PNAS RESEARCH ARTICLE | NEUROSCIENCE APPLIED MATHEMATICS

Causal connectivity measures for pulse-output network reconstruction: Analysis and applications

Zhong-qi K. Tian[a,b,c,1], Kai Chen[a,b,c,1], Songting Li[a,b,c,2], David W. McLaughlin[d,e,f,g,2], and Douglas Zhou[a,b,c,h,2]

Contributed by David W. McLaughlin; received April 3, 2023; accepted March 3, 2024; reviewed by David Hansel and Dario L. Ringach

The causal connectivity of a network is often inferred to understand network function. It is arguably acknowledged that the inferred causal connectivity relies on the causality measure one applies, and it may differ from the network's underlying structural connectivity. However, the interpretation of causal connectivity remains to be fully clarified, in particular, how causal connectivity depends on causality measures and how causal connectivity relates to structural connectivity. Here, we focus on nonlinear networks with pulse signals as measured output, e.g., neural networks with spike output, and address the above issues based on four commonly utilized causality measures, i.e., time-delayed correlation coefficient, time-delayed mutual information, Granger causality, and transfer entropy. We theoretically show how these causality measures are related to one another when applied to pulse signals. Taking a simulated Hodgkin–Huxley network and a real mouse brain network as two illustrative examples, we further verify the quantitative relations among the four causality measures and demonstrate that the causal connectivity inferred by any of the four well coincides with the underlying network structural connectivity, therefore illustrating a direct link between the causal and structural connectivity. We stress that the structural connectivity of pulse-output networks can be reconstructed pairwise without conditioning on the global information of all other nodes in a network, thus circumventing the curse of dimensionality. Our framework provides a practical and effective approach for pulse-output network reconstruction.

causality | correlation | mutual information | Granger causality | transfer entropy

Significance

Inferring network connectivity is a key challenge in many diverse scientific fields. We investigate networks with pulse signals as measured output and solve the above reverse-engineering issue by establishing a direct link between the network's causal connectivity and structural connectivity. Here, the causal connectivity can be inferred by any one of the four causality measures, i.e., time-delayed correlation coefficient, time-delayed mutual information, Granger causality, and transfer entropy. We analytically reveal the relationship among these four measures, and then we show that they can be used for network connectivity reconstruction. Our work provides a framework to potentially reconstruct the structural connectivity in general pulse-output nonlinear networks.

The structural connectivity of a network, such as a cortical network, is of great importance in understanding the cooperation and competition among nodes in the network (1–3). However, it is often difficult to measure directly the structural connectivity of a network. On the other hand, with the development of experimental techniques in neuroscience, it has become feasible to record simultaneously, with high temporal resolution, the activities of many nodes in a network (4–6). This provides a possibility to reveal the underlying connectivity of the network by analyzing the nodes' activity data and identifying the causal interactions (connectivity) among them (7–12).

There are difficulties in measuring these causal interactions: One of the most widely used statistical indicators for interaction identification is the correlation coefficient (13–15) that characterizes linear dependence between two nodes. The correlation coefficient is symmetric, and thus cannot distinguish the driver-recipient relation to recover the causal connectivity (16). To solve this, the time-delayed correlation coefficient (TDCC) (15, 17) is used to detect the direction of causal connectivity. However, TDCC, as a linear measure, may fail to capture causal interactions in nonlinear networks. As a nonlinear model-free generalization of TDCC, time-delayed mutual information (TDMI) (18–20) was proposed to measure the flow of information in nonlinear networks. Despite their mathematical simplicity and computational efficiency, TDCC and TDMI cannot exclude the historical effect of signals and may encounter the issue of overestimation (19–21). Granger causality (GC) (22–24) and transfer entropy (TE) (19, 25, 26) were two other measures that were introduced to detect causal connectivity with the exclusion of the signal's own historical effects. GC is based on linear regression models and assumes the causal relation can be revealed by analyzing low-order statistics of signals (up to the variance). Consequently, the validity of GC for nonlinear networks is in general questionable (27). In contrast, TE is a nonparametric information-theoretic measure that quantifies the causal interactions with no assumption of interaction models. However, it requires the estimation of the probability distribution of dynamical variables conditioned on the historical information in networks, which makes TE suffer from the curse of dimensionality in practical applications to network systems with many nodes (28–30). It is known that the causal connectivity inferred by different causality measures can

Author contributions: Z.-q.K.T., K.C., S.L., D.W.M., and D.Z. designed research; performed research; analyzed data; and wrote the paper.

Reviewers: D.H., Centre de Neurophysique, Université Paris Descartes; and D.L.R., University of California Los Angeles.

The authors declare no competing interest.

Copyright © 2024 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]Z.-q.K.T. and K.C. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: songting@sjtu.edu.cn, david.mclaughlin@nyu.edu, or zdz@sjtu.edu.cn.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2305297121/-/DCSupplemental.

Published March 28, 2024.

Downloaded from https://www.pnas.org by Kai Chen on March 31, 2024 from IP address 58.247.22.126.

be inconsistent with each other (31–33). With one exception [TE has been proven to be equivalent to GC for Gaussian variables (34)], there is little understanding of the relationships between these measurement techniques, including when they accurately (or inaccurately) predict causal interaction (connectivity).

The determination of structural connectivity from causal interactions faces even more fundamental difficulties: First, causal interactions may not be the result of direct structural connections between the nodes of the network, but rather the result of correlations arising from common external stimuli, or correlations arising from indirect connections (for example, there are two direct connections, from X to Y and from Y to Z, denoted by $X \to Y$ and $Y \to Z$, respectively. This results in correlations between X and Z, yet no direct connection from X to Z). For clarity, a pair of neurons Y and X with a synaptic connection from Y to X, denoted as $Y \to X$, are called d-conn (directly connected) (Y, X) pair. Otherwise, we term them an id-conn (indirectly connected) (Y, X) pair. Second, causal connectivity, as inferred from any of the above four measures, is statistical rather than structural (35–39), i.e., causal connectivity quantifies direct statistical correlation among network nodes, whereas structural connectivity corresponds to physical connections among network nodes. In general, it is not clear whether structural connectivity can be reconstructed from causal connectivity.

In this work, we address the above issues for an important class of nonlinear networks, which we term pulse-output networks and which include spiking neural networks. The activity of each node in a pulse-output network can be described as a stochastic binary time series for the presence/absence of a pulse (spike) in each time window, e.g., spike train. Under this description, we illustrate mathematically that the four causal connectivity measures (TDCC, TDMI, GC, and TE) can be represented by one another, to leading order in a perturbation expansion. Next, both by simulations of a Hodgkin–Huxley (HH) neural network and with experimentally measured data from mouse cortical network (40, 41), we verify that the mathematical relations among the four causality measures are valid for representative samples of pulse-output networks. More importantly, for pulse-output networks, we demonstrate that the underlying structural connectivity can be recovered from the causal connectivity, itself inferred from any one of four causality measures. For pulse-output networks of both simulated Hodgkin–Huxley neural networks and a real mouse cortical network (40, 41), we show that the potential problems with the recovery of structural connectivity from causal connectivity, e.g., confounders and hidden nodes, can be resolved.

We emphasize that the pulse-output nature of a spiking neuronal network allows one to represent the neuronal signal as a binary time series, with random spike times. With the utility of this stochastic binary representation, our analytical framework shows, using only pairwise information between neurons, i) the establishment of mathematical relationships between four common measures of causal connectivity, and ii) the accurate pairwise predictions of structural connectivity from causal connectivity, without conditioning on the global information from all other nodes. Thus, the reconstruction can circumvent the curse of dimensionality and can be applied to the reconstruction of structural connectivity in large-scale pulse-output nonlinear systems or subsystems.

Results

Concepts of Generalized Pairwise TDCC, TDMI, GC, and TE. Consider a nonlinear network of N nodes with dynamics given by

$$\frac{d\mathbf{Z}}{dt} = \mathbf{F}(\mathbf{Z}, t), \qquad [1]$$

where $\mathbf{Z} = (Z_1, Z_2, ..., Z_N)$. We focus on the application of TDCC, TDMI, GC, and TE to each pair of nodes without conditioning on the rest of nodes in the network, accounting for the practical constraint that conditional causality measures in general require the information of the whole network that is often difficult to observe. For the ease of illustration, we denote a pair of nodes as $X = Z_i$ and $Y = Z_j$, and their measured time series as $\{x_n\}$ and $\{y_n\}$, respectively.

TDCC (15, 17), as a function of time delay m, is defined by

$$C(X, Y; m) = \frac{\mathrm{cov}(x_n, y_{n-m})}{\sigma_x \sigma_y},$$

where "cov" represents the covariance, σ_x and σ_y are the standard deviations of $\{x_n\}$ and $\{y_n\}$, respectively. A positive (negative) value of m indicates the calculation of causal value from Y to X (from X to Y), and nonzero $C(X, Y; m)$ indicates the existence of causal interaction between X and Y. Without loss of generality, we consider the case of positive m in the following discussions, that is, the causality measure from Y to X.

In contrast to the linear measure TDCC, TDMI is a model-free method being able to characterize nonlinear causal interactions (18–20). TDMI from Y to X is defined by

$$I(X, Y; m) = \sum_{x_n, y_{n-m}} p(x_n, y_{n-m}) \log \frac{p(x_n, y_{n-m})}{p(x_n) p(y_{n-m})}, \qquad [2]$$

where $p(x_n, y_{n-m})$ is the joint probability distribution of x_n, y_{n-m}, $p(x_n)$ and $p(y_{n-m})$ are the corresponding marginal probability distributions. $I(X, Y; m)$ is nonnegative and vanishes if and only if x_n and y_{n-m} are independent (20). Nonzero $I(X, Y; m)$ implies the existence of causal interaction from Y to X for a positive m.

It has been noted that TDCC and TDMI could overestimate the causal interactions when a signal has a long memory (19–21). As an alternative, GC was proposed to overcome the issue of overestimation based on linear regression (22, 23, 42). The autoregression for X is represented by $x_{n+1} = a_0 + \sum_{i=1}^{k} a_i x_{n+1-i} + \epsilon_{n+1}$, where $\{a_i\}$ are the estimated autoregression coefficients and ϵ_{n+1} is the residual. By including the historical information of Y with a message length l and a time-delay m, the joint regression for X is represented by $x_{n+1} = \tilde{a}_0 + \sum_{i=1}^{k} \tilde{a}_i x_{n+1-i} + \sum_{j=1}^{l} \tilde{b}_j y_{n+2-m-j} + \eta_{n+1}$, where $\{\tilde{a}_i\}$ and $\{\tilde{b}_j\}$ are the estimated joint regression coefficients, and η_{n+1} is the corresponding residual. If there exists a causal interaction from Y to X, then the prediction of X using the linear regression models shall be improved by additionally incorporating the historical information of Y. Accordingly, the variance of residual η_{n+1} is smaller than that of ϵ_{n+1}. Based on this concept, the GC value from Y to X is defined by

$$G_{Y \to X}(k, l; m) = \log \frac{\mathrm{Var}(\epsilon_{n+1})}{\mathrm{Var}(\eta_{n+1})}.$$

The GC value is also nonnegative and vanishes if and only if $\{\tilde{b}_j\} = 0$, i.e., the variance of residual ϵ_{n+1} for X cannot be reduced by including the historical information of Y. Note that, by introducing the time-delay parameter m, the GC analysis defined above generalizes the conventional GC analysis, as the latter corresponds to the special case of $m = 1$.

Downloaded from https://www.pnas.org by Kai Chen on March 31, 2024 from IP address 58.247.22.126.

GC assumes that the causal interaction can be fully captured by the variance reduction in the linear regression models, which is valid for Gaussian signals but not for more general signals. As a nonlinear extension of GC, TE was proposed to describe the causal interaction from the information theoretic perspective (19). The TE value from Y to X is defined by

$$T_{Y\to X}(k,l;m)=\sum_{x_{n+1},x_n^{(k)},y_{n+1-m}^{(l)}} p(x_{n+1},x_n^{(k)},y_{n+1-m}^{(l)}) \cdot \log\frac{p(x_{n+1}|x_n^{(k)},y_{n+1-m}^{(l)})}{p(x_{n+1}|x_n^{(k)})}, \quad [3]$$

where the shorthand notation $x_n^{(k)}=(x_n,x_{n-1},\ldots,x_{n-k+1})$ and $y_{n+1-m}^{(l)}=(y_{n+1-m},y_{n-m},\ldots,y_{n+2-m-l})$, k, l indicate the length (order) of historical information of X and Y, respectively. Similar to GC, the time-delay parameter m is introduced that generalizes the conventional TE, the latter of which corresponds to the case of $m=1$. TE is nonnegative and vanishes if and only if $p(x_{n+1}|x_n^{(k)},y_{n+1-m}^{(l)})=p(x_{n+1}|x_n^{(k)})$, i.e., the uncertainty of x_{n+1} is not affected regardless of whether the historical information of Y is taken into account.

In this work, we investigate the mathematical relations among TDCC, TDMI, GC, and TE by focusing on nonlinear networks described by Eq. **1** with pulse signals as measured output, e.g., the spike trains measured in neural networks. Consider a pair of nodes X and Y in the network of N nodes, and denote their pulse-output signals by

$$w_x(t)=\sum_l \delta(t-\tau_{xl}) \quad \text{and} \quad w_y(t)=\sum_l \delta(t-\tau_{yl}), \quad [4]$$

respectively, where $\delta(\cdot)$ is the Dirac delta function, and $\{\tau_{xl}\}$ and $\{\tau_{yl}\}$ are the output time sequences of nodes X and Y determined by Eq. **1**, respectively. With the sampling resolution of Δt, the pulse-output signals are measured as binary time series $\{x_n\}$ and $\{y_n\}$, where $x_n=1$ ($y_n=1$) if there is a pulse signal, e.g., a spike generated by a neuron, of X (Y) occurred in the time window $[t_n,t_n+\Delta t)$, and $x_n=0$ ($y_n=0$) otherwise, i.e.,

$$x_n=\int_{t_n}^{t_n+\Delta t} w_x(t)dt \quad \text{and} \quad y_n=\int_{t_n}^{t_n+\Delta t} w_y(t)dt, \quad [5]$$

and $t_n=n\Delta t$. Note that the value of Δt is often chosen to make sure that there is at most one pulse signal in a single small enough time window. In the stationary state, the responses x_n and y_n can be viewed as stochastic processes when the network is driven by stochastic external inputs. In such a case, for the sake of simplicity, we denote $p_x=p(x_n=1)$, $p_y=p(y_n=1)$, and define $\Delta p_m=\frac{p(x_n=1,y_{n-m}=1)}{p(x_n=1)p(y_{n-m}=1)}-1$, measuring the dependence between x_n and y_{n-m}.

Mathematical Relation between TDMI and TDCC. For the relation between TDCC and TDMI when applied to nonlinear networks with pulse-output signals, we prove the following theorem:

Theorem 1. *For nodes X and Y with pulse-output signals given in Eqs. **4** and **5**, we have*

$$I(X,Y;m)=\frac{C^2(X,Y;m)}{2}+O(\Delta t^2\Delta p_m^3), \quad [6]$$

where the symbol "O" stands for the order.

Proof: The basic idea is to Taylor expand TDMI in Eq. **2** with respect to the term $\frac{p(x_n,y_{n-m})}{p(x_n)p(y_{n-m})}-1$ (the detailed derivation can be found in *SI Appendix*, Supporting Information Text 1B), then we arrive at the following expression:

$$\begin{aligned} I(X,Y;m)&=\sum_{x_n,y_{n-m}} p(x_n,y_{n-m})\log\left(1+\frac{p(x_n,y_{n-m})}{p(x_n)p(y_{n-m})}-1\right)\\ &=\frac{\left[p(x_n=1,y_{n-m}=1)-p_xp_y\right]^2}{2(p_x-p_x^2)(p_y-p_y^2)}+O(\Delta t^2\Delta p_m^3). \end{aligned}$$

Since TDCC can be written as

$$C(X,Y;m)=\frac{p(x_n=1,y_{n-m}=1)-p_xp_y}{\sqrt{(p_x-p_x^2)(p_y-p_y^2)}}, \quad [7]$$

we have

$$I(X,Y;m)=\frac{C^2(X,Y;m)}{2}+O(\Delta t^2\Delta p_m^3).$$

□

Mathematical Relation between GC and TDCC. We next derive the relation between GC and TDCC as follows:

Theorem 2. *For nodes X and Y with pulse-output signals given in Eqs. **4** and **5**, we have*

$$G_{Y\to X}(k,l;m)=\sum_{i=m}^{m+l-1} C^2(X,Y;i)+O(\Delta t^3\Delta p_m^2). \quad [8]$$

Proof: From the definition, GC can be represented by the covariances of the signals (34) as

$$G_{Y\to X}(k,l;m)=\log\frac{\Gamma(x_{n+1}|x_n^{(k)})}{\Gamma(x_{n+1}|x_n^{(k)}\oplus y_{n+1-m}^{(l)})}, \quad [9]$$

where $\Gamma(\mathbf{x}|\mathbf{y})=\mathrm{cov}(\mathbf{x})-\mathrm{cov}(\mathbf{x},\mathbf{y})\mathrm{cov}(\mathbf{y})^{-1}\mathrm{cov}(\mathbf{x},\mathbf{y})^T$ for random vectors $\mathbf{x}$ and $\mathbf{y}$, $\mathrm{cov}(\mathbf{x})$ and $\mathrm{cov}(\mathbf{y})$ denote the covariance matrix of $\mathbf{x}$ and $\mathbf{y}$, respectively, and $\mathrm{cov}(\mathbf{x},\mathbf{y})$ denote the cross-covariance matrix between $\mathbf{x}$ and $\mathbf{y}$. The symbol T is the transpose operator and $\oplus$ denotes the concatenation of vectors.

We first prove that the autocorrelation function (ACF) of binary time series $\{x_n\}$ as a function of time delay takes the order of Δt (see *SI Appendix*, Supporting Information Text 1C for the proof, *SI Appendix*, Fig. S1). Accordingly, we have

$$\mathrm{cov}(x_n^{(k)})=\sigma_x^2(\mathbf{I}+\hat{\mathbf{A}}),$$

where $\hat{\mathbf{A}}=(\hat{a}_{ij})$, $\hat{a}_{ij}=O(\Delta t)$, and $\mathbf{I}$ is the identity matrix. Hence,

$$\Gamma(x_{n+1}|x_n^{(k)})=\sigma_x^2-\frac{1}{\sigma_x^2}\mathrm{cov}(x_{n+1},x_n^{(k)})(\mathbf{I}-\hat{\mathbf{A}}) \cdot \mathrm{cov}(x_{n+1},x_n^{(k)})^T+O(\Delta t^5). \quad [10]$$

Downloaded from https://www.pnas.org by Kai Chen on March 31, 2024 from IP address 58.247.22.126.

In the same way, we have

$$\begin{aligned}\Gamma(x_{n+1}|x_n^{(k)} \oplus y_{n+1-m}^{(l)}) &= \sigma_x^2 - \frac{1}{\sigma_x^2}\mathrm{cov}(x_{n+1}, x_n^{(k)})(\mathbf{I}-\hat{\mathbf{A}})\mathrm{cov}(x_{n+1}, x_n^{(k)})^T \\ &\quad - \frac{1}{\sigma_y^2}\mathrm{cov}(x_{n+1}, y_{n+1-m}^{(l)})(\mathbf{I}-\hat{\mathbf{B}})\mathrm{cov}(x_{n+1}, y_{n+1-m}^{(l)})^T \\ &\quad + O(\Delta t^4 \Delta p_m^2). \end{aligned} \quad [11]$$

where $\hat{\mathbf{B}} = (\hat{b}_{ij})$, $\hat{b}_{ij} = O(\Delta t)$. Substituting Eqs. **10** and **11** into Eq. **9** and Taylor expanding Eq. **9** with respect to Δt, we can obtain

$$G_{Y\to X}(k, l; m) = \sum_{i=m}^{m+l-1} C^2(X, Y; i) + O(\Delta t^3 \Delta p_m^2). \quad [12]$$

The detailed derivation of Eqs. **10**, **11**, and **12** can be found in *SI Appendix*, Supporting Information Text 1C. □

Mathematical Relation between TE and TDMI. From the definitions of TE and TDMI, TE can be regarded as a generalization of TDMI conditioning on the signals' historical information additionally. To rigorously establish their relationship, we require that $\|x_{n+1}^{(k+1)}\|_0 \le 1$ and $\|y_{n+1-m}^{(l)}\|_0 \le 1$ in the definition of TE given in Eq. **3**, where $\|\cdot\|_0$ denotes the l_0 norm of a vector, i.e., the number of nonzero elements in a vector. This assumption indicates that the length of historical information used in the TE framework is shorter than the minimal time interval between two consecutive pulse-output signals. With this condition, we mathematically establish the following theorem:

Theorem 3. *For nodes X and Y with pulse-output signals given in Eqs. **4** and **5**, under the assumption that $\|x_{n+1}^{(k+1)}\|_0 \le 1$ and $\|y_{n+1-m}^{(l)}\|_0 \le 1$, we have*

$$T_{Y\to X}(k, l; m) = \sum_{i=m}^{m+l-1} I(X, Y; i) + O(\Delta t^3 \Delta p_m^2), \quad [13]$$

*where $T_{Y\to X}$ is defined in Eq. **3**.*

Proof: To simplify the notation, we denote $x_n^{(k)} = (x_n, x_{n-1}, \ldots, x_{n-k+1})$ and $y_{n+1-m}^{(l)} = (y_{n+1-m}, y_{n-m}, \ldots, y_{n+2-m-l})$ as x^- and y^-, respectively. From Eq. **3**, we have

$$\begin{aligned} T_{Y\to X}(k, l; m) &= \sum_{x_{n+1}, x^-, y^-} p(x_{n+1}, x^-, y^-) \log \frac{p(x_{n+1}|x^-, y^-)}{p(x_{n+1}|x^-)} \\ &= \sum_{i=m}^{m+l-1} I(X, Y; i) + \mathcal{A} + \mathcal{B}, \end{aligned} \quad [14]$$

where

$$\mathcal{A} = \sum_{x_{n+1}, y^-} p(x_{n+1}, y^-) \log \frac{p(y^-|x_{n+1})}{\prod_j p(y_j|x_{n+1})} \frac{\prod_j p(y_j)}{p(y^-)}$$

and

$$\mathcal{B} = \sum_{x_{n+1}, x^-, y^-} p(x_{n+1}, x^-, y^-) \log \frac{p(x_{n+1}|x^-, y^-)}{p(x_{n+1}|y^-)} \frac{p(x_{n+1})}{p(x_{n+1}|x^-)}$$

and $\prod_j$ in $\mathcal{A}$ represents $\prod_{j=n+2-m-l}^{n+1-m}$. The detailed derivation of Eq. **14** can be found in *SI Appendix*, Supporting Information Text 1D. $\mathcal{A}$ and $\mathcal{B}$ in Eq. **14** consist of multiple terms, and the leading order of each term can be analytically calculated. For the sake of illustration, we derive the leading order of one of these terms and the leading order of the rest terms can be estimated in a similar way. Under the assumption that $\|x_{n+1}^{(k+1)}\|_0 \le 1$ and $\|y_{n+1-m}^{(l)}\|_0 \le 1$, the number of nonzero components is at most one in $x_{n+1}^{(k+1)}$ and $y_{n+1-m}^{(l)}$. Without loss of generality, we assumed $x_{n+1} = 1$ and $y_{n+1-m} = 1$. In such a case, we can obtain the following expression in $\mathcal{A}$,

$$\begin{aligned} &p(x_{n+1}, y^-) \log \frac{p(y^-|x_{n+1})}{\prod_j p(y_j|x_{n+1})} \frac{\prod_j p(y_j)}{p(y^-)}\Big|_{x_{n+1}=1, y_{n+1-m}=1} \\ &\quad = p_x p_y^2 \sum_{j\neq n+1-m} \Delta p_{n+1-j} + O(\Delta t^3 \Delta p_m^2). \end{aligned}$$

We can further show that the leading order of all the terms in $\mathcal{A}$ cancel each other out (*SI Appendix*, Supporting Information Text 1D), thus we have $\mathcal{A} = O(\Delta t^3 \Delta p_m^2)$. Similarly, we can also show $\mathcal{B} = O(\Delta t^3 \Delta p_m^2)$ (*SI Appendix*, Supporting Information Text 1D), and thus

$$T_{Y\to X}(k, l; m) = \sum_{i=m}^{m+l-1} I(X, Y; i) + O(\Delta t^3 \Delta p_m^2).$$

□

Mathematical Relation between GC and TE. From Theorems 1–3, we can prove the following theorem (see details in *SI Appendix*, Supporting Information Text 1E):

Theorem 4. *For nodes X and Y with pulse-output signals given in Eqs. **4** and **5**, under the assumption that $\|x_{n+1}^{(k+1)}\|_0 \le 1$ and $\|y_{n+1-m}^{(l)}\|_0 \le 1$, we have*

$$G_{Y\to X}(k, l; m) = 2T_{Y\to X}(k, l; m) + O(\Delta t^2 \Delta p_m^3), \quad [15]$$

*where $T_{Y\to X}$ is defined in Eq. **3**.*

Note that in Theorems 3 and 4, we require the condition $\|x_{n+1}^{(k+1)}\|_0 \le 1$ and $\|y_{n+1-m}^{(l)}\|_0 \le 1$ to establish the relations of causality measures. However, the relations of the four causality measures remain to hold approximately in the absence of this condition, as will be discussed below.

We summarize the relations among the four causality measures in Fig. 1. Note that the relations depend on the condition of weak dependence between X and Y described by small Δp_m (see derivations in *SI Appendix*, Supporting Information Text 1). We note that in some extreme case with very strong inhibition, Δp_m can approach -1 so that the relations no longer hold. In such case of $\Delta p_m \approx -1$, different equivalence relations among the four causality measures apply, as developed in *SI Appendix*, Supporting Information Text 2. We describe the numerical verification of these relations in the *Discussion*.

Downloaded from https://www.pnas.org by Kai Chen on March 31, 2024 from IP address 58.247.22.126.

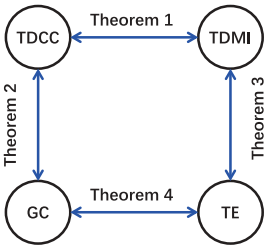

Fig. 1. Mathematical relations among TDCC, TDMI, GC, and TE established by four theorems.

Mathematical Relations of Causality Measures Verified in HH Neural Networks. To verify the relations among the causality measures derived above, as an illustrative example, we apply generalized pairwise TDCC, TDMI, GC, and TE to the HH neural network described in *Materials and Methods*. We first consider a d-conn pair of neurons $Y \to X$ in an HH network containing 10 excitatory neurons driven by independent homogeneous Poisson inputs. Let $\{\tau_{xl}\}$ and $\{\tau_{yl}\}$ be the ordered spike times of neuron X and Y in the HH network respectively and denote their spike trains as $w_x(t) = \sum_l \delta(t - \tau_{xl})$ and $w_y(t) = \sum_l \delta(t - \tau_{yl})$, respectively. With a sampling resolution of Δt, the spike train is measured as a binary time series as described above. To numerically verify the above theorems, we then check the order of the remainders in Eqs. **6**, **8**, **13**, and **15** in terms of Δt and Δp_m. Note that Δp_m, the measure of the dependence between X and Y, is insensitive to sampling resolution Δt (*SI Appendix*, Fig. S2). Therefore, by varying sampling interval Δt and coupling strength S (linearly related to Δp_m), respectively, the orders of the remainders are consistent with those derived in Eqs. **6**, **8**, **13**, and **15** as shown in Fig. 2. In addition, Fig. 3 verifies the relations among the causality measures by changing other parameters. For example, in Eqs. **8**, **13**, and **15**, the four causality measures are proved to be independent of the historical length k, which is numerically verified in Fig. 3*A*. And although the values of GC and TE rely on the historical length l in Fig. 3*B*, the mathematical relations among the four causality measures revealed by Theorems 1–4 hold for a wide range of l. Recall that, in order to establish rigorously the relations of the causality measures of Theorems 3 and 4, our proofs required the assumption $\left\|x_{n+1}^{(k+1)}\right\|_0 \leq 1$ and $\left\|y_{n+1-m}^{(l)}\right\|_0 \leq 1$. This assumption can often be satisfied: For example, in the above simulations, the memory time of the neuron is about 20 ms, while the interspike interval is around

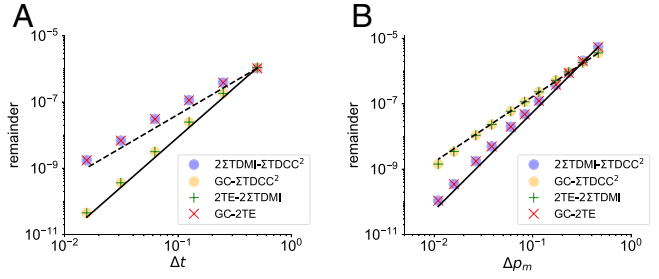

Fig. 2. Numerical verification of the convergence order of the remainders in terms of (*A*) Δt and (*B*) Δp_m. The convergence orders for Δt in (*A*) and Δp_m in (*B*) agree well with Theorems 1–4 ($R^2 > 0.998$). The gray dashed and solid lines indicate the 2nd-order and 3rd-order convergence, respectively. The four causality measures are calculated from a d-conn pair in an HH network of 10 excitatory neurons randomly connected with probability 0.25. The parameters are set as $k = l = 5$ and $m = 6$ (time delay is 3 ms), $S = 0.02$ mS cm^{-2} in (*A*), and $\Delta t = 0.5$ ms in (*B*).

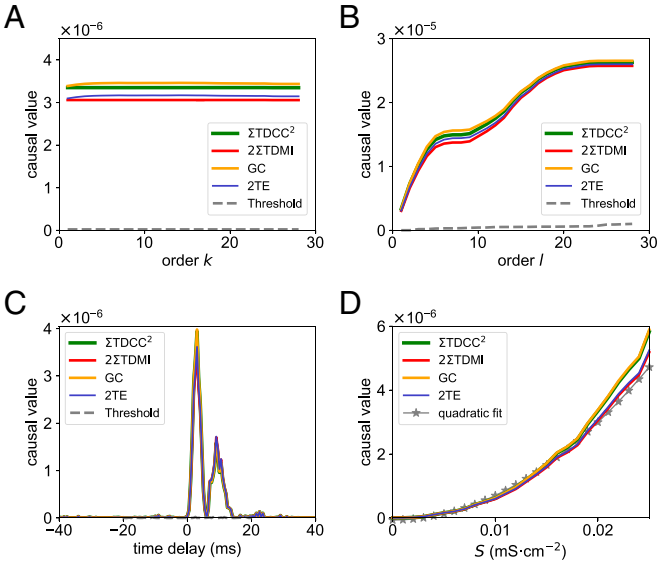

Fig. 3. Dependence of causal values on parameters of (*A*) order k, (*B*) order l, (*C*) time delay, and (*D*) coupling strength S obtained from the same pair of neurons in the HH network in Fig. 2. In (*C*), a positive (negative) time delay indicates the calculation of causal values from neuron Y to neuron X (from X to Y). The green curve represents the summation of squared TDCC $C(X, Y; m)$, the red curve represents twice of the summation of TDMI $I(X, Y; m)$, the orange curve stands for GC $G_{Y\to X}(k, l; m)$, and the blue curve stands for twice of TE $T_{Y\to X}(k, l; m)$. The curves virtually overlap in (*A*)–(*D*) (all significantly greater than those of randomly surrogate time series, $P < 0.05$). The gray dashed curve in (*A*)–(*C*) is the significance level of causality for an id-conn pair in the same HH network. The gray star curve in (*D*) is the quadratic fit of causal values with respect to different S. The parameters are set as (*A*): $l = 1$, $S = 0.02$ mS cm^{-2}, $\Delta t = 0.5$ ms, $m = 6$; (*B*): $k = 1$, $S = 0.02$ mS cm^{-2}, $\Delta t = 0.5$ ms, $m = 6$; (*C*): $k = l = 1$, $S = 0.02$ mS cm^{-2}, $\Delta t = 0.5$ ms; (*D*): $k = l = 1$, $\Delta t = 0.5$ ms, $m = 6$.

100 ms. However, even when the assumption breaks down in the regime of high firing rate, the mathematical relations among the four causality measures remain to hold approximately, as examined numerically in *SI Appendix*, Fig. S3.

We next verify the mathematical relations among the causality measures for the parameter of time delay m by fixing parameters k and l. In principle, the value of k and l in GC and TE shall be determined by the historical memory of the system. To reduce the computational cost (15, 43, 44), we take $k = l = 1$ for all the results below. It turns out that this parameter choice works well for pulse-output networks because of the short memory effect in general, as will be further discussed later. Fig. 3*C* shows the mathematical relations hold for a wide range of the time-delay parameter used in computing the four causality measures (see more examples in *SI Appendix*, Fig. S4).

We further examine the robustness of the mathematical relations among TDCC, TDMI, GC, and TE by scanning the parameters of the coupling strength S between the HH neurons and external Poisson input strength and rates. As shown in Fig. 3*D*, the values of the four causality measures with different coupling strength are very close to one another. Their relations also hold for a wide range of external Poisson input parameters (*SI Appendix*, Fig. S5). From the above, the mathematical relations among TDCC, TDMI, GC, and TE described in Theorems 1–4 are verified in the HH network.

Relation between Structural Connectivity and Causal Connectivity in HH Neural Networks. We next discuss the relation between the inferred causal connectivity and the structural

Downloaded from https://www.pnas.org by Kai Chen on March 31, 2024 from IP address 58.247.22.126.

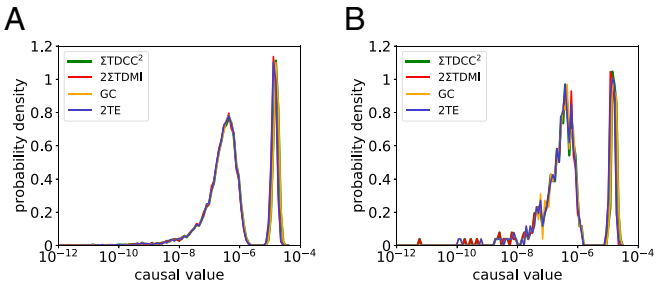

Fig. 4. Distributions of causal values in an HH network of 100 excitatory neurons randomly connected with probability 0.25. (*A*) The distribution of causal values of each pair of neurons in the whole network. (*B*) The distribution of causal values of each pair of neurons in an HH subnetwork of 20 excitatory neurons. The parameters are set as $k = l = 1$, $S = 0.02$ mS cm^{-2}, $\Delta t = 0.5$ ms, and $m = 6$. The colors are the same as those in Fig. 3 and the curves nearly overlap.

connectivity. Note that the causal connectivity inferred by these measures is statistical rather than structural (35–37), i.e., the causal connectivity quantifies the direct statistical correlation or dependence among network nodes, whereas the structural connectivity corresponds to physical connections among network nodes. Therefore, a precise relationship between causal connectivity and structural connectivity has been unclear. In Fig. 3*C*, the peak causal value from Y to X (at time delay around 3 ms, $m = 6$) is greater than the significance level (the gray dashed line in Fig. 3*C*), while the causal value from X to Y is not. Based on this, the inferred direct causal connections between X and Y are consistent with the underlying structural connections. From now on, we adopt peak causal values, $m = 6$, to represent the causal connectivity unless noted explicitly.

To investigate the validity of this consistency in larger networks, we further investigate a larger HH network (100 excitatory neurons) with random connectivity structure and homogeneous coupling strength (*SI Appendix*, Fig. S6*A*). As shown in Fig. 4*A*, the distributions of all four causal values across all pairs of neurons virtually overlap, which again verifies their mathematical relations given by Theorems 1–4. In addition, as the network size increases from 10 to 100 neurons, the distributions of the causality measures in Fig. 4*A* exhibit a bimodal structure with a clear separation of orders. By mapping the causal values with the structural connectivity, we find that the right bump of the distributions with larger causal values corresponds to d-conn pairs, while the left bump, with smaller causal values, corresponds to id-conn pairs. The well separation of the two modals indicates that the underlying structural connectivity in the HH network can be accurately estimated from the causal connectivity.

The performance of this reconstruction approach can be quantitatively characterized by the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) (31, 45, 46) (*Materials and Methods*). It is found that the AUC value becomes 1 when applying any of these four causality measures (*SI Appendix*, Fig. S6*B*), which indicates that the structural connectivity of the HH network could be reconstructed with 100% accuracy. We point out that the reconstruction of network connectivity based on causality measures is achieved by calculating the causal values between each pair of neurons that requires no access to the activity data of the rest of neurons. Therefore, this inference approach can be applied to a subnetwork when the activity of neurons outside the subnetwork is not observable. For example, when a subnetwork of 20 excitatory HH neurons is observed, the structural connectivity of the subnetwork can still be accurately reconstructed without knowing the information of the rest 80 neurons in the full network as shown in Fig. 4*B*. In such a case, the AUC values corresponding to the four causality measures are 1 (*SI Appendix*, Fig. S6*C*). In addition, we have also shown that, for HH networks with *heterogeneous* coupling strength (for example, the coupling strength follows a log-normal distribution), accurate reconstructions of network structural connectivity can still be achieved (*SI Appendix*, Fig. S7). To increase biological plausibility, we have enlarged network size to 10^4 neurons activated in a balanced excitation-inhibition state (Networks in balanced states in the *Discussion*). In such a large balanced network, accurate reconstruction performance is preserved.

Mechanism Underlying Network Connectivity Reconstruction by Causality Measures. We next demonstrate the mechanism underlying the validity of pairwise inference of pulse-output signals in the reconstruction of network structural connectivity. It has been frequently noticed that pairwise causal inference may potentially fail to distinguish d-conn from id-conn pairs (16, 47). For example, in the case that $Y \to W \to X$, the id-conn (Y, X) pair, may possibly be mis-inferred as a d-conn pair via pairwise causality measures. However, pulse-output signals circumvent such spurious inferences as explained below. Here we take TDCC as an example to elucidate the underlying reason for successful reconstruction. If we denote $\delta p_{Y\to X} = p(x_n = 1|y_{n-m} = 1) - p(x_n = 1|y_{n-m} = 0)$ as the increment of probability of generating a pulse output for neuron X induced by a pulse-output signal of neuron Y at m time step earlier, we have TDCC $C(X, Y; m) = \delta p_{Y\to X}\sqrt{\frac{p_y - p_y^2}{p_x - p_x^2}}$ through Eq. 7. For the case of $Y \to W \to X$, we can derive $\delta p_{Y\to X} = O(\delta p_{Y\to W} \cdot \delta p_{W\to X})$, and further $C(X, Y; m) = O(C(W, Y; m) \cdot C(X, W; m))$ (see derivation details about the relations among δp, Δp_m, and causality measures in *SI Appendix*, Supporting Information Text 3 and Fig. S8*A*). Because the influence of a single pulse-output signal is often small (e.g., in the HH neural network with physiologically realistic coupling strengths, we obtain $|\delta p| < 0.01$ from simulation data), the causal value $C(X, Y; m)$ for the id-conn (Y, X) pair is significantly smaller than $C(W, Y; m)$ or $C(X, W; m)$ for d-conn (Y, W) and (W, X) pairs as depicted in Fig. 4*A*. Another key concern in pairwise inference arises from correlated activity triggered by shared upstream drivers, termed "confounders." They will possibly result in the misidentification of id-conn pairs as d-conn pairs, an issue observed in traditional inference methodologies (47–50). For neuron pairs influenced by a common driver, represented as $Y \leftarrow W \to X$, where W acts as the confounder for X and Y, a similar relation can be derived: $\delta p_{Y\to X}$ (or $\delta p_{X\to Y}$) $= O(\delta p_{W\to Y} \cdot \delta p_{W\to X})$ (*SI Appendix*, Supporting Information Text 3 and Fig. S8*B*). As a consequence, the indirect causality resulting from confounders is also significantly diminished when compared to that from direct interactions.

We also note that the increment δp is linearly dependent on the coupling strength S (*SI Appendix*, Fig. S8*C*), thus we establish a mapping between the causal and structural connectivity, in which the causal value of TDCC is proportional to the coupling strength S between two d-conn neurons. The mapping between causal and structural connectivity for TDMI, GC, and TE is also established in a similar way, in which the corresponding causal values are proportional to S^2 as shown in Fig. 3*D*. Therefore, the application of pairwise causality measures to pulse-output signals is able to successfully reveal the underlying structural connectivity of a network. Furthermore, leveraging this

Downloaded from https://www.pnas.org by Kai Chen on March 31, 2024 from IP address 58.247.22.126.

quadratic relation between causal value and coupling strength, it is possible to estimate the coupling strength rather than "coupled or not coupled" based on the causal value, which requires more future studies. In summary, for pulse-output networks, structural connectivity can be accurately inferred from causal connectivity in a pairwise manner that overcomes computational issues of high dimensionality. Thus, the method is potentially applicable to experimentally measured data from large-scale biological networks or subnetworks as discussed below.

Network Connectivity Reconstruction with Physiological Experimental Data. Next, we apply all four causality measures to experimental data to address the issue of validity of their mathematical relations and reconstruction of the network structural connectivity. Here, we analyze the in vivo spike data recorded in the mouse cortex from Allen Brain Observatory (40, 41) (*Materials and Methods*). By applying the four causality measures, we infer the causal connectivity of those cortical networks. As the underlying structural connectivity of the recorded neurons in experiments is unknown, we first detect putative connected links from the distribution of causality measures, and then follow the same procedures as previously described in the HH model case using ROC and AUC in signal detection theory (31, 45, 46) (*Materials and Methods*) to quantify the reconstruction performance.

Because we have demonstrated the equivalence of the four causality measures, we will use TE as a representative causality measure to demonstrate a way to detect putative connections. As we have shown above, the TE values are proportional to S^2. In addition, previous experimental works observed that the coupling strength S follows the log-normal distribution both for intra-areal and interareal cortical networks in mouse and monkey brains (51). Thus, the distribution of TE for d-conn pairs should follow the log-normal distribution. In addition, we assume that the distribution of TE for id-conn pairs also obeys the log-normal distribution (*Materials and Methods*).

To validate the above assumption of two log-normal distributions of TE values, we consider an HH network where neurons are randomly connected and the network coupling strength follows the log-normal distribution as observed in experiment (51). In this case, although the distribution of log TE values ($\log_{10} T_{Y\to X}$) in Fig. 5*A* for d-conn and id-conn pairs overlap with each other, it can be well fitted by the summation of two log-normal distributions of TE values. Importantly, the fitted distributions of TE of d-conn and id-conn pairs agree well with those of the true connectivity setup in simulation as shown in Fig. 5*B*. Thus, we take the log-normal fitted distributions of causality measures as the ground truth of structural connectivity to evaluate the performance of network connectivity reconstruction, e.g., the AUC value is 0.997 in Fig. 5*B*. Additionally, we evaluate the performance of network reconstruction by using an experimentally measured log-normal distribution of synaptic weights (52). A slightly degraded reconstruction performance can be achieved with AUC value of 0.90, as depicted in *SI Appendix*, Fig. S7*E*. Practically, an optimal inference threshold determined as the intersection of the two fitted distribution curves (53) can be used for network connectivity reconstruction, which is indicated by the vertical solid line in Fig. 5*B*. In addition, the above results are robust for a variety of coupling strength distributions, including uniformly distributed and Gaussian distributed coupling strength with AUC values higher than 0.95 (*SI Appendix*, Fig. S7).

We then apply the same ROC analysis to all the four causality measures when analyzing experimental data under different visual stimuli conditions. As shown in the *Insets* of Fig. 5 *C*–*F*, the distribution of TDCC, TDMI, GC, and TE values are close to each other, which again verifies their mathematical relations given

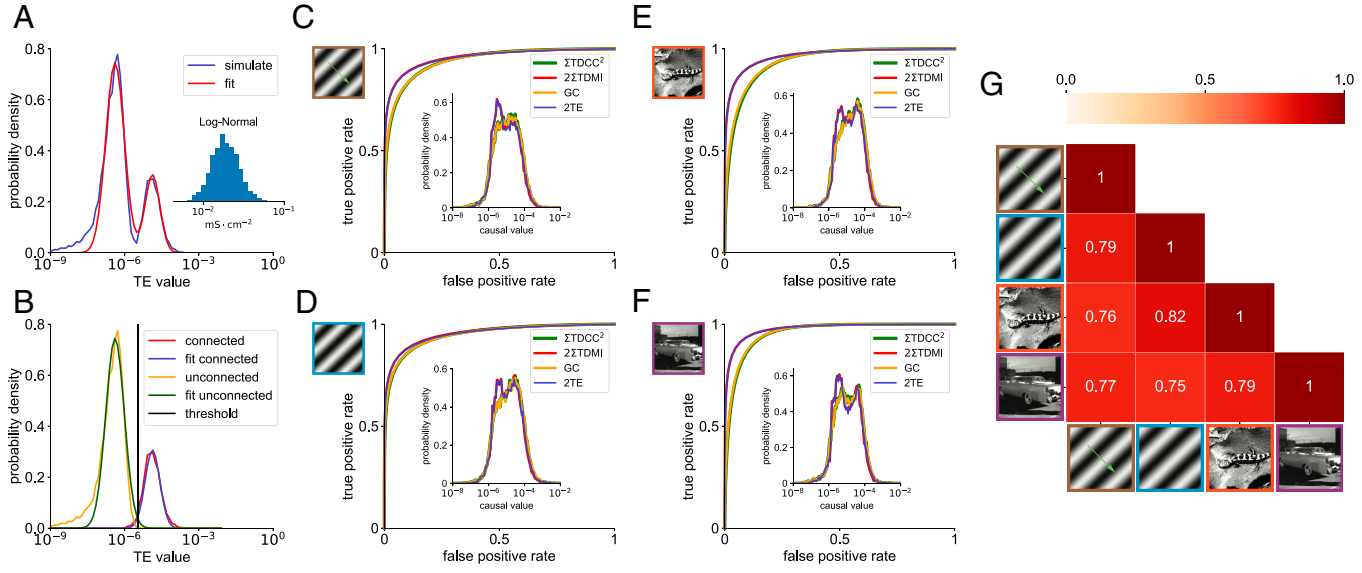

Fig. 5. Reconstruction of structural connectivity by the assumption of mixed log-normal distribution of causal values. (*A*) Distribution of TE values in an HH network of 100 excitatory neurons. The entry A_{ij} in the adjacency matrix follows a Bernoulli distribution with probability of 0.25 being 1. For the d-conn pairs, e.g., $A_{ij} = 1$, the corresponding coupling strength from neuron j to neuron i is sampled from a log-normal distribution. The blue and red curves are the simulated and fitted distributions, respectively. Inset: Histogram of structural coupling strength, S. (*B*) Distributions of TE values from d-conn and id-conn pairs of HH neurons in (*A*). The red and orange curves are the simulated TE values from d-conn and id-conn pairs respectively, while the blue and green curves are the fitted TE values from d-conn and id-conn pairs respectively which are obtained from the fitting in (*A*). Other parameters are the same as those in Fig. 4. (*C*–*F*) ROC curves of the network composed by the recorded neurons in experiment under visual stimuli of (*C*) drifting gratings, (*D*) static gratings, (*E*) natural scenes, and (*F*) natural movie, with AUC values equal to 0.950, 0.951, 0.953, 0.967. *Insets*: The distribution of causal values of each pair of neurons in the whole network. The parameters are set as $k = 1, l = 5, \Delta t = 1$ ms, and $m = 1$. The colors are the same as those in Fig. 3. (*G*) The similarity matrix of TE values over all neuron pairs across different stimuli conditions. For each pair of stimuli conditions, the similarity between the inferred TE values is measured by their correlation coefficient. The similarity matrix is symmetric with respect to its diagonal of ones, and the lower triangular part of the matrix is shown in the heatmap.

Downloaded from https://www.pnas.org by Kai Chen on March 31, 2024 from IP address 58.247.22.126.

by Theorems 1 to 4. Furthermore, to confirm the assumption of small $|\Delta p_m|$ for experimental data, we examined the distribution of $|\Delta p_m|$. As detailed in *SI Appendix* (*SI Appendix*, Fig. S9, *Top* panel), a substantial majority of the distribution adheres to the expansion assumption, specifically where $|\Delta p_m| < 1$. Under the assumption of mixed log-normal distribution (*SI Appendix*, Fig. S9, *Middle* panel), we first infer the ground truth of the network structural connectivity, and then evaluate the reconstruction performance of all four causality measures by the ROC curves as shown in Fig. 5 *C*–*F* with AUC values greater than 0.95. To investigate the consistency of reconstruction across different stimuli conditions, we compute the correlation coefficient of causal values for each pair of stimuli. As shown in Fig. 5*G*, our reconstruction achieves relatively high consistency across conditions with a correlation coefficient higher than 0.75. Note that almost all of the causal values of inconsistently inferred connections fall into the overlapping region of the distributions of d-conn and id-conn pairs (green curves in *SI Appendix*, Fig. S9, *Bottom* panel) which are generally error prone. We point out that such error prone generally exists as we can see similar phenomena for HH networks (*SI Appendix*, Fig. S10).

Discussion

In this work, inspired by the pulse-output nature of spiking neuronal networks, we have built an analytical framework upon a stochastic binary representation of neuronal interactions. This framework allows us, using only pairwise information between neurons, to i) establish the mathematical relationships between four widely used measures (TDCC, TDMI, GC, and TE) of causal connectivity, and ii) accurately reconstruct physical (structural) connectivity from causal connectivity in simulated HH neuronal networks. In addition, our framework provides the clarification and reduction of the major possible obstacles induced by confounders and unmeasured hidden nodes in conventional network reconstructions. Finally, we have used this analysis of pulse-output signals to reconstruct the structural connectivity of the real neuronal network in the mouse brain from experimentally measured spike-train data and have achieved promising performances.

We emphasize two key features of pulse-output signals in eliminating of the curse of dimensionality that lead to effective network reconstruction via our framework: i) the short timescale of autocorrelation, and ii) the weakness of indirect causalities. On the one hand, with small time step Δt commonly used in experiment, the short autocorrelation timescale protects the inferred causality from the corruption of the self-memory of time series. Therefore, the short autocorrelation length overcomes the curse of dimensionality in the estimation of probability density function by reducing the order of conditioned signal history (e.g., $k = l = 1$). On the other hand, because of the spike nature of the pulse-output signal, the causal values of d-conn pairs are several orders of magnitude larger than those of id-conn pairs, providing a clear distinction between the two. These two features make our framework a practical approach for pulse-coupled network reconstruction. In contrast, if these causality measures are directly applied to continuous-valued signals, e.g., voltage time series of a neuronal network, the mathematical relations derived in our theorems do not hold (*SI Appendix*, Fig. S11*A*) and the network reconstruction may also fail. For instance, TDCC and TDMI give incorrect reconstruction of the structural connectivity due to the strong self-correlation of continuous-valued time series (*SI Appendix*, Fig. S11*B*).

In the main text, we have illustrated the effectiveness of the four causality measures by taking the examples of an excitatory HH neural network receiving uncorrelated external Poisson drive (*Materials and Methods*). In fact, as discussed below, these methods apply much more broadly for pulse-output networks, including networks in synchronized states, networks receiving correlated inputs, networks with different neuronal models, networks comprising both excitatory and inhibitory neurons, networks in balanced states, and networks with output by surrogate calcium imaging data.

Synchronized States. Oscillations and synchronizations are commonly observed in the biological brain network, as shown in *SI Appendix*, Fig. S12*A*. Due to the fake causality between neurons introduced by the strong synchronous state, conventional reconstruction frameworks fail to capture the true structural connectivity. However, with our framework, high inference accuracy (AUC > 0.88) can still be achieved (*SI Appendix*, Fig. S12 *B* and *C*). Furthermore, by applying a desynchronized sampling method that only samples the pulse-output signals in asynchronous time intervals (*SI Appendix*, Fig. S12*D*), we can again perfectly reconstruct the network (AUC > 0.99, *SI Appendix*, Fig. S12 *E*- and *F*). The relation between percentages of desynchronized downsampling and AUC values is shown in *SI Appendix*, Fig. S13.

Correlated Inputs. External inputs to the network in the brain can often be correlated. In such a case, the synchronized states may be observed similarly as in previous cases (*SI Appendix*, Fig. S14*A*). Nevertheless, our framework can still achieve high inference accuracy (AUC > 0.99 with desynchronized downsampling methods, or AUC > 0.88 without downsampling, *SI Appendix*, Fig. S14 *B*–*F*) if the external inputs are moderately correlated, e.g., correlation coefficient less than 0.30 in our simulation case.

Different Neuronal Network Models. We apply our framework of reconstruction to other types of neuronal networks, including leaky integrate-and-fire (LIF) network (54), Izhikevich network (55), FitzHugh–Nagumo network (56), and Morris–Lecar network (57) (*Materials and Methods*). The results of all these networks can be seen in *SI Appendix*, Fig. S15. Our framework works well, with clear two-modal distributions of causal values (*SI Appendix*, Fig. S15 *B*, *E*, *H*, and *K*) and high reconstruction performance (AUC > 0.98, *SI Appendix*, Fig. S15 *C*, *F*, *I*, and *L*).

Networks Comprising Excitatory and Inhibitory Neurons. While the computational studies above have focused on networks of excitatory neurons, we now extend our framework to the case of HH networks consisting both excitatory and inhibitory neurons. In this context, we analyze a 100-neuron HH network configuration, comprised of 80 excitatory neurons and 20 inhibitory neurons. Recognizing the inherent tendency of inhibitory neurons to induce synchronization, we incorporate the downsampling preprocessing step prior to our methods application and achieve similarly accurate reconstructions (AUC > 0.99 with desynchronized downsampling methods, or AUC > 0.71 without downsampling, *SI Appendix*, Fig. S16).

Networks in Balanced States. Given that the inclusion of inhibitory neurons does not impact the performance of our method, we further extend our approach by applying it to the LIF network operating in the balanced dynamical regime (58).

Downloaded from https://www.pnas.org by Kai Chen on March 31, 2024 from IP address 58.247.22.126.

The balanced state is a widely recognized feature of cortical network dynamics, characterized by the cancelation of strong excitatory and inhibitory inputs for each individual neuron in the network. In the balanced state, neurons display fluctuation-driven behavior, and the entire network is in the asynchronous state. We simulate the balanced state of LIF neuronal networks using parameters from previous works (59) (*Materials and Methods*) with varying network sizes and sparsity levels. As illustrated in *SI Appendix*, Fig. S17*A*, our framework exhibits robust reconstruction performance for balanced networks with a fixed number of connections $K = 40$ for each neuron. As the network size N varies, AUC values always remain close to 1. Moreover, for a fixed network size $N = 40{,}000$, we vary the number of connections K from 20 to 4,000, changing the network connection density from 0.1 to 20%. Our findings indicate that for networks with connection density below 10%, our reconstruction framework performs well, as depicted in *SI Appendix*, Fig. S17*B*. Note that for relatively small K, the strong inhibition might lead to the case of $\Delta p_m \approx -1$ due to the synaptic coupling strength being $O(1/\sqrt{K})$. In such a case, another version of mathematical relations among the four causality measures has been developed and numerically verified in *SI Appendix*, Supporting Information Text 2 and in *SI Appendix*, Fig. S18*A*. For networks with fixed connection density, $p = K/N$, as the network size N increases, the proportion of neuronal pairs with $\Delta p_m \approx -1$ decreases (*SI Appendix*, Fig. S18 *C–F*). Therefore, the relations in Theorems 1–4 are valid for large E-I balanced networks with realistically large degree such as $K = 1{,}000$. In addition, for large sparse E-I balanced networks, the separation between distributions of d-conn pairs and id-conn pairs remains valid, guaranteeing good reconstruction performances (*SI Appendix*, Fig. S18*B*).

Surrogate Calcium Imaging Data. In addition to its application to spike-train data, our framework introduces the potential for reconstructing structural connectivity from calcium imaging data. Calcium imaging is widely used to visualize individual neuronal activities. However, compared to spike-train data, calcium signals have slower dynamics, which results in lower temporal resolution of the signals. This restricts the ability to capture precise spike timings for each neurons, thus, challenging the accuracy of our framework. To evaluate the validity of our framework in such a case, we first simulate calcium imaging data by coarsening the precise spike-train data of simulated HH networks, which mimics the spike train obtained by deconvoluting the calcium imaging data in experiments. As illustrated in *SI Appendix*, Fig. S19, given adequate data length, our framework can still successfully identify the bimodal distribution of causal values, achieving an AUC value being 0.97. Consequently, provided with sufficient recording length, our framework may be capable of reconstructing the intrinsic structural connectivity of neuronal networks with measured calcium imaging data.

In practice, a separation in the distribution of causal values between the contribution of d-conn and id-conn pairs calculated from data indicates the effectiveness of our method. To ensure good performance, our method requires i) a sufficient number of recorded firing events, ii) asynchrony (or only moderate synchrony), and iii) sparse connectivity. Somewhat more quantitatively, first, our method relies on the detection of accumulative information transfer in the pulse-output signals. Hence, it requires sufficient time length of recording signals. As shown in *SI Appendix*, Fig. S20*A*, by scanning various coupling strengths, our method is effective across the majority of coupling strengths with 1.2×10^5 spikes per neuron (corresponding to a duration of roughly a few hours (10^7 ms) of recording at a 12 Hz firing rate) in the pulse-output signals. We note that the effectiveness of our method relies on the magnitude disparity between d-conn and id-conn pairs. In application, it is better to choose the causality measure which depicts a clearer double-peak feature in the distribution of causal values. When considering two independent neurons with limited recording data length, the causal value lies at approximately $O(1/L)$ with L denoting the length of time series. If the causal values, being quadratically related to the coupling strength, are not considerably greater than $O(1/L)$, weak connections might remain undetected. Taking TDCC as an illustrative point, based on Eqs. **3** and **6** from *SI Appendix*, Supporting Information Text 1, and given $p_x \approx p_y = r\Delta t \ll 1$, it implies $C(X, Y; m) \approx r\Delta t \Delta p_m$. This further indicates that $C(X, Y; m)^2$ should surpass the baseline of independent pairs, i.e., $C(X, Y; m)^2 \approx r^2 \Delta t^2 \Delta p_m^2 > O(1/L) = \alpha/L$, where α is a system-dependent coefficient. Such criteria impose constraints on the spike-train data: the mean firing rate, r, the recording data length, L, and neuronal coupling strength (proportional to Δp_m), cannot be very small concurrently, i.e., $r^2 L \Delta p_m^2 \Delta t^2 > \alpha$. Therefore, the weaker the coupling strength, the longer the length of data that is required.

In addition, our approach works for neuronal networks in the asynchronous regime, but it can withstand a moderate amount of synchronization. Networks displaying extensive synchronization or correlation (60) have consistently posed challenges for structural connectivity reconstruction (47). We have previously highlighted that, with the aid of downsampling techniques, our method remains effective in quasisynchronized networks (*SI Appendix*, Figs. S12, S14, and S16). For HH dynamics with input correlation less than 30%, our method works well after downsampling, as shown in *SI Appendix*, Fig. S20*B*. Nevertheless, if the synchronization level rises further, the unsynchronized periods after downsampling will possess very few spikes, and, given the constraints discussed above, our method may be ineffective. In addition, high-density connectivity poses another challenge. We find that performance drops when connection density exceeds 30% in HH networks, and 40% after downsampling (*SI Appendix*, Fig. S20*C*). This can be attributed to the compounded causal effects via indirect interaction pathways (48–50). As the network connection density increases, the number of indirect pathways also increases and the contribution from indirect pathways toward causality will accumulate monotonically, as formulated by Eq. **40** in *SI Appendix*, Supporting Information Text 3. When these indirect pathways exert influence commensurate to (or even surpassing) direct connections, the reconstruction performance drops.

Materials and Methods

Reconstruction Performance Evaluation. For binary inference of structural connectivity, analysis based on receiver operating characteristic (ROC) curves is adopted to evaluate the reconstruction performance in this work. The following two scenarios are considered. (Unless otherwise specified, the length of simulated spike-train data used for reconstruction analysis in this work is 10^7 ms.)

With known structural connectivity. The conventional procedures for ROC curves analysis can be naturally applied toward data with true labels, i.e., structural connectivity in our binary reconstruction case. The area under the ROC curve (AUC) quantifies how well causality measures can distinguish d-conn pairs from id-conn pairs. If AUC is close to 1, the distribution of the causal values of those two kinds of pairs (d-conn and id-conn pairs) is well distinguishable, i.e., the performance of binary reconstruction is good. If the AUC is close to

Downloaded from https://www.pnas.org by Kai Chen on March 31, 2024 from IP address 58.247.22.126.

0.5, the distributions of causal values of those two types of pairs are virtually indistinguishable, meaning the performance of reconstruction is close to a random guess.

Without known structural connectivity. Conceptually, the causality measures for d-conn and id-conn pairs can be fitted by the log-normal distribution with different parameters, i.e., $\log_{10} T_{Y\to X}^{(\text{d-conn})} \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $\log_{10} T_{Y\to X}^{(\text{id-conn})} \sim \mathcal{N}(\mu_2, \sigma_2^2)$, respectively. So motivated, we fit the overall distribution with the summation of two log-normal distributions

$$\frac{p_c}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) + \frac{1-p_c}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right),$$

where p_c is the proportional coefficient between two Gaussian distributions $\log_{10} T_{Y\to X}^{(\text{d-conn})}$ and $\log_{10} T_{Y\to X}^{(\text{id-conn})}$. After that, the fitted distribution of two types of pairs is regarded as the true labels of structural connectivity. The similar procedures, as the previous case with known structural connectivity, are applied to obtain the ROC curve and the AUC value to evaluate the reconstruction performance.

HH Model. The dynamics of the ith neuron of an HH network is governed (61)

$$\begin{aligned} C\frac{dV_i}{dt} &= -G_L(V_i - V_L) + I_i^{\text{Na}} + I_i^{\text{K}} + I_i^{\text{input}} \\ I_i^{\text{Na}} &= -G_{\text{Na}} m_i^3 h_i (V_i - V_{\text{Na}}) \\ I_i^{\text{K}} &= -G_{\text{K}} n_i^4 (V_i - V_{\text{K}}), \\ \frac{dz_i}{dt} &= (1-z_i)\alpha_z(V_i) - z_i\beta_z(V_i), \quad \text{for } z = m, h, n, \end{aligned}$$

where C and V_i are the neuron's membrane capacitance and membrane potential (voltage), respectively; m_i, h_i, and n_i are gating variables; V_{Na}, V_{K}, and V_{L} are the reversal potentials for the sodium, potassium, and leak currents, respectively; G_{Na}, G_{K}, and G_{L} are the corresponding maximum conductances; and α_z and β_z are the rate variables. The detailed dynamics of the gating variables m, h, n and the choice of parameters can be found in ref. 53 and *SI Appendix*, Supporting Information Text 4. The input current $I_i^{\text{input}} = -G_i(t)(V_i - V_E)$ where $G_i(t)$ is the input conductance defined by $G_i(t) = f\sum_l H(t - s_{il}) + \sum_j A_{ij} S \sum_l H(t - \tau_{jl})$ and V_E is the reversal potential of excitation. Here, s_{il} is the lth spike time of the external Poisson input with strength f and rate ν, $\mathbf{A} = (A_{ij})$ is the adjacency matrix with $A_{ij} = 1$ indicating a direct connection from neuron j to neuron i and $A_{ij} = 0$ indicating no connection there. S is the coupling strength, and τ_{jl} is the lth spike time of the jth neuron. The spike-induced conductance change $H(t)$ is defined by $H(t) = \frac{\sigma_d\sigma_r}{\sigma_d - \sigma_r}\left[\exp\left(-\frac{t}{\sigma_d}\right) - \exp\left(-\frac{t}{\sigma_r}\right)\right]\Theta(t)$, where σ_d and σ_r are the decay and rise time scale, respectively, and $\Theta(\cdot)$ is the Heaviside function. When the voltage V_j reaches the firing threshold, V_{th}, the jth neuron generates a spike at this time, say τ_{jl}, and it will induce the ith neuron's conductance change if $A_{ij} = 1$.

LIF Model. The dynamics of the ith neuron in a leaky integrate-and-fire (LIF) network is governed (59, 62)

$$\begin{aligned} C\frac{dV_i}{dt} &= -G_L(V_i - V_L) + I_i^{\text{input}} \\ V_i(t) &= V_{\text{reset}} \quad \text{if} \quad V_i(t) \geq V_{\text{th}}, \end{aligned}$$

where C and V_i are the membrane capacitance and membrane potential. V_L and G_L are the reversal potential and conductance for leak currents. Compared with HH model, LIF model drops terms of nonlinear sodium and potassium current, and the input current is given by $I_i^{\text{input}} = f\sum_l \delta(t - s_{il}) + \sum_j A_{ij} S \sum_l \delta\left(t - \tau_{jl}\right)$, where $\delta(\cdot)$ is the Dirac delta function, s_{il} is the lth spike time of the external Poisson input with strength f, and rate ν, τ_{jl} is the lth spike time of jth neuron with strength S. And $\mathbf{A} = (A_{ij})$ is the adjacency matrix defined the same as that in the HH model. When the voltage reaches the threshold V_{th}, the ith neuron will emit a spike to all its connected postsynaptic neurons, and then reset to V_{reset} immediately. In numerical simulation, we use quantities: $C = 1$, $V_{\text{reset}} = -65$ mV, $V_{\text{th}} = -40$ mV, and the leakage conductance is set to be $G_L = 0.05\,\text{ms}^{-1}$ corresponding to the membrane time constant of 20 ms.

LIF Network Model in the Balanced Regime. The dynamics of the ith neuron with type Q in a LIF network is given below:

$$\begin{aligned} C\frac{dV_i^Q}{dt} &= -G_L(V_i^Q - V_L) + I_i^{QE} + I_i^{QI}, \\ V_i^Q(t) &= V_{\text{reset}} \quad \text{if} \quad V_i^Q(t) \geq V_{\text{th}}^Q, \end{aligned}$$

where $Q \in \{E, I\}$ represents the excitatory or the inhibitory neuron type. V_i^Q and V_{th}^Q are the membrane potential and spike threshold potential, respectively. And the excitatory input current to the ith neuron with type Q is given by $I_i^{QE}(t) = f^Q\sum_l \delta\left(t - s_{il}^Q\right) + S^{QE}\sum_{j=1}^{N^E} A_{ij}^{QE}\sum_l \delta\left(t - \tau_{jl}^E\right)$, where $\delta(\cdot)$ is the Dirac delta function, s_{il}^Q is the lth spike time of the external Poisson input for the ith neuron with type Q, f^Q is the input strength and ν^Q is the input rate. τ_{jl}^E is the lth spike time of the jth excitatory neuron. S^{QE} is the strength of the recurrent excitatory input to the neuron with type Q. $\mathbf{A}^{QE} = (A_{ij}^{QE})$ is the adjacency matrix from excitatory neurons to Q-type neurons. And the inhibitory input current is given by $I_i^{QI}(t) = -S^{QI}\sum_{j=1}^{N^I} A_{ij}^{QI}\sum_l \delta\left(t - \tau_{jl}^I\right)$, where τ_{jl}^I is the lth spike time of jth inhibitory neuron with strength S^{QI}, and $\mathbf{A}^{QI} = (A_{ij}^{QI})$ is the adjacency matrix from inhibitory neurons to Q-type neurons. In the numerical simulation of *E-I* balanced networks, we set parameters as $C = 1$, $V_{\text{reset}} = -65$ mV, $V_{\text{th}}^E = -40$ mV, $V_{\text{th}}^I = -54.5$ mV, $G_L = 0.05\,\text{ms}^{-1}$, and the coupling strengths are set to be $S^{EE} = S^{IE} = 15\,\text{mV}/\sqrt{K}$, $S^{EI} = 30\,\text{mV}/\sqrt{K}$, $S^{II} = 27\,\text{mV}/\sqrt{K}$, and parameters related to the Poisson drive are $f^E = 15\,\text{mV}/\sqrt{K}$, $f^I = 12\,\text{mV}/\sqrt{K}$, $\nu^E = \nu^I = K \cdot 50$ Hz, where K is the in-degree of a single neuron for both E and I types, i.e., each neuron receives K excitatory projections and K inhibitory connections from other neurons in the network.

Izhikevich Model. The dynamics of the ith neuron in an Izhikevich network is governed (55)

$$\begin{aligned} \tau\frac{dV_i}{dt} &= 0.04V_i^2 + 5V_i + 140 - u_i + I_i^{\text{input}} \\ \frac{du_i}{dt} &= a(bV_i - u_i), \end{aligned}$$

where V_i is the membrane potential and u_i is the recovery variable describing the force that drives V_i toward resting state. τ is the time constant of V_i. a and b describe the time scale and sensitivity (with respect to V_i) of u_i. I_i^{input} is the driving current containing the external Poisson input and the synaptic input from other neurons, defined by $I_i^{\text{input}} = f\sum_l \exp(t - s_{il})\Theta(t - s_{il}) + \sum_j A_{ij} S \sum_l \exp\left(t - \tau_{jl}\right)\Theta\left(t - \tau_{jl}\right)$, where $\Theta(\cdot)$ is the Heaviside function and parameters are defined similarly as those in the LIF model. When the voltage reaches the threshold V_{th}, the neuron emits a spike to all its postsynaptic neurons, and then reset V_i to c, and u_i to $u_i + d$. In numerical simulation, we use quantities: $\tau = 1$ ms, $a = 0.02\,\text{ms}^{-1}$, $b = 0.2\,\text{mV}^{-1}$, $c = -65$ mV, $d = 8$, $V_{\text{th}} = 30$ mV.

FitzHugh–Nagumo Model. The dynamics of the ith neuron in a FitzHugh–Nagumo network is governed (56)

$$\begin{aligned} \tau_V\frac{dV_i}{dt} &= V_i - \frac{V_i^3}{3} - W_i + I_i^{\text{input}} \\ \tau_W\frac{dW_i}{dt} &= V_i + a - bW_i, \end{aligned}$$

Downloaded from https://www.pnas.org by Kai Chen on March 31, 2024 from IP address 58.247.22.126.

where V_i and W_i describe the membrane potential and recovery variable, respectively. τ_V and τ_W are their corresponding time constants. I_i^{input} is the driving current containing the external Poisson input and the synaptic input from other neurons, defined by $I_i^{\text{input}} = f\sum_l \delta(t-s_{il}) + \sum_j A_{ij} S \sum_l \delta\left(t-\tau_{jl}\right)$, where parameters are defined similarly as those in the LIF model. When the voltage reaches the threshold V_{th}, the neuron emits a spike to all its postsynaptic neurons. In numerical simulation, we set quantities: $a = 0.7$, $b = 0.8$, $V_{\text{th}} = 0$, $\tau_V = 1$ ms, and $\tau_W = 12.5$ ms.

Morris–Lecar Model. The dynamics of the ith neuron in a Morris–Lecar network is governed by (57)

$$C\frac{dV_i}{dt} = -G_{\text{Ca}}M_i^{\infty}(V_i - V_{\text{Ca}}) - G_K W_i (V_i - V_K) - G_L (V_i - V_L) + I_i^{\text{input}}$$
$$\frac{dW_i}{dt} = \frac{1}{\tau_i^W}\left(W_i^{\infty} - W_i\right),$$

where C and V_i are the neuron's membrane capacitance and membrane potential, respectively; V_{Ca}, V_K, and V_L are the reversal potentials for the calcium, potassium, and leak currents, respectively; G_{Ca}, G_K and G_L are the corresponding maximum conductances; and W_i is the neuron's recovery variable (normalized K^+ conductance). M_i^{∞} and W_i^{∞} are the voltage-dependent equilibrium value of the normalized conductance of calcium and potassium, respectively, defined by

$$M_i^{\infty} = 0.5\left(1 + \tanh\left[(V_i - V_1)/V_2\right]\right)$$
$$W_i^{\infty} = 0.5\left(1 + \tanh\left[(V_i - V_3)/V_4\right]\right),$$

where V_1, V_2, V_3, and V_4 are the constant parameters. τ_i^W is a voltage-dependent time constant of W_i, defined by

$$\tau_i^W = \tau_0\left(\cosh\frac{V_i - V_3}{2V_4}\right)^{-1},$$

where τ_0 is a temperature-dependent parameter, fixed as a constant in simulation. I_i^{input} is the driving current containing the external Poisson input and the synaptic input from other neurons, defined by $I_i^{\text{input}} = f\sum_l \delta(t-s_{il}) + \sum_j A_{ij} S \sum_l \delta\left(t-\tau_{jl}\right)$, where parameters are defined similarly as those in LIF model. When the voltage reaches threshold V_{th}, the neuron emits a spike to all its postsynaptic neurons. In numerical simulation, we set parameters as $G_{\text{Ca}} = 4\ \text{mS}\cdot\text{cm}^{-2}$, $V_{\text{Ca}} = 120$ mV, $G_K = 8\ \text{mS}\cdot\text{cm}^{-2}$, $V_K = -80$ mV, $G_L = 2\ \text{mS}\cdot\text{cm}^{-2}$, $V_L = -60$ mV, $C = 20\ \text{mF}\cdot\text{cm}^{-2}$, $V_1 = -1.2$ mV, $V_2 = 18$ mV, $V_3 = 12$ mV, $V_4 = 17.4$ mV, $\tau_0 = 15$ ms, and $V_{\text{th}} = 0$ mV.

Neurophysiological Data. The public spike-train data are from Allen brain observatory (40, 41), accessed via the Allen Software Development Kit (AllenSDK) (63). Specifically, the data labeled with sessions-ID 715093703 were analyzed in this work. The 118-d-old male mouse passively received multiple visual stimuli from one of four categories, including drift gratings, static gratings, natural scenes, and natural movies. The single neuronal activities, i.e., spike trains, were recorded from multiple brain areas, including APN, CA1, CA3, DG, LGd, LP, PO, VISam, VISl, VISp, VISpm, and VISrl, using 6 Neuropixel probes. For each category of stimulus, the recording lasts for more than 20 min. 884 sorted spike trains were recorded and 156 of those were used for causality analysis with signal-to-noise ratio greater than 4 and mean firing rate greater than 0.08 Hz. The length of spike-train data used in analysis $\sim 10^6$ ms for each stimulus condition.

Data, Materials, and Software Availability. Implementation of causality estimation, network simulation, and processing scripts for Allen Brain Observatory data are all available through GitHub (https://github.com/neoneuron/causal4reconstruction). Previously published data were used for this work (40, 41).

ACKNOWLEDGMENTS. This work was supported by Science and Technology Innovation 2030 - Brain Science and Brain-Inspired Intelligence Project with Grant No. 2021ZD0200204 and the Lingang Laboratory Grant No. LG-QS-202202-01 (S.L., D.Z.); National Natural Science Foundation of China Grant 12271361, 12250710674 (S.L.); National Natural Science Foundation of China with Grant No. 12071287, 12225109 (D.Z.), Shanghai Municipal Science and Technology Major Project 2021SHZDZX0102, and the Student Innovation Center at Shanghai Jiao Tong University (Z.-q.K.T., K.C., S.L., and D.Z.).

Author affiliations: [a]School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai 200240, China; [b]Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai 200240, China; [c]Ministry of Education Key Laboratory of Scientific and Engineering Computing, Shanghai Jiao Tong University, Shanghai 200240, China; [d]Courant Institute of Mathematical Sciences, New York University, New York, NY 10012; [e]Center for Neural Science, New York University, New York, NY 10012; [f]Institute of Mathematical Sciences, New York University Shanghai, Shanghai 200122, China; [g]Neuroscience Institute of New York University Langone Health, New York University, New York, NY 10016; and [h]Shanghai Frontier Science Center of Modern Analysis, Shanghai Jiao Tong University, Shanghai 200240, China

1. B. Mišić, O. Sporns, From regions to connections and networks: New bridges between brain and behavior. *Curr. Opin. Neurobiol.* **40**, 1–7 (2016).
2. D. W. McLaughlin, R. Shapley, M. Shelley, D. J. Wielaard, A neuronal network model of macaque primary visual cortex (V1): Orientation selectivity and dynamics in the input layer 4C alpha. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8087–8092 (2000).
3. D. Zhou, A. V. Rangan, D. W. Mclaughlin, D. Cai, Spatiotemporal dynamics of neuronal population response in the primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 9517–9522 (2013).
4. C. Stringer *et al.*, Spontaneous behaviors drive multidimensional, brainwide activity. *Science* 80–364 (2019).
5. B. C. Wheeler, Y. Nam, In vitro microelectrode array technology and neural recordings. *Crit. Rev. Biomed. Eng.* **39**, 45–61 (2011).
6. C. Stosiek, O. Garaschuk, K. Holthoff, A. Konnerth, In vivo two-photon calcium imaging of neuronal networks. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 7319–7324 (2003).
7. A. K. Seth, Causal networks in simulated neural systems. *Cognit. Neurodyn.* **2**, 49–64 (2008).
8. C. J. Honey *et al.*, Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 2035–2040 (2009).
9. C. J. Honey, J. P. Thivierge, O. Sporns, Can structure predict function in the human brain? *Neuroimage* **52**, 766–776 (2010).
10. L. E. Suárez, R. D. Markello, R. F. Betzel, B. Misic, Linking structure and function in macroscale brain networks. *Trend. Cognit. Sci.* **24**, 302–315 (2020).
11. Y. Chen, B. Q. Rosen, T. J. Sejnowski, Dynamical differential covariance recovers directional network structure in multiscale neural systems. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2117234119 (2022).
12. G. Camps-Valls *et al.*, Discovering causal relations and equations from data. *Phys. Rep.* **1044**, 1–68 (2023).
13. J. Benesty, J. Chen, Y. Huang, I. Cohen, Pearson correlation coefficient in Springer Top. *Signal Process. (Springer)* **2**, 1–4 (2009).
14. M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein, Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14863–14868 (1998).
15. S. Ito *et al.*, Extending transfer entropy improves identification of effective connectivity in a spiking cortical network model. *PLoS One* **6**, e27431 (2011).
16. I. H. Stevenson, J. M. Rebesco, L. E. Miller, K. P. Körding, Inferring functional connections between neurons. *Curr. Opin. Neurobiol.* **18**, 582–588 (2008).
17. P. Bedenbaugh, G. L. Gerstein, Multiunit normalized cross correlation differs from the average single-unit normalized correlation. *Neural Comput.* **9**, 1265–1275 (1997).
18. J. A. Vastano, H. L. Swinney, Information transport in spatiotemporal systems. *Phys. Rev. Lett.* **60**, 1773–1776 (1988).
19. T. Schreiber, Measuring information transfer. *Phys. Rev. Lett.* **85**, 461–464 (2000).
20. S. Frenzel, B. Pompe, Partial mutual information for coupling analysis of multivariate time series. *Phys. Rev. Lett.* **99**, 204101 (2007).
21. P. W. McLaughlin *et al.*, The use of mutual information in registration of CT and MRI datasets post permanent implant. *Brachytherapy* **3**, 61–70 (2004).
22. C. W. J. Granger, Investigating causal relations by econometric models and cross-spectral methods. *Econ.?: J. Econ. Soc.* **37**, 424 (1969).
23. S. L. Bressler, A. K. Seth, Wiener-granger causality: A well established methodology. *Neuroimage* **58**, 323–329 (2011).
24. L. Barnett, A. B. Barrett, A. K. Seth, Solved problems for Granger causality in neuroscience: A response to Stokes and Purdon. *Neuroimage* **178**, 744–748 (2018).
25. T. Bossomaier, L. Barnett, M. Harré, J. T. Lizier, *An Introduction to Transfer Entropy: Information Flow in Complex Systems* (Springer, 2016).

Downloaded from https://www.pnas.org by Kai Chen on March 31, 2024 from IP address 58.247.22.126.

26. J. Borge-Holthoefer *et al*., The dynamics of information-driven coordination phenomena: A transfer entropy analysis. *Sci. Adv.* **2**, e1501158 (2016).
27. S. Li, Y. Xiao, D. Zhou, D. Cai, Causal inference in nonlinear systems: Granger causality versus time-delayed mutual information. *Phys. Rev. E: Stat. Phys. Plasmas Fluids Related Interdiscipl. Top.* **97**, 52216 (2018).
28. J. Runge, J. Heitzig, V. Petoukhov, J. Kurths, Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Phys. Rev. Lett.* **108**, 258701 (2012).
29. E. W. Newell, Y. Cheng, Mass cytometry: Blessed with the curse of dimensionality. *Nat. Immunol.* **17**, 890–895 (2016).
30. F. Bach, Breaking the curse of dimensionality with convex neural networks. *J. Mach. Learn. Res.* **18**, 1–53 (2017).
31. D. Marbach *et al*., Revealing strengths and weaknesses of methods for gene network inference. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 6286–6291 (2010).
32. R. De Smet, K. Marchal, Advantages and limitations of current network inference methods. *Nat. Rev. Microbiol.* **8**, 717–729 (2010).
33. C. Zou, K. J. Denby, J. Feng, Granger causality vs. dynamic Bayesian network inference: A comparative study. *BMC Bioinf.* **10**, 122 (2009).
34. L. Barnett, A. B. Barrett, A. K. Seth, Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.* **103**, 238701 (2009).
35. M. A. Koch, D. G. Norris, M. Hund-Georgiadis, An investigation of functional and anatomical connectivity using magnetic resonance imaging. *Neuroimage* **16**, 241–250 (2002).
36. A. Seth, Causal connectivity of evolved neural networks during behavior. *Netw. Comput. Neural Syst.* **16**, 35–54 (2005).
37. F. Schiele *et al*., A specific antidote for dabigatran: Functional and structural characterization. *Blood* **121**, 3554–3562 (2013).
38. D. Zhou, Y. Xiao, Y. Zhang, Z. Xu, D. Cai, Causal and structural connectivity of pulse-coupled nonlinear networks. *Phys. Rev. Lett.* **111**, 54102 (2013).
39. D. Zhou, Y. Xiao, Y. Zhang, Z. Xu, D. Cai, Granger causality network reconstruction of conductance-based integrate-and-fire neuronal systems. *PloS One* **9**, e87636 (2014).
40. Allen Institute MindScope Program, Allen Brain Observatory – Neuropixels Visual Coding [dataset] (2019). Available from brain-map.org/explore/circuits. Accessed 10 May 2021.
41. J. H. Siegle *et al*., Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* **592**, 86–92 (2021).
42. S. Guo, J. Wu, M. Ding, J. Feng, Uncovering interactions in the frequency domain. *Plos Comput. Biol.* **4**, e1000087 (2008).
43. B. Gourévitch, J. J. Eggermont, Evaluating information transfer between auditory cortical neurons. *J. Neurophysiol.* **97**, 2533–2543 (2007).
44. R. Vicente, M. Wibral, M. Lindner, G. Pipa, Transfer entropy-a model-free measure of effective connectivity for the neurosciences. *J. Comput. Neurosci.* **30**, 45–67 (2011).
45. T. Fawcett, An introduction to ROC analysis. *Pattern Recogn. Lett.* **27**, 861–874 (2006).
46. J. V. Carter, J. Pan, S. N. Rai, S. Galandiuk, ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. *Surg. (United States)* **159**, 1638–1645 (2016).
47. A. Das, I. R. Fiete, Systematic errors in connectivity inferred from activity in strongly recurrent networks. *Nat. Neurosci.* **23**, 1286–1296 (2020).
48. D. Q. Nykamp, Revealing pairwise coupling in linear-nonlinear networks. *SIAM J. Appl. Math.* **65**, 2005–2032 (2005).
49. D. Q. Nykamp, Exploiting history-dependent effects to infer network connectivity. *SIAM J. Appl. Math.* **68**, 354–391 (2007).
50. D. Q. Nykamp, A mathematical framework for inferring connectivity in probabilistic neuronal networks. *Math. Biosci.* **205**, 204–251 (2007).
51. G. Buzsáki, K. Mizuseki, The log-dynamic brain: How skewed distributions affect network operations. *Nat. Rev. Neurosci.* **15**, 264–278 (2014).
52. S. Song, P. J. Sjöström, M. Reigl, S. Nelson, D. B. Chklovskii, Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* **3**, e68 (2005).
53. P. Dayan, L. F. Abbott, *Theoretical Neuroscience* (MIT Press, Cambridge, MA, 2001), vol. 806.
54. L. F. Abbott, Lapicque's introduction of the integrate-and-fire model neuron (1907). *Brain Res. Bull.* **50**, 303–304 (1999).
55. E. M. Izhikevich, Simple model of spiking neurons. *IEEE Trans. Neural Netw.* **14**, 1569–1572 (2003).
56. R. FitzHugh, Impulses and physiological states in theoretical models of nerve membrane. *Biophys. J.* **1**, 445–466 (1961).
57. C. Morris, H. Lecar, Voltage oscillations in the barnacle giant muscle fiber. *Biophys. J.* **35**, 193–213 (1981).
58. C. Van Vreeswijk, H. Sompolinsky, Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* **274**, 1724–1726 (1996).
59. Q. Gu, Z. Tian, G. Kovacic, D. Cai, The dynamics of balanced spiking neuronal networks under Poisson drive is not chaotic. *Front. Comput. Neurosci.* **12**, 47 (2018).
60. I. Ginzburg, H. Sompolinsky, Theory of correlations in stochastic neural networks. *Phys. Rev. E* **50**, 3171–3191 (1994).
61. Y. Sun, D. Zhou, A. V. Rangan, D. Cai, Library-based numerical reduction of the Hodgkin-Huxley neuron for network simulation. *J. Comput. Neurosci.* **27**, 369–390 (2009).
62. K. A. Newhall, G. Kovačič, P. R. Kramer, D. Cai, Cascade-induced synchrony in stochastically driven neuronal networks. *Phys. Rev. E - Stat. Nonlinear Soft Matter Phys.* **82**, 41903 (2010).
63. Allen Institute for Brain Science, Allen Software Development Kit [software package] (2018). Available from https://allensdk.readthedocs.io/en/latest/. Accessed 10 May 2021.

Downloaded from https://www.pnas.org by Kai Chen on March 31, 2024 from IP address 58.247.22.126.