Letter

Fourier-domain transfer entropy spectrum

Yang Tian [1,*] Yaoyuan Wang,[2,†] Ziyang Zhang,[2,‡] and Pei Sun [1,§]

[1]Department of Psychology & Tsinghua Laboratory of Brain and Intelligence, Tsinghua University, Beijing 100084, China

[2]Data Center Technology Laboratory, Central Research Institute, 2012 Laboratories, Huawei Technologies Company Limited, Beijing 100084, China

(Received 12 October 2021; accepted 30 November 2021; published 15 December 2021)

We propose the Fourier-domain transfer entropy spectrum, a generalization of transfer entropy, as a model-free metric of causality. For arbitrary systems, this approach systematically quantifies the causality among their different system components rather than merely analyzing systems as entireties. The generated spectrum offers a rich-information representation of time-varying latent causal relations, efficiently dealing with nonstationary processes and high-dimensional conditions. We demonstrate its validity in the aspects of parameter dependence, statistical significance tests, and sensibility. An open-source multiplatform implementation of this metric is developed and computationally applied on neuroscience data sets and diffusively coupled logistic oscillators.

DOI: 10.1103/PhysRevResearch.3.L042040

Measuring causality between systems is critical in numerous scientific disciplines, ranging from physics [1] to neuroscience [2]. The challenge of such a measurement arises from a widespread situation where only a limited set of time series of systems are given, and the prior knowledge of their coupling relations remains absent. Over the last decades, identifying and quantifying causal relations with insufficient information has become one of the frontier problems in physics, mathematics, and statistics [3]. Substantial progress has been accomplished in measuring causality from statistics-theoretical (e.g., Granger causality [4] and its generalizations [5–10]) and information-theoretical (e.g., transfer entropy [11] and its generalizations [12–20]) perspectives. Although the nature of causality remains controversial [21], these two perspectives have been demonstrated to have nonparametric connections in mathematics [22,23] and share equivalent ideas in potential causality detection. Following their ideas, the existence of causality between systems X and Y (e.g., X is the cause of Y) can be interpreted as

$$\mathcal{P}(\mathsf{Y}(t), \mathsf{X}_{[t,\beta,\Delta t]} \mid \mathsf{Y}_{[t,\beta,\Delta t]}) \neq \mathcal{P}(\mathsf{Y}(t) \mid \mathsf{Y}_{[t,\beta,\Delta t]})\mathcal{P}(\mathsf{X}_{[t,\beta,\Delta t]} \mid \mathsf{Y}_{[t,\beta,\Delta t]}), \quad (1)$$

where $\mathcal{P}(\cdot)$ denotes the probability. We mark $\mathsf{X}_{[t,\beta,\Delta t]} = [\mathsf{X}(t-\beta\Delta t), \ldots, \mathsf{X}(t-\Delta t)]$ as the history of X with time lag unit Δt and maximum lag number β, representing the information of system X during a time interval $[t-\beta\Delta t, t-\Delta t]$. In general, inequality (1) means that the uncertainty of Y at moment t given its own historical information (term $\mathsf{Y}_{[t,\beta,\Delta t]}$) will be regulated if the historical information of X (term $\mathsf{X}_{[t,\beta,\Delta t]}$) is added.

Among various causality metrics, transfer entropy [11] features plentiful applications (e.g., in neuroscience [24–28] and economics [29–33]) because of its robust capacity to capture nonlinear causality and intrinsic connections with dynamics theory and information theory [3,11]. The initial version of transfer entropy can be defined as

$$\mathcal{T}(\mathsf{X},\mathsf{Y}) = \mathbb{E}_{t\in[\tau+1,n]}\{\mathcal{D}[\mathcal{P}(\mathsf{Y}(t), \mathsf{X}_{[t,\beta,\Delta t]} \mid \mathsf{Y}_{[t,\beta,\Delta t]}) \| \mathcal{P}(\mathsf{Y}(t) \mid \mathsf{Y}_{[t,\beta,\Delta t]}) \otimes \mathcal{P}(\mathsf{X}_{[t,\beta,\Delta t]} \mid \mathsf{Y}_{[t,\beta,\Delta t]})]\}, \quad (2)$$

actually equivalent to conditional mutual information [3,34]. Equation (2) quantifies the regulating effects in inequality (1) in terms of the expectation of the mutual information between $\mathsf{Y}(t)$ and $\mathsf{X}_{[t,\beta,\Delta t]}$ given $\mathsf{Y}_{[t,\beta,\Delta t]}$. While the model-free property of transfer entropy suggests its general applicability, its application in many real situations is still limited by various deficiencies, such as the dimensionality curse [15] and noise sensitivity [35]. Although extensive efforts have been devoted to optimizing transfer entropy estimation in real data (e.g., through symbolization [12], phase time series [13], ensemble evaluation [16], Lempel-Ziv complexity [18], artificial neural networks [19], and k nearest neighbor [36]), several perspectives remain to be improved:

(I) Previous works primarily analyze systems as entireties, yet it is more ubiquitous that causality only originates from specific system components (e.g., there only exists causality between the low-frequency components of X and Y). Such causality may be veiled if we do not distinguish between different system components.

(II) Causality is time varying. Although the initial version of transfer entropy can be calculated accumulatively or in sliding time windows, a more natural and accurate measurement remains absent.

*tiany20@mails.tsinghua.edu.cn

†Corresponding author: wangyaoyuan1@huawei.com

‡Corresponding author: zhangziyang11@huawei.com

§Corresponding author: peisun@tsinghua.edu.cn

(III) Probability density estimation, a prerequisite of transfer entropy calculation, may be costly (e.g., critically requires a large sample size) or inaccurate (e.g., in nonstationary systems and high-dimensional spaces). Although some approaches (e.g., ensemble evaluation in cyclostationary systems [16,37] and k-nearest-neighbor estimation [36] in high-dimensional spaces) are developed to optimize the estimation, a low-cost transfer entropy generalization born to deal with nonstationary systems is demanded.

In this Letter, we discuss the possibility of generalizing transfer entropy into a form that maintains its original advantages and makes progress in aspects of (I)–(III). To devote to open science, we release a multiplatform code implementation of our metric and demonstrate its validity on multiple data sets in neuroscience and physics.

Fourier-domain representation of systems. An arbitrary system X can be subdivided into a set of components according to different criteria. To maintain the model-free property of transfer entropy measurements, we concentrate on a universal criterion subdividing X into different frequency components, namely a Fourier-domain (frequency-domain) representation. Such a representation features widespread applications in physics [38], mathematics [39], neuroscience [40], and engineering [41], partly ensuring the general applicability of our approach.

To avoid the stationary assumption, one can derive the Fourier-domain representation $\mathcal{F}(\mathsf{X}, t, \omega)$ of X (here, ω denotes frequency) applying wavelet (WT) [42] or Hilbert-Huang transform (HHT) [43] rather than the Fourier transform (FT) owing to the limitations of FT for nonstationary processes [44]. In general, WT convolves X with oscillatory kernels of different frequency bands to realize time-frequency localization [42] while HHT directly measures the instantaneous frequency of X [43]. Although WT and HHT support time-frequency analysis irrespective of stationary assumption, they are still not ideal owing to certain limitations and should be selected according to circumstances [43,45]. Here, we choose WT to conduct our derivations owing to its complete mathematical foundations, yet other frameworks (e.g., HHT and its variant [46]) can also be applied.

In terms of a continuous WT spectrum, the Fourier-domain representation $\mathcal{F}(\mathsf{X}, t, \omega)$ can be obtained through [47–49]

$$\mathcal{F}(\mathsf{X}, t, \omega) = \mathsf{W}_\psi(\mathsf{X}, t, \zeta_\psi(\omega))\mathsf{W}^*_\psi(\mathsf{X}, t, \zeta_\psi(\omega)), \quad (3)$$

$$\mathsf{W}_\psi(\mathsf{X}, t, s) = \frac{1}{\sqrt{s}} \int \psi^*\Big(\frac{\tau - t}{s}\Big)\mathsf{X}(\tau)d\tau, \quad (4)$$

where $\psi \in L^2(\mathbb{R})$ denotes a mother wavelet function (e.g., Morlet wavelets) and notion $*$ represents a complex conjugate. Notions s and τ denote the scaling (frequency localization) and the translation (time localization) parameters, respectively. Function $\zeta_\psi(\cdot)$ is the bijective mapping from frequency ω to scale s, varying according to the selection of the mother wavelet function ψ. Because the approaches in Eqs. (3) and (4) have been extensively studied and applied [47–49], we do not repeat their derivations.

The Fourier-domain representation $\mathcal{F}(\mathsf{X}, t, \omega)$ can function as a time-frequency spectrum (t, ω) of system X, where the spectrum value denotes the power. In Fig. 1(a), we show the Fourier-domain representation of an open-source near-infrared spectroscopy (NIRS, a kind of electromagnetic-spectrum-based brain imaging approach) data set recorded in the superior frontal cortex [50,51]. Function ψ is set as the Morlet wavelet. Apart from $\mathcal{F}(\mathsf{X}, t, \omega)$ (subject 1) and $\mathcal{F}(\mathsf{Y}, t, \omega)$ (subject 2), we also present $\mathcal{F}(\mathsf{XY}, t, \omega) = \mathsf{W}_\psi(\mathsf{X}, t, \zeta_\psi(\omega))\mathsf{W}^*_\psi(\mathsf{Y}, t, \zeta_\psi(\omega))$.

Fourier-domain transfer entropy spectrum. Once $\mathcal{F}(\mathsf{X}, t, \omega)$ and $\mathcal{F}(\mathsf{Y}, t, \omega)$ have been given, we turn to quantify the time-varying latent causality between them in terms of transfer entropy. Below, we discuss one possibility of this measurement.

To overcome the deficiency of probability density estimation [see (III)], we transform these spectra by symbolization. Specifically, we generalize the approach in Ref. [52] to implement a two-dimensional symbolization $\widehat{\mathcal{F}}(\mathsf{X}, t, \omega)$ with a time symbolization order λ and a frequency symbolization order θ,

$$\widehat{\mathcal{F}}(\mathsf{X}, t, \omega) = [\Lambda_\mathsf{X}(t, \omega), \Theta_\mathsf{X}(t, \omega)],\ \forall t \geqslant \lambda,\ \omega \geqslant \theta, \quad (5)$$

$$\Lambda_\mathsf{X}(t, \omega) = \sum_{i=1}^{\lambda} [\Gamma_\Lambda(i) - 1]\lambda^{\lambda - i}, \quad (6)$$

$$\Theta_\mathsf{X}(t, \omega) = \sum_{i=1}^{\theta} [\Gamma_\Theta(i) - 1]\theta^{\theta - i}, \quad (7)$$

$$\Gamma_\Lambda = \gamma(\mathcal{F}(\mathsf{X}, t - \lambda, \omega), \ldots, \mathcal{F}(\mathsf{X}, t, \omega)), \quad (8)$$

$$\Gamma_\Theta = \gamma(\mathcal{F}(\mathsf{X}, t, \omega - \theta), \ldots, \mathcal{F}(\mathsf{X}, t, \omega)). \quad (9)$$

The above formulas map each $\mathcal{F}(\mathsf{X}, t, \omega)$ to a two-dimensional coordinate $[\Lambda_\mathsf{X}(t, \omega), \Theta_\mathsf{X}(t, \omega)]$. The first coordinate component $\Lambda_\mathsf{X}(t, \omega)$ corresponds to the symbolization along the time line while the second one corresponds to the symbolization along the frequency line. Both coordinate components are decimal numbers. The symbolization relays on a permutation $\gamma(\cdot)$ that returns the order of a sequence in ascending sort [e.g., $\gamma(1, 6, 3, 10) = (1, 3, 2, 4)$ because $1 < 3 < 6 < 10$]. If the order minus one [e.g., $(1, 3, 2, 4)$ becomes $(0, 2, 1, 3)$], then it becomes a number in λ base (or θ base) and can be transformed to a decimal number following Eqs. (6) and (7). In Fig. 1(b), we show the symbolization results of the NIRS data under two different conditions.

The symbolization benefits to our research by embedding $\mathcal{F}(\mathsf{X}, t, \omega)$, a space with a large cardinal number [there are a mass of possibilities of the spectrum value of $\mathcal{F}(\mathsf{X}, t, \omega)$ because the power in Eq. (3) is a continuous variable], to a space with a much smaller cardinal number [$\widehat{\mathcal{F}}(\mathsf{X}, t, \omega)$ is a discrete space whose cardinal number is no more than $\lambda!$][52]. It functions as a coarse-graining approach to control noises and supports a more adequate sampling of the historical information [e.g., $\mathsf{X}_{[t,\beta,\Delta t]}$ in Eq. (2)] during probability density estimation (because the variable becomes discrete and its variability is reduced) [17]. These benefits should be emphasized especially when X is nonstationary and a repetitive sampling is not infeasible. In sum, the symbolization in Eqs. (5)–(9) can, to a certain extent, alleviate deficiency (III). Parameters λ and θ control the granularity of coarse graining [e.g., see Fig. 1(b)]. They should not be too large. Otherwise, the symbolization is too fine grained and unnecessary [53].

After symbolization, the Fourier-domain transfer entropy spectrum can be defined as Eq. (10), in which we adopt the

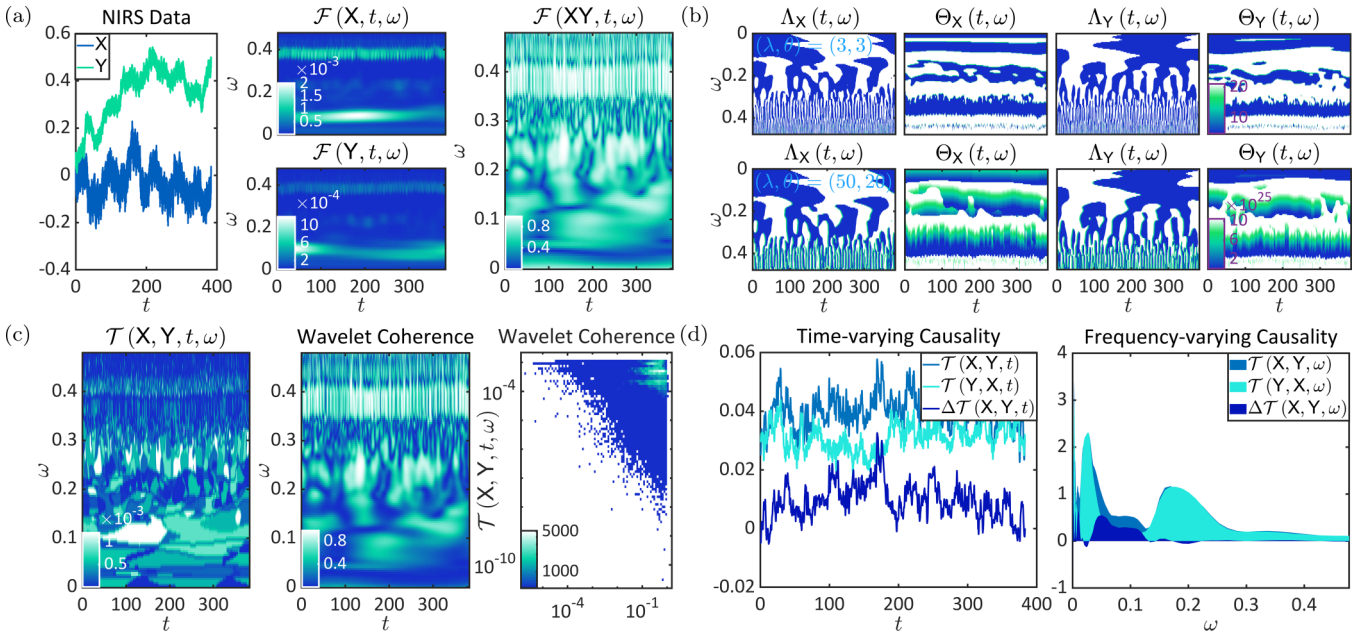

FIG. 1. Fourier-domain transfer entropy spectrum analysis. (a) The Fourier-domain representations $\mathcal{F}(\mathsf{X},t,\omega)$, $\mathcal{F}(\mathsf{Y},t,\omega)$, and $\mathcal{F}(\mathsf{XY},t,\omega)$ are generated from the NIRS data X and Y and presented as heat maps. Note that the frequency ω is given as cycles per sample (normalized frequency) in our Letter. (b) A coarse-grained symbolization with $(\lambda,\theta)=(3,3)$ (upper parallel) and a fine-grained symbolization with $(\lambda,\theta)=(50,20)$ (bottom parallel) are illustrated for comparison between different granularities. Compared with the coarse-grained symbolization, a fine-grained one maintains a higher expressiveness of the variability in spaces Λ and Θ. (c) The Fourier-domain transfer entropy spectrum (left) and the wavelet coherence (middle) are shown for comparison. Their distribution patterns shown by a planar plot (right) suggest that causality has no trivial accordant relation with the wavelet coherence. (d) Time-varying (left) and frequency-varying (right) transfer entropy as well as directional information transfer (denoted by Δ) are shown. The dominant information transfer direction, no matter whether it is analyzed from time-varying or frequency-varying perspective, is primarily suggested as $\mathsf{X}\to\mathsf{Y}$ rather than $\mathsf{Y}\to\mathsf{X}$

notions of inequality (1) to define an abbreviative representation $\widehat{\mathcal{F}}(\mathsf{X}_{[t,\beta,\Delta t]},\omega)=[\widehat{\mathcal{F}}(\mathsf{X},t-\beta\Delta t,\omega),\ldots,\widehat{\mathcal{F}}(\mathsf{X},t,\omega)]$. Although Eq. (10) is consistent with Eq. (2), it does not average or sum transfer entropy across time and frequency for a single value. Instead, the transfer entropy in Eq. (10) becomes a function of time and frequency, partly overcoming deficiencies (I) and (II). Note that $\mathcal{T}(\mathsf{X},\mathsf{Y},t,\omega)$ in Eq. (10) can be either positive or negative, respectively corresponding to the case where the added information of $\widehat{\mathcal{F}}(\mathsf{X}_{[t,\beta,\Delta t]},\omega)$ reduces or increases the uncertainty of $\widehat{\mathcal{F}}(\mathsf{Y},t,\omega)$ given $\widehat{\mathcal{F}}(\mathsf{Y}_{[t,\beta,\Delta t]},\omega)$ [there is no limitation of the directionality of regulating the effects in inequality (1)]. To keep consistency with the Granger causality that identifies potential causes through uncertainty (error) reduction [4], we refer to $\widehat{\mathcal{F}}(\mathsf{X}_{[t,\beta,\Delta t]},\omega)$ as a cause of $\widehat{\mathcal{F}}(\mathsf{Y},t,\omega)$ when $\widehat{\mathcal{F}}(\mathsf{X}_{[t,\beta,\Delta t]},\omega)$ is positive. A negative $\widehat{\mathcal{F}}(\mathsf{X}_{[t,\beta,\Delta t]},\omega)$ corresponds to a decoupling effect on $\widehat{\mathcal{F}}(\mathsf{Y},t,\omega)$ and $\widehat{\mathcal{F}}(\mathsf{X}_{[t,\beta,\Delta t]},\omega)$ provided by $\widehat{\mathcal{F}}(\mathsf{X}_{[t,\beta,\Delta t]},\omega)$:

$$\mathcal{T}(\mathsf{X},\mathsf{Y},t,\omega)=\mathcal{P}(\widehat{\mathcal{F}}(\mathsf{Y},t,\omega),\widehat{\mathcal{F}}(\mathsf{X}_{[t,\beta,\Delta t]},\omega),\widehat{\mathcal{F}}(\mathsf{Y}_{[t,\beta,\Delta t]},\omega))\times\log\frac{\mathcal{P}(\widehat{\mathcal{F}}(\mathsf{Y},t,\omega),\widehat{\mathcal{F}}(\mathsf{X}_{[t,\beta,\Delta t]},\omega)\mid\widehat{\mathcal{F}}(\mathsf{Y}_{[t,\beta,\Delta t]},\omega))}{\mathcal{P}(\widehat{\mathcal{F}}(\mathsf{Y},t,\omega)\mid\widehat{\mathcal{F}}(\mathsf{Y}_{[t,\beta,\Delta t]},\omega))\mathcal{P}(\widehat{\mathcal{F}}(\mathsf{X}_{[t,\beta,\Delta t]},\omega)\mid\widehat{\mathcal{F}}(\mathsf{Y}_{[t,\beta,\Delta t]},\omega))}. \tag{10}$$

In Fig. 1(c), we show the Fourier-domain transfer entropy spectrum of the NIRS data. The symbolization is implemented under a relatively coarse-grained condition $(\lambda,\theta)=(3,3)$. Each $\mathcal{T}(\mathsf{X},\mathsf{Y},t,\omega)$ is measured with $\beta=5$. For convenience, we set each negative $\mathcal{T}(\mathsf{X},\mathsf{Y},t,\omega)$ as 0 (this setting can be excluded when the decoupling effects are of interest). In our results, we can see relatively strong and oscillating causality between X and Y around $\omega\sim0.1$, $\omega\sim0.25$, and $\omega\sim0.4$. The causality in other frequency regions is relatively weak. We also compare $\mathcal{T}(\mathsf{X},\mathsf{Y},t,\omega)$ with the well-known wavelet coherence spectrum [54], which quantifies the time-varying and linear coherence between X and Y. Consistent with common knowledge, the causality measured by $\mathcal{T}(\mathsf{X},\mathsf{Y},t,\omega)$ cannot be trivially simplified as linear co-

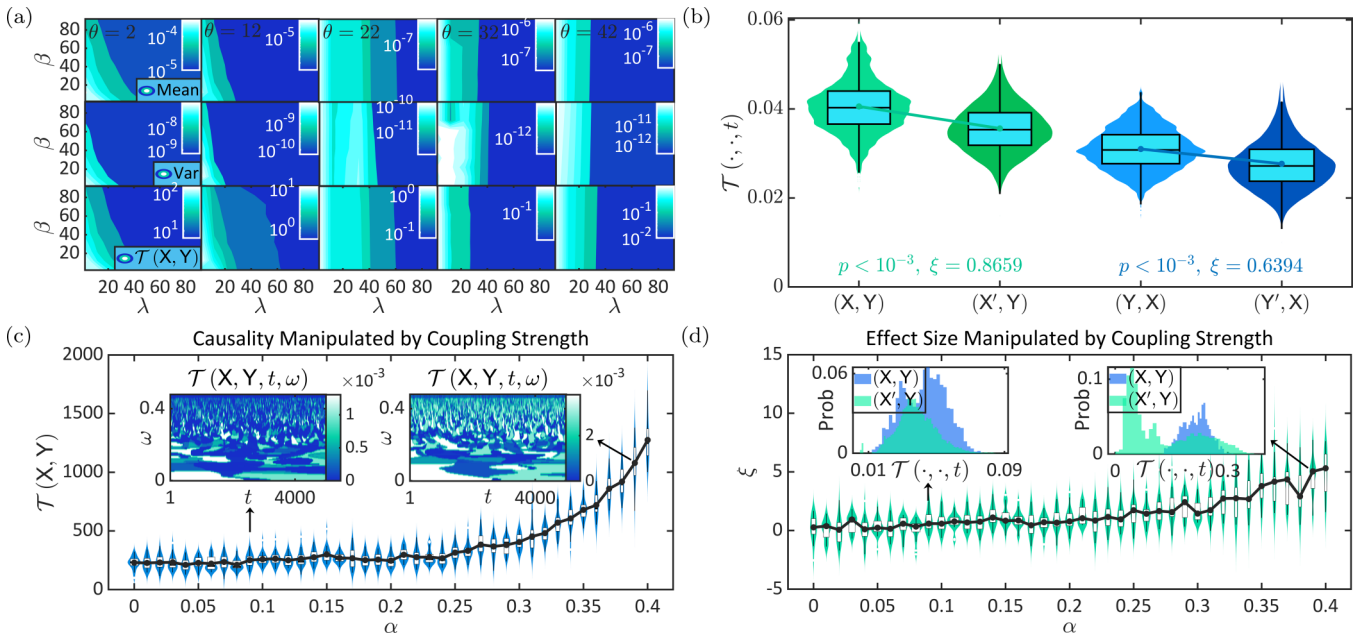

FIG. 2. Validity analysis of the Fourier-domain transfer entropy spectrum. (a) The dependence of parameter selections is analyzed in terms of expectation $\mathbb{E}_{t,\omega}\mathcal{T}(\mathsf{X},\mathsf{Y},t,\omega)$ (upper parallel), variance $\mathrm{Var}_{t,\omega}\mathcal{T}(\mathsf{X},\mathsf{Y},t,\omega)$ (middle parallel), and $\mathcal{T}(\mathsf{X},\mathsf{Y})$ (bottom parallel). While these quantities are sensitive to the increases of λ and θ, they are relatively robust to the variation of β. (b) Statistical significance analysis of the Fourier-domain transfer entropy spectrum in Figs. 1(c) and 1(d) is illustrated, where surrogates X' and Y' are randomly generated as references. Time-varying causality quantities $\mathcal{T}(\mathsf{X},\mathsf{Y},t)$ and $\mathcal{T}(\mathsf{Y},\mathsf{X},t)$ are both statistically significant. Consistent with our finding that the dominant information transfer direction in the NIRS data is $\mathsf{X}\to\mathsf{Y}$ [see Fig. 1(d)], $\mathcal{T}(\mathsf{X},\mathsf{Y},t)$ has a larger effect size ξ than $\mathcal{T}(\mathsf{Y},\mathsf{X},t)$. (c), (d) The Fourier-domain transfer entropy spectrum and its effect size is manipulated by the coupling strength α. As α increases (logistic oscillators X and Y become more tightly coupled), the causality measured by $\mathcal{T}(\mathsf{X},\mathsf{Y})$ becomes stronger and more statistically significant (with a larger effect size ξ).

herence because these two metrics have no clear accordant pattern:

$$\mathcal{T}(\mathsf{X},\mathsf{Y},t)=\sum_{\omega}\mathcal{T}(\mathsf{X},\mathsf{Y},t,\omega),\ \forall t, \tag{11}$$

$$\mathcal{T}(\mathsf{X},\mathsf{Y},\omega)=\sum_{t}\mathcal{T}(\mathsf{X},\mathsf{Y},t,\omega),\ \forall \omega, \tag{12}$$

$$\mathcal{T}(\mathsf{X},\mathsf{Y})=\sum_{t,\omega}\mathcal{T}(\mathsf{X},\mathsf{Y},t,\omega). \tag{13}$$

Meanwhile, we can sum $\mathcal{T}(\mathsf{X},\mathsf{Y},t,\omega)$ over frequency at each moment [see Eq. (11)], over time at each frequency [see Eq. (12)], or over both time and frequency [see Eq. (13)].

Equations (11)–(13) denote the varying Fourier-domain transfer entropy across time, frequency, and all possibilities, respectively. Note that $\mathcal{T}(\mathsf{X},\mathsf{Y})$ is always non-negative. Combining Eqs. (11)–(13) with the asymmetry property of transfer entropy, we define

$$\Delta\mathcal{T}(\mathsf{X},\mathsf{Y},t)=\mathcal{T}(\mathsf{X},\mathsf{Y},t)-\mathcal{T}(\mathsf{Y},\mathsf{X},t),\ \forall t, \tag{14}$$

$$\Delta\mathcal{T}(\mathsf{X},\mathsf{Y},\omega)=\mathcal{T}(\mathsf{X},\mathsf{Y},\omega)-\mathcal{T}(\mathsf{Y},\mathsf{X},\omega),\ \forall \omega, \tag{15}$$

$$\Delta\mathcal{T}(\mathsf{X},\mathsf{Y})=\mathcal{T}(\mathsf{X},\mathsf{Y})-\mathcal{T}(\mathsf{Y},\mathsf{X}), \tag{16}$$

to quantify the directional information transfer between X and Y. In Fig. 1(d), we illustrate these quantities based on the results in Fig. 1(c). A stronger information transfer from X to Y can be observed in both time line and frequency line, surpassing the transfer from Y to X. Such a directional transfer is significant when $\omega\sim 0.1$.

Validity of Fourier-domain transfer entropy spectrum. After proposing the Fourier-domain transfer entropy spectrum, we turn to analyze its validity.

A valid causality metric is demanded to have clear dependence relations with all involved parameters. Although the Fourier-domain transfer entropy spectrum is a model-free metric and has no preassumption on system properties, the symbolization and the transfer entropy definition itself still introduce three parameters: λ, θ, and β. Serving as coarse-graining parameters, λ and θ determine the granularity of symbolization. Sufficiently large quantities of λ and θ may not only make symbolization meaningless [53] but also cause obstacles in probability density estimation (because the variability of variables increase). An inexhaustive sampling frequently underestimates specific parts of the probability density, especially in high-dimensional spaces. Therefore, we speculate that $\mathcal{T}(\mathsf{X},\mathsf{Y},t,\omega)$ may be underestimated when λ and θ become too large. In Fig. 2(a), we calculate $\mathcal{T}(\mathsf{X},\mathsf{Y},t,\omega)$ in the NIRS data with different parameters. For convenience, we set each negative $\mathcal{T}(\mathsf{X},\mathsf{Y},t,\omega)$ as 0. Consistent with our inference, quantities $\mathbb{E}_{t,\omega}\mathcal{T}(\mathsf{X},\mathsf{Y},t,\omega)$ (expectation), $\mathrm{Var}_{t,\omega}\mathcal{T}(\mathsf{X},\mathsf{Y},t,\omega)$ (variance), and $\mathcal{T}(\mathsf{X},\mathsf{Y})$ approximate to 0 as λ and θ increase. Similar effects can be observed on the lag parameter β but become much more slight. In sum, the Fourier-domain transfer entropy spectrum is expected to be implemented with relatively small λ and θ. The selection of β is

less limited and can be carried out according to different demands.

Moreover, a valid causality metric should be equipped with a statistical significance test. Information-theoretical metrics, including our Fourier-domain transfer entropy spectrum, frequently involve biases in finite data sets [55]. Therefore, the absolute value of $\mathcal{T}(\mathsf{X},\mathsf{Y},t,\omega)$ only has a limited meaning until being statistically tested. Combining the time-shifting surrogates [56–58] with the permutation test [59,60], we design the significance test as follows: (1) Generate a set of surrogates $\{\mathsf{X}'_\delta|\mathsf{X}'_\delta(t)=\mathsf{X}(t+\delta),\ \delta\in\mathbb{N}^+\}$ and calculate $\{\mathcal{T}(\mathsf{X}'_\delta,\mathsf{Y},t)\}_t$ for each surrogate X'_δ. (2) Implement a permutation test for the difference in mean value between $\{\mathcal{T}(\mathsf{X},\mathsf{Y},t)\}_t$ and $\bigcup_\delta\{\mathcal{T}(\mathsf{X}'_\delta,\mathsf{Y},t)\}_t$ to calculate the statistical significance p and the effect size ξ. A similar idea can be seen in Ref. [61]. In practice, a weakest condition with $p\leqslant 10^{-2}$ and $\xi\geqslant 0.2$ (in a comparison of mean values, $\xi\sim 0.2$, $\xi\sim 0.5$, and $\xi\sim 0.8$ correspond to small, medium, and high effect sizes, respectively) should be satisfied if the Fourier-domain transfer entropy spectrum is statistically significant. In Fig. 2(b), we test the results in Figs. 1(c) and 1(d), where 50 surrogates are generated by randomizing $\delta\in[3,100]$. The generated $\bigcup_\delta\{\mathcal{T}(\mathsf{X}'_\delta,\mathsf{Y},t)\}_t$ features a large sample size of 191 350, supporting a robust statistical test. We implement a large-scale permutation test with 5000 permutations, demonstrating the statistic significance of $\mathcal{T}(\mathsf{X},\mathsf{Y},t)$ and $\mathcal{T}(\mathsf{Y},\mathsf{X},t)$ with ideal effect sizes. Meanwhile, the effect size on $\mathcal{T}(\mathsf{Y},\mathsf{X},t)$ is smaller than that on $\mathcal{T}(\mathsf{X},\mathsf{Y},t)$, corroborating our observations in Fig. 1(d) about the directional information transfer in the NIRS data.

Finally, a valid causality metric is expected to be sensitive to the changes of coupling strength between X and Y. In other words, $\mathcal{T}(\mathsf{X},\mathsf{Y},t,\omega)$ should be regulated as consequences if we manipulate the coupling strength. To realize the manipulation, we consider two diffusively coupled logistic oscillators $\mathsf{X}\to\mathsf{Y}$ in Figs. 2(c) and 2(d). Notion $\to$ denotes the direction of coupling. Oscillators X and Y are defined with a bifurcation parameter $\mu=4$, and their coupling strength is denoted by α. For more mathematical details, we refer to Ref. [62]. We manipulate the coupling strength α from 0 to 0.4 and apply an open-source algorithm [63] to generate the time series of X and Y under each α condition. The generation is repeated 50 times in each case, corresponding to different random initial values. For each pair of time series, we calculate the Fourier-domain transfer entropy spectrum with $\lambda=\theta=3$ and $\beta=1$. Meanwhile, five surrogates are generated by randomizing $\delta\in[3,100]$, leading to a large sample size 24 985 in the statistical significance test. The number of permutations is set as 5000. In Figs. 2(c) and 2(d), we demonstrate that the Fourier-domain transfer entropy spectrum and its effect sizes (with $p<10^{-2}$) increase with the coupling strength α. In other words, the quantified latent causality increases as a consequence of coupling strength enhancement.

Conclusion. We present the Fourier-domain transfer entropy spectrum, a model-free metric of causality, to extract the time-varying latent causality among different system components of an arbitrary pair of systems in the Fourier domain. Built on the wavelet transform and symbolization, our approach can partly alleviate deficiencies (I)–(III), generally applicable in various nonstationary or high-dimensional processes (e.g., neuroscience data). We have demonstrated the validity of our approach in the aspects of parameter dependence, statistical significance tests, and sensibility. An open-source release of our algorithm can be seen in Ref. [64].

Acknowledgments. This project is supported by the Artificial and General Intelligence Research Program of Guo Qiang Research Institute at Tsinghua University (2020GQG1017) as well as the Tsinghua University Initiative Scientific Research Program.

[1] A. Pikovsky, J. Kurths, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences*, Cambridge Nonlinear Science Series No. 12 (Cambridge University Press, Cambridge, U.K., 2003).

[2] E. Pereda, R. Q. Quiroga, and J. Bhattacharya, Nonlinear multivariate analysis of neurophysiological signals, Prog. Neurobiol. **77**, 1 (2005).

[3] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, Causality detection based on information-theoretic approaches in time series analysis, Phys. Rep. **441**, 1 (2007).

[4] C. W. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica: J. Eco. Soc. **37**, 424 (1969).

[5] D. Marinazzo, M. Pellicoro, and S. Stramaglia, Kernel Method for Nonlinear Granger Causality, Phys. Rev. Lett. **100**, 144103 (2008).

[6] N. Ancona, D. Marinazzo, and S. Stramaglia, Radial basis function approach to nonlinear Granger causality of time series, Phys. Rev. E **70**, 056221 (2004).

[7] M.-C. Ho, Y.-C. Hung, and I.-M. Jiang, Phase synchronization in inhomogeneous globally coupled map lattices, Phys. Lett. A **324**, 450 (2004).

[8] A. Arnold, Y. Liu, and N. Abe, Temporal causal modeling with graphical Granger methods, in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, New York, 2007), pp. 66–75.

[9] H. Liu, J. Lafferty, and L. Wasserman, The nonparanormal: Semiparametric estimation of high dimensional undirected graphs, J. Mach. Learn. Res. **10**, 2295 (2009).

[10] M. Dhamala, G. Rangarajan, and M. Ding, Estimating Granger Causality from Fourier and Wavelet Transforms of Time Series Data, Phys. Rev. Lett. **100**, 018701 (2008).

[11] T. Schreiber, Measuring Information Transfer, Phys. Rev. Lett. **85**, 461 (2000).

[12] M. Staniek and K. Lehnertz, Symbolic Transfer Entropy, Phys. Rev. Lett. **100**, 158101 (2008).

[13] M. Lobier, F. Siebenhühner, S. Palva, and J. M. Palva, Phase transfer entropy: A novel phase-based measure for directed connectivity in networks coupled by oscillatory interactions, Neuroimage **85**, 853 (2014).

[14] M. Lungarella, A. Pitti, and Y. Kuniyoshi, Information transfer at multiple scales, Phys. Rev. E **76**, 056117 (2007).

[15] J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths, Escaping the Curse of Dimensionality in Estimating Multivariate Transfer Entropy, Phys. Rev. Lett. **108**, 258701 (2012).

[16] P. Wollstadt, M. Martínez-Zarzuela, R. Vicente, F. J. Díaz-Pernas, and M. Wibral, Efficient transfer entropy analysis of non-stationary neural time series, PLoS One **9**, e102833 (2014).
[17] M. Porfiri and M. Ruiz Marín, Transfer entropy on symbolic recurrences, Chaos **29**, 063123 (2019).
[18] J. F. Restrepo, D. M. Mateos, and G. Schlotthauer, Transfer entropy rate through Lempel-Ziv complexity, Phys. Rev. E **101**, 052117 (2020).
[19] J. Zhang, O. Simeone, Z. Cvetkovic, E. Abela, and M. Richardson, ITENE: Intrinsic transfer entropy neural estimator, arXiv:1912.07277.
[20] R. Silini and C. Masoller, Fast and effective pseudo transfer entropy for bivariate data-driven causal inference, Sci. Rep. **11**, 8423 (2021).
[21] J. Pearl, Causality: Models, Reasoning and Inference, (Cambridge University Press, Cambridge, U.K., 2000).
[22] C. Diks and V. Panchenko, A note on the Hiemstra-Jones test for Granger non-causality, Stud. Nonlinear Dyn. Econom. **9**(2) (2005).
[23] C. Diks and J. DeGoede, A general nonparametric bootstrap test for Granger causality, in *Global Analysis of Dynamical Systems*, edited by H. W. Broer, B. Krauskopf, and G. Vegter (Institute of Physics Publishing (IOP), London, 2001), pp. 393–405.
[24] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, Transfer entropy—A model-free measure of effective connectivity for the neurosciences, J. Comput. Neurosci. **30**, 45 (2011).
[25] J. D. Victor, Binless strategies for estimation of information from neural data, Phys. Rev. E **66**, 051903 (2002).
[26] V. A. Vakorin, N. Kovacevic, and A. R. McIntosh, Exploring transient transfer entropy based on a group-wise ICA decomposition of EEG data, Neuroimage **49**, 1593 (2010).
[27] R. E. Spinney, M. Prokopenko, and J. T. Lizier, Transfer entropy in continuous time, with applications to jump and neural spiking processes, Phys. Rev. E **95**, 032319 (2017).
[28] M. Ursino, G. Ricci, and E. Magosso, Transfer entropy as a measure of brain connectivity: A critical analysis with the help of neural mass models, Front. Comput. Neurosci. **14**, 45 (2020).
[29] T. Dimpfl and F. J. Peter, Using transfer entropy to measure information flows between financial markets, Stud. Nonlinear Dyn. Econom. **17**, 85 (2013).
[30] A. Papana, C. Kyrtsou, D. Kugiumtzis, and C. Diks, Detecting causality in non-stationary time series using partial symbolic transfer entropy: Evidence in financial data, Comput. Econ. **47**, 341 (2016).
[31] F. Toriumi and K. Komura, Investment index construction from information propagation based on transfer entropy, Comput. Econ. **51**, 159 (2018).
[32] M. Camacho, A. Romeu, and M. Ruiz-Marin, Symbolic transfer entropy test for causality in longitudinal data, Econ. Model. **94**, 649 (2021).
[33] Q. Ji, H. Marfatia, and R. Gupta, Information spillover across international real estate investment trusts: Evidence from an entropy-based network analysis, North Am. J. Econ. Finance **46**, 103 (2018).
[34] T. M. Cover, *Elements of Information Theory* (Wiley, Hoboken, NJ, 1999).
[35] D. A. Smirnov, Spurious causalities with transfer entropy, Phys. Rev. E **87**, 042917 (2013).
[36] A. Kraskov, H. Stögbauer, and P. Grassberger, Estimating mutual information, Phys. Rev. E **69**, 066138 (2004).
[37] W. A. Gardner, A. Napolitano, and L. Paura, Cyclostationarity: Half a century of research, Signal Process. **86**, 639 (2006).
[38] R. J. Marks II, *Handbook of Fourier Analysis & its Applications* (Oxford University Press, Oxford, U.K., 2009).
[39] P. Brémaud, *Mathematical Principles of Signal Processing: Fourier and Wavelet Analysis* (Springer, Berlin, 2013).
[40] A. Bruns, Fourier-, Hilbert-and wavelet-based signal analysis: Are they really different approaches? J. Neurosci. Methods **137**, 321 (2004).
[41] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1975).
[42] O. Rioul and M. Vetterli, Wavelets and signal processing, IEEE Signal Process. Mag. **8**, 14 (1991).
[43] N. E. Huang, *Hilbert-Huang Transform and its Applications*, Vol. 16 (World Scientific, Singapore, 2014).
[44] G. Kaiser and L. H. Hudgins, *A Friendly Guide to Wavelets*, Vol. 300 (Springer, Berlin, 1994).
[45] G. B. Folland and A. Sitaram, The uncertainty principle: A mathematical survey, J. Fourier Anal. Appl. **3**, 207 (1997).
[46] Z. Peng, W. T. Peter, and F. Chu, An improved Hilbert–Huang transform and its application in vibration signal analysis, J. Sound Vib. **286**, 187 (2005).
[47] R. T. Ogden, *Essential Wavelets for Statistical Applications and Data Analysis* (Springer, Berlin, 1997).
[48] C.-L. Liu, A Tutorial of the Wavelet Transform (NTUEE, Taiwan, 2010).
[49] R. R. Coifman, Y. Meyer, and V. Wickerhauser, Wavelet analysis and signal processing, in *Wavelets and their Applications* (Jones and Barlett, Boston, 1992).
[50] X. Cui, D. M. Bryant, and A. L. Reiss, NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation, Neuroimage **59**, 2430 (2012).
[51] X. Cui, D. M. Bryant, and A. L. Reiss, The NIRS data in a built-in directory of Matlab (2012), https://ww2.mathworks.cn/help/wavelet/ug/wavelet-coherence-of-brain-dynamics.html.
[52] C. Bandt and B. Pompe, Permutation Entropy: A Natural Complexity Measure for Time Series, Phys. Rev. Lett. **88**, 174102 (2002).
[53] J. Xie, J. Gao, Z. Gao, X. Lv, and R. Wang, Adaptive symbolic transfer entropy and its applications in modeling for complex industrial systems, Chaos **29**, 093114 (2019).
[54] C. Torrence and G. P. Compo, A practical guide to wavelet analysis, Bull. Am. Meteorol. Soc. **79**, 61 (1998).
[55] S. Panzeri, R. Senatore, M. A. Montemurro, and R. S. Petersen, Correcting for the sampling bias problem in spike train information measures, J. Neurophysiol. **98**, 1064 (2007).
[56] R. Andrzejak, A. Ledberg, and G. Deco, Detecting event-related time-dependent directional couplings, New J. Phys. **8**, 6 (2006).
[57] T. Wagner, J. Fell, and K. Lehnertz, The detection of transient directional couplings based on phase synchronization, New J. Phys. **12**, 053031 (2010).
[58] M. Martini, T. A. Kranz, T. Wagner, and K. Lehnertz, Inferring directional interactions from transient signals with symbolic transfer entropy, Phys. Rev. E **83**, 011919 (2011).
[59] B. Phipson and G. K. Smyth, Permutation p-values should never be zero: Calculating exact p-values when permutations are randomly drawn, Stat. Appl. Genet. Mol. Biol. **9**(1) (2010).

[60] E. Maris and R. Oostenveld, Nonparametric statistical testing of EEG- and MEG-data, J. Neurosci. Methods **164**, 177 (2007).

[61] M. Lindner, R. Vicente, V. Priesemann, and M. Wibral, TRENTOOL: A Matlab open source toolbox to analyse information flow in time series data with transfer entropy, BMC Neurosci. **12**, 119 (2011).

[62] A. L. Lloyd, The coupled logistic map: A simple model for the effects of spatial heterogeneity on population dynamics, J. Theor. Biol. **173**, 217 (1995).

[63] A. Valencio and M. d. S. Baptista, Coupled logistic maps: Functions for generating the time-series from networks of coupled logistic systems (2018), open source codes for MATLAB. available at https://github.com/artvalencio/coupled-logistic-maps.

[64] Y. Tian, Y. Wang, Z. Zhang, and P. Sun, Transfer entropy spectrum in the Fourier domain (2021), open source codes available at https://github.com/doloMing/Transfer-entropy-spectrum-in-the-Fourier-domain.