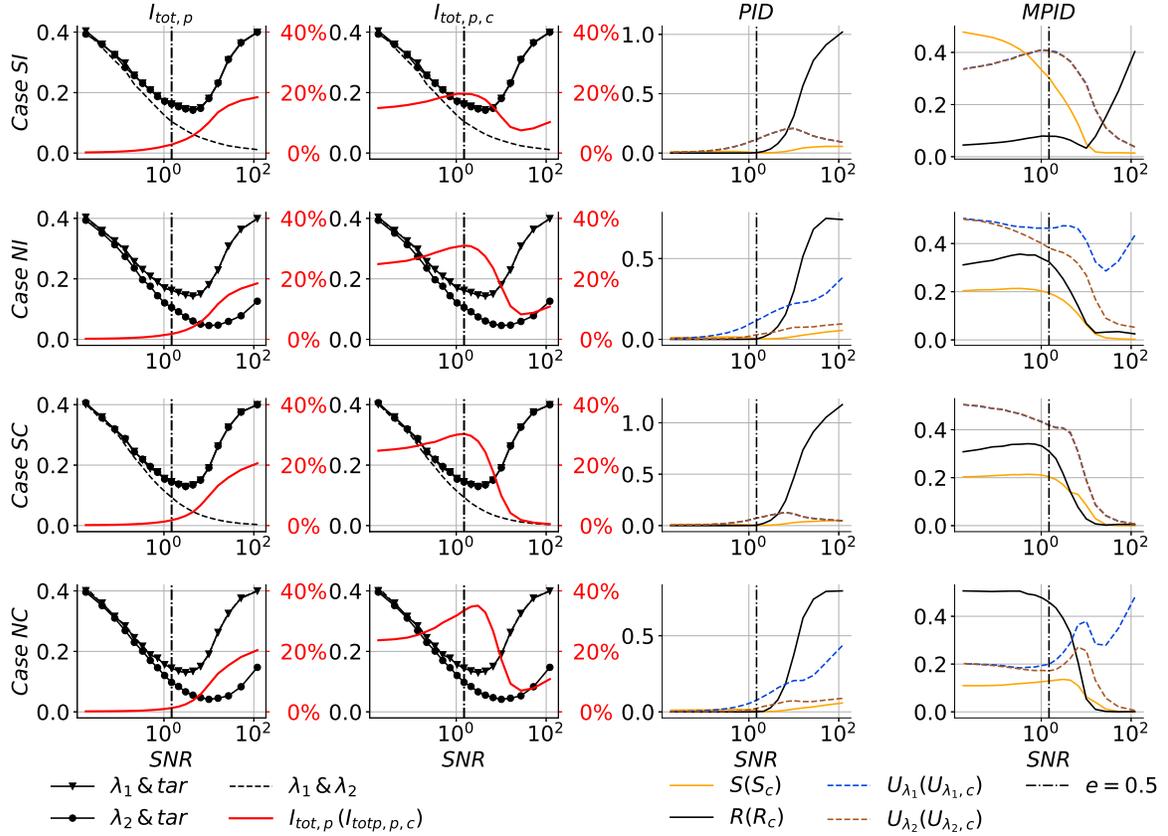
FIG. 7. The averaged percentages of total information, PID, and MPID for the four cases based on the two coupled logistic equations in Eqs.(32)-(33) in terms of different signal-to-noise ratio (SNR). The first and second columns plot the percentages of total information in Eqs.(35)-(36) (the red line) and the similarity functions in Eq.(34) between the first source and the target, λ_1 & *tar* (the black line marked with triangles), the second source and the target, λ_2 & *tar* (the black line marked with solid cirles), and the two sources, λ_1 & λ_2 (the black dashed line), for PID and MPID, respectively. The first and second columns plot the synergistic (the orange line), the redundant (the black line), and the two unique information (the blue and sienna dashed lines) for PID and MPID, respectively.

with $R_c > S_c$ for SNR < 1, and both start to decrease with SNR and reach around zero when SNR is close to 100. In the condition of a smaller SNR (< 1), a higher R_c results from (1) the non-separable causal paths in Cases of NI and NC, which results in an overlapped information transfer from the two sources, and (2) the fact that the target is determined by three causes in Case SC so that the synergistic information S_c given by two sources is not large enough. In the case of a larger SNR (> 1), S_c and R_c decline because of the the decreased total information $I_{tot,p,c}$. With regards to the two unique information, $U_{\lambda_1,c}$ and $U_{\lambda_2,c}$ are almost identical for the cases with separable causal paths (i.e., Cases SI and SC) since that both sources impact the target directly through the same logistic equation. For the other cases with non-separable causal path (i.e., Cases NI and NC), the two unique information are close for a noise-dominant system with a smaller SNR and start to diverge with a higher $U_{\lambda_1,c}$ when the system becomes less stochastic with a

larger SNR. The increasing difference between the two unique information with SNR illustrates that in a more deterministic system, compared with the second source $X_{3,t-2}$ indirectly controlling the target $X_{3,t}$, the first source $X_{1,t-1}$ is able to provide more information to the target through its direct influence on $X_{3,t}$.

VII. SUMMARY AND CONCLUSIONS

By employing a time series graph-based approach, where the dynamics among the components at each time step are explicitly represented, we propose the *momentary partial information decomposition* (MPID). It allows us to dissect the information transfer to a target through a preceding causal subgraph, which comprises multiple causal paths from multiple sources to the target, into synergistic, redundant and unique information.

Different from the original partial information decomposition (PID), whose quantification includes the information from the entire history, MPID is able to exclude the influence from this history or the complementary causal subgraph, through conditioning, for any direct cause of the subgraph of interest in the estimation. PID and MPID together provide two ways for investigating information partitioning (with and without the influence from the complementary causal subgraph), and the comparison between them draws out different behaviors of a process network.

The adopted rescaled method for estimating redundancy (Eq.(12)) in information partitioning [12] proves to be effective in MPID in excluding the influence of the complementary causal subgraph. For instance, when the two causal paths are separable, meaning that neither of the sources belongs to the causal path of the other, the redundant information R_c in MPID is minimized. It makes sense that when two sources influence the target through different causal paths, the redundancy is reduced. Also, because there is no universal agreement on the appropriate PID method, we provide the estimations for the MPID frameworks based on the three alternative redundancy measures in Section V.

Further, we investigate MPID and PID of a three-node dynamics under different causality structures and both separable and non-separable causal paths as well as the noise effect on the information partitioning. Application of PID and MPID for two coupled logistic equation models shows that compared with separable causal paths, non-separable causal paths generate more redundant information since the two sources have overlap in their causal paths towards the target. Also, the difference in causality structure gives rise to different MPID results. For instance, under two separable causal paths, more synergistic information is observed, when the target is entirely controlled by the two sources in the model without self-dependency, than the fully-coupled model where the target is also driven by an external node. Furthermore, the influence of noise on PID and especially MPID is more complex. In the two coupled logistic equations, when the system is noise dominant (low signal-to-noise ratio (SNR)), the decline of a strong noise influence is able to enhance the lag synchronization between each source and the target, which results in the growth of the total information given by the two sources towards the target for both PID and MPID (i.e., $I_{tot,p}$ and $I_{tot,p,c}$). Meanwhile, in a weak stochastic system (high signal-to-noise ratio), $I_{tot,p,c}$ might either decrease due to a higher proportion of information explained by the condition set $\vec{W}$ or increase because of the decrease of the desynchronization rate between the sources and the target. In short, the influence of noise on the estimates of PID and MPID is determined by the stochastic degree of the system as well as the causality structures of both the system and the causal subgraph of interest. Also, it is noted that the empirical results of MPID in the coupled logistic equations, which adopts the conditional rescaled redundancy measure, may differ if other conditional redundancy measures, such as those proposed in Section V, were used.

Although the momentary information approach is able to exclude most of the influences from the complementary causal subgraph of interest, as pointed out by Runge [2, 6], the time series graph-based approach has the following limitations. First, the coupling strength autonomy property is only analytically established when both *linearity* and *additivity* hold, and not guaranteed for nonlinear cases. However, in some cases, such as the nonlinear common driver model, this *momentary* approach can still significantly reduce the impact from the history and thus better reveal the internal dynamics among the nodes of interest in terms of information partitioning. Furthermore, the Markovian conditional independence property of the graph facilitates an approximate estimation of an automonous information partitioning. Therefore, when the functional dynamics of a complex system are unknown, MPID can provide at least a general picture of the autonomous mechanism in a causal subgraph. Second, the estimation of a high-dimensional probability distribution function, resulting from the potentially many external drivers (i.e., $\vec{W}$), requires a large amount of data, and would potentially result in biased estimation of cause-effect relationships for short datasets.

Finally, the proposed momentary partial information decomposition, which is a Granger causality-oriented framework, provides a new perspective in exploring complex systems, especially in natural systems where systems are complex, self-organized, and hard to be intervened. With the increasing availability of observational data recorded in finer resolutions [30, 31], a lot of investigations based on different data analysis approaches have been conducted in understanding the dynamics of different aspects in nature, such as exploring the self-organization in various ecohydrological system by using transfer entropy [32], the quantification of the strength and delay in climatic interaction through the causal network [33], etc. The proposed framework, anchored on the information partitioning of a causal subgraph, has the potential to enable the investigation of the dynamics of multiple lagged components in terms of different types of information transfer from the sources.

APPENDIX

In this appendix, some important notations used throughout this paper are listed. In addition, we provide the proof of the coupling strength autonomy property for MPID shown in Sect. IV as well as the derivation for the analytical solutions of the linear common driver model in Eq.(20). Some properties of Shannon entropy and mutual information are also given to facilitate the proof.

Appendix A: List of important notations

ABBREVIATIONS

MII	momentary interaction information
MII-SCP	momentary interaction information for separable causal paths
MIT	momentary information transfer
MITP	momentary information transfer along causal paths
MPID	momentary partial information decomposition
PID	partial information decomposition

KEY SYMBOLS

ϵ	noise coupling strength in Eqs. (32) and (33)
$\eta_{t-\tau}^K$	noise of the node $K_{t-\tau}$
λ_1, λ_2	the first and second source nodes, respectively
$\wedge$	logical AND
$\Delta I^M_{X_{t-\tau_X} \to Z_t \mid Y_{t-\tau_Y}}$	MII of $Y_{t-\tau_Y}$ on $C_{X_{t-\tau_X} \to Z_t}$
$\Delta I^{MSCP}_{\{X_{t-\tau_X}, Y_{t-\tau_2}\} \to Z_t}$	MII-SCP of the sources $X_{t-\tau_X}$ and $Y_{t-\tau_2}$ and the target Z_t
$f_K()$	function of the dependencies $P^{\vec{B}}_{K_{t-\tau}}$ for $K_{t-\tau}$
$\tilde{f}_K()$	linear combinations of all the noise terms in $\vec{B}$ for $K_{t-\tau}$
$g_K()$	function of the dependencies $P_{K_{t-\tau}} \backslash P^{\vec{B}}_{K_{t-\tau}}$ for $K_{t-\tau}$
$\vec{B}$	union of two causal paths $P(C_{X_{t-\tau_X} \to Z_t})$ and $P(C_{Y_{t-\tau_Y} \to Z_t})$ and the target node Z_t
$C_{X_{t-\tau} \to Z_t}$	causal path from $X_{t-\tau}$ to Z_t
G	time series graph
$H(X_t)$	Shannon entropy of X_t
$I(X_{t-\tau_X}; Y_{t-\tau_Y})$	mutual information between $X_{t-\tau_X}$ and $Y_{t-\tau_Y}$
$I(X_{t-\tau_X}, Y_{t-\tau_Y}; Z_t)$	mutual information between Z_t and the union of $X_{t-\tau_X}$ and $Y_{t-\tau_Y}$
$I(X_{t-\tau_X}; Y_{t-\tau_Y} \mid Z_t)$	conditional mutual information between $X_{t-\tau_X}$ and $Y_{t-\tau_Y}$ given Z_t
$I^{MIT}_{X_{t-\tau} \to Z_t}$	MIT from $X_{t-\tau}$ to Z_t
$I^{MITP}_{X_{t-\tau} \to Z_t}$	MITP from $X_{t-\tau}$ to Z_t
$I_{tot,p}$	percentage-wise total information for PID
$I_{tot,p,c}$	percentage-wise total information for MPID
$II(X_{t-\tau_X}; Y_{t-\tau_Y}; Z_t)$	interaction information among $X_{t-\tau_X}$, $Y_{t-\tau_Y}$ and Z_t
$II(X_{t-\tau_X}; Y; Z \mid \vec{W})$	conditional interaction information among $X_{t-\tau_X}$, $Y_{t-\tau_Y}$ and Z_t given $\vec{W}$
P_{X_t}	parents of the node X_t
$P_{C_{X_{t-\tau} \to Z_t}}$	parents of the causal path from $X_{t-\tau}$ to Z_t
$P^{\vec{B}}_{X_{t-\tau}}$	union set between $P_{X_{t-\tau}}$ and $\vec{B}$
P	the original joint distribution of the sources and the target
$\tilde{P}$	the maximum entropy distribution estimated from Eq.(31)
Q	the candidate distribution for $\tilde{P}$
R	redundant information in PID
R_S	the redundancy measure based on the rescaled approach
R_{MSI}	the redundancy measure based on the minimum specific information approach
R_{MMI}	the redundancy measure based on the minimum mutual information approach
R_{CCS}	the redundancy measure based on the pointwise common change in surprisal approach
R_c	redundant information in MPID
S	synergistic information in PID
S_c	synergistic information in MPID
$SF_{ij}(\tau)$	similarity function between $X_{i,t+\tau}$ and $X_{j,t}$
U	unique information in PID
U_c	unique information in MPID
$\vec{W}$	condition set for MPID
X	subprocess X
X_t	node representing a subprocess X at a specific time t
$\vec{X}_t$	all the nodes at time step t
$\vec{X}_t^-$	all the nodes at time steps preceding t
$\vec{X}$	multivariate process

Appendix B: Some properties of Shannon entropy and mutual information

The conditional Shannon entropy holds the translationally invariant property [2] such that

$$
\begin{aligned}
H(X &+ f(W) \mid V, W) \\
&= \int p(w) H(X + f(w) \mid V, W = w) dw \\
&= \int p(w) H(X \mid V, W = w) dw \\
&= H(X \mid V, W),
\end{aligned} \tag{B1}
$$

where X, V and W are random variables and f is an arbitrary function. Such translational invariance is also valid for the conditional mutual information and can be given as

$$
\begin{aligned}
I(X + & f(W); Y | V, W) \\
&= H(X + f(W)|V, W) - H(X + f(W)|Y, V, W) \\
&\overset{\text{Eq.(B1)}}{=} H(X|V, W) - H(X|Y, V, W) \\
&= I(X; Y|V, W),
\end{aligned} \tag{B2}
$$

where Y is a random variable. Moreover, the translational invariance for the conditional mutual information can be generalized as

$$
I(X + f(W); Y + g(V)|V, W) = I(X; Y|V, W), \tag{B3}
$$

where g is also an arbitrary function. The proof of Eq.(B3) is similar to the proof for $I(X + f(W); Y|V, W) = I(X; Y|V, W)$ such that we emphasize the translational invariance of the conditions V this time. Another way for generalizing the translational invariance for the conditional mutual information is

$$
\begin{aligned}
I(X + & f(W), Z + h(W); Y | V, W) \\
&= H(X + f(W), Z + h(W)|V, W) - \\
&\quad\ H(X + f(W), Z + h(W)|Y, V, W) \\
&\overset{\text{Eq.(B1)}}{=} H(X, Z|V, W) - H(X, Z|Y, V, W) \\
&= I(X, Z; Y|V, W),
\end{aligned} \tag{B4}
$$

where Z and h are a random variable and an arbitrary function, respectively.

Moreover, if all the variables are Gaussian, the entropy of a d-dimensional process $\vec{X}$ conditional on another multivariate Gaussian process $\vec{Y}$ can be expressed as [14]

$$
H(\vec{X} \mid \vec{Y}) = \frac{1}{2} \ln \left\{ (2\pi e)^d \frac{|\Gamma_{\vec{X}, \vec{Y}}|}{|\Gamma_{\vec{Y}}|} \right\}, \tag{B5}
$$

where $|\Gamma_{\vec{Y}}|$ and $|\Gamma_{\vec{X}, \vec{Y}}|$ are the determinant of the covariance matrix of $\vec{Y}$ and $(\vec{X}, \vec{Y})$, respectively.

Appendix C: Proof of the coupling strength autonomy property for MPID

To prove the coupling strength autonomy property for MPID shown in Section IV A, we start from the derivation for MII-SCP and then give the solutions for the synergistic, unique and redundant information. We denote $\vec{W}$ as the condition adopted in MPID such that

$$
\vec{W} = \vec{W}_1 \cup \vec{W}_2 \cup \vec{W}_3 \tag{C1a}
$$

where

$$
\vec{W}_1 = P_{Z_t} \backslash (C_{X_{t-\tau_X} \to Z_t} \cup C_{Y_{t-\tau_Y} \to Z_t}), \tag{C1b}
$$

$$
\vec{W}_2 = P_{C_{X_{t-\tau_X} \to Z_t}} \backslash C_{Y_{t-\tau_Y} \to Z_t}, \tag{C1c}
$$

$$
\vec{W}_3 = P_{C_{Y_{t-\tau_Y} \to Z_t}} \backslash C_{X_{t-\tau_X} \to Z_t}. \tag{C1d}
$$

The interaction information in MPID can be written as

$$
\begin{aligned}
\Delta I^{MSCP}_{\{X_{t-\tau_X}, Y_{t-\tau_Y}\} \to Z_t} \\
&= II(X_{t-\tau_X}; Y_{t-\tau_Y}; Z_t \mid \vec{W}) \\
&= I(X_{t-\tau_X}; Y_{t-\tau_Y} \mid \{Z_t\} \cup \vec{W}) - I(X_{t-\tau_X}; Y_{t-\tau_Y} \mid \vec{W}),
\end{aligned} \tag{C2}
$$

which means the expressions for the two conditional mutual information values $I(X_{t-\tau_X}; Y_{t-\tau_Y} \mid \{Z_t\} \cup \vec{W})$ and $I(X_{t-\tau_X}; Y_{t-\tau_Y} \mid \vec{W})$ are required.

(i) Assume the *additivity* condition holds. Then, $I(X_{t-\tau_X}; Y_{t-\tau_Y} \mid \vec{W})$ can be expressed as

$$
\begin{aligned}
I(X_{t-\tau_X}; & Y_{t-\tau_Y} \mid \vec{W}) \\
&\overset{\text{Eq.(15)}}{=} I(f_X(P^{\vec{B}}_{X_{t-\tau_X}}) + g_X(P_{X_{t-\tau_X}} \backslash P^{\vec{B}}_{X_{t-\tau_X}}) + \eta^X_{t-\tau_X}; \\
&\quad f_Y(P^{\vec{B}}_{Y_{t-\tau_Y}}) + g_Y(P_{Y_{t-\tau_Y}} \backslash P^{\vec{B}}_{Y_{t-\tau_Y}}) + \eta^Y_{t-\tau_2} \mid \vec{W}) \\
&\overset{\text{Eq.(B2)}}{=} I(f_X(P^{\vec{B}}_{X_{t-\tau_X}}) + \eta^X_{t-\tau_X}; f_Y(P^{\vec{B}}_{Y_{t-\tau_Y}}) + \eta^Y_{t-\tau_Y} \mid \vec{W})
\end{aligned} \tag{C3}
$$

The first equality holds due to the *additivity* condition defined in Eq.(15), while the second equality is obtained based on the translational invariance in Eq.(B2) because $P_{K_{t-\tau}} \backslash P^{\vec{B}}_{K_{t-\tau}} \in \vec{W}$ (where $K_{t-\tau} \in \vec{B}$).

Furthermore, due to the chain rule, $I(X_{t-\tau_X}; Y_{t-\tau_Y} \mid \{Z_t\} \cup \vec{W})$ in Eq.(C2) can be expanded as

$$
\begin{aligned}
I(X_{t-\tau_X}; & Y_{t-\tau_Y} \mid \{Z_t\} \cup \vec{W}) \\
&= I(X_{t-\tau_X}, Z_t; Y_{t-\tau_Y} \mid \vec{W}) - I(Y_{t-\tau_Y}; Z_t \mid \vec{W}).
\end{aligned} \tag{C4}
$$

$I(X_{t-\tau_X}, Z_t; Y_{t-\tau_Y} \mid \vec{W})$ in Eq.(C4) can be further expressed as

$$
\begin{aligned}
I(X_{t-\tau_X}, & Z_t; Y_{t-\tau_Y} \mid \vec{W}) \\
&\overset{\text{Eq.(15)}}{=} I(f_X(P^{\vec{B}}_{X_{t-\tau_X}}) + g_X(P_{X_{t-\tau_X}} \backslash P^{\vec{B}}_{X_{t-\tau_X}}) + \eta^X_{t-\tau_X}, \\
&\quad f_Z(P^{\vec{B}}_{Z_t}) + g_Z(P_{Z_t} \backslash P^{\vec{B}}_{Z_t}) + \eta^Z_t; \\
&\quad f_Y(P^{\vec{B}}_{Y_{t-\tau_Y}}) + g_Y(P_{Y_{t-\tau_Y}} \backslash P^{\vec{B}}_{Y_{t-\tau_Y}}) + \eta^Y_{t-\tau_Y}) \mid \vec{W}) \\
&\overset{\text{Eq.(B2)}}{=} I(f_X(P^{\vec{B}}_{X_{t-\tau_X}}) + g_X(P_{X_{t-\tau_X}} \backslash P^{\vec{B}}_{X_{t-\tau_X}}) + \eta^X_{t-\tau_X}, \\
&\quad f_Z(P^{\vec{B}}_{Z_t}) + g_Z(P_{Z_t} \backslash P^{\vec{B}}_{Z_t}) + \eta^Z_t; \\
&\quad f_Y(P^{\vec{B}}_{Y_{t-\tau_Y}}) + \eta^Y_{t-\tau_Y} \mid \vec{W}) \\
&\overset{\text{Eq.(B4)}}{=} I(f_X(P^{\vec{B}}_{X_{t-\tau_X}}) + \eta^X_{t-\tau_X}, f_Z(P^{\vec{B}}_{Z_t}) + \eta^Z_t; \\
&\quad f_Y(P^{\vec{B}}_{Y_{t-\tau_Y}}) + \eta^Y_{t-\tau_Y} \mid \vec{W}).
\end{aligned} \tag{C5}
$$

The first equality is because of the *additivity* condition. For the second and third equalities, notice that g represents the function of the parents not in $\vec{B}$, $P_{K_{t-\tau}} \backslash P^{\vec{B}}_{K_{t-\tau}}$, which are a part of the condition set $\vec{W}$ (i.e., $P_{K_{t-\tau}} \backslash P^{\vec{B}}_{K_{t-\tau}} \in \vec{W}$). Therefore, the translational invariance in Eqs.(B2) and (B4) are applicable in the last

two equalities in Eq.(C5). Similarly, the second term on the right hand side of Eq.(C4) can be changed into

$$I(Y_{t-\tau_Y}; Z_t \mid \vec{W})$$
$$\overset{\text{Eq.(15)}}{=} I(f_Y(P_{Y_{t-\tau_Y}}^{\vec{B}}) + g_Y(P_{Y_{t-\tau_Y}} \backslash P_{Y_{t-\tau_Y}}^{\vec{B}}) + \eta_{t-\tau_Y}^Y;$$
$$f_Z(P_{Z_t}^{\vec{B}}) + g_Z(P_{Z_t} \backslash P_{Z_t}^{\vec{B}}) + \eta_t^Z \mid \vec{W})$$
$$\overset{\text{Eq.(B3)}}{=} I(f_Y(P_{Y_{t-\tau_Y}}^{\vec{B}}) + \eta_{t-\tau_Y}^Y; f_Z(P_{Z_t}^{\vec{B}}) + \eta_t^Z \mid \vec{W}). \tag{C6}$$

Combining Eqs.(C3), (C4), (C5) and (C6) into Eq.(C2), we get

$$\Delta I_{\{X_{t-\tau_X}, Y_{t-\tau_Y}\} \to Z_t}^{MSCP}$$
$$= I(f_X(P_{X_{t-\tau_X}}^{\vec{B}}) + \eta_{t-\tau_X}^X, f_Y(P_{Y_{t-\tau_Y}}^{\vec{B}}) + \eta_{t-\tau_Y}^Y;$$
$$f_Z(P_{Z_t}^{\vec{B}}) + \eta_t^Z \mid \vec{W}) -$$
$$I(f_X(P_{X_{t-\tau_X}}^{\vec{B}}) + \eta_{t-\tau_X}^X; f_Z(P_{Z_t}^{\vec{B}}) + \eta_t^Z \mid \vec{W}) -$$
$$I(f_Y(P_{Y_{t-\tau_Y}}^{\vec{B}}) + \eta_{t-\tau_Y}^Y; f_Z(P_{Z_t}^{\vec{B}}) + \eta_t^Z \mid \vec{W})$$
$$= II(f_X(P_{X_{t-\tau_X}}^{\vec{B}}) + \eta_{t-\tau_X}^X; f_Y(P_{Y_{t-\tau_2}}^{\vec{B}}) + \eta_{t-\tau_Y}^Y;$$
$$f_Z(P_{Z_t}^{\vec{B}}) + \eta_t^Z \mid \vec{W}). \tag{C7}$$

(ii) If *linearity* and *additivity* hold, which means all the f functions are linear such that all the nodes on the two causal paths and the target node are linearly dependent on each other, the interaction information in Eq.(C7) can be reduced to

$$\Delta I_{\{X_{t-\tau_X}, Y_{t-\tau_Y}\} \to Z_t}^{MSCP}$$
$$= I(\tilde{f}_X(\bullet) + \eta_{t-\tau_X}^X, \tilde{f}_Y(\bullet) + \eta_{t-\tau_Y}^Y; \tilde{f}_Z(\bullet) + \eta_t^Z \mid \vec{W}) -$$
$$I(\tilde{f}_X(\bullet) + \eta_{t-\tau_X}^X; \tilde{f}_Z(\bullet) + \eta_t^Z \mid \vec{W}) -$$
$$I(\tilde{f}_Y(\bullet) + \eta_{t-\tau_Y}^Y; \tilde{f}_Z(\bullet) + \eta_t^Z \mid \vec{W}) \qquad (linearity)$$
$$= II(\tilde{f}_X(\bullet) + \eta_{t-\tau_X}^X; \tilde{f}_Y(\bullet) + \eta_{t-\tau_Y}^Y; \tilde{f}_Z(\bullet) + \eta_t^Z \mid \vec{W}), \tag{C8a}$$

where both $\tilde{f}_X(\bullet)$, $\tilde{f}_Y(\bullet)$ and $\tilde{f}_Z(\bullet)$ are the linear functions of all the noise terms η_t in the union of the two causal paths and the target node (i.e., $\vec{B}$). The first equality results from the fact that each parent in the linear function f can be iteratively decomposed into the summation of both the noise terms η and the function g of the external nodes (i.e, the parents of the two causal paths and the target node), and therefore the translational invariance in Eqs.(B2)-(B4) can be used for canceling out the g functions. Furthermore, because all the noise terms η are i.i.d., the condition $\vec{W}$ is independent of η and Eq.(C8a) can be simplified as

$$\Delta I_{\{X_{t-\tau_X}, Y_{t-\tau_Y}\} \to Z_t}^{MSCP}$$
$$= II(\tilde{f}_X(\bullet) + \eta_{t-\tau_X}^X; \tilde{f}_Y(\bullet) + \eta_{t-\tau_Y}^Y; \tilde{f}_Z(\bullet) + \eta_t^Z), \tag{C9}$$

yielding Eq.(17a). Similarly, the translationally invariant property can be used to simplify the momentary interaction partitioning in Eqs.(13a)-(13b) as

$$R_c = R_{min,c} + I_{s,c}(R_{MMI,c} - R_{min,c}) \tag{C10a}$$
$$S_c = \Delta I^{MSCP} + R_c \tag{C10b}$$
$$U_{X,c} = I(\tilde{f}_X(\bullet) + \eta_{t-\tau_X}^X; \tilde{f}_Z(\bullet) + \eta_t^Z) - R_c \tag{C10c}$$
$$U_{Y,c} = I(\tilde{f}_Y(\bullet) + \eta_{t-\tau_Y}^Y; \tilde{f}_Z(\bullet) + \eta_t^Z) - R_c, \tag{C10d}$$

where

$$R_{MMI,c} = \min[I(\tilde{f}_X(\bullet) + \eta_{t-\tau_X}^X; \tilde{f}_Z(\bullet) + \eta_t^Z),$$
$$I(\tilde{f}_Y(\bullet) + \eta_{t-\tau_Y}^Y; \tilde{f}_Z(\bullet) + \eta_t^Z)]$$
$$I_{s,c} = \frac{I(\tilde{f}_X(\bullet) + \eta_{t-\tau_X}^X; \tilde{f}_Y(\bullet) + \eta_{t-\tau_2}^Y)}{\min[H(\tilde{f}_X(\bullet) + \eta_{t-\tau_X}^X), H(\tilde{f}_Y(\bullet) + \eta_{t-\tau_Y}^Y)]}$$
$$R_{min,c} = \begin{cases} 0, & \text{if } \Delta I^{MSCP} \geq 0 \\ -\Delta I^{MSCP}, & \text{otherwise.} \end{cases}$$

It is obviously that under the *linearity* and *additivity* conditions, MPID is independent of the condition $\vec{W}$, resulting in the coupling strength autonomy property.

(iii) If *separability* and *additivity* hold, $P_{X_{t-\tau_X}}^{\vec{B}}$ and $P_{Y_{t-\tau_Y}}^{\vec{B}}$ are empty because X (or Z) is not in the causal path $C_{Y_{t-\tau_Y} \to Z_t}$ (or $C_{X_{t-\tau_X} \to Z_t}$). Therefore, $f_X(P_{X_{t-\tau_X}}^{\vec{B}})$ and $f_Y(P_{Y_{t-\tau_Y}}^{\vec{B}})$ are zero, which allows Eq.(C7) to be revised as

$$\Delta I_{\{X_{t-\tau_X}, Y_{t-\tau_Y}\} \to Z_t}^{MSCP}$$
$$= II(\eta_{t-\tau_X}^X; \eta_{t-\tau_Y}^Y; f_Z(P_{Z_t}^{\vec{B}}) + \eta_t^Z \mid \vec{W})$$
$$= I(\eta_{t-\tau_X}^X; \eta_{t-\tau_Y}^Y \mid \{f_Z(P_{Z_t}^{\vec{B}}) + \eta_t^Z\} \cup \vec{W})$$
$$\quad - I(\eta_{t-\tau_X}^X; \eta_{t-\tau_Y}^Y \mid \vec{W})$$
$$= I(\eta_{t-\tau_X}^X; \eta_{t-\tau_Y}^Y \mid \{f_Z(P_{Z_t}^{\vec{B}}) + \eta_t^Z\} \cup \vec{W}), \tag{C11}$$

yielding Eq.(19a). The final equality holds because η_t^X and η_t^Y are i.i.d. and the nodes in the $\vec{W}$ are not the common children of $\eta_{t-\tau_X}^X$ and $\eta_{t-\tau_Y}^Y$, leading to $I(\eta_{t-\tau_X}^X; \eta_{t-\tau_Y}^Y \mid \vec{W}) = 0$. Also, notice that due to the *separability* condition, $I(X_{t-\tau_X}; Y_{t-\tau_Y} \mid \vec{W}) = I(\eta_{t-\tau_X}^X; \eta_{t-\tau_Y}^Y \mid \vec{W}) = 0$ (here, g_X and g_Y cancel out because of the translational invariance). Hence, $I_{s,c}$ in Eq.(13b) is reduced to zero (since all the noises η are i.i.d.), resulting in the minimum value of the redundancy such that $R_c = R_{min,c}$ according to Eq.(13b). Therefore, the synergistic and redundant information can be expressed as

$$R_c = R_{min,c} \tag{C12}$$
$$S_c = \Delta I^{MSCP} + R_c. \tag{C13}$$

The two unique information can also be obtained based

on the *separability* condition such that

$$U_{X,c} = I(\eta_{t-\tau_X}^X; f_Z(P_{Z_t}^{\vec{B}}) + \eta_t^Z \mid \vec{W}) - R_c \qquad \text{(C14)}$$

$$U_{Y,c} = I(\eta_{t-\tau_Y}^Y; f_Z(P_{Z_t}^{\vec{B}}) + \eta_t^Z \mid \vec{W}) - R_c. \qquad \text{(C15)}$$

(iv) If all the three conditions hold, we can easily obtain MPID by combining the results of both the situations (ii) and (iii) such that

$$\Delta I_{\{X_{t-\tau_X}, Y_{t-\tau_Y}\} \to Z_t}^{MSCP}$$

$$= I(\eta_{t-\tau_X}^X; \eta_{t-\tau_Y}^Y \mid \tilde{f}_Z(\bullet) + \eta_t^Z) \qquad \text{(C16a)}$$

$$R_c = R_{min,c} \qquad \text{(C16b)}$$

$$S_c = \Delta I^{MSCP} + R_c \qquad \text{(C16c)}$$

$$U_{X,c} = I(\eta_{t-\tau_X}^X; \tilde{f}_Z(\bullet) + \eta_t^Z) - R_c \qquad \text{(C16d)}$$

$$U_{Y,c} = I(\eta_{t-\tau_Y}^Y; \tilde{f}_Z(\bullet) + \eta_t^Z) - R_c. \qquad \text{(C16e)}$$

Appendix D: Analytical solutions of the linear common driver model

In the following, we show the derivation of the analytical solutions of $\Delta I_{\{X_{t-1}, Y_{t-1}\} \to Z_t}^{MSCP}$ and $II(X_{t-1}; Y_{t-1}; Z_t)$ in Eqs.(21c) and (22) for the linear common driver model in Eq.(20).

a. Variances and covariances

We derive some variances and covariances of the linear model Eq.(20) for the further usage in the derivation of the analytical solutions of $\Delta I_{\{X_{t-1}, Y_{t-1}\} \to Z_t}^{MSCP}$ and $II(X_{t-1}; Y_{t-1}; Z_t)$.

The variances of the four sub-processes (i.e., V, X, Y, Z) can be expressed as

$$\Gamma_V = \sigma_V^2$$
$$\Gamma_X = c_{VX}^2 \Gamma_V + \sigma_X^2$$
$$\Gamma_Y = c_{VY}^2 \Gamma_V + \sigma_Y^2$$
$$\Gamma_Z = c_{XZ}^2 \Gamma_X + c_{YZ}^2 \Gamma_Y + \sigma_Z^2.$$

Also, some of the used covariances between two variables (e.g., X and Y) with a lag τ, denoted as $\Gamma_{XY}(\tau) = E[X_{t+\tau} Y_t]$ (where E represents the expectation function), are given by

$$\Gamma_{XY}(0) = c_{VX} c_{VY} \Gamma_V$$
$$\Gamma_{ZX}(1) = c_{XZ} \Gamma_X + c_{YZ} \Gamma_{XY}(0)$$
$$\Gamma_{ZY}(1) = c_{YZ} \Gamma_Y + c_{XZ} \Gamma_{XY}(0)$$
$$\Gamma_{XV}(1) = c_{VX} \Gamma_V$$
$$\Gamma_{YV}(1) = c_{VY} \Gamma_V$$
$$\Gamma_{ZV}(2) = c_{XZ} \Gamma_{XV}(1) + c_{YZ} \Gamma_{YV}(1).$$

b. The momentary interaction information for separable causal paths $\Delta I_{\{X_{t-1}, Y_{t-1}\} \to Z_t}^{MSCP}$

As shown in Fig. 3b, the condition set $\vec{W}$ for $\Delta I_{\{X_{t-1}, Y_{t-1}\} \to Z_t}^{MSCP}$ is $\{V_{t-2}\}$. Therefore, $\Delta I_{\{X_{t-1}, Y_{t-1}\} \to Z_t}^{MSCP}$ can be written as

$$\Delta I_{\{X_{t-1}, Y_{t-1}\} \to Z_t}^{MSCP}$$
$$= II(X_{t-1}; Y_{t-1}; Z_t \mid V_{t-2})$$
$$= I(X_{t-1}; Y_{t-1} \mid Z_t, V_{t-2}) - I(X_{t-\tau_1}; Y_{t-\tau_2} \mid V_{t-2})$$
$$= H(X_{t-1} \mid Z_t, V_{t-2}) - H(X_{t-1} \mid Y_{t-1}, Z_t, V_{t-2})$$
$$\quad - H(X_{t-1} \mid V_{t-2}) + H(X_{t-1} \mid Y_{t-1}, V_{t-2}). \qquad \text{(D1)}$$

Because all the processes in the linear model are Gaussian, based on Eq.(B5), each component in $\Delta I_{\{X_{t-1}, Y_{t-1}\} \to Z_t}^{MSCP}$ is given by

$$H(X_{t-1} \mid Z_t, V_{t-2}) = \frac{1}{2} \ln \left\{ 2\pi e \frac{|\Gamma_{X_{t-1}, Z_t, V_{t-2}}|}{|\Gamma_{Z_t, V_{t-2}}|} \right\}$$

$$H(X_{t-1} \mid Y_{t-1}, Z_t, V_{t-2}) = \frac{1}{2} \ln \left\{ 2\pi e \frac{|\Gamma_{X_{t-1}, Y_{t-1}, Z_t, V_{t-2}}|}{|\Gamma_{Y_{t-1}, Z_t, V_{t-2}}|} \right\}$$

$$H(X_{t-1} \mid V_{t-2}) = \frac{1}{2} \ln \left\{ 2\pi e \frac{|\Gamma_{X_{t-1}, V_{t-2}}|}{|\Gamma_{V_{t-2}}|} \right\}$$

$$H(X_{t-1} \mid Y_{t-1}, V_{t-2}) = \frac{1}{2} \ln \left\{ 2\pi e \frac{|\Gamma_{X_{t-1}, Y_{t-1}, V_{t-2}}|}{|\Gamma_{Y_{t-1}, V_{t-2}}|} \right\}. \qquad \text{(D2)}$$

To derive the analytical solutions of the above conditional entropies, we need to solve the determinants involved in the above equations. Consider the example of $H(X_{t-1} \mid V_{t-2})$. The determinant of the covariance matrix $|\Gamma_{X_{t-1}, V_{t-2}}|$ can be expressed as

$$|\Gamma_{X_{t-1}, V_{t-2}}| = \begin{vmatrix} \Gamma_X & \Gamma_{XV}(1) \\ \Gamma_{XV}(1) & \Gamma_V \end{vmatrix}$$
$$= \Gamma_X \Gamma_V - \Gamma_{XV}(1)^2$$
$$= \sigma_X^2 \Gamma_V. \qquad \text{(D3)}$$

Also, because $|\Gamma_{V_{t-2}}| = \Gamma_V$, the analytical solution of $H(X_{t-1} \mid V_{t-2})$ is given by

$$H(X_{t-1} \mid V_{t-2}) = \frac{1}{2} \ln\{2\pi e \sigma_X^2\}. \qquad \text{(D4)}$$

Therefore, by solving all the determinants above and putting all the analytical solutions of the four conditional entropy back to Eq.(D1), we obtain the analyticial solution of $\Delta I_{\{X_{t-1}, Y_{t-1}\} \to Z_t}^{MSCP}$ as shown in Eq.(21c).

It is noted that because all the three conditions (i.e., *separability*, *linearity* and *additivity*) hold for the linear common driver model, the solution in Eq.(21c) can also be achieved by solving $\Delta I_{\{X_{t-\tau_1}, Y_{t-\tau_2}\} \to Z_t}^{MSCP} = II(\eta_{t-\tau_1}^X; \eta_{t-\tau_2}^Y; \tilde{f_Z}(\bullet) + \eta_t^Z)$ directly, which is not shown here.

c. Interaction information $II(X_{t-1}; Y_{t-1}; Z_t)$

$II(X_{t-1}; Y_{t-1}; Z_t)$ can be expanded as

$$
\begin{aligned}
II&(X_{t-1}; Y_{t-1}; Z_t) \\
&= I(X_{t-1}; Y_{t-1} \mid Z_t) - I(X_{t-\tau_1}; Y_{t-\tau_2}) \\
&= H(X_{t-1} \mid Z_t) - H(X_{t-1} \mid Y_{t-1}, Z_t) - H(X_{t-1}) \\
&\quad + H(X_{t-1} \mid Y_{t-1})
\end{aligned} \tag{D5}
$$

Similarly, the Gaussian process-based $II(X_{t-1}; Y_{t-1}; Z_t)$ can be further revised according to Eq.(B5) as,

$$
\begin{aligned}
II&(X_{t-1}; Y_{t-1}; Z_t) \\
&= \frac{1}{2} \ln\left\{ 2\pi e \frac{|\Gamma_{X_{t-1}, Z_t}|}{|\Gamma_{Z_t}|} \right\} - \frac{1}{2} \ln\left\{ 2\pi e \frac{|\Gamma_{X_{t-1}, Y_{t-1}, Z_t}|}{|\Gamma_{Y_{t-1}, Z_t}|} \right\} - \\
&\quad \frac{1}{2} \ln\left\{ 2\pi e |\Gamma_{X_{t-1}}| \right\} + \frac{1}{2} \ln\left\{ 2\pi e \frac{|\Gamma_{X_{t-1}, Y_{t-1}}|}{|\Gamma_{Y_{t-1}}|} \right\} \\
&= \frac{1}{2} \ln\left\{ \frac{|\Gamma_{X_{t-1}, Z_t}| |\Gamma_{Y_{t-1}, Z_t}| |\Gamma_{X_{t-1}, Y_{t-1}}|}{|\Gamma_{Z_t}| |\Gamma_{X_{t-1}, Y_{t-1}, Z_t}| |\Gamma_{X_{t-1}}| |\Gamma_{Y_{t-1}}|} \right\}.
\end{aligned} \tag{D6}
$$

By solving all the determinants in Eq.(D6), we can obtain the solution in Eq.(22).

REFERENCES

[1] B. Pompe and J. Runge, Phys. Rev. E **83**, 051122 (2011).

[2] J. Runge, J. Heitzig, N. Marwan, and J. Kurths, Phys. Rev. E **86**, 061121 (2012).

[3] M. Eichler, Probability Theory and Related Fields **153**, 233 (2012).

[4] J. Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge University Press, New York, NY, USA, 2000).

[5] C. W. J. Granger, Econometrica **37**, 424 (1969).

[6] J. Runge, Phys. Rev. E **92**, 062829 (2015).

[7] B. L. Ruddell and P. Kumar, Water Resources Research **45**, n/a (2009), w03419.

[8] B. L. Ruddell and P. Kumar, Water Resources Research **45**, n/a (2009), w03420.

[9] J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths, Phys. Rev. Lett. **108**, 258701 (2012).

[10] P. L. Williams and R. D. Beer, CoRR **abs/1004.2515** (2010).

[11] A. B. Barrett, Phys. Rev. E **91**, 052802 (2015).

[12] A. E. Goodwell and P. Kumar, Water Resources Research **53**, 5920 (2017).

[13] A. E. Goodwell and P. Kumar, Water Resources Research **53**, 5899 (2017).

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)* (Wiley-Interscience, 2006).

[15] P. Spirtes, C. Glymour, and R. Scheines, "Causation, prediction, and search. adaptive computation and machine learning," (2000).

[16] M. Harder, C. Salge, and D. Polani, Phys. Rev. E **87**, 012130 (2013).

[17] R. A. A. Ince, Entropy **19**, 318 (2017).

[18] N. Bertschinger, J. Rauh, E. Olbrich, J. Jost, and N. Ay, Entropy **16**, 2161 (2014).

[19] V. Griffith and C. Koch, "Quantifying synergistic mutual information," in *Guided Self-Organization: Inception*, edited by M. Prokopenko (Springer Berlin Heidelberg, Berlin, Heidelberg, 2014) pp. 159–190.

[20] E. Olbrich, N. Bertschinger, and J. Rauh, Entropy **17**, 3501 (2015).

[21] V. Griffith and T. Ho, Entropy **17**, 4644 (2015).

[22] N. Bertschinger, J. Rauh, E. Olbrich, and J. Jost, in *Proceedings of the European Conference on Complex Systems 2012*, edited by T. Gilbert, M. Kirkilionis, and G. Nicolis (Springer International Publishing, Cham, 2013) pp. 251–269.

[23] The conditional redundancy proposed by Bertschinger et al. [22] suggests a "left chain rule", which is a generalization of the chain rule of mutual information. The "left chain rule" property is not fulfilled in most redundancy measures but it's promising in a recent paper about pointwise information decomposition [34].

[24] B. W. Silverman, *Density estimation for statistics and data analysis*, Vol. 26 (CRC press, 1986).

[25] M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, Phys. Rev. Lett. **78**, 4193 (1997).

[26] A. C. Martí, M. Ponce, and C. Masoller, Pramana **70**, 1117 (2008).

[27] C. Masoller and F. Atay, The European Physical Journal D **62**, 119 (2011).

[28] G. Paredes, O. Alvarez-Llamoza, and M. G. Cosenza, Phys. Rev. E **88**, 042920 (2013).

[29] J. Aguirre, R. Sevilla-Escoboza, R. Gutiérrez, D. Papo, and J. M. Buldú, Phys. Rev. Lett. **112**, 248701 (2014).

[30] D. Baldocchi, E. Falge, L. Gu, R. Olson, D. Hollinger, S. Running, P. Anthoni, C. Bernhofer, K. Davis, R. Evans, J. Fuentes, A. Goldstein, G. Katul, B. Law, X. Lee, Y. Malhi, T. Meyers, W. Munger, W. Oechel, K. T. Paw, K. Pilegaard, H. P. Schmid, R. Valentini, S. Verma, T. Vesala, K. Wilson, and S. Wofsy,

Bulletin of the American Meteorological Society **82**, 2415 (2001).

[31] C. Neal, B. Reynolds, J. W. Kirchner, P. Rowland, D. Norris, D. Sleep, A. Lawlor, C. Woods, S. Thacker, H. Guyatt, C. Vincent, K. Lehto, S. Grant, J. Williams, M. Neal, H. Wickham, S. Harman, and L. Armstrong, Hydrological Processes **27**, 2531 (2013).

[32] P. Kumar and B. L. Ruddell, Entropy **12**, 2085 (2010).

[33] J. Runge, V. Petoukhov, and J. Kurths, Journal of Climate **27**, 720 (2014).

[34] C. Finn and J. T. Lizier, arXiv preprint arXiv:1801.09010 (2018).