Population coding in neuronal systems with correlated noise

Haim Sompolinsky, Hyoungsoo Yoon, Kukjin Kang, and Maoz Shamir
Racah Institute of Physics and Center for Neural Computation, The Hebrew University of Jerusalem, Jerusalem 91904, Israel
(Received 5 April 2001; published 17 October 2001)

Neuronal representations of external events are often distributed across large populations of cells. We study the effect of correlated noise on the accuracy of these neuronal population codes. Our main question is whether the inherent error in the population code can be suppressed by increasing the size of the population N in the presence of correlated noise. We address this issue using a model of a population of neurons that are broadly tuned to an angular variable in two dimensions. The fluctuations in the neuronal activities are modeled as Gaussian noises with pairwise correlations that decay exponentially with the difference between the preferred angles of the correlated cells. We assume that the system is broadly tuned, which means that both the correlation length and the width of the tuning curves of the mean responses span a substantial fraction of the entire system length. The performance of the system is measured by the Fisher information (FI), which bounds its estimation error. By calculating the FI in the limit of a large N, we show that positive correlations decrease the estimation capability of the network, relative to the uncorrelated population. The information capacity saturates to a finite value as the number of cells in the population grows. In contrast, negative correlations substantially increase the information capacity of the neuronal population. These results are supplemented by the effect of correlations on the mutual information of the system. Our analysis provides an estimate of the effective number of statistically independent degrees of freedom, denoted N_{eff}, that a large correlated system can have. According to our theory N_{eff} remains finite in the limit of a large N. Estimating the parameters of the correlations and tuning curves from experimental data in some cortical areas that code for angles, we predict that the number of effective degrees of freedom embedded in localized populations in these areas is less than or of the order of $\approx 10^2$.

DOI: 10.1103/PhysRevE.64.051904 PACS number(s): 87.18.Sn

I. INTRODUCTION

In many neural systems, information regarding sensory inputs or (intended) motor outputs is distributed across large populations of neurons [1–6]. It is generally believed that one of the main features of neuronal population codes is their redundancy [7–9], and that this redundancy serves to compensate for the inherent noise caused by the stochasticity of the neuronal responses. The ability to overcome neuronal noise by pooling a large number of responses generally holds for an ensemble of neurons whose variabilities are statistically independent. This is far from obvious in situations where the ''noisy'' part of the neuronal activity is correlated within the population.

Naively it seems reasonable to assume that when the noise is correlated, population averaging would not be effective in suppressing the noise, hence the information capacity of the population should be limited even when the number of participating cells is large. This intuition has been supported by recent studies which found that averaging the responses of a uniformly correlated population does not suppress the inherent error [10]. On the other hand, a recent analytical investigation by Abbott and Dayan [11] showed that uniform positive correlations increase the information capacity of the population. They further concluded that more generally, even if the information is not enhanced compared to the uncorrelated situation, it still increases linearly with the size of the population, so that pooling a large population is effective in improving the accuracy of the extracted information even in the presence of correlations. Other studies based on simulations have also found that positive correlations enhance the accuracy of the population codes [12]. Since cross correlations in neuronal activity are frequently observed in sensory and motor cortical areas [13–15], understanding the effect of noise correlation in biologically relevant situations is of great importance.

In this paper we present an analytical study of the effect of noise correlations on the population coding of a pool of cells that encode a single variable, which for convenience is taken to represent an angle, such as the orientation of a visual stimulus or the direction of an arm movement in a plane. We assume that the correlation in the noisy activity of the neurons follows the multivariate Gaussian distribution. As will be shown, the effect of correlations depends crucially on their spatial dependence. Here, we will assume that the correlations between a pair of neurons decay exponentially as the difference between their functional spatial coordinates increases. In the present case, the functional coordinate of a neuron is its preferred angle, which is the angle that elicits the strongest mean response. We investigate the accuracy of the information in the population in biologically relevant parameter regimes, using the frameworks of the Fisher information (FI) and the Shannon mutual information (MI).

The paper is organized as follows: We begin by analyzing a simple example in Sec. II that clarifies the special case of uniform correlations, a topic that has been discussed extensively in the literature [10,11,16]. To relate directly to some of the previous treatments of this problem, we will consider the task of discrimination between two values of a signal by a linear rule, rather than the problem of general stimulus estimation that will be the focus of the rest of the paper. In Sec. III we describe the main model investigated in this

work. This model consists of neurons coding an angle with spatially dependent correlations. The properties of the FI of this system in various parameter regimes are studied in Sec. IV. In Sec. V we present the results for the effect of correlations on the MI of the system. Discussion of the results and their implications for concrete experimental systems is given in Sec. VI.

Partial preliminary results of this work have been reported elsewhere [17].

II. UNIFORM NOISE CORRELATIONS

A. Linear discrimination

We begin our analysis of the effect of correlations by considering the problem of discriminating two stimuli on the basis of noisy responses of a correlated pool of N neurons. We assume that each neuron fires on average at a rate f_i^+, $(i=1,2,\ldots,N)$, for one stimulus, and at f_i^- for the other. We further assume uniform pairwise correlations among neurons in the pool, namely,

$$C_{ij}=a[\delta_{ij}+c(1-\delta_{ij})], \tag{1}$$

where a is the variance of the responses of each neuron and c is the correlation coefficient of the pairwise fluctuations. Assuming Gaussian statistics, the probability distribution of an activity profile of the population is

$$P(\mathbf{r}|\pm)=Z^{-1}\exp\left(-\frac{1}{2}\sum_{ij}(r_i-f_i^{\pm})C_{ij}^{-1}(r_j-f_j^{\pm})\right), \tag{2}$$

where $\mathbf{r}=\{r_1,r_2,\ldots,r_N\}$, r_i is the response of the ith neuron, and Z is a normalization constant.

A simple discrimination rule is based on a linear readout

$$L_{\mathbf{W}}(\mathbf{r})=\sum_{i=1}^{N}W_i r_i. \tag{3}$$

The distribution of this quantity consists of a mixture of two Gaussians with equal variance and means $L_{\mathbf{W}}(\mathbf{f}^{\pm})$,

$$L_{\mathbf{W}}(\mathbf{f}^{\pm})=\sum_{i=1}^{N}W_i f_i^{\pm}, \tag{4}$$

where $L_{\mathbf{W}}(\mathbf{f}^{\pm})$ corresponds to the case of a $+$ and $-$ stimulus, respectively. A maximum likelihood discrimination rule based on $L_{\mathbf{W}}(\mathbf{r})$ makes a $+$ decision if $L_{\mathbf{W}}(\mathbf{r})$ is closer to $L_{\mathbf{W}}(\mathbf{f}^+)$. Otherwise, the inputs are classified as $-$. Note that we have assumed that the correlations are independent of the stimulus, hence the distributions of $L_{\mathbf{W}}(\mathbf{r})$ have the same variance for both stimuli. The error of this decision rule, P_e, for a case where the stimulus was $+$ is given by

$$P_e=\int_{\sqrt{S}}^{\infty}\frac{dx}{\sqrt{2\pi}}e^{-(x^2/2)}, \tag{5}$$

where S is the (squared) signal to noise ratio (SNR),

$$S=\frac{[L_{\mathbf{W}}(\mathbf{f}^+)-L_{\mathbf{W}}(\mathbf{f}^-)]^2}{\langle[L_{\mathbf{W}}(\mathbf{r})-L_{\mathbf{W}}(\mathbf{f}^+)]^2\rangle}, \tag{6}$$

where $\langle\cdots\rangle$ denotes the average over fluctuations, in this case, with respect to $P(\mathbf{r}|+)$, Eq. (2). This SNR is the variance of the signal divided by the variance of the noise. An identical result holds for the SNR in the case of a $-$ stimulus. The average over fluctuations, which is a Gaussian integration, can be performed easily,

$$\langle[L_{\mathbf{W}}(\mathbf{r})-L_{\mathbf{W}}(\mathbf{f}^+)]^2\rangle=\sum_{ij}W_iW_j\langle(r_i-f_i^+)(r_j-f_j^+)\rangle,$$

$$=\sum_{ij}W_iW_jC_{ij}. \tag{7}$$

C_{ij} of Eq. (1) gives

$$S=\frac{1}{a}\frac{\left(\sum_i g_iW_i\right)^2}{(1-c)\sum_i W_i^2+c\left(\sum_i W_i\right)^2}, \tag{8}$$

where

$$g_i=f_i^+-f_i^-. \tag{9}$$

B. Uniform pooling

A simple majority rule is a uniform pooling in which

$$W_i=W \tag{10}$$

for all i. In this case, Eq. (8) reduces to

$$S=\frac{N}{(1-c)+cN}S_0, \tag{11}$$

where S_0 denotes the SNR per neuron in an independent population, which is

$$S_0=\frac{\overline{g}^2}{a}, \tag{12}$$

where $\overline{x}$ represents an average of x_i over the population, i.e., $\overline{x}\equiv 1/N\Sigma_{i=1}^{N}x_i$. Equation (11) predicts that S saturates to a finite value in large systems,

$$S\simeq\frac{S_0}{c},\quad N\gg\frac{1}{c}. \tag{13}$$

A linear increase of S with N will be seen when the correlations are weak in the regime of a small N,

$$S\simeq S_0N,\quad 1\ll N\ll\frac{1}{c} \tag{14}$$

C. Optimal linear readout

Next we consider a linear readout with weights that are optimized to maximize the accuracy of its discrimination. By differentiating SNR with respect to $\{\mathbf{W}\}$ it can be shown that the following vector $\mathbf{W}$ maximizes the SNR [18].

$$W_i \propto \sum_{j=1}^{N} C_{ij}^{-1} g_j, \tag{15}$$

$$\propto g_i - \frac{cN}{(1-c)+cN}\overline{g}. \tag{16}$$

Note that W_i changes from g_i to $g_i - \overline{g}$ as N increases beyond $N_C = 1/c$ unless $c = 0$. For the optimal linear readout the SNR becomes

$$S = N\frac{\overline{g^2}}{a}\left[1 + c\left(-1 + N\frac{(1-c)(1-\kappa)}{1-c(1-N\kappa)}\right)\right]^{-1}, \tag{17}$$

where κ, $0 < \kappa < 1$, defined by

$$\kappa = \frac{\overline{g^2} - \overline{g}^2}{\overline{g^2}}, \tag{18}$$

represents the inhomogeneity of the population. The dependence of Eq. (17) on N is very different from that of the uniform pooling, Eq. (11). Most importantly unless $\kappa = 0$, S does not saturate for a large N, but diverges linearly with N as

$$S \simeq \left(\frac{\kappa}{1-c}\right) N S_0, \quad N \gg \frac{1-c}{c\kappa}. \tag{19}$$

Thus, for a large N, S increases faster than NS_0 if $c > 1-\kappa$, which means that for sufficiently strong uniform correlations, the performance of the optimal linear readout in a large population is superior to the independent case. Finally, for weak positive correlations ($0 < c \ll 1$), there is another linear regime for S, which is given by

$$S \simeq S_0 N, \quad 1 \ll N \ll \frac{1}{c}, \tag{20}$$

as in the uniform pooling case. These properties are shown in Fig. 1.

This analysis clarifies some of the apparently conflicting conclusions drawn previously with regard to the effect of uniform positive correlations. We show here that in the presence of a substantial inhomogeneity of the mean responses, accurate information can be extracted from a large uniformly correlated ensemble. However, a simple majority rule is incapable of extracting this information.

Naively, we would expect that the behavior exhibited in the case of uniform correlations should be indicative of the more general case. In particular, we would expect that even if the correlations are not uniform but rather decay in space, an optimal readout will be able to extract the information from the system with an accuracy that increases linearly with N. This in fact is not the case, as is shown in the analysis below. Nevertheless, the simple example discussed in this section illustrates one of the important features in correlated populations: the interplay between the correlation of the fluctuation, which in this example is denoted by c, and the homogeneity or diversity of the response characteristics of the neurons in the population, which in the present case is represented by the parameter κ.

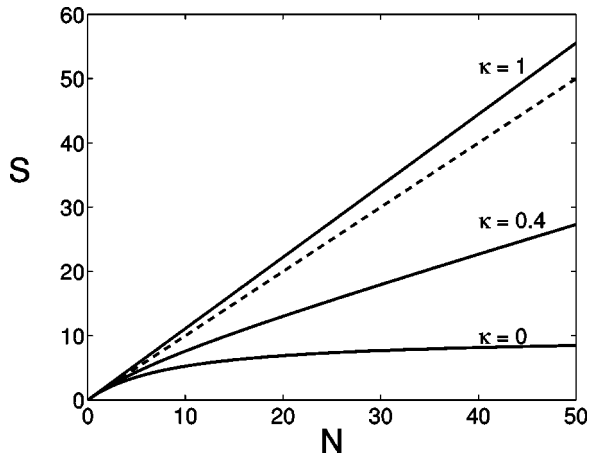

FIG. 1. Squared signal to noise ratio as a function of the number of neurons N that are uniformly correlated with $c = 0.1$. The dashed line represents the signal to noise Ratio of the independent population. The topmost curve is for $\kappa = 1$, and from the bottom $\kappa = 0$ and $\kappa = 0.4$.

III. MODEL: ESTIMATION OF AN ANGULAR VARIABLE

The remainder of this paper deals with a population of N neurons that responds to a stimulus characterized by an angle θ, where $-\pi < \theta \leq \pi$. The activity of each neuron is again assumed to be Gaussian with a mean $f_i(\theta)$ that represents its tuning curve, and a uniform variance a. The noise is assumed to be pairwise-correlated throughout the population. Hence the activity profile of the whole population, $\mathbf{r} = \{r_1, r_2, \cdots, r_N\}$, given a stimulus θ, follows the following multivariate Gaussian distribution

$$P(\mathbf{r}|\theta) = \frac{1}{Z}\exp\left(-\frac{1}{2}\sum_{i,j}[r_i - f_i(\theta)]C_{ij}^{-1}[r_j - f_j(\theta)]\right), \tag{21}$$

where Z is a normalization constant.

The tuning curves of all the neurons are assumed to be unimodal and identical in form but peaked at different angles, i.e.,

$$f_i(\theta) = f(\theta - \phi_i). \tag{22}$$

Here the angles at peaks or preferred angles ϕ_i are distributed *uniformly* from $-\pi$ to π, that is, $\phi_j = -\pi(N+1)/N + j\omega$, $j = 1, \ldots, N$. The lattice spacing, ω, between two neighboring preferred angles is

$$\omega = 2\pi/N. \tag{23}$$

For concreteness, we use the following tuning curve:

$$f(\theta)=(f_{\max}-f_{\mathrm{ref}})\exp((\cos\theta-1)/\sigma^2)+f_{\mathrm{ref}}, \quad (24)$$

for our numerical calculations. Here σ is the tuning width. The correlations are assumed to be independent of the stimulus, but dependent on the angular coordinates,

$$C_{ij}=a[\delta_{ij}+C(\phi_i-\phi_j)(1-\delta_{ij})]. \quad (25)$$

We assume that $C(\phi_i-\phi_j)$ decreases with a decay length ρ as the difference in the preferred angles of the correlated pair increases. A decrease in the magnitude of neuronal correlations with the dissimilarity in the preferred angles is often observed in cortical areas [10,14,19,20]. A concrete example is

$$C(\phi_i-\phi_j)=c\exp\left(-\frac{\|\phi_i-\phi_j\|}{\rho}\right) \quad \text{for } i\neq j, \quad (26)$$

where $\|\phi_1-\phi_2\|$ is the relative angle difference between ϕ_1 and ϕ_2, hence its range is from 0 to π. The coefficient c measures the correlations between neighboring cells. The positive-definiteness of the correlation matrix C_{ij} implies the following bounds on c:

$$-\frac{1}{N}\frac{\pi/\rho}{1-e^{-\pi/\rho}}<c<1. \quad (27)$$

This equation implies that stability puts strong limits on the magnitude of broadly tuned negative correlations in a large population. In the present context, it implies that for a given strength $|c|$ of negative correlations, the population size cannot be larger than $O(1/|c|)$. Unless otherwise stated, we will assume that c is positive.

The properties of the system for a large N are crucially dependent on the way in which the distance constants (in angular space) scale with N. In this paper we focus on a broadly tuned system that we believe is the most relevant case in biology. We define broad tuning to mean that all distance constants span a substantial fraction of the total range of the system. In our model the distance constants are the tuning width σ and the correlation length ρ. We thus assume that a given stimulus generates a response in a substantial fraction of the whole population, and that a given neuron is correlated with a substantial fraction of the population. Mathematically, this means that as we take the large N limit, the parameters σ and ρ stay finite. In addition, we assume that $f(\theta)$ and $C(\phi)$ are smooth, differentiable functions.

IV. THE FISHER INFORMATION

A useful way of measuring the efficiency of the population coding is the FI [9,11,17],

$$J(\theta)=\left\langle -\frac{\partial^2}{\partial\theta^2}\ln P(\mathbf{r}|\theta)\right\rangle. \quad (28)$$

For the Gaussian ensemble, Eq. (21), the FI reads

$$J(\theta)=\sum_{ij} f_i'(\theta)[C^{-1}]_{ij}f_j'(\theta), \quad (29)$$

where $f_i'(\theta)=\partial f_i(\theta)/\partial\theta$. Diagonalizing this quadratic form yields

$$J(\theta)=N\sum_n \frac{|g_n|^2}{C_n}, \quad (30)$$

where $\{C_n\}$ are the eigenvalues of the covariance matrix. Note that the eigenvectors of the correlation matrix **C** defined in Eqs. (25) and (26) have the form of harmonic functions, $e^{-in\phi_i}$. Hence, C_n is one of the Fourier components of $C(\|\phi_{i-}\phi_j\|)$.

$$C_n\equiv a(1-c)+a\sum_{j=1}^{N}\exp[-in(\phi_i-\phi_j)]\,C(\phi_i-\phi_j), \quad (31)$$

g_n is the Fourier transform of the derivative of $f_j(\theta)$, defined by

$$g_n=\frac{1}{N}\sum_{j=1}^{N} e^{-in\phi_j}f_j'(\theta). \quad (32)$$

The mode index n is an integer running from $-(N-1)/2$ to $(N-1)/2$ (for odd N). In the case of an uncorrelated population ($c=0$), the FI is given by

$$J=NJ_0, \quad J_0=\sum_n\frac{|g_n|^2}{a}. \quad (33)$$

It should be noted that in the uncorrelated case, the FI per neuron J_0 is of the order of one, since the particular normalization for $|g_n|^2$, Eq. (32), ensures that $\Sigma|g_n|^2\sim O(1)$. The FI is a local measure of the sensitivity of the response probability of the population to small changes in the stimulus value, and may therefore depend on the stimulus value. However, in our case the isotropy of the system ensures that J is the same for all θ.

Figure 2 displays the FI of the above model for various values of c as a function of ρ. The number of neurons N is 1000 and Eq. (24) is used for the tuning curves. The results clearly demonstrate several distinct regimes for J as we vary the correlation length ρ, as explained below.

A. The Fisher information for a system with broadly tuned correlations

The behavior of J, Eq. (29), for a large N depends on the range of correlations and the width of the tuning curves. Obviously, if the correlations are short ranged so that each neuron is correlated with only a few of its neighbors, the amount of information in the system will still grow linearly with N. Here we analyze the biologically interesting limit of broadly tuned systems. The broad tuning of f implies that $|g_n|^2$ decays rapidly as n increases beyond a characteristic value, which is proportional to the inverse of the tuning width σ so that the signal resides in the first few Fourier

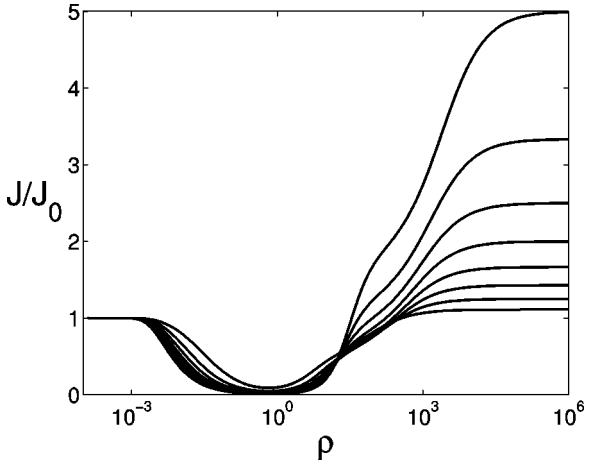

FIG. 2. Normalized Fisher information as a function of correlation length ρ. The correlation coefficient varies from $c=0.8-0.1$ with a uniform spacing of 0.1 from top to bottom when ρ is large. Equation (24) was used for the tuning curves, with $\sigma=\pi/4$, $f_{max}=25$, and $f_{ref}=5$. The number of neurons N is 1000. The ratio J/J_0 does not depend on the variance of the cells, a.

modes. At the same time, the broad tuning of C implies that the noise, namely, the variance of these modes, C_n, are of order N, resulting in the FI, which is of order unity [see Eq. (36) below]. To illustrate this important point, we evaluate C_n, for the example of Eq. (26), assuming a large N and $\rho \sim O(1)$, and obtain (see Appendix A)

$$J=N\sum_n \frac{|g_n|^2}{a}\left[1+\frac{N}{N_n}\right]^{-1}, \tag{34}$$

where

$$N_n=\frac{\pi\rho}{c}\,\frac{\rho^{-2}+n^2}{1-(-1)^n e^{-\pi/\rho}}. \tag{35}$$

B. Positive correlations

In the case of positive c, N_n is positive and can be interpreted as representing the number of degrees of freedom in each mode in the nth term (see below). The large N limit can be taken in Eq. (34) for $N\gg 1/c$, yielding

$$J=\sum_n \frac{|g_n|^2}{a}N_n, \tag{36}$$

which is $O(1)$. Hence the FI of the entire population does not scale linearly with the population size N but saturates to a size-independent finite limit (see Fig. 3, with $c=0.1$).

It is useful to introduce $N_{\rm eff}$ as the effective number of independent degrees of freedom in the correlated system. In general, it can be defined as the ratio between the FI of the system and the FI per neuron in the absence of correlations

$$N_{\rm eff}=\frac{J}{J_0}. \tag{37}$$

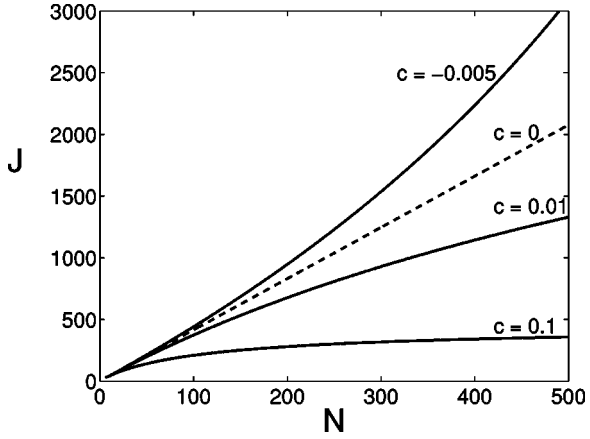

FIG. 3. Fisher information of the population as a function of the number of neurons N in the pool. The linear curve (dashed line) represents J_0 for an independent population. J with negative correlations lies above J_0, whereas J with positive correlations lies below. For all cases, $\rho=1$, $a=15$, $\sigma=\pi/4$, $f_{max}=25$, and $f_{ref}=5$.

In our case for a large N we obtain that $N_{\rm eff}$ is finite and is equal to

$$N_{\rm eff}=\langle N_n\rangle_n, \tag{38}$$

where $\langle\cdots\rangle_n$ is defined as a weighted average

$$\langle X_n\rangle_n \equiv \frac{\sum_n |g_n|^2 X_n}{\sum_n |g_n|^2}. \tag{39}$$

While it is easy to see from Eq. (35) that $N_{\rm eff}$ is inversely proportional to c, the dependence of $N_{\rm eff}$ on ρ is slightly more complicated. This is shown in Fig. 4. As a function of ρ, $N_{\rm eff}$ has a minimum around $\rho\sim 1$.

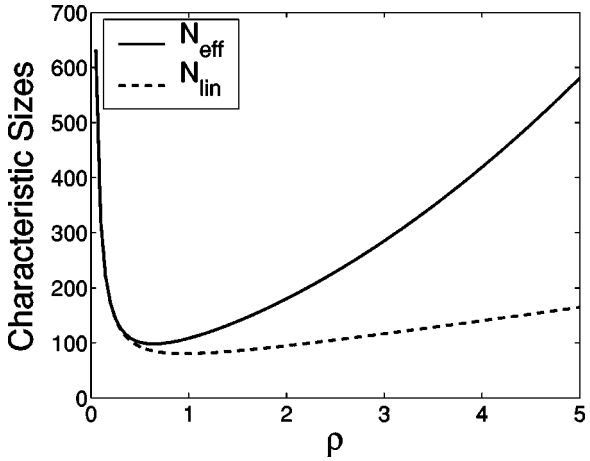

FIG. 4. $N_{\rm eff}$ and $N_{\rm lin}$ as a function of the correlation length ρ. $N=1000$ and $c=0.1$. Equation (24) was used for the tuning curves with $\sigma=\pi/4$, $f_{max}=25$, and $f_{ref}=5$.

C. Negative correlations

When $c<0$, J is larger than J_0N because $N_n<0$ [see Eq. (34)]. As mentioned above, the lower bound of c [Eq. (27)] implies that when the correlations are negative and ρ is of order of one there is an upper limit on the size of the population. As N increases and approaches this limit the system reaches an instability that increases its sensitivity to small changes in the stimulus. Consequently, J diverges in this limit, as shown in Fig. 3, with $c=-0.005$.

D. Weak correlations

When the magnitude of c is small there is an intermediate regime where $N\gg 1$ but nevertheless, J increases linearly with N. In this case, the FI can be approximated as

$$J\approx NJ_0. \tag{40}$$

From Eq. (34) it is readily seen that this behavior occurs for $1\ll N\ll N_{\rm lin}$, where

$$\frac{1}{N_{\rm lin}}=\left\langle \frac{1}{|N_n|}\right\rangle_n. \tag{41}$$

This holds for both positive and negative correlations. The dependence of $N_{\rm lin}$ on the correlation length constant ρ is shown in Fig. 4.

E. The effect of correlations on the population vector

Equation (30) can be interpreted as the FI of the collective modes of the system, which in our case are

$$z_n=\frac{1}{N}\sum_{i=1}^{N} e^{in\phi_i}r_i.$$

Since these modes are statistically independent (given a stimulus), the entire FI is simply the sum of the FI of each of them. Each mode contributes a term which is its SNR, where the signal of the nth mode is $|g_n|^2$ and its noise, i.e., its variance is C_n. According to our analysis above, because of the broad tuning of f_i only a few of these modes contribute significantly to J. On the other hand, because of the long range of the correlations, the contributions of these modes are of order unity, leading to the saturations of J.

Of particular interest is the accuracy of an estimation that is based on the first mode, z_1. This component can be written as a two-dimensional vector

$$\vec{z}=\frac{1}{N}\sum \vec{\varphi}_i r_i, \tag{42}$$

where $\vec{\varphi}_i$ is a unit vector along the preferred angle of the ith neuron. The vector $\vec{z}$ is the well-known population vector introduced originally for the decoding of the direction of reaching arm movement in two dimensions [21]. The accuracy of the population vector depends on the details of how exactly an angle estimation is constructed from it. Following Seung and Sompolinsky [9], a useful bound on the square error of the population vector is provided by the FI of $\vec{z}$. In our model this quantity is simply the first term in Eq. (34) which is

$$J[\vec{z}]=2N\frac{|f_1|^2}{a}\left[1+\left(\frac{cN}{\pi}\right)\left(\frac{1+e^{-\pi/\rho}}{\rho^{-1}+\rho}\right)\right]^{-1}, \tag{43}$$

where f_1, the first Fourier transform of $f_j(\theta)$, equals $-ig_1$. Thus, the effect of the correlations on $J[\vec{z}]$ is qualitatively similar to their effect on the full FI, which is illustrated in Fig. 3.

V. THE MUTUAL INFORMATION OF A CORRELATED POPULATION

In this section we analyze the effect of noise correlations on the MI of the system, which can be written as

$$I(\mathbf{r},\theta)=\int d\mathbf{r}\int d\theta\, P(\mathbf{r}|\theta)p(\theta)\log_2\left\{\frac{P(\mathbf{r}|\theta)}{P(\mathbf{r})}\right\}, \tag{44}$$

where $p(\theta)$ is the prior density of the stimuli and

$$P(\mathbf{r})=\int d\theta\, P(\mathbf{r}|\theta)p(\theta). \tag{45}$$

The MI measures the statistical dependence of $\mathbf{r}$ and θ. When $\mathbf{r}$ and θ are independent of each other, that is, $P(\mathbf{r}|\theta)$ is the same for different values of θ, the MI is zero. In the case of a statistically independent population coding for a continuous stimulus, the MI of a large system is related to the log of the FI ([22–24] and see below). Such a relation does not exist, in general, in the presence of correlations. Nevertheless, below we derive useful analytical results regarding the MI in a correlated population in special limits.

The analytical study of the MI is complicated even in cases where $P(\mathbf{r}|\theta)$ has a simple form. The reason is that evaluating the average over the marginal distribution, $P(\mathbf{r})$ in Eq. (45) is difficult to perform in general. The evaluation of the MI is facilitated by the use of the replica method [25]. As has been shown previously [24,26,27], the replica method is useful in analytical calculations of mutual information in model systems. In the present case, it enables us to perform the average over $\mathbf{r}$ in Eq. (44) before the averaging over θ, which is inherent to the definition of $P(\mathbf{r})$. Using this method, we show in Appendix B that

$$I=-[\ln 2]^{-1}\lim_{n\to 0}\frac{\partial}{\partial n}\int d\theta\, p(\theta)\int\prod_{\alpha=1}^{n} d\theta_\alpha\, p(\theta_\alpha) \times\exp\left\{\sum_{\alpha\neq\beta}^{n} G_\theta(\theta_\alpha,\theta_\beta)\right\}, \tag{46}$$

where

$$G_\theta(\psi,\phi)=\frac{1}{2}\sum_{i,j=1}^{N}\delta f_i(\psi)[C^{-1}]_{ij}\delta f_j(\phi) \tag{47}$$

$$\delta f_i(\psi)=f_i(\theta)-f_i(\psi).$$

Note that G_θ vanishes if the rates are not tuned, namely, if f_i is independent of θ, as expected. Equation (47) has the general structure of Eq. (29). The main difference between the two is that the FI depends on the local sensitivity of the mean responses, namely, their derivatives with respect to θ, whereas the MI depends on a global sensitivity measure, which is average of δf. However, the effect of correlations is qualitatively the same in both expressions. Thus, it follows from our previous analysis that in the case of broadly tuned positive correlations, G_θ approaches a finite value as $N\to\infty$. Unfortunately, this means that Eq. (46) cannot be evaluated by a saddle point method. Below we evaluate the MI in the two separate limits.

A. The mutual information for weak correlations

In this section we study the limit in which the strength of the correlations c is small. Formally, we first take the $N\to\infty$ limit and then the limit $c\to 0$. In the large N limit G_θ is proportional to $1/c$ [see, e.g., Eqs. (35) and (36)]. Therefore in the limit of small c, the integrals over θ_α are determined by their saddle point value. The saddle point equations of Eq. (46) are

$$-\frac{\partial}{\partial\theta_\alpha}G_\theta(\theta_\alpha,\theta_\alpha)+\sum_{\beta=1}^{n}\frac{\partial}{\partial\theta_\alpha}G(\theta_\alpha,\theta_\beta)=0, \tag{48}$$

the solution of which is

$$\theta_\alpha^*=\theta,\quad \alpha=1,\ldots,n. \tag{49}$$

Expanding $G_\theta(\theta_\alpha,\theta_\beta)$ around this saddle point yields

$$G_\theta(\theta_\alpha,\theta_\beta)=\frac{J(\theta)}{2}\delta\theta_\alpha\,\delta\theta_\beta, \tag{50}$$

where $J(\theta)$ is the FI [Eq. (29)], and in our case is $O(1/c)$. Substituting in Eq. (46) yields

$$I\simeq-[\ln 2]^{-1}\lim_{n\to 0}\frac{\partial}{\partial n}\int d\theta\,p(\theta)\int\prod_{\alpha=1}^{n}d\theta_\alpha \times\exp\left\{-\frac{J(\theta)}{2}\sum_{\alpha\neq\beta}\delta\theta_\alpha\,\delta\theta_\beta+n\ln p(\theta)\right\}.$$

Evaluating the Gaussian integrals over $\{\theta_\alpha\}$ yields

$$I\simeq H_\theta-\frac{1}{2}\int d\theta\,p(\theta)\log_2\left[\frac{2\pi e}{J(\theta)}\right],$$

where $H_\theta=-\int d\theta\,p(\theta)\log_2 p(\theta)$, which is the continuum ''entropy'' of θ. This result extends the log relation between the MI and the FI in the independent case to the case of weak broad correlations. In our particular case of a large isotropic system coding for an angle, $J(\theta)=J$ is independent of θ, and is given by Eq. (34), hence,

$$I\simeq H_\theta+\frac{1}{2}\log_2\left(\sum_n\frac{|g_n|^2}{a}N_n\right)-\frac{1}{2}\log_2(2\pi e). \tag{51}$$

This relation between I and the log of J is in agreement with previous results that have been derived for large independent, as well as for certain weakly correlated, populations [22–24]. Here we have shown that it is valid in the present case as long as c is small, i.e., that the N_ns [Eq. (35)] are large. Equation (51) implies that broadly tuned positive correlations saturate the MI so that I does not grow logarithmically with N as in the case of independent populations, but reaches a finite level as $N\to\infty$. This value represents the maximum information about θ that a system with this architecture can carry.

B. The mutual information for a noisy system

Another useful limit where the MI of a correlated population can be calculated is the case where the total amount of information in the system is small, namely, that the tuning of the responses to the stimulus is weak. In this case, we can expand the argument of the exponent in Eq. (46) in powers of G_θ. Retaining only the first order term in G_θ yields

$$I\simeq\frac{1}{2\ln 2}\langle G_\theta(\psi,\phi)\rangle_{\theta,\psi,\phi} =\frac{1}{2\ln 2}\sum_{i,j}\langle\overline{\delta f_i(\theta)}[C^{-1}]_{ij}\overline{\delta f_j(\theta)}\rangle_\theta, \tag{52}$$

where $\langle\cdots\rangle_\theta=\int d\theta\,p(\theta)\cdots$, and $\overline{\delta f_i(\theta)}=\langle\delta f_i(\psi)\rangle_\psi=f_i(\theta)-\langle f_i(\psi)\rangle_\psi$. Note that Eq. (52) is similar in form to Eq. (29) for the FI. Besides the 1/2 prefactor, the difference between the two expressions is that the signals in the FI are the derivative of f_i, $f_i'(\theta)$, whereas in the MI they are their global modulations, $\overline{\delta f_i(\theta)}$. Diagonalization of Eq. (52), and taking the large N limit leads to

$$I\simeq\frac{1}{2\ln 2}\sum_{n\neq 0}\frac{|f_n|^2}{a}N_n, \tag{53}$$

where we have used the fact that the Fourier transforms of $\overline{\delta f_i(\theta)}$ are equal to those of $f_i(\theta)$ for $n\neq 0$. This result holds as long as $I\ll 1$, which means that δf_i are small. Note that Eq. (53) can be interpreted as a sum of the MI of the statistically independent Fourier modes of the population responses. In general, the MI of an independent variable is linear with the number of variables in a regime where $I\ll 1$. Comparison of Eq. (53) and (36) reveals that the MI is less sensitive to the higher modes of the tuning curve, since $f_n=-ig_n/n$. This reflects the fact that the FI is a local measure of efficiency but the MI is not. Our high noise results hold for a general case where the stimulus modulation of the

responses is weak. A special case in which the low modulation is caused by low firing rates has been studied by Panzeri *et al.* [28].

VI. DISCUSSION

A. Summary of the results

In this paper we analyzed the effect of noise correlations on the performance of neural population codes. We have shown that whereas negative correlations improve the performance of these codes, positive correlations of the noise may have a drastic detrimental effect on their performance. In cases where both the mean responses and the correlations are broadly tuned, the noise generates global fluctuations in the response profile of the population that may result in a large error in the estimation of the stimulus. Because these fluctuations are correlated across the population, increasing the size of the sampled population may reduce the error somewhat but not suppress it completely even in large populations. We reached these conclusions by analyzing the FI of a correlated Gaussian population. We have shown that under the assumptions of broadly tuned system the FI reaches a saturation level given by Eq. (36). We also studied the MI of this system and showed that it saturates to a finite value as well. We derived the saturation value of the MI in the case of weak positive correlations, Eq. (51), and in the limit of high noise, Eq. (53). In addition, we showed that the case of uniform positive correlations is special in that the noise is restricted to a single collective mode (i.e., the uniform mode) whereas the remaining nonuniform modes carry most of the signal. Thus, in this case, the overall performance improves with increasing N, and can also be enhanced relative to the independent case. However, it can be shown that as long as the correlation length ρ is smaller than $O(\sqrt{N})$ [note that we are using normalization such that the total length of the system is $O(1)$] behavior is similar to the case of $\rho \sim O(1)$ considered above.

To illustrate the effect of correlated noise on the system we simulated the responses of our model population for three cases (see Fig. 5). The first corresponds to decaying correlations with $\rho = 1.18$, and the second to uniform correlations, $\rho \to \infty$. We also show a typical profile of responses for an independent population ($c = 0$). The data are generated according to Eq. (21) with a stimulus angle $\theta = 0°$. For the purpose of illustration, strong correlation coefficient $c = 0.9$ has been used in this simulation. To assess the error produced by these patterns, we assume a maximum-likelihood estimation. To show the performance of a ML estimator of θ, tuning curves centered at θ_{ML} were compared with the response profiles of the populations. As is clear from Fig. 5, in the case of the uniform correlations the overall deviation of the data from the tuning curve centered at $\theta = 0°$ can be very large. However, the deviation of the response profile is mostly in its amplitude. In fact, the deviation along the ''important'' direction is actually significantly smaller than the independent case. In contrast, in the case where $\rho = 1.18$, the lateral fluctuation is large due to the rough equality of two length scales, the correlation decay length and the tuning width. This results in a noisy firing pattern that looks as though the whole response profile had been shifted to the left (in this particular noise realization) relative to the ideal profile.

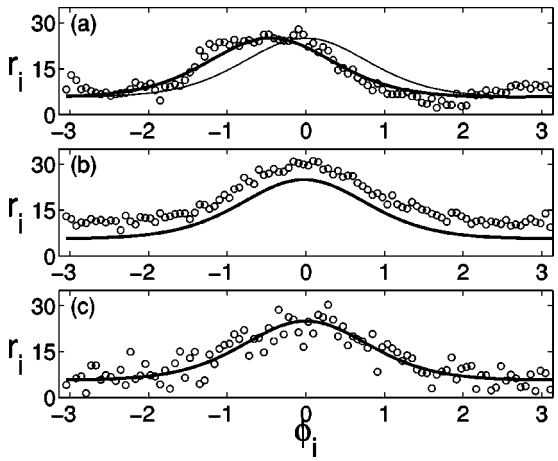

FIG. 5. Typical data generated according to Eq. (21) for a system with $N = 100$ neurons. Exponentially decaying correlations, Eq. (26), were used with: (a) $\rho = 1.18$, (b) $\rho \to \infty$ — uniform correlations, and (c) $\rho = 0$ — independent population. Each circle represents the number of spikes of a single neuron in the population, during a 500 msec time interval. The thick lines show $f(\theta_{ML} - \phi_i)$, where θ_{ML} is the maximum likelihood estimate of θ. The thin lines are the true mean responses of the neurons. Note that in (b) and (c) the thick and thin lines overlap. The parameters for the tuning curves are $f_{\max} = 25$, $f_{\text{ref}} = 5$, and $\sigma = \pi/4$. The variance a was taken to be 15.

B. Application to cortical populations

In order to assess the relevance of our theoretical results to cortical population codes, we consider the experimental values from Zohary *et al.* [10], who recorded the responses of pairs of neurons in the middle temporal visual area in monkeys to moving random dot patterns. Neurons in this area have a substantial degree of selectivity to the direction of movement of visual patterns. Since there were only 100 pairs of neurons and a limited number of repetitions (10 to 40 times), the data were averaged over different stimulus conditions. Then the pairs were divided into two groups depending on whether their preferred angles differed by less than 90° or not. The average correlation coefficient of the group with the similar preferred angles was 0.18 and the other was 0.04, with an overall average of 0.12. Although the precise dependence of the correlations on the difference between the preferred angles was not measured, the above results are consistent with smoothly decaying correlations. Fitting our model of exponentially decaying correlations, Eq. (26) to the average numbers quoted above yields the following parameter values:

$$c = 0.38, \tag{54}$$

$$\rho = 1. \tag{55}$$

In addition, we adopt the following values for our model tuning curve parameters, Eq. (24),

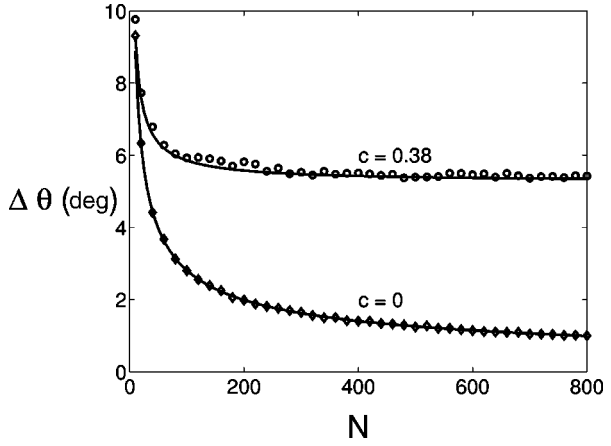

FIG. 6. Average error of the ML estimator and the Cramér-Rao bound as a function of the number of neurons N in the pool. The error of the ML estimator is nicely shown to saturate the bound. The error is represented in degrees. The parameters of the tuning curves are $f_{\max}=25$, $f_{\mathrm{ref}}=5$, and $\sigma=\pi/4$, for the correlations $a=15$, $c=0.38$, and $\rho=1$. The lower line corresponds to the error of the ML estimator for an independent population.

$$f_{\max}\sim 25 \ \text{spikes}, \tag{56}$$

$$f_{\mathrm{ref}}\sim 5 \ \text{spikes}, \tag{57}$$

$$\sigma\sim 45^{\circ}. \tag{58}$$

These parameters are typical for the cortical mean responses, assuming a time window of $T=500$ msec. Similarly, an estimate of the mean variance in the single neuron responses yields

$$a\sim 15 \ \text{spikes}^2. \tag{59}$$

Another system of interest is the primary visual cortex (V1). Neurons in V1 are broadly tuned to the orientation of a bar or a grating in their receptive field. A substantial fraction of these neurons show positive correlations that are stronger in amplitude for neurons with similar preferred orientations. The above parameter range also represents a reasonable estimate for the properties of neuronal ensembles in V1. Many neurons in motor cortex are sensitive to the direction of arm movement [14,21]. In this area the tuning curves of many cells are well approximated by a cosine function, which corresponds to $\sigma\simeq 90^{\circ}$. However, the degree of dependence of the correlations on the preferred angles of the neurons is still unclear.

Using the above parameters, we simulated the performance of a maximum-likelihood (ML) estimator in our model. The responses are generated randomly according to Eq. (21) with $\theta=0^{\circ}$. Then the mean squared error $\delta\theta=\theta^{ML}-\theta$ is obtained by averaging $\|\delta\theta\|^2$ over many repetitions with different random realizations of the responses. The standard deviation of the ML error is plotted against the number of neurons N in the pool in Fig. 6. The points represent averages over 4000 runs. We also present the Cramér-Rao bound of the error, which is simply the square root of the inverse of the FI calculated in Sec. IV. As shown in Fig. 6, the ML error closely follows the Cramér-Rao bound. Both saturate at a value around 5°. For reference, $\Delta\theta$ is around 5° for independent populations with $N\approx 30$. Thus, we conclude that the number of effective degrees of freedom in the correlated population cannot increase beyond $N_{\mathrm{eff}}\approx 30$. This is indeed confirmed by evaluating Eq. (37) with the above parameters. Changing slightly the parameter values, one can obtain slightly larger values of N_{eff}, but $N_{\mathrm{eff}}\leqslant 100$ seems to be a reasonable bound for this system. Finally, for comparison we also show the results for an uncorrelated population, which yields an error that decreases as $1/\sqrt{N}$, as expected.

C. Extensions of the present study

In conclusion, we briefly discuss some of the restrictions of the present model. We have focused here on systems with broadly tuned correlations. This means that a single neuron is significantly correlated with a fraction of the population. In the notation of our model, Eq. (26), we assumed that the correlation length ρ is of order unity [whereas the entire length of the system is also $O(1)$, in our case, 2π]. Short-range correlations correspond to the limit of $\rho=O(1/N)$, as studied by Abbott and Dayan [11] who show that the information is still extensive. This is consistent with the results shown in Fig. 2 where for fixed N and small ρ, the Fisher information is the same as in the uncorrelated case. Indeed, unlike the long-range case, in the limit of $\rho=O(1/N)$ each neuron is correlated with a finite number of neighbors, hence, the effective number of statistically independent degrees of freedom is proportional to N, and the information remains extensive.

We assumed in our study that the shape of the single-unit tuning curves is identical, but their peaks are displaced along the axis of the preferred angles. Adding inhomogeneity in the shapes of the tuning curves will not affect the qualitative behavior of the system, provided this inhomogeneity varies smoothly, namely, its space constant is of order unity. On the other hand, if the tuning curves of different neurons vary randomly across different neurons, the behavior may be quite different. The reason is that in such a case, all the modes including those with high Fourier wave numbers carry the signal, whereas the correlations (assuming they are not random) only concentrate in the long wavelength channels. This case will be analogous to the example of uniform correlations with inhomogeneous responses studied in Sec. II C above. However, a realistic model should include randomness both in the tuning and in the correlations. Investigating the information content of such a system and the resources required for reading out this information is beyond the scope of this paper.

Throughout this paper, multivariate Gaussian statistics was assumed for the neuronal responses. The most common candidate for neuronal responses are spike counts that obviously cannot be strictly Gaussian. It would be useful to extend the current study to other distributions, e.g., Poisson neurons. However, it is hard to generalize other single neuron statistical models to a correlated population, except for special cases, such as uniform positive correlations. This important issue deserves further study.

Finally, in this work we have analyzed examples of all positive or all negative correlations. In reality, correlations within a population may vary both in magnitude and in sign. However, reliable estimates of the distribution of correlations in cortical networks are hard to obtain. Another important restriction on the present work is our assumption that the correlations do not depend on θ. If the correlations depend on the stimulus, one must take into account not only the modification of the noise statistics by the correlations but also the additional information about the stimulus carried by them. This topic is investigated elsewhere [29].

ACKNOWLEDGMENTS

H.S. acknowledges helpful discussions with Larry Abbott and Sebastian Seung. This research was partially supported by a grant from the Israeli Science Foundation, and by a grant from the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel.

APPENDIX A: THE EIGENVALUES OF THE CORRELATION MATRIX

We calculate here the eigenvalues of **C**, as defined by Eq. (26). Denoting $\omega = 2\pi/N$, the lattice spacing, we can write

$$C_n = \frac{1}{N}\sum_{jk} e^{in(\phi_j-\phi_k)}C_{jk} = a\left\{1-c+\frac{c}{N}\sum_{jk} \times \exp\left[in\omega(j-k)-\frac{\omega}{\rho}\|j-k\|\right]\right\}, \tag{A1}$$

where $\|j-k\|$ ranges from 0 to $(N-1)/2$ for odd N. Summing the geometric series yields

$$C_n = a\left\{1+2c\,\frac{-\lambda^2+\lambda\cos(n\omega)+(-1)^n\lambda(N-1)/2\cos((n\omega)/2)[\lambda-1]}{1+\lambda^2-2\lambda\cos(n\omega)}\right\}, \tag{A2}$$

where $\lambda = e^{-\omega/\rho}$ and $n=-(N-1)/2,\ldots,(N-1)/2$. Taking the large N limit we expand to leading order in ω and obtain

$$C_n = a\left\{1+\frac{cN}{\pi\rho}\,\frac{1-(-1)^n e^{-\pi/\rho}}{\rho^{-2}+n^2}\right\} = a\left(1+\frac{N}{N_n}\right),$$

$$N_n = \frac{\pi\rho}{c}\,\frac{\rho^{-2}+n^2}{1-(-1)^n e^{-\pi/\rho}}. \tag{A3}$$

APPENDIX B: REPLICA CALCULATION OF THE MUTUAL INFORMATION

The replica trick is to calculate the average of the logarithm of a random variable using the following identity:

$$\langle \ln x\rangle = \left.\frac{\partial}{\partial n}\langle x\rangle^n\right|_{n=0}. \tag{B1}$$

The MI defined in Eq. (44) can be written in the following way:

$$I = \frac{1}{\ln 2}\left\langle\left\langle \ln\frac{P(\mathbf{r}|\theta)}{P(\mathbf{r})}\right\rangle_{r|\theta}\right\rangle_{\theta}, \tag{B2}$$

$$= -\frac{1}{\ln 2}\langle\langle \ln\langle \exp(X_{\theta_1\theta})\rangle_{\theta_1}\rangle_{r|\theta}\rangle_{\theta}, \tag{B3}$$

$$X_{\theta_1\theta} = \ln P(\mathbf{r}|\theta_1) - \ln P(\mathbf{r}|\theta) = -\frac{1}{2}\sum_{ij}\delta f_i(\theta_1)[C^{-1}]_{ij}\delta f_j(\theta_1) - \sum_{ij}\delta f_i(\theta_1)[C^{-1}]_{ij}[r_i - f_i(\theta)], \tag{B4}$$

where $\langle\cdots\rangle_\theta = \int d\theta\, P(\theta)\cdots$ and $\langle\cdots\rangle_{\mathbf{r}|\theta} = \int d\mathbf{r}\, P(\mathbf{r}|\theta)\cdots$. $\delta f_i(\theta_1)$ is $f_i(\theta) - f_i(\theta_1)$.

After applying the replica trick, we obtain the following form:

$$I = -\frac{1}{\ln 2}\frac{\partial}{\partial n}\langle\langle\langle \exp(X_{\theta_1\theta})\rangle_{\theta_1}^n\rangle_{r|\theta}\rangle_{\theta}\Big|_{n=0}, \tag{B5}$$

$$= -\frac{1}{\ln 2}\frac{\partial}{\partial n}\int d\theta\, p(\theta)\prod_{\alpha=1}^{n} d\theta_\alpha\, p(\theta_\alpha) \times\left\langle \exp\left(\sum_\alpha X_{\theta_\alpha\theta}\right)\right\rangle_{r|\theta}\Big|_{n=0}, \tag{B6}$$

$$\sum_{\alpha} X_{\theta_\alpha \theta} = -\sum_{\alpha} G_{\theta}(\theta_\alpha, \theta_\alpha) - \sum_{\alpha} \sum_{ij} \delta f_i(\theta_\alpha)[C^{-1}]_{ij} \times [r_i - f_i(\theta)], \tag{B7}$$

where $G_\theta(\theta_\alpha, \theta_\beta) = \frac{1}{2}\Sigma_{ij}\delta f_i(\theta_\alpha)[C^{-1}]_{ij}\delta f_i(\theta_\beta)$.

Note that the average over r_i, $\langle \cdots \rangle_{r|\theta}$ in Eq. (B6) is easy to perform because $\Sigma_\alpha X_{\theta_\alpha \theta}$ is linear with respect to r_i. Performing the average over r_i yields

$$I = -\frac{1}{\ln 2}\frac{\partial}{\partial n}\int d\theta\, p(\theta) \prod_{\alpha=1}^{n} d\theta_\alpha\, p(\theta_\alpha) \times \exp\left[\sum_{\alpha \neq \beta} G_\theta(\theta_\alpha, \theta_\beta)\right]\Bigg|_{n=0}, \tag{B8}$$

where $\Sigma_{\alpha \neq \beta}$ stands for summation over $\alpha = 1, \ldots, n$ and $\beta = 1, \ldots, n$, with the restriction $\alpha \neq \beta$. This completes the derivation of Eq. (46).

[1] P.S. Churchland and T.J. Sejnowski, *The Computational Brain* (MIT Press, London, 1992), Chap. 4.
[2] A.P. Georgopoulos, A.R. Schwartz, and R.E. Kettner, Science **233**, 1416 (1986).
[3] C. Lee, W.H. Roher, and D.L. Sparks, Nature (London) **332**, 357 (1988).
[4] M.A. Wilson and B.L. McNaughton, Science **261**, 1055 (1993).
[5] D.C. Fitzpatric, R. Batra, T.R. Stanford, and S. Kuwada, Nature (London) **388**, 871 (1997).
[6] M.P. Young and S. Yamane, Science **256**, 1327 (1992).
[7] M.A. Paradiso, Biol. Cybern. **58**, 35 (1988).
[8] H.P. Snippe and J.J. Koenderink, Biol. Cybern. **66**, 543 (1992).
[9] H.S. Seung and H. Sompolinsky, Proc. Natl. Acad. Sci. U.S.A. **90**, 10 749 (1993).
[10] E. Zohary, M.N. Shadlen, and W.T. Newsome, Nature (London) **370**, 140 (1994).
[11] L.F. Abbott and P. Dayan, Neural Comput. **11**, 91 (1999).
[12] R. Vogels, Biol. Cybern. **64**, 25 (1990).
[13] E. Fetz, K. Yoyoma, and W. Smith, in *Cerebral Cortex*, edited by A. Peters and E. Jones (Plenum, New York, 1991), Vol. 9.
[14] D. Lee, N.L. Port, W. Kruse, and A.P. Georgopoulos, J. Neurosci. **18**, 1161 (1998).
[15] E.M. Maynard, N.G. Hatsopoulos, C.L. Ojakangas, B.D. Acuna, J.N. Sanes, R.A. Normann, and J.P. Donoghue, J. Neurosci. **19**, 8083 (1999).
[16] M.N. Shadlen, K.H. Britten, W.T. Newsome, and J.A. Movshon, J. Neurosci. **16**, 1486 (1996).
[17] H. Yoon and H. Sompolinsky, Adv. in Neural Inf. Proc. Systems **11**, 167 (1999).
[18] E. Salinas and L.F. Abbott, J. Comput. Neurosci. **1**, 89 (1994).
[19] P.L.E. van Kan, R.P. Scobey, and A.J. Gabo, Exp. Brain Res. **60**, 559 (1985).
[20] D.N. Mastronarde, J. Neurophysiol. **49**, 303 (1983).
[21] A.P. Georgopoulos, J.F. Kalaska, R. Caminiti, and J.T. Massey, J. Neurosci. **2**, 1527 (1982).
[22] J. Rissanen, IEEE Trans. Inf. Theory **42**, 40 (1996).
[23] N. Brunel and J.P. Nadal, Neural Comput. **10**, 1731 (1998).
[24] K. Kang and H. Sompolinsky, Phys. Rev. Lett. **86**, 4958 (2001).
[25] M. Mezard, G. Parisi, and M.A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
[26] J.P. Nadal and N. Parga, Network **4**, 295 (1993).
[27] I. Samengo and A. Treves, Phys. Rev. E **63**, 011910 (2001).
[28] S. Panzeri, S.R. Schultz, A. Treves, and E.T. Rolls, Proc. R. Soc. London, Ser. B **266**, 1001 (1999).
[29] M. Shamir and H. Sompolinsky (unpublished).