

Irreversibility as a signature of non-equilibrium phase transition in large-scale human brain networks: An fMRI study

Jing Wang(王菁), Kejian Wu(吴克俭), Jiaqi Dong(董家奇), and Lianchun Yu(俞连春)
Citation: Chin. Phys. B, 2025, 34 (**5**): 058703. DOI: 10.1088/1674-1056/adbd27
http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Dynamics of zebrafish locomotion being independent of spatial size

Zhen Wang(王震), Jian Gao(高见), Yongshang Long(龙永尚), Huaping Lv(吕华平), and Qun Wang(王群)
Chin. Phys. B, 2025, 34 (**3**): 038702. DOI: 10.1088/1674-1056/ada54c

Two innovative equivalent statements of the third law of thermodynamics

Xiaohang Chen(陈晓航), Yinghui Zhou(周颖慧), and Jincan Chen(陈金灿)|
Chin. Phys. B, 2024, 33 (**6**): 060504. DOI: 10.1088/1674-1056/ad39c8

Effect of cognitive training on brain dynamics

Guiyang Lv(吕贵阳), Tianyong Xu(徐天勇), Feiyan Chen(陈飞燕), Ping Zhu(朱萍), Miao Wang(王淼), and Guoguang He(何国光)
Chin. Phys. B, 2024, 33 (**2**): 028704. DOI: 10.1088/1674-1056/ad09c8

Symmetric Brownian motor subjected to Lévy noise

Kao Jia(贾考), Lan Hu(胡兰), and Linru Nie(聂林如)
Chin. Phys. B, 2024, 33 (**2**): 020502. DOI: 10.1088/1674-1056/ace8fa

Experimental investigations of the functional morphology of dragonfly wings

H. Rajabi, A. Darvizeh
Chin. Phys. B, 2013, 22 (**8**): 088702. DOI: 10.1088/1674-1056/22/8/088702

Irreversibility as a signature of non-equilibrium phase transition in large-scale human brain networks: An fMRI study

Jing Wang(王菁), Kejian Wu(吴克俭), Jiaqi Dong(董家奇), and Lianchun Yu(俞连春)†

School of Physical Science and Technology, Lanzhou Center for Theoretical Physics, Key Laboratory of Theoretical Physics of Gansu Province, Key Laboratory of Quantum Theory and Applications of Ministry of Education, Gansu Provincial Research Center for Basic Disciplines of Quantum Physics, Lanzhou University, Lanzhou 730000, *China*

(Received 10 December 2024; revised manuscript received 26 January 2025; accepted manuscript online 6 March 2025)

It has been argued that the human brain, as an information-processing machine, operates near a phase transition point in a non-equilibrium state, where it violates detailed balance leading to entropy production. Thus, the assessment of irreversibility in brain networks can provide valuable insights into their non-equilibrium properties. In this study, we utilized an open-source whole-brain functional magnetic resonance imaging (fMRI) dataset from both resting and task states to evaluate the irreversibility of large-scale human brain networks. Our analysis revealed that the brain networks exhibited significant irreversibility, violating detailed balance, and generating entropy. Notably, both physical and cognitive tasks increased the extent of this violation compared to the resting state. Regardless of the state (rest or task), interactions between pairs of brain regions were the primary contributors to this irreversibility. Moreover, we observed that as global synchrony increased within brain networks, so did irreversibility. The first derivative of irreversibility with respect to synchronization peaked near the phase transition point, characterized by the moderate mean synchronization and maximized synchronization entropy of blood oxygenation level-dependent (BOLD) signals. These findings deepen our understanding of the non-equilibrium dynamics of large-scale brain networks, particularly in relation to their phase transition behaviors, and may have potential clinical applications for brain disorders.

Keywords: large-scale brain networks, fMRI, irreversibility, non-equilibrium phase transition

PACS: 87.18.–h, 05.70.Ln

DOI: 10.1088/1674-1056/adbd27 **CSTR:** 32038.14.CPB.adbd27

1. Introduction

Many physical, chemical, and biological processes operate far from thermodynamic equilibrium.[1] From molecules to organisms and machines, these non-equilibrium systems maintain their states by exchanging matter and energy with their environment. In contrast to equilibrium systems, whose dynamics adhere to detailed balance where net fluxes of transitions between different states vanish, non-equilibrium systems exhibit net fluxes between states or configurations, thereby breaking detailed balance.[2,3] This breaking of detailed balance inherently leads to irreversible dynamics that establish an arrow of time: by observing the system's evolution, one can distinguish whether time flows forward or backward.[4] Modern non-equilibrium statistical physics has demonstrated that the rate of evidence accumulation for the arrow of time equals the rate of entropy production in the system's environment.[5–7] This concept has been applied to detect signatures of irreversibility in various living systems through experimental data analysis.[6,8,9]

The human brain, which accounts for about 20% of the body's energy consumption despite representing only 2% of its weight,[10,11] relies on constant exchanges of energy and matter to maintain its activity levels. The brain's metabolic and enzymatic activities drive various non-equilibrium processes at the microscale, such as neuronal firing,[12] molecular cycles,[13] and cellular maintenance.[14] Given these facts, it is unlikely that cognitive functions can be supported by neural dynamics close to thermodynamic equilibrium. Consequently, it can be inferred that the brain must inherently break detailed balance at larger scales, leading to natural temporal asymmetry in brain activity.[8] Emerging evidence increasingly supports the notion that brain dynamics operate outside thermodynamic equilibrium, characterized by a strictly positive entropy production rate, making neural activity time series temporally irreversible.[15] This temporal irreversibility not only underscores the intricate complexity of the brain's functional architecture,[16] but is also empirically associated with levels of consciousness,[9,17] cognitive performance,[8,15,18] and various mental disorders.[19]

Modern neuroscience is enhancing our understanding of the structure and function of the human brain by employing principles from network science.[20,21] The "critical brain" hypothesis suggests that neural networks operate near a critical or phase transition point between order and disorder,[22] facilitating optimal information processing and adaptability.[23–26] At the level of large-scale brain networks, critical dynamics have been shown to maximize the diversity and flexibility of functional connectivity patterns,[27–29] support the bal-

†Corresponding author. E-mail: yulch@lzu.edu.cn

© 2025 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

ance between network segregation and integration,[30,31] and influence aspects of cognitive performance, such as fluid intelligence and working memory capacity.[32] However, previous studies on non-equilibrium phase transitions in the brain have often approached them by analogizing them to equilibrium systems, such as using order parameters.[32,33] Yet, this approach overlooks the irreversible nature of the dynamics, as well as its influence on phase transition properties. Entropy production has been revealed as an important concept to address this limitation, as it vanishes identically for equilibrium systems and is positive for non-equilibrium cases.[34–36] Furthermore, it remains unclear whether high-order interactions in brain networks contribute to orchestrating activity across brain areas, or if pairwise interactions suffice.[37] The recently developed irreversibility decomposition method may offer insights from a thermodynamic perspective.[7] Also, while various methods for measuring irreversibility in neural data have been proposed, they show significant inconsistency on the same dataset, highlighting their inaccuracy. Thus, evaluating the new method on the same dataset is crucial for confirming its reliability.

In this study, we investigate the characterization of phase transitions in large-scale human brain networks in terms of their irreversibility. By analyzing open-source fMRI data from resting and seven cognitive tasks of the Washington University–University of Minnesota Consortium (WU-Minn) Human Connectome Project (HCP),[38] we found that large-scale brain networks violate detailed balance by exhibiting significant irreversibility and generating positive entropy, with these violations increasing alongside physical and cognitive exertion. By decomposing irreversibility into independent and interaction terms of different orders, we found that irreversibility in large-scale brain networks is primarily driven by interactions among pairs of brain regions. Furthermore, we observed that the entropy production rate in large-scale brain networks exhibited significant non-equilibrium characteristics. For example, the entropy production rate increased monotonically as the networks became more synchronized, and the derivative of entropy production with respect to synchronization among regional BOLD signals was maximized at the phase transition point.

2. Materials and methods

2.1. fMRI data acquisition and preprocessing

The fMRI data were obtained from the HCP 1200-subject release.[38] Informed consent was obtained from all participants by the WU-Minn Consortium, and the research procedures were approved by the Washington University Institutional Review Board. This dataset includes fMRI scans for both resting state and seven different task conditions: emotional processing, working memory, social inference, language processing, relational matching, gambling, and motor execution.

In this study, we used two sets of fMRI data (Fig. 1(a)). The first dataset comprised eight cohorts: one cohort consisted of resting-state fMRI (rfMRI) data from 270 subjects, obtained from the resting-state fMRI 1 FIX-denoised package. The other seven cohorts included task fMRI (tfMRI) data for these same subjects from the HCP task scans (dataset 1 in Fig. 1(a)). To ensure consistency in data size across resting and task states, only the first 176 time points (which correspond to the scan length of the emotional processing task, the shortest among all tasks) were used in the analysis. The second dataset included rfMRI data from 801 subjects, drawn from the first resting-state dataset, with both left-to-right and right-to-left acquisitions (dataset 2 in Fig. 1(a)). For the second dataset, all 1200 time points were included in the analysis.

The rfMRI data were processed using the HCP minimal preprocessing pipeline[39,40] and the denoising procedures,[41] while the tfMRI data underwent the same HCP minimal preprocessing pipeline without filtering to retain potential high-frequency information. Additionally, a general linear regression model was used to remove the average signal associated with the tasks. Global brain signals were retained in this analysis due to the ongoing debate over their removal as a preprocessing step.[42] After preprocessing, 100 regional BOLD signals were extracted by parcellating the cortex into 100 regions using the Schaefer-100 atlas (Fig. 1(b)),[43] with each region mapped to one of the seven functional networks identified by Yeo *et al.*[44]

2.2. fMRI data analysis

2.2.1. Estimating the irreversibility of regional BOLD signals

Measuring the deviation of a system from detailed balance is essential for quantifying irreversibility. In an irreversible system, fluxes from one state to another break detailed balance. For such a system with joint transition probabilities, the Kullback–Leibler divergence between its forward and reverse state transition probabilities can be calculated to quantify irreversibility:

$$\dot{I} = \sum_{x,x'} P\left(x \to x'\right) \log \left[\frac{P\left(x \to x'\right)}{P\left(x' \to x\right)}\right], \tag{1}$$

where $P(x \to x') \equiv \mathrm{Prob}(x_t = x,\ x_{t+1} = x')$ is the joint transition probability between states x and x', and x_t is the state of the system at time t. If the system obeys detailed balance (i.e., $P(x \to x') = P(x' \to x)$ for all pairs of states x and x'), the irreversibility vanishes. Conversely, any violation of detailed balance (i.e., any flux of transitions such that $P(x \to x') \neq P(x' \to x)$) results in increased irreversibility. Thus, calculating the state transition probability $P(x \to x')$ is a critical step in quantifying a system's deviation from detailed balance.

In this study, we employed the methods proposed by Lynn *et al*. to estimate the joint transition probability $P(x \to x')$ and evaluate irreversibility in large-scale brain networks.[8,45] Notably, if the states x and x' encompass all microscopic degrees of freedom in a system, then equation (1) represents the physical rate of entropy production.[5,46] However, for macroscopic systems, equation (1) provides an information-theoretic measure of entropy production, serving as a lower bound on the physical entropy production inferred from coarse-grained dynamics.

As shown in Fig. 1(c), the regional BOLD signal for each brain region was extracted and normalized by subtracting its mean and dividing by its standard deviation. A threshold of 1 standard deviation was then applied to transform the signal into point processes by identifying threshold-crossing events from below. Figure 1(d) illustrates the raster plot of these threshold-crossing events in the resting-state brain of an arbitrarily chosen subject. For the analysis, we arbitrarily selected only N regions. Including too many regions could complicate the estimation of irreversibility (as discussed at the end of this subsection). Subsequently, the state transition matrix was constructed using a sliding window approach, similar to prior work analyzing spike trains generated by neurons in the salamander retina.[45] Briefly, considering a window of length ΔT sliding along the raster plots, a brain region transitions to the active state when an event enters the front of the window (Fig. 1(e), left window). Conversely, the region transitions to the inactive state when an event exits the back of the window (Fig. 1(e), center window). Self-transitions occur when a brain region is activated twice consecutively within the window (Fig. 1(e), right window). However, the all-silent state $\{0,\ldots,0\}$ does not exhibit self-transitions, since no events enter or exit the sliding window. Figure 1(f) shows an example of a state transition probability matrix for three regions. After calculating the state transition probability matrix for each subject within each brain state, these matrices were summed to provide a robust estimation of state transition probabilities across different brain states.

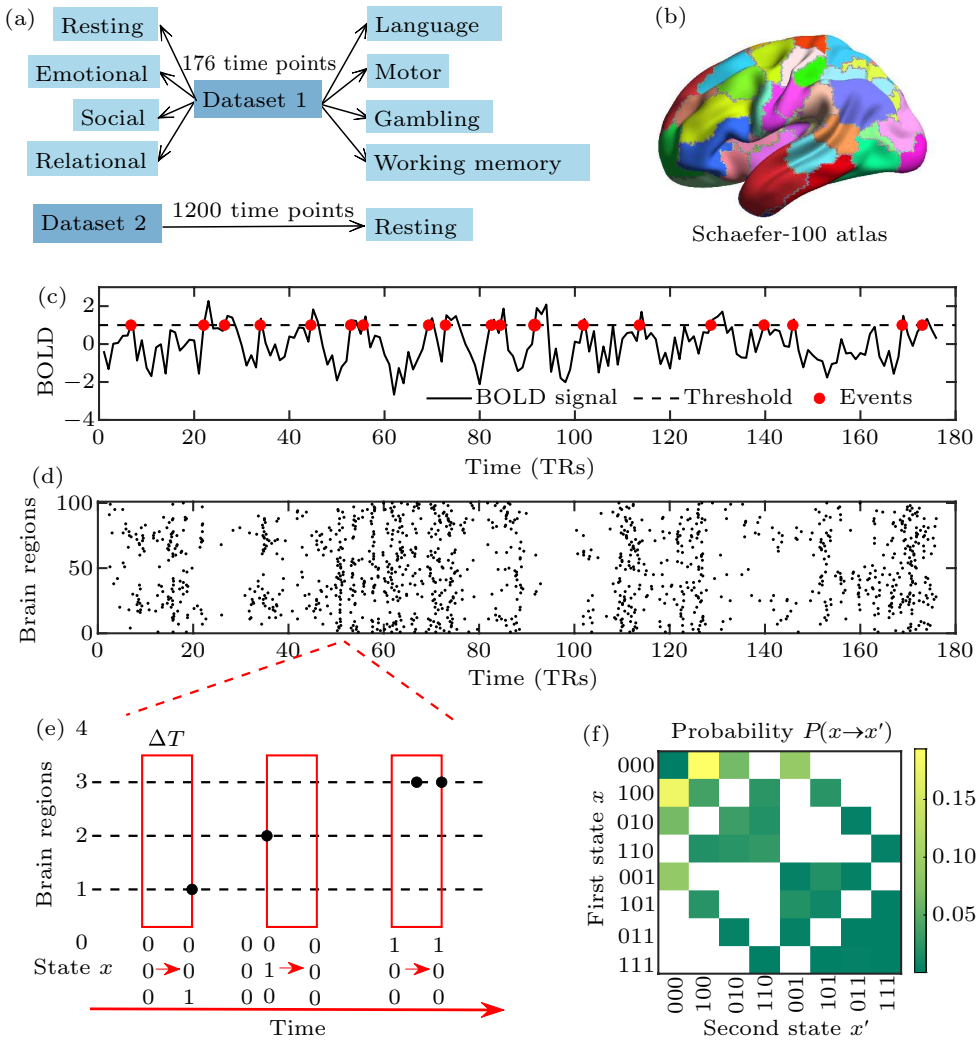


Fig. 1. The workflow to estimate the irreversibility from the regional BOLD signals. (a) The two datasets involved in the study. (b) The Schaefer-100 atlas that was used to parcellate the whole cortex into 100 regions. (c) An example demonstrated the conversion of normalized BOLD signals into a point process using a threshold of 1 standard deviation by defining the threshold crossing events. (d) The raster plot showed the events for an arbitrarily chosen subject. (e) State transitions were defined from the raster plots of events using a sliding window of length ΔT. (f) An example illustrated the state transition probability matrix for three regions.

Figure 2(a) illustrates the estimated irreversibility across different brain states as a function of the sliding window length. When the window is too short, random noise dominates the estimates, reducing irreversibility. Conversely, overly long windows tend to smooth out important state transitions, leading to a nearly symmetric state transition matrix and reduced irreversibility. Therefore, the optimal window length, corresponding to the peak irreversibility, minimizes noise influence while capturing the system's dynamic changes. Moreover, this peak also reflects the greatest asymmetry in the state transition matrix, representing the most prominent directional dynamics of the system. Based on this observation, we selected $\Delta T = 4$ s for irreversibility estimation in this study.

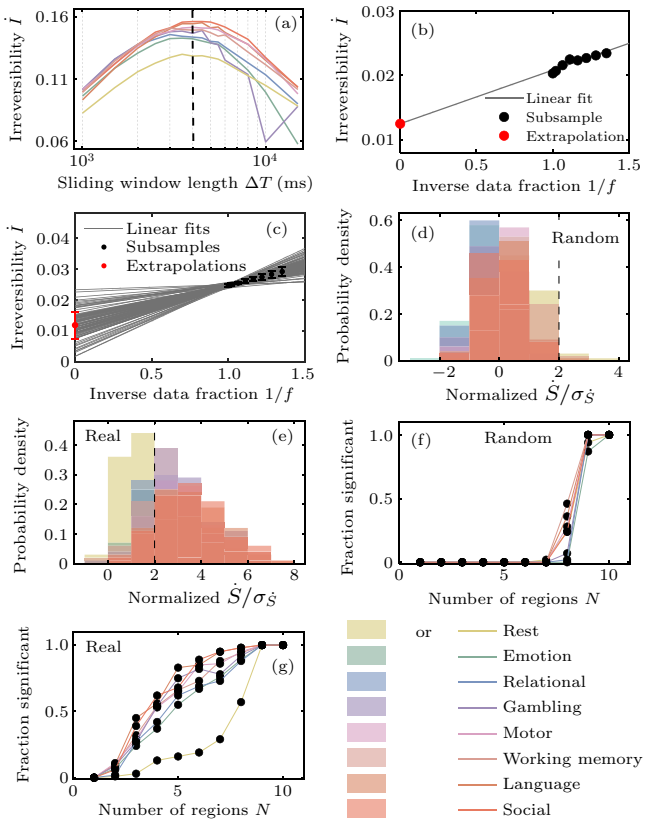


Fig. 2. The strategies used to account for limited data effects. (a) The impact of estimating different time window lengths ΔT on the calculation of irreversibility. Based on the observation, $\Delta T = 4$ s was selected for further analysis. (b) The irreversibility for the infinite-data (red dot) was estimated by extrapolating from values of irreversibility for subsamples (black dots) to the infinite-data limit. (c) The process in panel (b) was repeated 100 times to obtain both the average (red dot) and standard deviation (vertical red line) for the infinite-data estimates of the irreversibility. (d)–(e) The probability distributions of irreversibility for randomized (d) and original data (e) across brain region groups in different brain states. The vertical dashed line indicates the 2 times standard deviation, which is the threshold for significance. (f)–(g) Fraction of different states of the brain with significant irreversibility as a function of the number of regions N for randomized (f) and original data (g).

Given the long-term temporal dependencies inherent in BOLD signals, it is crucial to correct for finite data effects in order to accurately calculate irreversibility.[47–49] To address this issue, for each brain state, we considered 100 randomly selected $N = 6$ region groups and enumerated all possible state transitions that could be observed within each group. We then hierarchically subsampled the state transitions (without replacement) for varying data fractions f, ensuring that each subsample was a strict subset of the larger dataset. Our analysis determined that the data fractions $\{0.98, 0.94, 0.90, 0.86, 0.82, 0.78, 0.74\}$ were sufficient for robust estimation. For each subsample, we calculated irreversibility (black dots in Fig. 2(b)) and extrapolated the results to the infinite-data limit (red dot in Fig. 2(b)). This process was repeated 100 times (black lines in Fig. 2(c)) to obtain the average irreversibility estimate (red dot in Fig. 2(c)) and its standard deviation (red vertical bar in Fig. 2(c)) at the infinite-data limit. To control for inter-group variability across each brain state, we standardized the distribution of irreversibility estimates for each brain state by dividing the mean irreversibility across all region groups by their standard deviation. (Figs. 2(d) and 2(e)). Then we used two times the standard deviation as the threshold of significance (dashed vertical lines). It was found that for the time-randomized data, almost all brain region groups exhibit irreversibility distributions centered around zero, with only a few showing significant irreversibility, suggesting that time randomization disrupts the temporal directionality of the brain across different states (Fig. 2(d)). However, after correcting for the limited data effects, most brain region groups demonstrate significant irreversibility in the real data (Fig. 2(e)).

Theoretically, the calculated irreversibility should approach zero under conditions of infinite data constraints for time-randomized data. However, our analysis revealed that even for randomized data, the estimated irreversibility remained effectively zero when the number of brain regions in the brain region group was fewer than 7. In contrast, it increased sharply when the number of regions exceeded 7 (Fig. 2(f)), underscoring the difficulty of fully accounting for limited data effects in this context. For the original real data, the proportion of region groups exhibiting significant irreversibility increased from 10% at $N = 2$ to 100% at $N = 9$ (Fig. 2(g)). Based on these findings, $N = 6$ was determined as the maximum number of regions within a brain region group that could be selected to ensure reliable irreversibility estimates.

2.2.2. Decomposing irreversibility in brain networks

The irreversibility of complex systems can often be decomposed because they typically arise from different, interacting subsystems or processes that contribute to the overall direction of time's "arrow" within that system. Lynn *et al.* proposed that irreversibility could be decomposed into non-negative contributions in terms of the interactions within complex systems, specifically, from individual elements in the system ($\dot{I}_{\text{ind}}$), interactions between pairs of elements ($\dot{I}^{(2)}_{\text{int}}$), inter-

actions among triplets ($\dot{I}^{(3)}_{\rm int}$), and so on:[7,45]

$$\dot{I} = \dot{I}_{\rm ind} + \dot{I}^{(2)}_{\rm int} + \dot{I}^{(3)}_{\rm int} + \cdots + \cdots + \dot{I}^{(k)}_{\rm int} + \cdots + \dot{I}^{(N)}_{\rm int}, \tag{2}$$

where the k-th-order interaction irreversibility $\dot{I}^{(k)}_{\rm int} = \dot{I}^{(k)} - \dot{I}^{(k-1)} \geq 0$, and irreversibility $\dot{I}^{(k)}$ corresponds to the dynamics of k elements ($k < N$), which can be estimated from the marginal dynamics of the k elements. More details on the decomposition of irreversibility can be found in Ref. [45].

2.2.3. Measuring the mean synchronization and synchronization entropy among BOLD signals

For the second dataset which contains only the resting-state BOLD signals, we would utilize the previously proposed method to identify the phase transition point in the large-scale human brain networks characterized by moderate mean synchronization and maximized synchronization entropy among regional BOLD signals. The mean synchronization and synchronization entropy are calculated below.

For the obtained resting-state regional BOLD signals $S_q(t)$ from the second dataset, the Hilbert transform was performed to calculate the phase trace $\theta_q(t)$ of the signal

$$\theta_q(t) = \arctan \frac{H[S_q(t)]}{S_q(t)}. \tag{3}$$

Then instantaneous synchronization $r(t)$ among all the 100 BOLD signals can be calculated as follows:

$$r(t) = \frac{1}{u}\left|\sum_{q=1}^{u} \mathrm{e}^{\mathrm{i}\theta_q(t)}\right|, \tag{4}$$

where $u = 100$ represents the number of BOLD signals.

Then the mean synchronization among BOLD signals is calculated as the Kuramoto order parameter by averaging the instantaneous synchronization $r(t)$ across time

$$\bar{r} = \frac{1}{l}\sum_{t=1}^{l} r(t), \tag{5}$$

where l represents the total number of the time points.

Finally, through the distribution of $r(t)$, the synchronization entropy which quantifies synchronization variability over time can be calculated as

$$H(r) = -\sum_{i=1}^{n} p_i \log_2 p_i, \tag{6}$$

where p_i is the probability that $r(t)$ falls into a bin between r_i and r_{i+1}, and $n = 30$ is the number of bins that divides the distribution of $r(t)$.

3. Results

In this section, we employed the approach outlined in Section 2 to evaluate the extent of violation of detailed balance in the brain network across both the resting state and seven different task states (Subsection 3.1). Next, we assessed the contribution of various orders of brain dynamics to irreversibility (Subsection 3.2). Subsequently, we used the moderate mean and maximized variability of synchronization among regional BOLD signals during the resting state to identify phase transitions in the brain network and explore potential signatures of critical points as reflected by irreversibility (Subsection 3.3).

3.1. The violation of detailed balance in brain networks

We employed the Schaefer-100 atlas to parcellate the whole brain into 100 regions (Fig. 1(b)). From dataset 1, the corresponding 100 regional BOLD signals were extracted for each subject in each brain state (Fig. 1(a)). For all BOLD signals, a threshold of 1 standard deviation was used to convert these signals into point processes (Fig. 1(c)). For each subject and brain state, we randomly selected 100 region groups from the temporal point processes. Each group consisted of $N = 6$ brain regions (Figs. 2(f) and 2(g)). A state transition probability matrix for each group was calculated using a sliding window approach (Fig. 1(e)). To achieve a stable estimation of irreversibility, the state transition probability matrices for all subjects were summed to create averaged state transition probability matrices for each brain state. Next, a stable estimation of irreversibility for the infinite-data limit was achieved by extrapolating from the values of irreversibility estimated from subsamples of finite data (Figs. 2(b) and 2(c)). Finally, the irreversibility values of the groups showing significant irreversibility were averaged to calculate the irreversibility of the brain in different states.

As shown in Fig. 3, across all tasks and at rest, brain dynamics exhibit irreversibility, suggesting that the brain generates entropy by breaking detailed balance at large scales. Furthermore, brain dynamics exhibited greater irreversibility during all physical and cognitive tasks compared to the resting state. Notably, the relative irreversibility for different tasks measured by the method in our study closely mirrors that assessed using the temporal evolution NETwork (TENET) framework as reported by Deco *et al.*,[15] except that the irreversibility assessed by the TENET framework is larger than that measured in our study for relational tasks. It is also noted that there are works employing other estimation methods on the same HCP data, such as estimates derived from coarse-grained states through clustering neuroimaging data,[8] or using the INSIDEOUT framework, which quantifies the breaking of detailed balance based on the level of trajectory asymmetry.[16] Although the irreversibility for the resting state estimated using these two methods was among the lowest, the irreversibility for the physical and task states showed substantial differences compared to the results in this study and those obtained with the TENET method. This implies that differ-

ent estimation methods for irreversibility from the fMRI data of brains have inherent limitations in terms of accuracy. Future studies should be mindful of these limitations when selecting or interpreting methods to assess irreversibility. It is also noted that although the BOLD signals were converted into point processes in this study, which could potentially reduce measurement accuracy, the results closely matched those from the TENET method, known for its high accuracy due to utilizing the full signals. This suggests that the point processes may retain most of the information in the BOLD signals.[50] Nonetheless, we can still argue that the degree to which the brain breaks detailed balance depends critically on the specific task being performed.

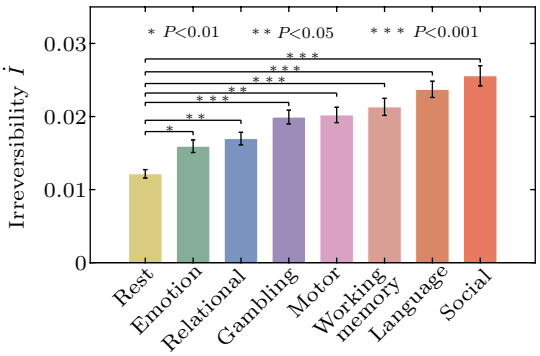


Fig. 3. Irreversibility in the brain varies with physical and cognitive demands. The stars indicate the level of significant difference with two-sample t-tests.

3.2. Irreversibility of brain networks emerges primarily from pairwise dynamics

Lynn *et al*. proposed that irreversibility can be decomposed into two components: an independent term reflecting the dynamics of individual elements and an interaction term driven by dependencies among elements. The interaction term can further be divided into contributions from pairs, triplets, and higher-order interactions among elements.[7]

Using this decomposition method, we analyzed the brain irreversibility in the resting state and across seven different task states (Fig. 4(a)). Our findings revealed that the pairwise dynamics contributed significantly more to irreversibility than higher-order dynamics. Specifically, the pairwise interactions contributed approximately 50% of the irreversibility in both resting and task states, which is comparable to the combined contributions of the third-, fourth-, fifth-, and sixth-order interactions. Notably, while the magnitude of the irreversibility varied significantly across tasks, the proportion of the irreversibility attributed to each order of dynamics remained relatively consistent (Fig. 4(a)). Lynn *et al*. observed that irreversibility in retinal neurons did not merely reflect the irreversibility of the stimulus, but consistently arose from the same combination of low-order dynamics, primarily driven by pairwise neuron interactions.[45] Thus, it is conjectured that pairwise interactions play a crucial role across various scales of brain networks.

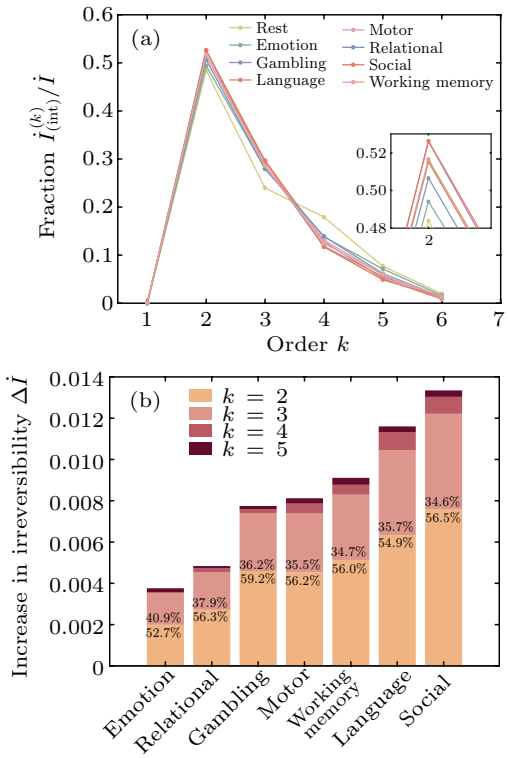


Fig. 4. Decomposition of the brain network irreversibility. (a) The fraction of different order interaction irreversibility, $\dot{I}^{(k)}_{\mathrm{int}}$, relative to the total irreversibility, $\dot{I}$, is presented as a function of the order (k) for the brain in different states. Inset: enlargement for $k=2$. (b) The increase in irreversibility was observed during the task states compared to the resting state. Different colored areas indicate the relative contribution of each order of interaction irreversibility ($k=2,3,\ldots,6$) to the increase in total irreversibility.

Additionally, we quantified how different orders of interaction dynamics contributed to the increase in brain irreversibility during task states compared to the resting state. The histograms in Fig. 4(b) illustrate the increase in irreversibility as brain networks transition from the resting state to various task states. In each column, the differently colored areas represent the relative contributions of each order of interaction dynamics to the overall increase in irreversibility. It is clear that the second- and third-order dynamics are key drivers, significantly enhancing brain irreversibility during the transition from the resting state to various task states (Fig. 4(b)). This result possibly suggested that the second- and third-order dynamics are enough to capture the main changes of the brain networks during transition among different brain states.

3.3. Irreversibility as a signature of phase transition in brain networks

Critical brain hypothesis suggests that, like other complex systems, healthy brain dynamics operate at the brink of a critical phase transition between order and disorder.[29] Previous studies have demonstrated that phase transitions in neural networks,[51] including large-scale brain networks, can be

characterized by moderate mean synchronization among regional BOLD signals and maximized synchronization entropy, where scale-free avalanche activities would emerge.[32,52] In this study, to investigate potential signatures of irreversibility during phase transitions in large-scale brain networks, we used regional BOLD signals extracted from the rfMRI data of the second dataset. First, we identified the phase transition by examining the mean synchronization and synchronization entropy of each subject's brain. For each subject, we calculated global mean synchronization and synchronization entropy using 100 regional BOLD signals. By plotting the values of $\bar{r}$ and $H(r)$ for each subject on the mean synchronization versus synchronization entropy plane, we observed a significant inverted U-shaped curve (Fig. 5(a)), consistent with previous studies.[32]

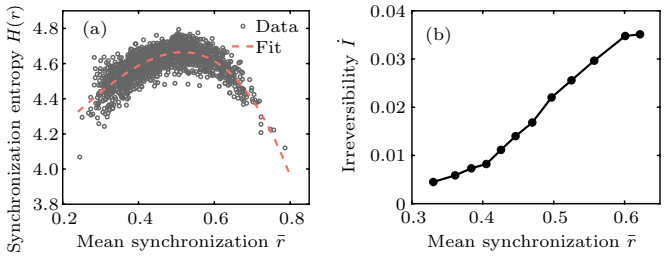


Fig. 5. The behavior of irreversibility during asynchronous–synchronous phase transition of large-scale brain networks. (a) The inverted U-shaped relationship between mean synchronization and synchronization entropy was observed. Each black circle represents the values for mean synchronization and synchronization entropy of one subject, measured through the BOLD signals from 100 regions for each subject. The red dashed line represents the quadratic fitting of the data ($R^2 = 0.5167$). The maximum value of the fitted data was at $\bar{r} = 0.52$. (b) The irreversibility of brain networks is presented as a function of mean synchronization among regional BOLD signals. Each data point indicates the value of the irreversibility calculated from the group of subjects with similar mean synchronization.

To obtain stable estimates of irreversibility in brain networks at varying levels of synchrony, we divided the subjects into 12 groups based on their mean synchronization values (group means were 0.33, 0.36, 0.38, 0.41, 0.43, 0.45, 0.47, 0.50, 0.53, 0.56, 0.60, and 0.62). Each group comprised 260 subjects. The irreversibility for each group was then calculated (Fig. 5(b)), revealing a monotonic increase in irreversibility with the increase of the mean synchronization. We note here that though many works have dedicated efforts to investigating the nontrivial behavior of irreversibility (entropy production) during phase transition,[36,53] a recent work on the asymmetric Sherrington–Kirkpatrick model, which exhibits order-disorder non-equilibrium phase transitions analogous to the paramagnetic–ferromagnetic phase transitions in the equilibrium Ising model, found that entropy production can reach its highest values in low-temperature regimes (ordered phases) with highly heterogeneous couplings, without necessarily peaking at the critical point of the phase transition.[35] Considering the substantial heterogeneity in both the structural connectivity and regional cyto-, myelo-, and chemoarchitecture of brain networks,[54,55] we argue that the high irreversibility observed in the synchronized phase of brain networks reflects this inherent heterogeneity. Furthermore, this result suggests that an increase in entropy production does not necessarily indicate that the brain is approaching a critical state.

Previous studies on the critical behavior of the Ising model driven by external field have shown that the first derivative of entropy production with respect to the order parameter exhibits divergent behavior near the phase transition point.[36,56] This is reflected in the peak of the first derivative of entropy production, which indicates the fastest increase in the entropy generation rate close to the critical point. Motivated by these findings, we computed the first-order derivative of the irreversibility with respect to the mean synchronization of regional BOLD signals. Due to significant fluctuations in the estimated irreversibility relative to mean synchronization, we employed three distinct approaches to calculate the derivative of the irreversibility with respect to mean synchronization to validate the results. Our first approach directly calculated the derivative of the irreversibility from the data on the irreversibility versus mean synchronization curve shown in Fig. 5(b) (also represented by the green dots in Fig. 6(a)). This approach, however, resulted in high variability due to the limited accuracy of the calculations (Fig. 6(b)). To address this limitation, we employed alternative methods. First, we fitted to the original the irreversibility versus mean synchronization curve with a cubic function (red line in Fig. 6(a)). The fitting process was performed using the least squares method. The function was expressed as $\dot{I} = p_1\bar{r}^3 + p_2\bar{r}^2 + p_3\bar{r} + p_4$, where p_1, p_2, p_3, and p_4 are the coefficients to be determined through the fitting procedure. The optimal values of these coefficients were obtained as follows: $p_1 = -2.1285$, $p_2 = 3.1266$, $p_3 = -1.3379$, and $p_4 = 0.1960$, yielding a coefficient of determination $R^2 = 0.9986$, indicating an excellent fit of the model to the data. Then we calculated the derivative of the irreversibility with respect to mean synchronization from this fitted curve (Fig. 6(c)). Additionally, we utilized a low-pass filtering method to eliminate high-frequency noise and ensure the smoothness of the data (Fig. 6(a), blue dashed line). Firstly, linear interpolation was applied to generate uniformly distributed new x-axis values, and corresponding y-axis values were calculated. Then, a low-pass finite impulse response filter was designed using the Hanning window method to ensure the filter's smoothness. Bidirectional filtering was employed to maintain the phase of the data, involving both the forward and reverse filtering processes. We then calculated the derivative of the irreversibility with respect to mean synchronization from the smoothed curve (Fig. 6(d)). As shown in Figs. 6(b)–6(d), regardless of the approach used, the derivative of the irreversibility consistently reached a maximum around moderate mean synchronization, where synchronization en-

tropy is also maximized. It is noted that the estimated maximal derivative of entropy production is in the range between $\bar{r} = (0.45, 0.47)$, whereas the estimated critical point is around $\bar{r} = 0.52$. This difference is mainly due to significant calculation errors in the realistic fMRI data, which is an indirect, low signal-to-noise measure of neural activity. Therefore, the conclusion requires further testing on additional datasets, and potentially a biologically realistic whole-brain model, to better understand the underlying mechanisms of irreversibility.[52,57] Anyway, echoing the theoretical predictions that the derivative of entropy production with respect to the order parameter exhibits divergent behavior near the phase transition point, our finding suggests that the derivative of entropy production may serve as an indicator of phase transitions in large-scale brain networks.

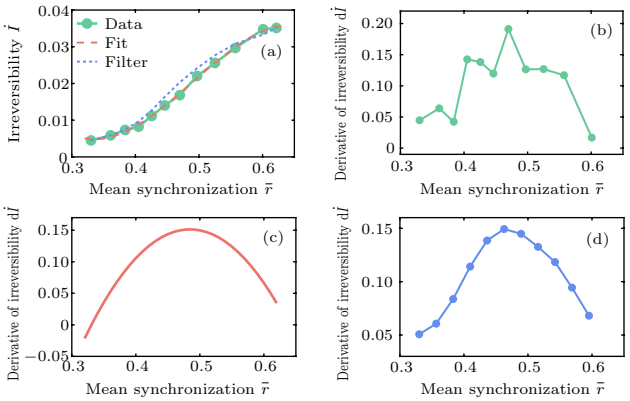


Fig. 6. The behavior of the first-order derivative of the irreversibility with respect to the mean synchronization during phase transition. (a) The irreversibility versus mean synchronization curve shown with original data (green dots), its fitted curve (red line), and its low-pass filtered curve (dashed blue line), respectively. (b) The first-order derivative of irreversibility was estimated from the original irreversibility versus mean synchronization data curve. The maximum value was at $\bar{r} = 0.47$. (c) The same as panel (b) but estimated from the fitted curve. The maximum value was at $\bar{r} = 0.48$. (d) The same as panel (b) but estimated from the filtered curve. The maximum value was at $\bar{r} = 0.46$.

4. Discussion and conclusion

In this study, we investigated the irreversibility of large-scale human brain networks using open-source fMRI data from the resting state and seven different task conditions. Our findings show that the human brain violates detailed balance by exhibiting significant irreversibility and producing entropy. Across different task states, irreversibility was higher than that observed in the resting state, suggesting that the degree to which large-scale brain networks break detailed balance depends critically on the specific task being performed.

We also explored the irreversibility of brain networks during phase transitions. Using a recently proposed characterization of phase transitions based on mean synchronization and synchronization entropy among regional BOLD signals, we found that irreversibility increased as global synchronization among regional activities grew. However, the derivative of irreversibility with respect to synchronization peaked around the phase transition point, where synchronization entropy was maximized. We argue that irreversibility could serve as a systematic description of the non-equilibrium nature of large-scale brain networks, including phase transitions.

Entropy production is often effective in describing how a structure self-organizes within a non-equilibrium thermodynamic system and in measuring the extent to which the system dissipates energy.[58] By quantifying the irreversibility, i.e., entropy production, of large-scale brain networks during rest and various task states, as well as during the phase transition process, this work demonstrates that the violation of detailed balance in large-scale brain networks supports more complex brain functions. Though several other methods have been applied to the same HCP dataset to estimate the irreversibility of brain networks during rest and task states, the method used in this study yield results that are similar to those from the deep learning-based TENET framework,[15] suggesting the effectiveness of this method in reliably estimating the irreversibility of brain networks. By decomposing the irreversibility into interaction terms with different orders, a previous study showed that the dynamics of paired interactions among neurons play a vital role at the level of neuronal networks.[45] In this study, we also reported that the irreversibility of large-scale brain networks consistently arises from pairwise dynamics, regardless of the brain states (rest or different tasks). This result echoes the previous findings that indicate the dominance of pairwise interactions in organizing coordinated activities among different brain areas to support various brain functions.[37,59]

Furthermore, we found that the irreversibility increased monotonically as the brain networks became more synchronized. However, the first-order derivative of entropy production relative to the synchronization of brain networks was maximized near the phase transition point. This implies that combining the irreversibility and its derivative could serve as an indicator of the brain networks approaching a critical point. Given the clinically relevant roles of brain criticality,[60] this finding could have potential clinical applications for brain disorders.

Acknowledgements

Project supported by the Fundamental Research Funds for the Central Universities (Grant Nos. lzujbky-2021-62 and lzujbky-2024-jdzx06), the National Natural Science Foundation of China (Grant No. 12247101), the Natural Science Foundation of Gansu Province, China (Grant Nos. 22JR5RA389 and 23JRRA1740), and the '111 Center' Fund (Grant No. B20063).

References

[1] Gnesotto F S, Mura F, Gladrow J and Broedersz C P 2018 *Rep. Prog. Phys.* **81** 066601

[2] Battle C, Broedersz C P, Fakhri N, Geyer V F, Howard J, Schmidt C F and MacKintosh F C 2016 *Science* **352** 604
[3] Zia R K P and Schmittmann B 2007 *J. Stat. Mech. Theory Exp.* **2007** P07012
[4] Seifert U 2012 *Rep. Prog. Phys.* **75** 126001
[5] Seifert U 2005 *Phys. Rev. Lett.* **95** 040602
[6] Kawai R, Parrondo J M R and den Broeck C V 2007 *Phys. Rev. Lett.* **98** 080602
[7] Lynn C W, Holmes C M, Bialek W and Schwab D J 2022 *Phys. Rev. Lett.* **129** 118101
[8] Lynn C W, Cornblath E J, Papadopoulos L, Bertolero M A and Bassett D S 2021 *Proc. Natl. Acad. Sci. USA* **118** e2109889118
[9] Sanz Perl Y, Bocaccio H, Pallavicini C, Pérez-Ipiña I, Laureys S, Laufs H, Kringelbach M, Deco G and Tagliazucchi E 2021 *Phys. Rev. E* **104** 014411
[10] Harris J J, Jolivet R and Attwell D 2012 *Neuron* **75** 762
[11] Yu L and Yu Y 2017 *J. Neurosci. Res.* **95** 2253
[12] Erecínska M and Silver I A 1989 *J. Cereb. Blood Flow Metab.* **9** 2
[13] Norberg K and Siejö B K 1975 *Brain Res.* **86** 45
[14] Du F, Zhu X H, Zhang Y, Friedman M, Zhang N, Uğurbil K and Chen W 2008 *Proc. Natl. Acad. Sci. USA* **105** 6409
[15] Deco G, Sanz Perl Y, de la Fuente L, Sitt J D, Yeo B T T, Tagliazucchi E and Kringelbach M L 2023 *Netw. Neurosci.* **7** 966
[16] Deco G, Sanz Perl Y, Bocaccio H, Tagliazucchi E and Kringelbach M L 2022 *Commun. Biol.* **5** 572
[17] de la Fuente L A, Zamberlan F, Bocaccio H, Kringelbach M, Deco G, Sanz Perl Y, Pallavicini C and Tagliazucchi E 2023 *Cereb. Cortex* **33** 1856
[18] Ibáñez A 2022 *Trends Cogn. Sci.* **26** 1031
[19] Zanin M, Güntekin B, Aktürk T, Hanoğlu L and Papo D 2019 *Front. Physiol.* **10** 1619
[20] Bassett D S and Sporns O 2017 *Nat. Neurosci.* **20** 353
[21] Lv G, Xu T, Chen F, Zhu P, Wang M and He G 2024 *Chin. Phys. B* **33** 028704
[22] Beggs J M 2022 *The Cortex and the Critical Point: Understanding the Power of Emergence* (Cambridge, MA: The MIT Press) pp. 51–71
[23] Beggs J M and Plenz D 2003 *J. Neurosci.* **23** 11167
[24] Haldeman C and Beggs J M 2005 *Phys. Rev. Lett.* **94** 058101
[25] Shew W L, Yang H, Petermann T, Roy R and Plenz D 2009 *J. Neurosci.* **29** 15595
[26] Beggs J M 2008 *Phil. Trans. R. Soc. A* **366** 329
[27] Wang R, Lin P, Liu M, Wu Y, Zhou T and Zhou C 2019 *Phys. Rev. Lett.* **123** 038301
[28] Song B, Ma N, Liu G, Zhang H, Yu L, Liu L and Zhang J 2019 *J. Neural Eng.* **16** 056002
[29] Haimovici A, Tagliazucchi E, Balenzuela P and Chialvo D R 2013 *Phys. Rev. Lett.* **110** 178101
[30] Zhou X, Ma N, Song B, Wu Z, Liu G, Liu L, Yu L and Feng J 2021 *Front. Comput. Neurosci.* **15** 641335
[31] Wang R, Liu M, Cheng X, Wu Y, Hildebrandt A and Zhou C 2021 *Proc. Natl. Acad. Sci. USA* **118** e2022288118
[32] Xu L, Feng J and Yu L 2022 *Hum. Brain Mapp.* **43** 2534
[33] Torres J J and Marro J 2015 *Sci. Rep.* **5** 12216
[34] Noa C E F, Harunari P E, de Oliveira M J and Fiore C E 2019 *Phys. Rev. E* **100** 012104
[35] Aguilera M, Igarashi M and Shimazaki H 2023 *Nat. Commun.* **14** 3685
[36] Zhang Y and Barato A C 2016 *J. Stat. Mech. Theory Exp.* **2016** 113207
[37] Huang X, Xu K, Chu C, Jiang T and Yu S 2017 *J. Neurosci.* **37** 10481
[38] Van Essen D C, Smith S M, Barch D M, Behrens T E J, Yacoub E, Uğurbil K and Consortium W U M H 2013 *NeuroImage* **80** 62
[39] Glasser M F, Smith S M, Marcus D S, Andersson J L, Auerbach E J, Behrens T E, Coalson T S, Harms M P, Jenkinson M, Moeller S, Robinson E C, Sotiropoulos S N, Xu J, Yacoub E, Uğurbil K and Van Essen D C 2016 *Nat. Neurosci.* **19** 1175
[40] Glasser M F, Sotiropoulos S N, Wilson J A, Coalson T S, Fischl B, Andersson J L, Xu J, Jbabdi S, Webster M, Polimeni J R, Van Essen D C, Jenkinson M and Consortium W U-M H 2013 *NeuroImage* **80** 105
[41] Smith S M, Nichols T E, Vidaurre D, Winkler A M, Behrens T E, Glasser M F, Uğurbil K, Barch D M, Van Essen D C and Miller K L 2015 *Nat. Neurosci.* **18** 1565
[42] Liu T T, Nalci A and Falahpour M 2017 *NeuroImage* **150** 213
[43] Schaefer A, Kong R, Gordon E M, Laumann T O, Zuo X N, Holmes A J, Eickhoff S B and Yeo B T T 2018 *Cereb. Cortex* **28** 3095
[44] Yeo B T, Krienen F M, Sepulcre J, Sabuncu M R, Lashkari D, Hollinshead M, Roffman J L, Smoller J W, Zöllei L, Polimeni J R, Fischl B, Liu H and Buckner R L 2011 *J. Neurophysiol.* **106** 1125
[45] Lynn C W, Holmes C M, Bialek W and Schwab D J 2022 *Phys. Rev. E* **106** 034102
[46] Ge H and Qian H 2013 *Phys. Rev. E* **87** 062125
[47] Palmer S E, Marre O, Berry M J and Bialek W 2015 *Proc. Natl. Acad. Sci. USA* **112** 6908
[48] Schneidman E, Berry M J, Segev R and Bialek W 2006 *Nature* **440** 1007
[49] Strong S P, Koberle R, de Ruyter van Steveninck R R and Bialek W 1998 *Phys. Rev. Lett.* **80** 197
[50] Tagliazucchi E, Balenzuela P, Fraiman D and Chialvo D R 2012 *Front. Physiol.* **3** 15
[51] Yang H, Shew W L, Roy R and Plenz D 2012 *J. Neurosci.* **32** 1061
[52] Zhang J, Wu K, Dong J, Feng J and Yu, L 2024 *Neural Netw.* **184** 107100
[53] Crochik L and Tomé T 2005 *Phys. Rev. E* **72** 057103
[54] Sterling P and Laughlin S 2015 *Principles of Neural Design* (Cambridge, MA: The MIT Press) pp. 57–104
[55] Balasubramanian V 2015 *Proc. IEEE Inst. Electr. Electron. Eng.* **103** 1346
[56] Tomé T and de Oliveira M J 2012 *Phys. Rev. Lett.* **108** 020601
[57] Lu W, Du X, Wang J, Zeng L, Ye L, Xiang S, Zheng Q, Zhang J, Xu N, Feng J and DTB Consortium 2024 *Nat. Comput. Sci.* **4** 890
[58] Kawazura Y and Yoshida Z 2010 *Phys. Rev. E* **82** 066403
[59] Niu W, Huang X, Xu K, Jiang T and Yu S 2019 *Neuroscience* **412** 190
[60] Zimmern V 2020 *Front. Neural Circuits* **14** 54