Supporting Online Material for

Decorrelated Neuronal Firing in Cortical Microcircuits

Alexander S. Ecker, Philipp Berens, Georgios A. Keliris, Matthias Bethge, Nikos K. Logothetis, Andreas S. Tolias*

*To whom correspondence should be addressed. E-mail: atolias@cns.bcm.edu

Published 29 January 2010, Science 327, 584 (2010)
DOI: 10.1126/science.1179867

This PDF file includes

Materials and Methods
SOM Text
Figs. S1 to S9
References
Supporting Online Material

Materials and Methods

Electrophysiological recordings
Recordings were made from three adult rhesus monkeys (*macaca mulatta*) weighing between 9.8 kg and 17 kg using chronically implanted arrays of 12 and 24 custom-build tetrodes. Tetrodes were custom-made out of either Nichrome or Platinum/Iridium wire (Fig. S1D, E) and electroplated with gold. Tips were cut at ~90° (Fig. S1C). The general methods, including implant design (Fig. S1A) and surgical procedures, have been described previously (24, p. 3781) for the 12-tetrode array, which was used in monkey D. In this array tetrodes are separated by approximately 200 µm (Fig. S1D). Data from the other two monkeys (B and H) were recorded in a different laboratory using a newer version of the drive that could hold 24 tetrodes spaced approximately 700 µm apart (Fig. S1B, E). The general procedures for the new drive followed the ones for the 12-tetrode drive. The main difference in the new setup was that data were recorded continuously using custom-build software (Siapas laboratory, Caltech, Pasadena, CA). The tetrode signals are preamplified immediately at the recording array by unity gain preamplifiers (HS-27, Neuralynx, Bozeman MT). Neural signals are digitized at 24 bits using analog acquisition cards with 30 dB of onboard gain (PXI-4498, National Instruments, Austin, TX) and recorded continuously at 32 KHz as broad-band signal (0.5 Hz to 16 kHz).

Spike detection and spike sorting
Spikes were detected offline when the signal on any of the four channels crossed a threshold of five times the noise amplitude on this channel. Spike sorting was done by fitting a mixture of Gaussians model to the detected and aligned spikes after reducing dimensionality by principal component analysis. The mixture model obtained from the spike sorting algorithm allows assessing single unit isolation quantitatively. Details of spike sorting and quantification of single unit isolation have been described previously (24, pp. 3781-3783). Briefly, one can generate a synthetic data set by sampling from the model and then apply the same maximum posterior classification rule as on the real data. Knowing ground truth for the sampled data, the estimated
number of false positive assignments and misses can be determined accurately. Throughout this paper, we only used cells for which the sum of estimated false positives and misses did not exceed 5% of their total number of spikes.

**Behavioral paradigm**

Visual stimuli were rendered by dedicated graphics workstations and displayed on CRT monitors. Trials were initiated by a sound and the appearance of a colored fixation target (~0.15°) at the center of the screen. 300 ms after the monkey acquired fixation, a sine wave grating was presented for 500 ms. After the grating was removed, the monkey had to maintain fixation for another 300 ms until the fixation target disappeared. Upon successful completion of a trial, monkeys were rewarded by a drop of juice.

The animals were implanted with a scleral search coil (S1). Eye movements were monitored online and also stored for offline analysis. Monkeys were required to fixate within a radius of 0.5–1° but typically fixated much more accurately due to prolonged training, as revealed by offline analysis.

**Visual stimulation**

As visual stimulus, we used static (27 sessions) and drifting (19 sessions) gratings in two monkeys (D & H; both types of gratings were presented to both monkeys). Static gratings were presented at eight different orientations and two different contrasts (contrast levels were changed between some sessions and ranged from 1% to 100% Michelson contrast). Drifting gratings were presented at 100% contrast and were drifting in the directions orthogonal to their orientation, resulting in 16 directions of motion. The majority of sessions were run with the following parameters: size: 4.5°; spatial frequency: 4 cycles/°; drift speed: 3.4 cycles/sec. In some sessions larger gratings were presented (9.3°: 8 sessions, 5.5°: 6 sessions). Gratings always covered and extended beyond the receptive fields of all neurons recorded from the array. For each condition we collected between 5 and 129 trials (mean: 33.4, S.D.: 24.7; small number of trials in some sessions are not problematic, see Fig. S2 and Section 1 below). In five sessions, only seven of the eight orientations were presented.

**Data set**

We recorded from a total of 917 single units (D: 483, H: 434). Of these, 60% or 569 single units (D: 368, H: 201) passed our strict criterion for single unit isolation of less than 5% misclassified
spikes as quantitatively determined from the parametric mixture model returned by our offline spike sorting algorithm (24). For analysis, only visually responsive and orientation-tuned neurons were considered (note that including all pairs of neurons irrespective of their response properties did not affect the results). Visual responsiveness was assessed by non-parametric repeated measures ANOVA (Friedman test) with factors orientation and time (with levels ‘fixation’ and ‘stimulus’) and only cells that showed a significant effect of time ($\alpha = 0.05$) were considered visually responsive. Orientation selectivity was assessed using the Rayleigh test (S2). Applying these criteria to all cells resulted in a final sample of 407 single units (D: 263, H: 146), which corresponds to 1907 (D: 1335; H: 572) simultaneously recorded pairs, in 406 (D: 361; H: 45) of which both neurons were recorded by the same tetrode.

In Fig. S9 we summarize the distribution of tuning function properties we observed in our population of neurons. Distributions of the tuning parameters half-width (Fig. S9A), orientation selectivity index (Fig. S9B, OTI = $\frac{f_{\text{post}}-f_{\text{fix}}}{f_{\text{post}}+f_{\text{fix}}}$), direction selectivity index (Fig. S9C, where applicable, $\text{DSI} = \frac{f_{\text{ref}}-f_{\text{anti}}}{f_{\text{ref}}+f_{\text{anti}}}$), baseline firing rate (Fig. S9D), and peak firing rate (Fig. S9E) are shown.

In a second series of experiments, we presented natural images to one monkey (H) in 10 sessions. Images were taken from the van Hateren database (S3) and presented under a circular aperture of 4.5°. They were shown four times for 200 ms, with pauses of 50 ms in between each presentation. For consistency with the grating experiments, we only analyzed the first 500 ms of each trial. In these experiments, we recorded a total of 92 well isolated single units, which resulted in 329 simultaneously recorded pairs.

In a third series of experiments, we presented single moving bars to a third monkey (B). Bars covered 0.15° × 5.2° and were presented at 50% contrast (Michelson contrast). On each trial, the bar moved horizontally in a randomly chosen direction at a speed of 14.7°/sec. Applying the same criteria for single unit isolation, we obtained 33 single units, resulting in 56 pairs.

**Orientation tuning**

Orientation tuning functions were fitted by least squares. A modified form of the circular von Mises distribution given by $f(\theta) = f_0 + f_m \exp(\kappa(\cos(2(\theta - \varphi)) - 1))$ was used, where $\varphi$ is the
cell’s preferred orientation, $\kappa$ defines the width of tuning (the greater $\kappa$ the narrower the curve), and $f_0$ and $f_m$ define baseline and amplitude. For directional data, this parametric form was extended to $f(\theta) = f_0 + f_m \exp\left(\kappa\left(\cos(\theta - \phi) - 1\right)\right) + f_n \exp\left(\kappa\left(\cos(\theta - \phi + \pi) - 1\right)\right)$ to account for different amplitudes $f_m$ and $f_n$ in the opposing directions of motion.

**Receptive fields**

In some recording session (gratings: $n = 12$, natural images: all 10 sessions, only monkey H), we quantitatively mapped receptive fields using a white noise random dot stimulus. A single square dot of size 0.11 to 0.13 degrees of visual angle was presented on a uniform gray background for 1 sec, changing location and color (black or white) randomly every three frames or 30 ms. Receptive field profiles were obtained using standard reverse correlation techniques. Receptive field centers were determined by fitting a two-dimensional Gaussian to the spatial receptive field profile at the time lag maximizing the signal-to-noise ratio.

**Spike count correlation**

To compute spike count correlations, spike counts $x$ and $y$ are first transformed into z-scores

$$u_\theta = z(x_\theta) = \frac{x_\theta - \mathbb{E}[x | \theta]}{\sqrt{\text{Var}[x | \theta]}}, \quad v_\theta = z(y_\theta)$$

for each condition individually. We use bold letters to denote vectors. Each element in $x$ and $y$ ($x_i$ and $y_i$) thus holds the spike count of a single trial. Spike count correlation $r_{sc}$ was calculated as the covariance of the z-scores $u$ and $v$ (of all trials concatenated):

$$r_{sc} = \mathbb{E}[uv]$$

Note that this is equivalent to first calculating $r_{sc}$ for each condition $\theta$ and then averaging over conditions:

$$r_{sc} = \left\langle \frac{\text{Cov}[x, y | \theta]}{\sqrt{\text{Var}[x | \theta] \text{Var}[y | \theta]}} \right\rangle_\theta.$$
1. Estimating correlations from limited data

Here we show that our findings are not an artifact of too few trials per experimental session. The two concerns commonly raised are (a) that with low average firing rates or spike counts it is hard to estimate correlation coefficients and (b) that one needs large numbers of trials to reliably estimate correlation coefficients.

To show that both concerns do not apply to our study, we generated artificial spike counts by sampling from a multivariate Poisson distribution (S4) with defined correlation coefficients. We matched both, the number of trials and the firing rates, to the values we measured in our experiments. We then used this artificial data to re-estimate the spike count correlations. Fig. S2A shows that the average correlation coefficients estimated in this way from limited data mimicking our dataset almost perfectly match the true population values. Of course the individual values measured can be quite variable because of finite data, but the population average is both accurate and unbiased (see section 2 below for a discussion of the variability).

In addition, there is no relationship between the number of trials in an experimental session and the average spike count correlations (Fig. S2B). Taken together, these results clearly rule out the possibility that we observe low correlations because we do not have enough data to estimate their true magnitude.

2. Variability of $r_{sc}$

The distributions of $r_{sc}$ (Figs. 2B, 3B, 4D) show scatter around the mean. Here we investigate whether this scatter reflects a property of the neural population or to what extent it is explained by the fact that correlation coefficients are estimated from finite data. To this end, we used the same sampling approach as above (S4). We fixed the average pairwise correlation to the experimentally observed mean of 0.01 and added different amounts of jitter to the individual correlation coefficients in the population. Fig. S3 shows that already without adding any jitter most of the scatter in the experimentally observed distribution is reproduced. While adding jitter with a standard deviation of ~0.06 matches the observed distribution best, a larger amount of jitter (> 0.08) results in a much broader distribution. This indicates that there is relatively little variability in the true correlation coefficients of pairs of neurons (Fig. S3A). Nevertheless, the
experimentally observed distribution does have somewhat heavier tails than what is expected without jitter.

When testing the statistical significance of individual correlation coefficients, ~14% of pairs have $P < 0.05$ compared to ~7% as would be expected from a population with constant $r_{sc} = 0.01$ (and the number of trials we recorded; Fig. S3B). This suggests that a small fraction of pairs do have significant (both positive and negative) spike count correlations. Whether this reflects a certain subpopulation of neurons, a specific combination of cell types or simply natural variability is unknown and requires further study.

Note that these observations do not contradict the main conclusion drawn from our study, namely that (a) most pairs are uncorrelated and (b) the overall average correlation is close to zero.

3. Firing rate-dependence and timescale of $r_{sc}$

There is a weak but significant relationship between $r_{sc}$ and the geometric mean firing rate of the two neurons in a pair (Fig. S4A; linear regression: slope $a = 0.0019/\text{Hz}$, $P = 6 \times 10^{-10}$, $F = 38.7$; $n = 1907$). Here we show that this relationship arises on a timescale much slower than one trial and is therefore unrelated to shared presynaptic noise.

We quantified the long-term component $r_{LT}$, which can be estimated from the trial cross-correlogram [see section 5 below and (6)]. Although we have taken great care to control for any variables that could conceivably lead to slow modulatory processes, controlled for drifts in electrode position (24) and assured excellent single unit isolation (see Materials and Methods), we find a significant fraction (~50%, $r_{LT} = 0.0053 \pm 0.0017$, mean ± S.E.M.) of spike count correlations to arise because of slow processes operating on the order of a few tens of seconds (Fig. S4B). The short-term component $r_{ST}$, which is likely to reflect shared presynaptic noise, is of the same order of magnitude ($r_{ST} = 0.0047 \pm 0.0021$, mean ± S.E.M.). It is the correlated variability on this shorter timescale that will affect the coding properties of the circuit. Therefore, we asked at what timescale the firing rate relationship described above arises. We re-analyzed the relationship between firing rates and noise correlations but replaced $r_{sc}$ by our estimate of $r_{ST}$. Consistent with the hypothesis that slow processes are generating the firing rate dependence of $r_{sc}$, we find that the dependence of $r_{ST}$ on firing rate decreases drastically. For both animals combined $r_{ST}$ does not depend on firing rates (Fig. S4C; linear regression: $P = 0.13$; $n = 1907$; in monkey H, $r_{ST}$ does not depend on firing rates; in monkey D the slope is significantly greater
than zero, $P = 0.006$). Because the firing rate dependence of $r_{sc}$ is significantly reduced at short time scales, we conclude that intrinsic variability due to shared presynaptic noise does not or only marginally depend on firing rates.

4. Effect of single unit isolation on $r_{sc}$

If multiple single units are isolated from the same electrode (2, 3, 5, 6), particular care needs to be taken that clusters of single units are well separated. We analytically analyzed the effects of false assignments of spikes between two single units on the measured value of $r_{sc}$ for different mean-variance relationships (Fano factors, $F = \text{variance} / \text{mean}$). Fig. S7A shows the relation between true $r_{sc}$ and measured $r_{sc}$ as a function of the Fano factor. For Fano factors greater than one and modest correlations (Fig. S7A third panel, which is what most experimental studies report), spike sorting errors create artificial spike count correlations.

To simulate the effect of poor single unit isolation, we perturbed the observed histogram of spike count correlations by the analytically derived curves under the assumption of a Fano factor of 1.5 and 10% contamination. As shown in Suppl. Fig. 7B, C even this modest single unit contamination would change the average spike count correlation by almost an order of magnitude from 0.01 to 0.07. Interestingly, if the percentage of falsely assigned spikes is not the same for all pairs but given by an exponential distribution (inset above Fig. S7G), this can give rise to limited range correlations, even if they are completely absent (Fig. S7D–G). This is because those pairs that are strongly contaminated will also become more similar in terms of their tuning functions. This can also lead to the impression that (a) nearby neurons are more similar and (b) have higher noise correlations than they truly do.

Finally, we tested the predictions of the above models directly on data. Indeed, as expected because the average Fano factor was slightly above 1 in our data set, the average $r_{sc}$ of pairs with very little cluster overlap (< 1% estimated false assignments, $n_1 = 8733$) was significantly smaller than that of less well isolated pairs (> 1%, $n_2 = 113$; Fig. S7J, two-sample t-test, $P = 0.014$, $df = 8844$). Note that for this analysis a different isolation criterion was used, which only took into account the direct overlap of the two clusters in each pair but ignored all other sources of contamination. The effect of contamination is even more striking if pairs are grouped according to their Fano factor: Cells with more than 3% pairwise contamination and Fano factors > 1 are more strongly positively correlated, while contaminated pairs with Fano factors < 1 have
slightly lower correlations (effect of Fano factor: $P < 10^{-3}$; interaction between level of contamination and Fano factor: $P < 0.02$; two factor ANOVA; Fig. S7K).

5. Effect of uncontrolled variables

**Spike correlation as the sum of short-term and long-term components**

Spike count correlations can arise on many timescales, indicating different generating mechanisms ranging from temporally precise synchrony or shared presynaptic noise to slow modulatory processes. We can write the spike count correlations $r_{sc}$ as the sum of a short-term component $r_{ST}$ (fast, $<1$ trial) and a long-term component $r_{LT}$ (slow, $>1$ trial)

$$r_{sc} = r_{LT} + r_{ST}.$$  

While the long-term component can be quantified by the trial cross-correlogram [TCC (6)], the short-term is usually studied by considering the (shuffle- or shift-) corrected cross-correlogram (CCG). It is important to realize that whenever slow processes are contributing, it is difficult to use the CCG to draw conclusions about the exact magnitude and timescale of $r_{ST}$ – a problem which has been pointed out previously (6, S5, S6) but is usually not explicitly addressed other than by arguing that the effect of $r_{LT}$ is negligibly small. As we show below, this is not true.

Here we analyze the effect of a slow common process operating on timescales ranging from a few hundred milliseconds to several seconds and minutes. We demonstrate that even very moderate modulations can cause $r_{LT}$ to be relatively large and confound the estimate of $r_{ST}$ significantly. Examples for such slow processes are uncontrolled variations of cognitive states (such as the precise attentional state, reward expectancy or task solving strategy), varying levels of anesthesia, changes in network state caused by damage due to movement of the electrode or an increased fraction of missed spikes because of deteriorating recording quality. For two cells whose short-term correlation $r_{ST}$ is small, we find (see below for the derivation) the following expression for $r_{sc}$:

$$r_{sc} \approx E_{XY} f_X \phi_Y v$$

where

$$B_{XY} = \frac{f_X \phi_Y}{1 + f_X \phi_Y v},$$

$v$ is the variance of the common gain-modulating process, and $f_X$ the tuning curve of neuron $X$.

Fig. S6 summarizes the results for a typical bell-shaped tuning function (Fig. S6A). It shows the
expected spike count correlation as a function of relative modulation (i.e. $\sqrt{v}$), peak firing rate and difference in preferred orientation between the two cells. Remarkably, firing rate modulations as low as 15% can cause correlations on the order of 0.2 and, thus, would suffice to reproduce previously observed results on the magnitude of correlations (Fig. S6B). Furthermore, the effect is particularly strong if stimuli are optimized for the cells under study (i.e. high firing rates, Fig. S6C) and affects cells with similar preferred orientations more than cells with different preferred orientations (Fig. S6D).

The above example shows that uncontrolled variables, which might have a seemingly minor effect at first glance, can significantly bias estimates of $r_{sc}$. It should be noted, however, that gain modulation is only one of many possibilities how an uncontrolled variable could affect firing rates and spike count correlations. Another possibility would be, for example, a purely additive effect on spike counts. In this case, the positive correlation between peak firing rates and resulting spike count correlations demonstrated above would be reversed. This could explain why there is disagreement concerning the relationship between firing rates and spike count correlations in the literature [compare, e.g. (6) and (7)].

**Derivation**

In the following, we will use the notation $E[X_i]$ for an expectation of a random variable $X_i$ in trial $t$; $E[X]$ is used for the expected value of $X$ over trials. We assume that the slow modulations can be reasonably well modeled by a common process, $g$, which is gain-modulating the firing rates of two cells and which is independent of the stimulus. Let $g(t) \equiv g_t$ be an arbitrarily shaped function with $E[g] = 1$ and $\text{Var}[g] = v$. Let $X_i$ and $Y_i$ be random variables representing the spike counts of two neurons in trial $t$. Spike counts are assumed to be drawn from a Poisson distribution with $E[X_i] = \text{Var}[X_i] = g_t f_X(\theta_i)$, where $f_X$ is the tuning function of neuron $X$. The average spike count of neuron $X$ for stimulus $\theta$ over the course of the experiment is

$$E[X | \theta] = \frac{1}{T} \sum_{t=1}^{T} E[X_i] = \frac{1}{T} \sum_{t=1}^{T} g_t f_X(\theta) = f_X(\theta) \frac{1}{T} \sum_{t=1}^{T} g_t = f_X(\theta).$$

Note that the summation is only over trials with stimulus $\theta$. The variance is given by
\[
\text{Var}_r[X | \theta] = E_r[X^2 | \theta] - E_r[X | \theta]^2
\]
\[
= \left( \frac{1}{T} \sum_{t=1}^{T} \text{Var}[X_t] + E[X_t]^2 \right) - E_r[X | \theta]^2
\]
\[
= \left( \frac{1}{T} \sum_{t=1}^{T} g_t f_X(\theta) + (g_t f_X(\theta))^2 \right) - f_X(\theta)^2
\]
\[
= f_X(\theta) \left( \frac{1}{T} \sum_{t=1}^{T} g_t \right) + f_X(\theta)^2 \left( \frac{1}{T} \sum_{t=1}^{T} g_t^2 \right) - f_X(\theta)^2
\]
\[
= f_X(\theta) + f_X(\theta)^2 \left( E[g^2] - 1 \right)
\]
\[
= f_X(\theta)(1 + f_X(\theta)v)
\]

Denote by \( U \) the z-scores of the spike counts \( X \) computed according to
\[
U_t = \frac{1}{\sqrt{f_X(\theta)(1 + f_X(\theta)v)}} (X_t - f_X(\theta)) \equiv A_X(\theta_t)(X_t - f_X(\theta_t)),
\]
where
\[
A_X = \frac{1}{\sqrt{f_X(\theta)(1 + f_X(\theta)v)}}
\]
and \( V \) the z-scores of \( Y \) calculated accordingly. Fixing two trials, \( s \) and \( t \), the covariance between \( U_s \) and \( V_t \) is given by
\[
\text{Cov}[U_s, V_t] = E[A_X(\theta_s)A_Y(\theta_t)(X_s - f_X(\theta_s))(Y_t - f_Y(\theta_t))]
\]
\[
= A_X(\theta_s)A_Y(\theta_t) \left( E[X_s Y_t] - f_X(\theta_s) E[Y_t] - f_Y(\theta_t) E[X_s] + f_X(\theta_s) f_Y(\theta_t) \right)
\]
\[
= A_X(\theta_s)A_Y(\theta_t) f_X(\theta_s) f_Y(\theta_t) \left( g_s - 1 \right) \left( g_t - 1 \right) + \delta_s \delta_t A_X(\theta_s)A_Y(\theta_t) \sqrt{f_X(\theta_s) f_Y(\theta_t)} g_s r_{ST}
\]
\[
= B_{X,s} B_{Y,t} (g_s - 1)(g_t - 1) + \delta_s \delta_t \tilde{B}_{X,s} \tilde{B}_{Y,t} g_s r_{ST}
\]
where we have used
\[
E[X_s Y_t] = E[X_s] E[Y_t] + \text{Cov}[X_s, Y_t]
\]
\[
= g_s g_t f_X(\theta_s) f_Y(\theta_t) + \delta_s \delta_t g_s \sqrt{f_X(\theta_s) f_Y(\theta_t)},
\]
\[
B_{X,s} = A_X(\theta_s) f_X(\theta_s) = \frac{\sqrt{f_X(\theta_s)}}{1 + f_X(\theta_s) v},
\]
\[
\tilde{B}_{X,s} = A_X(\theta_s) \sqrt{f_X(\theta_s)} = \frac{1}{1 + f_X(\theta_s) v},
\]
\( \delta_s \) is the Kronecker delta and \( r_{ST} \) the short-term noise correlation between \( X \) and \( Y \) which does not extend beyond one trial. The trial-cross-correlogram can now be written as
The spike count correlation \( r_{sc} \) is thus given by

\[
r_{sc} = TCC(0) = E_t[B_{X,t} B_{Y,t}] + r_{ST} E_t[\tilde{B}_{X,t} \tilde{B}_{Y,t}].
\]

### 6. Impact of correlation structure on information processing

To quantify the effect the correlation structure we observed has on information processing in the brain, we studied a simple model of a neural population encoding the orientation of a grating. We varied both the network size and its correlation structure and calculated the Fisher information \( J \) of the population:

\[
J = (f')^T Q^{-1} f' + \frac{1}{2} \text{Tr} \left[ Q' Q^{-1} Q' Q^{-1} \right],
\]

where \( f \) is the vector of tuning functions, \( Q \) the noise covariance matrix, and \( f' \) and \( Q' \) their derivatives with respect to the orientation \( \theta \). For notational convenience, we have omitted the dependences of \( f \) and \( Q \) on \( \theta \). Via the Cramér-Rao inequality

\[
E \left[ (\hat{\theta} - \theta)^2 \right] > \frac{1}{J}
\]

we can obtain a lower bound on the decoding error any unbiased estimator reading out the population activity can achieve. The Cramér-Rao bound is widely used as a measure of encoding accuracy of a neural population (8, 9, 19-21).

Our model neurons had identical tuning curves with equally spaced preferred orientations (Fig. S8A for \( n = 10 \) neurons). The tuning curves were given by modified von Mises functions as described in Materials and Methods above with the parameters \( f_0 = 1, \ f_m = 19, \ \kappa = 2 \), which were chosen to match the average tuning function of our population of recorded cells. The spike counts of the neurons were assumed to have variances equal to their means (Poisson), which is in good agreement with our data. The noise correlation coefficient between any pair of neurons was a function of the difference in the neurons’ preferred orientations and independent of the stimulus:
\[ r_{ij} = r_0 \exp\left(-\frac{(\varphi_i - \varphi_j)}{d}\right). \]

The constant \( r_0 \) is the correlation coefficient of two neurons with identical preferred orientations and \( d \) controls the steepness of the spatial decay of correlations. The covariance matrix was therefore

\[ Q_{g}(\theta) = \sqrt{f_i(\theta) f_j(\theta)} \left( \delta_{ij} + (1 - \delta_{ij}) r_{ij} \right). \]

We studied four network models with different correlation structures (Fig. S8B):

(a) completely independent population of neurons: \( r_{\text{avg}} = 0 \), (b) the correlation structure we observed: \( r_{\text{avg}} = 0.02 \), \( d = 2.0 \), (c) a correlation structure resembling that reported in (5, 6): \( r_0 = 0.36 \), \( d = 1.2 \), and (d) a correlation structure resembling that reported in (9): \( r_0 = 0.33 \), \( d = 3.0 \).

For each correlation structure, we calculated the Cramér-Rao bound for networks of different sizes (Fig. S8C). Consistent with previous reports (19-21), limited range correlations negatively affect the accuracy of a population code compared to an independent population of the same size and with equal tuning functions. Since the noise correlations we find are an order of magnitude smaller than previously reported, their negative effects are much smaller and the population is closer to independence. Specifically, if correlations were as previously reported, in order to achieve a decoding accuracy of 2° (RMS error), a downstream decoder would have to pool around four to five times as many neurons (322 vs. 1445) compared to the level of correlations we find. Note that the network with higher correlations (9) performs slightly better than the one from (5, 6) because it has relatively strong correlations even for neurons with orthogonal preferred orientations. This correlation structure thus contains a uniform correlation component, which is known to improve encoding (19, 20). This partly counteracts the negative effects of strong limited range correlations.
**Fig. S1.** (A) 3D rendering of the implant on the monkey’s skull. The shape of the bone is extracted from a structural MRI scan. The implant is custom-built to closely fit on the monkey’s skull. Thus, it only needs to be secured with screws. 24-tetrode drive on monkey H is shown. Image courtesy of Axel Klug (MPI Tübingen). (B) Photograph of the implanted drive with pre-amplifiers plugged in. (C) Electron microscopy image of tetrode tip (before electroplating). Wire diameters are 12.7 µm (just the wire) and 17.8 µm (entire strand). (D) Grid layout for 12-tetrode drive (viewed from above). Large bold numbers are tetrode numbers. Small numbers below are average impedances (kΩ) of the four channels after electroplating. Tetrodes were cut at an angle of ~90°. All tetrodes were made of Nichrome wire. (E) Same as (D) but for the 24-tetrode drive used in monkeys H and B. Impedances and materials are for monkey H. Tetrodes were made of Nichrome (white) and Platinum/Iridium (Pt/Ir, gray shading) wires.
Fig. S2. Effects of estimating correlations from finite data. (A) Artificial spike counts were generated by sampling from a multivariate Poisson distribution with specified correlations coefficients. The number of trials and the firing rates were matched to the values observed in our experiments. Black dots show the correlation coefficient estimated from data (y axis) plotted against ground truth (x axis). The gray line is the identity. (B) Average (± S.E.M.) $r_{sc}$ for sessions with different numbers of trials. There is no relationship between the number of trials and the measured spike count correlations.
**Fig. S3.** Variability of $r_{sc}$. (A) Distribution of spike count correlations when estimated from finite data. Black is the observed distribution. Colored lines are obtained by sampling spike counts from a multivariate Poisson distribution with an average correlation coefficient of 0.01 and different amounts of jitter on the correlation coefficients of each individual pair. Firing rates and number of trials are matched to our experimental data. The empirical data is consistent with a moderate amount (S.D. ~0.06) of variability in $r_{sc}$. (B) Distribution of $p$-values (of $r_{sc}$) for empirical (black) and sampled (colors) data.
Fig. S4. (A) Relationship between geometric mean firing rate and $r_{sc}$. Data are binned. Dots denote averages, error bars are ± S.E.M. Gray line is linear regression with 95% confidence intervals (gray shaded region). (B) Average (of all pairs) trial cross correlogram. Time lags are shown in seconds. Each dot corresponds to one trial (~4 sec). The continuous line is a smoothed version of the raw TCC (with the point at zero lag excluded). $r_{sc}$ can be split into a short-term ($r_{ST}$) and a long-term component ($r_{LT}$), both contributing approximately equally to $r_{sc}$. (C) Short-term component ($r_{ST}$) does not depend on firing rates (linear regression, $p = 0.13$).
Fig. S5. Distribution of spike count correlations for moving bar stimulus (mean ± S.E.M. = 0.015 ± 0.013) in monkey B. The mean is not significantly different from zero (t-test, $p = 0.25$, $n = 56$ pairs).
**Fig. S6.** Effect of common gain modulation on the long-term component ($r_{LT}$) of spike count correlations. Such a gain modulation could be caused by uncontrolled variables (e.g. shifts in attentional state etc.). (A) Tuning curve model used for the simulations in (B-D). In all simulations, both neurons have equal tuning curve shapes and peak firing rates as indicated. (B) Resulting $r_{LT}$ as a function of the strength of common gain modulation for different peak firing rates. The arrow indicates 15% modulation, which is used for (C) and (D). (C, D) Dependence of $r_{LT}$ on peak firing rate and difference in preferred orientation of two neurons.
Fig. S7. Effect of single unit isolation of spike count correlations. (A) Change of $r_{sc}$ for different percentages of spikes exchanged between two neurons for different Fano factors, $F$. (B) Histogram of $r_{sc}$ reproduced from Fig. 2B. (C) Histogram of $r_{sc}$ after perturbation by formulas from (A) with 10% contamination and Fano factor of 1.5. (D–I) Distribution of $r_{sc}$ (D, G), average $r_{sc}$ (E, H), and distribution of pairs (F, I) as a function of signal correlations $r_{signal}$ before (D–F) and after (G–I) contamination. Average contamination is 10%; individual pairs have different levels of contamination drawn form the exponential distribution shown in the inset. (J) $r_{sc}$ for extremely well isolated cells (<1% contamination) and less well isolated cells (>1% contamination). (K) Dependence of $r_{sc}$ on contamination and Fano factor, $F$. As predicted in (A), contaminated cells with $F > 1$ are affected most strongly by contamination and cells with $F < 1$ have slightly decreased. Note that almost all cells in the right bin (>3% contamination) are excluded from the analyses in the paper.
**Fig. S8.** (A) Population model: preferred orientations are uniformly spaced. Except for their preferred orientations, all model neurons have equal tuning curves. Baseline: 1 Hz, peak firing rate: 20 Hz. One tuning curve is highlighted to make it easier to see the shape of the tuning curves. (B) Limited range correlation structure. Black line is a population with correlation structure adapted from (5, 6), blue line adapted from (9), red line our data. (C) Cramér-Rao bound (minimum achievable decoding error) for previous reports (black and blue), our data (red), and an independent population of neurons (black dashed). Dotted lines highlight the number of neurons that would be necessary to achieve a decoding error of 2°. With higher correlations, four to five times as many neurons would be needed to read out orientation equally well. Note that the model from (9) slightly outperforms that from (5, 6) because it has a larger uniform component (see SOM text, section 6). (D) Same as (C) but replotted relative to independent population.
**Fig. S9.** Distribution of various tuning curve parameters for the grating data set ($n = 407$). (A) Half-width at half maximum. The half maximum is defined as $(f_{\text{pref}} - f_{\text{orth}}) / 2$, where $f_{\text{pref}}$ is the firing rate for the preferred orientation and $f_{\text{orth}}$ the one orthogonal to the preferred. (B) Orientation tuning index $(f_{\text{pref}} - f_{\text{orth}}) / (f_{\text{pref}} + f_{\text{orth}})$. (C) Direction selectivity index $(f_{\text{pref}} - f_{\text{anti}}) / (f_{\text{pref}} + f_{\text{anti}})$, where $f_{\text{anti}}$ is the same orientation as $f_{\text{pref}}$ but the opposite direction of motion. (D) Baseline firing rate $f_{\text{orth}}$. (E) Peak firing rate $f_{\text{pref}}$. 
References


