

The Effect of Noise Correlations in Populations of Diversely Tuned Neurons

Alexander S. Ecker,^{1,2,3,4} Philipp Berens,^{1,2,3,4} Andreas S. Tolias,^{4,5,6} and Matthias Bethge^{1,2,3}

¹Centre for Integrative Neuroscience and Institute for Theoretical Physics, University of Tübingen, ²Max Planck Institute for Biological Cybernetics, and ³Bernstein Centre for Computational Neuroscience, 72076 Tübingen, Germany, ⁴Department of Neuroscience, Baylor College of Medicine, and ⁵Michael E. DeBakey Veterans Affairs Medical Center, Houston, Texas 77030, and ⁶Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005

The amount of information encoded by networks of neurons critically depends on the correlation structure of their activity. Neurons with similar stimulus preferences tend to have higher noise correlations than others. In homogeneous populations of neurons, this limited range correlation structure is highly detrimental to the accuracy of a population code. Therefore, reduced spike count correlations under attention, after adaptation, or after learning have been interpreted as evidence for a more efficient population code. Here, we analyze the role of limited range correlations in more realistic, heterogeneous population models. We use Fisher information and maximum-likelihood decoding to show that reduced correlations do not necessarily improve encoding accuracy. In fact, in populations with more than a few hundred neurons, increasing the level of limited range correlations can substantially improve encoding accuracy. We found that this improvement results from a decrease in noise entropy that is associated with increasing correlations if the marginal distributions are unchanged. Surprisingly, for constant noise entropy and in the limit of large populations, the encoding accuracy is independent of both structure and magnitude of noise correlations.

Introduction

The accuracy of information processing in the cortex depends strongly on how sensory stimuli are encoded by a population of neurons. Two key factors influence the quality of a population code: (1) the shape of the tuning functions of individual neurons and (2) the structure of interneuronal noise correlations. Although the magnitude of noise correlations is debated, a common finding is that they are strongest for neurons with similar tuning properties (Zohary et al., 1994; Bair et al., 2001; Cohen and Newsome, 2008; Smith and Kohn, 2008; Ecker et al., 2010). Interestingly, such a limited range correlation structure seems to be highly detrimental for a population code, even if correlations are small. If correlations are unavoidable, it is generally believed that reducing them improves a population code (Zohary et al., 1994; Abbott and Dayan, 1999; Sompolinsky et al., 2001; Wilke and Eurich, 2002; Ecker et al., 2010). In line with this notion, a

number of recent experimental studies find reduced spike count correlations under experimental conditions in which improved coding is expected, such as under attention (Cohen and Maunsell, 2009; Mitchell et al., 2009), adaptation (Gutnisky and Dragoi, 2008), or after learning (Gu et al., 2011).

Most previous theoretical studies of population coding use homogeneous population models, in which all neurons have tuning functions that differ only in their preferred stimulus and are otherwise identical (Snippe and Koenderink, 1992; Abbott and Dayan, 1999; Sompolinsky et al., 2001; Wilke and Eurich, 2002). In these models, limited range correlations introduce a strong noise component in the subspace in which the stimulus is encoded, impairing the population code (Sompolinsky et al., 2001). However, a very prominent feature of cortical neurons is the diversity of their tuning functions. This heterogeneity changes the way the stimulus is encoded and can critically alter the properties of a population code (Shamir and Sompolinsky, 2006). Unfortunately, this has not been sufficiently appreciated. We extended the approach pioneered by Sompolinsky and colleagues (Sompolinsky et al., 2001; Shamir and Sompolinsky, 2006) to study population codes with heterogeneous tuning functions, arbitrary mean/variance relationships (Fano factors), and a broad range of correlation structures. To characterize the encoding accuracy, we used Fisher information and maximum-likelihood decoding and studied a simple model with heterogeneous tuning functions, Poisson-like noise, and limited range correlation structure. We found that, in contrast to current belief, decreasing correlations does not necessarily lead to increased information. Instead, if correlations are strong enough, increasing them can substantially

Received May 21, 2011; revised July 28, 2011; accepted Aug. 9, 2011.

Author contributions: A.S.E., P.B., A.S.T., and M.B. designed research; A.S.E., P.B., A.S.T., and M.B. performed research; A.S.E., P.B., A.S.T., and M.B. wrote the paper.

This work was supported by the Bernstein Award (M.B.) by the German Ministry of Education, Science, Research, and Technology (Bundesministerium für Bildung und Forschung) Grant FKZ 01GQ0601, the German Excellence Initiative through the Centre for Integrative Neuroscience Tübingen (M.B., A.S.E.), the Max Planck Society (M.B., A.S.E., P.B.), the German National Academic Foundation (P.B.), National Eye Institute–National Institutes of Health Grant R01 EY018847 (A.S.T.), The Arnold and Mabel Beckman Foundation Young Investigator Award (A.S.T.), the Veterans Affairs Merit Award (A.S.T.), and The McKnight Endowment Fund for Neuroscience Scholar Award (A.S.T.). We thank R. Haefner and S. Gerwin for discussions and comments on an earlier version of this manuscript.

This article is freely available online through the *JNeurosci* Open Choice option.

Correspondence should be addressed to Alexander S. Ecker, Max Planck Institute for Biological Cybernetics, Spemannstrasse 41, 72076 Tübingen, Germany. E-mail: aecker@tuebingen.mpg.de.

DOI:10.1523/JNEUROSCI.2539-11.2011

Copyright © 2011 the authors 0270-6474/11/3114272-12\$15.00/0

Table 1. List of symbols

Symbol	Meaning
n	Number of neurons
θ	Stimulus direction
$f_k(\theta)$	Tuning function of neuron k (average spike count)
\mathbf{f}	Vector of tuning functions (for $\theta = 0$; $n \times 1$)
$\tilde{\mathbf{f}}$	Discrete Fourier transform of \mathbf{f}
\mathbf{y}	Vector of spike counts
ϕ_k	Preferred direction of neuron k
$\sigma_k(\theta)$	SD of spike counts of neuron k
$\boldsymbol{\sigma}$	Vector of SDs (for $\theta = 0$; $n \times 1$)
$\mathbf{S} = \text{Diag}(\boldsymbol{\sigma})$	Diagonal matrix of SDs
$g_k = f'_k/\sigma_k$	Normalized tuning curve derivatives
$c(\phi_j \ominus \phi_k)$	Spatial structure of correlation coefficients (\ominus is the angular difference)
$c_0 = c(0)$	Correlation coefficient of two neurons with identical preferred directions
\mathbf{R}	Matrix of correlation coefficients ($n \times n$)
r_{jk}	Correlation coefficient of neurons j and k
$\langle r \rangle$	Average correlation coefficient of the population, $\langle r \rangle \approx 0.3 c_0$.
\mathbf{r}	First column of \mathbf{R} , $r_k = c_\omega$ except for $r_0 = 1$
\mathbf{Q}	Covariance matrix (for $\theta = 0$; $n \times n$)
\mathbf{U}	Fourier basis, $u_{jk} = \exp(-i\omega_j k)/\sqrt{n}$
$\tilde{\mathbf{R}}$	Diagonal matrix containing eigenvalues of \mathbf{R} . Diagonal elements are $\tilde{\mathbf{r}}$
J	Fisher information
J_{mean}	Linear part of Fisher information (Eq. 8)
J_{cov}	Non-linear part of Fisher information (Eq. 9)
$J_{\text{mean}}^{\text{hom}}$	J_{mean} of a homogeneous population of neurons
J_{indep}	Fisher information of an independent population of neurons
κ	Degree of amplitude variability, $\kappa = \text{Var}[\sqrt{a_k}]$.
a_k	Amplitude of tuning function of neuron k
F	Fano factor (variance/mean)

We use boldface lowercase letters to denote (column) vectors (e.g., \mathbf{v}) and boldface uppercase letters for matrices (e.g., \mathbf{M}). Scalars or scalar elements of vectors and matrices are set in italics (e.g., a , v_j or M_{jk}). The j th column of \mathbf{M} is referred to as \mathbf{m}_j .

increase the encoding accuracy, even to the point at which a population with limited range correlations is more accurate than an independent one.

We show that this increase in encoding accuracy with higher correlations is due to a decrease in noise entropy. If the entropy is kept constant, the encoding accuracy can be improved substantially by reducing correlations below a critical value. Surprisingly, in large neural populations, the quality of a code is mainly determined by the amount of heterogeneity in the tuning functions and by the noise entropy, while correlations play only a minor role.

Materials and Methods

Table 1 contains a list of symbols used throughout this article.

Population model. We consider a population of n neurons responding to a stimulus, which is characterized by its direction of motion $\theta \in [0, 2\pi)$. The response of neuron j is given by

$$y_j(\theta) = f_j(\theta) + \eta_j(\theta), \quad (1)$$

where $f_j(\theta)$ is the tuning curve of neuron j and $\eta_j(\theta)$ is the trial-to-trial variability in the neural responses. The variability is assumed to follow a multivariate normal distribution with zero mean and covariance $\mathbf{Q}(\theta)$. We assume that the preferred directions of the neurons are uniformly spaced around the circle (i.e., $\phi_j = 2\pi j/n$, where $j = 0, \dots, n-1$). Because of the circular nature of the stimulus, it is sometimes convenient (both for presentation and for calculations) to use negative indices ranging from $-n/2$ to $n/2$. These indices are simply understood modulo n . Thus, for example, if $n = 32$, then f_{-10} is the same as f_{22} .

Homogeneous population model. In the homogeneous population model, all neurons have identical tuning functions except for their preferred directions. In other words,

$$f_j(\theta) = f(\theta - \phi_j). \quad (2)$$

In this model, the population is invariant under rotation. This means that any shift in the stimulus can be translated into a renumbering of the cells in the population. The average population activity has the same shape as the tuning function centered on θ . Throughout the paper, we use von Mises tuning functions given by the following:

$$f_j(\theta) = \alpha + \beta \exp[\gamma(\cos(\theta - \phi_j) - 1)]. \quad (3)$$

We use the parameter values $\alpha = 1$, $\beta = 19$, and $\gamma = 2$. These parameters closely resemble the average values found in our recordings from monkey V1 (Ecker et al., 2010) and result in tuning curves with a maximum amplitude of 20 Hz (see Fig. 1E).

Random amplitude model. In the random amplitude model, all neurons are assumed to have identically shaped tuning functions but potentially different amplitudes:

$$f_j(\theta) = a_j f(\theta - \phi_j). \quad (4)$$

Here, $f(\theta - \phi_j)$ is as in the homogeneous model above. The amplitudes a_j are assumed to be drawn independently for each neuron, have mean $\langle a_j \rangle = 1$ and their square root has variance $\kappa = \text{Var}[\sqrt{a_j}]$. Our results do not require any specific distribution for the amplitudes. Note that, in the random amplitude model, all neurons have the same tuning width because a multiplicative gain does not change the tuning width.

Other forms of heterogeneity. We numerically simulated two additional types of heterogeneous populations. In the first model (see Fig. 8A), we randomly drew the tuning widths, γ , from a lognormal distribution with mean 2 and variance 4 while keeping all other parameters constant as in the homogeneous model described above. In the second model (see Fig. 8B), we drew (with replacement) triples of tuning parameters (α , β , γ) from a database of 408 orientation tuning curves measured in V1 of awake monkeys (Ecker et al., 2010). We assigned preferred directions uniformly spaced around the circle as above. We did not combine parameters from different neurons independently (i.e., the number of possible tuning curve shapes was 408 rather than 408^3).

Correlation structure. In our model, we assume the correlation coefficient of two neurons to be independent of the stimulus. This allows us to parameterize the covariance matrix as follows:

$$\mathbf{Q}_{jk}(\theta) = \sigma_j(\theta)\sigma_k(\theta)r_{jk}, \quad (5)$$

where $\sigma_j^2(\theta)$ is the variance of neuron j and r_{jk} is the correlation coefficient of neurons j and k . The covariance matrix \mathbf{Q} can be written as $\mathbf{Q} = \mathbf{SRS}$, where \mathbf{R} is the correlation coefficient matrix, which is premultiplied and postmultiplied by a diagonal matrix \mathbf{S} of standard deviations.

The correlation coefficient of two neurons depends only on $|\phi_j \ominus \phi_k|$, the angular difference between their preferred directions (we use \ominus to express the fact that it is a difference between two circular quantities, in other words, $\phi_j \ominus \phi_k = \arg \exp[i(\phi_j - \phi_k)]$),

$$r_{jk} = c(|\phi_j \ominus \phi_k|) + \delta_{jk}(1 - c(0)). \quad (6)$$

Here, δ_{jk} is the Kronecker delta ($\delta_{jk} = 1$ if $j = k$ and $\delta_{jk} = 0$ otherwise). We do not require any specific form for $c(|\phi_j \ominus \phi_k|)$, other than that it must lead to a valid covariance matrix. For the large n case, this is equivalent to requiring it to be bounded between -1 and 1 and all its Fourier components to be positive. While the former condition is a requirement for correlation coefficients, the latter ensures that the covariance matrix remains positive definite in the limit of large populations. We further assume that the variances are Poisson-like, which means $\sigma_j^2(\theta) = f_j(\theta)$. This is sometimes referred to as “proportional noise” (Wilke and Eurich, 2002), while the case where σ_j^2 does not depend on the stimulus is referred to as “additive noise.” Because the correlation coefficients depend only on the difference between the preferred directions of two neurons, the correlation matrix, \mathbf{R} , is circulant. We therefore have $r_{jk} = r_{j-k}$, where the vector \mathbf{r} is the first column of \mathbf{R} . We often refer to \mathbf{r} simply as the correlation structure.

For all examples, we assume that the correlation matrix has limited range structure. This means that the correlation between two neurons is

maximal if they have identical preferred directions and decreases with increasing difference in preferred direction. The parametric form we use is the following:

$$c(|\phi_j \ominus \phi_k|) = c_0 \exp\left(-\frac{|\phi_j \ominus \phi_k|}{L}\right), \quad (7)$$

where c_0 is the correlation of two neurons with identical preferred directions and L controls the spatial scale (the larger L the longer the range of correlations). We use $L = 1$ for all figures in this paper. Qualitatively, the results do not depend on the exact choice of L (within reasonable limits) and this value is in good agreement with our V1 data (Ecker et al., 2010) and previously published studies (Zohary et al., 1994; Bair et al., 2001). The parameter c_0 controls the average level of correlations, $\langle r \rangle$, which also depends on L . For $L = 1$, we have $\langle r \rangle \approx 0.3 c_0$. These values are included in the figures for reference. Note that larger values of L lead to higher average correlations relative to c_0 (for instance, for $L = 2$, we have $\langle r \rangle \approx 0.5 c_0$).

Fisher information. To quantify the encoding accuracy of the population, we use Fisher information (Cover and Thomas, 1991). If a Gaussian distribution is assumed for the noise, the Fisher information can be written as the sum of two terms, $J = J_{\text{mean}} + J_{\text{cov}}$, where (Kay, 1993)

$$J_{\text{mean}}(\theta) = \mathbf{f}'(\theta)^T \mathbf{Q}(\theta)^{-1} \mathbf{f}'(\theta) \quad (8)$$

$$J_{\text{cov}}(\theta) = \frac{1}{2} \text{Tr}[(\mathbf{Q}'(\theta) \mathbf{Q}(\theta)^{-1})^2], \quad (9)$$

and \mathbf{f}' and \mathbf{Q}' are the derivatives of the tuning curve and the covariance matrix with respect to the stimulus direction, θ . The term J_{mean} can be thought of as the information that is encoded in changes in the mean firing rates of the population [i.e., the term $\mathbf{f}'(\theta)$]. In contrast, J_{cov} is the information encoded by changes in the covariances [i.e., the term $\mathbf{Q}'(\theta)$].

To evaluate the above two terms, we have to invert the covariance matrix. Because \mathbf{R} is circulant, it can be diagonalized by changing to the Fourier basis as follows:

$$\mathbf{R} = \mathbf{U} \tilde{\mathbf{R}} \mathbf{U}^*. \quad (10)$$

The matrix \mathbf{U} is the Fourier basis, given by

$$u_{jk} = \frac{1}{\sqrt{n}} \exp(-i\omega jk), \quad (11)$$

(with $\omega = 2\pi/n$), \mathbf{U}^* is the hermitian (complex transpose) of \mathbf{U} , and $\tilde{\mathbf{R}}$ is a diagonal matrix containing the eigenvalues of \mathbf{R} , which can be calculated by the discrete Fourier transform of \mathbf{r} (the first column of \mathbf{R}), given by

$$\tilde{r}_k = \sum_{j=0}^{n-1} r_j \exp(-i\omega jk). \quad (12)$$

Using this factorization, $\mathbf{Q}(\theta)^{-1}$ reads

$$\mathbf{Q}(\theta)^{-1} = \mathbf{S}(\theta)^{-1} \mathbf{U} \tilde{\mathbf{R}}^{-1} \mathbf{U}^* \mathbf{S}(\theta)^{-1}, \quad (13)$$

which is easy to calculate because it contains only inverses of diagonal (\mathbf{S} , $\tilde{\mathbf{R}}$) matrices.

Calculation of J_{mean} . A model similar to ours but with the restriction to additive noise (i.e., with stimulus independent covariance matrix, \mathbf{Q}) has been studied by Sompolinsky et al. (2001). Although the additive case is somewhat simpler because J_{cov} is zero ($\mathbf{Q}' = 0$), the approach can be generalized to obtain analytic expressions for J_{mean} and J_{cov} in the non-additive case. For J_{mean} , we use the above factorization of \mathbf{Q}^{-1} and substitute it into Equation 8. After substituting $g_k = f'_k/\sigma_k$, we obtain the following:

$$J_{\text{mean}} = \frac{1}{n} \sum_{k=0}^{n-1} \frac{|\tilde{g}_k|^2}{\tilde{r}_k}, \quad (14)$$

where \tilde{g} is the discrete Fourier transform of \mathbf{g} (see Eq. 12; equivalently $\tilde{g} = \sqrt{n} \mathbf{U}^* \mathbf{g}$). Note that because \mathbf{g} depends on the stimulus, θ , also J_{mean} is stimulus dependent. We usually omit this dependence on θ for clarity.

J_{mean} in homogeneous population model. To illustrate the formula, we consider a homogeneous population of neurons (see Fig. 1E) with limited range correlation structure (see Fig. 1A–D). In the general case, J_{mean} is a function of the stimulus, θ . However, for a homogeneous population of neurons, where the tuning functions are broad compared with the spacing between the preferred directions, J_{mean} can be treated as independent of θ . We can therefore restrict our analysis to the case $\theta = 0$. In this case, the average population activity, \mathbf{f} , is given by $f_j = f(-\phi_j)$, where ϕ_j is the preferred direction of neuron j (see Fig. 1F). The covariance matrix \mathbf{Q} is shown in Figure 1B. Figure 1, C and G, show \mathbf{r} and \mathbf{g} , whose Fourier transforms (Fig. 1D, H) are the two main quantities entering Equation 14. Because both \mathbf{f} and \mathbf{r} are smooth and slowly varying, \mathbf{g} has almost all of its power in low frequencies, the power spectrum converging to zero for higher frequencies. We can write \mathbf{r} as $r_j = \delta_j(1 - c_0) + c_j$. Its Fourier transform can be split up into two parts as well:

$$\tilde{\mathbf{r}} = \tilde{\mathbf{c}} + (1 - c_0) \quad (15)$$

Because \mathbf{c} (Eq. 7) is also a smooth and slowly varying function of the differences in preferred direction, it also has most of its power in the low frequencies, power in higher frequencies quickly converging to zero. The delta peak at zero (because each neuron is correlated with itself by 1, while neighboring neurons with similar preferred directions have a correlation of c_0) has a constant Fourier transform of magnitude $1 - c_0$. Together, this results in a power spectrum with high power in low frequencies decaying to a constant offset $1 - c_0$ at high frequencies.

Each of the terms $|\tilde{g}_k|^2/\tilde{r}_k$ in Equation 14 can be seen as a signal-to-noise ratio of the k th Fourier mode of the population (Fig. 1I). J_{mean} is then simply the sum over the individual signal-to-noise ratios. For the homogeneous case, the only difference to the additive noise case studied by Sompolinsky et al. (2001) is that, for other mean-variance relationships, the tuning curve derivatives are normalized by the standard deviations. Because of this, the scaling behavior of J_{mean} for large populations is similar to the additive case: the low-frequency Fourier components of signal and noise grow at the same rate with n , leading to a saturation of J_{mean} for large networks (see Fig. 3A).

J_{mean} in random amplitude model. The above considerations suggest that the saturation of J_{mean} can be avoided by introducing high-frequency components into the signal, $\tilde{\mathbf{g}}$, for which the noise amplitude is small. In fact, this is naturally the case for any realistic population of neurons. The vanishing power in high frequencies for the homogeneous population model is a result of the simplifying assumption of identical tuning functions for all neurons. This results in the mean population activity \mathbf{f} having the same shape as the tuning function f evaluated at the preferred directions of the neurons. In realistic populations of neurons, however, tuning curves display a significant amount of heterogeneity between neurons—such as different amplitudes, widths, or baselines of their tuning functions—introducing high-frequency components into \mathbf{f} and, hence, also in \mathbf{g} .

To illustrate this point, we ignored all types of heterogeneity except for the overall amplitudes of the neurons and constructed a model population by assigning each neuron a tuning curve with a peak amplitude randomly drawn from a distribution of amplitudes that has the same mean as the homogeneous population (see Fig. 1J). Figure 1, K and L, shows the resulting mean population activity \mathbf{f} as well as the normalized derivatives \mathbf{g} for this population at stimulus $\theta = 0$. Because of the randomly selected amplitudes for each neuron, \mathbf{g} has power in all frequencies (see Fig. 1M). Intuitively, this affects J_{mean} positively in two ways. First, the overall number of Fourier components significantly contributing to J_{mean} is increased. Second, the signal-to-noise ratio is better for the high-frequency components \tilde{g}_k because \tilde{r}_k is small for large k (it converges to $1 - c_0$).

The exact value of J_{mean} depends on the specific set of amplitudes a_j that are drawn at random. We here generalize the results of Shamir and

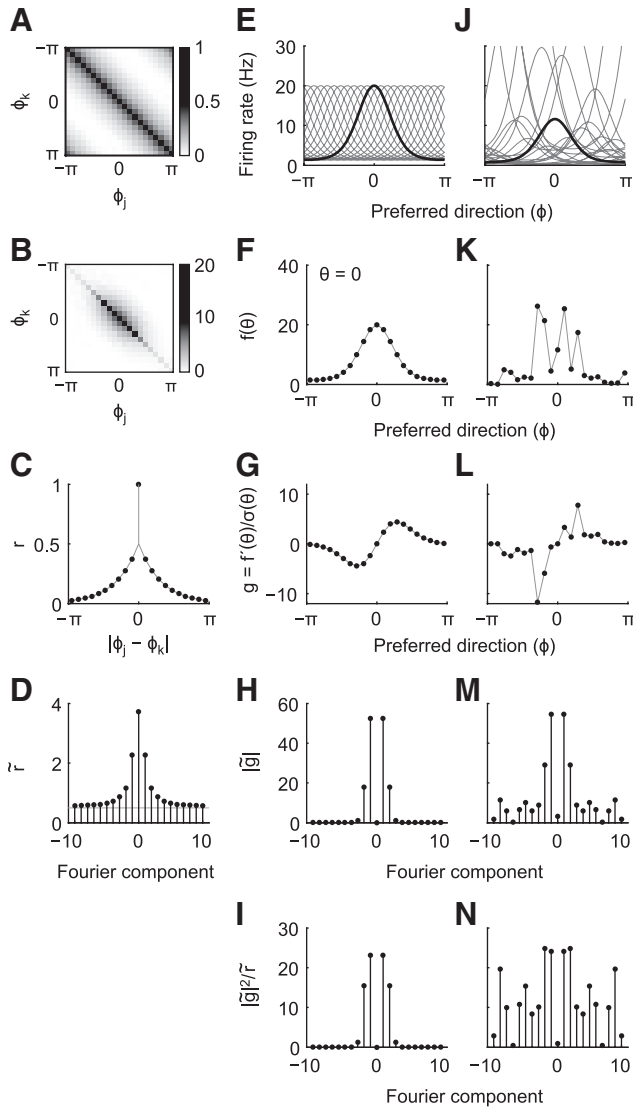


Figure 1. Illustration of the calculation of the linear Fisher information. **A–D**, Terms related to correlation structure. **A**, Matrix of correlation coefficients. The population has a limited range correlation structure. The correlation coefficient of two neurons depends only on the difference between their preferred directions. **B**, Covariance matrix. Variances are equal to mean spike counts. **C**, Correlation structure \mathbf{r} . This corresponds to a single slice through the correlation coefficient matrix shown in **A**. **D**, Fourier transform of correlation structure shown in **C**. This is the power spectrum of the noise after normalizing to unit variances. **E–I**, Terms related to tuning curves in homogeneous population model. **E**, The tuning curves of the neurons. For clarity, one tuning curve (with preferred direction $\phi = 0$) is highlighted. **F**, Average population response for $\theta = 0$ given by $f_j = f(-\phi_j)$. All the following panels are evaluated at $\theta = 0$. **G**, Normalized derivative \mathbf{g} of the tuning curve. It looks flipped about the y -axis because $g_j = f'(-\phi_j)/\sigma(-\phi_j)$ and f' is not symmetric. **H**, Power spectrum of normalized derivatives shown in **G**. **I**, Signal-to-noise ratio of individual Fourier components. J_{mean} is the sum over these terms. **J–N**, Terms related to tuning curves in heterogeneous population model. The panels are analogous to **E–I**. **J**, The tuning curves of the neurons. The average firing rate over all neurons and conditions is the same as for the homogeneous population. The tuning curve of one neuron is highlighted. **K**, Average population response for stimulus $\theta = 0$. Because each neuron has a different peak amplitude, the population hill looks scattered around the mean tuning curve. Note that this scatter does not reflect noise. **L**, Normalized tuning curve derivatives \mathbf{g} . **M**, Power spectrum of normalized tuning curve derivatives shown in **L**. Note that there is substantial power also in the higher frequencies (compare with **H**). **N**, Signal-to-noise ratio of all Fourier components.

Sompolinsky (2006) to the case with nonadditive, stimulus-dependent noise. In the random amplitude model with Poisson-like noise,

$$g_j = \sqrt{a_j} \frac{f'(-\phi_k)}{\sqrt{f(-\phi_k)}} \equiv \sqrt{a_j} q_j. \quad (16)$$

We define $\sqrt{a_j} = \mu + b_j$, which splits $\sqrt{a_j}$ into a nonrandom component μ and a random component b_j with mean zero and variance $\kappa = \text{Var}[\sqrt{a_j}]$. Plugging this into the formula for J_{mean} , we obtain

$$\begin{aligned} \langle J_{\text{mean}} \rangle &= \langle [(\mu + \mathbf{b}) \cdot \mathbf{q}]^T \mathbf{R}^{-1} [(\mu + \mathbf{b}) \cdot \mathbf{q}] \rangle \\ &= \mu^2 \mathbf{q}^T \mathbf{R}^{-1} \mathbf{q} + 2\mu \mathbf{q}^T \mathbf{R}^{-1} \langle \mathbf{b} \cdot \mathbf{q} \rangle + \langle [\mathbf{b} \cdot \mathbf{q}]^T \mathbf{U} \tilde{\mathbf{R}}^{-1} \mathbf{U}^* [\mathbf{b} \cdot \mathbf{q}] \rangle \\ &= \mu^2 J_{\text{mean}}^{\text{hom}} + n\kappa d, \end{aligned} \quad (17)$$

where $J_{\text{mean}}^{\text{hom}}$ is the linear Fisher information of a homogeneous population and d converges to a constant independent of n and is defined as follows:

$$d = \frac{1}{n^2} \sum_j q_j^2 \sum_k \frac{1}{\tilde{r}_k}, \quad (18)$$

To arrive at Equation 17, we used

$$\langle [\mathbf{b} \cdot \mathbf{q}]^T \mathbf{u}_k \rangle^2 = \sum_i \sum_j \langle b_i b_j \rangle q_i q_j u_{ik} u_{jk}^* = \frac{\kappa}{n} \sum_j q_j^2, \quad (19)$$

which holds because b_j is white noise with variance κ (i.e., $\langle b_i b_j \rangle = \delta_{ij} \kappa$) and $u_{jk} u_{jk}^* = 1/n$.

The above calculations can be generalized to non-Poisson mean-variance relationships. For instance, if $\sigma(\theta) = f(\theta)^\alpha$, define $a_j^{1-\alpha} = \mu + b_j$ and $\kappa = \text{Var}[a_j^{1-\alpha}]$ and Equation 17 will still be valid. By setting $\alpha = 0$, the result of Shamir and Sompolinsky (2006) is obtained. Unless $\alpha = 1$ (i.e., the standard deviations are equal to the means), populations with amplitude variability will have J_{mean} asymptotically proportional to n even in the presence of limited range correlations.

Furthermore, from Equation 17, one can see that for an independent population of Poisson-like neurons, amplitude variability does not affect J_{mean} on average. Because for an independent population $\tilde{r}_k = 1$, we have

$$\langle J_{\text{mean}} | c_0 = 0 \rangle = (\mu^2 + \kappa) \sum_j q_j^2 = \sum_j q_j^2, \quad (20)$$

which is independent of κ and identical with that of a homogeneous population.

Considering the large n limit of J_{mean} of a correlated population relative to that of an independent population, we obtain

$$\lim_{n \rightarrow \infty} \frac{\langle J_{\text{mean}} | c_0 > 0 \rangle}{\langle J_{\text{mean}} | c_0 = 0 \rangle} = \lim_{n \rightarrow \infty} \frac{\mu^2 J_{\text{mean}}^{\text{hom}} + \frac{\kappa}{n} \sum_j q_j^2 \sum_k \frac{1}{\tilde{r}_k}}{\sum_j q_j^2} = \frac{\kappa}{1 - c_0}. \quad (21)$$

The first term in the numerator above saturates to a finite value and can therefore be ignored. For the second term, note that \mathbf{r} is assumed to be smooth and slowly varying, in which case only $O(\sqrt{n})$ Fourier components \tilde{r}_k are large and for the remaining components $\tilde{r}_k \rightarrow 1 - c_0$. Thus, $(1/n) \sum_k 1/\tilde{r}_k \rightarrow 1/(1 - c_0)$ for large n .

Calculation of J_{cov} . Using similar methods as above, we also derived an expression for J_{cov} in terms of Fourier transforms that does not contain an inverse of the covariance matrix anymore:

$$J_{\text{cov}} = \sum_{k=0}^{n-1} h_k^2 + \frac{1}{n^2} \sum_{k=0}^{n-1} \tilde{h}_k^2 \cdot \left[\tilde{\mathbf{r}} * \frac{1}{\tilde{\mathbf{r}}} \right]_k, \quad (22)$$

where $h_k = \sigma'_k/\sigma_k$ and \tilde{h} is the discrete Fourier transform of \mathbf{h} . The expression $[\tilde{\mathbf{r}} * 1/\tilde{\mathbf{r}}]_k$ is the k th component of the circular convolution of $\tilde{\mathbf{r}}$ with its pointwise inverse.

We briefly outline the derivation in the following. First, note that the derivative of the covariance matrix with respect to θ is $\mathbf{Q}' = \mathbf{S}'\mathbf{R}\mathbf{S} + \mathbf{S}\mathbf{R}\mathbf{S}'$, where $\mathbf{S}' = \text{Diag}(\boldsymbol{\sigma}')$. Substituting this expression into Equation 9, expanding the square, and using the matrix trace identity $\text{Tr}[\mathbf{ABC}] = \text{Tr}[\mathbf{BCA}]$, we obtain

$$J_{\text{cov}} = \text{Tr}[\mathbf{S}'\mathbf{S}^{-1}\mathbf{S}'\mathbf{S}^{-1}] + \text{Tr}[\mathbf{S}'\mathbf{R}\mathbf{S}'\mathbf{S}^{-1}\mathbf{R}^{-1}\mathbf{S}^{-1}]. \quad (23)$$

Because \mathbf{S} is diagonal, the first term reduces to

$$\text{Tr}[\mathbf{S}'\mathbf{S}^{-1}\mathbf{S}'\mathbf{S}^{-1}] = \sum_{k=0}^{n-1} h_k^2. \quad (24)$$

To simplify the second term, let $\mathbf{V} = \mathbf{U}^*\mathbf{S}^{-1}\mathbf{S}'\mathbf{U}\tilde{\mathbf{R}}$ and $\mathbf{W} = \mathbf{U}^*\mathbf{S}'\mathbf{S}^{-1}\mathbf{U}\tilde{\mathbf{R}}^{-1}$. Because $\mathbf{S}^{-1}\mathbf{S}'$ is diagonal, $\mathbf{U}^*\mathbf{S}^{-1}\mathbf{S}'\mathbf{U}$ is circulant with the first column being the inverse Fourier transform of the diagonal elements. A right multiplication by the diagonal matrices $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{R}}^{-1}$ scales the columns, resulting in the following:

$$v_{jk} = \frac{1}{n} \tilde{r}_k \tilde{h}_{k-j} \quad w_{jk} = \frac{1}{n} \frac{\tilde{h}_{k-j}}{\tilde{r}_k}. \quad (25)$$

Substituting \mathbf{V} and \mathbf{W} into the second term for J_{cov} , we obtain

$$\text{Tr}[\mathbf{S}'\mathbf{R}\mathbf{S}'\mathbf{S}^{-1}\mathbf{R}^{-1}\mathbf{S}^{-1}] = \text{Tr}[\mathbf{V}\mathbf{W}] = \frac{1}{n^2} \sum_{l=0}^{n-1} |\tilde{h}_l|^2 \cdot \left[\tilde{r} * \frac{1}{\tilde{r}} \right]_l. \quad (26)$$

For an independent population of neurons, $\tilde{r}_k = 1$, and by using Parseval's theorem, the simple formula for J_d , the J_{cov} of an independent population of neurons (Shamir and Sompolinsky, 2001) is recovered:

$$J_d = 2 \sum_{k=0}^{n-1} \left(\frac{\sigma'_k}{\sigma_k} \right)^2. \quad (27)$$

Note that, for correlated populations, J_{cov} can be bounded from above and below by the following:

$$J_d \leq J_{\text{cov}} \leq \frac{1}{2} \left(1 + \frac{1}{1-c_0} \right) J_d. \quad (28)$$

Thus, asymptotically J_{cov} always grows linearly with n , regardless of the correlation structure. In addition, unlike for J_{mean} , small correlations do not substantially alter J_{cov} compared with independence (because the upper and lower bound become equal for $c_0 \rightarrow 0$). In addition, it is easy to see that J_{cov} is unaffected by amplitude variability because the amplitudes a_k appear both in the numerator and the denominator of h_k and cancel.

Effect of correlation structure under constant noise entropy. To study the effect of the average level of correlations under the constraint of constant amount of total noise, we relax the assumption of Poisson-like noise and adjust the Fano factors of the neurons such that the noise entropy is kept constant. We define

$$\sigma_j^2(c_0) = F(c_0) f_j. \quad (29)$$

Note that, in this section, we write most quantities as a function of c_0 as we are interested in their behavior with varying c_0 . We can now adjust $F(c_0)$ such that the noise entropy remains constant as we vary c_0 . For Gaussian noise, the differential entropy is given by the following:

$$h(\mathbf{y}) = \ln \sqrt{(2\pi e)^k |\mathbf{Q}|}. \quad (30)$$

In the above formula, the only quantity that depends on c_0 is $|\mathbf{Q}|$. Thus, to have constant noise entropy, we need $|\mathbf{Q}(c_0)| = |\mathbf{Q}(0)|$, where $\mathbf{Q}(0)$ is the covariance matrix of an independent population of neurons ($c_0 = 0$). We can write for the determinant $|\mathbf{Q}| = |\mathbf{S}|^2 \cdot |\mathbf{R}| = |\mathbf{V}| \cdot |\mathbf{R}|$, where $\mathbf{S} = \text{Diag}(\boldsymbol{\sigma})$ is the diagonal matrix containing the standard deviations and $\mathbf{V} = \text{Diag}(\boldsymbol{\sigma}^2)$. For the independent population, we have $\mathbf{Q}(0) = \mathbf{V}(0)$.

Writing the determinants as functions of c_0 and requiring constant entropy, we obtain

$$|\mathbf{V}(0)| = F(c_0)^n |\mathbf{V}(0)| \cdot |\mathbf{R}(c_0)|. \quad (31)$$

Solving for the Fano factor $F(c_0)$ results in

$$F(c_0) = |\mathbf{R}(c_0)|^{-1/n} = \left(\prod_{k=0}^{n-1} \tilde{r}_k \right)^{-1/n}, \quad (32)$$

which is the inverse of the geometric mean of the Fourier coefficients \tilde{r}_k of the correlation structure.

Applying this constant entropy constraint and considering the limit of large populations ($n \rightarrow \infty$), we find for the dependence of J_{mean} on c_0 the following:

$$\lim_{n \rightarrow \infty} \frac{\langle J_{\text{mean}} | c_0 > 0 \rangle}{\langle J_{\text{mean}} | c_0 = 0 \rangle} = \lim_{n \rightarrow \infty} \frac{\mu^2 J_{\text{mean}}^{\text{hom}} + \frac{\kappa}{nF(c_0)} \sum_j q_j^2 \sum_k \frac{1}{\tilde{r}_k}}{\sum_j q_j^2} = \kappa. \quad (33)$$

As before (Eq. 21), $J_{\text{mean}}^{\text{hom}}$ saturates to a finite value and therefore the first term in the numerator does not play a role. For the second term, note that, for large populations, $F(c_0) \rightarrow 1/(1-c_0)$ and $(1/n) \sum_k 1/\tilde{r}_k \rightarrow 1/(1-c_0)$, which leads to the above result.

Maximum-likelihood estimation. Under a Gaussian noise model, the log-likelihood function is

$$l(\theta) = \frac{n}{2} \log(2\pi) + \frac{1}{2} \log |\mathbf{Q}(\theta)| + \frac{1}{2} (\mathbf{y} - \mathbf{f}(\theta))^T \mathbf{Q}(\theta)^{-1} (\mathbf{y} - \mathbf{f}(\theta)). \quad (34)$$

Using the fact that $\mathbf{Q}(\theta) = \mathbf{S}(\theta)\mathbf{R}\mathbf{S}(\theta)$, we can rewrite the terms that depend on θ as follows:

$$\begin{aligned} \log |\mathbf{Q}(\theta)| &= \log |\mathbf{R}| + 2 \log |\mathbf{S}(\theta)| \\ &= \log |\mathbf{R}| + 2 \sum_{k=0}^{n-1} \log \sigma_k(\theta) \end{aligned} \quad (35)$$

and

$$\begin{aligned} (\mathbf{y} - \mathbf{f}(\theta))^T \mathbf{Q}(\theta)^{-1} (\mathbf{y} - \mathbf{f}(\theta)) &= (\mathbf{y} - \mathbf{f}(\theta))^T \mathbf{S}(\theta)^{-1} \mathbf{R}^{-1} \mathbf{S}(\theta)^{-1} (\mathbf{y} - \mathbf{f}(\theta)) \\ &= \sum_{k=0}^{n-1} \frac{|\tilde{z}_k(\theta)|^2}{\tilde{r}_k}, \end{aligned} \quad (36)$$

where we have defined $\mathbf{z}(\theta) = \mathbf{S}(\theta)^{-1} (\mathbf{y} - \mathbf{f}(\theta))$. Combining all steps and dropping terms that do not depend on θ , we arrive at

$$l(\theta) \sim \sum_{k=0}^{n-1} \log \sigma_k(\theta) + \frac{1}{2} \sum_{k=0}^{n-1} \frac{|\tilde{z}_k(\theta)|^2}{\tilde{r}_k}. \quad (37)$$

With the additional assumption of Poisson-like noise (i.e., $\sigma_k^2(\theta) = f_k(\theta)$) and using the same diagonalization of the quadratic forms as above, we obtain for the first and second derivatives

$$l'(\theta) = \frac{1}{2} \sum_{k=0}^{n-1} \frac{f'_k(\theta)}{f_k(\theta)} + \sum_{k=0}^{n-1} \frac{\tilde{z}_k^*(\theta) \tilde{z}'_k(\theta)}{\tilde{r}_k} \quad (38)$$

and

$$I''(\theta) = \frac{1}{2} \sum_{k=0}^{n-1} \frac{f_k''(\theta) f_k(\theta) - f_k'(\theta)^2}{f_k^2(\theta)} + \sum_{k=0}^{n-1} \frac{\bar{z}_k'(\theta) \bar{z}_k'(\theta) + \bar{z}_k^*(\theta) \bar{z}_k''(\theta)}{\bar{r}_k}. \quad (39)$$

Here, \bar{z}_k' is the k th component of the discrete Fourier transform of the first derivative of $\mathbf{z}(\theta)$ with respect to θ .

We evaluated the maximum-likelihood estimator (MLE) for homogeneous and heterogeneous populations. We used Newton's method to numerically find the maximum-likelihood estimate $\hat{\theta}$. For the heterogeneous populations, the amplitudes of the neurons a_k were drawn from a lognormal distribution with mean 1 and variance adjusted manually such that it resulted in $\kappa = 0.25$. All parameters for tuning functions and correlations were defined as above. For each population size (minimum, 32; maximum, 4096 neurons), we generated 4096 realizations of heterogeneous populations. The error of the MLE was evaluated at $m = 32$ regularly spaced stimulus values $\theta_k = 2\pi k/m$ for 64 random samples drawn from a normal distribution with mean $\mathbf{f}(\theta)$ and covariance $\mathbf{Q}(\theta)$. Squared errors $(\hat{\theta} - \theta)^2$ for all populations (of equal size), stimuli, and samples were averaged to calculate the mean squared error, the inverse of which was taken as the efficiency of the maximum-likelihood estimator.

Optimal linear decoding. For linear decoding, it is convenient to rewrite the stimulus as a vector $\mathbf{x} = [\cos(\theta), \sin(\theta)]^T$ on the unit circle. The direction θ is easily recovered from \mathbf{x} via $\theta = \text{atan2}(x_2, x_1)$. The optimal linear estimator (OLE) is defined as the linear estimator minimizing the mean squared error $\langle (\hat{\mathbf{x}} - \mathbf{x})^2 \rangle$. It minimizes the mean squared error for any kind of noise distribution (Salinas and Abbott, 1994) (no Gaussian assumption is necessary) and its weights are given by

$$\mathbf{w} = \mathbf{Q}_{yy} \mathbf{Q}_{yy}^{-1}, \quad (40)$$

where \mathbf{Q}_{yy} is the response covariance matrix over all stimuli,

$$\mathbf{Q}_{yy} = \int_{\theta} [\bar{\mathbf{f}}(\theta) \bar{\mathbf{f}}(\theta)^T + \mathbf{Q}(\theta)] d\theta, \quad (41)$$

where $\bar{\mathbf{f}}(\theta) = \mathbf{f}(\theta) - \int_{\theta} \mathbf{f}(\theta) d\theta$, and \mathbf{Q}_{xy} is the cross-covariance between stimulus \mathbf{x} and neural response \mathbf{y} , given by

$$\mathbf{Q}_{xy} = \int_{\theta} \mathbf{x}(\theta) \bar{\mathbf{f}}(\theta)^T d\theta. \quad (42)$$

We numerically estimated bias, variance, and mean squared error of the optimal linear estimator at $m = 32$ different stimulus values $\theta_k = 2\pi k/m$. For each population size (minimum, 64; maximum, 4096 neurons), we generated 8192 realizations of heterogeneous populations with the same parameters as defined above.

Results

We consider a simple model in which n neurons encode a one-dimensional circular stimulus (e.g., direction of motion) through bell-shaped tuning functions (for details on the model, see Materials and Methods). We introduce heterogeneity in the population by allowing the neurons to have different peak firing rates (amplitudes, a_j) but identical tuning widths. The degree of heterogeneity is controlled by the amount of variability in the amplitudes, controlled by the parameter $\kappa = \text{Var}[\sqrt{a_j}]$. Setting $\kappa = 0$ results in a homogeneous population in which all neurons have identical tuning curves. Setting $\kappa > 0$ results in a heterogeneous population (amplitude variability model). We set $\kappa = 0.25$ as determined from a dataset of 408 orientation tuning curves of neurons recorded in macaque primary visual cortex (Ecker et al., 2010). We further assume that the noise correlation between two neurons depends only on their difference in preferred direction (limited range structure) and that variances of the neurons are Poisson-like

(i.e., the variances are equal to the mean firing rates). In our model, the level of correlations is controlled by c_0 , the correlation coefficient of two neurons with identical preferred directions. The relationship of c_0 to the average level of correlations $\langle r \rangle$ depends on the decay constant in the correlation structure (Fig. 1C). For the set of parameters we used, $\langle r \rangle$ is $\sim 0.3 c_0$. Our choice of parameterization by c_0 is motivated by the fact that, for large populations, it is the more relevant quantity compared with the average correlations.

Although we have to choose a specific set of parameters for the figures, our results hold more generally, as long as both tuning curves and correlation structure do not change as a function of the population size and are sufficiently smooth and slowly varying.

Dependence of Fisher information on population size

To quantify the accuracy of a population code, we calculate the Fisher information, J , which under the assumption of Gaussian noise can be written as the sum of two terms as follows:

$$J = J_{\text{mean}} + J_{\text{cov}}. \quad (43)$$

J_{mean} can be thought of as the information that is encoded in the average population activity, while J_{cov} is the information contained in the variances and covariances.

The difficulty in evaluating the Fisher information for large populations of neurons lies in inverting the $n \times n$ covariance matrix. Following the approach of Sompolinsky and coworkers (Sompolinsky et al., 2001; Shamir and Sompolinsky, 2006), we obtained an analytic expression for this inverse in our model, leading to an expression for the Fisher information that is considerably easier to study, even for populations with tens of thousands of neurons (for details, see Materials and Methods; Fig. 1; Eqs. 14, 22).

We first study how the Fisher information depends on the number of neurons in the population. This question has been addressed by a number of authors who reported different results (Abbott and Dayan, 1999; Sompolinsky et al., 2001; Wilke and Eurich, 2002; Shamir and Sompolinsky, 2006). The apparent discrepancy arises from subtle differences in the assumptions that were made about the population activity, such as for instance the noise model (additive vs Poisson-like). In the following, we provide a comprehensive treatment of this problem using our framework that includes most of the previous studies as special cases.

Figure 2 shows the total Fisher information, J , as a function of the population size, while Figure 3 splits J into its two components, J_{mean} and J_{cov} . For homogeneous populations, the total Fisher information grows with increasing population size and does not saturate to a finite bound, even in the presence of limited range correlations (Fig. 2A) (Wilke and Eurich, 2002). This is because the second term, J_{cov} , increases linearly with n (Fig. 3A) if the variances of the neurons are stimulus dependent (a property of Poisson-like noise). In contrast, J_{mean} saturates to a finite value if neurons are correlated (Fig. 3A) (Sompolinsky et al., 2001). Interestingly, for independent neurons with Poisson-like noise, the degree of amplitude variability does not affect the Fisher information (Fig. 2, compare black lines in A, B; see Materials and Methods, Eq. 20). If neurons are correlated, however, the total Fisher information is generally higher for heterogeneous populations than for homogeneous ones (Fig. 2, compare A, B). Responsible for this difference is J_{mean} , which no longer saturates in the presence of correlations if neurons have heterogeneous tuning functions (Fig. 3B, Eq. 17). In contrast, J_{cov} is unaffected by heterogeneity (Fig. 3, dashed lines).

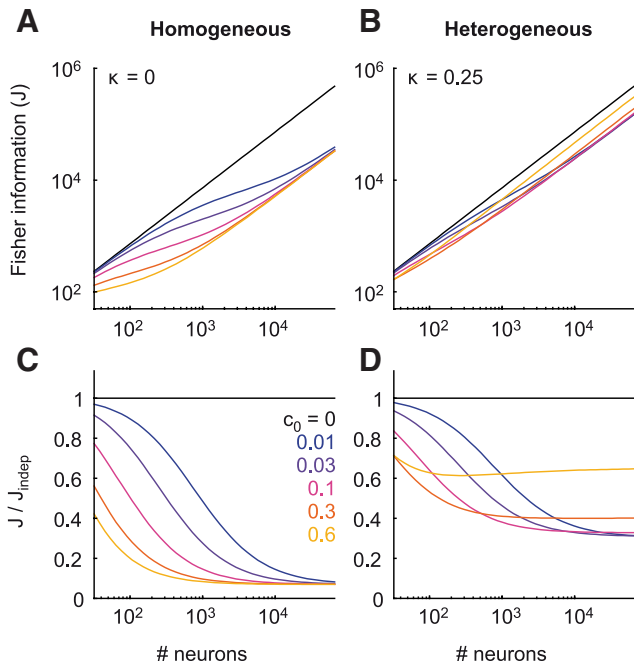


Figure 2. *A*, Fisher information as a function of population size in a homogeneous population of neurons (black line, independent population; colored lines, correlated populations; see legend in *C*). *B*, Same as in *A* but for heterogeneous population of neurons ($\kappa = 0.25$). *C*, Fisher information relative to independent population (J/J_{indep}) for a homogeneous population. *D*, Same as in *C* but for heterogeneous population of neurons.

In summary, the Fisher information saturates to a finite bound only if the variances of the neurons do not depend on the stimulus and all neurons have identical tuning curves. If one of these two conditions is not satisfied, the Fisher information increases linearly with the population size. While for large homogeneous populations it is dominated by J_{cov} (Fig. 3C), most of the information in a heterogeneous population is contributed by J_{mean} (Fig. 3D).

Dependence of Fisher information on magnitude of correlations

We now investigate how the magnitude of noise correlations affects the accuracy of a population code. Generally, small limited-range correlations decrease the accuracy compared with the independent case, consistent with previous reports (Zohary et al., 1994; Abbott and Dayan, 1999; Sompolinsky et al., 2001; Wilke and Eurich, 2002). While in homogeneous populations this detrimental effect becomes stronger with increasing correlations (Fig. 2C), in heterogeneous populations it is both non-monotonic and population size dependent (Fig. 2D). These differences between homogeneous and heterogeneous populations are exclusively due to differences in J_{mean} , as J_{cov} is independent of both the degree of heterogeneity and the level of correlations (Fig. 3, dashed lines).

To characterize the effect of varying the magnitude of correlations, we quantified J_{mean} of a correlated population relative to the total Fisher information of an independent population. We refer to this quantity ($J_{\text{mean}}/J_{\text{indep}}$) as the relative J_{mean} . While in homogeneous populations the relative J_{mean} decreases with increasing n and increasing correlations (Fig. 4A), in the presence of tuning curve variability a higher level of limited range correlations can also increase J_{mean} substantially, in particular if the population is large (Fig. 4B).

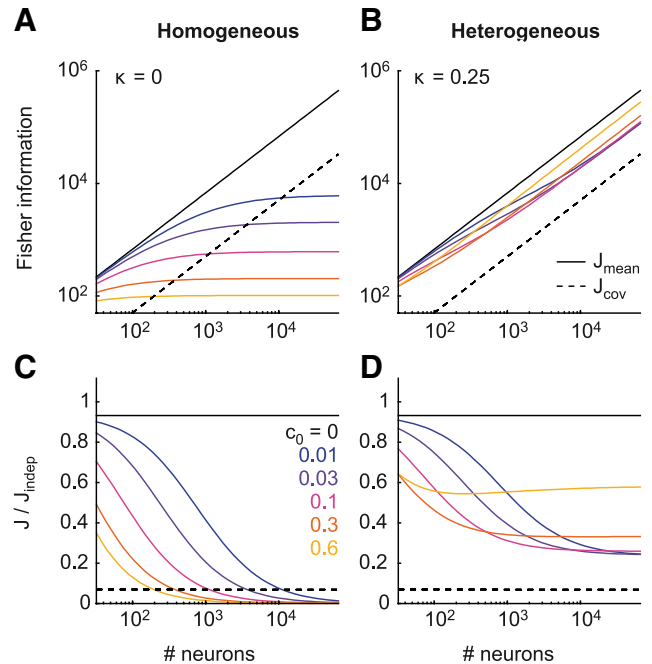


Figure 3. Relative contributions of J_{mean} and J_{cov} . The panels are analogous to those of Figure 2. The solid lines represent J_{mean} ; the dashed lines represent J_{cov} . Colors represent different levels of correlation (see legend in *C*). *A*, J_{mean} and J_{cov} as a function of population size in a homogeneous population of neurons. *B*, Same as in *A* but for heterogeneous population of neurons ($\kappa = 0.25$). *C*, J_{mean} and J_{cov} for a homogeneous population relative to the total Fisher information of an independent population. *D*, Same as in *C* but for heterogeneous population of neurons.

There are two regimes for the effect of changes in correlation strength: one in which reducing correlation improves and one in which it impairs the performance. Which of the two applies to a given population code depends on whether correlations are smaller or larger than the minimum in the curves in Figure 4B. We term this value the “maximally detrimental” correlation strength c_{min} . If correlations are below c_{min} , decreasing them further improves the population code. In this regime, which we refer to as the “low-correlation regime,” the behavior is similar to that of homogeneous populations. In the “high-correlation regime,” in contrast, increasing the correlation strength improves the population code. The value of c_{min} is inversely related to the population size and decreases with $O(1/\sqrt{n})$ (Fig. 5). For large populations, it converges to zero and, hence, increasing correlations is almost always advantageous. In particular, in the large n limit the value of J_{mean} relative to the independent case is given by the following (see Materials and Methods, Eq. 21):

$$\frac{\langle J_{\text{mean}} | c_0 > 0 \rangle}{\langle J_{\text{mean}} | c_0 = 0 \rangle} \rightarrow \frac{\kappa}{1 - c_0}. \tag{44}$$

Thus, asymptotically the relative performance depends only on the correlation coefficient of two neurons with identical preferred orientations, c_0 , and the degree of heterogeneity, κ . The more heterogeneous a population is, the higher J_{mean} becomes (Fig. 6). However, it should be noted that J_{mean} cannot be increased arbitrarily by simply increasing the heterogeneity, κ , because the amplitudes a_i are constrained to be positive, have an average of 1, and neurons have a maximum possible firing rate due to biophysical constraints. At the same time, the value $\kappa = 0.25$ that we used for the above figures is likely to be an underestimate because in real neural populations sources

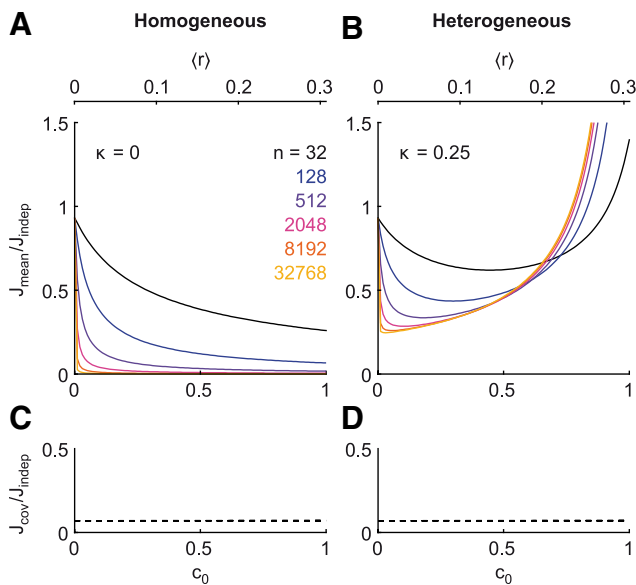


Figure 4. *A*, Relative J_{mean} as a function of correlation strength (c_0) in a homogeneous population of neurons (different colors indicate population sizes). *B*, Same as in *A* but for heterogeneous population of neurons ($\kappa = 0.25$). *C*, Relative J_{cov} in homogeneous population. *D*, Relative J_{cov} in heterogeneous population.

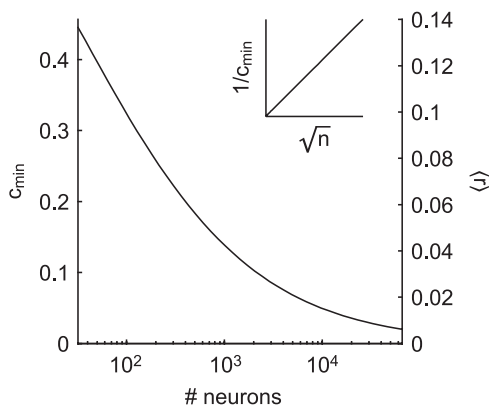


Figure 5. Maximally detrimental level of correlations, c_{min} . This is the minimum of the curves in Figure 4*B*. Inset, The inverse of c_{min} is proportional to \sqrt{n} (i.e., c_{min} decreases as $O(1/\sqrt{n})$).

of heterogeneity other than amplitudes exist. We explore this issue further below.

Maximum-likelihood estimator attains Cramér–Rao bound

One potential caveat that has to be addressed when using Fisher information is that it provides only a bound (the Cramér–Rao bound) on the accuracy of a population code. Unfortunately, under some conditions potentially relevant for neural coding this bound is not tight, which means there exists no estimator that achieves this performance (Bethge et al., 2002; Berens et al., 2011). If certain assumptions are placed on the statistics of population activity (e.g., independent and identically distributed samples), the maximum-likelihood estimator can be proven to converge (for large n) to the Cramér–Rao bound. However, under the conditions studied here, the responses are neither independent nor identically distributed and we do not know whether and how fast the bound is attained. Thus, it is unclear whether comparing Fisher information under different conditions (such as different levels of correlation or different amounts of hetero-

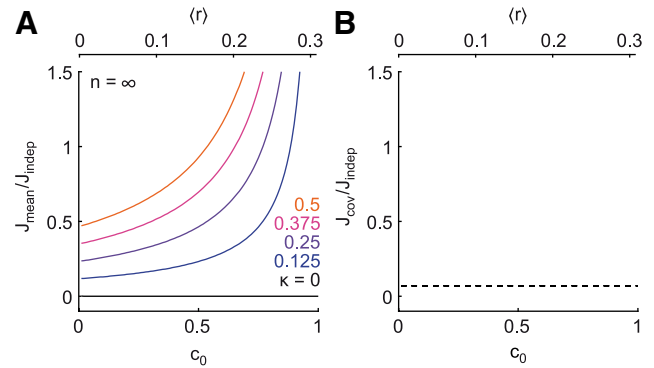


Figure 6. Asymptotic (large n limit) Fisher information relative to independent population. *A*, Relative J_{mean} (black line, homogeneous population; colored lines, degree of heterogeneity, κ). Asymptotically, increased correlations always increase encoding accuracy (unless the neurons are independent). *B*, Relative J_{cov} is unaffected by heterogeneity in our model.

geneity) can provide insights into the accuracy of population codes because comparing loose upper bounds is meaningless. To address this problem, we additionally evaluated the efficiency of the maximum-likelihood estimator numerically (for details, see Materials and Methods). For both homogeneous and heterogeneous populations, it attains the Cramér–Rao bound very quickly (Fig. 7), even in the case of nonidentical and nonindependent samples. Interestingly, for homogeneous populations, the rate of convergence depends more on the level of correlations than for heterogeneous populations (Fig. 7). More importantly, for both types of population models, the performance of the maximum-likelihood estimator is within 5% of the Cramér–Rao bound for all population sizes >64 neurons. Thus, the bound is sufficiently tight and the use of Fisher information is well justified.

Other sources of heterogeneity

Another issue that has to be addressed is whether our findings above generalize to other forms of tuning curve heterogeneity, such as variable widths and baseline firing rates. Because only the amplitude variability model is analytically tractable, we ran numerical simulations to estimate the Fisher information for populations with other forms of heterogeneity. First, we varied the tuning widths while keeping all other parameters constant (Fig. 8*A, C*). Second, we created populations of neurons by randomly drawing sets of tuning parameters from a dataset of orientation tuning curves in monkey V1 (Fig. 8*B, D*). In both cases, the dependence of the Fisher information on the level of correlations is similar to that in the amplitude variability model. One notable difference is that J_{cov} is not completely independent of correlation strength if parameters other than tuning amplitude are variable. For moderate levels of correlation, however, the differences are relatively small.

In the above analysis, we assumed that the preferred directions of the neurons are arranged on a regular grid around the circle. As this is not the case in real neural populations, we numerically analyzed the effect of this assumption on the results presented above. We found that the Fisher information is virtually unaffected by randomly assigning preferred directions compared with equally spacing them (data not shown).

In addition, in real neural populations not only the tuning curves are heterogeneous but also the pairwise correlation coefficients. This case has been studied by adding independent Gaussian noise (with variance σ^2) on each correlation coefficient (Wilke and Eurich, 2002; Shamir and Sompolinsky, 2006). Un-

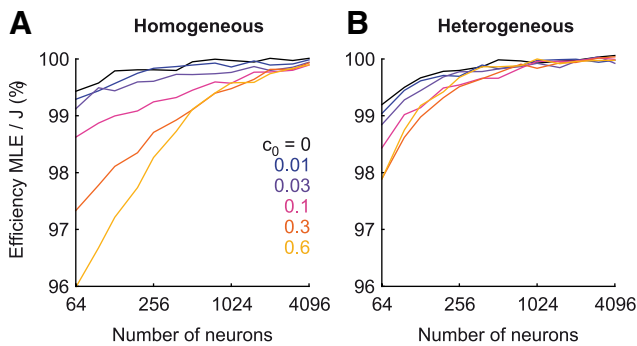


Figure 7. MLE attains the Cramér–Rao bound. **A**, Efficiency of MLE relative to Fisher information for homogeneous population of neurons. The colors indicate different levels of correlation. **B**, As in **A** but for heterogeneous population.

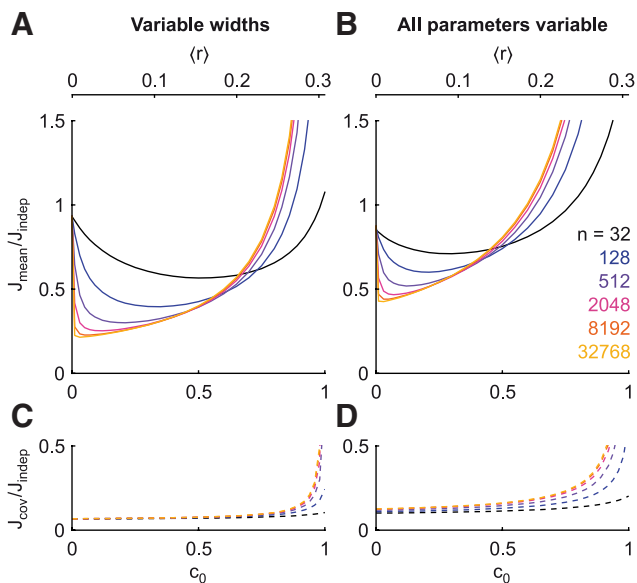


Figure 8. Fisher information for other types of tuning curve heterogeneity. **A**, J_{mean} for populations with random tuning widths. **B**, J_{mean} for populations with tuning parameters (baseline, amplitude, and width) sampled with replacement from a dataset of 408 orientation tuning curves from monkey V1. **C**, J_{cov} for populations with random tuning widths. **D**, J_{cov} for populations with all tuning parameters variable.

fortunately, if the variance of the correlation coefficients is fixed independent of the population size, the covariance matrix will be valid (positive definite) only up to $n \approx 1/(2\sigma)^2$. Thus, as the population size is increased, from a critical n on, the model will not be valid any more. One solution to this problem is to scale the variance in the correlation coefficients by $1/n$. We simulated this scenario numerically and found that, although it increases the Fisher information by a small constant factor, it does not affect the results qualitatively (data not shown).

Effect of noise entropy on encoding accuracy

So far, we considered the level of correlations a free parameter. We now investigate the implications of this assumption. In general, increasing correlations between neurons while keeping the variances of the individual neurons fixed reduces the noise entropy of the population. Because a lower noise entropy reduces the variance in most directions, it is likely to improve the encoding. To understand this argument intuitively, consider the two-neuron toy example shown in Figure 9. The top row shows the firing rate (marginal) distribution of each neuron, while the bot-

tom row depicts the two-dimensional joint distribution. The entropy of a normal distribution is closely related to the area enclosed by the 1 SD ellipse (it is linearly related to the logarithm of the area). Figure 9A shows an uncorrelated reference distribution with marginal SDs of 1. The distribution in Figure 9B has the same marginal distributions but a correlation coefficient of 0.8. The entropy of this correlated distribution is ~ 0.5 bits smaller than that of the uncorrelated distribution with identical marginals. To generate a distribution that has the same entropy as our reference distribution but a correlation of 0.8, we have to increase the marginal SDs to 1.29 (Fig. 9C).

Because the accuracy of a population code depends on how the signal is encoded relative to the noise, a lower noise entropy by itself does not necessarily imply an improvement in coding accuracy. For instance, in the toy example in Figure 9B, the variance along the main diagonal is larger than in the independent case in Figure 9A. Therefore, estimating the mean activity of the two neurons is less accurate than, for instance, their difference. In a heterogeneous population of neurons, the signal encoding is distributed across all directions (Fig. 1M). Because the Fisher information is the sum over the signal-to-noise ratios in each frequency component, a reduction in noise entropy should lead to an improvement in coding accuracy in this case.

To understand how the level of correlations affects the noise entropy in the multineuron (high-dimensional) case, we calculated the noise entropy as a function of the average correlation strength. As c_0 approaches 1, the noise entropy diverges to $-\infty$ (Fig. 10A). In other words, there is a subspace in which the system is effectively noise-free. Any signal in this subspace can be decoded with infinite precision. This explains why the relative Fisher information diverges as c_0 approaches 1 (Fig. 4B).

In real neural populations, however, there is a certain amount of independent noise in the system, due to input noise (e.g., photoreceptors), channel noise (e.g., unreliable synaptic transmission), or other sources. This noise cannot be removed by any type of encoding. This implies that most models used in previous studies (including our above model) are not well constrained if the correlation strength is considered a free parameter because they allow for a degenerate case in which the noise entropy becomes arbitrarily small (by simply increasing the level of correlations until $c_0 = 1$).

To disentangle the effect of the noise correlation structure from that of the noise entropy, we here introduce an additional constraint and fix the noise entropy as the correlations are changed. A convenient way to do so is to relax the Poisson assumption on the variances of the neurons, such that in the modified model, the neurons may have Fano factors $F \neq 1$. For simplicity, we keep the Fano factor constant across the population and define the variances of the neurons as follows:

$$\sigma_j^2(\theta) = F f_j(\theta). \quad (45)$$

Figure 10B shows the Fano factors necessary to maintain constant noise entropy when increasing the correlation strength. Changing the Fano factor affects only J_{mean} but leaves J_{cov} unchanged. Figure 10, C and D, show J_{mean} for heterogeneous populations as a function of the population size for different levels of correlation, analogous to Figure 3, B and D, but with identical noise entropy among populations of equal size.

Similar to the results presented above, under the constant entropy constraint there are two regimes for the effect of correlations. The low-correlation regime applies when populations are small or correlations are low. In this regime, the noise entropy is

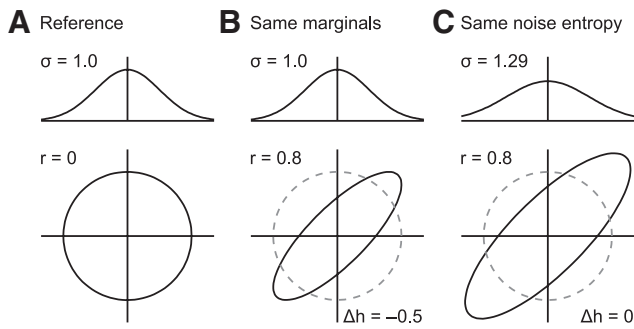


Figure 9. Changing correlations affects noise entropy. The top row shows marginal distributions; the bottom row shows isoprobitability contours (2σ from the mean) of two-dimensional joint distributions. **A**, An uncorrelated Gaussian distribution with $\sigma = 1$. **B**, Gaussian distribution with correlation coefficient $r = 0.8$ and marginals equal to the distribution in **A** ($\sigma = 1$). Its entropy (which is linearly related to the area enclosed by the ellipse) is 0.5 bits smaller than that of the distribution in **A**. **C**, Gaussian distribution with the same correlation coefficient and the same entropy as that in **A**. Its marginal SD is $\approx 30\%$ larger ($\sigma = 1.29$) as that in **A**. Note that the ellipse has the same area as the circle in **A**.

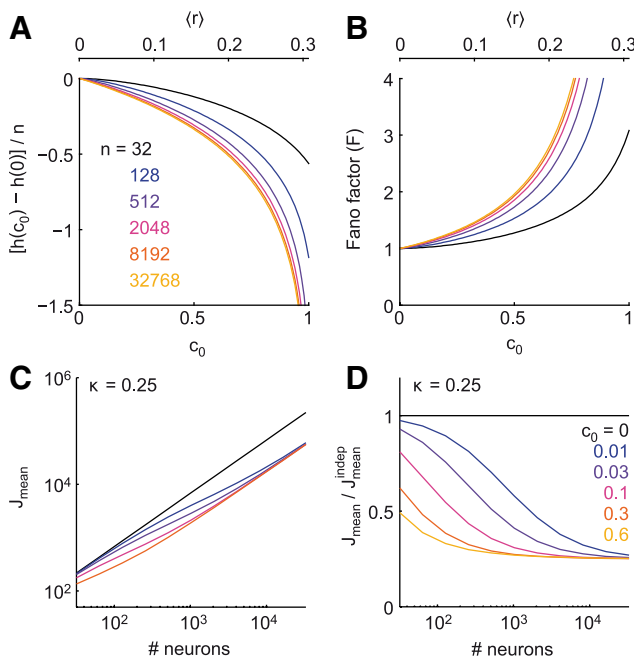


Figure 10. Effect of limited range correlations under constant noise entropy. **A**, Differential entropy (relative to an independent population and normalized by the number of neurons) as a function of correlation strength, c_0 . **B**, Fano factor necessary to maintain constant noise entropy when increasing correlation strength. **C**, Linear Fisher information, J_{mean} , as a function of population size in heterogeneous populations with identical noise entropy (identical across different levels of correlation for the same n). Analogous to Figure 3B. **D**, J_{mean} relative to independent population (as in Fig. 3D).

very similar to the independent case (Fig. 10A) and, thus, reducing correlations can lead to a substantial improvement (Figs. 10D, 11B). As before, the critical level of correlations that separates the low-correlation regime from the high-correlation regime converges to zero for increasing population sizes. In the high-correlation regime, the exact level of correlations is not important if the noise entropy is constant (Fig. 11B). This supports the idea that the improvement with increased correlations we observed above is due to a reduction of the noise entropy. In the large n limit the expected value for J_{mean} of a correlated population is κ times that of an independent population (see Materials and Methods, Eq. 33):

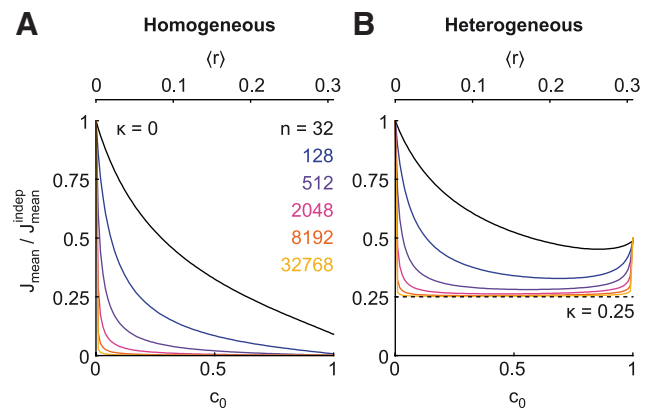


Figure 11. Relative J_{mean} for different correlation strengths under the constraint of constant noise entropy. **A**, Homogeneous population, $\kappa = 0$. **B**, Heterogeneous population, $\kappa = 0.25$. Asymptotically ($n \rightarrow \infty$), the relative J_{mean} converges to κ .

$$\frac{\langle J_{\text{mean}} | c_0 > 0 \rangle}{\langle J_{\text{mean}} | c_0 = 0 \rangle} \rightarrow \kappa. \quad (46)$$

Thus, J_{mean} depends only on the amount of amplitude variability, κ . It is independent of both the structure and the magnitude of correlations as none of the parameters of the correlation structure (e.g., c_0 or L) enter the right-hand side of Equation 46.

Discussion

Many theoretical studies have investigated the effect of noise correlations on encoding accuracy (Snippe and Koenderink, 1992; Zohary et al., 1994; Abbott and Dayan, 1999; Sompolinsky et al., 2001; Wilke and Eurich, 2002; Shamir and Sompolinsky, 2004, 2006; Josić et al., 2009). One of the main conclusions has been that limited range correlations are detrimental for a population code compared with independence or other correlation structures (e.g., uniform) (Abbott and Dayan, 1999; Sompolinsky et al., 2001; Wilke and Eurich, 2002). Because of this, it is often assumed that decreasing correlations improves any population code—as it becomes more similar to the independent case (Zohary et al., 1994; Abbott and Dayan, 1999; Sompolinsky et al., 2001).

In this paper, we showed that these results hold only under the assumption of homogeneous populations of neurons or very small numbers of neurons. In the range of biologically plausible parameters (heterogeneous tuning curves, nonadditive noise, thousands of neurons), increasing correlations can substantially increase the Fisher information (Fig. 4B). This increase in accuracy is mainly due to an overall reduction of the noise entropy that is associated with stronger correlations. When correlations are increased, the noise power increases in a few low-frequency Fourier components while it decreases in all higher-frequency components (it quickly asymptotes to $1 - c_0$ with increasing frequency; Fig. 1D). Because in homogeneous populations the stimulus encoding is confined to the low-frequency Fourier components, higher correlations have detrimental effects. In heterogeneous populations, in contrast, the stimulus encoding is distributed among all frequencies and the high-frequency components have a better signal-to-noise ratio if correlations are strong. These two mechanisms compete, leading to the non-monotonic correlation dependence of the Fisher information of heterogeneous populations. As the population size is increased, the high-frequency components dominate because their number grows linearly with the population size. As a consequence, in the

large n case, increasing correlations is almost always beneficial while a decrease is beneficial only for small population sizes and small enough correlations.

This result raises two important questions. First, if higher correlations improve the accuracy, why does the brain not implement a population code with strong correlations? Second, why do experimental studies find reduced noise correlations under experimental conditions, in which an improved population code is expected (Gutnisky and Dragoi, 2008; Cohen and Maunsell, 2009; Mitchell et al., 2009; Gu et al., 2011)?

With regard to the first question, we suggest that the level of correlations should not be interpreted as a free parameter that can be optimized independently while all other parameters, such as tuning functions and variances, are kept fixed. Under such assumptions, increasing the level of correlations to the maximum (i.e., $c_0 = 1$) leads to maximum Fisher information (Figs. 4B, 6). The noise power in the high frequencies vanishes (as it asymptotes to $1 - c_0$; Fig. 1D), allowing for effectively noise-free decoding of any signal that is encoded in these directions. However, in real neural populations, the amount of noise in the response cannot be arbitrarily small because input noise cannot be removed by processing. Thus, finding the optimal level of correlations under the assumption of fixed tuning functions and variances leads to a degenerate solution, which is not biologically plausible.

To avoid the noise-free case and better constrain the problem, we also analyzed population codes in which the total amount of noise (the noise entropy) is kept constant and only the correlation structure is changed. Our results demonstrate that, under this constraint, the performance of population codes in large populations is primarily determined by the amount of heterogeneity in the tuning functions of the neurons and by the overall noise entropy (Fig. 10D, Eq. 46), while the specific structure of the noise does not appear to be as important as commonly assumed. Whether the noise entropy is of the same importance for predicting the accuracy of a population code also for other scenarios not considered explicitly in this study—such as stimulus-dependent (Josić et al., 2009) and heterogeneous correlations, or simultaneous encoding of multiple stimulus dimensions (Zhang and Sejnowski, 1999)—remains to be investigated.

The second question raised above was how to reconcile our theoretical results with empirical findings of reduced correlations (Gutnisky and Dragoi, 2008; Cohen and Maunsell, 2009; Mitchell et al., 2009; Gu et al., 2011). One possibility is that the size of the population that is read out is small enough to be in the low-correlation regime. Although we did not assess this in detail by taking into account all relevant parameters from these studies, Figures 5 and 8B suggest that the relevant population size would have to be on the order of at most a few hundred cells. Given the number of cells even in a single cortical column, this seems rather unlikely.

A second explanation is related to the neural readout mechanism. Because we currently do not know how information is read out by downstream neurons (or populations thereof), we quantified the maximum amount of information that can be extracted from the population response. Of course, the effect of correlations can be different for other, computationally constrained readouts. If, for instance, a linear readout is assumed, the conclusions would be different. Even though J_{mean} is often referred to as the portion of the information that can be read out by linear methods, this notion is problematic—at least in the framework of stimulus reconstruction considered here. To illustrate this point, we estimated the performance of the OLE decoding the activity of a heterogeneous population of neurons. The efficiency of the

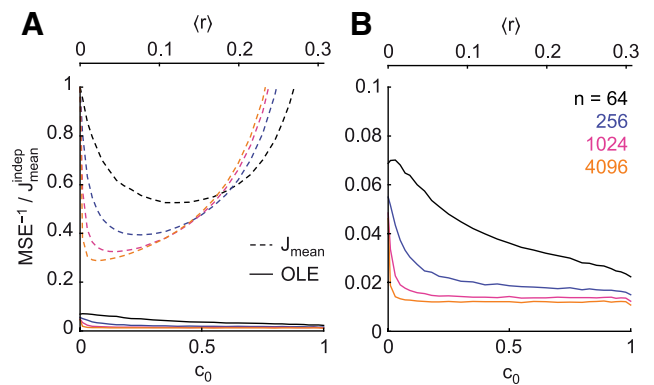


Figure 12. Dependence of OLE on correlations. Heterogeneous populations ($\kappa = 0.25$) are considered. This is similar to Figure 4. Dashed lines, J_{mean} ; solid lines, OLE. The colors represent different population sizes (see legend in B). A, J_{mean} and the efficiency (inverse of the mean squared error) of the OLE relative to J_{mean} of an independent population. B, Zoomed-in version of A.

OLE does not converge to the inverse of J_{mean} (Fig. 12A). In addition, in contrast to the Fisher information and the maximum-likelihood estimator, the accuracy of the OLE does not increase with increased correlations, not even for large population sizes (Fig. 12B). The reason for this behavior is the fact that the OLE is a biased estimator, for which the Cramér–Rao bound is not simply the inverse of the Fisher information but the bias has to be taken into account. Because the estimator bias depends on the correlations, the dependence of the mean squared error on correlations is not captured well by the Fisher information. Consequently, if downstream areas are confined to linear readout mechanisms, reducing pairwise correlations increases the readout accuracy.

Assuming optimal readout, a third possibility is that the improved performance is not exclusively due to reduced correlations. For instance, under attention Fano factors decrease and firing rates increase (Cohen and Maunsell, 2009; Mitchell et al., 2009). Although it has been argued that these changes are small compared with the relative changes in correlation strength and their effect is negligible for large populations, this argument is problematic. The effect of different factors was assessed using highly suboptimal pooling rules (Cohen and Maunsell, 2009; Mitchell et al., 2009) and the conclusions derived from these pooling models do not generalize to other (e.g., optimal) readout mechanisms—as our above results show. Our analysis suggests an alternative interpretation: the higher firing rates under attention increase the signal while the reduced Fano factors and correlations indicate the suppression of a common noise source, which reduces the noise entropy and therefore leads to improved coding accuracy. For example, if the response of a population of neurons with unit variances and weak correlations ($\langle r \rangle = 0.05$) is confounded by a common noise source with variance 0.05, removing this common noise source reduces noise correlations from ~ 0.1 to ~ 0.05 (a 50% change) while reducing the variances from ~ 1.05 to 1 (a 5% change). In this situation, however, the changes in correlations cannot be separated from the changes in variances and neither of both is more important than the other. In fact, considering each one in isolation is not meaningful.

In summary, the notion that reducing correlations leads to a more accurate encoding is not a general principle but is only true under certain conditions. Assumptions about the size of the population and the way information is read out can strongly affect the conclusions. For optimal decoding of large populations, the total

amount of noise—as measured by the noise entropy—is more important than the specific noise correlation structure.

Notes

Supplemental material for this article is available at <http://bethgelab.org/code/ecker2011>. It contains MATLAB code to reproduce all figures and numerical simulations in this paper. This material has not been peer reviewed.

References

- Abbott LF, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. *Neural Comput* 11:91–101.
- Bair W, Zohary E, Newsome WT (2001) Correlated firing in macaque visual area MT: time scales and relationship to behavior. *J Neurosci* 21:1676–1697.
- Berens P, Ecker AS, Gerwinn S, Tolias AS, Bethge M (2011) Reassessing optimal neural population codes with neurometric functions. *Proc Natl Acad Sci U S A* 108:4423–4428.
- Bethge M, Rotermund D, Pawelzik K (2002) Optimal short-term population coding: when Fisher Information fails. *Neural Comput* 14:2317–2351.
- Cohen MR, Maunsell JH (2009) Attention improves performance primarily by reducing interneuronal correlations. *Nat Neurosci* 12:1594–1600.
- Cohen MR, Newsome WT (2008) Context-dependent changes in functional circuitry in visual area MT. *Neuron* 60:162–173.
- Cover TM, Thomas JA (1991) *Elements of information theory*. New York: Wiley-Interscience.
- Ecker AS, Berens P, Keliris GA, Bethge M, Logothetis NK, Tolias AS (2010) Decorrelated neuronal firing in cortical microcircuits. *Science* 327:584–587.
- Gu Y, Liu S, Fetsch CR, Yang Y, Fok S, Sunkara A, DeAngelis GC, Angelaki DE (2011) Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron* 71:750–761.
- Gutnisky DA, Dragoi V (2008) Adaptive coding of visual information in neural populations. *Nature* 452:220–224.
- Josić K, Shea-Brown E, Doiron B, de la Rocha J (2009) Stimulus-dependent correlations and population codes. *Neural Comput* 21:2774–2804.
- Kay SM (1993) *Fundamentals of statistical signal processing, Vol I: Estimation theory*. Upper Saddle River, NJ: Prentice Hall.
- Mitchell JF, Sundberg KA, Reynolds JH (2009) Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* 63:879–888.
- Salinas E, Abbott LF (1994) Vector reconstruction from firing rates. *J Comput Neurosci* 1:89–107.
- Shamir M, Sompolinsky H (2001) Correlation codes in neuronal networks. In: *Advances in neural information processing systems, Vol 14* (Dietterich TG, Becker S, Ghahramani Z, eds), pp 277–284. Cambridge, MA: MIT.
- Shamir M, Sompolinsky H (2004) Nonlinear population codes. *Neural Comput* 16:1105–1136.
- Shamir M, Sompolinsky H (2006) Implications of neuronal diversity on population coding. *Neural Comput* 18:1951–1986.
- Smith MA, Kohn A (2008) Spatial and temporal scales of neuronal correlation in primary visual cortex. *J Neurosci* 28:12591–12603.
- Snippe HP, Koenderink JJ (1992) Information in channel-coded systems: correlated receivers. *Biol Cybern* 67:183–190.
- Sompolinsky H, Yoon H, Kang K, Shamir M (2001) Population coding in neuronal systems with correlated noise. *Phys Rev E Stat Nonlin Soft Matter Phys* 64:051904.
- Wilke SD, Eurich CW (2002) Representational accuracy of stochastic neural populations. *Neural Comput* 14:155–189.
- Zhang K, Sejnowski TJ (1999) Neuronal tuning: to sharpen or broaden? *Neural Comput* 11:75–84.
- Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370:140–143.