Explicit mutual information for simple networks and neurons with lognormal activities

Maurycy Chwiłka and Jan Karbowski*

Department of Mathematics, Informatics, and Mechanics, Institute of Applied Mathematics and Mechanics, University of Warsaw, Ulica Banacha 2, 02-097 Warsaw, Poland

(Received 22 June 2023; revised 14 September 2023; accepted 13 December 2023; published 16 January 2024)

Networks with stochastic variables described by heavy-tailed lognormal distribution are ubiquitous in nature, and hence they deserve an exact information-theoretic characterization. We derive analytical formulas for mutual information between elements of different networks with correlated lognormally distributed activities. In a special case, we find an explicit expression for mutual information between neurons when neural activities and synaptic weights are lognormally distributed, as suggested by experimental data. Comparison of this expression with the case when these two variables have short tails reveals that mutual information with heavy tails for neurons and synapses is generally larger and can diverge for some finite variances in presynaptic firing rates and synaptic weights. This result suggests that evolution might prefer brains with heterogeneous dynamics to optimize information processing.

DOI: 10.1103/PhysRevE.109.014117

I. INTRODUCTION

Lognormal distribution is arguably one of the most common continuous probability distributions describing naturally occurring phenomena in nature [1,2]: from environment and geology [1,3] to stock market fluctuations and finance [4,5] to brain activity [6–8]. Some other examples are included in Refs. [9–11]. Lognormal distribution has a skewed shape and is characterized by a long tail [2]. This means that the likelihood of huge size stochastic events with this distribution is small, but still significantly higher than for those described by short-tailed distributions, e.g., by Gaussian density.

Two or more stochastic variables can be correlated, and one of the most useful tools describing such correlation is mutual information (MI). MI is an information-theoretic concept [12,13] that measures the average number of bits we gain about behavior of one variable by observing the variability of another, correlated variable. Obviously, MI has a wide applicability in many different areas of science, from information science [12,13] and rapidly developing machine and deep learning [14,15] through different branches of physics (as diverse as material science, stochastic thermodynamics, and cosmology) [16–19] to molecular biology [20,21] and neuroscience [22]. Despite this huge popularity in using MI, it is important to stress that its exact analytical form for nontrivial continuous distributions is known only in few cases [23], which are, however, unexplored. It seems that in practical applications only MI for short-tailed Gaussian distribution is used (e.g., Refs. [12,21,22,24]). On the other hand, the exact form of MI for heavy-tailed lognormal distribution is rarely mentioned and virtually not used (but see a recent notable exception in Ref. [25]). Thus, clearly, there is a need for exploration of general properties of MI for lognormal variables in different settings.

In this paper, we present and use the general formula for MI between vectors of random variables with lognormal distributions to obtain explicit analytical expressions for MI in specific networks with simple topology, which can appear in many applications. Furthermore, we apply this formalism to neuroscience and derive analytical MI for information transfer between neurons in the neural networks relevant for brain cerebral cortex, i.e., when a neuron receives many correlated synaptic inputs with heavy tails.

II. PRELIMINARIES

A. Lognormal distribution

Let $\mathbf{z} = (z_1, z_2, \ldots, z_N)$ be an N-dimensional Gaussian random variable with the mean vector $\mu = (\mu_1, \ldots, \mu_N)$ and the covariance matrix Σ, which is positive definite (symmetric with real and positive eigenvalues). We define a new multidimensional random variable $\mathbf{x} = (x_1, x_2, \ldots, x_N)$, such that $x_i = e^{z_i}$ for every $i = 1, \ldots, N$. Then $\mathbf{x}$ is lognormally distributed and has the following probability density $\rho(\mathbf{x})$:

$$\rho(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^N \det(\Sigma)} \prod_{i=1}^N x_i} \times \exp\left(-\frac{1}{2}[\ln(\boldsymbol{x}) - \boldsymbol{\mu}]^T \Sigma^{-1} [\ln(\boldsymbol{x}) - \boldsymbol{\mu}]\right), \quad (1)$$

with $\langle \ln(x_i)\rangle = \mu_i$ and the covariance matrix $\Sigma_{ij} = \langle \ln(x_i)\ln(x_j)\rangle - \langle \ln(x_i)\rangle\langle \ln(x_j)\rangle$, where $\langle \ldots \rangle$ denotes averaging with respect to the density in Eq. (1). We denote the variance of $\ln x_i$ as $\Sigma_{ii} = \langle \ln^2(x_i)\rangle - \langle \ln(x_i)\rangle^2 \equiv \sigma_i^2$. For $N = 1$, the density in Eq. (1) reduces to $\rho(x) = e^{-[\ln(x)-\mu]^2/2\sigma^2}/(\sqrt{2\pi\sigma^2}x)$.

The major moments of the lognormal distribution, including variance $\mathrm{Var}(x_i) = \langle x_i^2\rangle - \langle x_i\rangle^2$ and covariance

*Corresponding author: jkarbowski@mimuw.edu.pl

$\mathrm{Cov}(x_i, x_j) = \langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle$, are related to the parameters μ_i and Σ_{ij} as [2]

$$\begin{aligned}\langle x_i\rangle &= e^{\mu_i+\frac{1}{2}\sigma_i^2},\\ \mathrm{Var}(x_i) &= \left(e^{\sigma_i^2}-1\right)\langle x_i\rangle^2,\\ \mathrm{Cov}(x_i,x_j) &= \left(e^{\Sigma_{ij}}-1\right)\langle x_i\rangle\langle x_j\rangle.\end{aligned}\tag{2}$$

B. Definition of mutual information

Mutual information between two lognormal random variables X and Y is defined as [12]

$$\mathrm{MI}(X,Y) = h(X) + h(Y) - h(X,Y),\tag{3}$$

where $h(X)$ and $h(Y)$ are differential entropies for X and Y variables, i.e.,

$$h(X) = -\frac{1}{\ln 2}\int \rho(\boldsymbol{x})\ln[\rho(\boldsymbol{x})]d\boldsymbol{x},$$

and $h(X,Y)$ is the joint differential entropy given by

$$h(X,Y) = -\frac{1}{\ln 2}\int \rho(\boldsymbol{x},\boldsymbol{y})\ln[\rho(\boldsymbol{x},\boldsymbol{y})]d\boldsymbol{x}d\boldsymbol{y},$$

where $\rho(\mathbf{x})$ is as in Eq. (1) and $\rho(\mathbf{x},\mathbf{y})$ is the joint probability density of $\mathbf{x}$ and $\mathbf{y}$ (see below). $\mathrm{MI}(X,Y)$ quantifies the amount of information we gain about X by observing Y, or vice versa.

C. Mutual information for lognormal stochastic vectors

In this section we present a nonstandard derivation of MI for two random vectors with lognormal distributions.

First we determine the entropy of the N-dimensional variable $\mathbf{x}$. To achieve this, we need the average of the argument of exp in Eq. (1). We have $\langle[\ln(\boldsymbol{x})-\mu]^T\Sigma^{-1}[\ln(\boldsymbol{x})-\mu)]\rangle = \sum_{ij}(\Sigma^{-1})_{ij}\langle[\ln(x_i)-\mu_i][\ln(x_j)-\mu_j]\rangle = \sum_{ij}(\Sigma^{-1})_{ij}\Sigma_{ji} = \sum_{ij}(\Sigma^{-1}\Sigma)_{ii} = N$, where we used the definition of the covariance matrix appearing after Eq. (1). After some additional straightforward steps, we obtain the formula for entropy of the N-dimensional lognormal random variable $\mathbf{x}$:

$$\begin{aligned}h(\mathbf{x}) &= -\frac{1}{\ln 2}\langle\ln[\rho(\boldsymbol{x})]\rangle\\ &= \frac{1}{\ln 2}\left[\frac{1}{2}\ln[(2\pi e)^N\det(\Sigma)] + \sum_{i=1}^{N}\langle\ln(x_i)\rangle\right].\end{aligned}\tag{4}$$

Now, we partition our dimensionality N into two parts, $N = n + k$, and define two new multidimensional lognormal random variables X and Y such that $\mathbf{x} = (X, Y)$, where $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_k)$, with $y_i \equiv x_{n+i}$. The lognormal distribution of X has the parameters $\mu_{\mathbf{X}}$ and Σ_X, defined as $\mu_{\mathbf{X}} = (\langle\ln x_1\rangle, \ldots, \langle\ln x_n\rangle)$ and $n\times n$ matrix $(\Sigma_X)_{ij} = \langle\ln(x_i)\ln(x_j)\rangle - \langle\ln(x_i)\rangle\langle\ln(x_j)\rangle$. The lognormal distribution of Y has the parameters $\mu_{\mathbf{Y}}$ and Σ_Y, defined as $\mu_{\mathbf{Y}} = (\langle\ln y_1\rangle, \ldots, \langle\ln y_k\rangle)$ and $k\times k$ matrix $(\Sigma_Y)_{ij} = \langle\ln(y_i)\ln(y_j)\rangle - \langle\ln(y_i)\rangle\langle\ln(y_j)\rangle$. The N-dimensional variable $\mathbf{x}$ is lognormally distributed with a vector of means $(\mu_{\mathbf{X}}, \mu_{\mathbf{Y}})$ and an $N\times N$ global covariance matrix Σ_{XY} dependent on Σ_X and Σ_Y matrices through the block matrix

$$\Sigma_{XY} = \begin{bmatrix}\Sigma_X & \mathrm{Cov}_{XY}\\ \mathrm{Cov}_{XY}^T & \Sigma_Y\end{bmatrix},\tag{5}$$

where Cov_{XY} is an $n\times k$ covariance matrix between variables $\ln X$ and $\ln Y$, i.e., $\mathrm{Cov}_{XY} \equiv \mathrm{Cov}(\ln X, \ln Y)$, and $(\mathrm{Cov}_{XY})_{ij} = \langle\ln(x_i)\ln(y_j)\rangle - \langle\ln(x_i)\rangle\langle\ln(y_j)\rangle$.

Consequently, from Eq. (4) it follows that we can write entropies for each of the three lognormal variables X, Y, and $\mathbf{x}$ as

$$\begin{aligned}h(X) &= \frac{1}{\ln 2}\left[\frac{1}{2}\ln[(2\pi e)^n\det(\Sigma_X)] + \sum_{i=1}^{n}\langle\ln(x_i)\rangle\right],\\ h(Y) &= \frac{1}{\ln 2}\left[\frac{1}{2}\ln[(2\pi e)^k\det(\Sigma_Y)] + \sum_{i=n+1}^{N}\langle\ln(x_i)\rangle\right],\\ h(X,Y) &= \frac{1}{\ln 2}\left[\frac{1}{2}\ln[(2\pi e)^N\det(\Sigma_{XY})] + \sum_{i=1}^{N}\langle\ln(x_i)\rangle\right].\end{aligned}$$

Now, using Eq. (3), we can observe that

$$\mathrm{MI}(X,Y) = \frac{1}{2\ln 2}\ln\left(\frac{\det(\Sigma_X)\det(\Sigma_Y)}{\det(\Sigma_{XY})}\right).\tag{6}$$

This is the general formula for mutual information of two random multidimensional lognormal variables, and it depends only on the matrices of covariance of corresponding Gaussian random vectors with appropriate parameters—both joint and marginal distributions. Thus, MI for lognormal variables is formally the same as MI for the underlying Gaussian variables [12], which is a consequence of the invariance of MI with respect to any bijective (one-to-one) transformation of the variables involved (see Ref. [25] for this type of approach).

The formula (6) can be rewritten in an equivalent form if we use a Schur's decomposition for the determinant of the block matrix Σ_{XY}, namely [26,27],

$$\begin{aligned}\det\Sigma_{XY} &= \det(\Sigma_X)\det\left(\Sigma_Y - \mathrm{Cov}_{XY}^T\Sigma_X^{-1}\mathrm{Cov}_{XY}\right)\\ &= \det(\Sigma_Y)\det\left(\Sigma_X - \mathrm{Cov}_{XY}\Sigma_Y^{-1}\mathrm{Cov}_{XY}^T\right).\end{aligned}\tag{7}$$

Then the mutual information is

$$\begin{aligned}\mathrm{MI}(X,Y) &= \frac{1}{\ln 4}\ln\left(\frac{\det(\Sigma_Y)}{\det\left(\Sigma_Y - \mathrm{Cov}_{XY}^T\Sigma_X^{-1}\mathrm{Cov}_{XY}\right)}\right)\\ &= \frac{1}{\ln 4}\ln\left(\frac{\det(\Sigma_X)}{\det\left(\Sigma_X - \mathrm{Cov}_{XY}\Sigma_Y^{-1}\mathrm{Cov}_{XY}^T\right)}\right)\\ &= \frac{1}{\ln 4}\ln\left(\frac{1}{\det\left(I - \Sigma_X^{-1}\mathrm{Cov}_{XY}\Sigma_Y^{-1}\mathrm{Cov}_{XY}^T\right)}\right),\end{aligned}\tag{8}$$

where I is the identity matrix, and the last equality follows from a known identity, $\det(AB) = \det(A)\det(B)$, valid for two arbitrary square matrices A and B. Equations (6) and (8) are the first major result of this paper and allow us to obtain exact analytical expressions for mutual information in some specific cases. Note that $\mathrm{MI}(X,Y)$ in Eq. (8) is nonzero only if there are correlations between X and Y, i.e., when the covariance matrix Cov_{XY} is nonzero. Depending on the structure of matrices Σ_X, Σ_Y, and Cov_{XY}, one can use either of the three expressions in Eq. (8), whichever is easier, for finding $\mathrm{MI}(X,Y)$.

III. ANALYTICAL RESULTS FOR MI IN SPECIFIC NETWORKS

In this section we consider several specific examples of networks with simple topology, and we give explicit formulas for MI in each case. These networks are relatively simple to analyze and are particular instances of the so-called bipartite networks with correlations that frequently appear in biological and social contexts [28–32]. Elements of the networks below are illustrated by graphs with nodes, which are connected by arrows. An arrow between the nodes indicates a correlation among particular elements with random lognormal activities.

A. Case $n = k = 1$

In this case, $X = x_1$ and $Y = y_1 = x_2$, and the global covariance matrix appearing in Eq. (5) is 2×2 and has the following form:

$$\Sigma_{XY} = \begin{bmatrix} \sigma_X^2 & \mathrm{Cov}_{XY} \\ \mathrm{Cov}_{XY} & \sigma_Y^2 \end{bmatrix}, \tag{9}$$

where $\sigma_X^2 = \mathrm{Var}(\ln X) = \langle \ln^2(X)\rangle - \langle \ln(X)\rangle^2$ and $\sigma_Y^2 = \mathrm{Var}(\ln Y) = \langle \ln^2(Y)\rangle - \langle \ln(Y)\rangle^2$. As a result

$$\mathrm{MI}(X,Y) = \frac{1}{2\ln 2}\ln\left(\frac{\sigma_X^2\sigma_Y^2}{\sigma_X^2\sigma_Y^2 - \mathrm{Cov}_{XY}^2}\right) = \frac{1}{2\ln 2}\ln\left(\frac{1}{1-c^2}\right), \tag{10}$$

where c denotes the correlation coefficient between variables $\ln(X)$ and $\ln(Y)$, defined as $c = \mathrm{Cov}_{XY}/\sigma_X\sigma_Y$. The coefficient c can be expressed by covariance of X and Y, $\mathrm{Cov}(X,Y)$, and their variances $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$ as

$$c = \frac{\ln\left[1 + \frac{\mathrm{Cov}(X,Y)}{\langle X\rangle\langle Y\rangle}\right]}{\sqrt{\ln\left[1+\frac{\mathrm{Var}(X)}{\langle X\rangle^2}\right]\ln\left[1+\frac{\mathrm{Var}(Y)}{\langle Y\rangle^2}\right]}}, \tag{11}$$

which follows from Eq. (2). Because of the Cauchy-Schwartz inequality for covariance and variance, we have $\mathrm{Cov}(X,Y) \leqslant \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}$, and this implies that c is bounded from above by c_0, which is

$$c_0 = \frac{\ln\left[1 + \frac{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}{\langle X\rangle\langle Y\rangle}\right]}{\sqrt{\ln\left[1+\frac{\mathrm{Var}(X)}{\langle X\rangle^2}\right]\ln\left[1+\frac{\mathrm{Var}(Y)}{\langle Y\rangle^2}\right]}}. \tag{12}$$

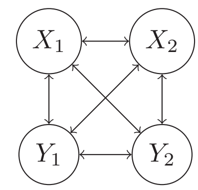
Consequently, we have an upper bound on $\mathrm{MI}(X,Y)$ in this case as

$$\mathrm{MI}(X,Y) \leqslant \frac{1}{2\ln 2}\ln\left(\frac{1}{1-c_0^2}\right). \tag{13}$$

This inequality may be useful in cases when we do not know covariances but we do know variances of original data.

B. Case $n = k = 2$

In this case, $X = (x_1, x_2)$ and $Y = (y_1, y_2)$, and the covariance matrix Σ_{XY} appearing in Eq. (5) is 4×4. We choose the components of Σ_{XY} with some degree of symmetry: $(\Sigma_X)_{ij} = \sigma_x^2\delta_{ij} + \gamma_x(1-\delta_{ij})$, $(\Sigma_Y)_{ij} = \sigma_y^2\delta_{ij} + \gamma_y(1-\delta_{ij})$, and $(\mathrm{Cov}_{XY})_{ij} = a\delta_{ij} + b(1-\delta_{ij})$, where $i,j = 1,2$. Here, all four variables x_1, x_2, y_1, and y_2 are mutually correlated (see the graph below).



Thus, we have $\det(\Sigma_Y) = \sigma_y^4 - \gamma_y^2$ and $\det(\Sigma_X) = \sigma_x^4 - \gamma_x^2$. Additionally, we obtain

$$\det\left[\Sigma_Y - \mathrm{Cov}_{XY}^T\Sigma_X^{-1}\mathrm{Cov}_{XY}\right] = \frac{\left[(\sigma_x^2-\gamma_x)(\sigma_y^2-\gamma_y) - (a-b)^2\right]}{\det\Sigma_X} \times \left[(\sigma_x^2+\gamma_x)(\sigma_y^2+\gamma_y) - (a+b)^2\right].$$

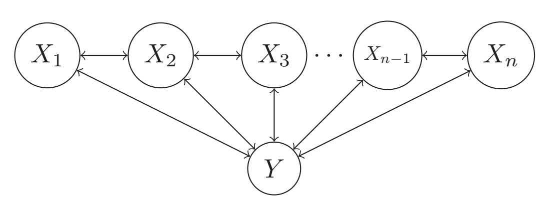
Using Eq. (8), this leads to mutual information as

$$\mathrm{MI}(X,Y) = \frac{-1}{\ln 4}\ln\left[\left(1 - \frac{(a-b)^2}{(\sigma_x^2+\gamma_x)(\sigma_y^2+\gamma_y)}\right) \times \left(1 - \frac{(a+b)^2}{(\sigma_x^2-\gamma_x)(\sigma_y^2-\gamma_y)}\right)\right]. \tag{14}$$

Equation (14) suggests that MI decreases monotonically with increasing variances σ_x^2 and σ_y^2 of $\ln(X)$ and $\ln(Y)$. Moreover, MI has a minimum as a function of correlations γ_x between x_1 and x_2, if $2\sigma_x^2(\sigma_y^2 - \gamma_y) \geqslant (a+b)^2$; see Fig. 1 (this analogously applies to γ_y). MI also diverges for large intracorrelations γ_x and γ_y and for sufficiently strong intercorrelations a and b between $\ln(X)$ and $\ln(Y)$ variables (either positive or negative; Fig. 1).

C. Case $n \geqslant 2$ and $k = 1$, with tridiagonal matrix for Σ_X

In this case we take Σ_X to be tridiagonal matrix, and we choose $(\Sigma_X)_{ii} = \sigma_x^2$ and $(\Sigma_X)_{i,i\pm1} = \gamma$, with all other elements of Σ_X to be zero. This means that correlated are only those x_i and x_j that are nearest neighbors in the "indices space." Examples of such short-range correlations are, e.g., correlations between spins in ferromagnets above a critical temperature, and density-density correlations in ideal and weakly nonideal gas (e.g., Ref. [33]). Matrix Σ_Y is just scalar, equal to σ_y^2. Covariance Cov_{XY} is a vector, and we take it as $(\mathrm{Cov}_{XY})_i = a$. This configuration is depicted by the graph below.



The matrix Σ_X is tridiagonal, and determinants of tridiagonal matrices are known [34,35]. The determinant of the arbitrary tridiagonal $n \times n$ matrix with all diagonal elements t and all off-diagonal elements s is given by the function

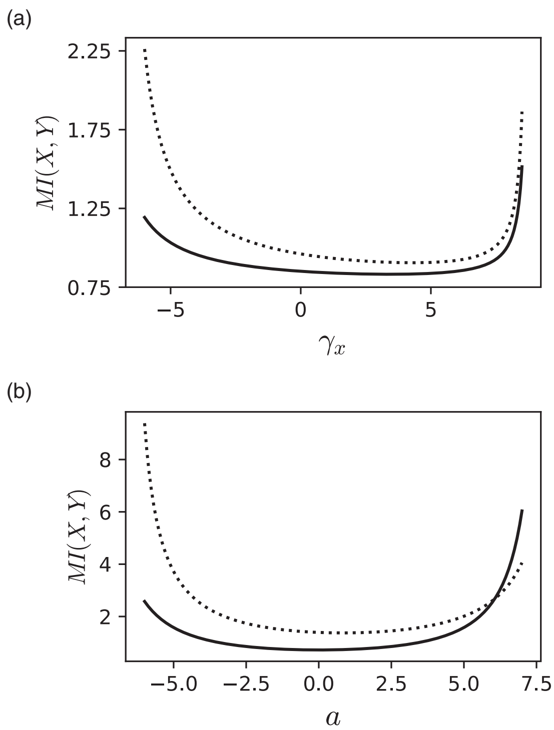


FIG. 1. MI in Eq. (14) as a function of γ_x and a. (a) Solid line: $\sigma_x^2 = 9$, $\sigma_y^2 = 9$, $a = 2.4$, $b = 1$, and $\gamma_y = 1$; dotted line: $\sigma_x^2 = 9$, $\sigma_y^2 = 9$, $a = 3$, $b = 1.5$, and $\gamma_y = 1$. (b) Solid line: $\sigma_x^2 = 9$, $\sigma_y^2 = 9$, $\gamma_x = 1$, $b = 0$, and $\gamma_y = 1$; dotted line: $\sigma_x^2 = 9$, $\sigma_y^2 = 16$, $\gamma_x = 1$, $b = 8.4$, and $\gamma_y = 1$.

$\theta_n(t, s)$, which takes the form [34,35]

$$\theta_n(t, s) = \frac{[(t + \sqrt{t^2 - 4s^2})^{n+1} - (t - \sqrt{t^2 - 4s^2})^{n+1}]}{2^{n+1}\sqrt{t^2 - 4s^2}}. \quad (15)$$

Hence, we have $\det(\Sigma_X) = \theta_n(\sigma_x^2, \gamma)$. Mutual information is determined from the first line in Eq. (8) and is given by

$$\mathrm{MI}(X, Y) = \frac{1}{\ln 4} \ln\left(\frac{\sigma_y^2}{\sigma_y^2 - \mathrm{Cov}_{XY}^T \Sigma_X^{-1} \mathrm{Cov}_{XY}}\right), \quad (16)$$

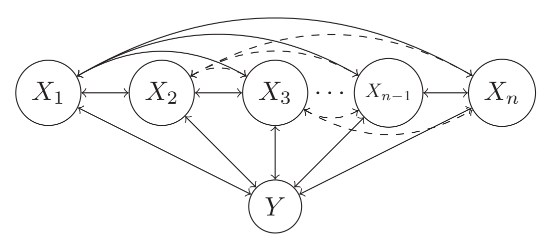
where $\mathrm{Cov}_{XY}^T \Sigma_X^{-1} \mathrm{Cov}_{XY}$ can be found after tedious calculations as (see the Appendix)

$$\begin{aligned}
&\mathrm{Cov}_{XY}^T \Sigma_X^{-1} \mathrm{Cov}_{XY} \\
&= \frac{na^2}{(\sigma_x^2 + 2\gamma)} + \frac{\gamma a^2}{2^n(\sigma_x^2 + 2\gamma)(\sigma_x^4 - 4\gamma^2)\det(\Sigma_X)} \\
&\quad \times \Big[(-1)^{n+1} 2^{n+1} \gamma^n (\sigma_x^2 - 2\gamma) + \Big(\sigma_x^2 + \sqrt{\sigma_x^4 - 4\gamma^2}\Big)^n \\
&\quad \times \Big(\sigma_x^2 - 2\gamma + \sqrt{\sigma_x^4 - 4\gamma^2}\Big) + \Big(\sigma_x^2 - \sqrt{\sigma_x^4 - 4\gamma^2}\Big)^n \\
&\quad \times \Big(\sigma_x^2 - 2\gamma - \sqrt{\sigma_x^4 - 4\gamma^2}\Big)\Big]. \quad (17)
\end{aligned}$$

The above complicated formula simplifies in two extreme cases. In the case when $n = 2$, we obtain $\mathrm{Cov}_{XY}^T \Sigma_X^{-1} \mathrm{Cov}_{XY} = 2a^2/(\sigma_x^2 + \gamma)$, and hence mutual information takes the simple form $\mathrm{MI}(X, Y) = \frac{1}{\ln 4} \ln(\frac{\sigma_y^2}{\sigma_y^2 - 2a^2/(\sigma_x^2+\gamma)})$. In the case of $n \gg 1$, the first term on the right in Eq. (17) dominates, and we obtain $\mathrm{Cov}_{XY}^T \Sigma_X^{-1} \mathrm{Cov}_{XY} \approx na^2/(\sigma_x^2 + \gamma) + O(1)$, and MI is given by a similar expression: $\mathrm{MI}(X, Y) \approx \frac{1}{\ln 4} \ln(\frac{\sigma_y^2}{\sigma_y^2 - na^2/(\sigma_x^2+\gamma)})$. The latter formula implies that MI diverges for the number of units $n \approx \sigma_y^2(\sigma_x^2 + \gamma)/a^2$. In both cases, it is easily seen that MI grows with correlations a between variables X and Y (both positive and negative), but it decreases monotonically with increasing the correlations γ between x_i variables. The latter implies that negative correlations among x_i carry more information than their positive correlations. This is a different situation than in Case B, where both strong positive and negative correlations between x_i enhance MI. As expected, MI also decreases with increasing variances σ_x^2 and σ_y^2. The general dependence of mutual information on γ, σ_x^2, and n, based on the exact Eqs. (16) and (17), is presented in Fig. 2.

D. Case $n \geqslant 2$ and $k = 1$, with Kac-Murdock-Szegö matrix for Σ_X

The only difference with Case C above is that now the matrix Σ_X has all nonzero elements, which are given by $(\Sigma_X)_{ij} = \sigma_x^2 \gamma^{|i-j|}$, where $|\gamma| \leqslant 1$ [36]. This means that all x_i are mutually correlated, but distant ones (with remote indices) are gradually less correlated (see below).



The inverse of the matrix Σ_X is [26,36]

$$\left(\Sigma_X^{-1}\right)_{ij} = \frac{\delta_{ij}[1 + \gamma^2(1 - \delta_{i1} - \delta_{in})] - \gamma(\delta_{i,j-1} + \delta_{i,j+1})}{\sigma_x^2(1 - \gamma^2)}, \quad (18)$$

which allows us to find

$$\mathrm{Cov}_{XY}^T \Sigma_X^{-1} \mathrm{Cov}_{XY} = \frac{a^2[n(1-\gamma) + 2\gamma]}{\sigma_x^2(1+\gamma)}. \quad (19)$$

Consequently, mutual information in this case is given by a simple explicit formula:

$$\mathrm{MI}(X, Y) = \frac{1}{\ln 4} \ln\left(\frac{\sigma_y^2}{\sigma_y^2 - \frac{a^2[n(1-\gamma)+2\gamma]}{\sigma_x^2(1+\gamma)}}\right). \quad (20)$$

It is easily seen that MI grows monotonically with the number of units n and diverges for $n = \frac{[\sigma_x^2\sigma_y^2(1+\gamma) - 2\gamma a^2]}{(1-\gamma)a^2}$. The dependence of MI on other parameters is similar to Case C above. In particular, MI decreases monotonically with increasing correlations γ among x_i variables from negative ($\gamma < 0$) to positive ($\gamma > 0$).

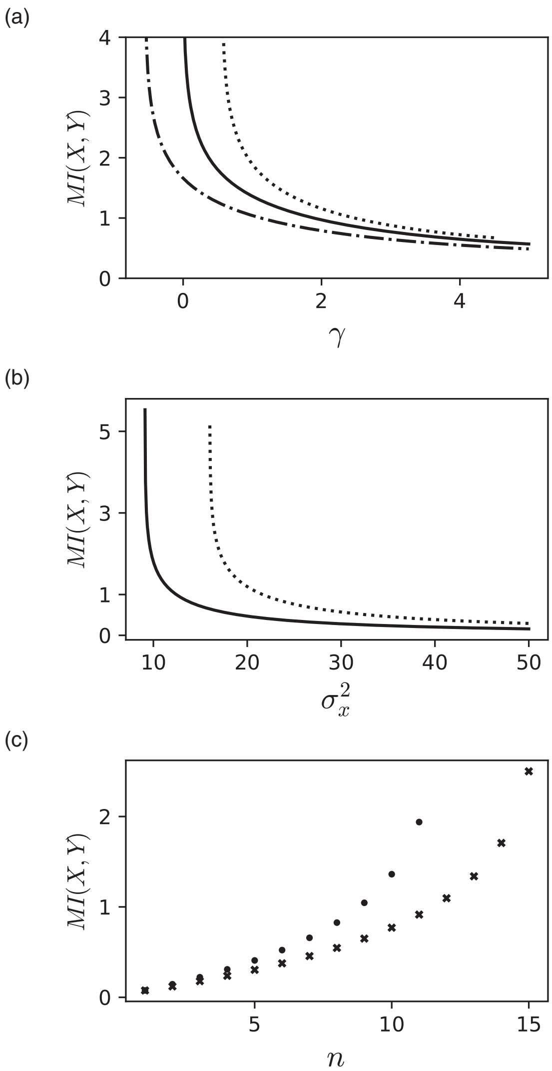
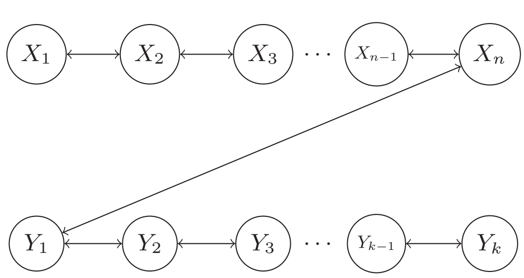


FIG. 2. MI in Eq. (16) as a function of γ, σ_x^2, and n. (a) Solid line: $\sigma_x^2 = 10$, $\sigma_y^2 = 1$, $a = 1$, and $n = 10$; dotted line: $\sigma_x^2 = 9$, $\sigma_y^2 = 9$, $a = 3$, and $n = 10$; dash-dotted line: $\sigma_x^2 = 10$, $\sigma_y^2 = 10$, $a = 3$, and $n = 10$. (b) Solid line: $\sigma_y^2 = 1$, $a = 1$, $\gamma = \frac{1}{2}$, and $n = 10$; dotted line: $\sigma_y^2 = 1$, $a = 1.3$, $\gamma = \frac{1}{2}$, and $n = 10$. (c) Circles: $\sigma_x^2 = 10$, $\sigma_y^2 = 1$, $a = 1$, and $\gamma = 1$; crosses: $\sigma_x^2 = 10$, $\sigma_y^2 = 1$, $a = 1$, and $\gamma = 3$.

E. Case $n \geqslant 2$ and $k \geqslant 2$, with nonsymmetric and nondiagonal Cov_{XY}

Let us consider the $n \times n$ matrix Σ_X and the $k \times k$ matrix Σ_Y to be both tridiagonal with elements $(\Sigma_X)_{ij} = \sigma_x^2\delta_{ij} + \gamma_x(\delta_{i,j-1} + \delta_{i,j+1})$ and $\Sigma_Y = \sigma_y^2\delta_{ij} + \gamma_y(\delta_{i,j-1} + \delta_{i,j+1})$. The covariance Cov_{XY} is an $n \times k$ sparse matrix taken with one nonzero element $(\mathrm{Cov}_{XY})_{n1} = a$ and the rest of the elements are 0 (see the graph below).



This case is relevant for two groups of cascade networks of interacting elements.

We use the last line in Eq. (8) for MI. In this configuration, it is easy to show that $(\Sigma_X^{-1}\mathrm{Cov}_{XY}\,\Sigma_Y^{-1}\mathrm{Cov}_{XY}^T)_{ij} = a^2(\Sigma_X^{-1})_{in}(\Sigma_Y^{-1})_{11}$ for $j = n$ and 0 for $1 \leqslant j \leqslant n-1$, for every i. This leads to the simple form for the determinant $\det(I - \Sigma_X^{-1}\mathrm{Cov}_{XY}\,\Sigma_Y^{-1}\mathrm{Cov}_{XY}^T) = 1 - a^2(\Sigma_X^{-1})_{nn}(\Sigma_Y^{-1})_{11}$, and consequently for MI

$$\mathrm{MI}(X, Y) = \frac{1}{\ln 4}\ln\left(\frac{1}{1 - a^2\left(\Sigma_X^{-1}\right)_{nn}\left(\Sigma_Y^{-1}\right)_{11}}\right). \tag{21}$$

The matrix elements $(\Sigma_X^{-1})_{nn}$ and $(\Sigma_Y^{-1})_{11}$ are given by the ratios of the function θ_n defined in Eq. (15), i.e., $(\Sigma_X^{-1})_{nn} = \theta_{n-1}(\sigma_x^2, \gamma_x)/\theta_n(\sigma_x^2, \gamma_x)$ and $(\Sigma_Y^{-1})_{11} = \theta_{k-1}(\sigma_y^2, \gamma_y)/\theta_k(\sigma_y^2, \gamma_y)$; see the Appendix.

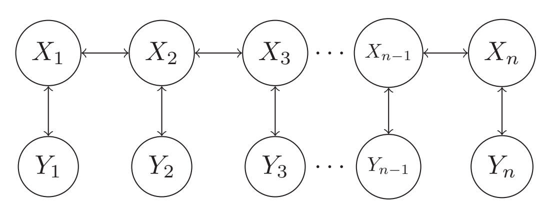
In the limits $n \gg 1$ and $k \gg 1$, and for $\sigma_x^2 > 2\gamma_x$ and $\sigma_y^2 > 2\gamma_y$, we obtain an approximated MI as

$$\mathrm{MI}(X, Y)_{n\gg 1} \approx \frac{-1}{\ln 4}\ln\left[1 - \frac{4a^2}{\left(\sigma_x^2 + \sqrt{\sigma_x^4 - 4\gamma_x^2}\right)} \times \left(\sigma_y^2 + \sqrt{\sigma_y^4 - 4\gamma_y^2}\right)^{-1}\right].$$

This equation shows that large variances of both σ_x^2 and σ_y^2 have a detrimental effect on MI. On the other hand, increasing the covariances among $\ln(X)$ ($|\gamma_x|$) and among $\ln(Y)$ ($|\gamma_y|$) causes an increase in MI.

F. Case $n = k \geqslant 2$, with symmetric and diagonal Cov_{XY}

In this case, we take Σ_X to be tridiagonal with elements $(\Sigma_X)_{ij} = \sigma_x^2\delta_{ij} + \gamma_x(\delta_{i,j-1} + \delta_{i,j+1})$, and Σ_Y to be diagonal with elements $(\Sigma_Y)_{ij} = \sigma_y^2\delta_{ij}$. Additionally, we choose the covariance matrix Cov_{XY} to be diagonal, $(\mathrm{Cov}_{XY})_{ij} = a\delta_{ij}$. This situation corresponds to the system depicted below.



Here, there are correlations between nearest neighbors of variables x_i, but there are no correlations between variables y_i. Additionally, x_i correlates directly only with y_i.

Elements of the matrix of interest $\Sigma_X - \mathrm{Cov}_{XY}\,\Sigma_Y^{-1}\mathrm{Cov}_{XY}^T$ appearing in Eq. (8) for mutual information are $(\Sigma_X - \mathrm{Cov}_{XY}\,\Sigma_Y^{-1}\mathrm{Cov}_{XY}^T)_{ij} = (\sigma_x^2 - \frac{a^2}{\sigma_y^2})\delta_{ij} + \gamma_x(\delta_{i,j-1} + \delta_{i,j+1})$. Thus, it is also the tridiagonal matrix, similar to Σ_X. Therefore, the determinants of both Σ_X and $\Sigma_X - \mathrm{Cov}_{XY}\,\Sigma_Y^{-1}\mathrm{Cov}_{XY}^T$ are given by Eq. (15), i.e., $\det(\Sigma_X) = \theta_n(\sigma_x^2, \gamma_x)$ and $\det(\Sigma_X - \mathrm{Cov}_{XY}\,\Sigma_Y^{-1}\mathrm{Cov}_{XY}^T) = \theta_n(\sigma_x^2 - \frac{a^2}{\sigma_y^2}, \gamma_x)$.

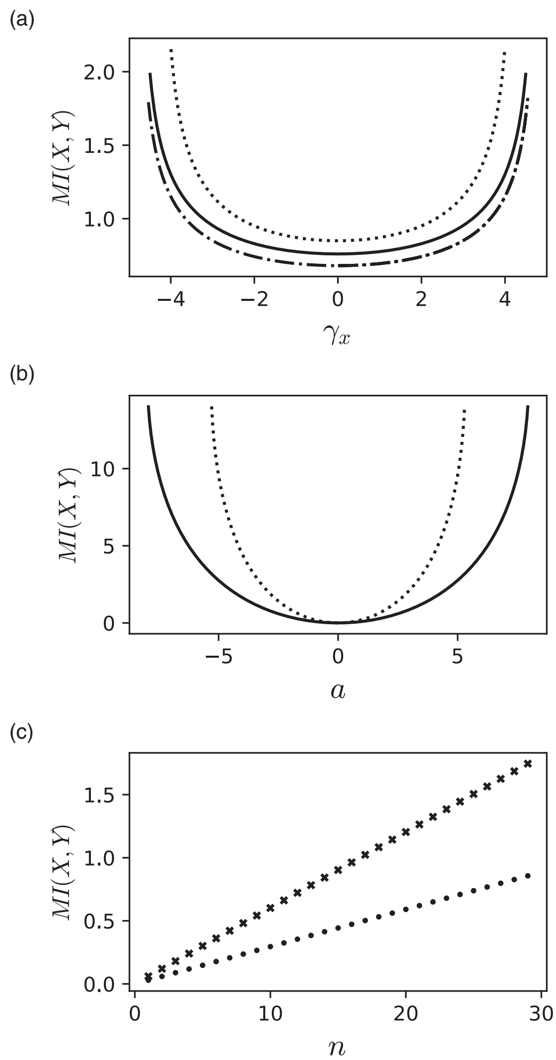


FIG. 3. MI in Eq. (22) as a function of γ_x, a, and n. (a) Solid line: $\sigma_x^2 = 10$, $\sigma_y^2 = 1$, $a = 1$, and $n = 10$; dotted line: $\sigma_x^2 = 9$, $\sigma_y^2 = 9$, $a = 3$, and $n = 10$; dash-dotted line: $\sigma_x^2 = 10$, $\sigma_y^2 = 10$, $a = 3$, and $n = 10$. (b) Solid line: $\sigma_x^2 = 9$, $\sigma_y^2 = 9$, $\gamma_x = 1$, and $n = 10$; dotted line: $\sigma_x^2 = 9$, $\sigma_y^2 = 4$, $\gamma_x = 1$, and $n = 10$. (c) Points: $\sigma_x^2 = 25$, $\sigma_y^2 = 1$, $a = 1$, and $\gamma_x = 1$; crosses: $\sigma_x^2 = 50$, $\sigma_y^2 = 1$, $a = 2$, and $\gamma_x = 2$.

The corresponding mutual information is given by the first line in Eq. (8) as

$$\mathrm{MI}(X, Y) = \frac{1}{\ln 4} \ln \left(\frac{\theta_n(\sigma_x^2, \gamma_x)}{\theta_n\left(\sigma_x^2 - \frac{a^2}{\sigma_y^2}, \gamma_x\right)} \right). \tag{22}$$

The dependence of MI on γ_x, a, and n is shown in Fig. 3. In the special case $n = k = 2$, we recover MI in Case B [see Eq. (14)], but with $b = \gamma_y = 0$.

IV. INFORMATION TRANSFER FOR NEURONS

Here, we apply the exact formula for MI in Eq. (8) to neuroscience; i.e., we derive mutual information relevant for information transfer in neural networks of a mammalian brain. Specifically, we find MI between activities of n weakly correlated presynaptic neurons with firing rates $\vec{f} = (f_1, f_2, \dots, f_n)$ and activity of a postsynaptic neuron with the firing rate Y. This situation can be depicted by the analogical graph as in Case D considered above. Below we perform computations and make comparison for two cases: (i) assuming that presynaptic firing rates $\vec{f}$ are lognormally distributed, and (ii) assuming that they are normally distributed. We note that although the mutual information between pre- and postsynaptic neural activities has been found many times in the past [22,24,37–40], none of those approaches used lognormally distributed variables (either Gaussian or a mixture of specific, like Poisson, and nonspecific discrete distributions was used).

The simplest model relating post- and presynaptic firing rates Y and $\vec{f}$ is [41,42]

$$Y = \sum_{i=1}^{n} w_i f_i, \tag{23}$$

where $\vec{w} = (w_1, w_2, \dots, w_n)$ are synaptic weights, and we assume that all of them are positive (all excitatory presynaptic neurons). For cortical neurons the number of synaptic contacts n is very large, typically in the range $n \sim 10^3$–10^4 [43]. The dependence of Y on $\vec{f}$ could, in principle, be nonlinear. We assume linearity, since it was shown that for a biophysically motivated nonlinear class of neurons (so-called class 2) the dependence of Y on $\vec{f}$ can become linear by incorporating adaptation in neural firing rates, which is often observed in brain networks [44–46].

A. MI for neural activities as lognormal variables

Our major assumption in this section is that both firing rates $\vec{f}$ and synaptic weights $\vec{w}$ have lognormal distributions, which is compatible with experimental data for neurons and synapses in the mammalian cerebral cortex [6–8]. Additionally, we assume that all f_i (and corresponding w_i) have the same parameters characterizing the distribution (uniformity assumption). Thus, the postsynaptic firing rate Y in Eq. (23) is a sum of lognormal random variables identically distributed, since each product $w_i f_i$ is lognormally distributed. It should be clearly said that Y has an unknown exact form of probability density. However, it has been numerically verified by others [47,48] that the sum of (uncorrelated and correlated) lognormal random variables can be approximated by lognormal distribution for large but finite n. Moreover, and more importantly, it has been proven that in the limiting case $n \mapsto \infty$ the sum of positively correlated lognormals also has a lognormal probability density [49,50]. (Standard central limit theorem does not apply here because of the correlations between the summands with heavy tails.) For these reasons, in this section, we assume that the postsynaptic neural activity Y has a lognormal distribution.

The goal is to find mutual information $\mathrm{MI}(\vec{f}, Y)$ between vectors $\vec{f}$ and Y. We assume that activities of presynaptic neurons [i.e., $\ln(f_i)$] are weakly correlated, and we take for their covariance matrix Σ_f the form given by the Kac-Murdock-Szegö matrix, i.e., $(\Sigma_f)_{ij} = \sigma_f^2 \gamma^{|i-j|}$, with $|\gamma| \ll 1$, where σ_f^2 is the variance for all $\ln(f_i)$. The corresponding variance for all $\ln(w_i)$ is denoted as σ_w^2, and the variance for postsynaptic activity $\ln(Y)$ is denoted as σ_Y^2. The latter depends on the parameters characterizing distributions of $\vec{f}$ and $\vec{w}$ (see below). Additionally, we assume that synaptic weights w_i are not correlated between themselves and that vectors $\vec{f}$ and $\vec{w}$ are not correlated either. These two assumptions are consistent with empirical observations, but are not essential for the derivation (in principle, they could be included). They make the final formula look simpler. Using Eq. (2), we can write expressions for various moments of f_i and w_i, which are used later:

$$\begin{aligned}\langle f_i\rangle &\equiv \langle f\rangle = e^{\mu_f+\frac{1}{2}\sigma_f^2},\\ \langle w_i\rangle &\equiv \langle w\rangle = e^{\mu_w+\frac{1}{2}\sigma_w^2},\\ \langle f_i^2\rangle &\equiv \langle f^2\rangle = e^{2\mu_f+2\sigma_f^2},\\ \langle w_i^2\rangle &\equiv \langle w^2\rangle = e^{2\mu_w+2\sigma_w^2},\\ \langle f_i f_j\rangle &= e^{2\mu_f+\sigma_f^2(1+c_{ij})},\end{aligned} \tag{24}$$

where we used the uniformity of the moments, and consequently the uniformity of the parameters $\mu_f = \langle \ln(f_i)\rangle$ and $\mu_w = \langle \ln(w_i)\rangle$. Additionally, c_{ij} is the correlation coefficient between $\ln(f_i)$ and $\ln(f_j)$, i.e., $c_{ij}\sigma_f^2 = \langle(\ln(f_i) - \mu_f)(\ln(f_j) - \mu_f)\rangle = (\Sigma_f)_{ij}$. The latter means that $c_{ij} = \gamma^{|i-j|}$.

To determine MI we need the variance of $\ln(Y)$ and the covariance matrix between $\ln(Y)$ and $\ln(f_i)$. This is accomplished by finding the first two moments of Y in terms of the parameters characterizing $\vec{f}$ and $\vec{w}$, and matching them to the first two moments of the assumed lognormal form of Y (so-called Wilkinson method; Refs. [47,48]). We have

$$\langle Y\rangle = \sum_{i=1}^{n}\langle w_i\rangle\langle f_i\rangle,$$

and

$$\langle Y^2\rangle = \sum_{i=1}^{n}\langle w_i^2\rangle\langle f_i^2\rangle + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\langle w_i\rangle\langle w_j\rangle\langle f_i f_j\rangle,$$

which after using Eqs. (24) yields

$$\begin{aligned}\langle Y\rangle &= n e^{\mu_f+\mu_w} e^{(\sigma_f^2+\sigma_w^2)/2},\\ \langle Y^2\rangle &= e^{2(\mu_f+\mu_w)} e^{\sigma_f^2+\sigma_w^2}\\ &\quad\times\left(n e^{\sigma_f^2+\sigma_w^2} + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} e^{\sigma_f^2 c_{ij}}\right).\end{aligned} \tag{25}$$

On the other hand, for lognormally distributed Y, we have $\langle Y\rangle = e^{\mu_Y+\frac{1}{2}\sigma_Y^2}$ and $\langle Y^2\rangle = e^{2\mu_Y+2\sigma_Y^2}$, where μ_Y and σ_Y^2 are the mean and the variance of $\ln(Y)$. This allows us to find σ_Y^2 in terms of the parameters for $\vec{f}$ and $\vec{w}$ as

$$\sigma_Y^2 = \ln\left(\frac{1}{n}e^{(\sigma_f^2+\sigma_w^2)} + \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} e^{\sigma_f^2 c_{ij}}\right). \tag{26}$$

To find the covariance $\mathrm{Cov}_{Y\vec{f}} \equiv \mathrm{Cov}[\ln(Y), \ln(\vec{f})]$ between $\ln(Y)$ and $\ln(f_i)$, we first determine the covariance $\mathrm{Cov}(Y, f_i)$ between Y and f_i. We have $\mathrm{Cov}(Y, f_i) = \langle Y f_i\rangle - \langle Y\rangle\langle f_i\rangle = \langle w_i\rangle\langle f_i^2\rangle - \langle Y\rangle\langle f_i\rangle + \sum_{j\neq i}^{n}\langle w_j\rangle\langle f_i f_j\rangle$. All these moments are given above, and we obtain

$$\begin{aligned}\mathrm{Cov}(Y, f_i) &= e^{m_w+\sigma_w^2/2} e^{2(m_f+\sigma_f^2)}\\ &\quad\times\left[1 - e^{-\sigma_f^2}\left(n - \sum_{j\neq i}^{n} e^{\sigma_f^2 c_{ij}}\right)\right],\end{aligned} \tag{27}$$

which after using Eq. (2) yields the covariance vector of interest:

$$(\mathrm{Cov}_{Y\vec{f}})_i = \ln\left[\frac{1}{n}\left(e^{\sigma_f^2} + \sum_{j\neq i}^{n} e^{\sigma_f^2 c_{ij}}\right)\right]. \tag{28}$$

The general formulas in Eqs. (26) and (28) can be expanded for weak presynaptic neural correlations, i.e., $e^{\sigma_f^2 c_{ij}} \approx 1 + \sigma_f^2 c_{ij}$ for $|\gamma| \ll 1$, with the help of $\sum_{j\neq i} c_{ij} = \gamma[\delta_{i1} + \delta_{in} + 2(1-\delta_{i1})(1-\delta_{in})] + O(\gamma^2)$. This leads to

$$\begin{aligned}\sigma_Y^2 &= \ln\left\{1 + \frac{1}{n}\left[e^{(\sigma_f^2+\sigma_w^2)} - 1\right]\right\}\\ &\quad + \frac{2\gamma\sigma_f^2\left(1-\frac{1}{n}\right)^2}{n + \left[e^{(\sigma_f^2+\sigma_w^2)} - 1\right]} + O(\gamma^2)\end{aligned} \tag{29}$$

and

$$\begin{aligned}(\mathrm{Cov}_{Y\vec{f}})_i &= \ln\left[1 + \frac{1}{n}\left(e^{\sigma_f^2} - 1\right)\right]\\ &\quad + \frac{\gamma\sigma_f^2[\delta_{i1} + \delta_{in} + 2(1-\delta_{i1})(1-\delta_{in})]}{n + \left[e^{\sigma_f^2} - 1\right]} + O(\gamma^2).\end{aligned} \tag{30}$$

Note that the covariance vector $\mathrm{Cov}_{Y\vec{f}}$ is finite even when there are no correlations between presynaptic neural activities ($\gamma \mapsto 0$); however, in the limit $n \mapsto \infty$ it vanishes.

The last step is to find the value of $\mathrm{Cov}_{Y\vec{f}}^T \Sigma_f^{-1} \mathrm{Cov}_{Y\vec{f}}$, using approximation to Eq. (18) for small γ, i.e., $(\Sigma_f^{-1})_{ij} = \frac{1}{\sigma_f^2}[\delta_{ij} - \gamma(\delta_{i,j-1} + \delta_{i,j+1})] + O(\gamma^2)$. The result is

$$\mathrm{Cov}_{Y\vec{f}}^T \Sigma_f^{-1} \mathrm{Cov}_{Y\vec{f}} = \frac{n(1-2\gamma)}{\sigma_f^2}\ln^2\left[1 + \frac{1}{n}\left(e^{\sigma_f^2} - 1\right)\right] + \frac{4(n-1)\gamma}{\left(e^{\sigma_f^2} + n - 1\right)}\ln\left[1 + \frac{1}{n}\left(e^{\sigma_f^2} - 1\right)\right] + O(\gamma^2). \tag{31}$$

This term also vanishes in the limit of very large number of presynaptic neurons, $n \mapsto \infty$.

Mutual information between pre- and postsynaptic neuronal activities is given by the first line of Eq. (8), with $\det(\Sigma_Y) = \sigma_Y^2$ and $\det(\Sigma_Y - \mathrm{Cov}_{Y\vec{f}}^T \Sigma_f^{-1} \mathrm{Cov}_{Y\vec{f}}) = \sigma_Y^2 - \mathrm{Cov}_{Y\vec{f}}^T \Sigma_f^{-1} \mathrm{Cov}_{Y\vec{f}}$, which are given by Eqs. (29) and (31). MI takes the form

$$\mathrm{MI}(\vec{f},Y)_{\ln} = \frac{1}{\ln 4}\ln\left(\frac{\ln\left\{1+\frac{1}{n}\left[e^{(\sigma_f^2+\sigma_w^2)}-1\right]\right\}}{\left(\ln\left\{1+\frac{1}{n}\left[e^{(\sigma_f^2+\sigma_w^2)}-1\right]\right\}-\frac{n}{\sigma_f^2}\ln^2\left[1+\frac{1}{n}\left(e^{\sigma_f^2}-1\right)\right]\right)}\left[1+\gamma g\left(n,\sigma_f^2,\sigma_w^2\right)+O(\gamma^2)\right]\right), \tag{32}$$

where the function $g(n, \sigma_f^2, \sigma_w^2)$ is defined as

$$g\left(n,\sigma_f^2,\sigma_w^2\right) = \frac{2n\ln^2\left(1+\frac{1}{n}\left[e^{\sigma_f^2}-1\right]\right)\left[\left(\frac{2(1-1/n)}{\left(e^{\sigma_f^2}+n-1\right)\ln\left[1+\frac{1}{n}\left(e^{\sigma_f^2}-1\right)\right]}-\frac{1}{\sigma_f^2}\right)\ln\left\{1+\frac{1}{n}\left[e^{(\sigma_f^2+\sigma_w^2)}-1\right]\right\}-\frac{(1-1/n)}{\left(e^{\sigma_f^2+\sigma_w^2}+n-1\right)}\right]}{\ln\left\{1+\frac{1}{n}\left[e^{(\sigma_f^2+\sigma_w^2)}-1\right]\right\}\left(\ln\left\{1+\frac{1}{n}\left[e^{(\sigma_f^2+\sigma_w^2)}-1\right]\right\}-\frac{n}{\sigma_f^2}\ln^2\left(1+\frac{1}{n}\left[e^{\sigma_f^2}-1\right]\right)\right)}, \tag{33}$$

with $\sigma_f^2 = \ln[1 + \mathrm{Var}(f)/\langle f\rangle^2]$ and $\sigma_w^2 = \ln[1 + \mathrm{Var}(w)/\langle w\rangle^2]$, where $\mathrm{Var}(f) = \langle f^2\rangle - \langle f\rangle^2$ and $\mathrm{Var}(w) = \langle w^2\rangle - \langle w\rangle^2$ [see Eq. (24)]. Equations (32) and (33) constitute the major result of this paper. These two equations with quite complex formulas give us $\mathrm{MI}_{\ln}$ for any number of presynaptic neurons n, as well as for arbitrary levels of variability in presynaptic firing rates σ_f and synaptic weights σ_w. Note that $\mathrm{MI}_{\ln}$ does not depend at all on μ_f and μ_w. For very noisy synapses, i.e., for $\sigma_w \mapsto \infty$, both factors under the logarithm in Eq. (32) tend to 1 (function $g \mapsto 0$), and hence mutual information $\mathrm{MI}_{\ln} \mapsto 0$. The function g in Eq. (33) is generally positive (unless presynaptic activity has very high variability, $\sigma_f^2 \gg 1$, but then MI already approaches infinity). Specifically, for $\sigma_f^2 \ll 1$ and arbitrary σ_w^2, we get a relatively simple form for this function, $g \approx 2\sigma_f^2/\{n\ln[1 + (e^{\sigma_w^2} - 1)/n]\}$, which implies that positive correlations between presynaptic neurons ($\gamma > 0$) increase MI, while their negative correlations ($\gamma < 0$) are detrimental for MI.

Of particular interest is the expression for MI when neuronal activities and synaptic weights have low variability, such that both $\sigma_f^2 \ll 1$ and $\sigma_w^2 \ll 1$. In this case, Eqs. (32) and (33) simplify significantly, and we obtain the following simple expression for MI:

$$\mathrm{MI}(\vec{f},Y)_{\ln} \approx \frac{\ln\left\{\left[1+\frac{\sigma_f^2}{\sigma_w^2}\right]\left[1+\frac{2\gamma\sigma_f^2}{(\sigma_w^2+\sigma_f^2)}+O(\gamma^2)\right]\right\}}{\ln 4}. \tag{34}$$

Note that MI in this limit is independent of the number of presynaptic neurons n. Moreover, Eq. (34) implies that MI grows with increasing the variance of presynaptic neurons σ_f^2, and it decays with increasing the synaptic weight variance σ_w^2. This suggests that the ratio σ_f^2/σ_w^2 can be interpreted as the signal-to-noise ratio, with the variability in synaptic weights serving as the noise.

Equation (34) can be rewritten in terms of means and variances of presynaptic firings $\langle f\rangle$, $\mathrm{Var}(f)$, and of synaptic weights $\langle w\rangle$, $\mathrm{Var}(w)$, using Eq. (24). For $\sigma_f^2 \ll 1$ and $\sigma_w^2 \ll 1$, we have $\sigma_f^2 \approx \mathrm{Var}(f)/\langle f\rangle^2$ and $\sigma_w^2 \approx \mathrm{Var}(w)/\langle w\rangle^2$, which leads to

$$\mathrm{MI}(\vec{f},Y)_{\ln} \approx \frac{1}{\ln 4}\ln\left[\left(1+\frac{\langle w\rangle^2\mathrm{Var}(f)}{\langle f\rangle^2\mathrm{Var}(w)}\right)\times\left(1+\frac{2\gamma\langle w\rangle^2\mathrm{Var}(f)}{[\langle f\rangle^2\mathrm{Var}(w)+\langle w\rangle^2\mathrm{Var}(f)]}+O(\gamma^2)\right)\right]. \tag{35}$$

From this formula it is clear that in the limit of low neuronal and synaptic variabilities, MI scales with their relative ratio.

B. Comparison between neuronal MI with lognormal and normal neural activities

In this section we take a more traditional viewpoint and derive neuronal MI treating neural activities as Gaussian variables. The resulting expression for Gaussian MI is then compared with the lognormal MI obtained in Eq. (32).

It should be said from the outset that treating positive firing rates as normal variables is a little unrealistic, since there is always some nonzero probability that the rates can become negative. Obviously, that likelihood is extremely small if neural activity variances are much smaller than their mean levels.

We assume explicitly that firing rates $\vec{f}$ in Eq. (23) have normal distribution with uniform mean $\langle f\rangle$ and variance $\mathrm{Var}(f)$, and they are weakly correlated. No specific distribution for synaptic weights $\vec{w}$ is assumed, except that it has uniform mean $\langle w\rangle$ and finite but not too large variance $\mathrm{Var}(w)$ (it could be Gaussian too, but it is not necessary for the arguments below). We also take the limit $n \mapsto \infty$, implying a very large number of presynaptic neurons. That limit allows us to use the central limit theorem in Eq. (23) and claim that the postsynaptic firing rate Y has a normal distribution. The central limit theorem is permissible here because, in contrast to the lognormal case, the summands in Eq. (23) have short tails [51].

The goal is to find mutual information between Gaussian $\vec{f}$ and Y, denoted as $\mathrm{MI}(\vec{f},Y)_g$, which is given by the first line of Eq. (8), where $X \mapsto \vec{f}$. This means that now $\Sigma_Y = \langle Y^2\rangle - \langle Y\rangle^2$, the covariance vector between Y and $\vec{f}$ is $\mathrm{Cov}(Y, f_i) = \langle Yf_i\rangle - \langle Y\rangle\langle f_i\rangle$, and the covariance matrix Σ_f

is between f_i and f_j, i.e., $(\Sigma_f)_{ij} = \langle f_i f_j\rangle - \langle f_i\rangle\langle f_j\rangle$. Again, for the latter matrix we take the Kac-Murdock-Szegö form, though with different coefficients, $(\Sigma_f)_{ij} = c_0\kappa^{|i-j|}$, where the variance in presynaptic activities $c_0 = \langle f^2\rangle - \langle f\rangle^2$ and $|\kappa| \ll 1$. The inverse of Σ_f is $(\Sigma_f^{-1})_{ij} = [\delta_{ij} - \kappa(\delta_{i,j-1} + \delta_{i,j+1})]/c_0 + O(\kappa^2)$. The calculation in many respects is similar to the one in the previous section, but a little easier. Specifically, we obtain

$$\Sigma_Y = n[\langle w^2\rangle\langle f^2\rangle - \langle w\rangle^2\langle f\rangle^2] + 2(n-1)\kappa\langle w\rangle^2\mathrm{Var}(f) \tag{36}$$

and

$$\mathrm{Cov}(Y, f_i) = \langle w\rangle\mathrm{Var}(f)(1 + \kappa[\delta_{i1} + \delta_{in} + 2(1-\delta_{i1})(1-\delta_{in})]) + O(\kappa^2). \tag{37}$$

With the help of these relations we find

$$\mathrm{Cov}(Y, \vec{f})^T\Sigma_f^{-1}\mathrm{Cov}(Y, \vec{f}) = \langle w\rangle^2\mathrm{Var}(f)[n + 2(n-1)\kappa] + O(\kappa^2), \tag{38}$$

which leads to the Gaussian MI in the form

$$\mathrm{MI}(\vec{f}, Y)_g \approx \frac{1}{\ln 4}\ln\left[\left(1 + \frac{\langle w\rangle^2\mathrm{Var}(f)}{\langle f^2\rangle\mathrm{Var}(w)}\right) \times\left(1 + \frac{2\kappa\langle w\rangle^2\mathrm{Var}(f)}{[\langle f^2\rangle\mathrm{Var}(w) + \langle w\rangle^2\mathrm{Var}(f)]} + O(\kappa^2)\right)\right]. \tag{39}$$

Comparing the above MI for Gaussian distributions with the corresponding MI for lognormal in Eq. (32), we can notice significant differences, as σ_f^2 and σ_w^2 depend in a nonlinear way on $\mathrm{Var}(f)$ and $\mathrm{Var}(w)$. Generally $\mathrm{MI}_{\ln}$ is greater than MI_g, and the discrepancy between them grows with increasing presynaptic firing rate variability and with decreasing synaptic noise (Fig. 4). Additionally, $\mathrm{MI}_{\ln}$ diverges for some finite $\sqrt{\mathrm{Var}(f)}/\langle f\rangle$ for a fixed level of $\sqrt{\mathrm{Var}(w)}/\langle w\rangle$ (and vice versa), whereas MI_g is always finite for finite means and variances. However, in the case when variabilities in f and w are weak, i.e., for $\sigma_f^2, \sigma_w^2 \ll 1$, we see that the lognormal formula for MI in Eq. (35) is exactly the same as the one in Eq. (39) for Gaussian MI (note that in this limit $\langle f^2\rangle \approx \langle f\rangle^2$ and $\kappa \mapsto \gamma$). This reflects a simple fact that, for small σ_f^2 and σ_w^2, lognormal distributions can be approximated by normal distributions [52].

C. Implications for information processing in cortical brain networks

The results shown in Fig. 4 indicate that high variability in neural activities fosters information transfer between neurons. Morever, there seems to be a fundamental difference between short-tailed and heavy-tailed distributions of neural activities, with the latter having much bigger transmission impact for a given ratio of standard deviation to the mean activity. This is also true for the variability in synaptic weights: synapses with broad heavy-tailed distributions generally provide higher mutual information than synapses with bell-shaped distributions [Fig. 4(b)].

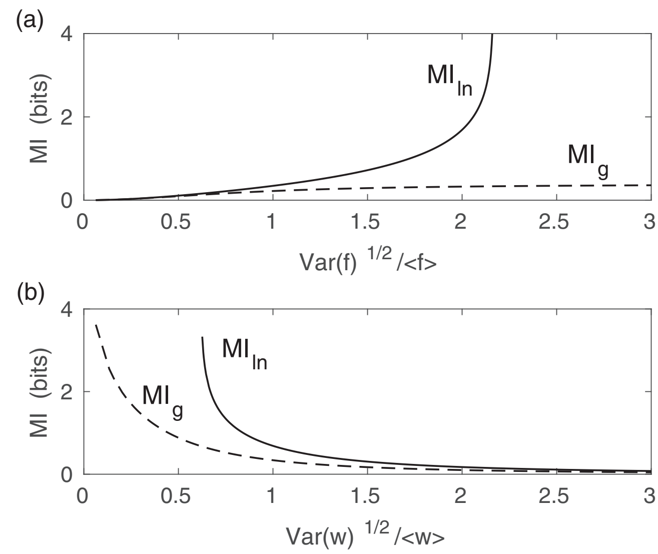


FIG. 4. Neuronal MI as a function of presynaptic activity (a) and synaptic weights (b). Parameters used are as follows: (a) $\mathrm{Var}(w)/\langle w\rangle^2 = 1.4$ [6]; (b) $\mathrm{Var}(f)/\langle f\rangle^2 = 1.5$ [7]; and for both $n = 8000$ [43].

Our theoretical result, presented in Fig. 4, and its interpretation are compatible with experimental data coming from cortical networks [53,54]. These papers demonstrated that diversity of brain dynamics, i.e., activity patterns with broad heavy-tailed distributions, tend to improve information transfer between different cortical regions [53,54]. More generally, these conclusions are also in line with empirical evidence showing that neuronal representations in mammalian brains are high dimensional, meaning that neurons exhibit a diversity of context-dependent activities, which is important functionally [55]. Taken together, this may suggest that evolution prefers brains with heterogeneous dynamics to optimize information processing (e.g., Ref. [8]).

V. CONCLUSIONS

In this paper we used analytical expression for mutual information between random vectors with lognormal distribution to obtain closed-form expressions of MI for different networks of interacting elements with specific covariance matrices, mostly sparse with a high degree of symmetry. These formulas may be helpful in many practical applications in engineering and biology. Additionally, we applied these results to information transfer in neural networks of the mammalian cerebral cortex. Specifically, we derived an analytical formula for MI of a neuron receiving many correlated synaptic inputs that are lognormally distributed, and we compared such MI with the case when the total synaptic input is normally distributed. Interestingly, mutual information in the first case of lognormal input can be significantly greater than that in the second case with Gaussian variables.

ACKNOWLEDGMENT

The work was supported by the Polish National Science Centre (NCN) under Grant No. 2021/41/B/ST3/04300 (J.K.).

APPENDIX: INVERSE AND DETERMINANT OF TRIDIAGONAL MATRIX AND PROOF OF EQ. (17)

To compute $\mathrm{Cov}_{XY}^T \Sigma_X^{-1} \mathrm{Cov}_{XY}$ in Eq. (17), we need first to find the inverse of the tridiagonal matrix Σ_X. Let us consider a general $n \times n$ tridiagonal matrix A with all diagonal elements equal to t and off-diagonal elements all equal to s. Inverse elements of A are given by [34,56]

$$(A^{-1})_{ij} = (-1)^{i+j} s^{j-i} \theta_{i-1} \phi_{j+1} / \theta_n \tag{A1}$$

for $i \leqslant j$, where $\theta_n = \det(A)$. Matrix A^{-1} is symmetric, and hence $(A^{-1})_{ij} = (A^{-1})_{ji}$. In Eq. (A1), θ_i and ϕ_i are functions of t and s. Additionally, functions θ_i satisfy the following recurrence relations:

$$\theta_i = t\theta_{i-1} - s^2 \theta_{i-2}, \tag{A2}$$

with $\theta_0 = 1$ and $\theta_1 = t$. Similarly, functions ϕ_i satisfy the following recurrence relations:

$$\phi_i = t\phi_{i+1} - s^2 \phi_{i+2}, \tag{A3}$$

with $\phi_{n+1} = 1$ and $\phi_n = t$.

Both of the recurrence relations, Eqs. (A2) and (A3), can be solved by a standard substitution of $\theta_i = r^i$, with unknown r. The solution for r is $r_\pm = \frac{1}{2}(t \pm \sqrt{t^2 - 4s})$. The specific solutions for θ_i and ϕ_i are given by linear combinations of r_- and r_+, with coefficients dependent on the boundary conditions in Eqs. (A2) and (A3). It is easy to show that

$$\theta_i(t, s) = \frac{[(t + \sqrt{t^2 - 4s^2})^{i+1} - (t - \sqrt{t^2 - 4s^2})^{i+1}]}{2^{i+1}\sqrt{t^2 - 4s^2}} \tag{A4}$$

and $\phi_i = \theta_{n-i+1}$, for $i = 1, 2, \ldots, n$. The functions $\theta_i(t, s)$ can be interpreted as determinants of reduced $i \times i$ matrices generated from the original matrix A, with all diagonal elements t and all off-diagonal elements s.

Now we can determine $\mathrm{Cov}_{XY}^T \Sigma_X^{-1} \mathrm{Cov}_{XY}$ in Eq. (17). We have

$$\begin{aligned}
&\sum_{i,j=1}^{n} (\mathrm{Cov}_{XY}^T)_i (\Sigma_X^{-1})_{ij} (\mathrm{Cov}_{XY})_j \\
&= a^2 \sum_{i,j=1}^{n} (\Sigma_X^{-1})_{ij} \\
&= a^2 \left(\sum_{i=1}^{n} (\Sigma_X^{-1})_{ii} + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (\Sigma_X^{-1})_{ij} \right),
\end{aligned}$$

where the symmetric matrix Σ_X^{-1} has the following elements:

$$(\Sigma_X^{-1})_{ij} = \frac{(-1)^{i+j} \gamma^{j-i}}{\det(\Sigma_X)} \theta_{i-1} \theta_{n-j}, \tag{A5}$$

for $i \leqslant j$. These sums can be executed, and after lengthy computations one can obtain Eq. (17) in the main text.

[1] E. Limpert, W. A. Stahel, and M. Abbt, Log-normal distributions across the sciences: Keys and clues, BioScience **51**, 341 (2001).

[2] *Lognormal Distributions: Theory and Applications*, edited by E. L. Crow and K. Shimizu (Dekker, New York, 1988).

[3] D. Sornette and L. Knopoff, The paradox of the expected time until the next earthquake, Bull. Seismol. Soc. Am. **87**, 789 (1997).

[4] F. Black and M. Scholes, The pricing of options and corporate liabilities, J. Political Econ. **81**, 637 (1973).

[5] J.-P. Bouchaud and M. Potters, *Theory of Financial Risks* (Cambridge University Press, Cambridge, England, 2000).

[6] S. Song, P. J. Sjöström, M. Reigl, S. Nelson, and D. B. Chklovskii, Highly nonrandom features of synaptic connectivity in local cortical circuits, PLoS Biol. **3**, e68 (2005).

[7] T. Hromadka, M. R. DeWeese, and A. M. Zador, Sparse representation of sounds in the unanesthetized auditory cortex, PLoS Biol. **6**, e16 (2008).

[8] G. Buzsaki and K. Mizuseki, The log-dynamic brain: How skewed distributions affect network operations, Nat. Rev. Neurosci. **15**, 264 (2014).

[9] A. Tang, H. Liu, G. Liu, Y. Zhong, L. Wang, Q. Lu, J. Wang, and Y. Shen, Lognormal distribution of local strain: A universal law of plastic deformation in material, Phys. Rev. Lett. **124**, 155501 (2020).

[10] A. Corral, F. Udina, and E. Arcaute, Truncated lognormal distributions and scaling in the size of naturally defined population clusters, Phys. Rev. E **101**, 042312 (2020).

[11] P. Sobkowicz, M. Thelwall, K. Buckley *et al.*, Lognormal distributions of user post lengths in internet discussions—a consequence of the Weber-Fechner law? EPJ Data Sci. **2**, 2 (2013).

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley & Sons, New York, 2006).

[13] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. **27**, 379 (1948).

[14] C. Lin, S. Xiong, and Y. Chen, Mutual information maximizing GAN inversion for real face with identity preservation, J. Visual Commun. Image Represent. **87**, 103566 (2022).

[15] M. Gabrie, A. Manoel, C. Luneau *et al.*, Entropy and mutual information in models of deep neural networks, J. Stat. Mech. (2019) 124014.

[16] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, Rep. Prog. Phys. **75**, 126001 (2012).

[17] G. De Tomasi, S. Bera, J. H. Bardarson, and F. Pollmann, Quantum mutual information as a probe for many-body localization, Phys. Rev. Lett. **118**, 016804 (2017).

[18] N. Iizuka and D. Kabat, Mutual information in Hawking radiation, Phys. Rev. D **88**, 084010 (2013).

[19] P. Golub, A. Antalik, P. Beran, and J. Brabec, Mutual information prediction for strongly correlated systems, Chem. Phys. Lett. **813**, 140297 (2023).

[20] Q. Wang, Y. Xiao, and D. Meng, Identification of structural key genes of mutual information gene networks of brain tumor, Phys. A (Amsterdam, Neth.) **608**, 128322 (2022).

[21] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, Information transduction capacity of noisy biochemical signaling networks, Science **334**, 354 (2011).

[22] F. Rieke, D. Warland, R. de Ruyter, and W. Bialek, *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, MA, 1999).

[23] G. A. Darbellay and I. Vajda, Entropy expressions for multivariate continuous distributions, IEEE Trans. Inf. Theory **46**, 709 (2000).

[24] L. Paninski, Estimation of entropy and mutual information, Neural Comput. **15**, 1191 (2003).

[25] C. Granero-Belinchon, S. G. Roux, P. Abry, and N. B. Garnier, Probing high-order dependencies with information theory, IEEE Trans. Signal Process. **67**, 3796 (2019).

[26] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, England, 1991).

[27] J. R. Silvester, Determinants of block matrices, Math. Gazette **84**, 460 (2000).

[28] R. Milo *et al.*, Network motifs: simple building blocks of complex networks, Science **298**, 824 (2002).

[29] R. Albert and A. L. Barabasi, Statistical mechanics of complex networks, Rev. Mod. Phys. **74**, 47 (2002).

[30] M. E. J. Newman, The structure and function of complex networks, SIAM Rev. **45**, 167 (2003).

[31] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).

[32] M. Peltomaki and M. Alava, Correlations in bipartite collaboration networks, J. Stat. Mech. (2006) P01010.

[33] J. P. Sethna, *Entropy, Order Parameters, and Complexity* (Oxford University, Oxford, 2006).

[34] G. Y. Hu and R. F. O'Connell, Analytical inversion of symmetric tridiagonal matrices, J. Phys. A **29**, 1511 (1996).

[35] L. G. Molinari, Determinants of block tridiagonal matrices, Lin. Alg. Its Appl. **429**, 2221 (2008).

[36] M. Kac, W. L. Murdock, and G. Szegö, On the eigenvalues of certain Hermitian forms, J. Ration. Mech. Anal. **2**, 767 (1953).

[37] A. Borst and F. E. Theunissen, Information theory and neural coding, Nat. Neurosci. **2**, 947 (1999).

[38] N. Brunel and J. P. Nadal, Mutual information, Fisher information and population coding, Neural Comput. **10**, 1731 (1998).

[39] S. Panzeri, S. R. Schultz, A. Treves, and E. T. Rolls, Correlations and the encoding of information in the nervous system, Proc. R. Soc. London, Ser. B **266**, 1001 (1999).

[40] W. B. Levy and R. A. Baxter, Energy-efficient neuronal computation via quantal synaptic failures, J. Neurosci. **22**, 4746 (2002).

[41] J. A. Hertz, A. S. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation* (Westview, Santa Fe, NM, 1991).

[42] P. Dayan and L. F. Abbott, *Theoretical Neuroscience* (MIT Press, Cambridge, MA, 2001).

[43] V. Braitenberg and A. Schuz, *Cortex: Statistics and Geometry of Neuronal Connectivity* (Springer, Berlin, 1998).

[44] G. B. Ermentrout, Linearization of F-I curves by adaptation, Neural Comput. **10**, 1721 (1998).

[45] G. B. Ermentrout and D. H. Terman, *Mathematical Foundations of Neuroscience* (Springer, New York, 2010).

[46] J. Karbowski, Energetics of stochastic BCM type synaptic plasticity and storing of accurate information, J. Comput. Neurosci. **49**, 71 (2021).

[47] A. A. Abu-Dayya and N. C. Beaulieu, Outage probabilities in the presence of correlated lognormal interferers, IEEE Trans. Veh. Technol. **43**, 164 (1994).

[48] N. B. Mehta, J. Wu, A. F. Molisch, and J. Zhang, Approximating a sum of random variables with a lognormal, IEEE Trans. Wireless Commun. **6**, 2690 (2007).

[49] S. S. Szyszkowicz and H. Yanikomeroglu, Limit theorem on the sum of identically distributed equally and positively correlated joint lognormals, IEEE Trans. Commun. **57**, 3538 (2009).

[50] N. C. Beaulieu, An extended limit theorem for correlated lognormal sums, IEEE Trans. Commun. **60**, 23 (2012).

[51] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 2007).

[52] M. Romeo, V. Da Costa, and F. Bardou, Broad distribution effects in sums of lognormal random variables, Eur. Phys. J. B **32**, 513 (2003).

[53] W. L. Shew *et al.*, Information capacity and transmission are maximized in balanced cortical networks with neuronal avalanches, J. Neurosci. **31**, 55 (2011).

[54] E. D. Fagerholm *et al.*, Cortical entropy, mutual information and scale-free dynamics in waking mice, Cereb. Cortex **26**, 3945 (2016).

[55] S. Fusi, E. K. Miller, and M. Rigotti, Why neurons mix: High dimensionality for higher cognition, Curr. Opin. Neurobiol. **37**, 66 (2016).

[56] R. A. Usmani, Inverse of a tridiagonal Jacobi matrix, Lin. Alg. Its Appl. **212-213**, 413 (1994).