Supplementary information

The origin of information-limiting noise correlations

I. Kanitscheider*, R. Coen-Cagli*, A. Pouget

Contents

1. Analytic expressions for tuning curves and noise covariance in the feedforward model

Here we derive analytical expressions for tuning curves and noise covariance in the feedforward model. The Gabor image presented to the retina is parameterized by orientation and can therefore be described by a map from orientation to pixel space

$$\begin{aligned}\mathbf{I} &: R \to R^P \\ \theta &\mapsto \mathbf{I}(\theta)\end{aligned}$$

where P is the number of pixels in the image. Due to input (e.g. photoreceptor) noise the image is corrupted by independent noise of variance σ_0^2, and the actual sensory input is given by

$$\tilde{\mathbf{I}} \,|\, \theta \sim \mathcal{N}(\mathbf{I}(\theta), \sigma_0^2 \mathbb{1}_P) \tag{1}$$

The V1 neurons filter the sensory input by a Gabor $\mathbf{F}_i \in R^P$ oriented according to their preferred orientation θ_i, followed by a rectifying non-linearity and a Poisson step. The neural response of neuron i is then sampled from

$$r_i \sim Poisson([\mathbf{F}_i \cdot \tilde{\mathbf{I}}]_+)\,, \tag{2}$$

where $[]_+$ denotes linear rectification. The tuning curves and covariance matrix are given by

$$f_i(\theta) = \langle r_i \rangle \approx [\mathbf{F}_i \cdot \mathbf{I}(\theta)]_+ \tag{3}$$

$$\Sigma_{ij}(\theta) = Cov(r_i, r_j) \approx \sigma_0^2 [\mathbf{F}_i \cdot \mathbf{F}_j]_+ + \delta_{ij} f_i(\theta)\,,$$

where the approximation is valid for small input noise and for neurons which are well above the firing threshold.

The correlation coefficient for two neurons $i \neq j$ is given by

$$\rho_{ij} = \frac{\sigma_0^2 \left[\mathbf{F}_i \cdot \mathbf{F}_j \right]_+}{\sqrt{\sigma_0^2 \,|\, \mathbf{F}_i \,|^2 + f_i(\theta)} \sqrt{\sigma_0^2 \,|\, \mathbf{F}_j \,|^2 + f_j(\theta)}} \tag{4}$$

From the expression (4) we see that the correlations between the neurons will be proportional to the overlap of their filters $\mathbf{F}_i$ and $\mathbf{F}_j$.

To get an intuition how tuning curves and correlations with realistic shape arise from eq. (3) let us plug in the example of Gabor images and Gabor filter. The Gabor filters are given by

$$F_{(x,y),i} \propto \exp\left(-\frac{x^2 + y^2}{2\sigma^2} \right) \cos\left(\frac{2\pi}{\lambda} (x \cos\theta_i + y \sin\theta_i) + \phi_i \right), \tag{5}$$

where (x, y) are the coordinates of the image, σ is the size of the Gaussian envelope, λ the preferred spatial wavelength, θ_i the preferred angle and ϕ_i the phase offset of the Gabor filter.

The Gabor image is given by

$$I_{(x,y)}(\theta) \propto \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)\cos\left(\frac{2\pi}{\lambda}(x\cos\theta + y\sin\theta) + \phi\right) \tag{6}$$

For now we assume that the image has the same Gaussian envelope and spatial wavelength as the filter. Given eq. (5) and (6) we can derive the tuning curves of the orientation population in the limit of a large number of pixels:

$$f_i(\theta) \propto \left[\sum_{x,y} F_{(x,y),i} I_{(x,y)}(\theta)\right]_+ \approx \left[\frac{P}{L^2}\int dxdy F_{(x,y),i} I_{(x,y)}(\theta)\right]_+ \tag{7}$$

$$= \frac{\pi\sigma^2}{2}\exp(-K)\left[\cos(\phi_i - \phi)\exp(K\cos(\theta_i - \theta)) + \cos(\phi_i + \phi)\exp(-K\cos(\theta_i - \theta))\right]_+,$$

where L is the receptive field size and we have defined $K = 2\pi^2\sigma^2/\lambda^2$. As we can see, the tuning curves are given by a combination of two von-Mises functions centered at $\theta = \theta_i$ and $\theta = \theta_i + \pi$, whose relative gains depend on the phase parameters ϕ_i and ϕ.

The covariance of two neurons $i \neq j$ has a very similar form:

$$\Sigma_{ij} \propto \sigma_0^2 \sum_{x,y} F_{(x,y),i} F_{(x,y),j}$$
$$= \frac{\pi\sigma^2\sigma_0^2}{2}\exp(-K)\left[\cos(\phi_i - \phi_j)\exp(K\cos(\theta_i - \theta_j)) + \cos(\phi_i + \phi_j)\exp(-K\cos(\theta_i - \theta_j))\right]_+ \tag{8}$$

In particular if the phase of the two filters are equal, $\phi_i = \phi_j$, the covariance is maximal if the filters have the same preferred tuning $\theta_i = \theta_j$. For $\theta_i \neq \theta_j$, the same von-Mises dependency as discussed above will produce the limited-range structure of the correlation matrix shown in Figure 1c. In general, when averaging across all possible pairs of phases, the rectification in eq. (8) will set negatives to zero, hence preserving the limited-range structure (for orientation differences between 0 and 90 degrees).

The approximations in eq. (7) and (8) are only valid for large number of pixels, which allows the replacement of the discrete dot product by a continuous integral. For this reason we use the more accurate expressions (3) to generate the figures, except when stated otherwise. As can be seen in Figure 1b and 1c, the limited range structure of tuning curves and covariance is also valid for small number of pixels.

The model given by eq. (3) can also incorporate heterogeneous tuning by introducing a heterogeneous scaling of the filters. The latter can be made explicit by representing the filters as $\mathbf{F}_i = a_i\hat{\mathbf{F}}_i$, where $a_i > 0$ is a random gain and $|\hat{\mathbf{F}}_i|^2 = 1$ is normalized.

2. Linear Fisher information in the input and the neural population

We would like to address how much information is contained in the neural population $\mathbf{r}$ compared to the image $\tilde{\mathbf{I}}$. Due to the input noise in the retina and the finite number of pixels, the Fisher information about orientation θ in the image $\tilde{\mathbf{I}}$ is limited. It is given by

$$FI_{input}(\theta) = \frac{1}{\sigma_0^2} |\, \mathbf{I}'(\theta) \,|^2 , \tag{9}$$

where $' = d/d\theta$ is the derivative with respect to orientation.

To analyze the information in the neural population, we will firstly consider deterministic neurons without Poisson step; instead of eq. (2) the neural response is given by $r_i = [\mathbf{F}_i \cdot \tilde{I}]_+$. In Figure 1d in the main text we illustrate that the Poisson noise can be averaged out if the neural population is large enough. The tuning curves of deterministic neurons is the same as (3), but the covariance matrix is only given by the first term in (3)

$$\Sigma_{ij} = Cov(r_i, r_j) \approx \sigma_0^2 [\mathbf{F}_i \cdot \mathbf{F}_j]_+ .$$

The linear Fisher information in the neural population is therefore given by

$$FI_{neural} = \sum_{ij} f_i'(\theta) \Sigma_{ij}^{-1} f_i'(\theta) = \frac{1}{\sigma_0^2} \mathbf{I}'(\theta) \cdot \sum_{ij} \mathbf{F}_i (\mathbf{F} \cdot \mathbf{F})_{ij}^{-1} \mathbf{F}_j \cdot \mathbf{I}'(\theta) . \tag{10}$$

The orthogonal projection matrix $P_{\mathbf{F}} = \sum_{ij} \mathbf{F}_i (\mathbf{F} \cdot \mathbf{F})_{ij}^{-1} \mathbf{F}_j$ is acting in the P-dimensional image space and projects to the vector space spanned by the filters. In particular for any vector $\mathbf{v} \in R^P$, $P_{\mathbf{F}} \mathbf{v} = \mathbf{v}$ if and only if the vector can be written as a linear combination of the filters.

Comparing eq. (9) and eq. (10) we can derive the optimality condition for the neural population. The neural Fisher information preserves all input information, if and only if

$$\mathbf{I}'(\theta) \in span_i \{\mathbf{F}_i\} , \tag{11}$$

that is if the derivative of the image with respect to the stimulus can be written as a linear combination of the filters. In the suboptimal case we obtain

$$FI_{neural} = FI_{input} \cos^2 \alpha , \tag{12}$$

where α is the principal angle between $\mathbf{I}'(\theta)$ and the vector space $span_i \{\mathbf{F}_i\}$.

Given an image $\mathbf{I}(\theta)$ parameterized by a stimulus like orientation, can we find a set of N linear filters $\mathbf{F}_i$ such that the optimality condition in eq. (11) is satisfied for all θ? One way to do this is to approximate $\mathbf{I}(\theta)$ in a piecewise linear manner by tiling the stimulus domain equally,

$0 \le \theta_1 < \ldots < \theta_i < \ldots < \theta_N < 2\pi$, and setting the filters to be proportional to the image at the respective stimulus value,

$$\mathbf{F}_i \propto \mathbf{I}(\theta_i)\,. \tag{13}$$

Given θ with $\theta_i < \theta < \theta_{i+1}$, one can approximate

$$\mathbf{I}'(\theta) \approx \frac{\mathbf{I}(\theta_{i+1}) - \mathbf{I}(\theta_i)}{\theta_{i+1} - \theta_i} = c_{i+1}\mathbf{F}_{i+1} - c_i\mathbf{F}_i\,,$$

where c_i and c_{i+1} are some constants of proportionality. For large N this approximation becomes arbitrarily good and eq. (11) is satisfied for all θ.

In the example of the Gabor image in eq. (6) and the Gabor filter in eq. (5) with identical envelope and spatial frequency the condition in eq. (13) is satisfied and we expect that deterministic neurons retrieve all input information.

Note also that the analysis leading to the optimality condition in eq. (11) can be generalized to the case of non-linear deterministic neurons in the limit of small noise variance σ_0^2. In this case the filters are given by the local gradient of the neural response with respect to the input image,

$$\mathbf{F}_i \propto \nabla_{\mathbf{I}} r_i(\mathbf{I}(\theta))\,,$$

and the mapping is optimal if the derivative of the image with respect to the stimulus lies in the tangent space of the neural response hypersurface,

$$\mathbf{I}'(\theta) \in span_i\{\nabla_{\mathbf{I}} r_i\}$$

3. Tuning curves for suboptimal filters

Here we consider a case where the bank of filters has a different spatial frequency than the image and the optimality condition in eq. (11) is not satisfied. We will show that if the spatial wavelength of the filters is smaller than the image, the tuning curves in this suboptimal population will be narrower than the optimal tuning curve. This is unlike the case of independent neuronal noise, where narrower tuning always leads to higher information.

Let us define Gabor filters and image with different spatial wavelengths λ_F and λ_I,

$$F_{(x,y),i} \propto \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)\cos\left(\frac{2\pi}{\lambda_F}(x\cos\theta_i + y\sin\theta_i) + \phi_i\right), \tag{14}$$

$$I_{(x,y)}(\theta) \propto \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)\cos\left(\frac{2\pi}{\lambda_I}(x\cos\theta + y\sin\theta) + \phi\right) \tag{15}$$

As in (7), we can calculate the tuning curves by replacing the dot product by a continuous integral. For simplicity we also assume $\phi_i = \phi = \frac{\pi}{4}$, a different choice of phases does not affect the tuning curve width. The tuning curves are given by

$$f_i(\theta) \propto \frac{\pi\sigma^2}{2} \exp\left(-\sigma^2\pi^2 \left(\frac{1}{\lambda_I{}^2} + \frac{1}{\lambda_F{}^2}\right)\right) \exp\left(2\frac{\sigma^2\pi^2}{\lambda_I \lambda_F} \cos(\theta_i - \theta)\right)$$

In particular, if the spatial wavelength of the filters is smaller than that of the image, $\lambda_F < \lambda_I$, the tuning curves will be sharper than for the optimal value $\lambda_F = \lambda_I$.

4. Stimulus dependence and contrast dependence of correlations

Consider again the model (3) in the case of heterogeneous tuning. As mentioned at the end of section 1, we can make the heterogeneity explicit by representing the filters as $\mathbf{F}_i = a_i \hat{\mathbf{F}}_i$, where $a_i > 0$ is a random gain and $|\hat{\mathbf{F}}_i|^2 = 1$ is normalized. We can then re-express the tuning curves and covariance matrix as

$$f_i(\theta) = a_i [\hat{\mathbf{F}}_i \cdot \mathbf{I}(\theta)]_+$$

$$\Sigma_{ij}(\theta) = \sigma_0^2 a_i a_j [\hat{\mathbf{F}}_i \cdot \hat{\mathbf{F}}_j]_+ + \delta_{ij} f_i(\theta)$$

The correlation coefficient between units $i \neq j$ is:

$$\rho_{ij}(\theta) = \frac{\sigma_0^2 a_i a_j [\hat{\mathbf{F}}_i \cdot \hat{\mathbf{F}}_j]_+}{\sqrt{\sigma_0^2 a_i^2 + f_i(\theta)}\sqrt{\sigma_0^2 a_j^2 + f_j(\theta)}} \tag{16}$$

From Equation (16) it is immediately clear that the correlation coefficient is stimulus dependent, due to the terms in the denominator. In particular, the stronger the stimulus θ drives the two neurons, the lower the correlation coefficient drops; in the limit $\sigma_0 \to 0$, the correlation is inversely proportional to the geometric mean of the firing rate of the neural pair. Furthermore, since changing the contrast of the input image $\mathbf{I}(\theta) \to c\mathbf{I}(\theta)$ linearly rescales the tuning curves by the same factor $f_i(\theta) \to c f_i(\theta)$, the correlation coefficient is also inversely proportional to contrast.

These results are somewhat at odds with experimental data: in primary visual cortex, noise correlations on long time scales (several hundreds of milliseconds) decrease with contrast but are found to be stimulus independent; on short time scales (tens of milliseconds), they increase with both contrast and stimulus drive. We will show in the next section that these failures of the linear model can be addressed by considering a simple nonlinearity.

Also, note that the correlations in eq. (16) increase with input noise level σ_0.

5. Static nonlinearity: squaring

Let us examine now the correlations in a population of nonlinear neurons, described by Gabor filters as above, followed by squaring and additive Poisson-like noise. To distinguish the responses, tuning curves, and covariances of the linear model from those of the nonlinear model, we will use the apexes $(\cdot)^{(L)}$ and $(\cdot)^{(NL)}$, respectively. The response to an arbitrary stimulus θ_k is:

$$r_i^{(NL)}(\theta_k) = \left(r_i^{(L)}(\theta_k)\right)^2 \tag{17}$$

and, from $\mathbf{r}^{(L)}(\theta_k) \sim \mathcal{N}\left(\mathbf{f}^{(L)}(\theta_k), \Sigma^{(L)}\right)$ and straightforward calculations, it follows that the tuning curves and covariance of the nonlinear model before the Poisson step are:

$$f_i^{(NL)}(\theta_k) \propto \left\langle \left(r_i^{(L)}(\theta_k)\right)^2 \right\rangle = \left(f_i^{(L)}(\theta_k)\right)^2 + \Sigma_{ii}{}^{(L)} \tag{18}$$

$$\begin{aligned} \Sigma_{ij}^{(NL)}(\theta_k) &\propto \left\langle \left(r_i^{(L)}(\theta_k)\right)^2 \left(r_j^{(L)}(\theta_k)\right)^2 \right\rangle - \left\langle \left(r_i^{(L)}(\theta_k)\right)^2 \right\rangle \left\langle \left(r_j^{(L)}(\theta_k)\right)^2 \right\rangle \\ &= 4\Sigma_{ij}{}^{(L)} f_i^{(L)}(\theta_k) f_j^{(L)}(\theta_k) + 2\left(\Sigma_{ij}{}^{(L)}\right)^2 \end{aligned} \tag{19}$$

After the addition of Poisson-like, independent noise, the covariance and correlations become:

$$\Sigma_{ij}^{(NL)}(\theta_k) = 4\Sigma_{ij}{}^{(L)} f_i^{(L)}(\theta_k) f_j^{(L)}(\theta_k) + 2\left(\Sigma_{ij}{}^{(L)}\right)^2 + \left[\left(f_i^{(L)}(\theta_k)\right)^2 + \Sigma_{ii}{}^{(L)}\right]\delta_{ij} \tag{20}$$

$$\rho_{ij}^{(NL)}(\theta_k) = \frac{4\Sigma_{ij}{}^{(L)} f_i^{(L)}(\theta_k) f_j^{(L)}(\theta_k) + 2\left(\Sigma_{ij}{}^{(L)}\right)^2}{\sqrt{\left(4\Sigma_{ii}{}^{(L)}+1\right)\left(f_i^{(L)}(\theta_k)\right)^2 + \left(2\Sigma_{ii}{}^{(L)}+1\right)\Sigma_{ii}{}^{(L)}}\sqrt{\left(4\Sigma_{jj}{}^{(L)}+1\right)\left(f_j^{(L)}(\theta_k)\right)^2 + \left(2\Sigma_{jj}{}^{(L)}+1\right)\Sigma_{jj}{}^{(L)}}} \tag{21}$$

Let us examine first the contrast dependence of the correlation coefficient. Remember that contrast simply rescales the tuning curves of the linear model: $f_i(\theta_k) \to c f_i(\theta_k)$. We can therefore rewrite more synthetically:

$$\rho_{ij}^{(NL)}(\theta_k, c) = \frac{A_{ij}c^2 + B_{ij}}{\sqrt{D_i c^2 + E_i}\sqrt{D_j c^2 + E_j}} = \begin{cases} \dfrac{B_{ij}}{\sqrt{E_i E_j}} & \text{for} \quad c \to 0 \\ \dfrac{A_{ij}}{\sqrt{D_i D_j}} & \text{for} \quad c \to \infty \end{cases} \tag{22}$$

There is therefore a monotonic relation between correlation and contrast, and it is easy to show that $\rho_{ij}^{(NL)}(\theta_k, c=0) < \rho_{ij}^{(NL)}(\theta_k, c \to \infty)$. A similar relation holds also for the dependence of the correlation coefficient on the mean rate of the pair, and both results are qualitatively in agreement with experimental data on noise correlations in cortex at short time scales.

We further consider the low input noise (or high contrast) regime, i.e. $f_i^{(L)} f_j^{(L)} \propto \Sigma_{ij}^{(L)}$, in which the covariance matrix becomes:

$$\Sigma_{ij}^{(NL)}\left(\theta_k\right) \approx 4\Sigma_{ij}{}^{(L)} f_i^{(L)}\left(\theta_k\right) f_j^{(L)}\left(\theta_k\right) + \left(f_i^{(L)}\left(\theta_k\right)\right)^2 \delta_{ij} \tag{23}$$

The correlation coefficient becomes stimulus and contrast-independent:

$$\rho_{ij}^{(NL)} \approx \frac{4\Sigma_{ij}{}^{(L)}}{\sqrt{4\Sigma_{ii}{}^{(L)}+1}\sqrt{4\Sigma_{jj}{}^{(L)}+1}} \tag{24}$$

The quadratic model also has advantages for decoding. Generally, a locally linear estimator in a fine discrimination task around θ_0 is an estimator of the form

$$\hat{\theta} = \mathbf{w}\cdot(\mathbf{r} - \mathbf{f}(\theta_0))$$

The optimal decoding weights with minimal variance for an unbiased estimator are given by (1, 2)

$$\mathbf{w}_{opt} = \frac{\Sigma^{-1}\mathbf{f}'}{\mathbf{f}'\Sigma^{-1}\mathbf{f}'} \tag{25}$$

In general, the decoding weights in eq. (25) depend on contrast through both tuning curves and covariance matrix. If contrast fluctuates over time, as is the case in natural circumstances, the decoding weights need to be adapted, for example by estimating contrast on the fly.

In the quadratic model at low noise, the covariance matrix in eq. (23) scales homogeneously with contrast, as do the tuning curves. The optimal decoding weights in eq. (25) only change by a constant of proportionality, but the relative weights do not change. If contrast fluctuates from trial to trial, as is the case in natural circumstances, the same relative decoding weights can be used without estimating contrast.

Moreover, the scaling of the covariance matrix and tuning curves are such that the decoding weights in the normalization

$$\mathbf{w}_{norm} = \Sigma^{-1}\mathbf{f}'$$

are contrast-invariant. This implies that the feedforward model with quadratic nonlinearity implements a linear probabilistic population code; such a code is particularly convenient to not only decode the stimulus but also the confidence about the stimulus, even if contrast fluctuates from trial to trial (3).

The same is however not true for the input noise level σ_0. The covariance matrix of both linear and quadratic model scale inhomogeneously with σ_0, and therefore the normalized decoding weights need to be adapted to a given level of input noise (Figure 4b,c)

6. Effects of internal global fluctuations on correlations and information

We consider now an extension of the feedforward network described above, that captures the doubly stochastic nature of response variability observed across the visual cortex (4), particularly under anesthesia (5). This has been previously described as a fluctuating gain, which is partly shared between neurons. The gain has a multiplicative effect on the deterministic, stimulus-driven part of the response, and the result of the multiplication sets the rate of the Poisson spike generator. Here we extend this model by assuming that the stimulus-driven response is produced according to our feedforward model, and is therefore itself variable. For simplicity, we assume just one gain factor that is shared between all neurons. The corresponding probabilistic graphical model is as follows:

$$\begin{aligned} &\mathbf{r} \sim \mathcal{N}\left(\mathbf{f}, \Sigma\right) \\ &g \sim \Gamma\left(\alpha, \beta\right) \quad ; \quad \left\langle g \right\rangle = 1 \\ &\mathbf{y} = g\mathbf{r} \\ &\mathbf{R} \sim \text{Poisson}\left(\mathbf{y}\right) \end{aligned} \tag{26}$$

where Σ denotes the covariance of the feedforward model (with or without nonlinearity), g is the gain factor and $\mathbf{R}$ denotes the vector of spike counts.

First, it is easy to show under mild assumptions that the tuning curves for this model remain unchanged w.r.t. the feedforward model, i.e.

$$\left\langle \mathbf{R} \right\rangle = \mathbf{f} \tag{27}$$

This follows simply from the assumption that $\mathbf{r}$ and g are independent and $\left\langle g \right\rangle = \alpha / \beta = 1$, hence $\left\langle \mathbf{y} \right\rangle = \left\langle \mathbf{r} \right\rangle$.

Second, we would like to compute the noise covariance matrix in this model, which we will denote by $\Sigma_{\mathbf{R}}$. We start by computing the covariance $\Sigma_{\mathbf{y}}$ of the intermediate rate $\mathbf{y}$:

$$\begin{aligned} \Sigma_{\mathbf{y},ij} &= \left\langle y_i y_j \right\rangle - \left\langle y_i \right\rangle \left\langle y_j \right\rangle \\ &= \left\langle g^2 \right\rangle \left\langle r_i r_j \right\rangle - \left\langle g \right\rangle^2 \left\langle r_i \right\rangle \left\langle r_j \right\rangle \\ &= \left(\sigma_g^2 + 1\right) \left\langle r_i r_j \right\rangle - \left\langle r_i \right\rangle \left\langle r_j \right\rangle \\ &= \left(\sigma_g^2 + 1\right) \Sigma_{ij} + \sigma_g^2 \left\langle r_i \right\rangle \left\langle r_j \right\rangle \end{aligned} \tag{28}$$

where $\sigma_g^2 = \alpha / \beta^2$ is the variance of the gain distribution. It follows that:

$$\Sigma_{\mathbf{y}} = \left(\sigma_g^2 + 1\right) \Sigma + \sigma_g^2 \mathbf{f}\mathbf{f}^\top \tag{29}$$

and, after the Poisson step:

$$\Sigma_{\mathbf{R}} = \left(\sigma_g^2 + 1\right) \Sigma + \sigma_g^2 \mathbf{f}\mathbf{f}^\top + \text{diag}\left(\mathbf{f}\right) \tag{30}$$

where $\text{diag}(\mathbf{x})$ denotes the diagonal matrix with diagonal entries given by the elements of $\mathbf{x}$.

Eq. (30) shows that global fluctuations affect the noise covariance in two ways: 1) by rescaling the amplitude of the feedforward covariance (first term on the r.h.s); and 2) by adding a rank-1 perturbation (second term on the r.h.s). What are the effects on noise correlations? To gain some insight, we expand the correlation between two neurons for the feedforward linear case (assuming, for readability, a homogeneous population with $a_i = 1$):

$$\begin{aligned}\rho_{\mathbf{R},ij} &= \frac{\Sigma_{\mathbf{R},ij}}{\sqrt{\Sigma_{\mathbf{R},ii}\Sigma_{\mathbf{R},jj}}} \\ &= \frac{\left(\sigma_g^2+1\right)\sigma_0^2\mathbf{F}_i\cdot\mathbf{F}_j}{\sqrt{\left(\sigma_g^2+1\right)\sigma_0^2+\sigma_g^2\left(\mathbf{F}_i\cdot\mathbf{I}\right)^2+\left(\mathbf{F}_i\cdot\mathbf{I}\right)}\sqrt{\left(\sigma_g^2+1\right)\sigma_0^2+\sigma_g^2\left(\mathbf{F}_j\cdot\mathbf{I}\right)^2+\left(\mathbf{F}_j\cdot\mathbf{I}\right)}} \\ &+ \frac{\sigma_g^2\left(\mathbf{F}_i\cdot\mathbf{I}\right)\left(\mathbf{F}_j\cdot\mathbf{I}\right)}{\sqrt{\left(\sigma_g^2+1\right)\sigma_0^2+\sigma_g^2\left(\mathbf{F}_i\cdot\mathbf{I}\right)^2+\left(\mathbf{F}_i\cdot\mathbf{I}\right)}\sqrt{\left(\sigma_g^2+1\right)\sigma_0^2+\sigma_g^2\left(\mathbf{F}_j\cdot\mathbf{I}\right)^2+\left(\mathbf{F}_j\cdot\mathbf{I}\right)}}\end{aligned} \tag{31}$$

This decomposition highlights the interplay between peripheral noise and global fluctuations, in shaping noise correlations. The first term in the last line displays a structure of noise correlation similar to the purely feedforward model, i.e. where correlations are determined by filter overlap (numerator). However the magnitude of such correlations is influenced by the magnitude of both the feedforward, σ_0^2, and internal, σ_g^2, parts: increasing either type of fluctuations increases the magnitude of noise correlations. The second term in the last line of eq. (31) has a different behavior. First, the numerator itself is highly stimulus-dependent, being determined by the overall response strength of the pair (product of individual responses) to the stimulus (4). Second, while the magnitude of this contribution to correlations increases with increasing global fluctuations, it can only decrease with increasing peripheral noise. Therefore, overall the interplay of peripheral noise and global fluctuations leads to the somewhat counterintuitive result that larger peripheral noise leads to larger noise correlations for some pairs, but smaller for others, depending on the relative size of the first and second term in eq. (31).

Given the predominance of global fluctuations in visual cortex, it is important to understand their impact on the information encoded in a large population. Intuitively, a global scaling of the population response affects only the height of the population hill of activity, leaving the location of the peak (and therefore the stimulus value extracted by a linear readout) unchanged. Therefore, one could expect that global fluctuations do not change information. We show here that this intuition is only partly correct: Global fluctuations do not, by themselves, limit information in a large population, but they do further decrease the asymptotic value when information is already limited in the input.

First, we show that global fluctuations do not limit information in large populations. Suppose we started from a population with tuning and covariance such that $\mathbf{f}'^{\top}\Sigma^{-1}\mathbf{f}'$ is O(N) (i.e. it grows indefinitely as we increase the population size). Now remember that in eq. (29) global fluctuations have two effects. First, they rescale the covariance by a factor that only depends on the amplitude of the fluctuations (and therefore does not change with population size). Second, they add a rank-1 perturbation proportional to $\mathbf{f}\mathbf{f}^{\top}$. As shown by Moreno et al. (6), a rank-1 perturbation only limits information if it is parallel to $\mathbf{f}'\mathbf{f}'^{\top}$ For $\mathbf{f}\mathbf{f}^{\top}$, this is the case only if the stimulus is encoded in the

amplitude of the population response; or, more precisely, if $\mathbf{f}(\theta)$ defines a zero-curvature line when θ is varied (e.g. $f_i = a_i\theta$). Therefore, global fluctuations do not limit information in general (and about orientation, in particular).

Nonetheless, we would like to assess more precisely the effect of global fluctuations on information, i.e. if there is any information loss, and how big. We start again from eq. (29). To compute linear Fisher information we need to invert $\Sigma_{\mathbf{y}}$, which can be done using the Morrison-Sherman lemma:

$$\begin{aligned}\Sigma_{\mathbf{y}}^{-1} &= \frac{1}{\sigma_g^2+1}\left(\frac{\Sigma_{\mathbf{y}}}{\sigma_g^2+1}\right)^{-1} \\ &= \frac{1}{\sigma_g^2+1}\left[\Sigma^{-1} - \frac{\sigma_g^2}{\sigma_g^2+1}\frac{\Sigma^{-1}\left(\mathbf{f}\mathbf{f}^\top\right)\Sigma^{-1}}{1+\frac{\sigma_g^2}{\sigma_g^2+1}\mathbf{f}\Sigma^{-1}\mathbf{f}^\top}\right]\end{aligned} \tag{32}$$

From this it follows that the linear Fisher information in the network with global fluctuations is:

$$\begin{aligned}FI_{\mathbf{y}} &= \mathbf{f}'^\top \Sigma_{\mathbf{y}}^{-1}\mathbf{f}' \\ &= \frac{1}{\sigma_g^2+1}\mathbf{f}'^\top \Sigma^{-1}\mathbf{f}' \\ &\quad -\frac{\sigma_g^2}{\sigma_g^2+1}\frac{\left(\mathbf{f}'^\top \Sigma^{-1}\mathbf{f}\right)\left(\mathbf{f}^\top\Sigma^{-1}\mathbf{f}'\right)}{1+\frac{\sigma_g^2}{\sigma_g^2+1}\mathbf{f}\Sigma^{-1}\mathbf{f}^\top}\end{aligned} \tag{33}$$

From this decomposition it is obvious that, whenever $\sigma_g^2 > 0$, information is reduced with respect to the purely feedforward network. The first term in the second line of eq. (33) coincides with the feedforward information, rescaled by a factor smaller than 1. Furthermore, the second term is always negative, and therefore further reduces information.

We conclude therefore that global fluctuations are not responsible for limiting information in large populations, but nonetheless they interact with feedforward-induced variability in a way that substantially affects the structure of noise correlations, and that reduces the information in the population.

7. The size of differential correlations

Here we show that any information-limiting covariance matrix can be split into a positive definite non-information-limiting part and an information-limiting part and analyze the size of differential correlations.

Let Σ be an information-limiting covariance matrix, that is the information saturates to a finite value, $\mathbf{f}'\cdot\Sigma^{-1}\mathbf{f}' \to FI_\infty$ as $N \to \infty$. To find the size of differential correlations, we would like to calculate the maximum amount of differential correlations we can subtract such that the remaining matrix is still positive definite. Defining

$$\Sigma_\epsilon = \Sigma - \epsilon \mathbf{f}'\mathbf{f}'^T,$$

we are looking for the maximum ϵ such that Σ_ϵ is still positive definite for all N.

Due to the matrix determinant lemma, the determinant of Σ_ϵ is given by

$$\det \Sigma_\epsilon = \left(1 - \epsilon n\, \mathbf{f}' \cdot \Sigma^{-1} \mathbf{f}'\right) \det(\Sigma). \tag{34}$$

For fixed N and $0 \le \epsilon < \frac{1}{\mathbf{f}' \cdot \Sigma^{-1} \mathbf{f}'}$, the determinant of Σ_ϵ is strictly positive. We would like to show that Σ_ϵ is also positive definite in that interval, or equivalently, that all eigenvalues are strictly positive. By assumption, we know that Σ_ϵ is positive definite for $\epsilon = 0$. If at least one eigenvalue has negative sign for some value of $\epsilon_1 \in [0, \frac{1}{\mathbf{f}' \cdot \Sigma^{-1} \mathbf{f}'}[$, the intermediate value theorem implies that there must be an ϵ_2, $0 \le \epsilon_2 < \epsilon_1 < \frac{1}{\mathbf{f}' \cdot \Sigma^{-1} \mathbf{f}'}$, such that the same eigenvalue is exactly zero, resulting in a zero determinant. This contradicts with the result above that the determinant is strictly positive for all $0 \le \epsilon < \frac{1}{\mathbf{f}' \cdot \Sigma^{-1} \mathbf{f}'}$. This implies that Σ_ϵ is strictly positive definite in this interval and only becomes positive semi-definite for $\epsilon = \frac{1}{\mathbf{f}' \cdot \Sigma^{-1} \mathbf{f}'}$.

Consequently, the maximum ϵ across N is given by

$$\epsilon_{max} = \frac{1}{sup_N \mathbf{f}' \cdot \Sigma^{-1} \mathbf{f}'} = \frac{1}{FI_\infty}, \tag{35}$$

where the last equality holds if the information as a function of N is well approximated by a monotonously increasing function, as is the case in most examples in the literature and in particular in our network (Figure 1d).

How large is the contribution of the differential correlations, $\epsilon_{max} \mathbf{f}'\mathbf{f}'^T$, to an element of the covariance matrix Σ? To get an intuition consider the case of homogeneous Gaussian tuning curves of width σ_t and a general stimulus parameter s. The contribution of the differential correlations is

$$\epsilon_{max} f_i'(s) f_j'(s) = \frac{\epsilon_{max}}{\sigma_t^2} \frac{(s - s_i)(s - s_j)}{\sigma_t^2} \exp\left(-\frac{1}{2\sigma_t^2} ((s - s_i)^2 + (s - s_j)^2) \right) \tag{36}$$

The scale of this contribution is set by the fraction $\frac{\epsilon_{max}}{\sigma_t^2}$. If the behavioral readout of the neural population is close to optimal, eq. (35) implies that $\epsilon_{max} \approx \theta_P^2$, where θ_P is the psychophysical threshold. This implies that the contribution of the differential correlations will be small whenever the psychophysical threshold is smaller than the tuning curve width.

As an example consider orientation discrimination at high contrast, where psychophysical thresholds are typically around $\theta_P = 2^\circ$ (7). In V1, tuning curve widths are typically around $\sigma_t = 20^\circ$ or larger (8). In this case the value of the prefactor is

$$\frac{\epsilon_{max}}{\sigma_t^2} = \frac{\theta_P^2}{\sigma_t^2} = 0.01,$$

giving rise to very small differential fluctuations. This explains why subtracting differential correlations in Figure 6 has such a small effect on the correlation structure.

8. Optimality of human subjects in orientation discrimination

Here we show that the thresholds for orientation discrimination reported for human subjects in perceptual learning studies (7, 9) are extremely close to the ideal observer. To do so, we will first compute the input information, under some simple assumptions. Second, we will explain the metric used in the psychophysics experiments, namely the threshold versus external noise contrast (TVC) curve; and we will show how to convert input information to a corresponding TVC curve. This will allow us to compare the TVC based on the input information (i.e., the ideal observer performance) to that measured experimentally.

The information in the input images, FI_{input}, can be computed using eq. (9). To do so, we need to estimate the parameters defining the images used in the experiment, and the amount of noise corrupting peripheral signals, which we denoted σ_0^2. In (7), Gabor images corrupted by white noise were used. The Gabor parameters are given in the table below.

Parameters of the Gabor images used in (7)		
Symbol	**Meaning**	**Value**
P	Side length (pixels)	64 (1.54 degrees)
θ	Orientation (degrees)	$[-12,12]$
σ	Gaussian envelope std (degrees)	0.385
ω	Preferred spatial frequency (cycles/degree)	2.3
σ_{ext}	External noise std (proportion of pixel range)	$[0, 0.02, 0.04, 0.08, 0.16, 0.25, 0.33]$
c	Contrast	Variable, see text

The images were presented parafoveally in the lower right visual field (3.1 degrees horizontally and 2.3 degrees vertically). At those eccentricities, the high spatial frequency cutoff for most V1 neurons is below 8 cycles/degree (10). This suggests that, effectively, the input to parafoveal V1 corresponds to approximately 16 independent pixels per degree. Therefore, before computing input information, we scaled the images to a size of 25 pixels (16 pixels/degree times 1.54 degrees), and similarly transformed ω and σ assuming 16 pixels/degree. To test the robustness to this assumption, we also considered other resolutions, namely 12 and 20 pixels/degree (Figure 9b). We computed the image derivative numerically, rather than using analytical derivative of eq. (6), because of the small

number of pixels. We did so by taking the difference between the images at the two test orientations (-12, 12 degrees), and dividing by the orientation difference. Lastly, to determine the level of the noise injected at the periphery, σ_0, we used the inflection point of the experimental TVC curves (11, 12). We varied σ_0 between 0.04 and 0.08 (Figure 9c).

We now specify the relation between input information and ideal observer TVC curves. In a typical discrimination task the subject is asked to distinguish two similar stimuli, $\theta^+ = \theta + d\theta$ and $\theta^- = \theta - d\theta$. Linear Fisher information is the inverse of the variance of the estimate of the stimulus based on the optimal linear decoder. Decoder performance, quantified by the proportion of correctly classified trials, and Fisher information are directly related. Specifically, proportion correct is given by $P = \Phi\left(\frac{d\theta\sqrt{FI}}{2}\right)$, where Φ is the Normal cumulative distribution function. During the experiments of (7), for each level of external noise, contrast was varied until subjects reached a prespecified performance level (79.3% correct). We did the same for the ideal observer, namely, for each noise level, we searched for the signal contrast level such that the input information corresponded to the desired percent correct, and so built a TVC curve for the ideal observer (Figure 9bc).

Comparison to the performance of the human subjects (Figure 9a) shows that, after perceptual learning, they were extremely close to the ideal observer.

9. Simulation details

The tables below provide implementation details and parameter values for the simulations used to generate the figures.

MODEL	ACRONYM	FIGURE USED
Linear-rectified-Poisson	LRP	1-6, 8; S1-S4
LRP with heterogeneous filters	LRPH	2, 5
LRP with suboptimal filters	LRPS	3
Linear-quadratic-Poisson	LQP	4; S2
LRP with global fluctuations	LRGP	5
LRP with global fluctuations, with heterogeneous filters	LRGPH	5
LRP minus differential correlations	LRPM	6, 8; S4
Synthetic tuning curves	SYNTH	7
LRP with larger filters	LRPV4	S1

Table 1 For each model variant, we provide in this table its acronym and the list of the figures based on that model.

Figures 1, 3, 6, 8, S2, S3, S4		
INPUT IMAGES		
Symbol	**Meaning**	**Value**
P	Side length (pixels)	12
θ	Orientation (degrees)	0
σ	Gaussian envelope std (degrees)	$P/3$
λ	Preferred spatial wavelength (pixel/cycle)	$P/1.5$
ϕ	Preferred spatial phase	0
c	Michelson contrast	1
σ_0	Input noise std (proportion of pixel range)	0.2
FILTERS		
Symbol	**Meaning**	**Value**
N	Number of neurons	$[10, 20, 50, 100, 200, 500, 1000, 2000, 10000]$
P	RF side length (pixels)	12
θ	Preferred orientation (degrees)	$\left[-180 : \frac{360}{N} : \frac{180(N-1)}{N}\right]$
σ	Gaussian envelope std (pixels)	$P/3$
λ (LRP, LRPM)	Preferred spatial wavelength (pixel/cycle)	$P/1.5$
λ (LRPS)	Preferred spatial wavelength (pixel/cycle)	$P/2$
ϕ	Preferred spatial phase	0
g_k	Tuning amplitude	$g_k = g a_k$ $g = 20$ $a_k \sim LogNormal\left(\mu = 0, \sigma^2 = 0.25\right)$

Table 2 Parameters of the orientation network and its inputs, used in Figures 1,3,6,8, S2, S3, S4. Simulations are repeated 50 times for each population size, then information values are averaged.

Figure 2,5

INPUT IMAGES		
Symbol	**Meaning**	**Value**
P	Side length (pixels)	48
θ	Orientation (degrees)	0
σ	Gaussian envelope std (degrees)	$P/12$
λ	Preferred spatial wavelength (pixel/cycle)	$P/6$
ϕ	Preferred spatial phase	0
c	Michelson contrast	1
σ_0	Input noise std (proportion of pixel range)	0.2
FILTERS		
Symbol	**Meaning**	**Value**
N	Number of neurons	$[10, 20, 50, 100, 200, 500, 1000, 2000, 10000]$
P	RF side length (pixels)	48
θ	Preferred orientation (degrees)	$\left[-180 : \frac{360}{N} : \frac{180(N-1)}{N}\right]$
(x_c, y_c) (LRP, LRGP)	Receptive field center coordinates	$x_c = 0 \quad ; \quad y_c = 0$
λ (LRP, LRGP)	Preferred spatial wavelength (pixel/cycle)	$P/6$
σ (LRP, LRGP)	Gaussian envelope std (pixels)	$P/12$
(x_c, y_c) (LRPH, LRGPH)	Receptive field center coordinates	$x_c, y_c \sim \text{unif}(-1,1)$
λ (LRPH, LRGPH)	Preferred spatial wavelength (pixel/cycle)	$\lambda = P/4\omega$ $\omega \sim LogNormal(\mu = \log(2), \sigma^2 = 0.25)$
σ (LRPH, LRGPH)	Gaussian envelope std (pixels)	$2\lambda/3$
ϕ	Preferred spatial phase	0
g_k	Tuning amplitude	$g_k = g a_k$ $g = 20$ $a_k \sim LogNormal(\mu = 0, \sigma^2 = 0.25)$
σ_g	Internal noise std	$0.03g$

Table 3 Parameters of the orientation network and its inputs, used in Figures 2 and 5. Simulations are repeated 50 times for each population size, then information values are averaged. In these simulations we used larger value for P to allow for the larger range of RF sizes and positions.

Figure 4		
INPUT IMAGES		
Symbol	**Meaning**	**Value**
P	Side length (pixels)	32
θ	Orientation (degrees)	0
σ	Gaussian envelope std (degrees)	$P/4.5$
λ	Preferred spatial wavelength (pixel/cycle)	$P/1.5$
ϕ	Preferred spatial phase	0
c	Michelson contrast	$[0.1, 0.2, 0.5, 1]$; c=0.2 in Figure 4c
σ_0	Input noise std (proportion of pixel range)	$\sigma_0 = 0.2$
σ_{ext}	External noise std (proportion of pixel range)	$[0, 0.05, 0.1, 0.2]$
FILTERS		
Symbol	**Meaning**	**Value**
N	Number of neurons	100
P	RF side length (pixels)	32
θ	Preferred orientation (degrees)	$\left[-180 : \frac{360}{N} : \frac{180(N-1)}{N}\right]$
σ	Gaussian envelope std (pixels)	$P/4.5$
λ	Preferred spatial wavelength (pixel/cycle)	$P/1.5$
ϕ	Preferred spatial phase	0
g_k (LRP)	Tuning amplitude	$g_k = g a_k$ $g = 20$ $a_k \sim LogNormal\left(\mu = 0, \sigma^2 = 0.25\right)$
$g_k^{(Q)}$ (LQP)	Tuning amplitude	$g_k^{(Q)} = \sqrt{3 g_k}$

Table 4 Parameters of the orientation network and its inputs, used in Figure 4. In Figure 4b,c decoding weights are averaged over 100 realizations of the network with random gains. The gains for the LQP model are chosen to approximately match the tuning amplitude of the LRP model at a contrast of 0.5. In these simulations we used larger value for P to avoid artifacts in the decoding weights due to inversion of the covariance matrix.

Figure 7		
SYNTHETIC TUNING CURVES		
Symbol	**Meaning**	**Value**
N	Number of neurons	$10^{[0:0.5:6]}$
f_k	Tuning curve	$g_k = b\cos(\theta - \theta_k)$ $b = 20$
Σ_{kl}	Noise covariance matrix	$\Sigma_{kl} = (1-c)\delta_{kl} + c\cos(\theta_k - \theta_l)$ $c = 0.12$

Table 5 Parameters of the model with synthetic tuning curves, used in Figure 7. The model uses synthetic tuning curves and covariance matrix, as in (6).

Figure S1

INPUT IMAGES		
Symbol	**Meaning**	**Value**
P	Side length (pixels)	24
θ	Orientation (degrees)	0
σ	Gaussian envelope std (degrees)	$P/4.5$
λ	Preferred spatial wavelength (pixel/cycle)	$P/3$
ϕ	Preferred spatial phase	0
c	Michelson contrast	1
σ_0	Input noise std (proportion of pixel range)	0.2

FILTERS		
Symbol	**Meaning**	**Value**
N	Number of neurons	1000
P	RF side length (pixels)	24
θ	Preferred orientation (degrees)	0
σ (LRP)	Gaussian envelope std (pixels)	$P/9$
σ (LRPV4)	Gaussian envelope std (pixels)	$P/4.5$
λ	Preferred spatial wavelength (pixel/cycle)	$P/3$
(x_c, y_c)	Coordinates of RF center	$x_c = \left[\frac{P}{4} : \frac{P}{2N} : \frac{3P}{4}\right]$ $y_c = 0$
ϕ	Preferred spatial phase	0
g_k	Tuning amplitude	$g_k = g a_k$ $g = 20$ $a_k \sim LogNormal\left(\mu = 0, \sigma^2 = 0.25\right)$

Table 6 Parameters of the orientation network and its inputs, used in Figure S1.

10. Supplementary Figures

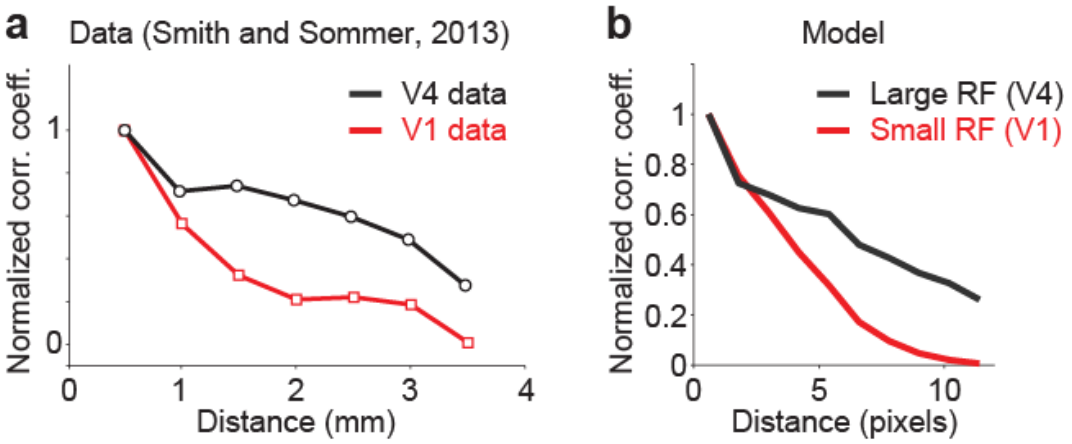

Figure S 1. Dependence of noise correlations on the filters spatial overlap. (**a**) Normalized noise correlations vs. cortical distance between electrode tips, in macaque V1 (red) and V4 (black). Reproduced after (13). The normalization is relative to the pairs with most overlapping receptive fields (RF), separately for each size. (**b**) Normalized noise correlations vs. center-to-center distance between filters, for small (red) and large (black) filters. Filter parameters are reported in Table S6.

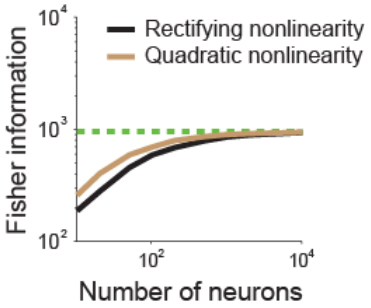

Figure S2. Effects of static nonlinearity on information. Linear Fisher information vs. population size for networks with half-rectification (black) or quadratic nonlinearity (brown). Parameters are set to the same values as Figure 1 (listed in Table S2), except for the neuronal response gain for the quadratic model, which is chosen to approximately match the tuning amplitude of the model with half-rectification at a contrast of 0.5.

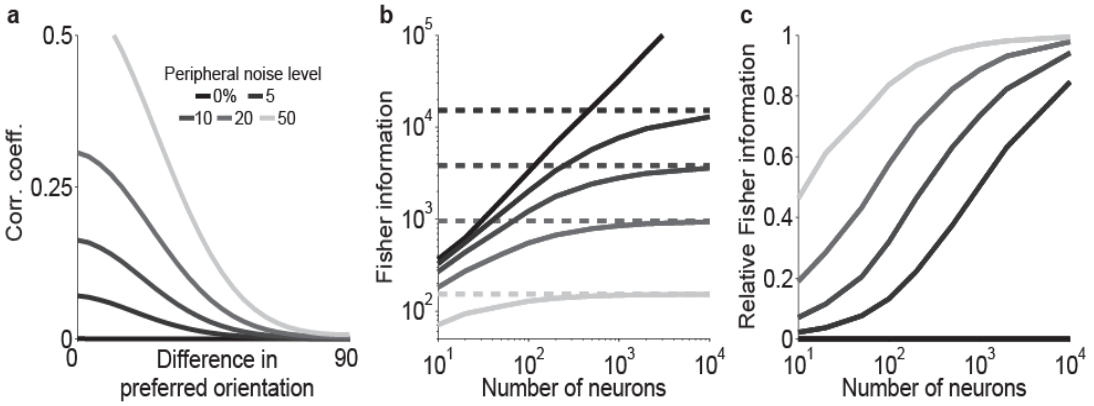

Figure S 3. Dependence of noise correlations and information on input noise level. (**a**) Noise correlations induced by feedforward connectivity increase with external noise level. (**b**) Information as a function of the number of neurons in the V1 layer (solid lines) and asymptotic information (dashed lines), for different levels of external noise. (**c**) Information in V1 populations of increasing size, relative to the asymptotic information. Population information does not simply scale with external noise level. At larger external noise, fewer neurons are required to achieve saturation. Parameters are set to the same values as Figure 1 (listed in Table S2), except for the peripheral noise level which is indicated in the legend in panel (a).

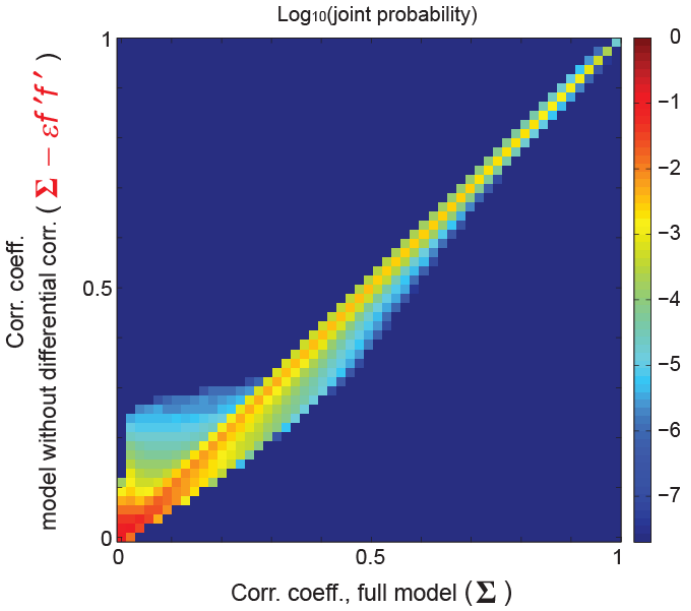

Figure S 4. Effects of removing differential correlations. Density plot of correlation coefficients across all pairs, before (abscissa) and after (ordinate) removing differential correlations. Simulations are based on the same parameters as main Figure 6, except N=10,000.

11. Supplementary references

1. Series P, Latham PE, & Pouget A (2004) Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations. *Nat Neurosci* 7(10):1129-1135.
2. Beck J, Bejjanki VR, & Pouget A (2011) Insights from a Simple Expression for Linear Fisher Information in a Recurrently Connected Population of Spiking Neurons. *Neural Comput* 23(6):1484-1502.
3. Ma WJ, Beck JM, Latham PE, & Pouget A (2006) Bayesian inference with probabilistic population codes. *Nat Neurosci* 9(11):1432-1438.
4. Goris RLT, Movshon JA, & Simoncelli EP (2014) Partitioning neuronal variability. *Nat Neurosci* 17(6):858-865.
5. Ecker Alexander S*, et al.* (2014) State Dependence of Noise Correlations in Macaque Primary Visual Cortex. *Neuron* 82(1):235-248.
6. Moreno-Bote R*, et al.* (2014) Information-limiting correlations. *Nat Neurosci* 17:1410–1417.
7. Dosher BA & Lu Z-L (1999) Mechanisms of perceptual learning. *Vision Res* 39(19):3197-3221.
8. Ringach DL, Shapley RM, & Hawken MJ (2002) Orientation Selectivity in Macaque V1: Diversity and Laminar Dependence. *The Journal of Neuroscience* 22(13):5639-5651.
9. Dosher BA & Lu Z-L (1998) Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proc Natl Acad Sci* 95(23):13988-13993.
10. Foster KH, Gaska JP, Nagler M, & Pollen DA (1985) Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. *J Physiol* 365(1):331-363.
11. Burgess A, Wagner R, Jennings R, & Barlow H (1981) Efficiency of human visual signal discrimination. *Science* 214(4516):93-94.
12. Lu Z-L & Dosher BA (2008) Characterizing observers using external noise and observer models: Assessing internal representations with external noise. *Psychological Review* 115(1):44-82.
13. Smith MA & Sommer MA (2013) Spatial and Temporal Scales of Neuronal Correlation in Visual Area V4. *J Neurosci* 33(12):5422-5432.