SCIENCE CHINA
Physics, Mechanics & Astronomy

• Article •

June 2024 Vol. 67 No. 6: 260511
https://doi.org/10.1007/s11433-024-2341-2

Energy-information trade-off induces continuous and discontinuous phase transitions in lateral predictive coding

Zhen-Ye Huang[1,2†], Ruyi Zhou[3†], Miao Huang[1,2†], and Hai-Jun Zhou[1,2,4*]

[1] *Key Laboratory for Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China;*
[2] *School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China;*
[3] *School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China;*
[4] *Minjiang Collaborative Center for Theoretical Physics, Minjiang University, Fuzhou 350108, China*

Received January 12, 2024; accepted February 4, 2024; published online May 7, 2024

Lateral predictive coding is a recurrent neural network that creates energy-efficient internal representations by exploiting statistical regularity in sensory inputs. Here, we analytically investigate the trade-off between information robustness and energy in a linear model of lateral predictive coding and numerically minimize a free energy quantity. We observed several phase transitions in the synaptic weight matrix, particularly a continuous transition that breaks reciprocity and permutation symmetry and builds cyclic dominance and a discontinuous transition with the associated sudden emergence of tight balance between excitatory and inhibitory interactions. The optimal network follows an ideal gas law over an extended temperature range and saturates the efficiency upper bound of energy use. These results provide theoretical insights into the emergence and evolution of complex internal models in predictive processing systems.

predictive coding, recurrent neural network, phase transition, internal model, free energy

PACS number(s): 87.19.L-, 89.70.-a, 87.15.Zg, 05.70.Ce, 89.75.Kd

Citation: Z.-Y. Huang, R. Zhou, M. Huang, and H.-J. Zhou, Energy-information trade-off induces continuous and discontinuous phase transitions in lateral predictive coding, Sci. China-Phys. Mech. Astron. **67**, 260511 (2024), https://doi.org/10.1007/s11433-024-2341-2

1 Introduction

Predictive coding is an influential theory in computational neuroscience with a long history tracing back to Kant and Helmholtz [1-4], and its recent conceptual advances include the Bayesian brain theory and the free energy principle of information processing and decision making [5-8]. Predictive coding theory regards the brain as a multilayered hierarchical neuron network that builds an internal model for the external world and employs it to perceive sensory signals and make predictions. To date, most studies on predictive coding have focused on feedforward and feedback interactions between different layers and have largely ignored lateral (horizontal, recurrent) interactions within the same layer of neurons. However, lateral interactions are abundant in biological neural networks and play important roles in perception and inference [9-13]. The brain consumes 50% of the body's total metabolic energy in children and 20% in adults [14]. As a large energy consumer, the brain must be capable of achieving energy-efficient information processing [15-19]. Lateral predictive coding (LPC) may be an elegant evolutionary solution to cope with the brain's tight energy budget.

One basic function of lateral interactions is to reduce the energy cost of internal representations. By adapting the

*Corresponding author (email: zhouhj@itp.ac.cn)
† These authors contributed equally to this work.

synaptic weights according to the input correlation between spatially adjacent neurons, the external drive to one neuron is partially canceled by the messages from its neighboring neurons, and its activity becomes weaker and sparser [20-23]. However, reducing the magnitude and redundancy of internal representation compromises its information content, making it less robust to external and internal noises [24-26]. How to balance the conflicting demands of energy reduction and information robustness is an important issue in LPC. Nonetheless, LPC has been rarely discussed in the physics community, in stark contrast to the large and enduring theoretical enthusiasm for other basic neural network models, such as the perceptron, Hopfield's model of associative memory, and restricted Boltzmann machine [27].

Here, we study phase transitions induced by the energy-information trade-off in the simplest model of linear LPC [28]. We introduce a temperature parameter T to incorporate the fitness effect of information robustness and define a free energy, which binds energy and function, as the minimization objective [29-31]. We find that even if the input signals are statistically symmetric among all N units, the optimal synaptic weight matrix will spontaneously break permutation (reciprocal) symmetry at a certain critical temperature and form cyclic-dominant interaction patterns. An ideal gas law $E = (N/2)T$ is then followed by the energy E of the multi-unit ($N \geq 3$) LPC networks, achieving upper-bound efficiency in energy use. At low temperatures, the weight matrix experiences several further phase transitions, one of which is a discontinuous transition to a tight balance of excitatory and inhibitory interactions. Internal symmetries within subgroups of units may also break, leading to multi-level functional differentiation.

Our theoretical results reveal the emergence of structural complexity in LPC systems. These collective properties of non-reciprocity, correlation reduction, and excitation-inhibition (EI) balance are qualitatively similar to the empirical observations of real-life recurrent neural networks. This simple theoretical model is helpful for understanding the qualitative features and organizational principles of the biological nervous system. As an initial attempt, we have left many important issues untouched, such as nonlinear synaptic interactions and nonquadratic energetic costs. More effort is needed from the physics community to fully appreciate LPC as a basic neural computing circuit.

2 Model and free energy

Consider a fully connected network of N units (each may be a single neuron or a small region of the brain). Denote an internal state vector as $\boldsymbol{x} = (x_1, \ldots, x_N)^\top$ and a sensory input as $\boldsymbol{s} = (s_1, \ldots, s_N)^\top$. Upon receiving an input $\boldsymbol{s}$, unit i responds by changing its state x_i according to linear LPC dynamics [20, 22, 28],

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = s_i - x_i - \sum_{j\neq i} w_{ij}x_j \ , \tag{1}$$

where positive (negative) synaptic weight w_{ij} means that unit j inhibits (excites) unit i. The steady-state encoding from sensory input to internal representation is $\boldsymbol{x_s} = (\boldsymbol{I} + \boldsymbol{W})^{-1}\boldsymbol{s}$, with $\boldsymbol{I}$ being the identity matrix and $\boldsymbol{W}$ the synaptic weight matrix. The convergence of the dynamics (1) requires that the real part of every eigenvalue of $(\boldsymbol{I} + \boldsymbol{W})$ be positive (we set a threshold value 10^{-5} and call it the eigenvalue bottom line). We interpret $\sum_{j\neq i} w_{ij}x_j$ as the predicted sensory input to unit i and x_i as the prediction error between the actual s_i and the predicted value [9]. Assuming that the energy cost of maintaining and transmitting a prediction error is quadratic in x_i, the mean energy E of an internal state is

$$E \equiv \sum_{\boldsymbol{s}} P(\boldsymbol{s})\boldsymbol{x_s}^2 = \mathrm{Tr}\Big[(\boldsymbol{I} + \boldsymbol{W})^{-1}\boldsymbol{C}(\boldsymbol{I} + \boldsymbol{W}^\top)^{-1}\Big], \tag{2}$$

where $P(\boldsymbol{s})$ is the probability distribution of sensory inputs, and $\boldsymbol{C}$ is the input correlation matrix with elements $c_{ij} \equiv \sum_{\boldsymbol{s}} P(\boldsymbol{s})s_i s_j$. The total energy cost of representing $\mathcal{M} \gg 1$ sensory inputs is then $\mathcal{M}E$.

An infinitesimal volume of the sensory space is transformed by $\boldsymbol{x_s}$ to an infinitesimal volume in the internal representation space with Jacobian $1/\mathrm{Det}(\boldsymbol{I} + \boldsymbol{W})$, which is the inverse determinant. The entropy of the internal states $\boldsymbol{x}$ is then

$$S = -\log[\mathrm{Det}(\boldsymbol{I} + \boldsymbol{W})] \tag{3}$$

plus a constant, and the mutual information between $\boldsymbol{s}$ and $\boldsymbol{x}$ is equal to this entropy up to a constant (see the derivation in Appendix A1). We can take $\mathcal{M}S$ as the total entropy of $\mathcal{M}$ internal states. A large entropy S is desirable so that the encoding will be sensitive to variations in $\boldsymbol{s}$ and be robust to noises in representing and transmitting $\boldsymbol{x}$ (the information maximization principle [29]).

Minimizing energy (cost) E and maximizing entropy (information) S are mutually conflicting goals. We introduce a temperature parameter T to quantify this energy-information trade-off. In the thermodynamic limit of an infinite number of sensory inputs ($\mathcal{M} \to \infty$), the optimization objective is the total free energy $\mathcal{M}F$, with the coefficient F being

$$F = E - TS \ , \tag{4}$$

which combines energetic and entropic effects [29-31]. The temperature embodies all external and internal fitness stresses. At high T values, information sensitivity and robustness are the main fitness drivers, and the adaptation of synaptic weights favors entropy S; at the other limit of low

T, reducing energy E becomes the dominant fitness concern, and the weight matrix evolves toward energy minimization. The thermodynamic equation accompanying free energy minimization is simply $T = \mathrm{d}E/\mathrm{d}S$.

We search for the global minimum of F at each fixed value of T using a slow simulated annealing process and determine the corresponding optimal weight matrix. The technical details are presented in Supplementary Information (these technical details could also be accessed from ref. [32]). We find that $E(T)$ and $S(T)$ are singular at several critical points of T, and the optimal weight matrix $\boldsymbol{W}$ changes its qualitative properties at these points. Discontinuities in energy susceptibility ($\mathrm{d}E/\mathrm{d}T$) imply continuous phase transitions, and discontinuities in energy itself manifest discontinuous phase transitions. To demonstrate most explicitly the endogenous nature of such phase transitions, in the following discussion we will focus on symmetric and homogeneous sensory inputs: the self-correlation $c_{ii} = 1$ for all inputs s_i and pair correlation $c_{ij} = c$ being identical ($c < 1$) for the inputs of any two different units. The default internal model $\boldsymbol{W}$ is then reciprocal ($w_{ij} = w_{ji}$) and permutation-symmetric (identical nondiagonal elements) [32].

The phase diagram for the simplest two-unit system (Figure 1) is a good starting point for understanding the optimal LPC. A continuous reciprocity-breaking phase transition occurs for $c < 2/3$, with w_{12} and w_{21} gradually deviating from each other at the critical temperature $T_2^{\mathrm{rb}} = (2-5c^2/2)/(1-c^2/4)^2$ [32]. When input correlation c exceeds $2/3$, a local minimum of the free energy with $w_{12} \neq w_{21}$ first emerges in the reciprocal ($w_{12} = w_{21}$) phase, and it then becomes the global minimum as T decreases to a certain critical value that is strictly higher than T_2^{rb}. A discontinuous transition then occurs from the reciprocal phase to a reciprocity-broken phase in which unit 2 inhibits unit 1 and unit 1 excites unit 2.

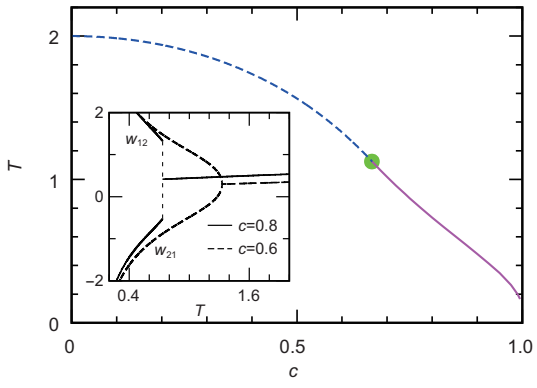

Figure 1 (Color online) Phase diagram for the $N = 2$ system. The symmetry-breaking transition to $w_{12} \neq w_{21}$ is continuous for $c < 2/3$ (blue dashed line) and discontinuous for $c > 2/3$ (red solid line). The green dot marks the tricritical point at $c = 2/3$ and $T = 9/8$. The inset shows the continuous transition for $c = 0.6$ and the discontinuous transition for $c = 0.8$.

3 Reciprocity-breaking and ideal gas law

For multi-unit systems containing $N \geq 3$ units, our numerical optimization results and local stability analysis reveal that the optimal internal model $\boldsymbol{W}$ is reciprocal ($w_{ij} = w_{ji}$) and permutation-symmetric (all w_{ij} being equal) at high temperatures. However, these properties break down spontaneously as T drops below the critical value [32]

$$T_N^{\mathrm{rb}} = 2\left(\sqrt{1+(N-1)c} + (N-1)\sqrt{1-c}\right)^2/N^2 . \tag{5}$$

However, eq. (5) does not apply for $N = 2$, indicating the emergence of new collective properties for $N \geq 3$ during this continuous reciprocity-breaking phase transition. Indeed, we find that the energy of the reciprocity-broken ($w_{ij} \neq w_{ji}$) optimal system exactly obeys the ideal gas law [32]

$$E = \frac{N}{2}T \quad (N \geq 3) \tag{6}$$

in an extended temperature interval (T_N^{iglb}, T_N^{rb}], with T_N^{iglb} being the lower-bound temperature at which this ideal gas law is deviated (Figure 2(a)).

One type of interaction pattern that realizes the ideal gas law is the rotation-symmetric solution with weight parameters w_i (we require $w_1 \leq \ldots \leq w_{N-1}$ to remove trivial degeneracies):

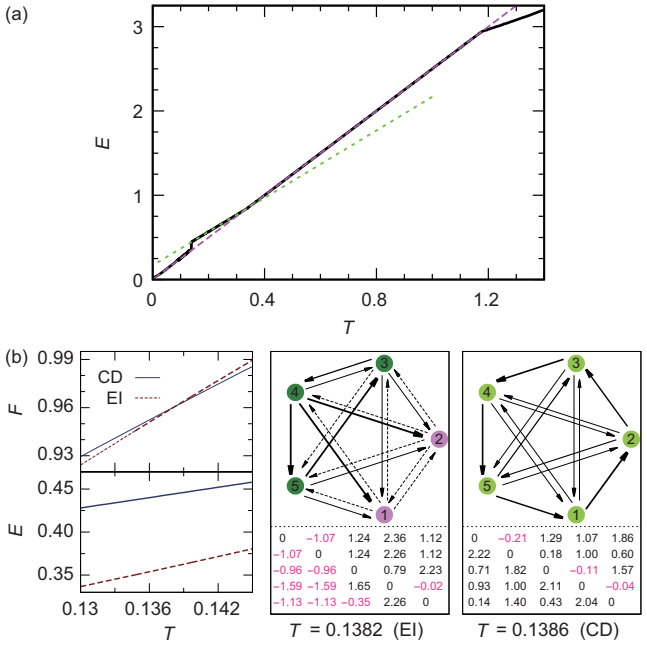

Figure 2 (Color online) Results for $N = 5$ and $c = 0.8$. (a) E versus T. The thin fitting lines have slopes of 2.5 and 2.0 (the critical temperatures $T_5^{\mathrm{rb}} = 1.1783$ and $T_5^{\mathrm{iglb}} = 0.3360$, respectively). (b) Examples of cyclic-dominant and EI-balanced optimal matrices at $T \approx 0.1383$, and $F(T)$ and $E(T)$ for these two branches of solutions. Solid (dashed) links in the interaction graphs indicate positive (negative) weights.

$$\boldsymbol{W}=\begin{bmatrix} 0 & w_1 & w_2 & \cdots & w_{N-1} \\ w_{N-1} & 0 & w_1 & \cdots & w_{N-2} \\ w_{N-2} & w_{N-1} & 0 & \cdots & w_{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_1 & w_2 & w_3 & \cdots & 0 \end{bmatrix}. \tag{7}$$

There is cyclic dominance (CD) among the N units, such that unit $i+1$ is most strongly inhibited by unit i and it most strongly inhibits unit $i+2$ and so on. Rotation-symmetric solutions are impossible for $N=2$, and there is only one such solution for $N=3,4$, but there are infinitely many degenerate solutions for $N\geq 5$ [32].

We can define a quantity $O_{\rm CD}$, which compares the most dominant global cycle of directed interactions with the reverse directed cycle of interactions, to measure the degree of cyclic dominance:

$$O_{\rm CD}=1-\min_{(p_1,\ldots,p_N)}\left|\frac{\sum_{i=1}^{N-1} w_{p_{i+1}\ p_i}}{\sum_{i=1}^{N-1} w_{p_i\ p_{i+1}}}\right|, \tag{8}$$

where $(p_1,\ldots,p_N)$ denotes a permutation of the N units. For example, the weight matrix shown in the right panel of Figure 2(b) has a very high value of $O_{\rm CD}=0.996$. Through numerical computations, we find that the ideal gas law can be achieved by cyclic-dominant weight matrices with or without rotational symmetry and also by two-component matrices composed of excitatory and inhibitory units [32].

What is the importance of the ideal gas law (6)? In terms of the eigenvalues ϵ_i of the symmetric matrix $(\boldsymbol{I}+\boldsymbol{W})^{-1}C(\boldsymbol{I}+\boldsymbol{W}^{\top})^{-1}$, we have $E=\sum_i \epsilon_i$ and $2S=\sum_i \ln\epsilon_i-\ln\mathrm{Det}(\boldsymbol{C})$. The entropy is then upper-bounded by $S\leq (N/2)\ln(E/N)-(1/2)\ln\mathrm{Det}(\boldsymbol{C})$ at each energy E, and the equality holds only if all the eigenvalues ϵ_i are identical (see Appendix A2). If this upper bound of entropy can be saturated within a continuous interval of E, then eq. (6) will be achieved due to energy-information competition, which means the most efficient use of energy to reach the upper bound of information robustness. An appealing statistical property is that the internal states are composed of independent components of equal magnitudes [20, 33], with $\langle x_i^2\rangle=E/N$ to achieve energy equipartition and $\langle x_i x_j\rangle=0$ for $j\neq i$ to eliminate internal pair correlations (here $\langle\cdot\rangle$ denotes averaging over all the samples). We notice a likely closely related biological fact that the neuron pair correlations of the visual cortex are very weak even for highly correlated sensory inputs [34].

As the temperature decreases to the lower critical value $T_N^{\rm iglb}$, it will no longer be possible for an optimal weight matrix to exactly obey eq. (6) and at the same time keep all its complex eigenvalues above the bottom line. The ideal gas law will then break down, and a kink of the energy function $E(T)$ will be observed. For $N=5$ and $c=0.8$, the energy slope changes from 2.5 to 2.0 at $T_5^{\rm iglb}=0.3360$ (Figure 2(a)). This continuous transition is not associated with any symmetry change.

4 Discontinuous phase transitions

The energy discontinuity at $T=0.1383$ in the example curve of Figure 2(a) signifies a discontinuous phase transition. Indeed, the free energy near this temperature has two minima, which are organized into two branches with distinct energies (Figure 2(b)). One branch corresponds to the one-component CD network, which is almost completely inhibitory. The other branch corresponds to a two-component EI-balanced network: The two excitatory units 1 and 2 of group g_E have permutation symmetry, while the three units 3, 4 and 5 of group g_I have cyclically dominant interactions within themselves and strongly inhibit group g_E. This pattern agrees with the vital biological fact of EI competition and balance [35-38].

A simple measure of EI balance is to compute the net input weights of the individual units,

$$O_{\rm EI}=\frac{1}{N}\sum_{i=1}^{N}\left[1-\frac{\left|\sum_{j\neq i} w_{ij}\right|}{\sum_{j\neq i}|w_{ij}|}\right]. \tag{9}$$

A value of $O_{\rm EI}$ substantially above zero indicates a highly EI-balanced network in which the excitatory inputs are largely canceled by the inhibitory inputs for most of the units. For example, $O_{\rm EI}=0.626$ for the weight matrix shown in the middle panel of Figure 2(b).

After the transition to the optimal EI-balanced network, the ideal gas law is then approximately (though not exactly) recovered at lower temperatures (Figure 2(a)), indicating that EI-balance is beneficial for energy usage efficiency. This benefit cannot be obtained if all synaptic weights are restricted to positive values (inhibitory) [32].

The phase diagram for the system containing $N=5$ units is shown in Figure 3. The β_1 and β_2 regions in this diagram correspond to one-component cyclic-dominant optimal networks (a subregion of β_1 also contains EI-balanced ones). The γ_1-γ_4 regions correspond to different types of EI-balanced optimal networks. The phase boundaries are precisely determined by the singularities of $E(T)$, and we can also define quantitative measures such as $O_{\rm CD}$ and $O_{\rm EI}$ to describe these different phases [32]. The ideal gas law is exact in regions β_1 and γ_1 but approximate in γ_2-γ_4. The discontinuous transition from γ_2 to γ_3 is associated with breaking the internal rotational symmetry of group g_I and leads to functional differentiation of the three units 3, 4 and 5 (see ref. [32] for more details). The phase diagram for the smallest multi-unit system ($N=3$) is qualitatively similar to Figure 3; and

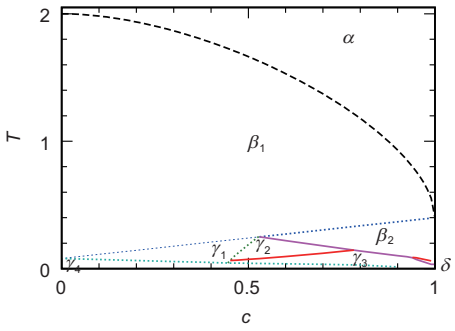

Figure 3 (Color online) Phase diagram for $N = 5$. The dashed line obeys eq. (5); the solid lines indicate discontinuous transitions, and the dotted lines mark continuous transitions attributed to the eigenvalue bottom line constraint. Region α, permutation-symmetric; β_1-β_2, cyclic dominant; γ_1-γ_4 (and part of β_1), EI-balanced; δ, an intermediate phase. β_1 and γ_1 obey the ideal gas law, eq. (6).

we also report numerical results for the larger system of $N = 10$ units to demonstrate the generality of our conclusions [32].

5 Discussion and conclusion

Our theoretical work attempted to consider energy and function simultaneously in a single model and revealed that the trade-off between energy reduction and information robustness can induce the spontaneous breaking of permutation symmetry in a lateral predictive coding system and drive the formation of cyclic dominance and the sudden emergence of excitation-inhibition balance among the network units. We discovered an ideal gas law (6) of internally representing the sensory inputs as equal-magnitude independent components. Our theoretical results qualitatively agree with some important biological facts, such as nonreciprocal interactions, excitation-inhibition balance, and weak internal pair correlations. These results may also have algorithmic implications.

For simplicity, we assumed that the input correlation matrix $\boldsymbol{C}$ is uniform with a single parameter c. A natural extension of this work is to study the ideal gas law (6) for heterogeneous sensory inputs. Our ongoing analytical and numerical computations confirm that eq. (6) and continuous and discontinuous phase transitions will also be observed for nonuniform matrices $\boldsymbol{C}$[1)] [32].

The existence of discontinuous phase transitions and the associated multiple free energy minima may render the optimal weight matrices difficult to acquire through local and gradual Hebbian learning processes [12, 39, 40]. This issue needs to be further explored. It is also very interesting to study phase transitions in nonlinear LPC models with energy cost being a sum $\sum_i |x_i|$ of absolute values and internal prediction being a rectified linear function $\max(0, \sum_{i\neq j} w_{ij}x_j)$ or another more complicated nonlinear function [41]. We expect to observe rich continuous and discontinuous phase transitions in these more realistic models. These nonlinear LPC systems may lead to enhanced sensitivity to fine details in typical input vectors.

This work was supported by the National Natural Science Foundation of China (Grant Nos. 12047503, 11747601 and 12247104), the National Innovation Institute of Defense Technology (Grant No. 22TQ0904ZT01025). The Numerical Simulations were carried out at the HPC cluster of ITP-CAS and also at the BSCC-A3 platform of the National Supercomputer Center in Beijing with the help of the TRNG random number generators [42].

Conflict of interest The authors declare that they have no conflict of interest.

Supporting Information

The supporting information is available online at http://phys.scichina.com and https://link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

1 L. R. Swanson, Front. Syst. Neurosci. **10**, (2016).
2 N. J. Wade, J. History Neurosci. **30**, 405 (2021).
3 H. B. Barlow, in *Possible principles underlying the transformations of sensory messages: Sensory Communication*, edited by W. A. Rosenblith (MIT Press, New York, 1961), pp. 217-234.
4 Y. Huang, and R. P. N. Rao, WIRES Cogn. Sci. **2**, 580 (2011).
5 C. S. Kim, Biol. Cybern. **115**, 87 (2021).
6 K. Friston, J. Kilner, and L. Harrison, J. Physiol.-Paris **100**, 70 (2006).
7 M. Aguilera, B. Millidge, A. Tschantz, and C. L. Buckley, Phys. Life Rev. **40**, 24 (2022), arXiv: 2105.11203.
8 V. Jirsa, and H. Sheheitli, J. Phys. Complex. **3**, 015007 (2022).
9 M. V. Srinivasan, S. B. Laughlin, and A. Dubs, Proc. R. Soc. Lond. B. **216**, 427 (1982).
10 H. Tang, M. Schrimpf, W. Lotter, C. Moerman, A. Paredes, J. Ortega Caro, W. Hardesty, D. Cox, and G. Kreiman, Proc. Natl. Acad. Sci. USA **115**, 8835 (2018), arXiv: 1706.02240.
11 Z. Pang, C. B. O'May, B. Choksi, and R. VanRullen, Neural Netw. **144**, 164 (2021).
12 F. Fumarola, B. Hein, and K. D. Miller, Phys. Rev. X **12**, 031024 (2022), arXiv: 2109.02048.
13 H. Zhao, Sci. China-Phys. Mech. Astron. **64**, 270511 (2021), arXiv:

1) M. Huang, Z.-Y. Huang, Y. Chen, and H.-J. Zhou, The ideal gas law of linear predictive coding networks (manuscript in preparation, 2024).

1905.08313.
14 D. D. Clarke, and L. Sokoloff, in *Circulation and energy metabolism in the brain: Basic Neurochemistry: Molecular, Cellular and Medical Aspects*, 6th ed., edited by G. J. Siegel, B. W. Agranoff, R. Wayne Albers, S. K. Fisher, and M. D. Uhler (Lippincott-Raven, Philadelphia, 1999), pp. 637-669.
15 P. Lennie, Curr. Biol. **13**, 493 (2003).
16 Y. Chen, S. Wang, C. C. Hilgetag, and C. Zhou, PLoS Comput. Biol. **9**, e1002937 (2013).
17 J. E. Niven, Curr. Opin. Neurobiol. **41**, 129 (2016).
18 L. Yu, and Y. Yu, J Neurosci. Res. **95**, 2253 (2017).
19 C. Metzner, M. E. Yamakou, D. Voelkl, A. Schilling, and P. Krauss, arXiv: 2301.12892.
20 C. Jutten, and J. Herault, Signal Process. **24**, 1 (1991).
21 B. A. Olshausen, and D. J. Field, Nature **381**, 607 (1996).
22 G. F. Harpur, and R. W. Prager, Network-Computat. Neural Syst. **7**, 277 (1996).
23 A. Ali, N. Ahmad, E. de Groot, M. A. Johannes van Gerven, and T. C. Kietzmann, Patterns **3**, 100639 (2022).
24 L. Yu, and L. Liu, Phys. Rev. E **89**, 032725 (2014), arXiv: 1308.4122.
25 Z. Padamsey, D. Katsanevaki, N. Dupuy, and N. L. Rochefort, Neuron **110**, 280 (2022).
26 L. Weninger, P. Srivastava, D. Zhou, J. Z. Kim, E. J. Cornblath, M. A. Bertolero, U. Habel, D. Merhof, and D. S. Bassett, Phys. Rev. E **106**, 014401 (2022), arXiv: 2110.13781.
27 H. Huang, *Statistical Mechanics of Neural Networks* (Higher Education Press, Beijing, 2022).
28 Z. Y. Huang, X. Y. Fan, J. Zhou, and H. J. Zhou, Commun. Theor. Phys. **74**, 095601 (2022), arXiv: 2207.09047.
29 A. J. Bell, and T. J. Sejnowski, Neural Computat. **7**, 1129 (1995).
30 N. Tishby, F. C. Pereira, and W. Bialek, in *The information bottleneck method: Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing* (IEEE Press, Piscataway, 1999), pp. 368-377.
31 B. He, and Y. Wang, Chin. Phys. B **26**, 030506 (2017).
32 Z.-Y. Huang, R. Zhou, M. Huang, and H.-J. Zhou, arXiv: 2302.11681.
33 A. Hyvärinen, and E. Oja, Neural Netw. **13**, 411 (2000).
34 E. Schneidman, M. J. Berry Ii, R. Segev, and W. Bialek, Nature **440**, 1007 (2006), arXiv: q-bio/0512013.
35 C. van Vreeswijk, and H. Sompolinsky, Science **274**, 1724 (1996).
36 D. P. Yang, H. J. Zhou, and C. Zhou, PLoS Comput. Biol. **13**, e1005384 (2017).
37 L. Yu, Z. Shen, C. Wang, and Y. Yu, Front. Cell. Neurosci. **12**, 123 (2018).
38 J. Liang, S. J. Wang, and C. Zhou, Natl. Sci. Rev. **9**, nwab102 (2022).
39 F. J. Pineda, Phys. Rev. Lett. **59**, 2229 (1987).
40 M. W. Spratling, Brain Cognition **112**, 92 (2017).
41 C. Curto, J. Geneson, and K. Morrison, Neural Computat. **31**, 94 (2019).
42 H. Bauke, and S. Mertens, Phys. Rev. E **75**, 066701 (2007), arXiv: cond-mat/0609584.

Appendix

A1 On the definition of entropy

Suppose there is noise in the mapping from the sensory input $\boldsymbol{s} = (s_1, \ldots, s_N)^\top$ to the internal representation $\boldsymbol{x} = (x_1, \ldots, x_N)^\top$,

$$x_i = g_i(\boldsymbol{s}) + \xi_i \,, \tag{a1}$$

where ξ_i for $i = 1, \ldots, N$ are independent Gaussian random variables with mean zero and variance σ^2. For the linear LPC system, the deterministic function $g_i(\boldsymbol{s})$ is

$$g_i(\boldsymbol{s}) = \sum_{j=1}^{N} \Big(\frac{\boldsymbol{I}}{\boldsymbol{I}+\boldsymbol{W}} \Big)_{ij} s_j \,, \tag{a2}$$

where $\boldsymbol{I}$ is the identity matrix, $I_{ii} = 1$ and $I_{ij} = 0$ for $i \neq j$. The mutual information between $\boldsymbol{x}$ and $\boldsymbol{s}$ is

$$\mathcal{I}[\,\boldsymbol{x}; \boldsymbol{s}\,] = H[\boldsymbol{x}] - H[\,\boldsymbol{x}|\boldsymbol{s}\,] \,, \tag{a3}$$

where $H[\boldsymbol{x}]$ and $H[\boldsymbol{s}]$ are the entropies of the internal states $\boldsymbol{x}$ and sensory inputs $\boldsymbol{s}$, respectively, and $H[\,\boldsymbol{x}|\boldsymbol{s}\,]$ is the conditional entropy of $\boldsymbol{x}$ given $\boldsymbol{s}$. Notice that $H[\boldsymbol{s}]$ depends only on the probability distribution $P(\boldsymbol{s})$ of the sensory inputs $\boldsymbol{s}$, so it can be considered a constant. Conditional entropy $H[\,\boldsymbol{x}|\boldsymbol{s}\,]$ only depends on the property of the noise vector $(\xi_1, \ldots, \xi_N)^\top$, and so it can also be considered a constant. Therefore, we have $\mathcal{I}[\,\boldsymbol{x}; \boldsymbol{s}\,] = H[\boldsymbol{x}] + \mathcal{I}_0$, with $\mathcal{I}_0$ being a constant. To increase this mutual information, we should increase the entropy $H[\boldsymbol{x}]$ of internal states $\boldsymbol{x}$.

The marginal distribution $Q(\boldsymbol{x})$ of internal states $\boldsymbol{x}$ is

$$\begin{aligned} Q(\boldsymbol{x}) &= \int \mathrm{d}\boldsymbol{s} P(\boldsymbol{s}) \prod_{i=1}^{N} \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - g_i(\boldsymbol{s}))^2}{2\sigma^2} \right) \right] \\ &\approx |\mathrm{Det}\,(\boldsymbol{I}+\boldsymbol{W})| \;\; P((\boldsymbol{I}+\boldsymbol{W})\boldsymbol{x}) \,. \end{aligned} \tag{a4}$$

Eq. (a4) is valid for small noises (σ^2 being sufficiently small). The entropy of $\boldsymbol{x}$ is then

$$\begin{aligned} H[\boldsymbol{x}] &\equiv -\int \mathrm{d}\boldsymbol{x} Q(\boldsymbol{x}) \ln Q(\boldsymbol{x}) \\ &= -\ln\Big|\mathrm{Det}(\boldsymbol{I}+\boldsymbol{W})\Big| + H[\boldsymbol{s}]. \end{aligned} \tag{a5}$$

Again because $H[\boldsymbol{s}]$ does not depend on the weight matrix $\boldsymbol{W}$, we see that $H[\boldsymbol{x}] = -\ln\Big|\mathrm{Det}(\boldsymbol{I}+\boldsymbol{W})\Big| + H_0$, with H_0 being a constant. For our LPC problem, the real parts of the eigenvalues of $\boldsymbol{I}+\boldsymbol{W}$ must be positive. The determinant of $\boldsymbol{I}+\boldsymbol{W}$ then must be positive. This requirement explains the entropy formula (3).

A2 On the ideal gas law

Let us define an auxiliary symmetric matrix $\boldsymbol{M}$ for a general weight matrix $\boldsymbol{W}$ as:

$$\boldsymbol{M} = \frac{\boldsymbol{I}}{\boldsymbol{I}+\boldsymbol{W}} \boldsymbol{C} \left(\frac{\boldsymbol{I}}{\boldsymbol{I}+\boldsymbol{W}} \right)^\top \,, \tag{a6}$$

and denote the N nonnegative eigenvalues of this matrix as $\epsilon_1, \ldots, \epsilon_N$. Then, the energy E and entropy S of the LPC system can be expressed as:

$$E = \sum_i \epsilon_i, \tag{a7}$$

$$S = \frac{1}{2}\sum_i \ln \epsilon_i - \frac{1}{2}\ln \mathrm{Det}(\boldsymbol{C}). \tag{a8}$$

Applying the sum-product inequality

$$\frac{1}{N}\sum_{i=1}^{N} \epsilon_i \geq \Big(\prod_i \epsilon_i\Big)^{\frac{1}{N}}, \tag{a9}$$

we obtain the following upper bound for entropy S as a function of energy E:

$$S \leq \frac{N}{2}\ln\Big(\frac{E}{N}\Big) - \frac{1}{2}\ln \mathrm{Det}(\boldsymbol{C})\ . \tag{a10}$$

This upper bound of S is attained if and only if all the eigenvalues ϵ_i take the same value E/N. If this is the case for a continuous range of E, then by applying $T = \frac{\mathrm{d}E}{\mathrm{d}S}$ as required by the minimization of the free energy F, we obtain the ideal gas law $E = \frac{N}{2}T$.

From the ideal gas law, we can easily derive that

$$\boldsymbol{C} = \frac{T}{2}(\boldsymbol{I}+\boldsymbol{W})(\boldsymbol{I}+\boldsymbol{W})^{\top}\ , \tag{a11}$$

which means that the general form of $\boldsymbol{I}+\boldsymbol{W}$ is

$$\boldsymbol{I}+\boldsymbol{W} \ = \ \sqrt{\frac{2}{T}}\boldsymbol{C}^{1/2}\boldsymbol{U}\ , \tag{a12}$$

where $\boldsymbol{U}$ is a certain real orthogonal matrix satisfying $\boldsymbol{U}\boldsymbol{U}^{\top} = \boldsymbol{U}^{\top}\boldsymbol{U} = \boldsymbol{I}$. When all nondiagonal elements of the correlation matrix $\boldsymbol{C}$ take the same value c, the square root of $\boldsymbol{C}$ is expressed as:

$$\boldsymbol{C}^{1/2} = \sqrt{1-c}\,\boldsymbol{I} + \frac{\big(\sqrt{1+(N-1)c} - \sqrt{1-c}\big)}{N}\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix}\begin{bmatrix}1 & 1 & \ldots & 1\end{bmatrix}. \tag{a13}$$

Because the diagonal elements of $\boldsymbol{W}$ are zero, we see from eq. (a12) that each column of the orthogonal matrix $\boldsymbol{U}$ must satisfy the additional condition of

$$\sum_{j=1}^{N}(\boldsymbol{C}^{1/2})_{ij}U_{ji} \ = \ \sqrt{\frac{T}{2}}, \quad (i = 1,\ldots,N). \tag{a14}$$

If the temperature T is too large, these N conditions (a14) cannot be satisfied by any orthogonal matrix $\boldsymbol{U}$. When T is lowered to the critical value T_N^{rb} as given by eq. (5), we find that these conditions are uniquely satisfied by the orthogonal matrix $\boldsymbol{U} = \boldsymbol{I}$. At this critical value T_N^{rb}, the matrix $\boldsymbol{W}$ is still permutation-symmetric, and all its nondiagonal elements are

$$w_{ij} = \frac{\sqrt{1+(N-1)c} - \sqrt{1-c}}{\sqrt{1+(N-1)c} + (N-1)\sqrt{1-c}}\ . \tag{a15}$$

For the ideal gas law to hold at $T < T_N^{\mathrm{rb}}$, the constraints (a14) require that the i-th column vector of $\boldsymbol{U}$ should be staying on an $(N-1)$-dimensional ellipsoid of an N-dimensional hypersphere of unit radius. These N constraints are compatible with the constraint of $\boldsymbol{U}$ being an orthogonal matrix, as long as $N \geq 3$ (details are given in ref. [32]).