Supporting Information

Berens et al. 10.1073/pnas.1015904108

SI Methods 1: Details on the Population Model

Covariance Matrix. To parameterize the covariance matrix $\sum(\theta)$, we follow Josić et al. (1) and set

$$\Sigma_{ij}(\theta) = \delta_{ij}\upsilon_i(\theta) + (1-\delta_{ij})\rho_{ij}(\theta)\sqrt{\upsilon_i(\theta)\upsilon_j(\theta)}.$$

Here, $\upsilon_i(\theta) = \bar{r}_i(\theta) = Tf_i(\theta)$ and is the variance of cell i; i.e., we assume a Fano factor of 1. $\rho_{ij}(\theta)$ is the correlation coefficient between cells i and j. We allow for both stimulus and spatial influences on ρ, by setting

$$\rho_{ij}(\theta) = s_i(\theta)\,s_i(\theta)\,c\left(\phi_i - \phi_j\right).$$

The function $s_i(\theta)$ models the influence of the stimulus-dependent component on the correlation structure, and the function c models the spatial component and is independent of θ. We use $s_i(\theta) = \kappa_1 + \kappa_2 a^2(\theta - \phi_i)$ with $a(\theta) = \frac{1}{2}(1 + \cos(\theta))$ and $c(\Delta\phi) = C \exp(-|\Delta\phi|/\alpha)$, where α controls the length of the spatial decay and C the average correlation. The four possible correlation shapes arising from this parameterization are illustrated in Fig. 4A. To obtain a desired mean level of correlations $\bar{\rho}$ in a population, we use the method described in appendix E of Josić et al. (1).

Signal-to-Noise Ratio. In our population model with Poisson-like noise the signal-to-noise ratio per neuron is proportional to the time available for decoding, because

$$\frac{S}{N} = \frac{\mathrm{Var}_\theta[\bar{r}_i(\theta)]}{E_\theta[\mathrm{Var}[r_i|\bar{r}_i]]} = \frac{T^2\mathrm{Var}_\theta[f_i(\theta)]}{E_\theta[\bar{r}_i(\theta)]} = \frac{T^2\mathrm{Var}_\theta[f_i(\theta)]}{TE_\theta[f_i(\theta)]} \sim T.$$

SI Methods 2: Numerical Computation of the MDE/IMDE

We approximate the integral of Eq. **3** numerically via Monte Carlo techniques (2, 3) by

$$\mathrm{MDE}(\theta, \theta + \Delta\theta) = \frac{1}{2}\int \min(p(r|\theta), p(r|\theta + \Delta\theta))\, dr$$
$$\approx \frac{1}{2M}\sum_{i=1}^{M} \min\left(p\left(r^{(i)}|\theta\right), p\left(r^{(i)}|\theta + \Delta\theta\right)\right)\Big/ p\left(r^{(i)}\right),$$

where $r^{(i)}$ are M samples, drawn from the mixture distribution $p(r) = \frac{1}{2}(p(r|\theta) + p(r|\theta + \Delta\theta))$. The factor $1/p(r^{(i)})$ corrects for the fact that by sampling from $p(r)$ we weigh each sample pattern with its probability. We used $M \geq 10^5$ and evaluated $\mathrm{MDE}_\theta(\Delta\theta)$ for 500 equally spaced points between 0° and 180°.

The IMDE and average neurometric functions $\mathrm{MDE}_\theta(\Delta\theta)$ were obtained by evaluating them at 20 different θ uniformly spaced between ϕ_1 and $\phi_1 + \Delta\phi/2$, where $\Delta\phi$ is the difference between two preferred orientations. This is sufficient because all codes considered here are shift symmetric with period $\Delta\phi = 2\pi/N$ and because tuning curves are symmetric about the preferred orientation, only half a period needs to be considered. We verified that 20 different reference directions were sufficient by repeating our simulations for >40 reference directions.

SI Methods 3: Numerical Estimation of the MMSE

The minimum mean squared error is achieved by the estimator that minimizes Eq. **10**. On the basis of a response r generated from the stimulus-conditional distribution for stimulus θ, it is given by

$$\widehat{\theta}(r) = \arg\min\int_0^{2\pi}\left(\widehat{\theta} \dot{-} \psi\right)^2 p(\psi|r)\, d\psi,$$

where

$$p(\psi|r) = \frac{p(r|\psi)p(\psi)}{p(r)}$$

is the posterior over stimuli given the response and $\alpha \dot{-} \beta = \min(|\alpha - \beta|, |2\pi - \alpha + \beta|)$ is the distance measured along the circle (4). The prior is uniform such that $p(\psi) = 1/2\pi$. We evaluate the above equations for L discrete, regularly spaced $\psi_i \in [0, 2\pi)$ and replace the integrals by sums. We obtain

$$p(r) \approx \frac{1}{2\pi}\sum_{i=1}^{L} p(r|\psi_i)\,\Delta\psi$$
$$\int_0^{2\pi}\left(\widehat{\theta} \dot{-} \psi\right)^2 p(\psi|r)\,d\psi \approx \frac{1}{2\pi}\sum_{i=1}^{L}\left(\widehat{\theta} \dot{-} \psi_i\right)^2 p(r|\psi_i)/p(r)\Delta\psi.$$

Simplifying we obtain

$$\widehat{\theta}(r) \approx \arg\min_{\theta_j} \frac{\sum_{i=1}^{L}\left(\theta_j \dot{-} \psi_i\right)^2 p(r|\psi_i)}{\sum_{i=1}^{L} p(r|\psi_i)},$$

which is solved by using again L discrete, uniformly spaced θ_j as candidates. This discretization limits the accuracy with which the MMSE can be estimated. The limited accuracy is a problem in particular for very good estimators, for which L must be very large. Here, we chose $L = 500$ and verified that the MMSE curves at the highest SNR did not change when L was substantially increased. Using this equation we can compute the MMSE as

$$\mathrm{MMSE} = E_{\theta, r}\left[\left(\theta \dot{-} \widehat{\theta}(r)\right)^2\right].$$

Similar procedures have been used in refs. 5 and 6. In some scenarios, approximation procedures like those presented in ref. 6 can be helpful.

SI Text

In this section, we formally show that for Fisher-optimal codes (i) a nonzero pedestal error exists in the large N limit and (ii) the saturation point $\Delta\theta$s goes to zero as the population size N increases for fixed T. Finally, we derive an approximation to the pedestal error to show that it depends on the available decoding time alone.

Preliminary Remarks. We first note that in Fisher-optimal codes the tuning width is inversely proportional to N (Fig. 3A), such that

$$w = \frac{c}{N}$$

for some constant c. Only a few cells are active for any given stimulus and this number does not depend on the population size

N (Fig. 3*G*). The tuning curve spacing can be expressed in terms of the population size as

$$\Delta\phi = \frac{2\pi}{N}.$$

Therefore, we can write w in terms of $\Delta\phi$ as

$$w = \frac{c\Delta\phi}{2\pi},$$

which holds for any N. Also, $\phi_i = i\Delta\phi$. We further note that the following relationship holds:

$$e_i(\theta) \equiv \left(\frac{1}{2} + \frac{1}{2}\cos(\theta - \phi_i)\right)^k \le \exp\left(-\frac{k}{4}(\theta - \phi_i)^2\right) \equiv \widehat{e}_i(\theta).$$

If the exponent k is sufficiently large, $e_i(\theta) \approx \widehat{e}_i(\theta)$. Thus, the tuning function in our model can be replaced by

$$f_i(\theta) = \lambda_1 + \lambda_2 e_i(\theta) \approx \lambda_1 + \lambda_2 \widehat{e}_i(\theta),$$

which is of Gaussian form. Therefore, we can rewrite the tuning functions as follows:

$$f_i(\theta) \approx g\left(\frac{\theta - \phi_i}{w}\right) = h\left(\frac{\theta - \phi_i}{\Delta\phi}\right) = h\left(\frac{\theta}{\Delta\phi} - i\right).$$

In this equation, i is the neuron index and the constants in w are absorbed into the function h. Note that the tuning functions g and h are fixed templates for which only the domain changes with N (Fig. 3 *G* and *H*). Whereas f_i is defined on $[-\pi, \pi]$, h is defined on $[-((N-1)/2) + 1, (N-1)/2]$, for even N. It follows that the Fisher-optimal tuning functions drawn in units of $\Delta\phi$ (instead of θ) are constant for different N (Fig. 3 *G* and *H*); the activity of a neuron depends only on $\phi_i/\Delta\phi$, that is, how many units of $\Delta\phi$ its preferred orientation is away from the stimulus, independent of N.

Existence of the Pedestal Error. We first show that there is a lower bound on the minimum discrimination error between any pair of stimuli, which is nonzero in the large N limit. To this end, we define an auxiliary population of neurons with additive Gaussian noise with variance λ_1, the parameter that determines the baseline firing rate of our tuning curves. The firing patterns of this population are distributed as

$$q(r|\theta) = N(\bar{r}, T\lambda_1 I_N),$$

where I_N is the identity matrix of dimension N. The minimum discrimination error of this population provides a lower bound on that of the populations with Poisson-like noise used in the main text, i.e.,

$$\mathrm{MDE}_p(\theta, \theta + \Delta\theta) \ge \mathrm{MDE}_q(\theta, \theta + \Delta\theta).$$

Here, the subscripts p and q indicate that the MDE is calculated with respect to the pattern distributions p and q, respectively. We can express the right-hand side of this equation as

$$\mathrm{MDE}_q(\theta, \theta + \Delta\theta) = 1 - \Psi(d'/2),$$

where $d' = \sqrt{\Delta\mu^T \Sigma^{-1} \Delta\mu}$. Equality holds because in the case of additive noise the linear discrimination error is equal to the MDE (*SI Discussion*). We now provide an upper bound for d':

$$d'^2 = \sum_i \frac{T^2(f_i(\theta) - f_i(\theta + \Delta\theta))^2}{T\lambda_1}$$

$$= \frac{T\lambda_2^2}{\lambda_1}\sum_i (e_i(\theta) - e_i(\theta + \Delta\theta))^2$$

$$\le \frac{T\lambda_2^2}{\lambda_1}\sum_i e_i(\theta)^2 + \sum_i e_i(\theta + \Delta\theta)^2$$

$$= \frac{2T\lambda_2^2}{\lambda_1}\sum_i e_i(\theta)^2.$$

Here we use e_i as defined above and the neuron index i ranges from $-((N-1)/2) + 1$ to $(N-1)/2$. We can now use the upper bound on $e(\theta)$ and use a Gaussian tuning function $\widehat{e}(\theta) = \exp(-(\theta - \phi_i)^2/w^2)$ instead. Now without loss of generality we assume $\theta = \phi_0$ and substitute w, ϕ_i, and $\Delta\phi$ from above. We obtain

$$\widehat{e}(\phi_0) = \exp\left(-\left(\frac{2\pi}{c}\cdot i\right)^2\right),$$

where the i indicates the neuron index, not the imaginary unit. Inserting into the above equation yields

$$d'^2 \le \frac{2T\lambda_2^2}{\lambda_1}\sum_i \widehat{e}_i(\theta)^2 = \frac{2T\lambda_2^2}{\lambda_1}\sum_i \exp\left(-2\left(\frac{2\pi}{c}\cdot i\right)^2\right)$$

$$= \frac{2T\lambda_2^2}{\lambda_1}\sum_i \exp\left(-\frac{1}{2}\frac{i^2}{\left(\frac{c}{4\pi}\right)^2}\right) \le \frac{T\lambda_2^2}{\lambda_1}\left(\frac{c}{\sqrt{8\pi}} + 1\right).$$

To arrive at the last inequality note that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{i^2}{2\sigma^2}\right)di = 1$$

is the area under the density function of a Gaussian with SD σ. We can approximate the integral by the lower Riemann sum, i.e., by rectangles $[i - 1, i]$ with height $\exp(-(i^2/2\sigma^2))$ for positive i and $[i, i + 1]$ with height $\exp(-(i^2/2\sigma^2))$ for negative i, respectively. Thus, we have

$$\sum_{i \ne 0} \exp\left(-\frac{i^2}{2\sigma^2}\right) \le \sqrt{2\pi}\sigma.$$

Substituting $\sigma = c/4\pi$ and including $i = 0$, we obtain the above inequality.

Thus, d' is bounded from above independent of N. Therefore,

$$\mathrm{MDE}_p(\theta, \theta + \Delta\theta) \ge \mathrm{MDE}_q(\theta, \theta + \Delta\theta)$$

$$= 1 - \Psi(d'/2)$$

$$= 1 - \Psi\left[\sqrt{\frac{T\lambda_2^2}{\lambda_1}\left(\frac{c}{\sqrt{8\pi}} + 1\right)}/2\right] > 0 \quad \textbf{[S1]}$$

independent of N and in particular also in the limit $N \to \infty$. This result shows that there is a nonvanishing pedestal error P for all N and for finite T.

Convergence of Saturation Point $\Delta\theta_S$ to Zero. Next we show that the saturation point $\Delta\theta_S$ converges to zero for $N \to \infty$. We define $\Delta\theta^*$ as

$$\Delta\theta^* = \min\{\Delta\theta : \mathrm{MDE}(\Delta\theta) - P \le \varepsilon/2\}.$$

We approximate the MDE of the whole population with N neurons by considering only two subsets of $2M_\varepsilon + 1$ neurons each that are most strongly activated by one of the two stimuli:

$$I_\varepsilon = \left\{ i : \left| \frac{\theta}{\Delta\phi} - i \right| \le M_\varepsilon \text{ or } \left| \frac{\theta + \Delta\theta}{\Delta\phi} - i \right| \le M_\varepsilon \right\}.$$

Here M_ε is chosen such that

$$\mathrm{MDE}_{I_\varepsilon}(\Delta\theta^*) \le \mathrm{MDE}(\Delta\theta^*) + \varepsilon/2.$$

Here, $\mathrm{MDE}_{I_\varepsilon}$ is the MDE achieved by the subpopulation with neurons in the set I_ε, for which holds

$$P \le \mathrm{MDE}(\Delta\theta) \le \mathrm{MDE}_{I_\varepsilon}(\Delta\theta).$$

Because the tuning curves are identical in units of $\Delta\phi$ for different N, M_ε does not change with N and therefore $\Delta\theta^*$ is also a constant in units of $\Delta\phi$:

$$\Delta\theta^* = c\Delta\phi \sim \frac{1}{N}.$$

Finally, we define $\Delta\theta_S$ as

$$\Delta\theta_S = \min\{\Delta\theta : \mathrm{MDE}(\Delta\theta) - P \le \varepsilon\}$$

and note that $\Delta\theta_S \le \Delta\theta^*$ and therefore $\Delta\theta_S \to 0$ as $N \to \infty$. Consequently, the area of the initial region will shrink to zero, too, as

$$A_{\mathrm{IR}} = \int_0^{\Delta\theta_S} [\mathrm{MDE}(\Delta\theta) - P] d\Delta\theta.$$

In particular, the neurometric functions for different N at fixed T are identical, when written as a function of $\Delta\phi$ (Fig. S4). Although they show a different pedestal error for different T, they reach their pedestal error at constant $\Delta\phi$ for all N and T considered ($\sim 2\Delta\phi$).

Approximation of the Pedestal Error P. Finally, we derive an analytically tractable approximation of the pedestal error. Looking only at $2 \times 2M_\varepsilon + 1$ neurons in a Fisher-optimal model population, we can approximate the pedestal error with arbitrary precision. We find for our model that $M_\varepsilon \sim 1$ such that only six cells suffice to achieve the same error as the entire population. We adopt the following notation: $\bar{r}_0$ is the activity by the maximally excited neuron and $\bar{r}_1$ and $\bar{r}_{-1}$ are the activities of the two neurons to the left and to the right. For the time being, we omit the dependence on θ and assume we place the stimulus at the peak of neuron 0. This assumption leads to the two average response vectors for the two stimuli θ and $\theta + \pi$,

$$\mu_1 = (\bar{r}_{-1}, \bar{r}_0, \bar{r}_1, \bar{r}_{\min}, \bar{r}_{\min}, \bar{r}_{\min})^T$$
$$\mu_2 = (\bar{r}_{\min}, \bar{r}_{\min}, \bar{r}_{\min}, \bar{r}_{-1}, \bar{r}_0, \bar{r}_1)^T,$$

and the respective stimulus conditional covariance matrices $\Sigma_1 = \mathrm{diag}(\mu_1)$ and $\Sigma_2 = \mathrm{diag}(\mu_2)$. To derive our approximation of the pedestal error, we calculate $d' = \sqrt{\Delta\mu^T \Sigma^{-1} \Delta\mu}$, taking advantage of the small subpopulation that needs to be considered, where $\Sigma = \frac{1}{2}(\Sigma_1 + \Sigma_2)$. We obtain

$$\Delta\mu = (\bar{r}_{-1} - \bar{r}_{\min}, \dots, \bar{r}_{\min} - \bar{r}_1)^T$$

$$\Sigma^{-1} = \mathrm{diag}\left(\frac{2}{\bar{r}_{-1} + \bar{r}_{\min}}, \dots, \frac{2}{\bar{r}_{\min} + \bar{r}_1}\right).$$

This yields

$$\Delta\mu^T \Sigma^{-1} \Delta\mu = 4 \sum_{i=-1}^{1} \frac{(\bar{r}_i - \bar{r}_{\min})^2}{\bar{r}_i + \bar{r}_{\min}}.$$

The error of the optimal linear classifier (7) in this situation is

$$\widehat{P}_\theta = 1 - \Psi\left[\sqrt{\sum_{i=-1}^{1} \frac{(\bar{r}_i - \bar{r}_{\min})^2}{\bar{r}_i + \bar{r}_{\min}}} \right],$$

where Ψ is the cumulative normal distribution function. This equation provides a good approximation of the pedestal error of the neurometric function of Fisher-optimal population codes (Fig. 3F). We can see the dependence on time by rewriting the above expression,

$$\widehat{P}_\theta = 1 - \Psi\left[F\sqrt{T} \right],$$

where $F = \sqrt{\sum_{i=-1}^{1} ((f_i - f_{\min})^2 / (f_i + f_{\min}))}$ depends only on the tuning curves of the individual neurons. In particular, $f_{\pm 1}$ are constant with growing N (as shown above), because we can rewrite the tuning function as a function of $\Delta\phi$. The above expression depends on the choice of the reference direction θ, so we average again over θ and obtain

$$\widehat{P} = \left\langle 1 - \Psi[F_\theta \sqrt{T}] \right\rangle_\theta, \qquad \textbf{[S2]}$$

where the subscript θ indicates the dependence of F on θ inherited from the tuning functions.

SI Discussion

Most other studies that investigated population codes in the discrimination framework measured the minimal linear discrimination error (LDE), such as refs. 8–11 as well as part 1 of ref. 7. Few others such as ref. 12 and part 2 of ref. 7 also consider nonlinear approximations of the minimal discrimination error. However, none of these studies computed the minimal discrimination error.

Linear Approaches. The studies by Johnson (8), Snippe and Koenderink (9, 10), and Averbeck and Lee (7) used the discriminability index d' from signal detection theory:

$$d' = \sqrt{\Delta\mu^T \Sigma^{-1} \Delta\mu}.$$

Here, $\Delta\mu = \bar{r}(\theta) - \bar{r}(\theta + \Delta\theta)$ is the difference in average firing rate profiles across the population and Σ is the noise covariance matrix. The first two studies (8, 9) evaluated this equation for constant Σ and in the limit $\Delta\theta \to 0$. Because $\Delta\mu \approx \bar{r}'(\theta)\Delta\theta$ is an approximation of the derivative of the population firing rate profile for small $\Delta\theta$,

$$d' = \sqrt{\Delta\mu^T \Sigma^{-1} \Delta\mu}$$
$$\approx \sqrt{\bar{r}'(\theta)^T \Delta\theta \Sigma^{-1} \bar{r}'(\theta) \Delta\theta} = \Delta\theta \sqrt{J_{\mathrm{mean}}}$$

so that the two studies effectively study the linear part of the Gaussian Fisher information. Similar approaches have also been used by refs. 9, 13, and 14).

Averbeck and Lee (7) used d' also for finite $\Delta\theta$ with $\Sigma = \frac{1}{2}(\Sigma_\theta + \Sigma_{\theta+\Delta\theta})$. They then proceeded to compute the minimum *linear* discrimination error

$$\text{LDE} = 1 - \Psi(d'/2),$$

where Ψ is the standard normal cumulative distribution function. It might not be immediately obvious why this computation really yields the minimal linear discrimination error. To see why this is the case, observe that for two normal distributions with means μ_1, μ_2 and covariance matrices Σ_1, Σ_2 and equal prior probabilities, Fisher's linear discriminant is the optimal linear classifier (15). Its weight vector is given by

$$w = \Sigma^{-1}(\mu_1 - \mu_2),$$

where $\Sigma = \Sigma_1 + \Sigma_2$. The discriminability index d' along w with

$$\bar{\mu}_i = w^T \mu_i$$
$$\bar{\sigma}_i^2 = w^T \Sigma_i w$$

is

$$d' = \frac{\bar{\mu}_1 - \bar{\mu}_2}{\sqrt{\bar{\sigma}_1^2 + \bar{\sigma}_2^2}} = \frac{w^T \Delta\mu}{\sqrt{w^T(\Sigma_1 + \Sigma_2)w}} = \sqrt{\Delta\mu^T \Sigma^{-1} \Delta\mu},$$

which is the same as the above expression. For one-dimensional data, the error can be computed from d' with the formula used above (see also ref. 7).

Whereas the LDE is equal to the MDE for additive Gaussian noise models, i.e., when $r = \bar{r}(\theta) + \varepsilon$ with $\varepsilon \sim N(0, \Sigma)$, it does not capture the coding properties of a population code in the general case with stimulus-dependent covariance matrices, e.g., for a Poisson-like Gaussian noise model or for stimulus-dependent correlation structures.

Nonlinear Approaches. As a second measure of coding quality, Averbeck and Lee (7) consider the Bhattacharyya distance (D_B). It is defined as

$$D_B(\Delta\theta) = -\log \int_r \sqrt{p(r|\theta) \cdot p(r|\theta + \Delta\theta) \, dr},$$

which is, in the general case, as difficult to compute as the MDE. For the Gaussian case it simplifies to

$$D_B(\Delta\theta) = \frac{1}{4}\Delta\mu^T(\Sigma_\theta + \Sigma_{\theta+\Delta\theta})^{-1}\Delta\mu$$
$$+ \frac{1}{2}\log\frac{|\Sigma_\theta + \Sigma_{\theta+\Delta\theta}|}{2\sqrt{|\Sigma_\theta + \Sigma_{\theta+\Delta\theta}|}}.$$

Previously, Kang et al. (12) used the Chernoff distance (D_C) as a measure of coding accuracy, which is defined as

$$D_\alpha(\Delta\theta) = -\log \int_r \sqrt{p^\alpha(r|\theta)p^{1-\alpha}(r|\theta+\Delta\theta) \, dr}$$
$$D_c(\Delta\theta) = \max_\alpha D_\alpha(\Delta\theta) \qquad \textbf{[S4]}$$

with $\alpha \in (0, 1)$. Interestingly, D_B is a special case of D_C obtained by setting $\alpha = \frac{1}{2}$. To compute the Chernoff distance, Kang et al. exploit the fact that they assume a Gaussian noise model and a population with independent neurons and show that for this case, the optimal $\alpha = \frac{1}{2}$, so that they effectively use D_B instead of D_C in their study.

In the Gaussian case, a simpler formula can be provided for computing D_α (16):

$$D_\alpha(\Delta\theta) = \frac{\alpha(1-\alpha)}{2}\Delta\mu^T(\alpha\Sigma_\theta + (1-\alpha)\Sigma_{\theta+\Delta\theta})^{-1}\Delta\mu$$
$$+ \frac{1}{2}\log\frac{|\alpha\Sigma_\theta + (1-\alpha)\Sigma_{\theta+\Delta\theta}|}{|\Sigma_\theta|^\alpha|\Sigma_{\theta+\Delta\theta}|^{1-\alpha}}.$$

The interest in D_B and D_C originates in the fact that both provide an upper bound on the MDE, the Chernoff bound (17, 16):

$$\text{MDE}(\Delta\theta) \leq \exp(-D_C(\Delta\theta)). \qquad \textbf{[S5]}$$

The identical bound for D_B is in general less tight than Eq. S5, as $D_C \geq D_B$ with equality if and only if the optimal $\alpha = \frac{1}{2}$ in Eq. S4. If both class-conditional distributions are Gaussians with $\Sigma_\theta = \Sigma_{\theta+\Delta\theta}$, the true optimum can be shown to lie at $\alpha = \frac{1}{2}$ (16). For arbitrary population codes and noise distributions, the question of whether the Chernoff bound is tight is not straightforward to answer. Kang et al. state that its tightness depends on the population size, the integration time, and the shape of the tuning curves (12). In summary, D_B and D_C provide useful upper bounds on the MDE but cannot be used to measure the MDE directly.

1. Josić K, Shea-Brown E, Doiron B, de la Rocha J (2009) Stimulus-dependent correlations and population codes. *Neural Comput* 21:2774–2804.
2. Berens P, Gerwinn S, Ecker AS, Bethge M (2009) Neurometric function analysis of population codes. *Advances in Neural Information Processing Systems 22: Proceedings of the 2009 Conference* (MIT Press, Cambridge, MA), pp 90–98.
3. Hershey J, Olsen P (2007) Approximating the Kullback Leibler divergence between Gaussian mixture models. *International Conference on Acoustics, Speech and Signal Processing, 2007.* (Institute of Electrical and Electronics Engineers, New York), pp IV-317–IV-320.
4. Berens P (2009) CircStat: A MATLAB toolbox for circular statistics, *J Stat Softw* 31: 10, Available at http://www.jstatsoft.org/v31/i10.
5. Bethge M, Rotermund D, Pawelzik K (2002) Optimal short-term population coding: When Fisher information fails. *Neural Comput* 14:2317–2351.
6. Yaeli S, Meir R (2010) Error-based analysis of optimal tuning functions explains phenomena observed in sensory neurons. *Front Comput Neurosci* 4:130.
7. Averbeck BB, Lee D (2006) Effects of noise correlations on information encoding and decoding. *J Neurophysiol* 95:3633–3644.
8. Johnson KO (1980) Sensory discrimination: Decision process. *J Neurophysiol* 43: 1771–1792.
9. Snippe HP, Koenderink JJ (1992) Information in channel-coded systems: Correlated receivers. *Biol Cybern* 67:183–190.
10. Snippe HP, Koenderink JJ (1992) Discrimination thresholds for channel-coded systems. *Biol Cybern* 66:543–551.
11. Pouget A, Thorpe SJ (1991) Connectionist models of orientation identification. *Connect Sci* 3:127–142.
12. Kang K, Shapley RM, Sompolinsky H (2004) Information tuning of populations of neurons in primary visual cortex. *J Neurosci* 24:3726–3735.
13. Seriès P, Stocker AA, Simoncelli EP (2009) Is the homunculus "aware" of sensory adaptation? *Neural Comput* 21:3271–3304.
14. Mato G, Sompolinsky H (1996) Neural network models of perceptual learning of angle discrimination. *Neural Comput* 8:270–299.
15. Duda RO, Hart PE, Stork DG (2000) *Pattern Classification* (Wiley, New York), 2nd Ed.
16. Fukunaga K (1990) *Introduction to Statistical Pattern Recognition* (Academic, New York).
17. Cover TM, Thomas JA (2006) *Elements of Information Theory* (Wiley-Interscience, New York).

Fig. S1. Tuning curves with width optimized for various criteria (black). For better visualization of the population structure, two additional tuning curves are shown in light gray. (*A* and *B*) MASE-optimal tuning curve for 10 and 1,000 ms, respectively. (*C* and *D*) MMSE-optimal tuning curve for 10 and 1,000 ms, respectively. (*E* and *F*) IMDE-optimal tuning curve for 10 and 1,000 ms, respectively.

Fig. S2. Fig. 2 shows that the optimal tuning width with regard to the MASE is almost independent of time, but varies slightly. The reason for this result is that Fisher Information has two parts, J_{mean} and J_{cov}, which have different time dependencies. For an independent population, we have

$$J_{mean} = \bar{r}' \Sigma^{-1} \bar{r}' = \sum_{i=1}^{N} \frac{\bar{r}'^2}{\bar{r}_i} \sim T$$

$$J_{cov} = \frac{1}{2} \text{Tr}\left[\Sigma' \Sigma^{-1} \Sigma' \Sigma^{-1}\right] = \sum_{i=1}^{N} \frac{\bar{r}'^2}{\bar{r}_i^2} \nsim T.$$

Thus J_{mean} is proportional to time and its optimum is fixed for varying T. J_{cov} is constant and does not depend on time. Therefore, the relative importance of the two terms changes with time: Whereas for small T J_{mean} and J_{cov} are roughly on the same order of magnitude, J_{mean} dominates for large T. When we plot the two extreme cases, $\langle 1/J_{mean}\rangle$ and $\langle 1/J_{cov}\rangle$, corresponding to $T \to \infty$ and $T \to 0$, respectively, we find that they lead to slightly different optimal tuning widths. The graph shows $\langle 1/J_{mean}\rangle$ (solid line) and $\langle 1/J_{cov}\rangle$ (dashed line). Note that this behavior is present only for the Poisson-like Gaussian but not for the discrete Poisson noise model. The Fisher information of an independent Poisson distribution is

$$J = \sum_{i=1}^{N} \frac{\bar{r}'^2_i}{\bar{r}_i},$$

which is equal to the first term of Fisher Information in the Gaussian case, J_{mean}. For the Poisson noise model, Fisher Information and therefore the MASE lead to a constant optimum completely independent of time (Fig. S3).

Fig. S3. (*A* and *B*) Replication of the results shown in Fig. 2 with Poisson noise (discrete spike counts). We set

$$p(r|\theta) = \prod_i \bar{r}_i^{r_i} \frac{\exp(\bar{r}_i)}{r_i!}.$$

As in Fig. 2, we compute the MASE (*A*) and the IMDE (*B*) as a function of the tuning width for short and long time intervals (*T* = 10, 100, 500, and 1,000 ms; light gray to black). The results are very similar to the Gaussian case: Fisher Information leads to a narrow tuning curve independent of time and the discrimination error to broad tuning functions for short time intervals and narrow ones for long time intervals. (*C–E*) IMDE for a population of 100 independent neurons with Poisson-like noise and variable Fano factor (Fano factor = 0.25, 1, and 4) as a function of tuning width at two different integration times (*T* = 10 and 500 ms; light gray and dark gray, respectively). We used *M* = 3 × 10⁴ samples for the numerical evaluation. (*F–H*) Same as in *C–E*, but MASE of the same population.

Fig. S4. Neurometric functions with rescaled *x* axis of populations (*n* = 10, ... ,190) with Fisher-optimal tuning functions for different integration times (*T* = 10–600 ms; light gray to dark gray) as a function of Δθ/Δφ. The rescaled neurometric functions for populations of different size and identical integration time are identical. Note the log-scale on the *y* axis. All neurometric functions level off at Δθ/Δφ ∼ 2, independent of the population size.
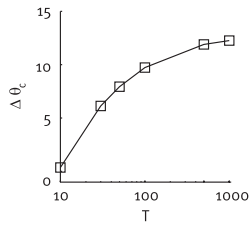
Fig. S5. Dependence of the critical $\Delta\theta_c$ on time. Starting at $\Delta\theta_c$, populations with uniform correlations outperform populations with stimulus-dependent correlations, on the available decoding time T. The value of $\Delta\theta_c$ was extracted from the smoothed, relative versions of the neurometric functions.
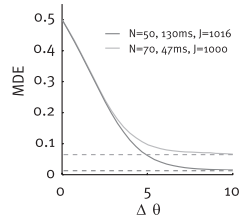

Fig. S6. Neurometric functions of two neural populations with independent noise and Fisher-optimal tuning functions (population 1, $n = 70$, $T = 47$ ms; population 2, $n = 50$, T = 130 ms). The Fisher information of both populations is almost equal (1,000 vs. 1,016) but the pedestal errors are quite different. Note that in this case Fisher information and neurometric functions were calculated for a stimulus located at the peak of one of the tuning functions and not averaged over stimuli.