Understanding Neural Population Communication with Latent Channels

Amanda Merkley and Pulkit Grover
Department of Electrical & Computer Engineering
Carnegie Mellon University, Pittsburgh, USA
{amerkley, pgrover}@andrew.cmu.edu

***Abstract*—Experiments in neuroscience often aim to understand how information about a message, e.g., an experimental stimulus, is communicated between neural populations, where each population can comprise millions of neurons. In such high-dimensional systems, dimension reduction is imperative for isolating message-relevant communication. Here, we address the problem of defining and characterizing latent representations of population-level communication. We introduce the concept of a latent channel (LC) to capture statistical communication between latent activity of two neural populations. Through simple, illustrative examples of population communication, we show how four kinds of LCs emerge from message-specific interactions between populations. We then propose an algorithm for identifying 'valid' LCs that optimally preserve information about the message. Finally, we use Partial Information Decomposition, a framework for describing how two populations interact with respect to the message, to provide insight into how LCs preserve and transform pieces of message-relevant information from the original, high-dimensional system. Through definitions and examples, we provide a principled approach for understanding latent representations of neural population communication.**

***Index Terms*—dimension reduction, communication, neuroscience**

I. Introduction

It is increasingly common in neuroscience experiments to probe the activity of *multiple* neural populations in response to a stimulus. Because each neural population is a group of anatomically localized neurons, numbering several orders of magnitude, these experimental paradigms can provide insight into how distinct mechanisms across multiple populations are jointly activated to produce a desirable behavior. A natural question of communication arises in these experiments: how does a neural population communicate information about a message (e.g., the stimulus) to another neural population?

To frame population communication concretely, we leverage the concept of 'M-forwarding' that formally defines one notion of statistical communication for neural populations [1]. Consider how one population $\mathbf{X}$ at time t_1 communicates to another population $\mathbf{Y}$ at time t_2. Then, $\mathbf{X}^{(t_1)}$ is said to forward the message $\mathbf{M}$ to $\mathbf{Y}^{(t_2)}$ if the Markov chain $\mathbf{M}^{(t_0)} - \mathbf{X}^{(t_1)} - \mathbf{Y}^{(t_2)}$ holds for some consecutive time steps. This conditional independence relation represents a channel through which $\mathbf{X}^{(t_1)}$ communicates $\mathbf{M}$ to $\mathbf{Y}^{(t_2)}$. Intuitively, this means all the information that $\mathbf{Y}^{(t_2)}$ has about $\mathbf{M}$ comes only from $\mathbf{X}^{(t_1)}$. But since each neural population is a high-dimensional system, it is unlikely that the stringent conditional independence relation of M-forwarding can hold. Instead, we are interested in finding low-dimensional M-forwarding structures from population activity.

A. M. was supported by NSF GRFP DGE2140739 and NIH T32 EB029365, and P. G. was supported by NSF-CCF-1763561.

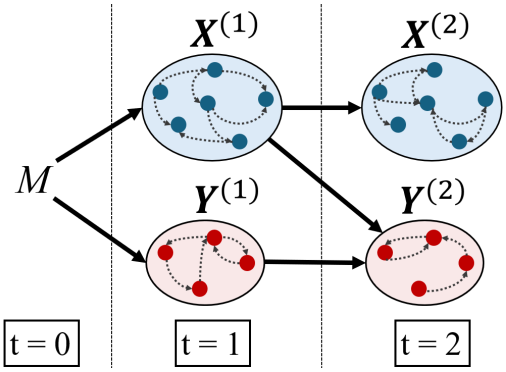

Fig. 1. Example of population communication measured in two time intervals between physically distinct populations $\mathbf{X}$ and $\mathbf{Y}$. Bold arrows indicate communication between populations while dashed arrows indicate communication between neurons within a population. Note that the arrows do not imply physical connectivity between the neurons and populations.

As detailed in Sec. II-C, previous approaches only consider isolated aspects of population communication, such as information flow between individual neurons [2] or population-level interaction in the absence of a message [3]. Unlike earlier works that do not directly address low-dimensional population communication, our previous work identified latent M-forwarding structures in two steps on real neural data [1]. We first applied message-relevant dimension reduction to each population separately. Then, we tested for M-forwarding in the inferred latent activity of both populations. The main limitation of this approach is that dimension reduction is applied to each population independently, which ignores joint activity that both populations may share with respect to M. Here, we formalize the problem of low-dimensional population communication to identify latent M-forwarding in a single step. We systematically define latent communication structures by highlighting simple, but non-trivial examples of population communication where M-forwarding can be observed in projected population activity. Specifically, our goal is to

characterize the $\mathbf{M}$-forwarding structure:

$$\mathbf{M}^{(t_0)} - f(\mathbf{X}^{(t_1)}) - g(\mathbf{Y}^{(t_2)}) \tag{1}$$

for projections f and g that map high-dimensional activity of $\mathbf{X}^{(t_1)}$ and $\mathbf{Y}^{(t_2)}$, respectively, to a low-dimensional space. We say that the structure formed in (1) is a latent channel[1] (LC).

The simplest setting of population communication occurs between two neural populations, $\mathbf{X}$ and $\mathbf{Y}$, over two time points. Fig. 1 illustrates this as a time-unrolled [2], directed causal graph [4]. We assume an event-related experimental paradigm [5] where the message is presented at time $t = 0$ and influences neural activity in populations $\mathbf{X}^{(1)}$ and $\mathbf{Y}^{(1)}$ at time $t = 1$. When $t = 2$, we observe the influence of each population from $t = 1$, as well as communication of $\mathbf{M}$-relevant information from $\mathbf{X}^{(1)}$ to $\mathbf{Y}^{(2)}$. Fig. 1 could represent, for example, a simplified model of the rodent whisker sensory system that includes a cortical population, $\mathbf{X}$, and a subcortical, deep-brain population, $\mathbf{Y}$, that jointly respond to and communicate information about whisker stimulation [6]. We assume the Causal Markov assumption holds between the high-dimensional populations [4], meaning that beyond direct causes (bold arrows), there are no additional causal influences on the activity after conditioning out the direct causes. However, note that LCs capture a statistical relationship, and do not imply a causal structure.

In the setting of Fig. 1, a central concern is defining what constitutes a 'valid' LC. Without additional constraints, innumerable LCs can be found that are not necessarily meaningful. Using an illustrative example, we first propose a procedure for finding projections f and g that satisfy (1) and optimally preserve the information about $\mathbf{M}$ (Sec. III-A), providing one criterion for defining a valid LC. Even after applying this procedure, we show by example that degenerate 'phantom' LCs can still exist, where zero information about $\mathbf{M}$ is forwarded (Sec. III-C). Thus, phantom LCs provide a key criterion for specifying valid LCs. We show that LCs can also be found through joint interactions between populations, resulting in four types of LCs (Sec. III-D). Based on these four types, we then define valid LCs in Sec. IV. Finally, we leverage bivariate Partial Information Decomposition (PID) to show how special structure in the high (low) dimensional populations can imply structure in the low (high) dimensional populations (Sec. V).

II. Background

Notation: Vector-valued random variables are denoted by boldface font, e.g., $\mathbf{A}$. Values observed at time point t are denoted by $\mathbf{A}^{(t)}$. Sets are denoted in calligraphic font, e.g., $\mathcal{F}$ is a set of projection functions. We use $\mathbf{A}$ and $\mathbf{B}$ to refer to arbitrary populations, and reserve $\mathbf{X}$ and $\mathbf{Y}$ to refer to anatomically distinct neural populations as in Fig. 1.

[1]Although we exemplify LCs with $f(\mathbf{X}^{(t_1)})$ and $g(\mathbf{Y}^{(t_2)})$, note that $\mathbf{M}^{(t_0)} - f(\mathbf{A}^{(t_1)}) - g(\mathbf{A}^{(t_2)})$ for the same population $\mathbf{A}$ is also an LC.

A. Sufficient Dimension Reduction

Dimension reduction is defined by some projection $f : \mathbb{R}^n \to \mathbb{R}^q$ of $\mathbf{A} \in \mathbb{R}^n$ to a lower dimensional space $q < n$. Thus, the Markov relation $\mathbf{M} - \mathbf{A} - f(\mathbf{A})$ holds. Sufficient dimension reduction is defined by a projection $f(\cdot)$ such that $\mathbf{M} - f(\mathbf{A}) - \mathbf{A}$ also holds [7]. If f is a sufficient projection of $\mathbf{A}$, it preserves information in $\mathbf{A}$ about $\mathbf{M}$ so that $I(\mathbf{M}; \mathbf{A}) = I(\mathbf{M}; f(\mathbf{A}))$.

B. Partial Information Decomposition (PID)

Bivariate PID[2] is a framework that specifies a decomposition of the total mutual information $\mathbf{X}$ and $\mathbf{Y}$ have about $\mathbf{M}$ into non-negative unique, redundant, and synergistic partial information terms [8]:

$$I(\mathbf{M}; \mathbf{X}, \mathbf{Y}) = UI(\mathbf{X}) + UI(\mathbf{Y}) + RI + SI \tag{2}$$

$$I(\mathbf{M}; \mathbf{X}) = UI(\mathbf{X}) + RI \tag{3}$$

$$I(\mathbf{M}; \mathbf{Y}) = UI(\mathbf{Y}) + RI. \tag{4}$$

Here, $UI(\mathbf{X})$ is the unique information $\mathbf{X}$ has about $\mathbf{M}$ that is not in $\mathbf{Y}$, $UI(\mathbf{Y})$ is the unique information $\mathbf{Y}$ has about $\mathbf{M}$, RI is the redundant information $\mathbf{X}$ and $\mathbf{Y}$ share about $\mathbf{M}$, and SI is the synergistic information $\mathbf{X}$ and $\mathbf{Y}$ have jointly about $\mathbf{M}$ (i.e. SI is computed through joint knowledge of both $\mathbf{X}$ and $\mathbf{Y}$). As there are four partial information terms but only three equations, a fourth equation measuring one of UI, RI, or SI must be specified to evaluate the decomposition. Many measures have been proposed to satisfy various desirable properties [9]. While we do not rely on a specific PID measure in this work, we limit our analysis to measures that are Blackwellian [10].

Blackwellian PID measures are defined by a key property that simplifies the computation of PID for systems $P(\mathbf{M}, \mathbf{X}, \mathbf{Y})$ that exhibit stochastic degradation [10], which we call degraded systems. Specifically, if $P_{\mathbf{Y}|\mathbf{M}}$ is stochastically degraded with respect to $P_{\mathbf{X}|\mathbf{M}}$, then Blackwellian PIDs measure zero unique information in $\mathbf{Y}$, i.e. $UI(\mathbf{Y}) = 0$. Thus, the computation of PID in degraded systems simply reduces to mutual information expressions derived from P since redundant information can be expressed:

$$RI = \min\{I(\mathbf{M}; \mathbf{X}), I(\mathbf{M}; \mathbf{Y})\}. \tag{5}$$

Jointly Gaussian $P(\mathbf{M}, \mathbf{X}, \mathbf{Y})$ was the first degraded system in which (5) was observed, and has since been extended to Poisson, multinomial, and, more generally, stable distributions [11], [12]. Note that (5) generally does not hold for arbitrary systems, even if the PID measure is Blackwellian.

We highlight a key relationship implied by (2)-(4):

$$I(\mathbf{M}; \mathbf{Y}|\mathbf{X}) = UI(\mathbf{Y}) + SI, \tag{6}$$

and formalize a useful property when the Markov chain $\mathbf{M} - \mathbf{X} - \mathbf{Y}$ holds (i.e. the system is also physically degraded [13]):

Lemma 1. *If $I(\mathbf{M}; \mathbf{Y}|\mathbf{X}) = 0$, then $UI(\mathbf{Y}) = SI = 0$.*

Proof. Proved by (6) and non-negativity of PID terms. □

[2]We henceforth refer to bivariate PID as PID.

C. Related work on neural communication

Existing work on neural population communication can be distinguished by three features: message relevance (implicit or explicit), measure of dependence (linear or general), and dimensionality (high or low). Ideally, a framework fully describing neural population communication explicitly depends on the message and summarizes communication in a low-dimensional space with a general measure of dependence. Besides [1], existing approaches only account for one or two of the three features. In computational neuroscience, a significant body of work has focused on quantifying low-dimensional neural interactions [3], [14]. These techniques share similar principles to Granger causal modeling [15], where 'information flow' is inferred from a time lag between latent activity in two or more populations and is typically quantified with linear measures of dependence.

Extensions to general dependence are often based on information-theoretic measures, including directed information [16], transfer entropy [17], and O-information [18]. However, these extensions do not consider low-dimensional representations of neural interaction. Furthermore, regardless of how dependence is quantified, these works discuss population *interactions* without referring to a specific message, and are thus not about *communication* of a message. Recently, PID [19], [20] has been applied to neural data to explicitly quantify $\mathbf{M}$-relevant interactions. In contrast to these works, $\mathbf{M}$-Information Flow ($\mathbf{M}$-IF) [2] is a theoretical framework that directly addresses message-dependent information flow. $\mathbf{M}$-IF specifies a computational system through which a message $\mathbf{M}$ is transmitted on edges between nodes that perform computations on $\mathbf{M}$. While PID and $\mathbf{M}$-IF are designed to explicitly account for a message, they do not describe *low-dimensional* representations of communication.

We contend that dimension reduction is a critical step for inferring communication in high-dimensional neural systems. The alternatives are inferring communication in the high-dimensional system, where it may not be possible to detect $\mathbf{M}$-forwarding directly, or communication between single neurons, which can miss $\mathbf{M}$-dependent high-order interactions within populations (Sec. III-D). We address the deficiencies of existing approaches by considering all three features of neural communication at once: explicit message relevance with a general measure of dependence for communication in low-dimensional representations.

III. Characterizing Latent Channels

We now describe examples modeling simplistic, but instructive scenarios of communication (via $\mathbf{M}$-forwarding) between populations of two neurons. In this section, we aim to propose desirable properties of the projections functions $f \in \mathcal{F}$ and $g \in \mathcal{G}$, and to characterize the types of LCs. $\mathcal{F}$ and $\mathcal{G}$ are arbitrary classes of projection functions, e.g., linear projections, autoencoders, etc. Based on this characterization, we propose an algorithm to find *valid* LCs. To simplify notation, let $t_0 = 0$, $t_1 = 1$, and $t_2 = 2$. As $\mathbf{M}$ is assumed to cause population activity, we omit t_0 notation on $\mathbf{M}$ unless otherwise noted.

A. Marginal Latent Channels

We first consider how LCs are formed marginally by projections of physically distinct neural populations (i.e. anatomically and/or temporally separated). Marginal LCs are $\mathbf{M}$-forwarding structures that embody how neural population communication questions are often formulated.

Definition 1. *A marginal LC is an LC,* $\mathbf{M} - f(\mathbf{X}, \mathbf{Y}) - g(\mathbf{X}, \mathbf{Y})$*, where* f *and* g *depend on only one of* $\mathbf{X}$ *or* $\mathbf{Y}$*.*

Two examples of marginal LCs include $\mathbf{M} - f(\mathbf{X}^{(1)}) - g(\mathbf{X}^{(2)})$, a marginal LC between the same population, but temporally separated, and $\mathbf{M} - f(\mathbf{X}^{(1)}) - g(\mathbf{Y}^{(2)})$, a marginal LC between spatially distinct populations over time. In this section, we focus on $\mathbf{M}$-forwarding between two spatially separated populations, e.g., between $\mathbf{X}^{(1)}$ and $\mathbf{Y}^{(2)}$ so we can ignore activity in $\mathbf{X}^{(2)}$. Example 1 illustrates a marginal LC that we also use to state desirable properties of the projection functions $f \in \mathcal{F}$ and $g \in \mathcal{G}$.

Example 1. Suppose $\mathbf{M} = \begin{bmatrix} M_1 & M_2 \end{bmatrix}$ has components marginally distributed as $P(M_1, M_2)$. Let ϵ_1 and ϵ_2 be noise terms, and suppose M_i, ϵ_j, $i, j \in \{1, 2\}$ are mutually independent. Define the population activity

$$\mathbf{X}^{(1)} = \begin{bmatrix} M_1 \\ \epsilon_1 \end{bmatrix}, \mathbf{Y}^{(1)} = \begin{bmatrix} M_2 \\ \epsilon_2 \end{bmatrix}, \mathbf{Y}^{(2)} = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix}.$$

The components of the message are split across the populations at time 1, then recombined in $\mathbf{Y}$ at time 2. Intuitively, a marginal LC from $f(\mathbf{X}^{(1)})$ to $g(\mathbf{Y}^{(2)})$ is given by the linear operators $f = g = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$ that select components in $\mathbf{X}^{(1)}$ and $\mathbf{Y}^{(2)}$, yielding the Markov chain $\mathbf{M}^{(0)} - M_1^{(1)} - M_1^{(2)}$.

Example 1 suggests desirable properties of the projecting functions. First, $f \in \mathcal{F}$ is a projection that maximizes total marginal information of the projected activity with the message, $I(\mathbf{M}; f(\mathbf{X}^{(1)}))$. Second, g maximizes the marginal information $I(\mathbf{M}; g(\mathbf{Y}^{(2)}))$ while respecting (1).

B. Case Study: Jointly Gaussian Systems

We illustrate the procedure discussed in Sec. III-A for identifying projections f and g to form marginal LCs in jointly Gaussian systems. Here, we restrict our analysis to linear projections U and V. Our goal is to find LCs in which $\mathbf{Y}^{(2)} \in \mathbb{R}^{n_y}$ receives information about a p-dimensional message $\mathbf{M} \in \mathbb{R}^p$ through $\mathbf{X}^{(1)} \in \mathbb{R}^{n_x}$. We let $p \leq \min\{n_x, n_y\}$ so the number of dimensions of the message is fewer than the number of recorded neurons, an assumption that is often satisfied in experiments[3]. Let the joint distribution of the message and neural activity be $P(\mathbf{M}, \mathbf{X}^{(1)}, \mathbf{Y}^{(2)}) = N(0, \Sigma)$, where

$$\Sigma = \begin{bmatrix} \Sigma_{\mathbf{M}} & \Sigma_{\mathbf{M}\mathbf{Y}^{(2)}} & \Sigma_{\mathbf{M}\mathbf{X}^{(1)}} \\ \Sigma_{\mathbf{Y}^{(2)}\mathbf{M}} & \Sigma_{\mathbf{Y}^{(2)}} & \Sigma_{\mathbf{Y}^{(2)}\mathbf{X}^{(1)}} \\ \Sigma_{\mathbf{X}^{(1)}\mathbf{M}} & \Sigma_{\mathbf{X}^{(1)}\mathbf{Y}^{(2)}} & \Sigma_{\mathbf{X}^{(2)}} \end{bmatrix}.$$

[3]We consider event-related experiments [5] which are often based on simple, low-dimensional stimuli. In naturalistic experimental paradigms, where stimuli may be naturally high-dimensional, we can restrict $\mathbf{M}$ to be a few features of the stimuli so the message is again low-dimensional.

The off-diagonal blocks are cross-covariance matrices between the variables in the subscripts while the diagonal blocks are square covariance matrices. We can now define a new joint distribution $S(\mathbf{M}, U^T\mathbf{X}^{(1)}, V^T\mathbf{Y}^{(2)}) = N(0, \Sigma_S)$ based on linear projections U and V of $\mathbf{X}^{(1)}$ and $\mathbf{Y}^{(2)}$, respectively. Now, the new covariance matrix is

$$\Sigma_S = \begin{bmatrix} \Sigma_{\mathbf{M}} & \Sigma_{\mathbf{M}\mathbf{Y}^{(2)}}V & \Sigma_{\mathbf{M}\mathbf{X}^{(1)}}U \\ V^T\Sigma_{\mathbf{Y}^{(2)}} & V^T\Sigma_{\mathbf{Y}^{(2)}}V & V^T\Sigma_{\mathbf{Y}^{(2)}\mathbf{X}^{(1)}}U \\ U^T\Sigma_{\mathbf{X}^{(1)}} & U^T\Sigma_{\mathbf{X}^{(1)}\mathbf{Y}^{(2)}}V & U^T\Sigma_{\mathbf{X}^{(2)}}U \end{bmatrix}.$$

We can identify the pairs (U, V) such that the LC, $\mathbf{M} - U^T\mathbf{X}^{(1)} - V^T\mathbf{Y}^{(2)}$, forms by setting the off-diagonal of the conditional covariance of $S(\mathbf{M}, \mathbf{Y}^{(2)}|\mathbf{X}^{(1)})$ to zero. This yields the condition:

$$h(U)V = \mathbf{0}, \text{ where} \tag{7}$$
$$h(U) = \Sigma_{\mathbf{M}\mathbf{Y}^{(2)}} - \Sigma_{\mathbf{M}\mathbf{X}^{(1)}}U(U^T\Sigma_{\mathbf{X}^{(1)}}U)^{-1}U^T\Sigma_{\mathbf{X}^{(1)}\mathbf{Y}^{(2)}}.$$

We first find the linear projection U that maximizes marginal information of projected $\mathbf{X}^{(1)}$ activity with $\mathbf{M}$. In jointly Gaussian systems, ordinary least squares regression, $U^* = \Sigma_X^{-1}\Sigma_{XM}$, is a sufficient projection [1]. Hence, by choosing U^* as the projection of $\mathbf{X}^{(1)}$ activity, marginal information is completely preserved after projection, i.e. $I(\mathbf{M}; (U^*)^T\mathbf{X}^{(1)}) = I(\mathbf{M}; \mathbf{X}^{(1)})$. Now that the projection of $\mathbf{X}^{(1)}$ is chosen, we select the subset of projections of $\mathbf{Y}^{(2)}$ such that (1) holds. This is given by condition (7), where the projection of $\mathbf{Y}^{(2)}$ must be in the $(n_y - p)$-dimensional nullspace of $h(U^*)$, so we choose V^* as the nullspace of $h(U^*)$. Thus, (U^*, V^*) is the joint set of linear projections satisfying (1). Finally, if we wish to further restrict the dimension of the receiving population, we would then select a subspace $V_S^* \subseteq V^*$ such that $I(\mathbf{M}; (V_S^*)^T\mathbf{Y}^{(2)})$ is maximized. Note that (U^*, V^*) itself also constitutes a pair of projection functions, and forms an LC between a p-dimensional transmitting population and a $(n_y - p)$-dimensional receiving population.

C. Phantom Latent Channels

Selecting projections f and g that maximize information of projected activity with $\mathbf{M}$ ensures that certain cases of degenerate LCs (e.g., those formed by constant projections) are avoided. However, we can still find degenerate LCs even if we follow the procedure for identifying projections outlined in Sec. III-A.

Example 2. Let $M \sim N(0, \sigma_M^2)$ and define mutually independent noise terms $\epsilon_i \sim N(0, \sigma_i^2)$ for $i = 1, 2, 3$. As in Sec. III-B, let $\mathcal{F}$ and $\mathcal{G}$ be sets of linear operators. We define the populations $\mathbf{X}^{(1)}, \mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}$ as

$$\mathbf{X}^{(1)} = \begin{bmatrix} M + \epsilon_1 \\ \epsilon_2 \end{bmatrix}, \ \mathbf{Y}^{(1)} = \mathbf{Y}^{(2)} = \begin{bmatrix} M \\ \epsilon_3 \end{bmatrix}$$

and seek an LC of the form $M - f(\mathbf{X}^{(1)}) - g(\mathbf{Y}^{(2)})$. Intuitively, there should be no marginal LC from $\mathbf{X}^{(1)}$ to $\mathbf{Y}^{(2)}$ since message-relevant activity in $\mathbf{X}^{(1)}$ is a noisier version of activity in $\mathbf{Y}^{(2)}$. But we can still find a marginal LC following the procedure of Sec. III-A. The regression vector $\mathbf{r} = \gamma \begin{bmatrix} 1 & 0 \end{bmatrix}^T$ for some constant γ is the sufficient projection of $\mathbf{X}^{(1)}$. Thus, we can choose the first projection as $\mathbf{u_1} = \mathbf{r}/\gamma$. Then, $\mathbf{v_1}$ must be in the nullspace of $h(\mathbf{u_1})$ to satisfy (7), where

$$h(\mathbf{u_1}) = \frac{\sigma_M^2\sigma_1^2}{\sigma_M^2 + \sigma_1^2} \cdot \begin{bmatrix} 1 & 0 \end{bmatrix}^T.$$

We select $\mathbf{v_1} = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$. Then, $\mathbf{u_1}^T\mathbf{X}^{(1)} = M + \epsilon_1$ and $\mathbf{v_1}^T\mathbf{Y}^{(2)} = \epsilon_3$. Here, the LC given by $M - (M + \epsilon_1) - \epsilon_3$ exists merely by copying successively noisier versions of M.

Although the LC of Example 2 satisfies the Markov property for $\mathbf{M}$-forwarding, it should not represent any communication of information about $\mathbf{M}$ since $I(M; g(\mathbf{Y}^{(2)})) = 0$. We formalize LCs that $\mathbf{M}$-forward no information about $\mathbf{M}$ as phantom LCs.

Definition 2. *For populations $\mathbf{A}, \mathbf{B}$, the LC given by $\mathbf{M} - f(\mathbf{A}) - g(\mathbf{B})$ is a phantom LC if $I(\mathbf{M}; g(\mathbf{B})) = 0$.*

Phantom LCs, together with the procedure of Sec. III-A, provide criteria for defining a 'valid' marginal LC. However, an LC need not be marginal, as shown in Sec. III-D.

D. Joint Latent Channels

Even if the high-dimensional populations marginally have no information about $\mathbf{M}$, e.g., $I(\mathbf{M}; \mathbf{X}^{(1)}) = 0$, there may exist LCs that are not phantoms but are still defined by information-maximizing projections.

Example 3. Suppose $M \sim Bern(0.5)$. Define the populations

$$\mathbf{X}^{(1)} = \begin{bmatrix} M \oplus \epsilon_1 \\ \epsilon_2 \end{bmatrix}, \mathbf{Y}^{(1)} = \begin{bmatrix} M \oplus \epsilon_2 \\ \epsilon_1 \end{bmatrix}, \mathbf{Y}^{(2)} = \begin{bmatrix} M \\ \epsilon_0 \end{bmatrix}$$

where ϵ_i for $i = 1, 2, 3$ are mutually independent $Bern(0.5)$ noise terms. The first component of $\mathbf{Y}^{(2)}$ is a deterministic function of components of $\mathbf{X}^{(1)}$ and $\mathbf{Y}^{(1)}$. Specifically, $Y_1^{(2)} = X_1^{(1)} \oplus Y_2^{(1)} = X_2^{(1)} \oplus Y_1^{(1)}$. Thus, $\mathbf{M}$-forwarding clearly holds in the high-dimensional system. However, we note that $I(M; \mathbf{X}^{(1)}) = I(M; \mathbf{Y}^{(1)}) = 0$, so there is no marginal LC that is not a phantom. Since the information required to recover $\mathbf{M}$ is split across both physical populations, we can find a non-degenerate LC as a function of both populations. We concatenate $\mathbf{X}^{(1)}$ and $\mathbf{Y}^{(1)}$ activity in

$$\mathbf{A}^{(1)} = (\mathbf{X}^{(1)}, \mathbf{Y}^{(1)}) = \begin{bmatrix} M \oplus \epsilon_1 & \epsilon_2 & M \oplus \epsilon_2 & \epsilon_1 \end{bmatrix}^T$$

and see that there are two sufficient projections of $\mathbf{A}^{(1)}$ given by the linear projections (with summation modulo 2) $f_1 = \begin{bmatrix} 1 & 0 & 0 & 1 \end{bmatrix}^T$ and $f_2 = \begin{bmatrix} 0 & 1 & 1 & 0 \end{bmatrix}^T$. Choosing $g = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$ for $\mathbf{Y}^{(2)}$, we see that f_1, f_2 are precisely the deterministic functions relating $Y_1^{(2)}$ to $\mathbf{X}^{(1)}$ and $\mathbf{Y}^{(1)}$.

To identify the LC in Example 3, we sufficiently project the *joint* neural activity at time 1, instead of the *marginal* activity of any one physical population. We formalize LCs whose projections depend on activity in both physical populations as joint LCs.

Definition 3. *A joint LC is an LC,* $\mathbf{M} - f(\mathbf{X}, \mathbf{Y}) - g(\mathbf{X}, \mathbf{Y})$*, where* f *or* g *depends on both* $\mathbf{X}$ *and* $\mathbf{Y}$*.*

Example 3 illustrates a joint LC formed from the combination of activity in the transmitting populations. Joint LCs can also form in the receiving populations, where we now must account for activity in both $\mathbf{X}^{(2)}$ and $\mathbf{Y}^{(2)}$.

Example 4. Suppose $M \sim Bern(0.5)$. The populations at time 1 are defined as

$$\mathbf{X}^{(1)} = \begin{bmatrix} M \\ \epsilon_1 \end{bmatrix}, \mathbf{Y}^{(1)} = \begin{bmatrix} M \\ \epsilon_2 \end{bmatrix}$$

and at time 2 are defined as

$$\mathbf{X}^{(2)} = \begin{bmatrix} \epsilon_0 \\ \epsilon_2 \end{bmatrix}, \mathbf{Y}^{(2)} = \begin{bmatrix} M \oplus \epsilon_0 \\ \epsilon_2 \end{bmatrix}.$$

Here, $\epsilon_i \sim Bern(0.5)$, for $i = 1, 2, 3$, are mutually independent noise terms. Note that $\mathbf{M}$-forwarding holds at the population level. As in Example 3, all marginal LCs are phantoms. If we take $f = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$ as the modulo 2 linear projection of $\mathbf{X}^{(1)}$ and $g = \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix}^T$ as the modulo 2 linear projection of the concatenated activity $(\mathbf{X}^{(2)}, \mathbf{Y}^{(2)})$, we obtain a joint LC.

The combination of the joint channel examples, e.g., the transmitting population of Example 3 with the receiving population of Example 4, reveals an additional LC resulting from joint population activity at both $t = 1$ and $t = 2$. As such, three distinct types of joint LCs can form, depending on whether there is a joint interaction between the populations at $t = 1$, $t = 2$, or both $t = 1$ and $t = 2$. These examples show that independent projections of marginal population activity to infer LCs, as in [1], can miss out on three possible joint LCs arising from different kinds of cross-population interactions about $\mathbf{M}$. We emphasize the distinction between marginal and joint LCs since many neuroscience questions often focus only on marginal LC structures, such as the questions in our earlier work [1]. However, joint LCs are also biologically relevant as they reveal a fundamentally different kind of neural communication arising from joint activity between neural populations.

IV. Valid Latent Channels

The examples in Sec. III encapsulate the four kinds of LCs that can arise due to message-relevant marginal and joint interactions between populations at a given time. This is summarized in Fig. 2. Based on this characterization, we state the General LC Algorithm for selecting projection functions $f \in \mathcal{F}$ and $g \in \mathcal{G}$ that project activity of arbitrary (i.e. marginal and joint) neural populations $\mathbf{A}^{(t_1)}$ and $\mathbf{B}^{(t_2)}$:

1) Choose the projection $f \in \mathcal{F}$ such that
$$f = \arg\max_{h \in \mathcal{F}} I(\mathbf{M}; h(\mathbf{A}^{(t_1)})).$$
2) Identify the subset of projections $\mathcal{G}_M \subseteq \mathcal{G}$ of $\mathbf{B}^{(t_2)}$ such that $\mathbf{M} - f(\mathbf{A}^{(t_1)}) - h(\mathbf{B}^{(t_2)})$ holds for $h \in \mathcal{G}_M$.
3) Choose the projection $g \in \mathcal{G}_M$ such that
$$g = \arg\max_{h \in \mathcal{G}_M} I(\mathbf{M}; h(\mathbf{B}^{(t_2)})).$$

The General LC Algorithm ensures that information about $\mathbf{M}$ is optimally preserved in each population given (1) and the chosen projection classes $\mathcal{F}$ and $\mathcal{G}$. We now use the General LC Algorithm to define a valid LC:

Definition 4. *A valid LC is an LC found by the General LC Algorithm that is not a phantom LC.*

Although we assume $\mathbf{M}$-forwarding in the original, high-dimensional system, $P(\mathbf{M}, \mathbf{X}^{(1)}, \mathbf{Y}^{(1)}, \mathbf{X}^{(2)}, \mathbf{Y}^{(2)})$, captures a causal mechanism in a directed graph [4], detection of a valid LC does not imply it also exists as a causal mechanism through which communication occurs [21]. Since valid LCs only specify statistical relationships, they simply identify possible mechanisms through which the populations may communicate. This is observed in Example 3, where $Y_1^{(2)}$ is related to the earlier population activity by two equivalent XOR operations: $Y_1^{(2)} = X_1^{(1)} \oplus Y_2^{(1)}$ and $Y_1^{(2)} = X_2^{(1)} \oplus Y_1^{(1)}$. Information about $\mathbf{M}$ could be communicated through one or both of the LCs. Without additional information, we can make no inferences about 'true' causal LCs with observational data.

V. PID in Latent Channels

A. PID across populations

The PID of cross-population activity, $I(\mathbf{M}; \mathbf{A}^{(i)}, \mathbf{B}^{(i)})$ at time points $i = 1, 2$, provides interesting insights into the four types of valid LCs. A valid marginal LC emerges from positive unique and/or redundant information between populations, e.g., Example 1, whose PID is $I(\mathbf{M}; \mathbf{X}^{(1)}, \mathbf{Y}^{(1)}) = UI(\mathbf{X}^{(1)}) + UI(\mathbf{Y}^{(1)})$. In fact, positivity of unique and/or redundant information in the transmitting populations is a necessary condition for the existence of valid marginal LCs. However, since the LC also depends on the projecting function class, which can be degenerate (e.g., class of projections to constants), it is not a sufficient condition.

Lemma 2. *For transmitting population* $\mathbf{X}^{(1)}$*,* $UI(\mathbf{X}^{(1)}) > 0$ *or* $RI > 0$ *is a necessary condition for the existence of a valid marginal LC.*

Proof. Suppose both $UI(\mathbf{X}^{(1)}) = RI = 0$. By (3) and the data processing inequality [22], $0 = I(\mathbf{M}; \mathbf{X}^{(1)}) \geq I(\mathbf{M}; f(\mathbf{X}^{(1)}))$ for all projections $f \in \mathcal{F}$. Thus, any LC $\mathbf{M} - f(\mathbf{X}^{(1)}) - g(\mathbf{Y}^{(2)})$ is such that $I(\mathbf{M}; g(\mathbf{Y}^{(2)})) = 0$. □

Valid joint LCs can arise from synergistic interactions between the transmitting, receiving, or both populations. The examples in Sec. III-D illustrate this with the XOR operation, the canonical example of pure synergistic information about $\mathbf{M}$ [8]. Specifically, in Example 3, the PID reduces to $I(M; \mathbf{X}^{(1)}, \mathbf{Y}^{(1)}) = SI$. Unlike positivity of unique/redundant information, positive synergistic information is *not* necessary for the existence of valid joint LCs. A joint LC can also emerge from fractional, non-synergistic information, as shown in Example 5.

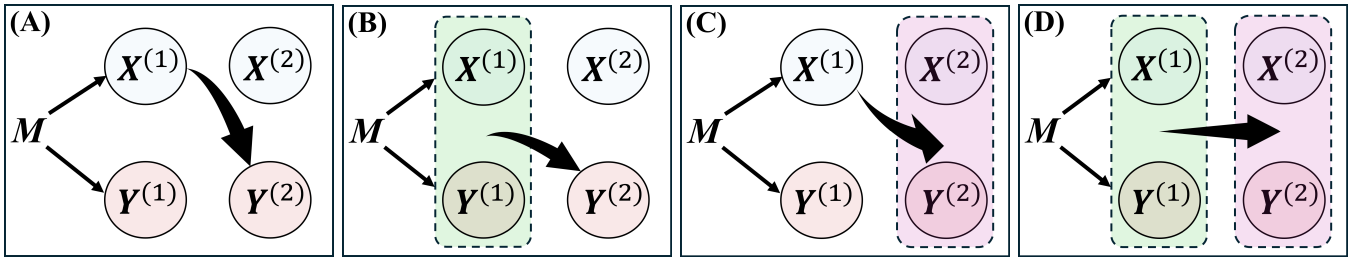

Fig. 2. The four kinds of valid LCs: (A) Marginal, (B) Joint in transmitting populations, (C) Joint in receiving populations, (D) Joint in transmitter and receiver. The green and pink shaded boxes represent joint information about $\mathbf{M}$ in transmitter and receiver populations, respectively.

Example 5. Suppose $M \sim P(M)$. The population activity is

$$\mathbf{X}^{(1)} = \mathbf{Y}^{(1)} = \mathbf{Y}^{(2)} = \begin{bmatrix} M \\ \epsilon \end{bmatrix}$$

where ϵ is an independent noise term. Define the joint projection of the concatenated activity $\mathbf{A}^{(1)} = (\mathbf{X}^{(1)}, \mathbf{Y}^{(1)})$ as

$$f_\alpha = \begin{bmatrix} \alpha & 0 & (1-\alpha) & 0 \end{bmatrix}^T, \text{ for } \alpha \in [0,1]$$

and let the projection of $\mathbf{Y}^{(2)}$ be $g = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$. If $\alpha = 1$, the valid marginal LC is $M - f_1(\mathbf{X}^{(1)}) - g(\mathbf{Y}^{(2)})$ while if $\alpha = 0$ the LC is $M - f_0(\mathbf{Y}^{(1)}) - g(\mathbf{Y}^{(2)})$. If $\alpha \in (0,1)$, there is a continuous range of valid input joint LCs: $M - f_\alpha(\mathbf{A}^{(1)}) - g(\mathbf{Y}^{(2)})$. Note that the PID is $I(M; \mathbf{X}^{(1)}, \mathbf{Y}^{(1)}) = RI$, so both populations are completely redundant about M and $SI = 0$.

In addition to demonstrating how redundant information can produce valid joint LCs, Example 5 shows how there can be multiple projection functions that each maximally preserve information about $\mathbf{M}$. Although the lack of a unique projection function can adversely affect the interpretation of LCs, identification of the possible valid LCs can also provide tangible suggestions for experiment designs.

B. PID across time

We now consider how the PID of $I_P(\mathbf{M}; \mathbf{A}^{(i)}, \mathbf{B}^{(j)}), i \neq j$, across time points can reveal basic properties of latent population communication, as well as how projections reallocate the quantity of partial information terms. In this section, we use subscripts P and S on information quantities to denote whether the quantity is computed for the original, high-dimensional system $P(\mathbf{M}, \mathbf{A}^{(i)}, \mathbf{B}^{(j)})$ or the projected, latent system $S(\mathbf{M}, f(\mathbf{A}^{(i)}), g(\mathbf{B}^{(j)}))$. We first analyze the PID of P. In our simplified model of communication (Fig. 1), the populations at time 1 jointly $\mathbf{M}$-forward to themselves at time 2. In other words, the following Markov chain holds:

$$\mathbf{M} - \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{Y}^{(1)} \end{bmatrix} - \begin{bmatrix} \mathbf{X}^{(2)} \\ \mathbf{Y}^{(2)} \end{bmatrix}$$

so $\mathbf{A}^{(2)} = (\mathbf{X}^{(2)}, \mathbf{Y}^{(2)})$ depends on $\mathbf{M}$ only through $\mathbf{A}^{(1)} = (\mathbf{X}^{(1)}, \mathbf{Y}^{(1)})$. Because $I_P(\mathbf{M}; \mathbf{A}^{(2)} | \mathbf{A}^{(1)}) = 0$, we can conclude with Lemma 1 that the PID is

$$I_P(\mathbf{M}; \mathbf{A}^{(1)}, \mathbf{A}^{(2)}) = UI_P(\mathbf{A}^{(1)}) + RI_P.$$

The redundant information RI_P between $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ about $\mathbf{M}$ can be interpreted as the information that is $\mathbf{M}$-forwarded from $\mathbf{A}^{(1)}$ to $\mathbf{A}^{(2)}$. This interpretation of redundant information also holds for LCs, and allows an equivalent definition of phantom LCs as LCs that have $RI_S = 0$ (i.e. zero information about $\mathbf{M}$ is forwarded). Moreover, since $SI_P = UI_P(\mathbf{A}^{(2)}) = 0$ by Lemma 1, LCs represent zero-synergy structures that isolate all the unique information to the transmitting population. Intuitively, this means the receiving populations are neither getting information about $\mathbf{M}$ from elsewhere (since $UI_P(\mathbf{A}^{(2)}) = 0$), nor internally generating information about $\mathbf{M}$ (since $SI_P = 0$). In this manner, PID provides interpretations of the information that is $\mathbf{M}$-forwarded in the examples of Sec. III. Namely, the information terms in the PID of $I_P(\mathbf{M}; \mathbf{A}^{(1)}, \mathbf{B}^{(1)})$ are converted into unique information $UI(\mathbf{A}^{(1)})$ in the PID of $I_P(\mathbf{M}; \mathbf{A}^{(1)}, \mathbf{B}^{(2)})$.

We now address the question of how information about $\mathbf{M}$ is modified in the latent system $S(\mathbf{M}, f(\mathbf{X}), g(\mathbf{Y}))$ after projections f and g on the original system $P(\mathbf{M}, \mathbf{X}, \mathbf{Y})$. We omit time in the remainder of this section. Since we focus on $\mathbf{M}$-forwarding structures, we assume that f and g are functions that satisfy $\mathbf{M} - f(\mathbf{X}) - g(\mathbf{Y})$. Without knowing anything about P or properties of f, g, we can only conclude that $I(\mathbf{M}; \mathbf{X}, \mathbf{Y}) \geq I(\mathbf{M}; f(\mathbf{X}), g(\mathbf{Y}))$ by the data processing inequality (DPI) [22] and $UI_S(\mathbf{Y}) = SI_S = 0$ by Lemma 1. However, certain special structure in the original system P can imply structure in the latent system S. Specifically, Prop. 1 shows that if P specifies a physically degraded system (i.e. $\mathbf{M}$-forwarding *does* hold in the high-dimensional system), an LC between the same populations is guaranteed.

Proposition 1. Suppose $\mathbf{M} - \mathbf{X} - \mathbf{Y}$ holds for $P(\mathbf{M}, \mathbf{X}, \mathbf{Y})$. If f is a sufficient projection of $\mathbf{X}$, then $\mathbf{M} - f(\mathbf{X}) - g(\mathbf{Y})$ holds for $S(\mathbf{M}, f(\mathbf{X}), g(\mathbf{Y}))$ for all projections g of $\mathbf{Y}$.

Proof. By $M - \mathbf{X} - \mathbf{Y}$ and sufficiency of f,

$$\begin{aligned} I_P(\mathbf{M}; \mathbf{X}, \mathbf{Y}) &= I_P(\mathbf{M}; \mathbf{X}) + I_P(\mathbf{M}; \mathbf{Y} | \mathbf{X}) \\ &= I_S(\mathbf{M}; f(\mathbf{X})). \end{aligned} \tag{8}$$

By the DPI and (8),

$$\begin{aligned} I_S(\mathbf{M}; f(\mathbf{X})) &= I_P(\mathbf{M}; \mathbf{X}, \mathbf{Y}) \\ &\geq I_S(\mathbf{M}; f(\mathbf{X}), g(\mathbf{Y})) \\ &= I_S(\mathbf{M}; f(\mathbf{X})) + I_S(\mathbf{M}; g(\mathbf{Y}) | f(\mathbf{X})). \end{aligned}$$

Thus, $I_S(\mathbf{M}; g(\mathbf{Y})|f(\mathbf{X})) = 0$ and $\mathbf{M} - f(\mathbf{X}) - g(\mathbf{Y})$ holds. □

Although Prop. 1 shows that the $\mathbf{M}$-forwarding structure in P is preserved in S, the PID is generally not invariant to these projections. This is because $UI_P(\mathbf{X}) + RI_P = UI_S(f(\mathbf{X})) + RI_S$, due to sufficiency of f, so both the quantity and proportion of unique and redundant information can change from the original system. However, since g is arbitrary, we can choose it to also be a sufficient projection of $\mathbf{Y}$. Then both the $\mathbf{M}$-forwarding structure and the PID remain the same after projection. Furthermore, if $I_P(\mathbf{M}; \mathbf{Y}) > 0$, the LC derived from sufficient projections f and g is also a valid LC.

In many experimental settings, observing $\mathbf{M}$-forwarding in the high-dimensional system, P, is unlikely. However, Prop. 2 shows that we can still make conclusions about how PID terms are affected by sufficient projections if we know that P forms a degraded system (defined in Sec. II-B).

Proposition 2. Let $P(\mathbf{M}, \mathbf{X}, \mathbf{Y})$ describe a degraded system where $I_P(\mathbf{M}; \mathbf{X}) \geq I_P(\mathbf{M}; \mathbf{Y})$. Let f and g be projections of $\mathbf{X}$ and $\mathbf{Y}$, respectively, and denote their joint distribution by $S(\mathbf{M}, f(\mathbf{X}), g(\mathbf{Y}))$.

1) If f is sufficient then: (a) $UI_S(f(\mathbf{X})) \geq UI_P(\mathbf{X})$, (b) $RI_S \leq RI_P$, and (c) $SI_S \leq SI_P$.
2) If g is sufficient then: (a) $UI_S(f(\mathbf{X})) \leq UI_P(\mathbf{X}) + C_1$, (b) $RI_S \leq RI_P$, and (c) $SI_S \leq SI_P + C_2$ for constants $C_1, C_2 \geq 0$.

Proof. Since P is a degraded system, $RI_P = I_P(\mathbf{M}; \mathbf{Y})$. We now show the inequalities for each case.

Case 1: f is sufficient.

Because g is a function of $\mathbf{Y}$, we apply the DPI:

$$I_P(\mathbf{M}; \mathbf{Y}) = RI_P \geq RI_S + UI_S(g(\mathbf{Y})) = I_S(\mathbf{M}; g(\mathbf{Y})) \tag{9}$$

which implies $RI_P \geq RI_S$. Now, by sufficiency of f,

$$\begin{aligned} I_P(\mathbf{M}; \mathbf{X}) &= UI_P(\mathbf{X}) + RI_P \\ &= UI_S(f(\mathbf{X})) + RI_S = I_S(\mathbf{M}; f(\mathbf{X})) \end{aligned}$$

so $RI_P = UI_S(f(\mathbf{X})) + RI_S - UI_P(\mathbf{X})$. Plugging this into (9),

$$UI_S(f(\mathbf{X})) + RI_S - UI_P(\mathbf{X}) \geq UI_S(g(\mathbf{Y})) + RI_S.$$

Then the following inequality holds:

$$UI_S(f(\mathbf{X})) \geq UI_P(\mathbf{X}) + UI_S(g(\mathbf{Y})) \geq UI_P(\mathbf{X}).$$

By sufficiency of f, $I_P(\mathbf{M}; \mathbf{Y}|\mathbf{X}) \geq I_S(\mathbf{M}; g(\mathbf{Y})|f(\mathbf{X}))$. Thus,

$$SI_P \geq UI_S(g(\mathbf{Y})) + SI_S \geq SI_S.$$

Case 2: g is sufficient.

The relationship $RI_P \geq RI_S$ is shown as in Case 1. We now apply DPI on $I_S(\mathbf{M}; f(\mathbf{X}))$ to obtain:

$$\begin{aligned} I_S(\mathbf{M}; f(\mathbf{X})) &= UI_S(f(\mathbf{X})) + RI_S \\ &\leq UI_P(\mathbf{X}) + RI_P = I_P(\mathbf{M}; \mathbf{X}). \end{aligned}$$

Thus, $UI_S(f(\mathbf{X})) \leq UI_P(\mathbf{X}) + C_1$, where $C_1 = RI_P - RI_S$. We now apply DPI to the total mutual information:

$$\begin{aligned} I_P(\mathbf{M}; \mathbf{X}, \mathbf{Y}) &= I_P(\mathbf{M}; \mathbf{X}) + I_P(\mathbf{M}; \mathbf{Y}|\mathbf{X}) \\ &\geq I_S(\mathbf{M}; f(\mathbf{X})) + I_S(\mathbf{M}; g(\mathbf{Y})|f(\mathbf{X})) \\ &= I_S(\mathbf{M}; f(\mathbf{X}), g(\mathbf{Y})). \end{aligned}$$

We use the PID equalities to conclude

$$SI_P + C_2 \geq UI_S(g(\mathbf{Y})) + SI_S \geq SI_S \tag{10}$$

where $C_2 = I_P(\mathbf{M}; \mathbf{X}) - I_S(\mathbf{M}; f(\mathbf{X}))$. □

Props. 1 and 2 show that structure in S is implied when there is additional knowledge of the original system P. In general, knowing that the latent system S forms an LC, $\mathbf{M} - f(\mathbf{X}) - g(\mathbf{Y})$, for some f and g does not imply anything about P. Trivially, if S specifies an LC where $g(\mathbf{Y})$ is the receiving population, then $SI_P \geq SI_S = 0$ and $UI_P(\mathbf{Y}) \geq UI_S(g(\mathbf{Y})) = 0$ by Lemma 1. However, if we know that both f and g are sufficient projections, this also implies a reallocation of the quantities of redundant information and unique information in $\mathbf{X}$ under P, summarized in Prop. 3.

Proposition 3. If $\mathbf{M} - f(\mathbf{X}) - g(\mathbf{Y})$ holds for sufficient projections f and g of $\mathbf{X}$ and $\mathbf{Y}$, respectively, then (a) $UI_S(f(\mathbf{X})) \leq UI_P(\mathbf{X})$ and (b) $RI_S \geq RI_P$.

Proof. By sufficiency of g and Lemma 1

$$\begin{aligned} I_P(\mathbf{M}; \mathbf{Y}) &= RI_P + UI_P(\mathbf{Y}) \\ &= RI_S = I_S(\mathbf{M}; g(\mathbf{Y})). \end{aligned} \tag{11}$$

So $RI_S \geq RI_P$. We now invoke the sufficiency of f:

$$\begin{aligned} I_P(\mathbf{M}; \mathbf{X}) &= RI_P + UI_P(\mathbf{X}) \\ &= RI_S + UI_S(\mathbf{X}) = I_S(\mathbf{M}; f(\mathbf{X})). \end{aligned} \tag{12}$$

We plug (11) into (12) to obtain

$$UI_P(\mathbf{X}) + RI_P = UI_S(f(\mathbf{X})) + RI_P + UI_P(\mathbf{Y})$$

from which we conclude that $UI_P(\mathbf{X}) \geq UI_S(f(\mathbf{X}))$. □

Prop. 3 states that when we know an LC is formed by sufficient projections on both populations, the redundant information RI_P in P can only decrease from S. For positive PID terms, this can be equivalently stated from inequality (b) in Prop. 3 as $1/RI_S \leq 1/RI_P$. Combining this with inequality (a) for unique information, we have $UI_S(f(\mathbf{X}))/RI_S \leq UI_P(\mathbf{X})/RI_P$. The new inequality states that the proportion of unique to redundant information increases in the high-dimensional system, so the sufficient projections necessarily reduce or eliminate the unique information from P.

Finally, if we know additional structure in both S and P, we can also infer how PID terms in S and P are related. Prop. 4 summarizes this for when f is sufficient and P is known to be a degraded system.

Proposition 4. If $P(\mathbf{M}, \mathbf{X}, \mathbf{Y})$ is a degraded system where $I_P(\mathbf{M}; \mathbf{X}) \geq I_P(\mathbf{M}; \mathbf{Y})$ and $\mathbf{M} - f(\mathbf{X}) - g(\mathbf{Y})$ holds for

sufficient f and arbitrary g, then (a) $UI_S(f(\mathbf{X})) \geq UI_P(\mathbf{X})$ and (b) $RI_S \leq RI_P$.

Proof. Since $P(\mathbf{M}, \mathbf{X}, \mathbf{Y})$ is a degraded system with $I_P(\mathbf{M}; \mathbf{X}) \geq I_P(\mathbf{M}; \mathbf{Y})$, we know that $RI_P = I_P(\mathbf{M}; \mathbf{Y})$. By DPI,

$$I_P(\mathbf{M}; \mathbf{Y}) = RI_P \geq RI_S = I_S(\mathbf{M}; g(\mathbf{Y})).$$

By sufficiency of f:

$$\begin{aligned} I_P(\mathbf{M}; \mathbf{X}) &= UI_P(\mathbf{X}) + RI_P \\ &= UI_S(f(\mathbf{X})) + RI_S = I_S(\mathbf{M}; f(\mathbf{X})). \end{aligned}$$

Thus, $UI_S(f(\mathbf{X})) = UI_P(\mathbf{X}) + RI_P - RI_S$, which implies that $UI_S(f(\mathbf{X})) \geq UI_P(\mathbf{X})$. □

Prop. 4 states that if P is known to be a degraded system and we seek the LC where the more informative variable transmits to the redundant variable, the proportion of unique to redundant information *increases* in the latent representation, in contrast to the result of Prop. 3. Interpreting RI as the amount of information that is forwarded, Prop. 3 can be loosely viewed as a manifestation of DPI for forwarded information. Specifically, after projection of a degraded system to an LC, the latent transmitting population, $f(\mathbf{X})$, cannot transmit more information about $\mathbf{M}$ to $g(\mathbf{Y})$ than its high-dimensional counterpart, $\mathbf{X}$, to $\mathbf{Y}$.

VI. Discussion

We formalize the problem of low-dimensional neural population communication of stimulus (i.e. message) information. In our simplified setting of population communication, we assume that $\mathbf{M}$-forwarding in the high-dimensional system specifies a directed causal graph, i.e. the message causes subsequent neural activity. As noted in Sec. IV, we *cannot* make causal inferences about valid LCs, which are fundamentally statistical objects. Nevertheless, identification of valid LCs can inform experimental designs to test the existence of causal mechanisms, e.g., through optogenetic inactivation [6]. The question of how to define a 'true LC', representing a latent causal mechanism, from a high-dimensional structural causal model remains a promising avenue for future theoretical work.

Here, we focus on defining the message as an experimental stimulus (or a function thereof), which causes subsequent neural activity. It is equally interesting to characterize LCs when the message is a behavioral feature resulting from neural activity. Mathematically, this is simply a reversal of the Markov chain of $\mathbf{M}$-forwarding. However, behavioral features can correlate to a multitude of neural processes that can be difficult to separate. For example, the analysis in [1] was based on an experiment design where mice were trained to lick left or right depending on whether they detected whisker stimulation [6]. The animal's choice is a behavioral feature, but is correlated to sensory detection, decision-making, and motor neural processes, sometimes all in a single population. As such, additional care is required to define LCs when the message results from neural activity.

Acknowledgment

We thank Y. Kate Hong and Alice Y. Nam for insightful discussions.

References

[1] A. Merkley, A. Y. Nam, Y. K. Hong, and P. Grover, "Message-relevant dimension reduction of neural populations," in *2024 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2024.

[2] P. Venkatesh, S. Dutta, and P. Grover, "Information flow in computational systems," *IEEE Transactions on Information Theory*, vol. 66, no. 9, pp. 5456–5491, 2020.

[3] E. Gokcen, A. I. Jasper, J. D. Semedo, A. Zandvakili, A. Kohn, C. K. Machens, and B. M. Yu, "Disentangling the flow of signals between populations of neurons," *Nature Computational Science*, vol. 2, no. 8, pp. 512–525, 2022.

[4] C. Glymour, K. Zhang, and P. Spirtes, "Review of causal discovery methods based on graphical models," *Frontiers in genetics*, vol. 10, p. 524, 2019.

[5] J. Samuels and N. D. Zasler, *Event-Related Paradigms*. Cham: Springer International Publishing, 2018, pp. 1346–1347.

[6] Y. K. Hong, C. O. Lacefield, C. C. Rodgers, and R. M. Bruno, "Sensation, movement and learning in the absence of barrel cortex," *Nature*, vol. 561, no. 7724, pp. 542–546, 2018.

[7] D. Ghosh, "Sufficient dimension reduction: an information-theoretic viewpoint," *Entropy*, vol. 24, no. 2, p. 167, 2022.

[8] P. L. Williams and R. D. Beer, "Nonnegative decomposition of multivariate information," *arXiv preprint arXiv:1004.2515*, 2010.

[9] J. T. Lizier, N. Bertschinger, J. Jost, and M. Wibral, "Information decomposition of target effects from multi-source interactions: Perspectives on previous, current and future work," *Entropy*, vol. 20, no. 4, p. 307, 2018.

[10] P. Venkatesh and G. Schamberg, "Partial information decomposition via deficiency for multivariate Gaussians," in *2022 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2022, pp. 2892–2897.

[11] C. Goswami, A. Merkley, and P. Grover, "Computing unique information for Poisson and multinomial systems," *arXiv preprint arXiv:2305.07013*, 2023.

[12] C. Goswami and A. Merkley, "Analytically computing partial information decomposition," *Advances in Neural Information Processing Systems*, 2024.

[13] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.

[14] H. Bong, Z. Liu, Z. Ren, M. Smith, V. Ventura, and R. E. Kass, "Latent dynamic factor analysis of high-dimensional neural recordings," *Advances in Neural Information Processing Systems*, vol. 33, pp. 16 446–16 456, 2020.

[15] M. Ding, Y. Chen, and S. L. Bressler, "Granger causality: basic theory and application to neuroscience," *Handbook of time series analysis: recent theoretical developments and applications*, pp. 437–460, 2006.

[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, pp. 17–44, 2011.

[17] M. Wibral, R. Vicente, and M. Lindner, "Transfer entropy in neuroscience," *Directed information measures in neuroscience*, pp. 3–36, 2014.

[18] F. E. Rosas, P. A. Mediano, M. Gastpar, and H. J. Jensen, "Quantifying high-order interdependencies via multivariate extensions of the mutual information," *Physical Review E*, vol. 100, no. 3, p. 032305, 2019.

[19] P. Venkatesh, C. Bennett, S. Gale, T. Ramirez, G. Heller, S. Durand, S. Olsen, and S. Mihalas, "Gaussian partial information decomposition: Bias correction and application to high-dimensional data," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[20] M. Celotto, J. Bím, A. Tlaie, V. De Feo, A. Toso, S. Lemke, D. Chicharro, H. Nili, M. Bieler, I. Hanganu-Opatz *et al.*, "An information-theoretic quantification of the content of communication between brain regions," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[21] C. A. López and O. I. Lombardi, "No communication without manipulation: A causal-deflationary view of information," *Studies in History and Philosophy of Science Part A*, vol. 73, pp. 34–43, 2019.

[22] T. M. Cover and J. A. Thomas, "Elements of information theory," *John Wiley & Sons*, 2006.