How should we define Information Flow in Neural Circuits?

Praveen Venkatesh*, Sanghamitra Dutta† and Pulkit Grover‡
Electrical & Computer Engineering, and the Center for the Neural Basis of Cognition, Carnegie Mellon University
*vpraveen@cmu.edu †sanghamd@andrew.cmu.edu ‡pulkit@cmu.edu

***Abstract*—We develop a theoretical framework for defining information flow in neural circuits, within the context of "event-related" experimental paradigms in neuroscience. Here, a neural circuit is modeled as a directed graph, with "clocked" nodes that send transmissions to each other along the edges of the graph at discrete points in time. We are interested in a definition that captures the flow of "stimulus"-related information, and which guarantees a continuous information path between appropriately defined inputs and outputs in the directed graph. Prior measures, including those based on Granger Causality and Directed Information, fail to provide clear assumptions and guarantees about when they correctly reflect stimulus-related information flow, due to the absence of a theoretical foundation with a mathematical definition. We take a methodical approach—iterating through candidate definitions and counterexamples—to arrive at a definition for information flow that is based on conditional mutual information, and which satisfies desirable properties, including the existence of information paths.**

I. Introduction

Neuroscientists often seek an understanding of how information flows in the brain while it performs a particular task [3]–[6]. As a concrete example, consider [3], where, crudely speaking, the authors are trying to discern whether information about images of common hand-held tools passes from visual cortex to motor cortex and then to the area responsible for object recognition, or to the object recognition area first, and only then to motor cortex. Distinguishing between these hypotheses involves understanding how information about a tool's identity flows in the brain. Experiments such as this—termed "event-related paradigms" [7] due to the stimulus, or "event", that occurs at the beginning of the task—are common in cognitive neuroscience, and form the basis for our model of neural computation. We use this model to develop a theoretical framework for understanding information flow in such experiments. Our framework is also general enough to analyze information flow in various kinds of Artificial Neural Networks. Ultimately, we believe that such a framework will help with understanding brain function, and hence with diagnosing and treating brain diseases [8]–[11].

Prior work on statistically inferring flows of information in the brain appears under the umbrella of "functional connectivity mapping" [12], [13]. These efforts have largely relied on measures of statistical causal influence such as Granger Causality [14], Massey's Directed Information [15]–[17], Transfer Entropy [18] and Partial Directed Coherence [19]. Despite widespread use, these measures have frequently been a subject of debate and disagreement within the neuroscientific community [20]–[26]. In part, these disagreements stem from the widely-acknowledged fact that under non-ideal measurement conditions (e.g. in the presence of hidden variables [27, p. 54], asymmetric noise [28], or limited sampling [29]), estimation of these quantities may be erroneous. While these non-idealities may eventually be overcome through technological improvements, we believe that more fundamental issues will still remain. For instance, in previous work, we demonstrated using a feedback communication network that even under *ideal* measurement conditions, the direction inferred by comparing Granger-causal influences can be opposite to the direction of information flow [30]. Another fundamental issue is that these measures do not directly capture the effect of the stimulus. Instead, it is often left in the hands of the experimentalist to find causal influences that are stimulus-dependent [4], [31].

Fundamental issues of this nature persist due to the absence of a formal theory which links a model for information flow with the signals that are actually recorded. The lack of a model and a mathematical definition has led to conflation of *defining* and *estimating* information flow. We believe that this field requires a Shannon-like approach, for modeling the underlying computational system and for defining information flow.

Based on prior art in the information theory literature [32], [33], we propose a model of neural computation consisting of nodes communicating to each other at discrete points in time on a directed graph. At every time instant, each node receives transmissions on its incoming edges and computes a function of these transmissions to send out on its outgoing edges. This model is sufficiently general to also encompass various kinds of Artificial Neural Networks. We will be interested in the flow of a particular random variable called the "message" (defined in Section II). Then, we state an intuitive property (Property 1), which we use to motivate a definition for information flow through counterexamples (Section III). Finally, we show that this definition satisfies several desirable properties, including our main result: the existence of "information paths" (Section IV). Proofs are deferred to the appendices in the full version of this paper [1]. An extended journal-length manuscript [2] builds on this paper and includes detailed discussions as well as example circuits such as the Network Coding Butterfly and the Fast Fourier Transform.

The full version of this paper, including appendices, can be found online [1]. Additionally, an extended journal-length manuscript of this paper is under peer review as of this writing, and a draft is available online [2].

We thank M. Bakshi, M. Behrmann, T. Coleman, E. Collins, U. Jagadisan, H. Jeong, R. Kass, G. Schamburg and T. Weissman for extremely useful discussions. PV was supported, in part, by a CMLH Fellowship in Digital Health. PG was supported, in part, by an NSF CAREER Award.

II. The Computational System

In this section, we define the computational system that is used to model neural circuits throughout this paper. But first, we start with the definition of a time-unrolled graph, upon which the computational system model is based.

Definition 1 (Complete directed graph): A *complete directed graph* $\mathcal{G}^* = (\mathcal{V}^*, \mathcal{E}^*)$ is described by a set of nodes and the set of all edges between those nodes (including self-edges). We denote the set of nodes by their indices, $\mathcal{V}^* = \{1, 2, \ldots, N\}$, where N is a positive integer denoting the number of nodes in the graph. The set of edges in the graph is the set of all ordered pairs of nodes, $\mathcal{E}^* = \mathcal{V}^* \times \mathcal{V}^*$.

Remarks: (i) Edges are directed, so the edge $(A, B) \in \mathcal{E}^*$ describes an edge *from* node A *to* node B. (ii) Nodes have self-edges. For every $A \in \mathcal{V}^*$, there is an edge (A, A) in $\mathcal{E}^*$. (iii) Moving forward, nodes shall be thought of as performing computations and possessing local memories. We shall interpret the transmission of a node to itself as the variable it stores within its memory.[1]

Definition 2 (Time-unrolled graph): Let $\mathcal{T} = \{0, 1, \ldots, T\}$ be a set of time indices, where T is a positive integer representing the maximum time index. Then, a *time-unrolled graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is constructed by indexing a complete directed graph $\mathcal{G}^*$ using the time indices $\mathcal{T}$ as follows: (i) The nodes $\mathcal{V}$ consist of all nodes $\mathcal{V}^*$ in $\mathcal{G}^*$, subscripted by time indices $\mathcal{T}$, i.e., $\mathcal{V} = \{A_t : A \in \mathcal{V}^*, t \in \mathcal{T}\}$; (ii) The edges $\mathcal{E}$ connect nodes of *successive* times in $\mathcal{V}$, so they can be written in terms of the edges in $\mathcal{E}^*$ as $\mathcal{E} = \{(A_t, B_{t+1}) : (A, B) \in \mathcal{E}^*, t \in \mathcal{T}\}$.

Remarks: (i) For brevity, we denote the set of all nodes at time t by $\mathcal{V}_t$, and the set of all (outgoing) edges at time t by $\mathcal{E}_t$. So, for example, we will have $A_1 \in \mathcal{V}_1$ and $(A_1, B_2) \in \mathcal{E}_1$. (ii) Edges at time t connect nodes at time t to nodes at time $t+1$. (iii) Since the original graph $\mathcal{G}^*$ had self-edges, there will always be an edge (A_t, A_{t+1}) in $\mathcal{E}_t$ for every node $A_t \in \mathcal{V}_t$.

Definition 3 (Computational System): A *computational system* $\mathcal{C} = (\mathcal{G}, X, W, f)$ is a time unrolled graph $\mathcal{G}$ that has *transmissions on its edges* which are constrained by *computations at its nodes*. The *input nodes* of the computational system compute a function of a *message*, M. We now elaborate upon these italicized terms:

3a) Transmissions on Edges

In a time-unrolled graph $\mathcal{G}$, let $X : \mathcal{E} \to \mathfrak{X}$ be a function that describes what random variable is being transmitted on a given edge, i.e., $X(E)$ is the random variable corresponding to the transmission on the edge E. Here, the range $\mathfrak{X}$ is the set of all random variables in some probability space.[2]

For convenience, we define X applied to a *set of edges* as the set of random variables produced by applying X to each of those edges individually, i.e., for *any* subset $\mathcal{E}' \subseteq \mathcal{E}$,

$$X(\mathcal{E}') = \{X(E) : E \in \mathcal{E}'\}. \tag{1}$$

We extend the use of this notation to other functions of nodes and edges that we define, going forward.

3b) Computation at a Node

Let $A_t \in \mathcal{V}_t$ be a node in the time-unrolled graph $\mathcal{G}$, at some time $t \geq 1$ (recall that $t \in \{0, 1, \ldots, T\}$). Let $\mathcal{P}(A_t)$ be the set of edges entering A_t, and $\mathbb{Q}(A_t)$ be the set of edges leaving A_t. Further, let us suppose that A_t is able to intrinsically generate the random variable[3] $W(A_t)$ at time t, where $W(A_t) \perp\!\!\!\perp W(\mathcal{V} \backslash \{A_t\}) \; \forall \; A_t \in \mathcal{V}$, $W(\mathcal{V}_t) \perp\!\!\!\perp \{M, X(\mathcal{E}_{t-1})\}$ and the symbol "$\perp\!\!\!\perp$" stands for independence between random variables. Then, the *computation* performed by the node A_t (for $t \geq 1$) is a deterministic function[4] f_{A_t} that satisfies

$$f_{A_t}\big(X(\mathcal{P}(A_t)), W(A_t)\big) = X(\mathbb{Q}(A_t)). \tag{2}$$

Here, $X(\mathcal{E}_{t-1})$, $W(\mathcal{V} \setminus \{A_t\})$, $W(\mathcal{V}_t)$, $X(\mathcal{P}(A_t))$ and $X(\mathbb{Q}(A_t))$ all make use of the notation described in (1). Note that the definition above does not apply when $t = 0$; this is a special case which is discussed below.

3c) The Message and the Input Nodes

The *message* is a random variable M, which is of interest to the observer, and for which we shall define information flow. We assume that the message enters the computational system at (and only at) time $t = 0$. We formally define the *input nodes* of the system as those nodes of $\mathcal{G}$, at time $t = 0$, whose transmissions statistically depend on the message M: $\mathcal{V}_{\text{ip}} \coloneqq \{A_0 \in \mathcal{V}_0 : I\big(M; X(\mathbb{Q}(A_0))\big) > 0\}$, where $\mathbb{Q}(A_0)$ represents the set of edges leaving the node A_0.

To remain consistent with Definition 3b, we define the computation performed by an input node $A_0 \in \mathcal{V}_{\text{ip}}$ as a function f_{A_0} that satisfies $f_{A_0}\big(M, W(A_0)\big) = X(\mathbb{Q}(A_0))$, and the computation performed by a non-input node at time $t = 0$, $A_0 \in \mathcal{V}_0 \backslash \mathcal{V}_{\text{ip}}$, as a function f_{A_0} that satisfies $f_{A_0}\big(W(A_0)\big) = X(\mathbb{Q}(A_0))$. As before, $W(A_0) \perp\!\!\!\perp W(\mathcal{V}_0 \backslash \{A_0\}) \; \forall \; A_0 \in \mathcal{V}_0$ and $W(\mathcal{V}_0) \perp\!\!\!\perp M$.

Remarks: (i) Informally, Definition 3 allows each node to generate a randomized function of its incoming transmissions for each of its outgoing transmissions. (ii) The randomization at each node is explicitly captured by its intrinsic random variable $W(\cdot)$, and is assumed to be independent across all nodes of the system. (iii) Each node is allowed to send a different transmission on each of its outgoing edges. (iv) The condition imposed by Equation (2) introduces dependence between the random variables in the set $X(\mathcal{E})$. (v) We will not be concerned with the precise form of the computation being performed by every node. We will only make use of information-theoretic measures applied to the random variables in the computational system.

III. Defining Information Flow

Before one can speak of *estimating* information flow in a network, it is important to first *define* what we seek to estimate.

[1] Graphs that are *not complete* and nodes with *no memory* are simply special cases of our model, where the respective edges' transmissions are set to zero.

[2] We assume that the measures admit well-defined mutual- and conditional-mutual-information between any sets of random variables [34, Sec. 2.6].

[3] $X(E_t)$ and $W(A_t)$ may also be random *vectors* instead of random variables, i.e., an edge may *transmit a vector*. The proofs remain unchanged.

[4] This kind of model is not new, and can be found in the causality literature for instance, under the name "Structural Equation Models" [27, Sec. 1.4.1].

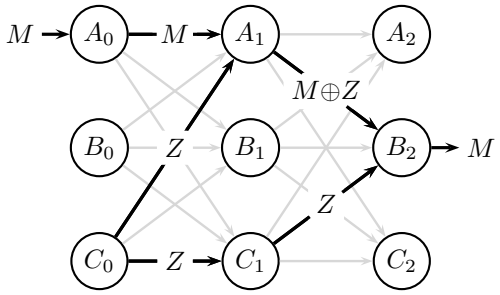

Figure 1: The computational system for Counterexample 1 (only relevant edges' transmissions are shown; transmissions on faded edges are assumed to be zero). Observe that no edge at time $t = 1$ has information flow as per Candidate Definition 1, yet the message reappears at time $t = 2$.

Towards finding the paths taken by the message while being processed in the computational system, we define information flow about the message on a single edge.

A. An Intuitive Property

Towards assessing and choosing between competing candidate definitions for information flow, we state a straightforward and intuitive property, which we would want any definition of information flow to satisfy.

Suppose that, at a given point in time, there is *no* flow of information about the message on *any* edge of a computational system (including self-edges). Then, we expect that information about the message has ceased to persist in the system, so the information flow about the message *must* be zero on all edges of the computational system, at all future points in time.

Property 1 (The Broken Telephone): Let $\mathscr{C}$ be a computational system, and let $\mathscr{F}_M : \mathscr{C} \to \{0, 1\}$ be an indicator of the presence of information flow about M on an edge. That is, $\mathscr{F}_M(E) = 1$, if information about M flows on the edge $E \in \mathscr{C}$ and $\mathscr{F}_M(E) = 0$, otherwise. The Broken Telephone Property states that if, at some time $t \in \mathscr{T}$, we have $\mathscr{F}_M(E_t) = 0 \; \forall \; E_t \in \mathscr{C}_t$, then $\mathscr{F}_M(E_{t'}) = 0 \; \forall \; E_{t'} \in \mathscr{C}_{t'} \;\; \forall \; t' \in \mathscr{T}, t' > t$.

B. Intuiting Information Flow through a Counterexample

A simple and intuitive definition for information flow might stem from dependence.

Candidate Definition 1: We say that information about the message M flows on an edge E if $I\big(M; X(E)\big) > 0$.

Counterexample 1: Consider the computational system depicted in Fig. 1. A_0 is the input node, which receives the message $M \sim \text{Ber}(1/2)$ at time $t = 0$. The system's goal is to communicate M to the node B (this is equivalent to computing the identity function, and making the output available at B). It uses the following strategy: (i) At $t = 0$, A_0 "transmits" M to A_1 (i.e., node A stores M in its memory). C_0 independently generates a random number, $W(C_0) = Z \sim \text{Ber}(1/2)$, $Z \perp\!\!\!\perp M$, and sends this variable to A_1 and C_1. (ii) At $t = 1$, A_1 computes $M \oplus Z$ and passes the result to B_2, while C_1 sends Z to B_2. Here, the symbol "$\oplus$" stands for XOR, the exclusive-OR operator on two bits. (iii) B_2 recovers M by once again XOR-ing its inputs, $M \oplus Z$ and Z.

Note that the output of B_2 depends on M, even though none of its inputs individually depends on M. That is, $I\big(M; X((A_1, B_2))\big) = I(M; M \oplus Z) = 0$, and $I\big(M; X((C_1, B_2))\big) = I(M; Z) = 0$, so by Candidate Definition 1, information about the message flows on *no* edge at time $t = 1$. However, information about the message *does* flow out of node B_2 at time $t = 2$. This violates Property 1. Thus, mere *dependence* on the message cannot be a valid definition for flow of information on a single edge. ■

Communication strategies such as the one in Counterexample 1 frequently arise in cryptography [35], to prevent an eavesdropper from reading confidential information, and in network coding [33], for achieving the communication capacity of a network (e.g., the butterfly network [33, Fig. 7b]). Furthermore, a complex computational network may have smaller sub-networks with such topologies.

Central to the idea of Counterexample 1 is a concept known as "synergy", which is well-studied in the literature on Partial Information Decomposition (PID; see [36] for a recent review). Even in neuroscience, the concept of synergy is recognized and well-understood [37]–[39], and some experimental evidence has appeared in the literature [40].

Counterexample 1 might appear to suggest that information flow ought to be defined for a *set* of edges. Indeed, this can be done in such a way as to be completely consistent with the single-edge definition that appears shortly (see Appendix D).

Attempts to correct Candidate Definition 1 by adding conditioning either on a *single* edge, or on *all* other edges' transmissions fail. Counterexamples to these can be found in the extended version of the paper [2]. Instead, as we shall see, one must condition on a *subset* of edges.

C. Information Flow on a Single Edge

Counterexample 1 motivates a new definition for information flow about a message on an edge. Given an edge E_t upon which we expect non-zero information flow, we observe: there is at least one subset of edges $\mathscr{C}'_t \subseteq \mathscr{C}_t \setminus \{E_t\}$ such that $X(E_t)$ is conditionally dependent on M, given $X(\mathscr{C}'_t)$.

Definition 4 (M-information Flow on a Single Edge): We say that information about the message M flows on an edge $E_t \in \mathscr{C}_t$ if

$$\exists \, \mathscr{C}'_t \subseteq \mathscr{C}_t \setminus \{E_t\} \quad \text{s.t.} \quad I\big(M; X(E_t) \,\big|\, X(\mathscr{C}'_t)\big) > 0. \tag{3}$$

Henceforth, we refer to "information flow about the message M" as *M-information flow*, and use the phrase "the edge E_t *has* M-information flow" or "the edge E_t *carries* M-information flow" to mean that information about M flows on E_t per this definition.

Definition 4 can also be stated using the concept of Synergistic information [36]: E_t has M-information flow if and only if $X(E_t)$ depends on M, or $X(E_t)$ has synergistic information about M, with respect to some subset $X(\mathscr{C}'_t)$. This equivalence is demonstrated in the extended version of this paper [2].

IV. Properties of Information Flow

Having defined information flow about a message for an edge, we demonstrate that Definition 4 satisfies several intuitively desirable properties, including Property 1.

A. The Broken Telephone Property

Theorem 1: M-information flow, as given by Definition 4, satisfies Property 1.

A proof of this theorem appears in Appendix A in the full version of this paper [1].

B. The Existence of Orphans

Definition 4 also has a very non-intuitive property: an edge leading out of a node may have M-information flow, even though *no* edge leading *into* that node has M-information flow.

Definition 5 (M-information Orphan): In a computational system $\mathscr{C}$, a node V_t is said to be an *M-information orphan* if there exists some $Q_t \in \mathbb{Q}(V_t)$ that has M-information flow, but no edge $P_t \in \mathscr{P}(V_t)$ has M-information flow.

Property 2: M-information orphans may exist in a computational system.

A detailed proof is provided in Appendix B in the full version of this paper [1]. Essentially, the node C_1 in Fig. 1 is an M-information orphan, because its outgoing edge (C_1, B_2) carries M-information flow, whereas none of its incoming edges has M-information flow.

The existence of M-information orphans—as well as the presence of M-information flow on (C_1, B_2) in Counterexample 1—may not be expected, since Z was never computed from M. But closer inspection reveals that, from the perspective of B_2, Z is statistically indistinguishable from $M \oplus Z$, and is therefore just as important for recovering M.

Information flow can thus be "created" at an M-information orphan, and can also be "destroyed" at a node (e.g., by omission), so M-information flow does not obey a law of conservation at nodes. In this sense, it is not a typical kind of "flow" defined on graphs (see, for example, [41, Sec. 26.1]), and well-known results such as the Max-flow Min-cut Theorem [41, Thm. 26.6] do not apply as-is.

It is worthwhile to note at this point that the existence of M-information orphans such as C_1 in Counterexample 1 does not violate the Data Processing Inequality (DPI) [42, Ch. 2].

Property 3 (Local Markov Property): For any given subset of nodes $\mathscr{V}_t' \subseteq \mathscr{V}_t$, the following Markov Chain holds: M—$X(\mathscr{P}(\mathscr{V}_t'))$—$X(\mathbb{Q}(\mathscr{V}_t'))$.

A proof of this property appears in Appendix C in the full version [1]. Given that this property is a direct consequence of Definition 3b, it may not be very surprising. However, it is worth noting that the Local Markov Property holds *even at an M-information orphan*: even if M-information flow spontaneously emerges from a node, the Local Markov Property at that node is preserved, so the DPI is not violated.

C. The Existence of Information Paths

We now come to our main result: if the outgoing transmissions of any given node depend on the message, then we can find a path leading to that node from one or more input nodes, along which M-information flows. Before we demonstrate this property, we formally define an "M-information path".

Definition 6 (Path): In any computational system $\mathscr{C}$, suppose $\mathscr{A}$ and $\mathscr{B}$ are two disjoint sets of nodes in $\mathscr{V}$. Then, a *path* from $\mathscr{A}$ to $\mathscr{B}$ is any ordered set of nodes $\{V^{(0)}, V^{(1)}, \ldots, V^{(L)}\}$ that satisfies (i) $V^{(0)} \in \mathscr{A}$; (ii) $V^{(L)} \in \mathscr{B}$; and (iii) $(V^{(i-1)}, V^{(i)}) \in \mathscr{E}$ for every $1 \leq i \leq L$, where L is a positive integer indicating the length of the path. We refer to the set $\{(V^{(i-1)}, V^{(i)})\}_{i=1}^{L}$ as the *edges of the path*.

Definition 7 (M-Information Path): Continuing from Definition 6, we define an *M-information path* from $\mathscr{A}$ to $\mathscr{B}$ as any path from $\mathscr{A}$ to $\mathscr{B}$, each of whose edges carries M-information flow. That is, if $(V^{(i-1)}, V^{(i)}) = E_{t_i} \in \mathscr{E}_{t_i}$ for some $t_i \in \mathscr{T}$, then for every $1 \leq i \leq L$,

$$\exists\, \mathscr{E}'_{t_i} \subseteq \mathscr{E}_{t_i} \quad \text{s.t.} \quad I\big(M; X(E_{t_i}) \,\big|\, X(\mathscr{E}'_{t_i})\big) > 0. \tag{4}$$

Property 4 (Existence of an Information Path): In any computational system $\mathscr{C}$, suppose that at some time $t_{\text{op}} \in \mathscr{T}$, there is an "output node" $V_{\text{op}} \in \mathscr{V}$ whose outgoing edges $\mathbb{Q}(V_{\text{op}})$ satisfy $I\big(M; X(\mathbb{Q}(V_{\text{op}}))\big) > 0$. Then, there must exist an M-information path from the input nodes $\mathscr{V}_{\text{ip}}$ to V_{op}.

Theorem 2: Definition 4 satisfies Property 4.

A rigorous proof of this theorem is provided in Appendix D, which appears in the full version of this paper [1]. Below, we provide a brief outline.

Proof outline: The proof shows the contrapositive of the theorem: if there exists no M-information path from $\mathscr{V}_{\text{ip}}$ to V_{op}, then there must be a cut separating $\mathscr{V}_{\text{ip}}$ and V_{op}, each of whose edges has zero M-information flow (see Fig. 2 in Appendix D [1]). The proof is non-trivial principally because this cut may stretch across multiple time instants, whereas our definition of M-information flow involves edges at a single time instant. The proof, therefore, needs to rely on induction to show that none of the transmissions on the V_{op}-side of the cut can statistically depend on M. The induction itself starts with the first nodes that come after the cut (temporally) and systematically demonstrates, time-point by time-point, that all nodes to the right of the cut have outgoing transmissions that are independent of the message M. ∎

D. On the Uniqueness of Our Definition of Information Flow

From the perspective of designing an axiomatic framework, it is desirable to find a minimal set of properties that gives rise to a unique definition of information flow. Note that Property 1 does not uniquely specify a definition: a definition that sets *all* edges to have information flow (or *no* edges to have information flow) would also be consistent with this property.

In this section, we provide a set of properties that uniquely leads to our definition of information flow. However, we must acknowledge that we arrived at these properties with the benefit of hindsight. As such, they are mathematically very similar to our definition, and a more abstract set of properties that leads to a unique definition would be desirable.

Property 5: In a computational system $\mathscr{C}$, let $\mathscr{F}_M : \mathscr{E} \to \{0, 1\}$ be an indicator of the presence of information flow about M on an edge (as in Property 1). We now state three conditions $\mathscr{F}_M$ must satisfy, which naturally lead to Definition 4:

5a) $\mathscr{F}_M(E_t) = 1$ if $I\big(M; X(E_t)\big) > 0$

5b) $\mathcal{F}_M(E_t) = 1$ if $\exists\ \mathscr{C}'_t \subseteq \mathscr{C}_t \setminus \{E_t\}$ s.t.
$$I\big(M; X(\mathscr{C}'_t) \,\big|\, X(E_t)\big) > I\big(M; X(\mathscr{C}'_t)\big)$$
5c) $\mathcal{F}_M(E_t) = 0$ if $I\big(M; X(E_t) \,\big|\, X(\mathscr{C}'_t)\big) = 0\ \forall\ \mathscr{C}'_t \subseteq \mathscr{C}_t$.

In the language of the PID literature [36], Property 5a states that if an edge has unique or redundant information about M, then it must carry information flow, while Property 5b states that if an edge has synergistic information about M along with some other set of transmissions, then it must carry information flow. Finally, Property 5c states that if all three of these components are absent, then that edge carries no information flow. This also explains how, if any one of these three properties is absent, our definition is no longer unique.

Proposition 3 (Uniqueness): If $\mathcal{F}_M$ is an indicator of information flow that satisfies the conditions in Property 5, then $\mathcal{F}_M(E_t) = 1$ if and only if E_t has M-information flow, per Definition 4.

A proof of this proposition appears in Appendix E in the full version of the paper [1].

V. Discussion

Estimation of information flow can be achieved using techniques for conditional independence testing that have been established in the literature. Once edges that have information flow have been identified, one can discover all information paths using a version of the depth-first search algorithm. These ideas, along with examples of information flow in simple systems, and detailed discussions on neuroscientific issues and on the limitations of tools based on statistical causal influence, appear in the extended version of this paper [2].

References

[1] Full version of this paper with appendices. [Online]. Available: https://praveenv253.github.io/publications#Venkatesh2019Define

[2] P. Venkatesh, S. Dutta, and P. Grover, "Information flow in computational systems," *arXiv:1902.02292 [cs.IT]*, 2019.

[3] J. Almeida *et al.*, "Tool manipulation knowledge is retrieved by way of the ventral visual object processing pathway," *Cortex*, vol. 49, no. 9, pp. 2334–2344, 2013.

[4] A. Brovelli *et al.*, "Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality," *PNAS*, vol. 101, no. 26, pp. 9849–9854, 2004.

[5] M. Bar *et al.*, "Top-down facilitation of visual recognition," *PNAS*, vol. 103, no. 2, pp. 449–454, 2006.

[6] A. S. Greenberg *et al.*, "Visuotopic cortical connectivity underlying attention revealed with white-matter tractography," *J. Neurosci.*, vol. 32, no. 8, pp. 2773–2782, 2012.

[7] J. Samuels and N. D. Zasler, *Event-Related Paradigms*. Springer, 2018, pp. 1346–1347.

[8] C. Hammond *et al.*, "Pathological synchronization in Parkinson's disease: networks, models and treatments," *Trends in neurosciences*, vol. 30, no. 7, pp. 357–364, 2007.

[9] Y. Smith *et al.*, "Microcircuitry of the direct and indirect pathways of the basal ganglia," *Neuroscience*, vol. 86, no. 2, pp. 353–387, 1998.

[10] A. A. Grace, "Gating of information flow within the limbic system and the pathophysiology of schizophrenia," *Brain Research Reviews*, vol. 31, no. 2-3, pp. 330–341, 2000.

[11] E. Lalo *et al.*, "Patterns of bidirectional communication between cortex and basal ganglia during movement in patients with Parkinson disease," *Journal of Neuroscience*, vol. 28, no. 12, pp. 3008–3016, 2008.

[12] K. J. Friston, "Functional and effective connectivity: a review," *Brain connectivity*, vol. 1, no. 1, pp. 13–36, 2011.

[13] A. M. Bastos and J.-M. Schoffelen, "A tutorial review of functional connectivity analysis methods and their interpretational pitfalls," *Frontiers in systems neuroscience*, vol. 9, p. 175, 2016.

[14] S. L. Bressler and A. K. Seth, "Wiener–Granger causality: A well established methodology," *NeuroImage*, vol. 58(2), pp. 323–329, 2011.

[15] J. Massey, "Causality, feedback and directed information," in *ISITA*, 1990, pp. 303–305.

[16] C. J. Quinn, T. P. Coleman, N. Kiyavash, and N. G. Hatsopoulos, "Estimating the directed information to infer causal relationships in ensemble neural spike train recordings," *Journal of Computational Neuroscience*, vol. 30, no. 1, pp. 17–44, Feb 2011.

[17] J. Jiao, H. H. Permuter, L. Zhao, Y. Kim, and T. Weissman, "Universal estimation of directed information," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6220–6242, Oct 2013.

[18] T. Schreiber, "Measuring information transfer," *Physical Review Letters*, vol. 85, pp. 461–464, Jul 2000.

[19] L. A. Baccalá and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biological Cybernetics*, vol. 84, no. 6, pp. 463–474, May 2001.

[20] O. David *et al.*, "Identifying neural drivers with functional MRI: an electrophysiological validation," *PLoS Biology*, vol. 6(12), p. e315, 2008.

[21] A. Roebroeck *et al.*, "The identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution," *NeuroImage*, vol. 58, no. 2, pp. 296–302, 2011.

[22] O. David, "fMRI connectivity, meaning and empiricism. comments on: Roebroeck et al. The identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution." *NeuroImage*, vol. 58, no. 2, pp. 306–309, 2011.

[23] P. A. Stokes and P. L. Purdon, "A study of problems encountered in Granger causality analysis from a neuroscience perspective," *PNAS*, vol. 114, no. 34, pp. E7063–E7072, 2017.

[24] L. Barnett *et al.*, "Solved problems for Granger causality in neuroscience: A response to Stokes and Purdon," *NeuroImage*, vol. 178, pp. 744–748, 2018.

[25] L. Faes *et al.*, "On the interpretability and computational reliability of frequency-domain Granger causality," *F1000Research*, Sep 2017.

[26] D. M. A. Mehler and K. P. Kording, "The lure of causal statements: Rampant mis-inference of causality in estimated connectivity," *arXiv:1812.03363 [q-bio.NC]*, 2018.

[27] J. Pearl, *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2009.

[28] J. Andersson, "Testing for Granger causality in the presence of measurement errors," *Economics Bulletin*, 2005.

[29] M. Gong *et al.*, "Discovering temporal causal relations from subsampled data," in *ICML*, 2015, pp. 1898–1906.

[30] P. Venkatesh and P. Grover, "Is the direction of greater Granger causal influence the same as the direction of information flow?" in *Allerton*, Sept 2015, pp. 672–679.

[31] M. Kamiński *et al.*, "Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance," *Biol. Cybern.*, vol. 85, no. 2, pp. 145–157, Aug 2001.

[32] C. D. Thompson, "A complexity theory for VLSI," Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, PA, USA, 1980, aAI8100621.

[33] R. Ahlswede, N. Cai, S. Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Th.*, vol. 46, no. 4, pp. 1204–1216, July 2000.

[34] Y. Polyanskiy and Y. Wu, "Lecture notes on information theory," August 2017. [Online]. Available: http://www.stat.yale.edu/~yw562/teaching/itlectures.pdf

[35] C. E. Shannon, "Communication theory of secrecy systems," *Bell system technical journal*, vol. 28, no. 4, pp. 656–715, 1949.

[36] J. T. Lizier *et al.*, "Information decomposition of target effects from multi-source interactions: Perspectives on previous, current and future work," *Entropy*, vol. 20, no. 4, p. 307, 2018.

[37] E. Schneidman *et al.*, "Synergy, redundancy, and independence in population codes," *J. Neurosci.*, vol. 23, no. 37, pp. 11 539–11 553, 2003.

[38] P. E. Latham and S. Nirenberg, "Synergy, redundancy, and independence in population codes, revisited," *J. Neurosci.*, vol. 25, no. 21, pp. 5195–5206, 2005.

[39] N. M. Timme and C. Lapish, "A tutorial for information theory in neuroscience," *eNeuro*, vol. 5, no. 3, 2018.

[40] I. Gat and N. Tishby, "Synergy and redundancy among brain cells of behaving monkeys," in *NIPS*, 1999, pp. 111–117.

[41] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 3rd ed. The MIT Press, 2009.

[42] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2012.

Appendix A
Proof of Theorem 1

Before we prove this theorem, we prove a simpler lemma which directly falls out of Definition 4 and the properties of mutual information.

Lemma 4: No edge in $\mathscr{E}_t$ has M-information flow if and only if $X(\mathscr{E}_t)$ is independent of M. In other words,

$$I\big(M; X(E_t) \,\big|\, X(\mathscr{E}'_t)\big) = 0 \quad \forall\, E_t \in \mathscr{E}_t,\ \mathscr{E}'_t \subseteq \mathscr{E}_t \backslash \{E_t\} \tag{5}$$

if and only if

$$I\big(M; X(\mathscr{E}_t)\big) = 0. \tag{6}$$

Equivalently, $X(\mathscr{E}_t)$ *depends* on M if and only if *at least* one edge in $\mathscr{E}_t$ has non-zero M-information flow.

Proof: ($\Rightarrow$) Suppose that the condition in (5) holds. Let $\mathscr{E}_t = \big\{E_t^{(1)}, E_t^{(2)}, \ldots, E_t^{(N^2)}\big\}$ be any ordering of the edges in $\mathscr{E}_t$. Then,

$$\begin{aligned}
&I\big(M; X(\mathscr{E}_t)\big) && (7)\\
&\overset{(a)}{=} I\big(M; X(E_t^{(1)})\big) + I\big(M; X(E_t^{(2)}) \,\big|\, X(E_t^{(1)})\big) && (8)\\
&\quad + I\big(M; X(E_t^{(3)}) \,\big|\, X(E_t^{(1)}), X(E_t^{(2)})\big) + \cdots \\
&= \sum_{i=1}^{N^2} I\Big(M; X(E_t^{(i)}) \,\Big|\, \bigcup_{j=1}^{i-1} \{X(E_t^{(j)})\}\Big) && (9)\\
&\overset{(b)}{=} \sum_{i=1}^{N^2} I\Big(M; X(E_t^{(i)}) \,\Big|\, X\Big(\bigcup_{j=1}^{i-1} \{E_t^{(j)}\}\Big)\Big) \overset{(c)}{=} 0, && (10)
\end{aligned}$$

where (a) follows from the chain-rule of mutual information [42, Ch. 2], (b) is simply the application of Equation (1), and (c) follows from the fact that each term in the summation is zero, by (5). This proves the forward implication.

($\Leftarrow$) Next, suppose $I\big(M; X(\mathscr{E}_t)\big) = 0$. Let E_t be any edge in $\mathscr{E}_t$ and let $\mathscr{E}'_t$ be any subset of $\mathscr{E}_t \setminus \{E_t\}$. Also, let $\mathscr{E}''_t = \mathscr{E}_t \setminus \big(\mathscr{E}'_t \cup \{E_t\}\big)$. Then,

$$\begin{aligned}
0 &= I\big(M; X(\mathscr{E}_t)\big) && (11)\\
&= I\big(M; X(\mathscr{E}'_t)\big) + I\big(M; X(E_t) \,\big|\, X(\mathscr{E}'_t)\big) && (12)\\
&\qquad + I\big(M; X(\mathscr{E}''_t) \,\big|\, X(\mathscr{E}'_t), X(E_t)\big)
\end{aligned}$$

by the chain rule. Since (conditional) mutual information is always non-negative [42, Ch. 2], all three terms on the right hand side must be zero. So in particular,

$$I\big(M; X(E_t) \,\big|\, X(\mathscr{E}'_t)\big) = 0. \tag{13}$$

Since E_t and $\mathscr{E}'_t$ are arbitrary, this proves the converse. ∎

Proof of Theorem 1: We need to prove that M-information flow, as given by Definition 4, satisfies Property 1. Explicitly stated, we need to show that if every edge at some time t has zero M-information flow, then every edge at all future times $t' > t$ must also have zero M-information flow. So suppose that, at time t, for every $E_t \in \mathscr{E}_t$ we have

$$I\big(M; X(E_t) \,\big|\, X(\mathscr{E}'_t)\big) = 0 \qquad \forall\, \mathscr{E}'_t \subseteq \mathscr{E}_t \setminus \{E_t\}. \tag{14}$$

By Lemma 4, this implies that

$$I\big(M; X(\mathscr{E}_t)\big) = 0. \tag{15}$$

Now, consider the first future time instant, $t' = t + 1$. For every node $A_{t+1} \in \mathscr{V}_{t+1}$, the definition of computation at a node (Definition 3b) states that

$$X(\mathbb{Q}(A_{t+1})) = f_{A_{t+1}}\big(X(\mathscr{P}(A_{t+1})), W(A_{t+1})\big), \tag{16}$$

where the reader may recall, $\mathscr{P}(A_{t+1})$ and $\mathbb{Q}(A_{t+1})$ are the edges entering and leaving A_{t+1} respectively. We can collect the individual functions $f_{A_{t+1}}$ across all nodes in $\mathscr{V}_{t+1}$ into a single joint function, $f_{\mathscr{V}_{t+1}}$, to obtain

$$X(\mathscr{E}_{t+1}) = f_{\mathscr{V}_{t+1}}\big(X(\mathscr{E}_t), W(\mathscr{V}_{t+1})\big). \tag{17}$$

Therefore,

$$\begin{aligned}
0 &\le I\big(M; X(\mathscr{E}_{t+1})\big) && (18)\\
&= I\big(M; f_{\mathscr{V}_{t+1}}\big(X(\mathscr{E}_t), W(\mathscr{V}_{t+1})\big)\big) && (19)\\
&\overset{(a)}{\le} I\big(M; X(\mathscr{E}_t), W(\mathscr{V}_{t+1})\big) && (20)\\
&= I\big(M; X(\mathscr{E}_t)\big) + I\big(M; W(\mathscr{V}_{t+1}) \,\big|\, X(\mathscr{E}_t)\big) && (21)\\
&\overset{(b)}{=} I\big(M; X(\mathscr{E}_t)\big) \overset{(c)}{=} 0, && (22)
\end{aligned}$$

where (a) follows from the DPI, (b) follows from the fact that $W(\mathscr{V}_{t+1}) \perp\!\!\!\perp \{M, X(\mathscr{E}_t)\}$, and (c) follows from (15). Once again, by non-negativity of mutual information we must have that $I\big(M; X(\mathscr{E}_{t+1})\big) = 0$. Applying Lemma 4 once again, we find that for $t' = t + 1$,

$$I\big(M; X(E_{t'}) \,\big|\, X(\mathscr{E}'_{t'})\big) = 0 \quad \forall\, E_{t'} \in \mathscr{E}_{t'}, \mathscr{E}'_{t'} \subseteq \mathscr{E}_{t'} \backslash \{E_{t'}\} \tag{23}$$

We have shown that (14) implies (23), hence, induction on t' yields that (23) holds for all future times $t' > t$, completing the proof. ∎

Appendix B
Proof of Property 2

Proof: Consider the computational system in Fig. 1 from Counterexample 1. The node C_1 is an M-information orphan, since the edge (C_1, B_2) carries M-information flow, whereas none of its incoming edges carries M-information flow. To see this, first consider the incoming edge (C_0, C_1):

$$\begin{aligned}
&I\big(M; X((C_0, C_1))\big) = I(M; Z) = 0, && (24)\\
&I\big(M; X((C_0, C_1)) \,\big|\, X((C_0, A_1))\big) = I(M; Z \,|\, Z) = 0,\\
&I\big(M; X((C_0, C_1)) \,\big|\, X((A_0, A_1))\big) = I(M; Z \,|\, M) = 0,\\
&I\big(M; X((C_0, C_1)) \,\big|\, X((C_0, A_1)), X((A_0, A_1))\big)\\
&\qquad\qquad = I(M; Z \,|\, Z, M) = 0.
\end{aligned}$$

Thus, (C_0, C_1) has no M-information flow. Next, consider (C_1, B_2):

$$I\big(M; X((C_1, B_2)) \,\big|\, X((A_1, B_2))\big) = I(M; Z \,|\, M \oplus Z) = 1. \tag{25}$$

This implies that $X((C_1, B_2))$ *has* M-information flow. Hence, by Definition 5, since C_1 has no M-information flow on its incoming edges, but has M-information flow on one of its outgoing edges, C_1 is an M-information orphan. ∎

Appendix C
Proof of Property 3

Proof: Since $X(\mathbb{Q}(\mathscr{V}_t')) = f_{\mathscr{V}_t'}\big(X(\mathbb{P}(\mathscr{V}_t')), W(\mathscr{V}_t')\big)$ by Definition 3b, the tuple $\big(X(\mathbb{P}(\mathscr{V}_t')), X(\mathbb{Q}(\mathscr{V}_t'))\big)$ is also a function of $X(\mathbb{P}(\mathscr{V}_t'))$ and $X(W(\mathscr{V}_t'))$. Hence, the following Markov chain holds:

$$M—\big(X(\mathbb{P}(\mathscr{V}_t')), W(\mathscr{V}_t')\big)—\big(X(\mathbb{P}(\mathscr{V}_t')), X(\mathbb{Q}(\mathscr{V}_t'))\big).$$

By the DPI, this implies that

$$\begin{aligned}
&I\big(M; X(\mathbb{Q}(\mathscr{V}_t')), X(\mathbb{P}(\mathscr{V}_t'))\big) && (26)\\
&\le I\big(M; X(\mathbb{P}(\mathscr{V}_t')), W(\mathscr{V}_t')\big) && (27)\\
&\overset{(a)}{=} I\big(M; X(\mathbb{P}(\mathscr{V}_t'))\big) + I\big(M; W(\mathscr{V}_t') \,\big|\, X(\mathbb{P}(\mathscr{V}_t'))\big) && (28)\\
&\overset{(b)}{=} I\big(M; X(\mathbb{P}(\mathscr{V}_t'))\big) + I\big(W(\mathscr{V}_t'); M, X(\mathbb{P}(\mathscr{V}_t'))\big) && (29)\\
&\qquad - I\big(W(\mathscr{V}_t'); X(\mathbb{P}(\mathscr{V}_t'))\big)\\
&\overset{(c)}{=} I\big(M; X(\mathbb{P}(\mathscr{V}_t'))\big) + 0 - 0, && (30)
\end{aligned}$$

where in (a) and (b), we have used the chain rule of mutual information in two different ways, and in (c) we have used the fact that $W(\mathscr{V}_t') \perp\!\!\!\perp \{M, X(\mathbb{P}(\mathscr{V}_t'))\}$. Therefore,

$$I\big(M; X(\mathbb{Q}(\mathscr{V}_t')) \,\big|\, X(\mathbb{P}(\mathscr{V}_t'))\big) = 0, \tag{31}$$

which implies the Markov chain in Property 3. ∎

Appendix D
Proof of Theorem 2

Before we proceed to the proof of Theorem 2, it is necessary to define a few quantities.

The definition of M-information flow for a single edge naturally generalizes to one for a set of edges, at a given time.

Definition 8 (M-information Flow on a Set of Edges): We say that information about the message M flows on a set of edges $\mathscr{E}_t' \subseteq \mathscr{E}_t$ if

$$\exists\, \mathscr{R}_t' \subseteq \mathscr{E}_t \quad \text{s.t.} \quad I\big(M; X(\mathscr{E}_t') \,\big|\, X(\mathscr{R}_t')\big) > 0. \tag{32}$$

The definition of M-information flow on a set of edges is nearly identical to its single-edge counterpart. Indeed, they are closely related, as the following proposition shows.

Proposition 5: A set $\mathscr{E}_t' \subseteq \mathscr{E}_t$ has M-information flow if and only if there exists an edge $E_t' \in \mathscr{E}_t'$ that has M-information flow.

Proof: ($\Rightarrow$) Suppose there exists some $E_t' \in \mathscr{E}_t'$ that has M-information flow. That is,

$$\exists\, \mathscr{E}_t'' \subseteq \mathscr{E}_t \setminus \{E_t'\} \quad \text{s.t.} \quad I\big(M; X(E_t') \,\big|\, X(\mathscr{E}_t'')\big) > 0. \tag{33}$$

Then,

$$\begin{aligned}
&I\big(M; X(\mathscr{E}_t') \,\big|\, X(\mathscr{E}_t'')\big) && (34)\\
&\quad = I\big(M; X(E_t') \,\big|\, X(\mathscr{E}_t'')\big)\\
&\qquad + I\big(M; X(\mathscr{E}_t' \setminus \{E_t'\}) \,\big|\, X(\mathscr{E}_t''), X(E_t')\big) && (35)\\
&\quad \overset{(a)}{\ge} I\big(M; X(E_t') \,\big|\, X(\mathscr{E}_t'')\big) \overset{(b)}{>} 0 && (36)
\end{aligned}$$

where (a) follows from the non-negativity of conditional mutual information and (b) from (33). Taking $\mathscr{R}_t' \coloneqq \mathscr{E}_t''$ in Definition 8, we see that set $\mathscr{E}_t'$ has M-information flow.

($\Leftarrow$) Next, suppose that the set $\mathscr{E}_t'$ has M-information flow, as per Definition 8. That is, there exists a set $\mathscr{R}_t' \subseteq \mathscr{E}_t$ such that

$$I\big(M; X(\mathscr{E}_t') \,\big|\, X(\mathscr{R}_t')\big) > 0. \tag{37}$$

Also, let $\{E_t^{(1)}, E_t^{(2)}, \dots E_t^{(K)}\}$ be any ordering of the nodes in $\mathscr{E}_t'$ (where $K = |\mathscr{E}_t'|$). Then by the chain rule of mutual information,

$$\begin{aligned}
&0 < I\big(M; X(\mathscr{E}_t') \,\big|\, X(\mathscr{R}_t')\big) && (38)\\
&\quad = \sum_{k=1}^{K} I\Bigg(M; X(E_t^{(k)}) \,\Bigg|\, X(\mathscr{R}_t'), X\Big(\bigcup_{j=1}^{k-1}\{E_t^{(j)}\}\Big)\Bigg). && (39)
\end{aligned}$$

By the non-negativity of conditional mutual information, at least one of the terms in the summation must be strictly positive. Let the index of this term be k^*. Hence, there exists $E_t' \coloneqq E_t^{(k^*)}$ and $\mathscr{E}_t'' \coloneqq \mathscr{R}_t' \cup \{E_t^{(1)}, \dots E_t^{(k^*-1)}\}$, such that

$$I\big(M; X(E_t') \,\big|\, X(\mathscr{E}_t'')\big) > 0. \tag{40}$$

In other words, there exists an edge $E_t' \in \mathscr{E}_t'$ that has M-information flow. ∎

Next, we formally define the counterpart of an M-information path, namely, a zero–M-information cut.

Definition 9 (Cut): In any computational system $\mathscr{C}$, suppose $\mathscr{A}$ and $\mathscr{B}$ are two disjoint sets of nodes in $\mathscr{V}$. Then, a *cut* separating $\mathscr{A}$ and $\mathscr{B}$ is any pair of sets $(\mathscr{V}^{\text{src}}, \mathscr{V}^{\text{sink}})$, such that (i) $\mathscr{V}^{\text{src}} \cup \mathscr{V}^{\text{sink}} = \mathscr{V}$; (ii) $\mathscr{V}^{\text{src}} \cap \mathscr{V}^{\text{sink}} = \emptyset$; (iii) $\mathscr{A} \subseteq \mathscr{V}^{\text{src}}$; and (iv) $\mathscr{B} \subseteq \mathscr{V}^{\text{sink}}$. We refer to the set of edges *going from* $\mathscr{V}^{src}$ *to* $\mathscr{V}^{sink}$, i.e. $\mathscr{E} \cap (\mathscr{V}^{\text{src}} \times \mathscr{V}^{\text{sink}})$, as the *edges in the cut set*.[5]

Definition 10 (Zero–M-information Cut): Continuing from Definition 9, we say that a cut $(\mathscr{V}^{\text{src}}, \mathscr{V}^{\text{sink}})$ is a *zero–M-information cut* if every edge in its cut set has zero M-information flow. That is, for every $E_t \in \mathscr{E} \cap (\mathscr{V}^{\text{src}} \times \mathscr{V}^{\text{sink}})$,

$$I\big(M; X(E_t) \,\big|\, X(\mathscr{E}_t')\big) = 0 \quad \forall\, \mathscr{E}_t' \subseteq \mathscr{E}_t \setminus \{E_t\}. \tag{41}$$

Remark. In Definition 10, we require that (41) hold for every edge E_t in $\mathscr{E} \cap (\mathscr{V}^{\text{src}} \times \mathscr{V}^{\text{sink}})$. However, the edges in this set may belong to several different time points, since the cut is not restricted to any particular time (e.g., see Fig. 2). The time t used in (41), therefore, is determined by the time of the edge E_t, and varies for each E_t that we check in $\mathscr{E} \cap (\mathscr{V}^{\text{src}} \times \mathscr{V}^{\text{sink}})$.

Proof outline: We shall prove the contrapositive of the theorem, i.e., we will show that if there exists no M-information path from $\mathscr{V}_{\text{ip}}$ to V_{op}, then the outgoing transmissions of V_{op} are independent of M. We first connect the absence of any M-information path with the presence of a zero–M-information cut. This is achieved in Lemma 6, which we present before the proof of Theorem 2.

[5]Note that it is not necessary for us to assume that, individually, $\mathscr{V}^{\text{src}}$ and $\mathscr{V}^{\text{sink}}$ are *connected* sets of nodes. For instance, there may be an isolated subset of $\mathscr{V}^{\text{sink}}$, surrounded only by nodes in $\mathscr{V}^{\text{src}}$. Our theorems and proofs remain unaffected, even in such a scenario.

The proof itself proceeds by induction over time. We divide the proof into two steps: initialization and continuation. Starting with the first nodes that come after the cut (temporally) in the initialization step, we systematically show that all nodes to the right of the cut have outgoing transmissions that are independent of the message M through induction. In this proof outline, we show these steps intuitively using Fig. 2, where the dashed black line denotes the cut.

Initialization. Here, node C_1 is the first node to the right of the cut, and all of its incoming edges must come from across the cut (depicted by lines in red). Because the cut is a zero–M-information cut, none of its incoming transmissions have M-information flow. Furthermore, the intrinsically generated random variable $W(C_1)$ is independent of M. Using these two facts along with the DPI, we can show that the transmissions on C_1's outgoing edges, $X(\mathbb{Q}(C_1))$, are also independent of M.

Continuation. At the second time instant to the right of the cut, nodes B_2 and C_2 receive their incoming transmissions from either C_1 (shown in orange) or from across the cut (shown in blue). Once again, the transmissions coming from across the cut can have no information flow, and we have shown that the transmissions coming from C_1 are independent of M. Also, $W(B_2)$ and $W(C_2)$ are independent of M and all incoming transmissions. This suffices to show that the outgoing transmissions of B_2 and C_2, $X\big(\mathbb{Q}(B_2) \cup \mathbb{Q}(C_2)\big)$, are independent of M. Applying this argument repeatedly over time shows that the transmissions of all nodes to the right of the cut are independent of M.

Therefore, if there is a node V_{op} whose outputs depend on M, we can be assured that there exists no zero–M-information cut separating $\mathcal{V}_{\text{ip}}$ from V_{op}. Therefore, by Lemma 6, there exists an M-information path from $\mathcal{V}_{\text{ip}}$ to V_{op}. □

A few nuances are omitted in this outline, such as how the definition of $\mathcal{V}_{\text{ip}}$ plays a role precisely. These subtleties are better elucidated in the full proof.

Before proceeding to the formal proof of Theorem 2, we first state and prove the lemma we alluded to earlier, which shows how the absence of an M-information path implies the presence of a zero–M-information cut, and vice versa.

Lemma 6: Let $\mathcal{A}$ and $\mathcal{B}$ be two disjoint sets of nodes in the computational system $\mathcal{C}$. There exists no M-information path from $\mathcal{A}$ to $\mathcal{B}$ if and only if there is a zero–M-information cut separating $\mathcal{A}$ and $\mathcal{B}$.

Proof: ($\Rightarrow$) Suppose there exists no M-information path from $\mathcal{A}$ to $\mathcal{B}$. Consider the set of all nodes to which there exists at least one M-information path from $\mathcal{A}$. Let $\mathcal{V}^{\text{src}}$ be the collection of all such nodes, along with the nodes in $\mathcal{A}$, i.e.,

$$\mathcal{V}^{\text{src}} := \mathcal{A} \cup \{V_t \in \mathcal{V} : \exists \text{ an } M\text{-information path from } \mathcal{A} \text{ to } V_t\}. \tag{42}$$

Let $\mathcal{V}^{\text{sink}} = \mathcal{V} \setminus \mathcal{V}^{\text{src}}$, so that $\mathcal{V}^{\text{sink}}$ consists of nodes to which there is no M-information path from $\mathcal{A}$. Then, we must have $\mathcal{B} \subseteq \mathcal{V}^{\text{sink}}$, since it is known that there are no M-information paths from $\mathcal{A}$ to $\mathcal{B}$. Therefore, $(\mathcal{V}^{\text{src}}, \mathcal{V}^{\text{sink}})$ is a cut that separates $\mathcal{A}$ and $\mathcal{B}$, such that no edge in the cut set has M-information flow. In other words, by Definition 10, this is a zero–M-information cut separating $\mathcal{A}$ and $\mathcal{B}$.

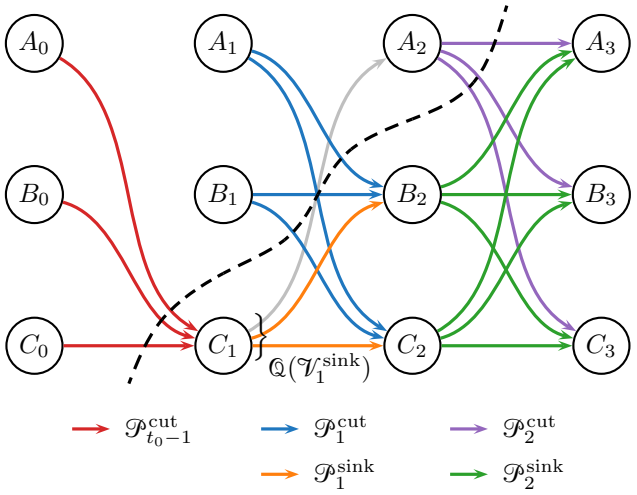

Figure 2: A generic computational system used in the proof outline and to explain certain steps in the proof of Theorem 2. For the purposes of the proof outline, it suffices to note that the black dashed line denotes the cut. All variable names can be ignored at this point of time.
For the purposes of the formal proof, note that in this figure, $\mathcal{E}^{\text{cut}}$ is essentially the union of the red, blue and purple edges, while $\mathcal{E}^{\text{sink}}$ is the union of the orange and green edges. From this, it is evident that $\mathcal{P}(\mathcal{V}_t^{\text{sink}}) = \mathcal{P}_{t-1}^{\text{cut}} \cup \mathcal{P}_{t-1}^{\text{sink}}$ for any time t, i.e., the incoming edges of $\mathcal{V}^{\text{sink}}$ at time t must either come from nodes in $\mathcal{V}^{\text{sink}}$ or from nodes across the cut. Secondly, it should be clear that $\mathcal{P}_{t-1}^{\text{sink}} = \mathbb{Q}(\mathcal{V}_{t-1}^{\text{sink}}) \cap \mathcal{E}^{\text{sink}}$, i.e., the *incoming* edges of $\mathcal{V}_t^{\text{sink}}$ that originate from nodes in $\mathcal{V}^{\text{sink}}$ are simply the *outgoing* edges of $\mathcal{V}_{t-1}^{\text{sink}}$ which terminate at nodes in $\mathcal{V}^{\text{sink}}$. This is seen best at time $t = 1$ in the graph above, where the orange and grey lines together represent $\mathbb{Q}(\mathcal{V}_1^{\text{sink}})$, the orange and green edges together make up $\mathcal{E}^{\text{sink}}$, and $\mathcal{P}_1^{\text{sink}}$ is given by the orange edges, which is the intersection of the two sets.

($\Leftarrow$) Next, suppose that there *is* an M-information path $\{V^{(i)}\}_{i=0}^{L}$ from $\mathcal{A}$ to $\mathcal{B}$. Then, we claim that there can exist no zero–M-information cut separating $\mathcal{A}$ and $\mathcal{B}$. Let $(\mathcal{V}^{\text{src}}, \mathcal{V}^{\text{sink}})$ be any cut separating $\mathcal{A}$ and $\mathcal{B}$. By Definition 6, we must have $V^{(0)} \in \mathcal{V}^{\text{src}}$ and $V^{(L)} \in \mathcal{V}^{\text{sink}}$. So, there must be at least one edge going from $\mathcal{V}^{\text{src}}$ to $\mathcal{V}^{\text{sink}}$ which lies on the path. This implies that at least one edge in the cut set carries M-information flow. Since the conditions of Definition 10 are not satisfied, this cut is *not* a zero–M-information cut. Finally, since this is true for every cut separating $\mathcal{A}$ and $\mathcal{B}$, the claim holds. ■

Proof of Theorem 2: As mentioned in the proof outline, we prove the contrapositive of the theorem. Suppose there exists no M-information path from the input nodes $\mathcal{V}_{\text{ip}}$ to V_{op}. Then, by Lemma 6, there exists a zero–M-information cut separating $\mathcal{V}_{\text{ip}}$ and V_{op}. We use this to prove that the transmissions of V_{op} are independent of M.

Setup. Let the cut separating $\mathcal{V}_{\text{ip}}$ and V_{op} be given by $(\mathcal{V}^{\text{src}}, \mathcal{V}^{\text{sink}})$, so that $\mathcal{V}_{\text{ip}} \subseteq \mathcal{V}^{\text{src}}$ and $V_{\text{op}} \in \mathcal{V}^{\text{sink}}$. Then, the cut divides $\mathcal{E}$ into the following sets: $\mathcal{E}^{\text{src}} = \mathcal{E} \cap (\mathcal{V}^{\text{src}} \times \mathcal{V}^{\text{src}})$, the edges between the nodes in $\mathcal{V}^{\text{src}}$; $\mathcal{E}^{\text{sink}} = \mathcal{E} \cap (\mathcal{V}^{\text{sink}} \times \mathcal{V}^{\text{sink}})$, the edges between nodes in $\mathcal{V}^{\text{sink}}$; and $\mathcal{E}^{\text{cut}} = \mathcal{E} \cap (\mathcal{V}^{\text{src}} \times \mathcal{V}^{\text{sink}})$, the edges going from $\mathcal{V}^{\text{src}}$ to $\mathcal{V}^{\text{sink}}$ (the edges going from $\mathcal{V}^{\text{sink}}$ to $\mathcal{V}^{\text{src}}$ will not be relevant

to our discussion). From the previous paragraph, Lemma 6 implies that $(\mathcal{V}^{\text{src}}, \mathcal{V}^{\text{sink}})$ is a zero–M-information cut, so by Definition 10, we have that for all $E_t \in \mathcal{E}^{\text{cut}}$,

$$I\big(M; X(E_t) \,\big|\, X(\mathcal{E}'_t)\big) = 0 \quad \forall\, \mathcal{E}'_t \subseteq \mathcal{E}_t \setminus \{E_t\}. \tag{43}$$

Note that the edges in $\mathcal{E}^{\text{cut}}$ may belong to different time instants. In particular, the time instant t in the equation above corresponds to the time of the edge E_t, whose flow is in question.[6]

Order the nodes in $\mathcal{V}^{\text{sink}}$ by time, and let $\mathcal{V}_t^{\text{sink}}$ be the subset of nodes in $\mathcal{V}^{\text{sink}}$ at time t. Let $\mathcal{P}(\mathcal{V}_t^{\text{sink}})$ and $\mathbb{Q}(\mathcal{V}_t^{\text{sink}})$ respectively be the sets of edges *collectively* entering and leaving all nodes in $\mathcal{V}_t^{\text{sink}}$. We shall prove that the outgoing transmissions of every node in $\mathcal{V}^{\text{sink}}$, including those of V_{op}, must be independent of the message, i.e.,

$$I\big(M; X(\mathbb{Q}(V))\big) = 0 \quad \forall\, V \in \mathcal{V}^{\text{sink}}. \tag{44}$$

Initialization. Let t_0 be the first time instant t for which $\mathcal{V}_t^{\text{sink}}$ is non-empty. Then, we encounter two cases: either $t_0 = 0$, in which case the nodes in $\mathcal{V}_{t_0}^{\text{sink}}$ have *no* incoming edges, or $t_0 > 0$, and the nodes in $\mathcal{V}_{t_0}^{\text{sink}}$ *have* incoming edges. We shall first prove that in *both* cases, the outgoing transmissions of $\mathcal{V}_{t_0}^{\text{sink}}$ are independent of the message, i.e. $I\big(M; X(\mathbb{Q}(\mathcal{V}_{t_0}^{\text{sink}}))\big) = 0$.

(Case I) When $t_0 = 0$, $\mathcal{V}_0^{\text{sink}} \cap \mathcal{V}_{\text{ip}} = \emptyset$. This is because the cut separates $\mathcal{V}_{\text{ip}}$ from V_{op}, with $\mathcal{V}_{\text{ip}} \subseteq \mathcal{V}^{\text{src}}$, so no nodes in $\mathcal{V}_0^{\text{sink}}$ can be input nodes. So, by the definition of (non-)input nodes (Definition 3c), we must have

$$\begin{aligned} I\big(M; X(\mathbb{Q}(\mathcal{V}_0^{\text{sink}}))\big) &= I\big(M; f_{\mathcal{V}_0^{\text{sink}}}(W(\mathcal{V}_0^{\text{sink}}))\big) && (45)\\ &\overset{(a)}{\leq} I\big(M; W(\mathcal{V}_0^{\text{sink}})\big) && (46)\\ &\overset{(b)}{=} 0, && (47)\end{aligned}$$

where step (a) uses the DPI and step (b) makes use of the fact that $W(\mathcal{V}_0) \perp\!\!\!\perp M$.

(Case II) When $t_0 > 0$, the definition of t_0 implies that all nodes at time $t_0 - 1$ are in $\mathcal{V}^{\text{src}}$, so all incoming edges of $\mathcal{V}_{t_0}^{\text{sink}}$ must lie in the cut set, i.e., $\mathcal{P}(\mathcal{V}_{t_0}^{\text{sink}}) \subseteq \mathcal{E}^{\text{cut}}$. Since the cut is a zero–M-information cut, we have that for all $E_{t_0-1} \in \mathcal{P}(\mathcal{V}_{t_0}^{\text{sink}})$,

$$I\big(M; X(E_{t_0-1}) \,\big|\, X(\mathcal{E}'_{t_0-1})\big) = 0 \quad \forall\, \mathcal{E}'_{t_0-1} \subseteq \mathcal{E}_{t_0-1}. \tag{48}$$

By the definition of M-information flow for a set of edges (Definition 8) and Proposition 5, we have

$$I\big(M; X(\mathcal{P}(\mathcal{V}_{t_0}^{\text{sink}})) \,\big|\, X(\mathcal{E}'_{t_0-1})\big) = 0 \quad \forall\, \mathcal{E}'_{t_0-1} \subseteq \mathcal{E}_{t_0-1}. \tag{49}$$

[6] In fact, this is one of the central factors that prevents us from recursively applying the DPI at every node, leading from $\mathcal{V}_{\text{ip}}$ to V_{op}.

Once again, considering $\mathbb{Q}(\mathcal{V}_{t_0}^{\text{sink}})$, we have

$$\begin{aligned} &I\big(M; X(\mathbb{Q}(\mathcal{V}_{t_0}^{\text{sink}}))\big) && (50)\\ &\quad= I\big(M; f_{\mathcal{V}_{t_0}^{\text{sink}}}\big(X(\mathcal{P}(\mathcal{V}_{t_0}^{\text{sink}})), W(\mathcal{V}_{t_0}^{\text{sink}})\big)\big) && (51)\\ &\quad\overset{(a)}{\leq} I\big(M; X(\mathcal{P}(\mathcal{V}_{t_0}^{\text{sink}})), W(\mathcal{V}_{t_0}^{\text{sink}})\big) && (52)\\ &\quad\overset{(b)}{=} I\big(M; X(\mathcal{P}(\mathcal{V}_{t_0}^{\text{sink}}))\big) \\ &\qquad + I\big(M; W(\mathcal{V}_{t_0}^{\text{sink}}) \,\big|\, X(\mathcal{P}(\mathcal{V}_{t_0}^{\text{sink}}))\big) && (53)\\ &\quad\overset{(c)}{=} 0, && (54)\end{aligned}$$

where (a) and (b) follow from the DPI and the chain rule of mutual information respectively. In step (c), the first expression in the sum goes to zero by taking $\mathcal{E}_{t_0-1} = \emptyset$ in (49) and the second expression is zero since $W(\mathcal{V}_{t_0}^{\text{sink}}) \perp\!\!\!\perp \{M, X(\mathcal{E}_{t_0-1})\}$, and $\mathcal{P}(\mathcal{V}_{t_0}^{\text{sink}}) \subseteq \mathcal{E}_{t_0-1}$ (refer Definition 3b). So, from equations (47) and (54), we have that for all values of t_0,

$$I\big(M; X(\mathbb{Q}(\mathcal{V}_{t_0}^{\text{sink}}))\big) = 0. \tag{55}$$

Continuation. Now, suppose that for some $t > t_0$, we have $I\big(M; X(\mathbb{Q}(\mathcal{V}_{t-1}^{\text{sink}}))\big) = 0$. We shall prove that this implies $I\big(M; X(\mathbb{Q}(\mathcal{V}_t^{\text{sink}}))\big) = 0$. First, observe that

$$\mathcal{P}(\mathcal{V}_t^{\text{sink}}) = (\mathcal{P}(\mathcal{V}_t^{\text{sink}}) \cap \mathcal{E}^{\text{cut}}) \cup (\mathcal{P}(\mathcal{V}_t^{\text{sink}}) \cap \mathcal{E}^{\text{sink}}) \tag{56}$$

For convenience, let $\mathcal{P}_{t-1}^{\text{cut}} := \mathcal{P}(\mathcal{V}_t^{\text{sink}}) \cap \mathcal{E}^{\text{cut}}$ and $\mathcal{P}_{t-1}^{\text{sink}} := \mathcal{P}(\mathcal{V}_t^{\text{sink}}) \cap \mathcal{E}^{\text{sink}}$. We have used the subscript $t-1$ here to remind the reader that $\mathcal{P}(\mathcal{V}_t^{\text{sink}})$, which are the *incoming* edges of $\mathcal{V}_t^{\text{sink}}$, are a subset of $\mathcal{E}_{t-1}$. Then, we have

$$\mathcal{P}(\mathcal{V}_t^{\text{sink}}) = \mathcal{P}_{t-1}^{\text{cut}} \cup \mathcal{P}_{t-1}^{\text{sink}}. \tag{57}$$

Since the cut is a zero–M-information cut, we have that for every $E_{t-1} \in \mathcal{P}_{t-1}^{\text{cut}}$,

$$I\big(M; X(E_{t-1}) \,\big|\, X(\mathcal{E}'_{t-1})\big) = 0 \quad \forall\, \mathcal{E}'_{t-1} \subseteq \mathcal{E}_{t-1}. \tag{58}$$

Therefore, by Definition 8 and Proposition 5,

$$I\big(M; X(\mathcal{P}_{t-1}^{\text{cut}}) \,\big|\, X(\mathcal{E}'_{t-1})\big) = 0 \quad \forall\, \mathcal{E}'_{t-1} \subseteq \mathcal{E}_{t-1}. \tag{59}$$

Secondly, $\mathcal{P}_{t-1}^{\text{sink}} = \mathbb{Q}(\mathcal{V}_{t-1}^{\text{sink}}) \cap \mathcal{E}^{\text{sink}}$. This is depicted in Fig. 2, and explained in the caption. So,

$$\begin{aligned} I\big(M; X(\mathcal{P}_{t-1}^{\text{sink}})\big) &= I\big(M; X(\mathbb{Q}(\mathcal{V}_{t-1}^{\text{sink}}) \cap \mathcal{E}^{\text{sink}})\big) && (60)\\ &\overset{(a)}{\leq} I\big(M; X(\mathbb{Q}(\mathcal{V}_{t-1}^{\text{sink}}))\big) \overset{(b)}{=} 0 && (61)\end{aligned}$$

where (a) follows from the fact that considering more random variables can only increase mutual information, and (b) follows from the induction assumption. Finally, consider how $X(\mathbb{Q}(\mathcal{V}_t^{\text{sink}}))$ depends on M:

$$\begin{aligned} &I\big(M; X(\mathbb{Q}(\mathcal{V}_t^{\text{sink}}))\big) && (62)\\ &\quad= I\big(M; f_{\mathcal{V}_t^{\text{sink}}}\big(X(\mathcal{P}_{t-1}^{\text{sink}} \cup \mathcal{P}_{t-1}^{\text{cut}}), W(\mathcal{V}_t^{\text{sink}})\big)\big) && (63)\\ &\quad\overset{(a)}{\leq} I\big(M; X(\mathcal{P}_{t-1}^{\text{sink}}), X(\mathcal{P}_{t-1}^{\text{cut}}), W(\mathcal{V}_t^{\text{sink}})\big) && (64)\\ &\quad\overset{(b)}{=} I\big(M; X(\mathcal{P}_{t-1}^{\text{sink}})\big) + I\big(M; X(\mathcal{P}_{t-1}^{\text{cut}}) \,\big|\, X(\mathcal{P}_{t-1}^{\text{sink}})\big) \\ &\qquad + I\big(M; W(\mathcal{V}_t^{\text{sink}}) \,\big|\, X(\mathcal{P}_{t-1}^{\text{sink}}), X(\mathcal{P}_{t-1}^{\text{cut}})\big) && (65)\\ &\quad\overset{(c)}{=} 0, && (66)\end{aligned}$$

where once again, (a) and (b) follow from the DPI and the chain rule respectively. In step (c), the first and second terms go to zero by equations (61) and (59) respectively, while the third term is zero since $W(\mathscr{V}_t^{\text{sink}}) \perp\!\!\!\perp \{M, X(\mathscr{E}_{t-1})\}$ and $\mathscr{P}_{t-1}^{\text{sink}} \cup \mathscr{P}_{t-1}^{\text{cut}} \subseteq \mathscr{E}_{t-1}$.

The proof follows from induction on t, so

$$I\big(M; X(\mathbb{Q}(\mathscr{V}_t^{\text{sink}}))\big) = 0 \quad \forall\, t \geq t_0, \tag{67}$$

which in turn implies that

$$I\big(M; X(\mathbb{Q}(V))\big) = 0 \quad \forall\, V \in \mathscr{V}^{\text{sink}}. \tag{68}$$

If there exists an output node whose transmissions depend on M, then there can exist no cut consisting of edges with zero M-information flow, and hence by Lemma 6, there must be a path consisting of edges that carry M-information flow between the input nodes and the output node in question. ■

APPENDIX E
PROOF OF PROPOSITION 3

Proof: ($\Rightarrow$) Suppose the edge E_t has no M-information flow per Definition 4. This directly implies the condition in Property 5c. Hence, $\mathscr{F}_M(E_t) = 0$. This proves that if $\mathscr{F}_M(E_t) = 1$, the edge E_t must have M-information flow.

($\Leftarrow$) Suppose the edge E_t *has* M-information flow per Definition 4. Then,

$$\exists\, \mathscr{E}_t' \subseteq \mathscr{E}_t \backslash \{E_t\} \quad \text{s.t} \quad I\big(M; X(E_t) \,\big|\, X(\mathscr{E}_t')\big) > 0. \tag{69}$$

If $\mathscr{E}_t' = \emptyset$, $I\big(M; X(E_t)\big) > 0$, so by Property 5a, $\mathscr{F}_M(E_t) = 1$. If $I\big(M; X(E_t)\big) = 0$, then (69) guarantees the existence of some $\mathscr{E}_t' \neq \emptyset$ such that

$$I\big(M; X(E_t) \,\big|\, X(\mathscr{E}_t')\big) > 0. \tag{70}$$

Hence,

$$\begin{aligned}
&I\big(M; X(\mathscr{E}_t')\big) && (71)\\
&\overset{(a)}{<} I\big(M; X(\mathscr{E}_t')\big) + I\big(M; X(E_t) \,\big|\, X(\mathscr{E}_t')\big) && (72)\\
&\overset{(b)}{<} I\big(M; X(E_t), X(\mathscr{E}_t')\big) && (73)\\
&\overset{(c)}{<} I\big(M; X(E_t)\big) + I\big(M; X(\mathscr{E}_t') \,\big|\, X(E_t)\big) && (74)\\
&\overset{(d)}{<} I\big(M; X(\mathscr{E}_t') \,\big|\, X(E_t)\big), && (75)
\end{aligned}$$

where in (a), we simply added $I\big(M; X(\mathscr{E}_t')\big)$ to both sides; in (b) and (c), we used the chain rule in two different ways; and in (d), we used the fact that $I\big(M; X(E_t)\big) = 0$. So, by Property 5b, we have that $\mathscr{F}_M(E_t) = 1$. This proves the converse. ■

Remark. It should be noted that Definition 4 only specifies *whether or not* a given edge has M-information flow. It does not *quantify* this flow. So Proposition 3 demonstrates the uniqueness of our definition up to an unspecified information volume. If we require that the conditions in Property 5 hold, then any quantitative definition of information flow will go to zero at an edge if and only if the M-information flow carried by that edge is zero.