nature neuroscience

Information-limiting correlations

Rubén Moreno-Bote[1,2], Jeffrey Beck[3], Ingmar Kanitscheider[4], Xaq Pitkow[3,5,6], Peter Latham[7] & Alexandre Pouget[3,4,7]

Computational strategies used by the brain strongly depend on the amount of information that can be stored in population activity, which in turn strongly depends on the pattern of noise correlations. *In vivo*, noise correlations tend to be positive and proportional to the similarity in tuning properties. Such correlations are thought to limit information, which has led to the suggestion that decorrelation increases information. In contrast, we found, analytically and numerically, that decorrelation does not imply an increase in information. Instead, the only information-limiting correlations are what we refer to as differential correlations: correlations proportional to the product of the derivatives of the tuning curves. Unfortunately, differential correlations are likely to be very small and buried under correlations that do not limit information, making them particularly difficult to detect. We found, however, that the effect of differential correlations on information can be detected with relatively simple decoders.

Neuronal responses are typically variable in the sense that the number and timing of the spikes in response to the same stimulus is never the same from one trial to the next[1]. This variability can greatly reduce the precision of the neural code, as several values of the encoded stimulus are typically compatible with the observation of a given spike count. When stimuli are encoded in large populations of neurons, this problem may be reduced by averaging. However, the effectiveness of averaging depends greatly on the pattern of correlations across neurons (note that we use the term correlations to refer to what are commonly called noise correlations, that is, the correlations among neurons for a fixed stimulus[2]).

When the response variability is independent (**Fig. 1**), information increases linearly with the number of neurons in the population ($\rho = 0$; **Fig. 1c**). This is a well-known case in which averaging helps. In contrast, when neurons have translation-invariant tuning curves (**Fig. 1a**) and correlations among neurons are positive and stronger for similarly tuned neurons than for dissimilarly tuned ones, as has been reported in multiple cortical areas[3–7] (**Fig. 1b**), the information saturates as the number of neurons increases (**Fig. 1c**)[3,8]. In this case, beyond a certain number of neurons, averaging does not help. Because the information decreases as correlations increase (**Fig. 1c**), it would seem advantageous to decorrelate neural activity, either by a passive process such as balancing excitation and inhibition[9], or an active one such as attention[10,11].

The supposed benefits of reducing noise correlations have, in fact, motivated a large number of studies[9–12]. However, are these benefits real? Can one simply go into a network that is receiving information about the outside world, reduce the correlations and expect the information to go up (**Fig. 1c**)? Notably, neither these results nor the large number of theoretical studies on which they are based[8,13–16] can answer this question. This is because all of those studies simply assumed a correlational structure, without taking into account the fact that the information must come from other spike trains, which are themselves variable.

So what changes when one considers the more realistic case of networks receiving noisy external input? An obvious change is that the input carries finite information. For example, because of variable distortions caused by the lens, micro–eye movements and ocular media, even an ideal observer of photoreceptors in bright light (where the noise is essentially nonexistent) would not know the orientation of a line exactly. And for more complex tasks, the problem is worse: when processing the speech of a person talking in a noisy street, the ability to recognize the words is limited by the physical mixing of the voice with the background noise, which in turn imposes a limit on the information conveyed by the sound stream. Thus, even if neurons were independent, adding more of them wouldn't increase information forever. This immediately rules out the $\rho = 0$ line in **Figure 1c** for large networks.

A less obvious effect of realistic input is that it changes the relationship between correlations and information. Information drops as correlations increase (**Fig. 1c**). But this assumes that only the correlations change. What would happen if the input was fixed and network parameters such as connectivity or single neuron properties were modified? Such a modification would typically change the correlations, but it would also change other aspects of the network, including tuning curves. It's not known in general what would happen to information in this case. However, we will show at least one realistic network in which the level of correlations has virtually no effect on the information. Thus, contrary to what **Figure 1c** suggests, smaller correlations do not necessarily imply more information.

This doesn't mean that correlations don't affect information. Indeed, they do. However, it's not typically the size of correlations that matters, it's the pattern. We found that large networks receiving finite

[1]Research Unit, Parc Sanitari Sant Joan de Déu and Universitat de Barcelona, Esplugues de Llobregat, Barcelona, Spain. [2]Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Esplugues de Llobregat, Barcelona, Spain. [3]Department of Brain and Cognitive Sciences, University of Rochester, Rochester, New York, USA. [4]Department of Basic Neuroscience, University of Geneva, Geneva, Switzerland. [5]Department of Neuroscience, Baylor College of Medicine, Houston, Texas, USA. [6]Department of Electrical and Computer Engineering, Rice University, Houston, Texas, USA. [7]Gatsby Computational Neuroscience Unit, University College London, London, UK. Correspondence should be addressed to A.P. (alexandre.pouget@unige.ch).

Received 20 May; accepted 14 August; published online 7 September 2014; doi:10.1038/nn.3807

information must contain correlations approximately proportional to the product of the derivatives of the tuning curves (referred to as differential correlations, see below), which are solely responsible for the information limitation. Thus, positive correlations between neurons with similar preferred stimuli (**Fig. 1b**) do not always limit information: they do so only when they contain differential correlations. We also found that these information-limiting correlations can be exceedingly difficult to measure directly, primarily because they can be very small and masked by other correlations. Fortunately, their effect on information can be detected with a realistic number of trials so long as the neurons are recorded simultaneously.

RESULTS

Why decorrelation does not imply more information: a simple case

To determine the relationship between correlations and information, we considered a network that receives finite information. We varied the parameters of the network in a way that caused the correlations to change without changing the input information, and examined how this affected the information in the network (**Fig. 2a**, Online Methods (equations (6–11)) and **Supplementary Modeling**).

The network consists of an all-to-all connected homogeneous population of leaky integrate-and-fire excitatory and inhibitory neurons. Connection strengths were chosen so that each neuron received large amounts of excitation and inhibition, the so-called balanced regime. In addition, each neuron received external input in the form of a common signal s corrupted by temporal white noise. The white noise has two components: an independent one, with variance σ^2_{ind}, and a shared one, common to all neurons, with variance σ^2_s. Because of the shared component of noise, the information entering the network is finite.

We used a particular measure of information known as Fisher information, which is inversely proportional to the square of the discrimination threshold of an ideal observer of the neural activity[17,18]. We focused on Fisher information because many animal experiments involve discrimination tasks. The rate at which Fisher information enters the network, the input information rate (Online Methods, equations (27) and (28)), is given by

$$I_{\text{in}} = \frac{1}{\sigma_s^2 + \frac{\sigma^2_{\text{ind}}}{N}} \tag{1}$$

and, in the limit that the leak time-constant goes to infinity, the output information rate is given by

$$I_{\text{out}} = \frac{1}{\sigma_s^2 + \frac{\sigma^2_{\text{ind}}}{N} + O(T^{-1/2})} \tag{2}$$

where N is the number of neurons in the network and T is the observation time window. We found that in the relevant regimes, time windows above 2 s and networks above 250 neurons, the dependence on T was relatively weak (**Supplementary Fig. 1**). Equation (2) implies that, for large time windows and a long leak time-constant, the output information is equal to the input information for all values of N: the network preserves all of its input information despite the spiking nonlinearity of integrate-and-fire neurons. It should be noted, however, that the information saturates as a function of the number of neurons, very much like what we saw in **Figure 1c**. However, unlike that case, equation (2) has no explicit dependence on the overall correlations in the network; instead, it depends only on the correlations inherited from the input. This implies that we should be able to change network parameters in such a way that correlations change without affecting information. This is indeed what we found: we could change the mean level of correlations by a factor 10 or more (**Fig. 2b,c**), simply by changing connectivity, without changing the information (**Fig. 2d**). Notably, this is true even for small networks ($N = 75$) for which the input information has not yet saturated to its maximum value (**Fig. 2d**).

For the more realistic case of leaky integrate-and-fire neurons, it is much more difficult to compute information analytically, but simulations revealed similar behavior: information was independent of the level of correlations for large networks (**Fig. 2d**). For small networks ($N = 75$), away from saturation, information decreased slightly with overall correlations, but this dependence was very weak compared to what we saw in **Figure 1c**. Indeed, for $N = 500$, information changed by a few percent when the correlations changed by a factor of 10 (**Fig. 2b,d**), whereas information changed by 1,000% for the same relative change in correlations in **Figure 1c**.

What this example shows is that noise decorrelation does not necessarily increase information in networks of spiking neurons receiving finite information. This doesn't mean, of course, that correlations have no effect on information. The saturation of information was indeed a result of correlations (**Figs. 1c** and **2d**). However, this saturation was a result of a very specific pattern of correlations. To show this, we left spiking networks and asked a general question. If the information in a network saturates as the number of neurons increases, what is the pattern of correlations that must be present? In other words, what do information-limiting correlations look like?

Information-limiting correlations in population codes

Consider a population composed of N neurons with bell-shaped tuning curves. For such a population, the mean activity in response to stimulus s takes the shape of a hill of activity (**Fig. 3a**). If the distribution of neuronal responses conditioned on the stimulus, $p(\mathbf{r}|s)$, follows the exponential family with linear sufficient statistics, which is known to provide a good approximation to neural responses *in vivo*[1,5,7,19–24], then Fisher information is given by

$$I = \mathbf{f}'^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{f}' \tag{3}$$

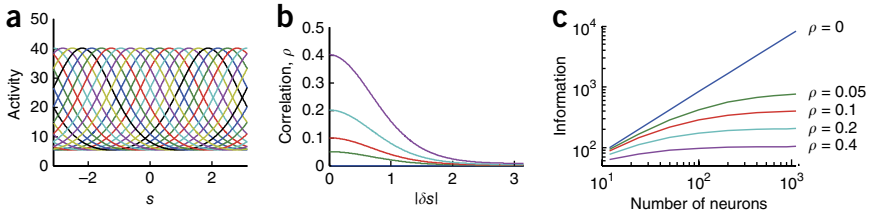

Figure 1 The effect of correlations on a population code with translation invariant tuning curves. (**a**) A neuronal population with translation invariant tuning curves to the stimulus s (arbitrary units). (**b**) Correlations *in vivo* often decrease as a function of the difference in preferred stimuli, δs. This decrease is often reasonably well described by a circular Gaussian (as shown here) or an exponential function of the stimulus. (**c**) Information in a population of neurons with the tuning curves and correlations shown in **a** and **b**. The different curves correspond to different maximum correlation values (the correlation coefficient, ρ, at $\delta s = 0$). When neurons are independent ($\rho = 0$), the information scales linearly with the number of neurons. For $\rho > 0$, the information saturates as N increases. This plot is often used to argue that correlations such as those in **b** limit information in population codes.

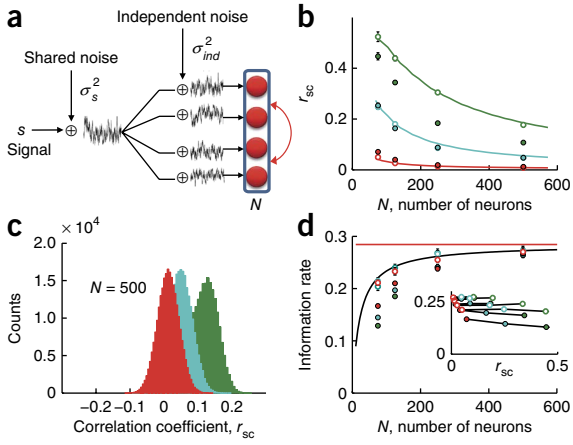

Figure 2 Decorrelation does not necessarily increase information. (**a**) Network architecture. Each neuron receives input from recurrent connections, and, in addition, external input with mean proportional to s, but corrupted by shared and independent noise (Online Methods, equation (9)). (**b**) Mean correlation coefficient as a function of the number of neurons. The three colors correspond to three networks that differ only in their connection strengths, but chosen so that, in all cases, the mean firing rate was close to 40 Hz. For a fixed number of neurons, the mean correlation coefficients can vary by more than a factor of 10 across the three networks. Solid lines show the analytical predictions for non-leaky integrate-and-fire neurons, and the open and closed circles show the results of the simulations with non-leaky and leaky integrate-and-fire neurons, respectively. Some of the open circles fall behind the closed ones and are therefore not visible. The observation time window was 10 s for the non-leaky integrate-and-fire neurons and 2 s for the leaky integrate-and-fire neurons. (**c**) Histogram of correlation coefficients ($N = 500$); same color code as in **b**. The red distribution has a mean very close to zero (0.013), whereas the green distribution has a mean of 0.108. (**d**) Information as a function of the number of output neurons (same color code as in **b**). The black solid line shows the input information (equation (1)) and the red solid line corresponds to the information in the input for infinite networks (equation (1) with N taken to be infinity). As in **b**, open and closed circles show the results of the simulations with non-leaky and leaky integrate-and-fire neurons, respectively. Inset, information as a function of the correlations for each network size. Unlike in **Figure 1c**, the information at which the network saturates (open dots, $N = 500$) has a very weak dependence on the mean correlations; instead, all three networks show nearly the same asymptotic value of information. Even for small networks away from the saturation, information is independent of overall correlations for non-leaky integrate-and-fire neurons, and only weakly dependent on correlations for leaky integrate-and-fire neurons. Thus, smaller correlations do not necessarily imply more information. Error bars correspond to s.e.m.

where $\mathbf{f}'$ is a vector of the derivatives of the tuning curve with respect to s ($\mathbf{f} = (f_1(s), \ldots, f_N(s))^T$ and $\mathbf{f}' = d\mathbf{f}/ds$), and Σ is the noise covariance matrix of the neural activity. The right side of equation (3) is sometimes referred to as the linear Fisher information; independent of the spike train statistics, its inverse is the variance of a locally optimal linear estimator[25].

What are the types of covariance matrices for which Fisher information saturates as N goes to infinity? Or equivalently, what are the covariance matrices that limit the ability to detect a small change in the stimulus? With population codes, a small change in the stimulus induces a shift in the hill of activity (**Fig. 3a**). One kind of noise that can limit performance is noise that similarly shifts the hill of activity sideways from trial to trial across multiple presentations of the same stimulus. In this case, the noise looks just like the signal and cannot be averaged away. To see this more explicitly, consider a space in which each axis corresponds to the activity of one neuron. A particular noiseless hill of activity corresponds to a point in that space, and the set of all possible noiseless hills (corresponding to all possible stimuli) lies on a curve in the same space (**Fig. 3a**,**b**). Noise that shifts only the position of the hill produces random shifts along the curve. This type of noise results in a distribution that lies only on the curve (**Fig. 3b**). In other words, given a true stimulus, the activity ends up somewhere along the curve corresponding to the set of noiseless hills. If the shifts are small compared with the curvature, this distribution can be approximated by one that lies along a line tangent to the curve (**Fig. 3b**). This distribution has a covariance matrix proportional to the product of the tuning curve derivatives, which we refer to as differential correlations; these are the correlations that limit information (**Supplementary Fig. 2**). For simplicity, here we focus on the approximate distribution.

To see more formally that differential correlations limit information, consider a covariance matrix of the form

$$\mathbf{\Sigma} = \mathbf{\Sigma}_0 + \varepsilon \mathbf{f}'\mathbf{f}'^T \tag{4}$$

where Σ_0 is a covariance matrix that does not limit information, that is, a matrix such that $I_0 = \mathbf{f}'^T\mathbf{\Sigma}_0^{-1}\mathbf{f}'$ does not saturate as N goes to infinity. The information in this case (Online Methods, equation (34)) is given by

$$I = \frac{I_0}{1 + \varepsilon I_0} \tag{5}$$

Because I_0 goes to infinity with N, this expression saturates at $1/\varepsilon$ as N goes to infinity. We found that this result is more general (Online Methods): differential correlations are the only correlations that can lead to information saturation in the large N limit. Notably, this result does not rely on any assumptions about the shape of the tuning curves or whether the tuning curves depend on a single variable. Thus, it holds for any shape of the tuning curves (that is, not just Gaussian tuning curves; **Fig. 3**), and for tuning curves that depend on multiple variables, including time (the only difference when there are multiple variables is that total derivatives must be replaced with partial derivatives). This last point implies that our analysis applies to dynamical networks in which tuning curves change over time, as is the case, for instance, in the motor system[26].

Although the effects of differential correlations are most prominent in the large N limit, they are also important for small populations, or for observation times that are so short that only a handful of neurons fire (thereby effectively reducing the population size). Even in these cases, increasing the level of differential correlations always lowers the information. This is because differential correlations effectively move the hill of activity to a different place on every trial, and so necessarily make it harder to accurately decode the stimulus, independent of how many neurons there are, the observation time window or the structure of the noise. This is true even if information is defined as the inverse of the variance of the locally optimal linear estimator, as opposed to Fisher information. In the large N limit and for long time windows, these two quantities are equal. However, for small time windows, Fisher information underestimates the discrimination threshold of an optimal estimator[27].

In addition, we found a subtle, yet important, result (Online Methods, equation (39)). Suppose correlations of the form $\mathbf{u}\mathbf{u}^T$ are added to the covariance matrix. If $\mathbf{u}$ is not parallel to $\mathbf{f}'$, then information decreases, but it does not saturate in the large N limit. Thus, only differential correlations can make information saturate as N increases; other correlations can decrease information, but cannot make it saturate.

Potential sources of differential correlations

Two factors contribute to the emergence of differential correlations in the brain: limited information in the world and approximate

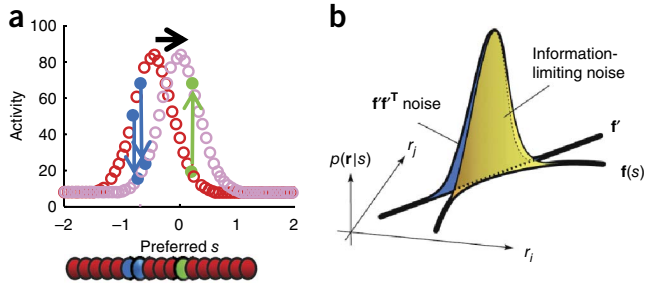

Figure 3 Differential correlations induced by a shifting hill. (**a**) Population activity for neurons with translation invariant tuning curves (as in **Fig. 1a**), with neurons ranked according to their preferred stimulus. The red and pink curves correspond to the population response to the same input on two different trials. We assumed that the variability is such that the hill simply translated sideways from trial to trial. When this was the case, correlations were mostly proportional to the product of the derivatives of the tuning curves. The two blue neurons were positively correlated because the derivatives of their tuning curves were negative for both neurons, and the product of their derivatives was therefore positive. In contrast, the green neuron was negatively correlated with either of the blue neurons because its derivative had the opposite sign as theirs. (**b**) Population patterns of activity, as in **a**, can be thought of as points in an N-dimensional space, in which each axis corresponds to the activity of one neuron. Only two neural dimensions are shown here, out of N. As the hill of activity shifts with s, the mean population activity traces out a curve (the black curve labeled $\mathbf{f}(s)$). Pure information-limiting noise looks like a sideways shift of the hill, corresponding to movement along the curve. In this case, the activity has a probability distribution that wraps along the curve $\mathbf{f}(s)$, as shown in yellow. If the variability is small compared with the curvature of the manifold, the yellow distribution can be approximated by the blue distribution that lies along the tangent to the curve, corresponding to the $\mathbf{f}'(s)$ direction, with a resultant covariance matrix proportional to $\mathbf{f}'(s)\mathbf{f}'(s)^{\mathsf{T}}$.

computations. The first factor is quite simple. If the information provided by the outside world about a stimulus, s, is finite, then information must saturate as we increase the number of neurons coding for s, just as in **Figures 1c** (with $\rho \neq 0$) and **2d**. This in turn implies the presence of differential correlations (assuming the tuning curves don't disappear altogether). The second factor is suboptimal computations[28]. If information saturates because of finite information in the world, and the brain performs suboptimal computation on this input, the value at which information saturates must decrease. This is because, by definition, suboptimal computation must induce an information loss. Given that the amount of information at saturation is controlled by differential correlations, suboptimal computations must also increase differential correlations (again assuming that the tuning curves don't change and that suboptimal computations introduce additional variance, not bias).

This tells us when differential correlations arise, but it doesn't tell us how they arise. A natural culprit is shared connectivity. Indeed, it has been proposed that shared connectivity among neurons with similar tuning properties induces positive correlations that limit information[29]. The intuition is quite simple. If neurons share input, the resulting shared variability in their response cannot be averaged out. Although this intuition is sometimes valid, it isn't always: multiple neurons may partially share different aspects of the variability, and those differences can be combined to eliminate the variability almost entirely[8,30]. To demonstrate this, we simulated a feedforward network, wired so that each neuron in the output layer received a large number of shared connections from neurons in the input layer (**Fig. 4a**). We used independent neurons in the input layer, and the input and output layers contained the same number of neurons, such that input information scaled with the number of neurons. The drive to the neurons in the output layer was chosen so that they all fired at close to 50 Hz; this produced near Poisson statistics (Fano factors of 0.9; **Fig. 4b**). Because of shared connections, the neurons in the output layer were correlated, with average correlation coefficients around 0.1 (**Fig. 4c** and **Supplementary Modeling**). Despite these correlations, information grows linearly with the number of neurons (**Fig. 4d**), with the same slope as the information in the input layer (data not shown). Thus, this network, similar to the one we studied in **Figure 2**, preserves the information it receives. Shared connectivity therefore does not necessarily induce information-limiting correlations.

Another potential source of information-limiting correlations is shared fluctuations in the excitability, or gain, of neurons. This has recently been shown to be a major source of correlations in cortex, particularly in anesthetized animals[31,32]. In the case of population codes like the one shown in **Figure 3**, this would induce fluctuations in the height of the hill of activity, but not in the position of this hill. As such, the correlations induced by these shared fluctuations have little effect on discrimination tasks. However, they could create differential correlations for detection task, thereby making it harder to determine whether or not an object is present or what its overall contrast is. Note that these arguments are not specific to bell-shaped tuning curves, but also apply to curves with other shapes, including sigmoidal.

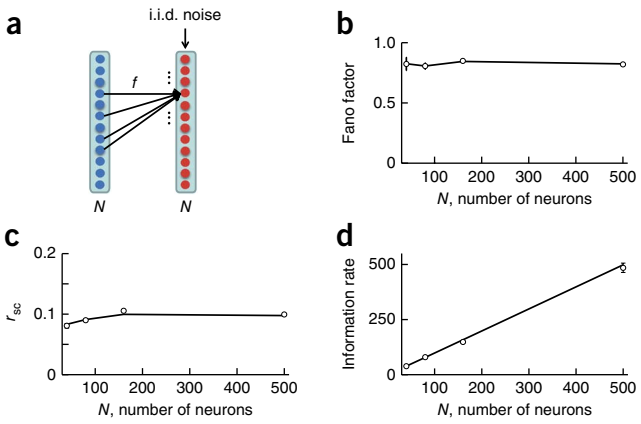

Figure 4 Shared connections do not necessarily induce information limiting correlations. (**a**) Network architecture. The network consists of two layers. The input layer contains N neurons. These are modeled as white noise process and they project to N output neurons. The output neurons, which were not recurrently connected, were modeled as non-leaky integrate-and-fire neurons. The output neurons also receive independent and identically distributed (i.i.d.) noise. A parameter f controls the probability that a connection is shared by two neurons. (**b**) The Fano factor was roughly constant at around 0.9. Not shown are the firing rates; these were held at 50 Hz, independent of network size, by adjusting the mean drive to each neuron. (**c**) Correlations were also roughly constant as the size of the network increases. (**d**) Information in the output layer (which was the same as in the input layer; data not shown). Unlike in **Figure 1c**, where positive correlations led to information saturation as a function of the number of neurons, information increased linearly with network size. Thus, the correlations induced by the shared connectivity do not limit information in this case. In **b**–**d**, the solid lines indicate analytical predictions, the dots show the results of simulations and error bars represent s.e.m.

Differential correlations might be small and masked

We then turned to the problem of detecting differential correlations in experimental data. One approach is to look for them directly. When differential correlations are the only correlations, they are easy to find. One can simply plot the correlation coefficient between any pair of neurons as a function of the difference in preferred stimulus, and average over preferred stimuli, as is often done in experiments[33]. For bell-shaped tuning curves, the resulting plots have a very characteristic shape (**Fig. 5a**): positive when the difference in preferred stimulus is small and negative when the difference is large. Moreover the correlations between pairs of cells are stimulus dependent and have a characteristic shape when plotted as a function of stimulus. We plotted the correlation coefficient versus stimulus for two pairs of neurons (**Fig. 5b**), one with preferred stimuli at ±0.5 and the other with preferred stimuli at ±0.25. The negative correlations at zero (where the stimulus falls on parts of the tuning curves that have opposite slope for the two neurons) and the positive side lobes (where the stimulus falls on parts of the tuning curves that have the same slope for the two neurons) were telltale signs of differential correlations.

Correlations such as these (**Fig. 5a**) have, to the best of our knowledge, been seen in only one study[34]. In that study, activity corresponded to a slowly diffusing hill of activity in a working memory—ideal conditions for the emergence of differential correlations. However, such correlations have not been observed in other experiments. Instead, correlations, particularly in sensory areas, look like those shown in **Figure 1b**; they do not have the stimulus dependence shown in **Figure 5b**[33,35]. This would appear to suggest that differential correlations *in vivo* are very rare, or at the very least show up under only the most favorable conditions. However, that is not the case. **Figure 5a**,**b** correspond to almost pure differential correlations (with additional independent noise), but it is quite likely that the nervous system contains additional correlations that do not limit information, and those correlations can mask the differential component. Thus, the pattern of correlations predicted by pure differential correlations may be very hard to see. Indeed, in **Figure 2** of ref. 36, the correlations look somewhat like those shown in **Figure 5b**, in the sense that they are slightly negative when the preferred stimulus is zero and slightly positive away from zero. However, the modulation was weak compared with the error bars, making it difficult to draw strong conclusions.

To see the masking effect in a simplified setting, consider tuning curves of the form

$$f_i(s) = a + b\cos(s - s_i)$$

with $a > b$ (to avoid negative or zero firing rates), and a covariance matrix given by

$$\Sigma_{ij} = (1 - c)\delta_{ij} + c\cos(s_i - s_j)$$

where δ_{ij} is 1 when $i = j$ and 0 otherwise and the preferred stimuli, s_i, are equally spaced. It can be shown for this case that information saturates with the number of neurons (**Supplementary Modeling**). Yet the correlations have no stimulus dependence (unlike in **Fig. 5a**), and there appears to be no differential component. This, however, is a bit of an illusion. We can rewrite the covariance matrix as

$$\Sigma_{ij} = (1 - c)\delta_{ij} + c\cos(s - s_i)\cos(s - s_j) + c\sin(s - s_i)\sin(s - s_j).$$

The last term, $c\ \sin(s - s_i)\ \sin(s - s_j)$, is proportional to $f_i'(s)f_j'(s)$, which is exactly the kind of correlation that limits information.

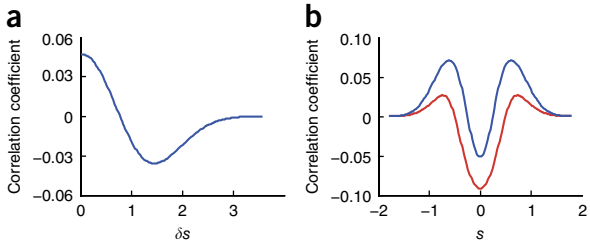

Figure 5 Differential correlations. (**a**) Average correlation coefficient as a function of the difference in preferred stimuli for a population of neurons with differential correlations (as shown in **Fig. 3a**) plus independent Poisson noise. The average was taken over all pairs of neurons with the same difference in preferred stimuli, δs. Correlations were negative for large δs, in contrast to the correlations found *in vivo* (**Fig. 1b**). (**b**) Correlation coefficient between two pairs of neurons as a function the stimulus. The blue line represents a pair of neurons with preferred stimuli −0.5 and 0.5. The red line represents a pair with preferred stimuli −0.25 and 0.25. Correlations were strongly modulated by the value of the stimulus.

Masking via the above mechanism is one way that differential correlations can be hidden. They may also simply be small compared with the other correlations. To illustrate this, we considered the more biologically realistic case of heterogeneous tuning curves, with amplitudes and widths varying from one neuron to the next (**Fig. 6a**). This case is important because, for heterogeneous tuning curves and the correlations shown in **Figure 1b** (and, in fact, for any correlational structure that is sufficiently independent of the tuning curves), information does not saturate with N[15,16] (**Fig. 6b**). However, adding even a small differential component causes information to saturate. This saturation is shown by the blue curve in **Figure 6b**, which is the information in a population with heterogeneous tuning curves (that is, a set of tuning curves with varying peak firing rates; **Fig. 6a**), with the covariance matrix set to $\Sigma_0 + \varepsilon \mathbf{f}'\mathbf{f}'^{\mathrm{T}}$. Here, Σ_0 corresponds to the kind of correlations shown in **Figure 1b**, and $\varepsilon = 0.0027$ (**Supplementary Modeling**). Notably, the presence of the differential component ($\mathbf{f}'\mathbf{f}'^{\mathrm{T}}$) cannot be revealed by plotting the correlations as a function of the difference in preferred stimuli. The correlation coefficients estimated empirically from 1,000 trials looked essentially the same whether or not there were information-limiting correlations (**Fig. 6c**).

These examples show that it is difficult to detect differential correlations simply by inspecting the noise correlations. This is because differential correlations can be very small and may be masked by non–information-limiting correlations. However, it is possible to detect the effect of differential correlations on information (see below). This must be done by estimating information directly, rather than by estimating correlations.

Discrete classification and other performance measures

Although our results thus far were derived for fine discrimination and Fisher information, they generalize to coarse discrimination between two classes. In this case the information-limiting correlations are proportional to $\Delta\mathbf{f}\Delta\mathbf{f}^{\mathrm{T}}$, where $\Delta\mathbf{f}$ is the mean difference in neural responses for the two classes; this is just the discrete version of correlations proportional to $\mathbf{f}'\mathbf{f}'^{\mathrm{T}}$. To illustrate this, we plotted the percent correct in a binary categorization task as a function of the number of neurons (**Fig. 6d**). As before, we compare two correlational structures: Σ_0 and $\Sigma_0 + \varepsilon\Delta\mathbf{f}\ \Delta\mathbf{f}^{\mathrm{T}}$, where Σ_0 corresponds to the kind of correlations shown in **Figure 1b** (the same covariance matrix used in **Fig. 6b**) and ε was set to 0.2742 (**Supplementary Modeling**). Asymptotic performance was 75% in the presence of correlations, but

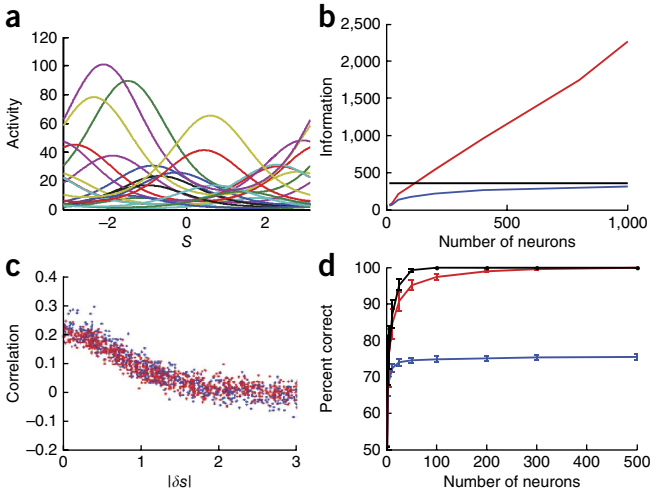

Figure 6 Small differential correlations can have a large effect on information. (**a**) A population code with tuning curves of varying amplitudes. (**b**) Information as a function of the number of neurons for the population code shown in **a**. The two curves correspond to two different covariance matrices, Σ_0 (red) and $\Sigma_0 + \varepsilon \mathbf{f}'\mathbf{f}'^T$ (blue), where Σ_0 is a covariance matrix in which pairwise covariances follow a decaying exponential function of the cosine of the difference in preferred stimuli (as in **Fig. 1b**; see **Supplementary Modeling**). Information saturated in the presence of differential correlations (the $\varepsilon \mathbf{f}'\mathbf{f}'^T$ component). (**c**) Empirically estimated correlations from 1,000 trials plotted as function of the difference in preferred stimuli for the two covariance matrices Σ_0 (red) and $\Sigma_0 + \varepsilon \mathbf{f}'\mathbf{f}'^T$ (blue). The two distributions of correlation coefficients were nearly indistinguishable even though the amount of information was very different (**b**). (**d**) Percentage correct in a binary classification task as a function of the number of neurons. The red and blue curves correspond to two different covariance matrices, Σ_0 (red) and $\Sigma_0 + \varepsilon \Delta\mathbf{f}\,\Delta\mathbf{f}^T$ (blue), where Σ_0 is as described in **b** and $\Delta\mathbf{f}$ is the mean difference of activity of the neurons for the two stimulus classes; the black curve corresponds to the performance of an independent population. Performance was limited to 72% when $\Delta\mathbf{f}\Delta\mathbf{f}^T$ correlations were present, but reached the value predicted for an independent code when there were no $\Delta\mathbf{f}\Delta\mathbf{f}^T$ correlations (**Supplementary Modeling**). Data points and error bars represent mean and s.e.m., respectively.

100% when these correlations were removed, once again illustrating that differential correlations limit performance.

Our results are also valid for two-alternative forced-choice tasks in which subjects are asked to effectively determine a boundary. A classic example is the well-known motion dot task[37], in which subjects are asked to discriminate upward versus downward motion. The difficulty of this task is determined by the coherence of the dots (the percentage of dots moving coherently upward or downward), and is hardest when the coherence is near zero. This task would seem to involve a coarse discrimination of direction of motion (up versus down), and is often referred to as a coarse discrimination task. However, subjects are effectively asked whether coherence is greater than or less than zero, with greater than zero corresponding to up, and less than zero to down, so this is really a fine discrimination task around a coherence of zero. Information-limiting correlations are, then, the ones proportional to the product of the derivative of the tuning curves with respect to coherence, evaluated at a coherence of zero. The fact that the derivative is with respect to coherence, rather than direction of motion, illustrates an important point: correlations that limit information are different for different stimuli and tasks.

Detecting differential correlations by directly measuring information

Although it is typically not possible to observe differential correlations directly, we can infer their existence by computing information versus the number of neurons, and determining where the information saturates. However, although it is straightforward in principle to compute information from data, in practice, care must be taken.

Correlations between spike trains are typically collected by recording a few neurons at a time with single electrodes, tetrodes or electrode arrays. Even with large electrode arrays, most neurons are not recorded simultaneously. Thus, if one built a covariance matrix based on that data, many entries would be empty. Unfortunately, estimating information from such an incomplete covariance matrix is prone to profound errors. This is because one must use interpolation techniques to fill in the missing entries. For example, a common approach is to assume that correlations decrease with the distance between preferred stimuli (**Fig. 1b**). However, this is almost guaranteed to lead to a serious mis-estimation of information: if the tuning curves are not translation invariant, as is the case *in vivo* (and as is shown in **Fig. 6a**), then correlations of this type do not limit information (**Fig. 6b**). One would mistakenly conclude that information is not limited even when the population code contains an information-limiting component. This is, in fact, a general problem: one usually relies on strong assumptions about the parametric form of the covariance matrix, and estimates of information will depend more on the parameterization than on the scarce covariance data.

To illustrate this point, we performed simulations in which we measured correlations among $N/2$ pairs of neurons and, as proposed previously[38] (**Supplementary Modeling**), we filled in the missing elements by approximately resampling the measured correlation coefficients. This approach greatly over-estimated the information and, not surprisingly[15,16], failed to reveal that the information saturates (**Fig. 7a**). It should be noted that, in a previous study[10], this approach did reveal saturating information. This is because the study used a suboptimal decoder (**Supplementary Modeling**).

Even when all the neurons are recorded simultaneously, such that all pairwise correlations can be measured, estimating information can be tricky because of the limited number of trials. In **Figure 7b**, we show the information obtained directly via equation (3) for the information-limited neuronal population simulated in **Figure 6a**. Here, 'directly' means that we estimated derivatives of the tuning curves and the correlations from data, inverted the covariance matrix, and used equation (3). Even for 2,400 trials, the information obtained in this way grew linearly with the number of neurons; it took 4,000 trials for a hint of saturation to emerge. Similar results were obtained with ridge regularization of the covariance matrix (**Supplementary Modeling**).

Fortunately, so long as enough neurons are recorded simultaneously, it is possible to obtain a reliable estimate of information. This can be done using a cross-validated decoder of neural activity (trained with gradient descent and early stopping[39], a method that bypasses the matrix inversion step of the direct method described above, making it much more robust; **Supplementary Modeling**). This approach revealed information saturation with only a few hundred trials (**Fig. 7c**). It should be noted, however, that although a linear decoder is guaranteed to reveal information saturation when the true information saturates, the reverse is not true: finding that the information estimated by a linear decoder saturates does not imply that the true information saturates, as that decoder may be suboptimal (**Supplementary Modeling**).

Perhaps surprisingly, although differential correlations almost exclusively determine information in the large N limit, the locally

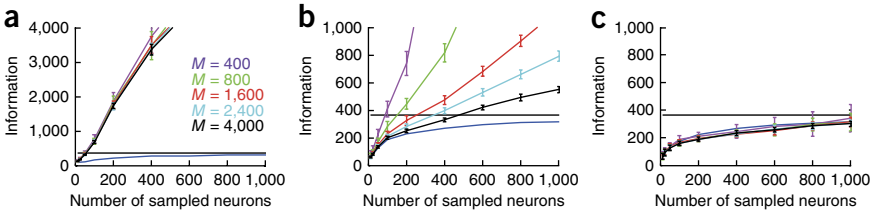

Figure 7 Estimating information from neuronal populations with differential correlations (**Supplementary Modeling**). (**a**–**c**) Empirical estimate of Fisher information versus the number of sampled neurons (N) and number of trials (M). In all panels, the solid blue line shows the average Fisher information as a function of the number of simultaneously sampled neurons, with the average calculated over 20 random sets of N neurons; the black horizontal line shows the true Fisher information for the entire population, which is assumed to have infinite size. Error bars represent s.e.m. The same color code for M is used in all panels. (**a**) Fisher information computed from equation (3), using a covariance matrix in which only $N/2$ correlation values were measured experimentally. The missing entries were estimated from the empirical measurements by requiring that they have approximately the same statistics as the observed correlation coefficients[27]. Regardless of the number of trials, the estimated information failed to reveal the saturation of information, thereby missing the presence of differential correlations. (**b**) Fisher information again computed from equation (3), but using a covariance matrix estimated from N simultaneously recorded neurons. The estimated information still missed the information saturation even for thousands of trials. (**c**) Fisher information estimated with a locally optimal linear estimator trained with early stopping to prevent overfitting. This method consistently returned a lower bound on the true information for the entire population (horizontal black line), even for a small number of trials and neurons. As a result, it revealed the presence of information limiting correlations. Data points represent mean values in all panels.

optimal linear decoder does not depend on them. Specifically, if the covariance matrix is $\Sigma_0 + \varepsilon \mathbf{f}'\mathbf{f}'^{\mathrm{T}}$, as in equation (4), the optimal decoder depends on Σ_0, but not on ε. Intuitively, this is because even an optimal decoder cannot get rid of the fluctuations resulting from the differential correlations, as they simply shift the hill of activity without leaving any trace of whether it was shifted by the stimulus or by noise. All the decoder can do is to ensure that the correlations resulting from Σ_0 have been properly taken into account (Online Methods, equations (33), (36) and (37)).

To conclude, in large populations of neurons, it is typically difficult to assess the patterns of correlated activity with sufficient accuracy to determine whether or not differential correlations exist, and so it is difficult to conclude anything about information in a population from measurements of correlations. Nonetheless, it is possible to accurately assess information and, if enough neurons are recorded, to find out whether sensory information in a neural population saturates. It is critical to always use neurons that are recorded simultaneously (that is, never fill in missing entries in the covariance matrix). But even with simultaneously recorded neurons, it is best to use a method such as a cross-validated decoder to obtain accurate estimates of information (**Fig. 7c** and **Supplementary Fig. 3**), as opposed to direct estimates obtained via equation (3) (**Fig. 7b**).

DISCUSSION

The fact that information can saturate in population codes with positive correlations, such as those shown in **Figure 1b**, has often been used to argue that positive noise correlations among neurons with similar tuning properties limit information. It has also been proposed that such positive correlations are the unavoidable consequence of shared input connections between neurons with similar tuning, thereby suggesting that shared connections might be the main cause of information limitation in neural circuits[29].

Our results contradict this perspective in several respects. First, we found that, when information is limited, the limit is a result of differential correlations; that is, correlations proportional to the product of the derivatives of the tuning curves. These correlations necessarily emerge in any (sufficiently large) network receiving finite information and/or performing suboptimal computations. Second, correlations induced by shared connections did not necessarily limit information. And third, information was not limited by the overall level of correlations or by correlations that fell off with the difference in preferred stimuli (also see refs. 15,16), from which it follows that decorrelation does not necessarily increase information.

It is crucial to keep in mind that these results, particularly the last one about decorrelation, are valid only for noise correlations (the correlations among neurons at fixed stimulus). In the case of signal correlations (the correlations between the mean responses of the neurons as the stimulus varies), the story is very different and is very well understood, thanks to previous studies (for a thorough discussion, see ref. 40).

We also found that it is very difficult to detect differential correlations *in vivo* by measuring the correlations directly, as the differential correlations are likely to be masked by other correlations that do not limit information. Ultimately, however, the main reason to look for such correlations is to determine whether information is limited in a particular area. Thus, a better approach is to measure information directly by decoding the neural activity. Notably, this decoding must be done on simultaneously recorded neurons, and not on sets of neurons recorded on different trials. In the latter case, even if one fills in the missing entries in the covariance matrix, it is very likely that the critical differential component will be missed, and the information will therefore be radically overestimated.

These results have an important implication: the effect of attention or perceptual learning on information in a particular area cannot be assessed by simply measuring pairwise correlations. Thus, the fact that pairwise correlations decrease when attention is engaged, which has been observed in several studies[10,11], cannot be taken as evidence that information must have gone up, nor as a mechanism that can explain any improved performance. The only way to determine the true effect of correlations in these kinds of studies is to record from large populations of neurons in parallel and decode the neural activity with and without attention, or before and after perceptual learning.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

R.M.-B. was supported by the Ramón y Cajal Spanish Award RYC-2010-05952 and by the Marie Curie FP7-PEOPLE-2010-IRG grant PIRG08-GA-2010-276795. X.P. was supported in part by US National Institutes of Health grant T32DC009974. P.L. was supported by the Gatsby Charitable Foundation. A.P. was supported by a grant from the Swiss National Science Foundation (#31003A_143707), and a grant from the Human Frontier Science Program.

AUTHOR CONTRIBUTIONS
R.M.-B., J.B., P.L. and A.P. conceived the project. R.M.-B., J.B., I.K., X.P., P.L. and A.P. developed the theory and wrote the manuscript.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Tolhurst, D.J., Movshon, J.A. & Dean, A.D. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res.* **23**, 775–785 (1983).
2. Averbeck, B.B., Latham, P.E. & Pouget, A. Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* **7**, 358–366 (2006).
3. Zohary, E., Shadlen, M.N. & Newsome, W.T. Correlated neuronal discharge rate and its implication for psychophysical performance. *Nature* **370**, 140–143 (1994).
4. Maynard, E.M. *et al.* Neuronal interactions improve cortical population coding of movement direction. *J. Neurosci.* **19**, 8083–8093 (1999).
5. Averbeck, B.B. & Lee, D. Neural noise and movement-related codes in the macaque supplementary motor area. *J. Neurosci.* **23**, 7630–7641 (2003).
6. Gu, Y. *et al.* Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron* **71**, 750–761 (2011).
7. Adibi, M., McDonald, J.S., Clifford, C.W. & Arabzadeh, E. Adaptation improves neural coding efficiency despite increasing correlations in variability. *J. Neurosci.* **33**, 2108–2120 (2013).
8. Abbott, L.F. & Dayan, P. The effect of correlated variability on the accuracy of a population code. *Neural Comput.* **11**, 91–101 (1999).
9. Renart, A. *et al.* The asynchronous state in cortical circuits. *Science* **327**, 587–590 (2010).
10. Cohen, M.R. & Maunsell, J.H. Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* **12**, 1594–1600 (2009).
11. Mitchell, J.F., Sundberg, K.A. & Reynolds, J.H. Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* **63**, 879–888 (2009).
12. Ecker, A.S. *et al.* Decorrelated neuronal firing in cortical microcircuits. *Science* **327**, 584–587 (2010).
13. Sompolinsky, H., Yoon, H., Kang, K. & Shamir, M. Population coding in neuronal systems with correlated noise. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **64**, 051904 (2001).
14. Yoon, H. & Sompolinsky, H. The effect of correlations on the Fisher information of population codes. in *Advances in Neural Information Processing Systems* (eds. Kearns, M.S., Solla, S. & Cohn, D.A.) 167–173 (MIT Press, Cambridge, Massachusetts, 1999).
15. Shamir, M. & Sompolinsky, H. Implications of neuronal diversity on population coding. *Neural Comput.* **18**, 1951–1986 (2006).
16. Ecker, A.S., Berens, P., Tolias, A.S. & Bethge, M. The effect of noise correlations in populations of diversely tuned neurons. *J. Neurosci.* **31**, 14272–14283 (2011).
17. Papoulis, A. *Probability, Random Variables and Stochastic Process* (McGraw-Hill, New York, 1991).
18. Paradiso, M.A. A theory of the use of visual orientation information which exploits the columnar structure of striate cortex. *Biol. Cybern.* **58**, 35–49 (1988).
19. Graf, A.B., Kohn, A., Jazayeri, M. & Movshon, J.A. Decoding the activity of neuronal populations in macaque primary visual cortex. *Nat. Neurosci.* **14**, 239–245 (2011).
20. Berens, P. *et al.* A fast and simple population code for orientation in primate V1. *J. Neurosci.* **32**, 10618–10626 (2012).
21. Maimon, G. & Assad, J.A. Beyond Poisson: increased spike-time regularity across primate parietal cortex. *Neuron* **62**, 426–440 (2009).
22. Gershon, E.D., Wiener, M.C., Latham, P.E. & Richmond, B.J. Coding strategies in monkey V1 and inferior temporal cortices. *J. Neurophysiol.* **79**, 1135–1144 (1998).
23. Qi, X.L. & Constantinidis, C. Variability of prefrontal neuronal discharges before and after training in a working memory task. *PLoS ONE* **7**, e41053 (2012).
24. Churchland, M.M. *et al.* Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat. Neurosci.* **13**, 369–378 (2010).
25. Beck, J., Bejjanki, V.R. & Pouget, A. Insights from a simple expression for linear fisher information in a recurrently connected population of spiking neurons. *Neural Comput.* **23**, 1484–1502 (2011).
26. Churchland, M.M. & Shenoy, K.V. Temporal complexity and heterogeneity of single-neuron activity in premotor and motor cortex. *J. Neurophysiol.* **97**, 4235–4257 (2007).
27. Berens, P., Ecker, A.S., Gerwinn, S., Tolias, A.S. & Bethge, M. Reassessing optimal neural population codes with neurometric functions. *Proc. Natl. Acad. Sci. USA* **108**, 4423–4428 (2011).
28. Beck, J.M., Ma, W.J., Pitkow, X., Latham, P.E. & Pouget, A. Not noisy, just wrong: the role of suboptimal inference in behavioral variability. *Neuron* **74**, 30–39 (2012).
29. Shadlen, M.N. & Newsome, W.T. The variable discharge of cortical neurons: Implications for connectivity, computation and information coding. *J. Neurosci.* **18**, 3870–3896 (1998).
30. Romo, R., Hernandez, A., Zainos, A. & Salinas, E. Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. *Neuron* **38**, 649–657 (2003).
31. Goris, R.L., Movshon, J.A. & Simoncelli, E.P. Partitioning neuronal variability. *Nat. Neurosci.* **17**, 858–865 (2014).
32. Ecker, A.S. *et al.* State dependence of noise correlations in macaque primary visual cortex. *Neuron* **82**, 235–248 (2014).
33. Kohn, A. & Smith, M.A. Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J. Neurosci.* **25**, 3661–3673 (2005).
34. Wimmer, K., Nykamp, D.Q., Constantinidis, C. & Compte, A. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat. Neurosci.* **17**, 431–439 (2014).
35. Huang, X. & Lisberger, S.G. Noise correlations in cortical area MT and their potential impact on trial-by-trial variation in the direction and speed of smooth-pursuit eye movements. *J. Neurophysiol.* **101**, 3012–3030 (2009).
36. Ponce-Alvarez, A., Thiele, A., Albright, T.D., Stoner, G.R. & Deco, G. Stimulus-dependent variability and noise correlations in cortical MT neurons. *Proc. Natl. Acad. Sci. USA* **110**, 13162–13167 (2013).
37. Newsome, W.T., Britten, K.H. & Movshon, J.A. Neuronal correlates of a perceptual decision. *Nature* **341**, 52–54 (1989).
38. Shadlen, M.N., Britten, K.H., Newsome, W.T. & Movshon, T.A. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J. Neurosci.* **16**, 1486–1510 (1996).
39. Series, P., Latham, P. & Pouget, A. Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations. *Nat. Neurosci.* **7**, 1129–1135 (2004).
40. Barlow, H. Redundancy reduction revisited. *Network* **12**, 241–253 (2001).

ONLINE METHODS

Here we describe the recurrent network of leaky integrate-and-fire neurons used in our simulations, and compute information analytically in that network in the limit of long time windows and small leak. We then switch to a more theoretical topic and show that differential correlations, and only differential correlations, limit information in large networks.

A network of leaky integrate-and-fire neurons. We computed the mean and covariance of the response of a network of N leaky integrate-and-fire neurons. Our starting point is the time evolution equation for the membrane potential, denoted V_i. For the leaky integrate-and-fire neuron, this is given by

$$\frac{dV_i}{dt} = -\frac{V_i}{\tau_m} + \sum_j J_{ij} \sum_l h_{ij}(t - t_j^l) + g_i(s) + \sum_k M_{ik}\xi_k(t) \tag{6}$$

Here τ_m is the membrane time constant (set to 20 ms), J_{ij} is the connectivity matrix (so the sum over j runs from 1 to N), t_j^l is the time of the lth spike of neuron j, $h_{ij}(t - t_j^l)$ is the synaptic response to that spike (its properties are given in equation (11) and preceding text, and the form for $h_{ij}(t)$ that we use in our simulations is given in **Supplementary Modeling**), $g_i(s)$ is the mean synaptic drive, which depends on a (potentially multi-dimensional, but time-independent) stimulus, s, and $\xi_k(t)$ is time-dependent Gaussian noise

$$\langle \xi_k(t)\xi_{k'}(t')\rangle = \delta_{kk'}C(t - t') \tag{7}$$

The autocorrelation function, $C(\tau)$, is chosen to integrate to 1,

$$\int_{-\infty}^{\infty} d\tau\, C(\tau) = 1 \tag{8}$$

and $\delta_{kk'}$ is the Kronecker delta. The last term in equation (6) corresponds to a mixing of noises, where M_{ik} is an arbitrary mixing matrix, taken in our simulations to consist of both independent and shared noise; the latter is taken into account by letting $M_{ik} = \sigma_{ind}\delta_{ik} + \sigma_s\delta_{k0}$. Consequently,

$$\sum_k M_{ik}\xi_k(t) = \sigma_{ind}\xi_i(t) + \sigma_s\xi_0(t) \tag{9}$$

The neuron emits a spike when the voltage reaches a threshold, denoted θ_i for neuron i, after which the voltage is reset to V_r, which we take to be 0. To take care of the reset, we introduce a negative self-current,

$$\begin{aligned} J_{ii} &= -\theta_i \\ h_{ii}(t) &= \delta(t) \end{aligned} \tag{10}$$

where $\delta(t)$ is the Dirac delta function. When $i \neq j$, $h_{ij}(t)$ corresponds to a brief current pulse; it is zero when $t < 0$ and, for convenience we choose it so that it integrates to 1,

$$\int_0^{\infty} dt\, h_{ij}(t) = 1 \tag{11}$$

This gives $h_{ij}(t)$ units of inverse time, and so J_{ij} has units of voltage. For now, we take J_{ij} to be an arbitrary matrix. In our simulations, however, we consider excitatory-inhibitory networks.

Fisher information for long time windows and small leak. Although we would like to compute Fisher information analytically for this network, that is not, as far as we know, possible. However, there is one limit in which we can compute it: the observation time window is long, and the leak time constant, τ_m, is infinite (corresponding to a non-leaky integrate-and-fire neuron). We consider that limit here.

Our starting point is to compute spike count in a window of size T. This is straightforward: simply integrate both sides of equation (6) from 0 to T. This gives

$$\Delta V_i(T) = \sum_j J_{ij} \int_0^T dt \sum_l h_{ij}(t - t_j^l) + Tg_i(s) + \sum_k M_{ik} \int_0^T dt\, \xi_k(t) \tag{12}$$

where $\Delta V_i(T) \equiv V_i(T) - V_i(0)$. The first integral is approximately the spike count at time T, denoted $n_j(T)$ for neuron j,

$$\int_0^T dt \sum_l h_{ij}(t - t_j^l) = n_j(T) + \delta n_j(T) \tag{13}$$

The term $\delta n_j(t)$ comes from the fact that $h_{ij}(t)$ has finite width in time (if $i \neq j$). Note that δn_j is at most 1. The mean of the second integral is zero, and, using equation (7), its variance is given by

$$\mathrm{Var}\left[\int_0^T dt\, \xi_k(t)\right] = \int_0^T dt \int_0^T dt' C(t - t') \tag{14}$$

If the correlation time is zero (corresponding to delta-correlated white noise), the double integral on the right is just T. However, for finite correlation time, there is a small reduction due to edge effects. To take that into account, we write

$$\mathrm{Var}\left[\int_0^T dt\, \xi_k(t)\right] = T_c \tag{15}$$

where T_c is smaller than T by about the correlation time. Thus,

$$\int_0^T dt\, \xi_k(t) = T_c^{1/2}\eta_k \tag{16}$$

where η_k are a set of uncorrelated, zero mean, unit variance Gaussian random variables,

$$\langle \eta_k \eta_{k'} \rangle = \delta_{kk'} \tag{17}$$

Inserting equations (13) and (16) into equation (12), yields

$$0 = \sum_j J_{ij} n_j(T) + Tg_i(s) + T_c^{1/2} \sum_k M_{ik}\eta_k + \gamma_i(T) \tag{18}$$

where the noise term, γ_i, is given by

$$\gamma_i(T) \equiv \sum_j J_{ij}\delta n_j(T) - \Delta V_i(T) \tag{19}$$

How big is $\gamma_i(T)$? In the large T limit, there are typically a large number of spikes, so δn_j is small compared to n_j. Moreover, ΔV_i is at most θ_i. Consequently, in this limit, $\gamma_i(T)$ is usually small compared with T. However, it isn't always small: although ΔV_i can be at most θ_i, it can be negative, and it can be large and negative. This happens whenever a neuron receives a consistently negative current, in which case $\Delta V_i(T)$ is proportional to $-T$. We can partially solve this problem by simply ignoring any neuron that doesn't spike. Unfortunately, this is only a partial solution, because it's possible for a neuron to fire once or twice due to noise, and then for the voltage to steadily decrease with time; this would again mean $\Delta V_i \propto -T$ in the large T limit. This can be taken care of by ignoring the transients, which we effectively do in our simulations: we run one very long simulation, divide them into intervals of either 2 or 10 s, and collect spike counts in those intervals.

Assuming that the connectivity matrix, J_{ij}, is invertible, we can solve equation (18) directly for spike count. To simplify notation we switch to vectors and matrices, for which we use bold font. Multiplying both sides of equation (18) by $\mathbf{J}^{-1}$, we have

$$\mathbf{n} = -T\mathbf{J}^{-1}\mathbf{g}(s) - \mathbf{J}^{-1}(T_c^{1/2}\mathbf{M}\boldsymbol{\eta} + \boldsymbol{\gamma}(T)) \tag{20}$$

The mean and covariance of the spike count are therefore given by

$$\begin{aligned} \mathrm{Mean}[\mathbf{n}] &= -\mathbf{J}^{-1}\left[T\mathbf{g}(s) + \langle \boldsymbol{\gamma} \rangle\right] \\ \mathrm{Covar}[\mathbf{n}] &= T_c\mathbf{J}^{-1}\left[\mathbf{M}\mathbf{M}^{\mathrm{T}} + T_c^{-1/2}(\mathbf{M}\langle \boldsymbol{\eta}\boldsymbol{\gamma}^{\mathrm{T}} \rangle + \langle \boldsymbol{\gamma}\boldsymbol{\eta} \rangle^{\mathrm{T}}\mathbf{M}^{\mathrm{T}}) + T_c^{-1}\langle \boldsymbol{\gamma}\boldsymbol{\gamma}^{\mathrm{T}} \rangle\right]\mathbf{J}^{-\mathrm{T}} \end{aligned} \tag{21}$$

where the superscript –T denotes transpose and inverse, and, for clarity, we suppress the fact that $\boldsymbol{\eta}$ and $\boldsymbol{\gamma}$ depend on T.

doi:10.1038/nn.3807

Linear Fisher information. Our next step is to compute the linear Fisher information, both in the input and in the output. In general, for a random variable with stimulus-dependent mean $\mathbf{f}(s)$ and covariance matrix $\Sigma(s)$, the linear Fisher information is given by[41]

$$I = \mathbf{f}'(s)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(s) \mathbf{f}'(s) \tag{22}$$

where a prime denotes a derivative with respect to s (if s were a vector, I would be the Fisher information matrix, but here we restrict ourselves to scalars). Using the expression of the covariance in equation (21), and working to lowest non-vanishing order in $1/T^{1/2}$, we find that the linear Fisher information in the spike train, denoted I_{out}, is given by

$$I_{\mathrm{out}} = T\mathbf{g}'(s)^{\mathrm{T}} \left[\mathbf{M}\mathbf{M}^{\mathrm{T}} + O(T^{-1/2}) \right]^{-1} \mathbf{g}'(s) \tag{23}$$

How does this compare to the information in the input? The latter is the information available to an observer that has direct access to the current, $\mathbf{g}(s) + \mathbf{M}\boldsymbol{\xi}$. After observing this current for time T, the mean is $T\mathbf{g}(s)$ and, to lowest non-vanishing order in $1/T$, the variance is $T\mathbf{M}\mathbf{M}^{\mathrm{T}}$. Thus, using equation (22), the Fisher information in the input, denoted I_{in}, is given by

$$I_{\mathrm{in}} = T\mathbf{g}'(s)^{\mathrm{T}} [\mathbf{M}\mathbf{M}^{\mathrm{T}}]^{-1} \mathbf{g}'(s) \tag{24}$$

This is almost the same as the Fisher information in the output, I_{out}. The difference is that there is extra noise, captured by the $O(T^{-1/2})$ correction, associated with spikes. Essentially, for small times, spikes inject additional noise into the estimate of the mean, but as time increases, that noise diminishes. Note, however, that in general there is one more potential loss of information: if some neurons never spike, they need to be taken out of the network. See, for example, the comments following equation (28).

Here, and in the simulations, we consider a simple model in which the noise is given in equation (9), so that

$$\mathbf{M}\mathbf{M}^{\mathrm{T}} = \sigma_{ind}^2 \mathbf{I} + \sigma_s^2 \mathbf{1}\mathbf{1}^{\mathrm{T}} \tag{25}$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{1}$ is a vector consisting of all 1's (so $\mathbf{1}\mathbf{1}^T$ is a matrix with all elements equal to 1), and the signal, $\mathbf{g}(s)$, is given by

$$\mathbf{g}(s) = s\mathbf{1} \tag{26}$$

In this case, as is straightforward to show, the input information is given by

$$I_{\mathrm{in}} = \frac{T}{\sigma_s^2 + \sigma_{ind}^2 / N} \tag{27}$$

The output information, I_{out}, is only slightly smaller,

$$I_{\mathrm{out}} = \frac{T}{\sigma_s^2 + \sigma_{ind}^2 / N + O(T^{-1/2})} \tag{28}$$

Note that here N should really be the number of neurons that fire. However, assuming that number is large, it doesn't really matter what it is—the Fisher information is determined mainly by the shared noise, σ_s^2.

Equations (27) and (28) correspond to equations (1) and (2), except that in the main text we report information rates, so we divide by T.

Theoretical analysis of differential correlations. In this section we leave neural networks, and turn to theoretical analysis of differential correlations. We analyze information when there is a 'pure' $\mathbf{f}'\mathbf{f}'^{\mathrm{T}}$ component and, just as importantly, when there is a not so pure component. We show that in the former case information saturates with N; in the latter case it doesn't. We also show, somewhat surprisingly, that the optimal decoder doesn't need to know about the $\mathbf{f}'\mathbf{f}'^{\mathrm{T}}$ component of the correlations. In the **Supplementary Modeling**, we provide further insight into differential correlations by expressing them in terms of the eigenvectors and eigenvalues of the covariance matrix, and we use that analysis to understand why, and when, it's hard to accurately estimate Fisher information.

Here we ask how the linear Fisher information scales with the number of neurons, N, when the covariance matrix contains a pure $\mathbf{f}'\mathbf{f}'^{\mathrm{T}}$ component (the second term in equation (31)). Our starting point is a covariance matrix, $\Sigma_0(s)$, that doesn't necessarily contain an $\mathbf{f}'\mathbf{f}'^{\mathrm{T}}$ component. As in equation (3), the (linear) Fisher information associated with $\Sigma_0(s)$, denoted I_0, is given by

$$I_0 = \mathbf{f}'(s)^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1}(s) \mathbf{f}'(s) \tag{29}$$

where, as usual, $\mathbf{f}(s)$ is a vector of tuning curves,

$$\mathbf{f}(s) \equiv (f_1(s), f_2(s), ..., f_N(s))^{\mathrm{T}} \tag{30}$$

and a prime denotes a derivative with respect to s. Note that the information also depends on stimulus, s; we suppress that dependence for clarity. To add a pure $\mathbf{f}'\mathbf{f}'^{\mathrm{T}}$ component, we define a new covariance matrix, $\boldsymbol{\Sigma}_\epsilon(s)$, via

$$\boldsymbol{\Sigma}_\epsilon(s) = \boldsymbol{\Sigma}_0(s) + \epsilon \mathbf{f}'(s) \mathbf{f}'^{\mathrm{T}}(s) \tag{31}$$

The new information, denoted I_ϵ, is given by

$$I_\epsilon = \mathbf{f}'(s)^{\mathrm{T}} \boldsymbol{\Sigma}_\epsilon^{-1}(s) \mathbf{f}'(s) \tag{32}$$

To compute I_ϵ, we need the inverse of $\boldsymbol{\Sigma}_\epsilon$. As is easy to verify, this inverse is given by

$$\boldsymbol{\Sigma}_\epsilon^{-1}(s) = \boldsymbol{\Sigma}_0^{-1}(s) - \frac{\epsilon}{1 + \epsilon I_0} \boldsymbol{\Sigma}_0^{-1}(s) \mathbf{f}'(s) \mathbf{f}'^{\mathrm{T}}(s) \boldsymbol{\Sigma}_0^{-1}(s) \tag{33}$$

Inserting equation (33) into (32), we arrive at

$$I_\epsilon = I_0 - \frac{\epsilon I_0^2}{1 + \epsilon I_0} = \frac{I_0}{1 + \epsilon I_0} \tag{34}$$

which is equation (5).

Perhaps surprisingly, although $\mathbf{f}'\mathbf{f}'^{\mathrm{T}}$ correlations have a critical role in determining information, they are irrelevant for decoding, in the sense that they have no effect on the locally optimal linear estimator. To see this explicitly, note first of all that the locally optimal linear estimator, denoted $\mathbf{w}^{\mathrm{T}}$, generates an estimate of the stimulus near some particular value, s_0, by linearly operating on neural activity,

$$\hat{s} = s_0 + \mathbf{w}^{\mathrm{T}} (\mathbf{r} - \mathbf{f}(s_0)) \tag{35}$$

In the presence of the covariance matrix given in equation (31), the optimal weight, $\mathbf{w}_{opt}^{\mathrm{T}}$ is given by

$$\mathbf{w}_{opt}^{\mathrm{T}} = \frac{\mathbf{f}'^{\mathrm{T}} (\boldsymbol{\Sigma}_0 + \epsilon \mathbf{f}'\mathbf{f}'^{\mathrm{T}})^{-1}}{\mathbf{f}'^{\mathrm{T}} (\boldsymbol{\Sigma}_0 + \epsilon \mathbf{f}'\mathbf{f}'^{\mathrm{T}})^{-1} \mathbf{f}'} \tag{36}$$

where we have dropped, for clarity, the explicit dependence on s_0. Using equation (33), this reduces to

$$\mathbf{w}_{opt}^{\mathrm{T}} = \frac{\mathbf{f}'^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1}}{\mathbf{f}'^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1} \mathbf{f}'} \tag{37}$$

Thus, the locally optimal linear decoder does not need to know the size of the $\mathbf{f}'\mathbf{f}'^{\mathrm{T}}$ correlations.

In hindsight this makes sense: $\mathbf{f}'\mathbf{f}'^{\mathrm{T}}$ correlations shift the hill of activity, and there is, quite literally, nothing any decoder can do about this. This suggests that these correlations are in some sense special. To determine just how special, we ask what happens when we add correlations in a different direction—say correlations of the form $\mathbf{u}\mathbf{u}^{\mathrm{T}}$, where $\mathbf{u}$ is not parallel to $\mathbf{f}'$. In that case, the covariance matrix becomes (with a normalization added for convenience only)

$$\boldsymbol{\Sigma}_u(s) = \boldsymbol{\Sigma}_0(s) + \epsilon \frac{\mathbf{f}'(s)^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1}(s) \mathbf{f}'(s)}{\mathbf{u}^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1}(s) \mathbf{u}} \mathbf{u}\mathbf{u}^{\mathrm{T}} \tag{38}$$

Repeating the steps leading to equation (34), we find that

$$I_u \equiv \mathbf{f}'(s)^{\mathrm{T}} \boldsymbol{\Sigma}_u^{-1}(s) \mathbf{f}'(s) = I_0 \sin^2\theta + \frac{I_0 \cos^2\theta}{1 + \epsilon I_0} \tag{39}$$

where I_0 is defined in equation (29) and

$$\cos\theta \equiv \frac{\mathbf{f}'(s)^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1}(s) \mathbf{u}}{\left[\mathbf{f}'(s)^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1}(s) \mathbf{f}'(s) \mathbf{u}^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1}(s) \mathbf{u}\right]^{1/2}} \tag{40}$$

Whenever $\theta \neq 0$—meaning $\mathbf{u}$ is not parallel to $\mathbf{f}'(s)$—information does not saturate as N goes to infinity. Thus, in the large N limit, $\mathbf{f}'(s)\mathbf{f}'(s)^{\mathrm{T}}$ correlations are the only ones that cause saturation.

A **Supplementary Methods Checklist** is available.

41. Beck, J., Bejjanki, V.R. & Pouget, A. Insights from a simple expression for linear Fisher information in a recurrently connected population of spiking neurons. *Neural Comput.* **23**, 1484–1502 (2011).

doi:10.1038/nn.3807