A Letters Journal Exploring the Frontiers of Physics

May 2019

EPL, **126** (2019) 38005 www.epljournal.org
doi: 10.1209/0295-5075/126/38005

A cautionary tale of entropic criteria in assessing the validity of the maximum entropy principle

Zhi-Qin John Xu[1(a)], Fang Xu[2], Guoqiang Bi[2], Douglas Zhou[3(b)] and David Cai[1,3,4†]

[1] *NYUAD Institute, New York University Abu Dhabi - Abu Dhabi, United Arab Emirates*
[2] *CAS Key Laboratory of Brain Function and Disease, CAS Center for Excellence in Brain Science and Intelligence Technology, and School of Life Sciences, University of Science and Technology of China - Anhui, China*
[3] *School of Mathematical Sciences, MOE-LSC and Institute of Natural Sciences, Shanghai Jiao Tong University Shanghai, PRC*
[4] *Courant Institute of Mathematical Sciences and Center for Neural Science, New York University New York, NY, USA*

received 8 November 2018; accepted in final form 16 May 2019
published online 18 June 2019

PACS 89.70.Cf – Entropy and other measures of information
PACS 87.19.lo – Information theory
PACS 87.19.ls – Encoding, decoding, and transformation

Abstract – The maximum entropy principle (MEP) has been applied to study various problems in equilibrium and nonequilibrium systems in physics and other disciplines. Through analyses of numerical and experimental data, we demonstrate that the widely used entropic criteria, an assessment of the validity of MEP, can be misleading indexes as they can often fail to reflect the important difference between the observed and the MEP predicted statistical distribution. Our work demonstrates the importance of high-order statistical structures that cannot be captured by the entropic criteria and provides a cautionary tale of over-interpretation of results of MEP.

This paper is dedicated to David Cai.

Many systems in science, engineering and sociology can be described as complex networks in which each node is considered to be one of two possible states, *e.g.*, increasing or decreasing of stock prices in market systems, susceptible or infected in epidemic spreading of a disease in human society, up or down of spins in ferromagnetic systems, and firing or quiescence of neurons in neuronal systems. To characterize the emergence of network collective behavior of these nodes, the ideas from statistical physics, *e.g.*, the maximum entropy principle (MEP), have been successfully applied to many disciplines, *e.g.*, complex networks [1], neuroscience [2–7], stock market [8], and functional magnetic resonance imaging [9]. These applications [2–4,8–13] have lent credence that the second-order MEP probability distribution, which is obtained by maximizing the entropy subject to the measured constraints of the means and the second-order correlations, succeeds in capturing 90%–99% of information (as measured by multi-information fraction (MIF)) in various systems. The information is defined as follows. Entropy indicates the uncertainty in knowledge about the state of a system. Given more constraints, *e.g.*, correlations, the uncertainty would decrease. Here the information captured by correlations is the decrement of entropy induced by the measured constraint of correlations under MEP [2–4]. The MIF is defined as the ratio between the information of the second-order correlations and all-order correlations (exact definitions can be found in eqs. (2), (3)). Intuitively, a MIF of 90% indicates that the uncertainty reduced by correlations higher than the second-order ones is less than 10% of the uncertainty reduced by the second-order correlations. However, as a coarse-grained characterization of probability distributions, a MIF of 90% does not directly measure how close is the second-order MEP distribution to the observed distribution.

The good performance of the second-order MEP model implies the universal low-dimensional, *i.e.*, up to the

(a) E-mail: zhiqinxu@nyu.edu
(b) E-mail: zdz@sjtu.edu.cn
† Deceased.

second-order, statistical correlation structure in these systems. However, the second-order MEP (pairwise sufficiency) could be insufficient in experimental data. For example, empirical studies found that the second-order MEP distribution is no longer sufficient for networks over fine-scales (*i.e.*, local clusters of neurons within $300\,\mu$m [14]). A series of studies investigated this pairwise sufficiency [12,15–18], such as how the pairwise sufficiency depends on the inputs [12,16]. Meanwhile, methods have also been developed to account for high-order effective interactions in the MEP analysis when the pairwise sufficiency fails [19–22], such as adaptively identifying the specific pairwise and higher-order moments in the MEP model [22] and using generalized entropies to incorporate correlations [23].

The MIF has been a widely used quantity to assess the performance of the second-order MEP analysis. For example, many experiments show that MIF is more than 90% in many neuronal networks [2–4,10] *e.g.*, vertebrate retina [2], in the stocks orientation data [8] and fMRI (functional magnetic resonance imaging) data. In addition, with MIF, previous works show that the effect of high-order interactions could be important in MEP analysis [11,12]. MIF is also used to study the sensitivity of MEP analysis with respect to various factors, such as common input [3], sampling bin size [4], and temporal correlation [10]. A perturbation analysis shows how the MIF depends on the network size in a regime in which the activities of different neurons in the network are approximately statistically independent [15].

Despite wide applications of the MEP analysis and the entropic criteria using MIF, there are several important issues that remain to be clarified. First, entropy is a coarse-grained characterization of uncertainty about the underlying distribution. It is yet to demonstrate how well the entropy decrement, described by MIF, can measure the discrepancy of fine structures in the estimated distribution with respect to the observed one. Second, high-order structures are often significant to generate synchronization dynamics [14,16], which is believed to be important for information processing in the brain [24]. It is yet to demonstrate whether high values of MIF can be construed as irrelevance of high-order structures.

In this work, we investigate the validity of entropic criteria in assessing the performance of the second-order MEP model by analyzing data from both constructed model systems and physiological experiments. We demonstrate that the widely used MIF [2–4,8–13,15] can be a misleading index as a high MIF can still often fail to reflect the important difference between the observed distribution of network states and the distribution predicted by the MEP model. The MIF can illude one to overestimate the power of MEP results because a high MIF has been often erroneously construed as irrelevance of high-order structures, notwithstanding their importance for capturing the true distribution of network states. Our numerical and experimental analyses indicate that instead of using MIF, it is better to assess the performance of the MEP analysis by comparing the probability distribution observed in the experiment to that predicted by the MEP model.

To begin with, we consider a network of n nodes and record its state in each sampling time bin. The state of the i-th node σ_i is either 0 or 1, indicating a silent or an active state. For all n recorded nodes, the state in each sampling time bin can be described by a binary vector $\Sigma = (\sigma_1, \cdots, \sigma_n) \in \{0,1\}^n$. By maximizing the entropy subject to measured correlations up to the m-th order ($m \leq n$) (see the subsect. "Maximum entropy principle (MEP) analysis" in the following), the m-th order MEP probability distribution can be obtained [25] in eq. (1):

$$P_m(\Sigma) = \frac{1}{Z} \exp\left(\sum_{i=1}^{n} J_i \sigma_i + \sum_{i<j}^{n} J_{ij}\sigma_i\sigma_j + \cdots + \sum_{i_1<\cdots<i_m}^{n} J_{i_1,\cdots,i_m}\sigma_{i_1}\cdots\sigma_{i_m} \right), \quad (1)$$

where Z is the normalization factor and $J_{i_1,\cdots,i_k}$ is the k-th order effective interaction ($1 \leq k \leq m$), which can be estimated by an iterative scaling algorithm [4,26]. The corresponding entropy for P_m is $S_m = -\sum_\Sigma P_m(\Sigma)\log P_m(\Sigma)$. Note that the highest-order MEP probability distribution is $P_n(\Sigma)$, which will be referred to as the full MEP distribution. One can show that $P_n(\Sigma)$ is identical to the experimentally observed distribution $P_{\rm obs}(\Sigma)$ for any recording time [27] because P_n is subject to all-order moments of $P_{\rm obs}(\Sigma)$.

The entropy decreases with the number of constrained correlations increasing [2,4], *i.e.*, $S_1 \geq S_2 \geq \cdots \geq S_n = S$, where S is the entropy of $P_{\rm obs}(\Sigma)$. The multi-information I_n is defined by $I_n = S_1 - S_n$, the amount of information accounted for by the second-order correlations is $I_2 = S_1 - S_2$ and the MIF [2,4] f is defined by

$$f = I_2/I_n. \quad (2)$$

Another closely related criterion [3,4] using the Kullback-Leibler divergence is

$$g = 1 - D_2/D_1, \quad (3)$$

where $D_m = \sum P_n \log(P_n/P_m)$ for $m = 1, 2$. Our results show that the conclusions using f and g are similar. For brevity, we only show the results using f below. Note that f and g are defined by ratios and not directly related to the absolute value of entropy or Kullback-Leibler divergence. The MIF is commonly regarded as the quantification of how much information is captured by the second-order correlations. As mentioned above, the MIF is found to be in the range of 90%–99% for states of various types of data [2–4,8–13,15].

We now discuss whether the MIF can be used as a criterion to assess the validity of the second-order MEP model. The following numerical examples of the full MEP distribution P_n are constructed using eq. (1) under the homogeneity condition by assuming an identical interaction strength for interactions of the same order. Therefore,

there are only $n+1$ possible probability intensities for 2^n states of n nodes. We denote $J^{(k)}$ as the effective interaction strength of the k-th order, $1 \le k \le n$. Figure 1(a) illustrates the result that in the representation of $\{0,1\}$ for binary states, the MIF can be very high for a distribution of 7 nodes with a vanishing interaction of the first two low orders but with nonvanishing high-order interactions of the third and fourth-order interactions only —$J^{(1)} = 0$, $J^{(2)} = 0$, but $J^{(3)} = 0.5$, $J^{(4)} = -0.2857$ while the strength of interaction of any-order higher than four here is set to zero. The MIF for this case is as high as 92%, notwithstanding the only nonvanishing orders are the third and fourth orders. With vanishing interactions of the first and second orders, the MEP probability distribution P_2 thereby fails to capture the full MEP distribution P_n for $n = 10$ (fig. 1(a)). Illustrated in fig. 1(b) is a case with an almost perfect performance of the second-order MEP probability to represent the full MEP distribution but with an extremely low MIF (12%) in a nearly independent system —$J^{(1)} = -5$, $J^{(2)} = 0.001$, $J^{(3)} = 0.1$ while the strength of the effective interaction of any order higher than three is set to zero for $n = 5$. Despite such a low MIF, surprisingly, the first- and the second-order MEP probability distributions are both in excellent agreement with the full MEP P_n.

We further study the functional dependence of MIF on the effective interaction strength. It turns out that it exhibits a rather complex behavior. An example is displayed in fig. 1(c), which shows the MIF as a function of the third-order interaction $J^{(3)}$ in the MEP distribution of seven nodes with $J^{(1)} = 0$, $J^{(2)} = 0.05$, $J^{(k)} \equiv 0$ for $4 \le k \le 7$. Clearly, the behavior of MIF is not a simple function of $J^{(3)}$, neither monotonically increasing nor decreasing, with its value varying widely from nearly 0 to 1. As $J^{(3)}$ becomes dominating over the low-order interactions by more than one order of magnitude in strength, it results in a poor characterization of P_n by P_2 but still with a relatively high MIF (greater than 90%). For example, $J^{(3)} = -1.9$ (fig. 1(d)), the MIF is 91.5%, whereas P_2 is not close to the full MEP distribution P_n at all. The above examples demonstrate that the entropic difference as characterized by MIF is unable to quantify the discrepancy between the second-order MEP probability distribution and the full MEP distribution as observed in the experiment.

Note that in the first case in fig. 1(a), if we represent the binary states by $\{-1,1\}$, then, the strength for interactions of the same order is still the same with the first order as 0.75, the second order as 0, the third order as -0.0625 and the fourth order as -0.0179. Therefore, although there is no pairwise interaction in both representations of $\{0,1\}$ and $\{-1,1\}$, we still have a high MIF, 0.92. In addition, to emphasize the similarity between indexes f in eq. (2) and g in eq. (3), we list their values for the first two examples. For fig. 1(a), $f = g = 0.920253$. For fig. 1(b), $f = 0.119343$ and $g = 0.119519$.

Next, we use data from our physiological experiments with calcium imaging [28] (see the subsect. "Calcium Imaging experiment" in the following) to show that high-order structures can be crucial to determine the distribution of network states even if the MIF is greater than 90%. The sampling time of spike trains maintains the tolerance ϵ in our entropy estimation within 2.5%. For any recorded neuron, *e.g.*, the i-th neuron, if it fires at least once during the sampling time bin, the neuronal state σ_i is set to be 1, otherwise it is set to 0. To avoid multiple firing in a time bin, the bin size is often chosen to be ∼10 ms in experiments [2–4]. We perform the MEP analysis for five selected neurons recorded in experiment using different bin sizes ranging from 4 ms to 27 ms. The MIF for all the bin sizes is greater than 89%, and can be even greater than 95% for a broad range of the bin size, as is displayed by the solid black line in fig. 2(a). In addition, we discover that the MIF depends sensitively on the size of the neuronal group analyzed. Figure 2(a) displays the MIF of a group of three neurons chosen from the above five neurons. For all bin sizes, the MIF decreases in general as the group size increases. The effect of the group size is consistent with the theoretical study in ref. [15] and the empirical study in ref. [20] that the performance of the second-order MEP analysis in a small-size network could be very different from that in a large-size network.

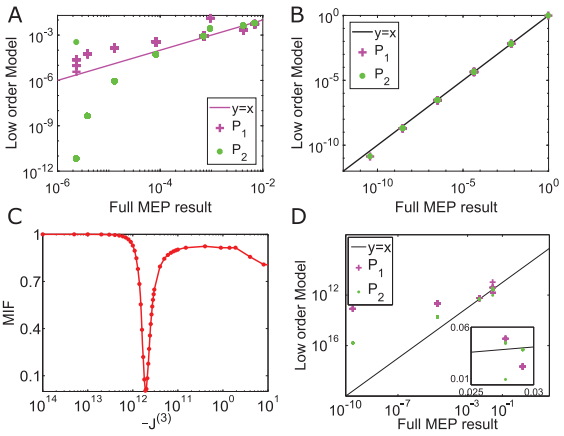

Fig. 1: MEP analysis using numerically constructed probability distribution. As in the text, $J^{(1)}$, $J^{(2)}$, $J^{(3)}$, $J^{(4)}$ are the first-, second-, third- and fourth-order effective interaction strength, respectively. For (a), (b) and (d), the MEP probability P_2 (green), and P_1 (magenta) of firing states is shown against the full MEP distribution P_n. Each data point here corresponds to a particular network state. The case in (a) has the MIF of 92% for 10 nodes. $J^{(1)} = 0$, $J^{(2)} = 0$, $J^{(3)} = 0.5$, $J^{(4)} = -0.2857$, $J^{(k)} \equiv 0$ for $5 \le k \le 10$. Points with probability smaller than 10^{-20} are not plotted. The case in (b) has the MIF of 12% for 5 nodes. $J^{(1)} = -5$, $J^{(2)} = 0.001$, $J^{(3)} = 0.1$, $J^{(k)} \equiv 0$ for $4 \le k \le 5$. The constructed MEP probabilities P_1 and P_2 using these J's overlap each other rather well. (c) MIF *vs.* the third-order effective interaction. Here we use 7 nodes, $J^{(1)} = 0$, $J^{(2)} = 0.05$, $J^{(k)} \equiv 0$ for $4 \le k \le 7$. The case in (d) is one of cases in (c) for $J^{(3)} = -1.9$. The MIF is 91.5%. Points with probability smaller than 10^{-10} are not plotted. The inset zooms in to a region of the comparison between the full MEP distribution P_n and P_2.

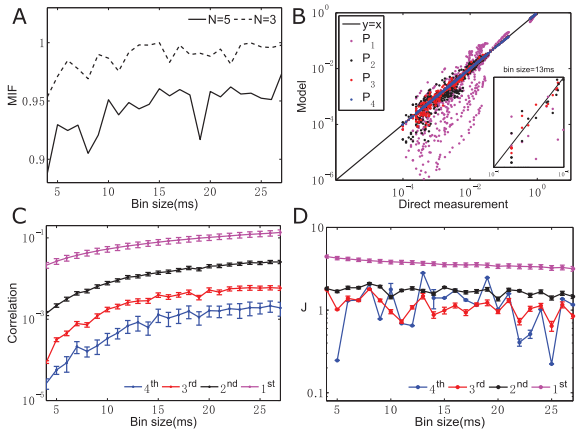

Fig. 2: MEP analysis using experimental data measured by calcium imaging (see the subsect. "Calcium Imaging experiment" in the following). Data from five neurons are collected to perform the MEP analysis. The firing rates are 2.8 Hz, 3.0 Hz, 3.9 Hz, 6.0 Hz, 7.2 Hz for these neurons, respectively. The total recording time is 45 s. (a) The MIFs for the neuronal group of size $N = 3$ and $N = 5$. The black dashed and solid lines are the MIFs for $N = 3$ and $N = 5$, respectively. (b) The predicted MEP probability distribution P_4 (blue), P_3 (red) P_2 (black), and P_1 (magenta) of each firing state against the probability directly measured from experiment. Each data point here corresponds to a particular neuronal firing pattern. Plotted is the data of all bin sizes aggregated. The inset is for the case of bin size 13 ms, for which the MIF is 94%. (c) Correlation strength *vs.* sampling time bin size. The data point is the mean absolute value of correlation strengths of the corresponding order for these five neurons, means (magenta), the second-order correlations (black), the third-order correlations (red) and the fourth-order correlations (blue). The standard deviation is indicated by the error bar. (d) Effective interaction strength *vs.* sampling time bin size. The data point is the mean absolute value of effective interaction strengths of the corresponding order for different sampling time bin sizes, the first order (magenta), the second order (black), the third order (red), and the fourth order (blue). The standard deviation is indicated by the error bar.

Therefore, great care should be taken to extrapolate MEP results from a small group of neurons to a large group [2]. Illustrated in fig. 2(b) is the MEP probability of each firing pattern predicted from the MEP analysis P_4, P_3, P_2, and P_1 of all bin sizes plotted against the probability $P_{\rm obs}(\Sigma)$ measured from the experiment. The second-order MEP probability P_2 deviates significantly from $P_{\rm obs}(\Sigma)$ for all bin sizes. The case in the inset of fig. 2(b), which displays only data of the bin size of 13 ms, emphasizes the fact that a P_2 of a MIF as high as 94% can still strongly deviate from the observed distribution.

As observed in experiments [2,29], most pairs of neurons are weakly correlated while the neuronal population as a whole is strongly correlated. To understand the relation between the correlation and the performance of the second-order MEP model, we next examine the role of the correlation strengths in the MEP probability reconstruction. Here, the k-th–order correlation is the expectation of $(\sigma_{i_1}-\bar{\sigma}_{i_1})\cdots(\sigma_{i_k}-\bar{\sigma}_{i_k})$ for neurons with respect to $P_{\rm obs}(\Sigma)$ of neuronal firing patterns, with $\bar{\sigma}_{i_j}$ being the mean value of σ_{i_j}. Even though the high-order correlations are nearly one order of magnitude smaller than the first- and the second-order correlations (fig. 2(c)), the second-order MEP probability distribution can be in drastic disagreement with the experimental measurement (fig. 2(b)). The case of the bin size of 13 ms, which has high-order correlations nearly one order of magnitude weaker than the first- and the second-order correlations, exemplifies the failure of the second-order MEP probability P_2 (black in the inset of fig. 2(b) to capture the observed distribution. Consequently, weak high-order correlations do not underpin the success of low-order MEP probabilities to represent the observed probability $P_{\rm obs}(\Sigma)$. We can use the following example to intuitively understand the effect of weak high-order correlation. Consider a high-synchronization network with high firing rates. The mean of a neuron's state, $\bar{\sigma}_{i_1}$, is close to 1. Then, $(\sigma_{i_1}-\bar{\sigma}_{i_1})$ is close to zero. The k-th–order correlation is close to $(\sigma_{i_1}-\bar{\sigma}_{i_1})^k$. In such case, higher-order correlations would be much smaller than low-order correlations. However, the high-order effective interactions could be comparable to low-order ones. Thus, a low-order MEP analysis cannot capture the observed probability distribution. A more rigorous study of this correlation effect can be found in ref. [5]. Therefore, an over-interpretation of a high MIF as a signifier to downplay the importance of high-order correlations (even though they can be weak) can cause an underestimation of the significant impact of high-order structures in neuronal coding.

Finally, we address the role of effective interactions in the MEP probability reconstruction. First, we note that strengths of all effective interactions in the full MEP distribution can be obtained using the observed distribution $P_{\rm obs}(\Sigma)$ (see the subsect. "The interactions in the n-th order MEP analysis" in the following). We display in fig. 2(d) the mean of the absolute values of effective interaction strengths of different orders in the full MEP distribution against different bin sizes for the five selected analyzed neurons above. For all bin sizes, it can be clearly seen that the third- and the fourth-order effective interactions are comparable to the first- and the second-order interactions. Note that the second-order MEP analysis takes into account the first two order effective interactions only while ignoring high-order structures. A theoretical analysis of the difference between the correlation and effective interaction can be found in ref. [5]. Therefore, a high MIF (as demonstrated in fig. 2(a)) or weak high-order correlation (as demonstrated in fig. 2(c)) entails neither a second-order MEP probability P_2 that can well capture the experimentally observed distribution nor negligible high-order effective interactions, as commonly assumed [2–4].

For information transmission, entropy is introduced to characterize the uncertainty associated with a distribution instead of the precise form of the distribution [30].

However, as a coarse-grained description of a distribution, an entropic criterion, *e.g.*, MIF, is often insufficient to characterize whether an estimated distribution can represent the experimentally observed distribution and often becomes problematic in assessing the performance of an MEP analysis. Note that the state variables of nodes in our study are chosen as 1 for active and 0 for inactive, the same as discussed in previous works [2,3,31]. However, we point out that our conclusion is also valid for other choices of binary-state variables, *e.g.*, 1 for active and -1 for inactive. In summary, the MIF may lead to an incorrect characterization of the distribution of neuronal firing patterns and thereby neuronal information coding. Considering that the MEP analysis has been widely used in diverse fields to infer the statistical distribution of system states, our results provide a cautionary tale of the over-interpretation of MEP results.

Methods. –

Maximum entropy principle (MEP) analysis. The MEP analysis has been described in detail in the previous works [3,4,25]. Given the expectation values of functions $f_r(x)$ for $r = 1, 2, \cdots m$, *i.e.*, $\langle f_r(x)\rangle = \sum_{i=1}^{L} p_i f_r(x_i)$, and the normalization constraint $\sum_{i=1}^{L} p_i = 1$, where p_i is the probability of x_i, the quantity x is assumed to take the discrete values x_i $(i = 1, 2, \cdots, L)$. x can be a scalar or a vector. The corresponding entropy is $S = -\sum_{i=1}^{L} p_i \ln p_i$. Maximizing the entropy S subject to the given constraints yields the MEP probability distribution $p_i = \exp\{-\lambda_0 - \sum_{k=1}^{m} \lambda_k f_k(x_i)\}$, where λ_k for $k = 0, 1, \cdots, m$ are the Lagrangian multipliers.

For an ensemble of n nodes, its state in each sampling time bin can be denoted by $\Sigma = (\sigma_1, \cdots, \sigma_n) \in \{0, 1\}^n$. Here, the k-th–order correlation is the expectation of $(\sigma_{i_1} - \bar{\sigma}_{i_1}) \cdots (\sigma_{i_k} - \bar{\sigma}_{i_k})$ for nodes with respect to the observed distribution $P_{\rm obs}(\Sigma)$ of network states, with $\bar{\sigma}_{i_j}$ being the mean value of σ_{i_j}. To obtain correlations up to the m-th order requires to evaluate all expectations of $\sigma_{i_1}\sigma_{i_2}\cdots\sigma_{i_M}$ for nodes with respect to $P_{\rm obs}(\Sigma)$, where $i_1 < i_2 < \cdots < i_M$ and $1 \le M \le m$. The m-th–order MEP probability is the distribution by maximizing the entropy subject to measured correlations up to the m-th order as in eq. (1) in the main text. Numerically, we use an iterative scaling algorithm [4,26] to estimate the parameters of the MEP model in eq. (1) in the main text for m less than n.

For the numerical examples in the MEP analysis, first, P_n is obtained using eq. (1) in the main text with given strengths of effective interactions. Then, the mean values and the second-order correlations are computed with respect to P_n and are used as constraints of the entropy maximization for obtaining P_2. Finally, we estimate parameters for P_2 as above.

Calcium Imaging experiment. *Cell culture*: Low-density co-cultures of dissociated hippocampal neurons and glial cells from E18 rat embryos were prepared as previously described [28]. Briefly, glass coverslips were evenly pre-coated with poly-L-lysine (PLL) and then covered with $\sim 200\,\mu$m thick membranes made of PDMS (Sylgard 184, Dow Corning). An array of 1 mm diameter holes were pre-made on the membrane following standard soft photolithographic procedures, forming wells on the PLL-coated coverslip. Dissociated hippocampal cells were evenly plated onto the coverslips. Each well supported the growth of 50 to 100 neurons into interconnected networks with glial cells usually forming a monolayer underneath. Cultures were used at 12 to 14 days *in vitro* when evoked reverberatory neuronal activity was frequently observed. All animal experiments were performed according to guidelines and protocols approved by the Animal Care and Use Committee of University of Science and Technology of China.

Electrophysiology: Perforated whole-cell recordings were carried out at room temperature using patch-clamp amplifiers (MultiClamp 700 B, Molecular Devices), as previously described [28]. Patch-clamp data were acquired using customized IGOR Pro (WaveMetrics) programs. Intracellular pipette solution (pH 7.3) contained: 136.5 mM potassium gluconate, 17.5 mM KCl, 9 mM NaCl, 1 mM $MgCl_2$, 10 mM Hepes, 0.2 mM EGTA, and $\sim 200\,\mu$g/ml amphotericin B. The external bath solution (pH 7.3) contained: 150 mM NaCl, 3 mM KCl, 3 mM $CaCl_2$, 2 mM $MgCl_2$, 10 mM Hepes, and 5 mM Glucose. Stock solutions of 10 mM BMI were prepared in water and used after dilution in external bath solution.

Fluorescent imaging: Cells were loaded with membrane permeable calcium indicator Fluo-8 AM (TEFLabs) for 15 min at room temperature. Stock solution of the indicator was prepared in DMSO with 20% pluronic F-127 (Invitrogen) and then diluted in bath solution to a final concentration of 1 or $2\,\mu$M. Calcium imaging of network activity was performed on an inverted microscope (Olympus IX71) with a 20× (NA 0.7) objective. Image series were acquired using an sCMOS camera (Andor Neo) at 200 frames per second. The camera and the light source were both synchronized with electrophysiological stimulation via custom-built hardware triggering circuits. After an experiment, high K^+ solution (same as that used in the patch pipette without amphotericin B) was perfused to the culture to activate all neurons in the network in order to obtain high-brightness fluorescence images of all neurons.

Image analysis: Offline analyses of fluorescence image series were carried out using ImageJ (NIH, `http://imagej.nih.gov`) and custom programs based on Matlab (MathWorks). After morphological segmentation based on high-contrast images acquired in high K^+ solution, the relative calcium fluorescence signal ($\Delta F/F$) of each neuron was measured. At the relatively high spiking rate during reverberation, neuronal calcium signals in response to each spike increases in a stepwise fashion. This is due to the relatively slow clearance of calcium that has a typical timescale of a few hundred milliseconds. To obtain the spike train of each neuron, we first detected the inflection

time points (T_0) of each step, and performed local single-exponential fitting in a time window (50 ms) before the points T_0 and linear fitting in 20 ms after T_0. The crossing timings of the exponential and linear fitting pairs were identified as spike timings.

The time duration in experimental recording: The tolerance ϵ in our entropy estimation is set to be 2.5%, the time duration in experimental recording for the full MEP analysis can be estimated as [32] $T_{\min} \geq 2^n/(2n\epsilon\bar{\nu})$, where $\bar{\nu}$ is the mean firing rate. In our data recorded by Calcium Imaging, the minimum firing rate of neurons (fig. 2) is 2.8 Hz. The minimum experimental time scales can be estimated as $T_{\min} \geq 45$ s. We recorded the neuronal spike trains for 45 s in our experiment.

The interactions in the n-th order MEP analysis. Since P_n is identical to $P_{\rm obs}(\Sigma)$, effective interactions in the n-th–order MEP analysis can be computed by $P_{\rm obs}(\Sigma)$. For a fixed sampling time bin, we partition the recording time by the bin size and record the state of nodes at every sampling time bin. We count the occurring frequency of all possible states of n nodes as the observed distribution $P_{\rm obs}(\Sigma)$ of network states. Substituting 2^n states of $\Sigma = (\sigma_1, \sigma_2, \cdots, \sigma_n)$ and $P_{\rm obs}(\Sigma)$ into the n-th–order MEP analysisas shown in eq. (1) in the main text, then taking the logarithm of both sides of eq. (1) in the main text, we obtain a system of linear equations for all the effective interactions in terms of $P_{\rm obs}(\Sigma)$,

$$\sum_{k=1}^{n} \sum_{i_1<\cdots<i_k}^{n} J_{i_1,\cdots,i_k}\sigma_{i_1}\cdots\sigma_{i_k} = \log P_{\rm obs}(\Sigma) + \log Z, \quad (4)$$

where $Z = 1/P_{\rm obs}(\Sigma = (0, 0, \cdots, 0))$. By solving the system of linear equations (4), we can obtain all the $2^n - 1$ effective interactions J's for the n-th–order MEP analysis in terms of $P_{\rm obs}(\Sigma)$.

* * *

This work was supported by NSFC-11671259, NSFC-11722107, NSFC-91630208 and Shanghai Rising-Star Program-15QA1402600 (DZ); by NSF DMS-1009575 and NSFC-31571071 (DC); by Shanghai 14JC1403800, 15JC1400104, and SJTU-UM Collaborative Research Program (DC and DZ); and by the NYU Abu Dhabi Institute G1301 (Z-QJX, DZ, and DC); Strategic Priority Research Program of the Chinese Academy of Sciences (XDB02050000 to GB), National Basic Research Program of China (2013CB835101 to GB).

REFERENCES

[1] Park J. and Newman M. E., *Phys. Rev. E*, **70** (2004) 066117.
[2] Schneidman E., Berry M. J., Segev R. and Bialek W., *Nature*, **440** (2006) 1007.
[3] Shlens J., Field G. D., Gauthier J. L., Grivich M. I., Petrusca D., Sher A., Litke A. M. and Chichilnisky E., *J. Neurosci*, **26** (2006) 8254.
[4] Tang A., Jackson D., Hobbs J., Chen W., Smith J. L., Patel H., Prieto A., Petrusca D., Grivich M. I. and Sher A. *et al.*, *J. Neurosci*, **28** (2008) 505.
[5] Xu Z.-Q. J., Bi G., Zhou D. and Cai D., *Commun. Math. Sci.*, **15** (2017) 665.
[6] Xu Z.-Q. J., Zhou D. and Cai D., *Entropy*, **21** (2019) 76.
[7] Xu Z.-Q. J., Crodelle J., Zhou D. and Cai D., *Phys. Rev. E*, **99** (2019) 022409.
[8] Bury T., *Eur. Phys. J. B*, **86** (2013) 89.
[9] Watanabe T., Hirose S., Wada H., Imai Y., Machida T., Shirouzu I., Konishi S., Miyashita Y. and Masuda N., *Nat. Commun.*, **4** (2013) 1370.
[10] Marre O., El Boustani S., Frégnac Y. and Destexhe A., *Phys. Rev. Lett.*, **102** (2009) 138101.
[11] Yu S., Yang H., Nakahara H., Santos G. S., Nikolić D. and Plenz D., *J. Neurosci.*, **31** (2011) 17514.
[12] Barreiro A. K., Gjorgjieva J., Rieke F. and Shea-Brown E., *Front. Comput. Neurosci.*, **8** (2014) 10.
[13] Martin E. A., Hlinka J. and Davidsen J., *Phys. Rev. E*, **94** (2016) 040301.
[14] Ohiorhenuan I. E., Mechler F., Purpura K. P., Schmid A. M., Hu Q. and Victor J. D., *Nature*, **466** (2010) 617.
[15] Roudi Y., Nirenberg S. and Latham P. E., *PLoS Comput. Biol.*, **5** (2009) e1000380.
[16] Macke J. H., Opper M. and Bethge M., *Phys. Rev. Lett.*, **106** (2011) 208102.
[17] Shimazaki H., Sadeghi K., Ishikawa T., Ikegaya Y. and Toyoizumi T., *Sci. Rep.*, **5** (2015) 9821.
[18] Merchan L. and Nemenman I., *J. Stat. Phys.*, **162** (2016) 1294.
[19] Ganmor E., Segev R. and Schneidman E., *Proc. Natl. Acad. Sci. U.S.A.*, **108** (2011) 9679.
[20] Tkačik G., Marre O., Amodei D., Schneidman E., Bialek W. and Berry II M. J., *PLoS Comput. Biol.*, **10** (2014) e1003408.
[21] Humplik J. and Tkačik G., *PLoS Comput. Biol.*, **13** (2017) e1005763.
[22] Cayco-Gajic N., Zylberberg J. and Shea-Brown E., *Entropy*, **20** (2018) 489.
[23] Jizba P. and Korbel J., *Phys. Rev. Lett.*, **122** (2019) 120601.
[24] Abeles M., Prut Y., Bergman H. and Vaadia E., *Prog. Brain Res.*, **102** (1994) 395.
[25] Jaynes E. T., *Phys. Rev.*, **106** (1957) 620.
[26] Darroch J. N. and Ratcliff D., *Ann. Math. Stat.*, **43** (1972) 1470.
[27] Amari S.-I., *IEEE Trans. Inf. Theory*, **47** (2001) 1701.
[28] Lau P.-M. and Bi G.-Q., *Proc. Natl. Acad. Sci. U.S.A.*, **102** (2005) 10333.
[29] Cohen M. R. and Kohn A., *Nat. Neurosci.*, **14** (2011) 811.
[30] Shannon C. E., *Bell Syst. Tech. J.*, **27** (1948) 379.
[31] Schneidman E., Still S., Berry M. J., Bialek W. *et al.*, *Phys. Rev. Lett.*, **91** (2003) 238701.
[32] Macke J. H., Murray I. and Latham P. E., *Entropy*, **15** (2013) 3109.