

¹Department of Biosystems Science and Engineering, ETH Zürich, 4058 Basel, Switzerland

²Department of Physics, Ecole Normale Supérieure, 75005 Paris, France

³Laboratoire de Physique Statistique, Centre National de la Recherche Scientifique, Université Pierre et Marie Curie, Université Denis Diderot, France

⁴Friedrich Miescher Institute, 4058 Basel, Switzerland

Structures of Neural Correlation and How They Favor Coding

SUPPLEMENTARY METHODS

1 Computation of the coding error and the percent improvement due to correlation

1.1 Estimation of the empirical covariance matrix

To compute the Cramér-Rao bound [1] in a group of neurons, we estimated the empirical covariance matrix, $C_{\text{empirical, raw}}(\theta)$, for each considered group of cells and for each stimulus, θ , independently. Even though we had 100 stimulus repetitions, the resulting covariance estimates could be noisy. We therefore decided to estimate the covariance in a more robust manner by considering the adjacent 2 stimuli in both directions, thus increasing the number of repetitions over which the covariance was estimated to 500, but smoothing the correlation structure over the stimulus. This procedure was implemented by circular boxcar filtering of $C_{\text{empirical, raw}}(\theta)$ with a window of length 5, to obtain $C_{\text{empirical}}(\theta)$. We deliberately decided to keep this procedure simple, since more involved smoothing techniques, like the Savitzky-Golai filtering, are not guaranteed to yield valid covariance matrices after smoothing (this can happen when the filter coefficients are negative). For some stimuli close to the null direction of a neuron, the neuron did not elicit a single spike within the 100 repetitions, yielding a vanishing variance. Since this would introduce a pathological eigenvalue of 0 into the covariance matrix, we set the variances of each covariance matrix to be at least equal to 0.1.

1.2 Fisher information and the Cramér-Rao bound

We then went on to compute the stimulus-dependent Fisher information in the Gaussian approximation (sometimes referred to as linear Fisher information), as

$$I_F(\theta) = F'(\theta)^T C^{-1}(\theta) F'(\theta), \quad (1)$$

where $F'(\theta) = (f'_1(\theta), \dots, f'_n(\theta))^T$ is the vector whose elements are the empirical derivatives of the tuning functions, $f_i(\theta)$, $i = 1, \dots, n$, of all n neurons in the group considered, the

superscript T denotes the transpose operation, and where we have neglected the positive trace term in the Fisher information (see also Secs. 3.1 and 3.2, below). According to the Cramér-Rao bound, the variance of the decoded stimulus direction is at least as large as the quantity defined as

$$\delta\theta^2(\theta) = \frac{1}{I_F(\theta)}. \quad (2)$$

1.3 Computation of the coding error and the percent improvement due to correlation

To assess the influence of correlations upon coding, we computed this bound on the variance of the coding error in three cases:

(i) The empirical, correlated case ($\delta\theta_{\text{correlated}}^2$), in which we use the empirical covariance matrix, $C_{\text{empirical}}(\theta)$.

(ii) The case of stimulus-independent correlation ($\delta\theta_{\text{stim-indep}}^2$), in which we use $C_{\text{stim-indep}}(\theta)$, a covariance matrix constructed from stimulus-independent correlations. To compute this covariance matrix, we first estimated the stimulus-dependent correlation matrix, $\chi(\theta)$, from the data, and then averaged it over θ , to obtain $\chi_{\text{mean}} = \langle \chi(\theta) \rangle_{\theta}$. Using the empirical stimulus-dependent variances, $V(\theta) = (v_i(\theta), \dots, v_n(\theta))$, we calculated $C_{\text{stim-indep}}(\theta) \equiv \sqrt{V(\theta)^T} \sqrt{V(\theta)} \odot \chi_{\text{mean}}$, where the square roots are applied over each element of the vector and \odot denotes element-wise matrix product.

(iii) The uncorrelated case ($\delta\theta_{\text{independent}}^2$), in which we use $C_{\text{independent}}(\theta)$, a covariance matrix equal to $C_{\text{empirical}}(\theta)$ but with all off-diagonal elements set to 0. Equivalently, $C_{\text{independent}}(\theta)$ is obtained by shuffling trials in the data.

We note that both covariance matrices, $C_{\text{independent}}(\theta)$ and $C_{\text{stim-indep}}(\theta)$, are dependent upon the stimulus (through the variances), while the underlying correlations are stimulus-independent.

To quantify the effect of the correlation structure upon the coding performance, we defined the percent improvement due to correlations as

$$\Delta R = \left(1 - \left\langle \frac{\delta\theta^2(\theta)}{\delta\theta_{\text{independent}}^2(\theta)} \right\rangle_{\theta} \right) \times 100 \quad (3)$$

when the comparison is with the uncorrelated case, and as

$$\Delta R_{\text{stim-indep}} = \left(1 - \left\langle \frac{\delta\theta_{\text{stim-indep}}^2(\theta)}{\delta\theta_{\text{independent}}^2(\theta)} \right\rangle_{\theta} \right) \times 100 \quad (4)$$

when the comparison is with the stimulus-independent case.

In the case of a two-dimensional stimulus space, the above treatment is generalized in a straightforward manner. We denote the two stimulus variables by θ and d , by analogy with

the variables used in our functional models (see below). The linear Fisher information, again ignoring the trace term, is promoted to a 2×2 matrix with elements

$$(I_F(\theta, d))_{\alpha\beta} = \frac{F(\theta, d)^T}{\partial\alpha} C^{-1}(\theta, d) \frac{F(\theta, d)}{\partial\beta}, \quad (5)$$

where $\alpha, \beta = \theta, d$. Here, $F(\theta, d)$ is the vector whose elements are the average responses of the cells in response to a stimulus direction θ and a stimulus position d , and $C(\theta, d)$ is the stimulus-dependent covariance matrix of the population. The variances and covariances of the output of an arbitrary, unbiased decoder are bounded below by the inverse of the Fisher information matrix. In particular, the variance of the decoded direction and the variance of the decoded position are bounded below by the Cramér-Rao bounds,

$$\delta\theta^2(\theta, d) = (I_F(\theta, d)^{-1})_{\theta\theta} \quad \text{and} \quad \delta d^2(\theta, d) = (I_F(\theta, d)^{-1})_{dd}, \quad (6)$$

respectively. We define the coding errors for direction and for position as the square roots of these two quantities, respectively.

In Fig. 5B, the coding improvement due to correlation was defined analogously to the case of a one-dimensional stimulus (direction coding only). In order to derive results for independent populations, the off-diagonal elements of the correlation matrix were set to zero; in order to derive results for populations with stimulus-independent correlation, the correlation matrix was replaced by its average over all stimulus directions. In Fig. 6C and D, we defined the coding extent for direction as the width of the range of positions in which the coding error for direction is inferior to 15° and the coding extent for position as the width of the range of positions for which the coding error for position is inferior to 0.04 mm.

2 Phenomenological encoding models of direction-selective neurons

2.1 Models of direction coding

We consider that the output of a neuron, labeled by index μ , can be described by a rate, λ_μ , and we assume that the mean firing is then λ_μ and that the variance about this mean is $\lambda_\mu^{1-\beta}$. Poisson variability corresponds to the special case with $\beta = 0$; retina direction-selective neurons display sub-Poisson variability, so that in our data we expect $\beta > 0$. We refer to this source of noise as ‘intrinsic’, as it can be interpreted to mimic the noise inherent to single-cell processing, and so we assume that it is uncorrelated among different neurons.

We then have to specify how the rate, λ_μ , is constructed for each neuron. Here, we assume a doubly stochastic process, i.e., we assume that this rate itself is a random

variable. In the spirit of phenomenological models, we consider the three simplest ways in which λ_μ can fluctuate; specifically, we assume that λ_μ can be obtained as

$$\lambda_\mu = (1 + \sigma_3 \eta_\mu^{(3)}) g \left((1 + \sigma_2 \eta_\mu^{(2)}) \varphi_\mu(\theta) + \sigma_1 \eta_\mu^{(1)} \right). \quad (7)$$

Here, $g(\cdot)$ is the non-linear transfer function of the neuron and $\varphi_\mu(\theta)$ is its stimulus-dependent input; the three positive constants, $\sigma_1, \sigma_2, \sigma_3$, define the magnitude of each source of noise; finally, the random variables, $\eta_\mu^{(1)}, \eta_\mu^{(2)}, \eta_\mu^{(3)}$, have vanishing means,

$$\langle \eta_\mu^{(i)} \rangle = 0, \quad (8)$$

are uncorrelated among each other,

$$\langle \eta_\mu^{(i)} \eta_\nu^{(j)} \rangle = 0 \text{ if } i \neq j, \quad (9)$$

but are individually correlated among neurons,

$$\langle \eta_\mu^{(i)} \eta_\nu^{(i)} \rangle = \delta_{\mu\nu} + \rho_{\mu\nu}^{(i)} (1 - \delta_{\mu\nu}). \quad (10)$$

Thus, $\eta_\mu^{(i)}$ has unit variance and is correlated among neurons with a coefficient $\rho_{\mu\nu}^{(i)}$. The *first* random term in Eq. (7), $\sigma_1 \eta_\mu^{(1)}$, can be interpreted as additive input noise. The *second* random term in Eq. (7), $\sigma_2 \eta_\mu^{(2)}$, can be interpreted as input gain modulation. The *third* random term in Eq. (7), $\sigma_3 \eta_\mu^{(3)}$, can be interpreted as output gain modulation. We refer to these three types of noise collectively as ‘modulation noise’. In the case of retinal direction-selective neurons, the second term can be thought of affecting amacrine cell inputs to the ganglion cells, while the third term can be thought of as arising from common bipolar input to different ganglion cells.

With these elements in hand, we can compute the first- and second-order moment of the neural activity. Hereafter, we denote by double brackets, $\langle\langle \cdot \rangle\rangle$, the double average over the intrinsic noise and over the modulation noise, i.e., over the random variables $\eta_\mu^{(1)}, \eta_\mu^{(2)}, \eta_\mu^{(3)}$. Assuming small amplitudes of the modulation noise and expanding to second order in $\sigma_1, \sigma_2, \sigma_3$, we obtain the mean response as

$$\begin{aligned} \langle\langle r_\mu \rangle\rangle &= \langle g(\varphi_\mu(\theta) + \eta) \rangle_{\eta_\mu^{(1)}, \eta_\mu^{(2)}, \eta_\mu^{(3)}} \\ &\approx g(\varphi_\mu(\theta)) + \frac{1}{2} g''(\varphi_\mu(\theta)) (\sigma_2^2 \varphi_\mu(\theta)^2 + \sigma_1^2); \end{aligned} \quad (11)$$

by construction, this mean corresponds to the tuning curve of neuron μ , i.e., $f_\mu(\theta) = \langle\langle r_\mu \rangle\rangle$. The second-order moments are obtained similarly; the variance reads

$$\begin{aligned} \langle\langle r_\mu^2 \rangle\rangle - \langle\langle r_\mu \rangle\rangle^2 &\approx g'(\varphi_\mu(\theta))^2 (\sigma_2^2 \varphi_\mu(\theta)^2 + \sigma_1^2) \\ &\quad + \sigma_1^2 \{g(\varphi_\mu(\theta))^2 \\ &\quad + [g'(\varphi_\mu(\theta))^2 + g(\varphi_\mu(\theta)) g''(\varphi_\mu(\theta))] (\sigma_2^2 \varphi_\mu(\theta)^2 + \sigma_1^2)\} \\ &\quad + \left[1 - \frac{\beta}{2} (1 - \beta) \sigma_3^2 \right] \left\{ g(\varphi_\mu(\theta))^{1-\beta} + \frac{1}{2} (1 - \beta) g(\varphi_\mu(\theta))^{-\beta} \right. \\ &\quad \left. \times [g''(\varphi_\mu(\theta)) - \beta g(\varphi_\mu(\theta))^{-1} g'(\varphi_\mu(\theta))^2] (\sigma_2^2 \varphi_\mu(\theta)^2 + \sigma_1^2) \right\}, \end{aligned} \quad (12)$$

and the covariance reads

$$\begin{aligned} \langle\langle r_\mu r_\nu \rangle\rangle - \langle\langle r_\mu \rangle\rangle \langle\langle r_\nu \rangle\rangle &\approx g'(\varphi_\mu(\theta)) g'(\varphi_\nu(\theta)) (\sigma_2^2 \varphi_\mu(\theta) \varphi_\nu(\theta) \rho_{\mu\nu}^{(2)} + \sigma_1^2 \rho_{\mu\nu}^{(1)}) \\ &+ \sigma_3^2 [f_\mu(\theta) f_\nu(\theta) + g'(\varphi_\mu(\theta)) g'(\varphi_\nu(\theta))] \\ &\times (\sigma_2^2 \varphi_\mu(\theta) \varphi_\nu(\theta) \rho_{\mu\nu}^{(2)} + \sigma_1^2 \rho_{\mu\nu}^{(1)}) \rho_{\mu\nu}^{(3)}. \end{aligned} \quad (13)$$

The correlation between neurons μ and ν , $c_{\mu\nu}$, is obtained by normalizing the covariance appropriately, as

$$c_{\mu\nu} = \frac{\langle\langle r_\mu r_\nu \rangle\rangle - \langle\langle r_\mu \rangle\rangle \langle\langle r_\nu \rangle\rangle}{\sqrt{\langle\langle r_\mu^2 \rangle\rangle - \langle\langle r_\mu \rangle\rangle^2} \sqrt{\langle\langle r_\nu^2 \rangle\rangle - \langle\langle r_\nu \rangle\rangle^2}}. \quad (14)$$

The noteworthy point, here, is that the modulation noise, although independent of stimulus ($\eta_\mu^{(1)}$, $\eta_\mu^{(2)}$ and $\eta_\mu^{(3)}$ come with vanishing mean and constant variances and covariances), invests the correlations in neural activity with stimulus-dependence.

In our study, we investigated the role of each of the three types of modulation noise separately, i.e., we considered each of the three cases with $\sigma_2 = \sigma_3 = 0$, $\sigma_1 = \sigma_3 = 0$, and $\sigma_1 = \sigma_2 = 0$. In the presence of *additive input noise* only ($\sigma_1 \neq 0$, $\sigma_2 = \sigma_3 = 0$), the tuning curve, variance, and covariance reduce to the forms

$$f_\mu(\theta) \approx g(\varphi_\mu(\theta)) + \frac{1}{2} g''(\varphi_\mu(\theta)) \sigma_1^2, \quad (15)$$

$$\begin{aligned} \langle\langle r_\mu^2 \rangle\rangle - \langle\langle r_\mu \rangle\rangle^2 &\approx g(\varphi_\mu(\theta))^{1-\beta} + \left\{ g'(\varphi_\mu(\theta))^2 + \frac{1}{2} (1-\beta) g(\varphi_\mu(\theta))^{-\beta} \right. \\ &\times \left. [g''(\varphi_\mu(\theta)) - \beta g(\varphi_\mu(\theta))^{-1} g'(\varphi_\mu(\theta))^2] \right\} \sigma_1^2, \end{aligned} \quad (16)$$

$$\langle\langle r_\mu r_\nu \rangle\rangle - \langle\langle r_\mu \rangle\rangle \langle\langle r_\nu \rangle\rangle \approx g'(\varphi_\mu(\theta)) g'(\varphi_\nu(\theta)) \sigma_1^2 \rho_{\mu\nu}^{(1)}. \quad (17)$$

In the presence of *input gain modulation* only ($\sigma_2 \neq 0$, $\sigma_1 = \sigma_3 = 0$), the tuning curve, variance, and covariance reduce to the forms

$$f_\mu(\theta) \approx g(\varphi_\mu(\theta)) + \frac{1}{2} g''(\varphi_\mu(\theta)) \varphi_\mu(\theta)^2 \sigma_2^2, \quad (18)$$

$$\begin{aligned} \langle\langle r_\mu^2 \rangle\rangle - \langle\langle r_\mu \rangle\rangle^2 &\approx g(\varphi_\mu(\theta))^{1-\beta} + \left\{ g'(\varphi_\mu(\theta))^2 + \frac{1}{2} (1-\beta) g(\varphi_\mu(\theta))^{-\beta} \right. \\ &\times \left. [g''(\varphi_\mu(\theta)) - \beta g(\varphi_\mu(\theta))^{-1} g'(\varphi_\mu(\theta))^2] \right\} \varphi_\mu(\theta)^2 \sigma_2^2, \end{aligned} \quad (19)$$

$$\langle\langle r_\mu r_\nu \rangle\rangle - \langle\langle r_\mu \rangle\rangle \langle\langle r_\nu \rangle\rangle \approx g'(\varphi_\mu(\theta)) g'(\varphi_\nu(\theta)) \varphi_\mu(\theta) \varphi_\nu(\theta) \sigma_2^2 \rho_{\mu\nu}^{(2)}. \quad (20)$$

In the presence of *output gain modulation* only ($\sigma_3 \neq 0$, $\sigma_1 = \sigma_2 = 0$), the tuning curve, variance, and covariance reduce to the forms

$$f_\mu(\theta) \approx g(\varphi_\mu(\theta)), \quad (21)$$

$$\langle\langle r_\mu^2 \rangle\rangle - \langle\langle r_\mu \rangle\rangle^2 \approx g(\varphi_\mu(\theta))^{1-\beta} + \left[g(\varphi_\mu(\theta))^2 - \frac{\beta}{2}(1-\beta)g(\varphi_\mu(\theta))^{1-\beta} \right] \sigma_3^2, \quad (22)$$

$$\langle\langle r_\mu r_\nu \rangle\rangle - \langle\langle r_\mu \rangle\rangle \langle\langle r_\nu \rangle\rangle \approx g(\varphi_\mu(\theta))g(\varphi_\nu(\theta))\sigma_3^2\rho_{\mu\nu}^{(3)}. \quad (23)$$

In the main text, we consider two incarnations of each of the three models. In Models I, II, III, we consider that all neurons share the same source of secondary noise, i.e., we set all $\rho_{\mu\nu}^{(i)}$ to unity. More precisely, these models are defined by the following sets of parameters.

- Model I: $\sigma_1 \neq 0, \sigma_2 = \sigma_3 = 0, \rho_{\mu\nu}^{(1)} = 1$.
- Model II: $\sigma_2 \neq 0, \sigma_1 = \sigma_3 = 0, \rho_{\mu\nu}^{(2)} = 1$.
- Model III: $\sigma_3 \neq 0, \sigma_1 = \sigma_2 = 0, \rho_{\mu\nu}^{(3)} = 1$.

In the models denoted by primed numbers, Models I', II', and III', we assume that the secondary source of noise is shared only partially by different neurons, so that the value of $\rho_{\mu\nu}$ can differ from unity. More precisely, these models are defined by the following sets of parameters.

- Model I': $\sigma_1 \neq 0, \sigma_2 = \sigma_3 = 0, -1 \leq \rho_{\mu\nu}^{(1)} \leq 1$.
- Model II': $\sigma_2 \neq 0, \sigma_1 = \sigma_3 = 0, -1 \leq \rho_{\mu\nu}^{(2)} \leq 1$.
- Model III': $\sigma_3 \neq 0, \sigma_1 = \sigma_2 = 0, -1 \leq \rho_{\mu\nu}^{(3)} \leq 1$.

2.2 Fitting the models of direction coding to data

Each of the six phenomenological models (Models I, II, III and Models I', II', III') were fit in a two-step procedure. In the first step of the fitting procedure, for each neuron, the neuron's tuning curve was fit by a flat-topped von Mises function [2],

$$f_\mu(\theta) = A \exp\left(\frac{\cos(\theta - \theta_\mu + Z \sin(\theta - \theta_\mu)) - 1}{W}\right) + B, \quad (24)$$

where $f_\mu(\theta)$ denotes the mean response of neuron μ to the presentation of stimulus θ and $A, B, W,$ and Z are positive constants. All parameters were fit simultaneously with the Nelder-Mead Simplex Method using Matlab's `fminsearch` function, which minimizes the squared error between the fit and the empirical tuning function.

In the second step of the fitting procedure, we assigned $\varphi_\mu(\theta)$ the same form as $f_\mu(\theta)$; we kept all parameters fixed, but allowed the scale of A and B to change (i.e., allowed a free multiplicative factor). Then, for each pair of neurons, labeled by μ and ν , within each experiment, the six phenomenological models were fit, again all parameters

simultaneously, including the two multiplicative scale parameters in $\varphi_\mu(\theta)$ and $\varphi_\nu(\theta)$, using Matlab’s `fminsearch`. Since all six models yield the tuning functions, $f_\mu(\theta)$, variances of each neuron, $V_\mu(\theta)$, and the correlation between two neurons, $\chi_{\mu\nu}(\theta)$, we optimized the average variance explained for all three of these quantities. That is, we computed each contribution,

$$E_{\text{tuning curve}} \equiv \frac{\sum_\mu [f_{\mu,\text{fit}}(\theta) - f_{\mu,\text{empirical}}(\theta)]^2}{\sum_\mu [f_{\mu,\text{empirical}}(\theta) - \langle f_{\mu,\text{empirical}}(\theta) \rangle_\theta]^2}, \quad (25)$$

$$E_{\text{variance}} \equiv \frac{\sum_\mu [V_{\mu,\text{fit}}(\theta) - V_{\mu,\text{empirical}}(\theta)]^2}{\sum_\mu [V_{\mu,\text{empirical}}(\theta) - \langle V_{\mu,\text{empirical}}(\theta) \rangle_\theta]^2}, \quad (26)$$

$$E_{\text{correlation}} \equiv \frac{[\chi_{\mu\nu,\text{fit}}(\theta) - \chi_{\mu\nu,\text{empirical}}(\theta)]^2}{[\chi_{\mu\nu,\text{empirical}}(\theta) - \langle \chi_{\mu\nu,\text{empirical}}(\theta) \rangle_\theta]^2}, \quad (27)$$

and we minimized the appropriately weighted summed error,

$$E_{\text{total}} = \frac{1}{2}E_{\text{tuning curve}} + \frac{1}{2}E_{\text{variance}} + E_{\text{correlation}}. \quad (28)$$

Here, $\langle \cdot \rangle_\theta$ denotes an average over stimuli, and the subscripts “ μ,fit ” and “ $\mu,\text{empirical}$ ” refer to the fitted and empirical quantities relative to neuron μ , respectively. From the residual error of the fitting procedure, we computed the percent variance explained, $(1 - E_{\text{total}}/3) \times 100$. For the nonlinearity that enters the models, we found that a form as simple as a pure power law, $g(x) = x^p$ yielded good fits. Thus, in this step of the fitting procedure, the number of fitting parameters was 5 (for Models I, II, and III) or 6 (for Models I’, II’, and III’).

The variance explained by the fits was larger than 50% (Models II and III) and larger than 47% (Model I) for over 75% of the fits; the means \pm standard deviations of the variance explained were 56% \pm 16% (Model I), 60% \pm 17% (Model II), and 60% \pm 17% (Model III) (Fig. S1). Overall, Models II and III appeared to explain a larger fraction of the data. While allowing for partial correlation of the secondary noise terms among neurons (i.e., $-1 < \rho_{\mu\nu}^{(i)} < 1$, Models I’, II’, III’) improved the fits, the improvement was modest. Furthermore, for more than 50% of the cell pairs, the fits yielded $\rho_{\mu\nu}^{(i)} > 0.95$ in Models I’ and II’, compared to 19% in Model III’. The means \pm standard deviations of the variance explained were 63% \pm 12% (Model I), 63% \pm 15% (Model II), and 63% \pm 15% (Model III) (Fig. S1). That is, the improvement in variance explained from partial correlation did not exceed 3% for Models II’ and III’ and 7% for Model I’. For the sake of simplicity and because its impact was modest, we neglected the possibility of partial correlation in the main text: for neighboring direction-selective cells, which share inputs from the same bipolar cells, high correlation is expected.

2.3 Models of direction and position coding, and fits to the data

For finer temporal coding, in which case the response depends not only on the direction of the stimulus but also on its position, we extended our phenomenological models to include the dependence upon position. We again develop six extended models (I, II, III, I', II', and III'), in analogy with the simpler models described above. As before, all models use a same tuning function (now two-dimensional) for individual neurons, defined as

$$f_\mu(\theta, d) = \ln(1 + \exp(f_\mu(\theta)\varphi_\mu(d) - D)) + \tilde{D}, \quad (29)$$

for neuron μ , where d is the distance between the leading edge of the stimulus bar and the stimulus location that elicits the strongest response when the bar moves in the preferred direction of the cell, and $f_\mu(\theta)$ is defined in Eq. (24). Here, the ‘tuning function’ $\varphi_\mu(d)$ provides the position dependence; we chose for it a simple Gaussian form,

$$\varphi_\mu(d) = \tilde{A} \exp\left(-\frac{1}{2} \frac{d^2}{\tilde{W}}\right) + \tilde{B}, \quad (30)$$

where the constants \tilde{A} , \tilde{B} , and \tilde{W} are named by analogy to those in Eq. (24). We introduced the exponential and the logarithm in Eq. (29) to have a soft rectifier, because, in the case of fine temporal coding, a given cell will be silent or nearly silent over a range of positions; the constant, D , controls the crossover from a regime with $f_\mu(\theta, d) \approx f_\mu(\theta)\varphi_\mu(d)$ to a regime with $f_\mu(\theta, d) \approx 0$, and the constant, \tilde{D} , corresponds to the spontaneous activity of the cell. Equation (30) allows only a single maximum because we model the ON response, i.e., the response of direction-selective cells to the leading edge of the stimulus.

We followed the same procedure as before when fitting the data; specifically, the ON responses of the cells. All parameters were fit simultaneously. The experimental two-dimensional tuning function, the two-dimensional variances, and the two-dimensional covariance were estimated by binning spike trains in non-overlapping, 50 ms bins; this time window, corresponds, given the speed of the stimulus bar (1.6 mm/s), to an 80 μm displacement of the leading edge of the bar. Since the cells are mostly silent until they are stimulated by the stimulus bar, it was not uncommon for a neuron to fire no spike at all in all 100 trials, for given position-direction bins, especially close to the onset of bar movement. For those bins, the correlation between two neurons is not defined; we excluded them from the fitting procedure. Specifically, we applied a ‘mask’ by thresholding the mean responses of two neurons in a pair: if either of the neurons had a mean response of below 5% of its maximal response, we excluded the corresponding bin in the calculation of the error function for the correlation (Eq. (27)). While some pairs exhibited negative correlations for some of the position-direction bins, we did not attempt to fit negative correlation values.

For all six extended models, the mean variance explained was close to 50% (all models). More than 60% of the fits had a variance explained larger than 50% (all models). The

means \pm standard deviations of the variance explained were $51\% \pm 9\%$ (Model I), $50\% \pm 8\%$ (Model II), and $50\% \pm 9\%$ (Model III) (Fig. S2). For Models II and III, the best 25% fits had a variance explained ranging from 56% to 70%. As in the simpler models, allowing for partial correlation of the secondary noise terms among neurons (i.e., $-1 < \rho_{\mu\nu}^{(i)} < 1$, Models I, II, III) did not yield any sensitive improvement of the fits. The corresponding means \pm standard deviations of the variance explained were $51\% \pm 8\%$ (Model I), $50\% \pm 9\%$ (Model II), and $50\% \pm 8\%$ (Model III) (Fig. S2).

2.4 Simulation of a mosaic of direction-selective neurons

For the simulation of 12 direction-selective neurons arranged as a retinal mosaic made up of three quadruplets (Figs. 5 and 6), we placed the points of maximal response of each cell on a circle with varying radius (‘mosaic spacing’, referred to in Fig. 6). The center of the circle was defined as the center of the coordinate system. The three quadruplets were spaced equidistantly. Within each quadruplet, the position preference of the four cells coincided and their preferred directions pointed along the four cardinal directions. In order to avoid having a particular symmetry, in which a preferred direction would be along a side of the triangle of quadruplets, the latter was rotated at an angle of 15° away from a cardinal direction.

For each model neuron, we used the same model parameters except for its preferred direction and its position on the retina. We used the parameters $\beta = 0$, $\sigma = 0.3000$, $p = 1.0060$, $\rho = 1$, to characterize the cells. In addition, in Eq. (24), we used $A = 12.4664$, $B = 0.1917$, $W = 5.8808$, and $Z = -0.3000$; in Eq. (29), we used $D = 197.4445$ and $\tilde{D} = 0.0200$; in Eq. (30), we used $\tilde{A} = 15.9331$, $\tilde{B} = 0$, and $\tilde{W} = 0.4308$.

3 Theoretical analyses of coding with correlated neurons

3.1 Definitions and notation

We consider a continuous, periodic stimulus, θ , which is normalized so as to vary between 0 and 2π , as is natural in the case of direction selectivity. The response of an individual neuron is characterized by the number of spikes it emits during a given time window. Each neuron has a stimulus preference: neuron μ responds most strongly to a stimulus θ_μ —the ‘preferred stimulus’—and gradually less strongly to stimuli that take values away from θ_μ . In the case of retinal direction-selective cells, fits to this tuning curve are provided by a von Mises form,

$$f_\mu(\theta) = A \exp\left(\frac{\cos(\theta - \theta_\mu)}{W}\right), \quad (31)$$

or a flat-topped von Mises form, Eq. (24), where $f_\mu(\theta)$ denotes the mean response of neuron μ to the presentation of stimulus θ and A and W are positive constants. In agreement with data, and following earlier theoretical work, we assume throughout that the width of the tuning curve is an appreciable fraction of the range of stimuli, 2π .

By and large, in earlier studies, the set of preferred stimuli in the population was arranged uniformly in the interval $[0, 2\pi]$ or with small perturbations to a uniform arrangement. Here, we relax this constraint and allow neurons to be arranged in pools: a total of $N_0 = N \cdot m$ neurons are divided into N ‘pools’ with m neurons each. In the most basic incarnation of the model, all neurons in a given pool have identical tuning curves: neuron i ($i = 1, \dots, m$) in pool μ ($\mu = 1, \dots, N$) responds according to Eq. (31) or Eq. (24), with $\theta_\mu = 2\pi\mu/N$. The limiting case with $m = 1$ reduces to a uniform arrangement; the other limiting case with $N = 1$ amounts to a population of identical neurons. The case of the four types of retinal direction-selective cells we recorded from corresponds to $N = 4$, as the preferred directions align along the four cardinal directions.

From trial to trial, the activity fluctuates about the mean given by Eqs. (31, 24), and this noise may be correlated among neurons. For the sake of simplicity, in the general theory we assume Gaussian noise with a covariance matrix $C_{\mu\nu ij} \equiv \langle (r_{\mu i} - \langle r_{\mu i} \rangle) (r_{\nu j} - \langle r_{\nu j} \rangle) \rangle$, where $r_{\mu i}$ is the response of neuron i in pool μ . Following the literature, we adopt the simple form of the covariance matrix,

$$C_{\mu\nu ij} = \sigma^2 v_\mu(\theta) v_\nu(\theta) [\delta_{\mu\nu} \delta_{ij} + c(0) \delta_{\mu\nu} (1 - \delta_{ij}) + c(\theta_\mu - \theta_\nu) (1 - \delta_{\mu\nu})], \quad (32)$$

where σ^2 is a constant which scales the variance (auto-correlation) of the activity of each individual neuron, $c(0)$ is the correlation between two different neurons in the same pool, and $c(\theta_\mu - \theta_\nu)$ is the correlation between two neurons in two different pools with stimulus preferences θ_μ and θ_ν respectively. Furthermore, $v_\mu(\theta)$ describes the dependence upon the stimulus, θ , of the standard deviation of the response of a neuron in pool μ ; the correlation coefficients, $c(0)$ and $c(\theta_\mu - \theta_\nu)$, are assumed to be stimulus-independent in the present analysis. Poisson variability corresponds to the choice $\sigma v_\mu(\theta) = \sqrt{f_\mu(\theta)}$, sub-Poisson variability can be fitted with a power law, $v_\mu(\theta) \propto v_\mu(\theta)^{(1-\beta)/2}$, and stimulus-independent (additive) variability corresponds to the choice $v_\mu(\theta) = 1$.

3.2 Coding error for a population of correlated neurons: general expressions

We calculate a lower bound to the variance of a deterministic decoder, $\langle \delta\theta^2 \rangle$, through the Cramér-Rao bound,

$$\langle \delta\theta^2 \rangle \geq \frac{1}{I_F(\theta)}, \quad (33)$$

where $I_F(\theta)$ is the Fisher information evaluated at a stimulus value θ . If the response of neurons, $r_{\mu i}$, where the index $\mu = 1, \dots, N$ labels the pools and the index $i = 1, \dots, m$

labels the neurons within a given pool, obey a probability density, $P(\{r_{\mu i}\} | \theta)$, with Gaussian fluctuation about the turning curves, then it is well known that the Fisher information, defined as

$$I_F(\theta) = \langle -\partial_\theta^2 \ln(P(\{r_{\mu i}\} | \theta)) \rangle_{P(\{r_{\mu i}\} | \theta)}, \quad (34)$$

reduces to the expression

$$I_F(\theta) = f'(\theta)^T C^{-1} f'(\theta) + \frac{1}{2} \text{Tr} [(\partial_\theta C) C^{-1} (\partial_\theta C) C^{-1}]. \quad (35)$$

Here, C is the covariance matrix defined in Eq. (32) and $f'(\theta)$ is the vector with elements $f'_{\mu i}(\theta)$, the derivatives of the tuning curve,

$$f'_{\mu i}(\theta) = \partial_\theta f_{\mu i}(\theta) = \partial_\theta f_\mu(\theta) = \partial_\theta f_\mu(\theta - \theta_\mu), \quad (36)$$

where θ_μ is the preferred stimulus in pool μ and $f_\mu(\theta - \theta_\mu) = f(\theta - \theta_\mu)$ depends upon μ only through its argument. The trace term in the Fisher information (the second term on the right-hand side) vanishes in the stimulus-independent case. More generally, this term is always positive, so we do not violate the bound if we ignore it. As, in our data analysis, we have found that this term is consistently negligible with respect to the first term on the right-hand side, we ignore the trace term henceforth.

The covariance matrix Eq. (32) can be rewritten as

$$C = V\chi V, \quad (37)$$

where V is the diagonal matrix given by

$$V_{\mu\nu ij} = \sigma v_\mu(\theta) \delta_{\mu\nu} \delta_{ij} \quad (38)$$

and the matrix χ has elements

$$\chi_{\mu\nu ij} = \delta_{\mu\nu} \delta_{ij} + c(0) \delta_{\mu\nu} (1 - \delta_{ij}) + c(\theta_\mu - \theta_\nu) (1 - \delta_{\mu\nu}). \quad (39)$$

In order to calculate the Fisher information,

$$I_F(\theta) = f'(\theta)^T V^{-1} \chi^{-1} V^{-1} f'(\theta), \quad (40)$$

we diagonalize χ with the orthonormal matrix, S , whose columns contain the elements of the eigenvectors of χ , $u^{(\kappa, k)}$, labeled by the index (κ, k) with $\kappa = 1, \dots, N$ and $k = 1, \dots, m$; i.e., $S_{\mu i \nu j} = u_{\mu i}^{(\nu, j)}$. Since $S^{-1} = S^\dagger$, we have

$$\begin{aligned} \chi^{-1} &= S S^\dagger \chi^{-1} S S^\dagger \\ &= S (S^\dagger \chi S)^{-1} S^\dagger \\ &= S \Lambda^{-1} S^\dagger, \end{aligned} \quad (41)$$

where Λ is a diagonal matrix with the eigenvalues of χ on its diagonal. Thus, we can rewrite the Fisher information as

$$\begin{aligned} I_F(\theta) &= f'(\theta)^T V^{-1} S \Lambda^{-1} S^\dagger V^{-1} f'(\theta) \\ &= F(\theta) \Lambda^{-1} F(\theta)^\dagger, \end{aligned} \quad (42)$$

where

$$F(\theta) = f'(\theta)^T V^{-1} S \quad (43)$$

or, in components,

$$F_{\kappa k}(\theta) = \sum_{\mu, i} \frac{f'_{\mu i}(\theta)}{\sigma v_\mu(\theta)} u_{\nu j}^{(\kappa, k)}. \quad (44)$$

Since the matrix, χ , defined in Eq. (39) is circulant, the elements of its eigenvectors are the roots of unity: the $(\nu, j)^{\text{th}}$ element of the eigenvector labeled by (κ, k) is

$$u_{\nu j}^{(\kappa, k)} \equiv \frac{1}{\sqrt{Nm}} e^{2\pi i \kappa \nu / N} e^{2\pi i k j / m}. \quad (45)$$

And its eigenvalues, $\Lambda_{\kappa k}$, are calculated from the identity

$$\sum_{\nu, j} \chi_{\mu \nu j} u_{\kappa k}^{\nu j} = \begin{cases} a [1 - c(0)] u_{\kappa k}^{\mu i} & \text{if } k \neq 0 \\ a [1 - c(0) + Nm \tilde{c}_\kappa] u_{\kappa k}^{\mu i} & \text{if } k = 0 \end{cases}, \quad (46)$$

where

$$\tilde{c}_\kappa = \frac{1}{N} \sum_{\mu=0}^{N-1} c(\theta_\mu) e^{-2\pi i \kappa \mu / N}, \quad (47)$$

as

$$\Lambda_{\kappa k} = 1 - c(0) + Nm \tilde{c}_\kappa \delta_{k0}. \quad (48)$$

Finally, combining the expressions of $F(\theta)$ and Λ^{-1} , we obtain the Fisher information as

$$\begin{aligned} I_F(\theta) &= \sum_{\kappa, \kappa', k, k'} F_{\kappa k}(\theta) \left[\frac{\delta_{k0}}{1 - c(0) + Nm \tilde{c}_\kappa} + \frac{1 - \delta_{k0}}{1 - c(0)} \right] \delta_{\kappa \kappa'} \delta_{kk'} F_{\kappa' k'}(\theta)^* \\ &= \frac{Nm}{\sigma^2} \sum_{\kappa=0}^{N-1} \frac{\|\tilde{F}_\kappa(\theta)\|^2}{1 - c(0) + Nm \tilde{c}_\kappa}, \end{aligned} \quad (49)$$

where

$$\tilde{F}_\kappa(\theta) = \frac{1}{N} \sum_{\mu=0}^{N-1} \frac{f'_{\mu i}(\theta)}{v_\mu(\theta)} e^{-2\pi i \kappa \mu / N}. \quad (50)$$

3.3 Coding error for a population of correlated neurons: direction-selective neurons

In the case of direction-selective neurons with preferred directions aligned along the four cardinal directions, where we assume that neurons with the same tuning preference are homogeneous in their properties, the correlations are described completely by three numbers, namely, the correlation between two neurons with identical preferred direction, $c(0)$; with preferred directions that differ by 90° , $c(\pi/2)$; with preferred directions that differ by 180° , $c(\pi)$. The denominators in the expression of the Fisher information (Eq. (49)) then reduce to

$$1 - c(0) + Nm\tilde{c}_\kappa = 1 + (m - 1)c(0) + \begin{cases} 2c(\pi/2) + c(\pi) & \text{if } \kappa = 0 \\ -c(\pi) & \text{if } \kappa = 1 \text{ or } 3 \\ c(\pi) - 2c(\pi/2) & \text{if } \kappa = 2 \end{cases}, \quad (51)$$

and the numerators are the corresponding Fourier transforms, as defined in Eq. (50), for $\kappa = 0, 1, 2, 3$.

3.4 Coding error for a population of correlated neurons: populations of N_0 neurons

As explained in the main text, we consider a non-monotonic form of the function which describes the correlations between pairs of neurons which belong to the same pool or to different pools. For the sake of analytical tractability, we choose the simple exponential form

$$c(\theta) = 4(1 - \beta)c_{\max} \left(e^{-|\theta|/\rho} - (1 - \beta)e^{-2|\theta|/\rho} \right), \quad (52)$$

where $c(\theta)$ is the correlation between two neurons whose preferred stimuli differ by θ (modulo 2π) and $0 \leq \beta < 1/2$. The constant c_{\max} is the maximum value reached by the correlation, for $\theta = \rho \ln(2(1 - \beta))$; thus, pairs of neurons whose stimulus preference differ on the order of ρ are the more correlated ones.

Since the Fourier transformation is a linear operation, we can transform each exponential separately. For the sake of clarity of notation, we give names to each Fourier transform, according to

$$\tilde{c}_\kappa = \frac{4(1 - \beta)c_{\max}}{N} [\beta + g_\kappa(\rho) - (1 - \beta)g_\kappa(\rho/2)]. \quad (53)$$

In this notation, the g_κ -functions are modified Fourier transforms of the exponentials, in which the component with $\mu = 0$ is omitted. These Fourier transforms are calculated

readily, by summing a geometric series. For N even, we have

$$\begin{aligned}
g_\kappa(\rho) &= \sum_{\mu=1}^{N/2-1} \left[e^{-|\theta_\mu|/\rho} e^{-2\pi i \kappa \mu/N} + e^{-|\theta_{-\mu}|/\rho} e^{2\pi i \kappa \mu/N} \right] + e^{-|\theta_{N/2}|/\rho} e^{-\pi i \kappa} \\
&= \sum_{\mu=1}^{N/2-1} \left[e^{(2\pi/N)(\rho^{-1}-i\kappa)\mu} + e^{(2\pi/N)(\rho^{-1}+i\kappa)\mu} \right] + e^{-\pi/\rho} (-1)^\kappa \\
&= e^{(2\pi/N)(\rho^{-1}-i\kappa)} \frac{1 - e^{(2\pi/N)(\rho^{-1}-i\kappa)N/2-1}}{1 - e^{(2\pi/N)(\rho^{-1}-i\kappa)}} \\
&\quad + e^{(2\pi/N)(\rho^{-1}+i\kappa)} \frac{1 - e^{(2\pi/N)(\rho^{-1}+i\kappa)N/2-1}}{1 - e^{(2\pi/N)(\rho^{-1}+i\kappa)}} + (-1)^\kappa e^{-\pi/\rho} \\
&= 2 \operatorname{Re} \left[e^{(2\pi/N)(\rho^{-1}+i\kappa)} \frac{1 - e^{(2\pi/N)(\rho^{-1}+i\kappa)N/2-1}}{1 - e^{(2\pi/N)(\rho^{-1}+i\kappa)}} \right] + (-1)^\kappa e^{-\pi/\rho} \\
&= 2\lambda \frac{\cos(2\pi\kappa/N) - \lambda + (-1)^\kappa \lambda^{N/2} \gamma_\kappa^{\text{even}}(\rho)}{\lambda^2 - 2\lambda \cos(2\pi\kappa/N) + 1} + \zeta_\kappa^{\text{even}}(\rho), \tag{54}
\end{aligned}$$

where

$$\gamma_\kappa^{\text{even}}(\rho) = \cos(2\pi\kappa/N) - \lambda^{-1}, \tag{55}$$

$$\zeta_\kappa^{\text{even}}(\rho) = (-1)^\kappa e^{-\pi/\rho}, \tag{56}$$

and where we have defined $\lambda = e^{2\pi/(N\rho)}$. For N odd, the calculation runs along a similar line, and we find the same form for $g_\kappa(\rho)$, but with $\gamma_\kappa^{\text{even}}(\rho)$ and $\zeta_\kappa^{\text{even}}(\rho)$ replaced by

$$\gamma_\kappa^{\text{odd}}(\rho) = (\lambda^{1/2} - \lambda^{-1/2}) \cos(2\pi\kappa/N), \tag{57}$$

$$\zeta_\kappa^{\text{odd}}(\rho) = 0. \tag{58}$$

Putting all this together, we obtain an expression for the Fisher information, as

$$I_F(\theta) = \frac{Nm}{\sigma^2} \sum_{\kappa=0}^{N-1} \frac{\left\| \tilde{F}_\kappa(\theta) \right\|^2}{1 + 4(1-\beta) c_{\max} \{ (m-1)\beta + m[g_\kappa(\rho) - (1-\beta)g_\kappa(\rho/2)] \}}, \tag{59}$$

where

$$g(\rho) = 2\lambda \frac{\cos(2\pi\kappa/N) - \lambda + (-1)^\kappa \lambda^{N/2} \gamma_\kappa(\rho)}{\lambda^2 - 2\lambda \cos(2\pi\kappa/N) + 1} + \zeta_\kappa(\rho), \tag{60}$$

$$\gamma_\kappa(\rho) = \begin{cases} (\lambda^{1/2} - \lambda^{-1/2}) \cos(2\pi\kappa/N) & \text{if } N \text{ is odd} \\ \cos(2\pi\kappa/N) - \lambda^{-1} & \text{if } N \text{ is even} \end{cases}, \tag{61}$$

$$\zeta_\kappa(\rho) = \begin{cases} 0 & \text{if } N \text{ is odd} \\ (-1)^\kappa \lambda^{N/2} & \text{if } N \text{ is even} \end{cases}, \tag{62}$$

and $\lambda = e^{-2\pi/(N\rho)}$.

References

- [1] T. Cover and J. Thomas. Elements of information theory. 1991.
- [2] Nicholas V Swindale. Orientation tuning curves: empirical description and estimation of parameters. *Biological cybernetics*, 78(1):45–56, 1998.

Figure Captions

Fig. S1. Percent variance explained by the fits of the models coding for direction, for models with perfect correlation in the secondary sources of noise (labeled as Models I, II, III) and for models with partial correlation in the secondary sources of noise (labeled as Models I', II', III').

Fig. S2. Percent variance explained by the fits of the models coding for direction and position, for models with perfect correlation in the secondary sources of noise (labeled as Models I, II, III) and for models with partial correlation in the secondary sources of noise (labeled as Models I', II', III').

- Submission 3/Supplement $varExplained.eps$