



Published in final edited form as:

Nature. 2010 July 29; 466(7306): 617–621. doi:10.1038/nature09178.

Sparse coding and high-order correlations in fine-scale cortical networks

Ifije E. Ohiorhenuan, Ferenc Mechler, Keith P. Purpura, Anita M. Schmid, Qin Hu, and Jonathan D. Victor

Department of Neurology and Neuroscience, Weill Cornell Medical College, New York, 10065, USA

Abstract

Connectivity in the cortex is organized at multiple scales¹⁻⁵, suggesting that scale-dependent correlated activity is particularly important for understanding the behavior of sensory cortices and their function in stimulus encoding. Here, we analyze the scale-dependent structure of cortical interactions by using maximum entropy models⁶⁻⁹ to characterize multiple-tetrode recordings from primary visual cortex of anesthetized monkeys (*Macaca mulatta*). We compare the properties of firing patterns among local clusters of neurons (<300 microns) with neurons separated by larger distances (600-2500 microns). We find that local firing patterns are distinctive: while multi-neuronal firing patterns at larger distances can be predicted by pairwise interactions, patterns within local clusters often show evidence of high-order correlations. Surprisingly, these local correlations are flexible and rapidly reorganized by visual input. While they modestly reduce the amount of information that a cluster conveys, they also modify the *format* of this information, creating sparser codes by increasing the periods of total quiescence, and concentrating information into briefer periods of common activity. These results imply a hierarchical organization of neuronal correlations: simple pairwise correlations link neurons over scales of tens to hundreds of minicolumns, but on the scale of a few minicolumns, ensembles of neurons form complex subnetworks whose moment-to-moment effective connectivity is dynamically reorganized by the stimulus.

Early cortical sensory areas create internal representations of the sensory world. At the level of individual neurons, this process is reasonably well understood. For instance, in the primary visual cortex (V1) neurons respond selectively to components or features of the sensory stimulus, such as orientation or spatial frequency. But, because the activity of pairs of cortical neurons is correlated¹⁰, the behavior of a network of cortical neurons cannot be fully understood from measurements of its individual responses. Understanding the functional role of correlations among groups of neurons is challenging because of the combinatorial explosion of possible interactions. However, the organization of cortical

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Author contributions. I.E.O. conceived the project and carried out the data analysis. I.E.O. and J.D.V. wrote the manuscript. J.D.V. supervised the project. I.E.O., F.M., K.P.P., A.S., Q.H., and J.D.V. collected experimental data. F.M., K.P.P., and A.S. provided feedback on the manuscript.

connectivity suggests that certain types of interactions are particularly relevant to cortical processing.

A striking anatomical feature of the neocortex is that connectivity between neurons is highly structured. Across the cortical sheet, neurons are organized over a range of spatial scales: fine scale networks (50-100 μ m) display specific, nonrandom connectivity¹⁻³. Neurons with similar responses are grouped into functional columns which span several hundred microns⁴, and long range horizontal connections link neurons together over several millimeters^{5, 11}. The prominence of this multi-scale organization argues that scale-dependent interactions between neurons shape the behavior of cortical networks, and the manner in which they encode sensory information. This view predicts that cortical neurons participate in multiple subnetworks whose characteristics vary with spatial scale.

Directly addressing this question requires *in vivo* sampling (with high temporal resolution) of neuronal populations at different spatial scales and a principled way to characterize multi-neuron activity. To do this, we combine multiple-tetrode recording with maximum entropy models^{6,9}. Multiple-tetrode recording in macaque primary visual cortex (Fig. 1a) enables sampling of cortical activity at different scales: each tetrode isolates several neurons within a radius of approximately 150 microns¹², and we separate the tetrodes by distances ranging from 600 microns to several millimeters. A complete characterization of the activity of a network of neurons is challenging, because the number of potential interactions grows exponentially as the network size increases. Even for small networks, it is infeasible to make enough measurements to accurately estimate multi-neuron joint histograms. Here, we record from small groups of neurons (3 to 6) and use maximum entropy models^{6,9} to provide an insightful summary of the many possible interactions between them. In the simple case that neurons in a population are independent, the frequency of joint activity of any set of neurons can be predicted from the product of the individual neurons' mean firing rates. Maximum entropy models allow us to explore more complex hypotheses, such as, whether pairwise interactions account for the observed firing patterns. For example, given recordings from three neurons, we can ask if the frequency of triplet firing (the pattern – '111') is predicted by the frequency of pairs of neurons firing (the patterns '011', '110' and '101'). Thus, we characterize a network by asking what kinds of simple interactions are sufficient to predict the observed distribution of firing patterns.

To test for the existence of scale-dependent functional subnetworks, we present a visual stimulus of a binary checkerboard stimulus, pseudorandom in space and time (Fig. 1b) and record from several neurons (typically 3) isolated on one or more tetrodes. We choose this stimulus because its lack of spatiotemporal correlations minimizes the possibility that neuronal correlations are merely driven by correlations within the stimulus itself⁶. We bin multi-neuron firing patterns (10-15 ms bins, depending on the frame rate of the stimulus) to create a distribution of firing pattern counts (Fig. 1c). We ask how well the joint activity of these neurons is predicted by two models that have proved useful for studying the retina^{6, 8}: an independent model, M_{ind} , which assumes that neurons are independent, and a pairwise model, M_{pair} , which takes into account interactions between pairs of neurons. Fig. 2 shows the extent to which these models account for observed firing patterns, both in terms of predicting the firing rates of specific configurations (left panels), and in terms of an overall

measure of goodness of fit (right panels). Although M_{ind} (Fig. 2a,b) fits well for a fraction of recordings, it often fails for distances $>300 \mu\text{m}$ and nearly always fails for nearby ($<300 \mu\text{m}$) recordings. In contrast, for distances $>300 \mu\text{m}$, M_{pair} fits well in 79/80 triplet recordings (Fig. 2c,d), but surprisingly, it often fits poorly for nearby recordings: 15/38 recordings have a log-likelihood ratio per minute of < -5 (5/23 triplets and 10/15 groups of four, five or six neurons - Supp. Fig. 6). A log-likelihood ratio per minute of -5 means that on average, after 60 seconds of data, neuronal responses are 32 times more likely (2^5) to have come from a perfect model (M_{obs}) than from M_{pair} . These results are robust to errors in estimating joint firing activity due to spike sorting and other statistical artifacts (see Supplementary Information), and the degree of failure was uncorrelated with the similarity of orientation tuning of the neurons (data not shown).

Our finding, that pairwise correlations account for the multi-neuronal activity of neurons separated by several hundred microns, is consistent with previous studies from area 17 of the anesthetized cat⁹ and studies of *ex vivo* retinal^{6, 8} and cortical tissue⁷. However, the failure of the pairwise model for nearby cortical neurons is novel and implies that complex local interactions distinguish the behavior of local cortical networks. The difference between local and long-range patterns of correlation shows that complex fine-scale anatomical connectivity¹⁻³ has an observable effect on network firing patterns.

To determine if local correlations play a role in encoding visual information, we analyzed how visual input affects the network behavior. We extend the maximum entropy approach to incorporate stimulus-dependent interactions by examining how the population firing pattern depends on the state of individual pixels within the overlap of the three neurons' receptive fields. We select the individual spatiotemporal pixel that maximally modulates the population response (Fig. 3a), subdivide the data into halves by conditioning on each of the two states of this "maximally informative pixel," and fit the pairwise maximum entropy model (M_{pair}) to each half to the data.

For each recording site, this procedure generates a fit of M_{pair} to each stimulus condition (pixel ON/OFF). In Fig. 3b we plot these fits against the fit of M_{pair} without stimulus conditioning. Typically, conditioning on maximally informative pixels (blue dots) often significantly improves the fit of M_{pair} for one pixel state, and simultaneously worsens the fit for the other pixel state. (As a control for the statistical effects of halving the data available for each fit, we also conditioned on random pixels (red dots); this has a negligible effect on the fit of M_{pair}). If stimulus-varying *pairwise* correlations could account for the network correlation patterns, we would have found a very different result: conditioning on the maximally informative pixel, the main determinant of visual responsiveness, would *improve* the fit of M_{pair} for both states, rather than improve it for one state, and worsen it for the other (as we observe). Further conditioning (on the second- or third- most-informative pixels) continues to reveal data subsets in which the pairwise model fit worsens, but analysis of further conditioning is limited by the successive halving of the amount of data available to build the models (data not shown).

Our observation that the extent of failure of M_{pair} depends on the pixel state (Fig.3b,c), suggests that the effective connectivity¹⁰ of local networks dynamically depends on visual

input, and can be modulated on a frame-by-frame basis (This need not mean changes in actual connections; more likely, it is the functional result of nonlinear interactions between the stimulus and the network.) To further support the notion that effective connectivity is rapidly modulated, we compare the interaction strengths between neurons when the neural response is conditioned on each of the two states of a pixel. These interaction strengths – the interaction parameters of the pairwise model – reflect correlated firing between two neurons which cannot be attributable to a third (unlike a peak in a cross-correlogram)⁸. As a robust measure of the dependence of functional network connectivity on the two states of the pixel, we compared the mean interaction strengths of the pairwise models fit to each condition (see Methods). As shown in Fig. 3d, conditioning on maximally informative pixels (blue bars), but not random pixels (red bars), can substantially modulate overall functional connectivity (see also Supp. Fig. 7)

Since multi-neuronal correlations form dynamically even in response to spatiotemporally uncorrelated stimuli, we hypothesized that multineuronal correlations will be even stronger when stimuli contain correlations – as do naturalistic stimuli. Supp. Fig. 1 shows that this is indeed the case. For neurons at $<300\ \mu\text{m}$, M_{pair} fails for 20/46 sites; overall, the fit of M_{pair} is worse for naturalistic stimuli (mean log likelihood ratio per minute, -11.7) than for pseudorandom ones (mean, -6.9). Moreover, M_{pair} occasionally fails at $600\ \mu\text{m}$ (22/481) for naturalistic stimuli (Supp. Fig. 1, right); no failures of M_{pair} were seen at this distance for pseudorandom stimuli (Fig. 2c,d). However, since the correlation structure of natural stimuli is spatially extensive and complex¹³, it is difficult to separate correlations that arise as a result of intrinsic network dynamics, from those which arise from simple (e.g. linear) filtering of the stimulus, or from nonlinear interactions that contours drive¹⁴.

Finally, we consider two key functional aspects of local correlations: their impact on the *amount* of information carried, and on the *format* of this information (i.e., the neural code). As described below, we find a substantial effect on the latter, but only a minor effect on the former.

To determine the impact of high-order interactions on the amount of information carried, we compared the mutual information between the informative pixels and the neural responses generated under M_{obs} and M_{pair} (see Methods). As shown in Fig. 4a, higher-than-second order correlations have little effect on the overall information content. However, comparing M_{pair} and M_{ind} (Fig. 4b) shows that there is a mild reduction in information content due to the second-order correlations (i.e., redundancy), as has been seen in previous studies in retina¹⁵, primary visual cortex^{16, 17}, and inferior temporal cortex¹⁸. Thus, while it has been suggested that fine-scale pairwise correlations might result in an increase in information content^{19, 20} (i.e., synergy), we find that redundancy dominates for larger neuronal populations – supporting the notion that it is a strategy the cortex employs to maintain the fidelity of information in the face of variable individual neural responses²¹, and that correlations do not increase the information conveyed by neurons²².

The effect of local correlations on the format of the visual information is represented is shown in Fig. 4c. Correlations sparsify the neural code – i.e., they decrease the fraction of time at which the population is active, without a proportional decrease in the amount of

information encoded. Specifically, as shown in Fig. 4c, the probability of total quiescence, p_{000} , is larger for M_{obs} than M_{ind} in local networks. This effect was driven nearly completely by pairwise correlations. However, information does not decrease proportionally. Instead, as shown in Fig. 4d, information during the non-quiescent periods is higher when pairwise (and higher) interactions are included. These functional consequences are specific to local correlations, and distinguish them from the longer-range correlations typically studied^{9, 11}.

We have analyzed correlations at three spatial scales, sampled from the continuum of scales that are present in cortex. The analysis shows that correlations in cortical networks have a specific scale-dependence. Fine-scale subnetworks are characterized by a prevalence of stimulus-dependent high order correlations and pairwise correlations which increase coding redundancy and response sparseness. In turn, these fine-scale networks are weakly synchronized by pairwise noise correlations at longer ranges. In contrast, responses of retinal networks to naturalistic stimuli⁸ and flickering checkerboards⁶ did not display high order correlations, and pairwise interactions nearly perfectly accounted for the behavior, even among adjacent neurons. Thus, complex scale-dependent patterns of correlations between neurons are an emergent property of cortical processing.

Cortical minicolumns have been proposed to form the smallest organizational unit in the cortex²³; in the macaque, they are approximately 40-60 μm in diameter²⁴. Since tetrodes typically isolate neurons up to of 70-150 μm ^{12, 25}, our measurements of local correlations reflect cortical processing that occurs on the scale of one to a few minicolumns. Our observation that stimulus-dependent correlations impact coding strengthens the concept that locally, minicolumns interact to form functional groups²⁴. Because, as we have shown, these interactions increase coding redundancy and concentrate the output of the network into short time epochs, they are potentially useful for transmitting information to higher order neurons in the face of noisy neuronal activity²¹ and frequent synaptic failures²⁶. Although we found that correlations at a scale of tens to hundreds of minicolumns produce significant interactions between pairs of neurons, the role of these correlations in cortical activity is still unclear. Correlations at these scales could reflect a global cortical state, such as that captured by electroencephalographic recordings. Alternatively, they may contribute to encoding of naturalistic visual input when the stimulus itself contains long-range correlations such as extended contours¹⁴, or high-order correlations¹³.

Methods Summary

Data

Recordings were made in V1 of 12 anesthetized macaque monkeys (*Macaca mulatta*)²⁷. Pseudorandom checkerboards²⁸ were presented at 67.6/100 Hz for 16-32 minutes.

Analysis

Spikes were sorted²⁹ and binned into 10 or 14.8 ms bins. Similar bin widths have been useful for exploring multi-neuronal correlations^{6,9}. Bins with two or more spikes (<3%) were replaced with one spike. A conservative spike count correction was applied to estimates of multi-neuron events from single tetrode recordings (see Methods).

Maximum Entropy Models

The models M_{ind} and M_{pair} have been previously described in detail^{6, 8}. Their performance was evaluated by the Kullback-Leibler divergence between the model-predicted firing pattern distributions and the observed distributions, M_{obs} , which also yields the log likelihood ratio between the maximum entropy models and a perfect model (M_{obs}).

Conditional maximum entropy models were similarly calculated based on firing probabilities that occurred following specific sets of stimuli. Stimuli were divided into sets based on the state of individual pixels. These (“maximally informative”) pixels were selected by the criterion that the mutual information between pixel state and firing pattern is maximized.

Encoding

The contribution of correlations to the mutual information between the neural response and the stimulus was evaluated by fitting M_{ind} and M_{pair} to firing patterns conditioned on the state of the maximally informative pixel. Mutual information in the absence of correlation was calculated as the Jensen-Shannon divergence between the two conditioned M_{ind} models. Mutual information with pairwise correlations included was calculated as the Jensen-Shannon divergence of the two conditioned M_{pair} models. We quantified sparseness by the frequency at which the network is silent (for three neurons, the ‘000’ pattern). Information transmitted when the network is active was measured by removing this “all-silent” firing pattern and calculating the Jensen-Shannon divergence between the remaining stimulus-conditioned firing patterns.

Methods

Stimulus

Pseudorandom binary checkerboards²⁸ at 100% contrast were presented at 67.6 Hz or 100 Hz. Each check typically subtended 0.25×0.25 degrees. 8-16 repeats of a 60 second stimulus were presented, along with its contrast inverse, for a total of 16-32 minutes at each recording site. Naturalistic stimuli consisted of vignettted natural movies and frame-shuffled natural movies sampling a range of natural environments and containing diverse sets of animals as well as man-made structures. Stimuli spanned the same spatial extent as the pseudorandom stimuli and were presented for 12 repeats at 100 Hz for a total of 20 minutes.

Physiology

Single and multi-tetrode extracellular recordings were made from V1 in 12 propofol/sulfentanil²⁷ anesthetized macaque monkeys (*Macaca mulatta*).

Analysis

Spikes were sorted using a principal components based algorithm²⁹, and binned into 10 or 14.8 ms bins, matching the frame rate of the stimulus. This bin width was chosen because pairs of V1 neurons are correlated on scales of tens of milliseconds; finer temporal resolution would reduce the accuracy of estimates of multi-neuron events. Similar bin

widths have been useful for exploring multi-neuronal correlations in the retina^{6, 8}, ex vivo cortex⁷ and cat area 17⁹.

Under pseudorandom stimulation, for single tetrode recordings (38 sites in 8 animals), 23 groups of 3 neurons, 8 groups of 4 neurons, 4 groups of 5 and 3 groups of 6 neurons were jointly analyzed. For multi-tetrode recordings at the 600 μm scale, 56 groups of 3 neurons (5 sites, 5 animals) were selected by choosing subsets of 3 neurons where two neurons were isolated on one tetrode and the other on a different tetrode. For recordings at $>1000 \mu\text{m}$, subsets of 3 neurons were chosen where each neuron was isolated on a separate tetrode (24 neurons in 5 sites, 4 animals).

Under naturalistic stimuli, the $<300 \mu\text{m}$ dataset consisted of 46 sites from 10 animals (28 groups of 3 neurons, 5 groups of 4 neurons, 5 groups of 5 neurons, and 8 groups of 6 neurons). The 600 μm dataset consisted of 25 recording sites from 7 animals and the $>1000 \mu\text{m}$ dataset consisted of 15 recording sites from 5 animals, with subsets of 3 neurons chosen as described above.

When multiple neurons are recorded on one tetrode, near-simultaneous spiking from multiple units can superimpose to generate waveforms that are not readily sorted. This prevents our software from detecting multiple spikes (at one tetrode) that occur within 1.2 ms. We correct this systematic underestimate as follows. First, we partition a 10 ms (or 14.8 ms) bin into $n=8$ (or $n=12$) slots of 1.2 ms, and assume that events are properly detected if they occur in separate slots, and are occluded (i.e., not detected) if they occur in the same slot. We then assume that within each analysis bin (10 or 14.8 ms), the k components of a multi-neuron spiking event will fall randomly into the n slots. For k simultaneously active

neurons, there are $\binom{n+k-1}{k}$ equally likely ways in which the spikes can fall into the n slots, but only $\binom{n}{k}$ are observable. We therefore multiply the observed occurrences of k -neuron events by the ratio of these two quantities, namely, $(n+1)/(n-1)$ for $k=2$, and $(n+2)(n+1)/(n-1)(n-2)$ for $k=3$. In the Supplement we consider alternative corrections that take into account tighter correlation at timescales of 1-2 ms than at 10 ms. These alternative corrections had little effect on the goodness of fit of the models considered (see Supp. Methods and Supp. Figs. 2-3).

Maximum Entropy Models

The maximum entropy models, M_{ind} and M_{pair} , have been previously described in detail^{6, 8}. Our implementation is similar. We solved for the maximum entropy distribution subject to the constraints of firing rate (M_{ind}) or firing rate and pairwise correlations (M_{pair}) using a gradient descent algorithm. For M_{pair} this procedure yields the Lagrange multipliers, h_i , which describes each neuron's intrinsic firing rate, and J_{ij} , which describes the strength of interaction between pairs of neurons.⁸ To characterize the strength of interactions between pairs of neurons for a recording site, we average J_{ij} over all possible pairs. We measure the effect of stimulus conditioning on functional network connectivity by taking the absolute value of the difference between the average J_{ij} in each stimulus condition (Fig. 3d). We

measure the goodness of fit as the Kullback-Leibler divergence between the model-predicted firing pattern distribution and the observed distributions (M_{obs}):

$D_{kl} M_{obs} || M_{model} = \sum_i m_{obs_i} \log_2 \left(\frac{m_{obs_i}}{m_{model_i}} \right)$, where m_{obs_i} is the observed probability of the i th firing pattern, and m_{model_i} is the corresponding prediction from a maximum entropy model. We calculate the \log_2 of the likelihood ratio (LLR) per minute between the maximum entropy models and the observed distributions (M_{obs}) via $\langle LLR \rangle = B - D_{kl} M_{obs} || M_{model}$, where B is the number of bins in 60 s (as in ⁶).

Conditional maximum entropy models

The above analysis was extended to determine maximum entropy distributions conditional on the state of a stimulus pixel. Pixels were chosen either at random, or to have maximal influence on the firing patterns (“maximally informative” pixels). The maximally informative pixels were identified as follows. For each stimulus pixel and each time lag τ (0-120 ms (10 ms bins) or 0-178 ms (14.8 ms bins)), we determined the distribution of firing patterns at a time τ after the pixel was ON or OFF. This yielded two conditional distributions: $P(r | s_{ON})$ and $P(r | s_{OFF})$. (M_{obs} is a 50:50 mixture of $P(r | s_{ON})$ and $P(r | s_{OFF})$, since the probability of a pixel being ON or OFF was 0.5). In the specific (and present) case that the two pixel states are equally likely, the mutual information $I(S,R)$ between the state of the pixel and the response is equal to the Jensen-Shannon divergence between the two conditional distributions ²¹:

$$I(S, R) = D_{js} P(r | s_{ON}) || P(r | s_{OFF}) = \frac{1}{2} D_{kl} P(r | s_{ON}) || M_{obs} + \frac{1}{2} D_{kl} P(r | s_{OFF}) || M_{obs}$$

Thus, the maximally informative pixel (in the sense of greatest mutual information between pixel state and firing pattern distribution) is also the pixel for which the two conditional distributions are maximally different in the Jensen-Shannon sense.

Random pixels were chosen by randomly choosing 50 pixels from the lower half of the distribution of informative pixels. These pixels generally lie outside the receptive fields of the neurons.

Confidence intervals on log likelihood ratios

We created simulated datasets of the same size as the real datasets by Markov Chain Monte Carlo sampling of a distribution based on a Dirichlet prior and the observed firing pattern counts ³⁰ We fit maximum entropy models to 200 such simulated datasets. The confidence intervals are the 95% range of the resulting distribution of LLR's, indicating the confidence with which we can specify the LLR of a particular model. We used three Dirichlet priors (Dirichlet parameter $\beta = 0, 0.5, \text{ and } 1$); these led to similar results and we quote the analysis based on $\beta = 0$. For random pixels (Fig. 3b,c), confidence limits indicate two standard errors of the mean LLR and demonstrate the effect of limited data.

As a test of the statistical methods, we created artificial datasets drawn from a true pairwise model, with a comparable number of spikes as in the real datasets. For these datasets, one-minute LLR's were >-1 (likelihood ratio of $>1/2$). Note that our criterion of a “failed” model was a LLR of <-5 (likelihood ratio of $<1/32$).

Encoding

To determine the contribution of stimulus dependent correlations to stimulus encoding, we first choose the maximally informative pixel. We fit M_{ind} or M_{pair} to firing patterns conditional on this pixel's state. We calculate the mutual information between the model population responses and the stimulus state via the Jensen-Shannon divergence of the model conditional distributions.

We measure the contribution of correlations to the sparseness of the population response as the frequency at which the network is silent (for three neurons, the ‘000’ pattern) under M_{ind} and M_{pair} . We measure the information transmitted when the network is active by removing this “all-silent” firing pattern and calculating the Jensen-Shannon divergence between the remaining stimulus-conditioned firing patterns.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank S. Nirenberg for helpful comments on a draft of the manuscript.

References

1. Yoshimura Y, Callaway EM. Fine-scale specificity of cortical networks depends on inhibitory cell type and connectivity. *Nat Neurosci.* 2005; 8:1552–1559. [PubMed: 16222228]
2. Yoshimura Y, Dantzker JL, Callaway EM. Excitatory cortical neurons form fine-scale functional networks. *Nature.* 2005; 433:868–873. [PubMed: 15729343]
3. Song S, Sjöström PJ, Reigl M, Nelson S, Chklovskii DB. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* 2005; 3:e68. [PubMed: 15737062]
4. Hubel DH, Wiesel TN. Sequence regularity and geometry of orientation columns in the monkey striate cortex. *J Comp Neurol.* 1974; 158:267–293. [PubMed: 4436456]
5. Das A, Gilbert CD. Long-range horizontal connections and their role in cortical reorganization revealed by optical recording of cat primary visual cortex. *Nature.* 1995; 375:780–784. [PubMed: 7596409]
6. Shlens J, et al. The structure of multi-neuron firing patterns in primate retina. *J Neurosci.* 2006; 26:8254–8266. [PubMed: 16899720]
7. Tang A, et al. A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. *J Neurosci.* 2008; 28:505–518. [PubMed: 18184793]
8. Schneidman E, Berry MJ, Segev R 2nd, Bialek W. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature.* 2006; 440:1007–1012. [PubMed: 16625187]
9. Yu S, Huang D, Singer W, Nikolic D. A Small World of Neuronal Synchrony. *Cereb Cortex.* 2008
10. Aertsen AM, Gerstein GL, Habib MK, Palm G. Dynamics of neuronal firing correlation: modulation of “effective connectivity”. *J Neurophysiol.* 1989; 61:900–917. [PubMed: 2723733]

11. Smith MA, Kohn A. Spatial and temporal scales of neuronal correlation in primary visual cortex. *J Neurosci*. 2008; 28:12591–12603. [PubMed: 19036953]
12. Buzsaki G. Large-scale recording of neuronal ensembles. *Nat Neurosci*. 2004; 7:446–451. [PubMed: 15114356]
13. Thomson MG. Beats, kurtosis and visual coding. *Network*. 2001; 12:271–287. [PubMed: 11563530]
14. Samonds JM, Zhou Z, Bernard MR, Bonds AB. Synchronous activity in cat visual cortex encodes collinear and cocircular contours. *J Neurophysiol*. 2006; 95:2602–2616. [PubMed: 16354730]
15. Puchalla JL, Schneidman E, Harris RA, Berry MJ. Redundancy in the population code of the retina. *Neuron*. 2005; 46:493–504. [PubMed: 15882648]
16. Reich DS, Mechler F, Victor JD. Independent and redundant information in nearby cortical neurons. *Science*. 2001; 294:2566–2568. [PubMed: 11752580]
17. Gawne TJ, Kjaer TW, Hertz JA, Richmond BJ. Adjacent visual cortical complex cells share about 20% of their stimulus-related information. *Cereb Cortex*. 1996; 6:482–489. [PubMed: 8670673]
18. Gawne TJ, Richmond BJ. How independent are the messages carried by adjacent inferior temporal cortical neurons? *J Neurosci*. 1993; 13:2758–2771. [PubMed: 8331371]
19. Gray CM, Konig P, Engel AK, Singer W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*. 1989; 338:334–337. [PubMed: 2922061]
20. Meister M. Multineuronal codes in retinal signaling. *Proc Natl Acad Sci U S A*. 1996; 93:609–614. [PubMed: 8570603]
21. Cover, TM.; Thomas, JA. *Elements of Information Theory*. Wiley; New York: 1991.
22. Nirenberg S, Carcieri SM, Jacobs AL, Latham PE. Retinal ganglion cells act largely as independent encoders. *Nature*. 2001; 411:698–701. [PubMed: 11395773]
23. Mountcastle VB. Modality and topographic properties of single neurons of cat's somatic sensory cortex. *J Neurophysiol*. 1957; 20:408–434. [PubMed: 13439410]
24. Mountcastle VB. The columnar organization of the neocortex. *Brain*. 1997; 120(Pt 4):701–722. [PubMed: 9153131]
25. Henze DA, et al. Intracellular features predicted by extracellular recordings in the hippocampus in vivo. *J Neurophysiol*. 2000; 84:390–400. [PubMed: 10899213]
26. Stevens CF, Wang Y. Facilitation and depression at single central synapses. *Neuron*. 1995; 14:795–802. [PubMed: 7718241]
27. Victor JD, Mechler F, Repucci MA, Purpura KP, Sharpee T. Responses of V1 neurons to two-dimensional hermite functions. *J Neurophysiol*. 2006; 95:379–400. [PubMed: 16148274]
28. Reid RC, Victor JD, Shapley RM. The use of m-sequences in the analysis of visual neurons: linear receptive field properties. *Vis Neurosci*. 1997; 14:1015–1027. [PubMed: 9447685]
29. Fee MS, Mitra PP, Kleinfeld D. Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability. *J Neurosci Methods*. 1996; 69:175–188. [PubMed: 8946321]
30. Kennel MB, Shlens J, Abarbanel HD, Chichilnisky EJ. Estimating entropy rates with Bayesian confidence intervals. *Neural Comput*. 2005; 17:1531–1576. [PubMed: 15901407]

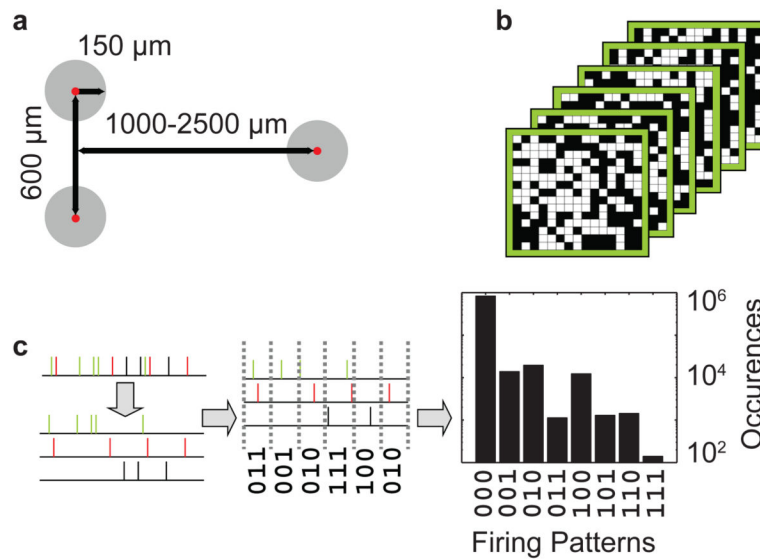


Figure 1. Analysis of multi-tetrode recordings

a, Multi-tetrode recording geometry: red dots indicate tetrode centers; grey circles indicate approximate recording areas of each tetrode. **b** Visual stimuli consisted of pseudorandom checker boards. **c**, Extracellularly-recorded spikes are sorted and binned (10 or 15 ms bin width). Multi-neuron firing patterns are identified and tallied to form a firing pattern distribution. Maximum entropy models are fit to this distribution.

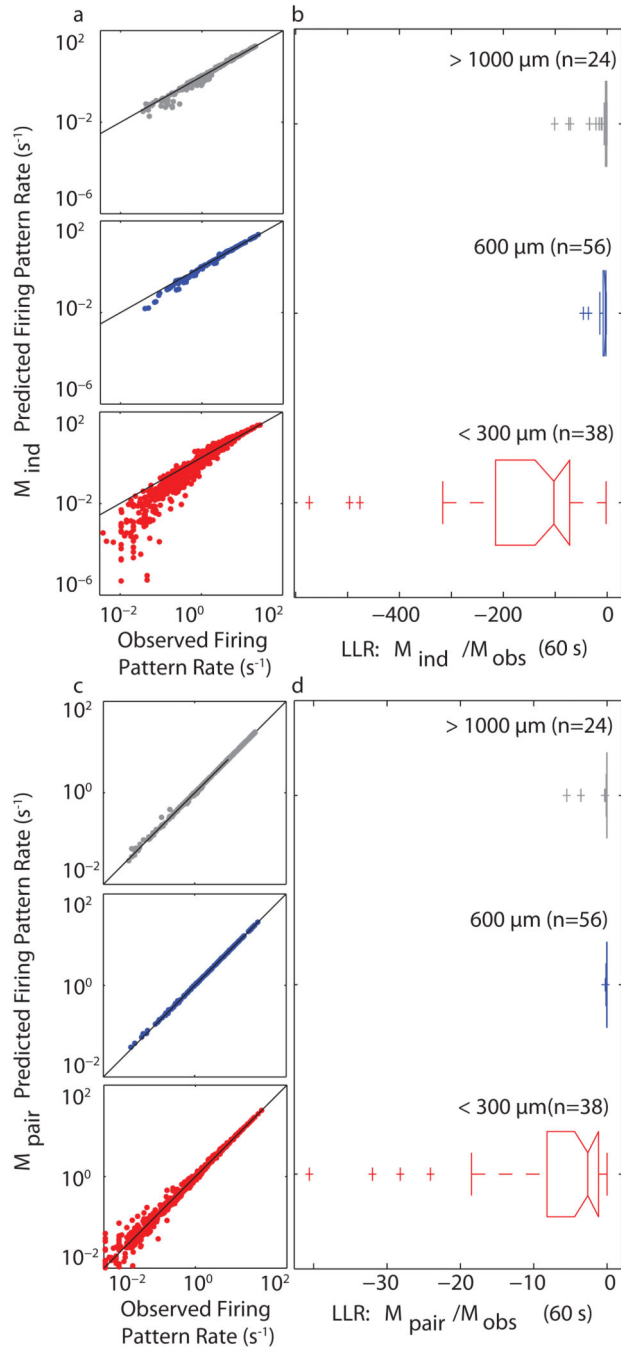


Figure 2. A pairwise maximum entropy model frequently fails for local networks of neurons
a, The independent model (M_{ind} , ordinate) fails to predict the observed (M_{obs} , abscissa) frequency of firing patterns over a range of spatial scales. **b**, Box plots (inter-quartile range) of log-likelihood ratios of the independent model (M_{ind}), referenced to a perfect model (M_{obs}) for 60 s of data. Notches indicate median values; outliers plotted as pluses. Across spatial scales, M_{ind} is a poor fit. **c**, The pairwise model (M_{pair} , ordinate) fails to predict the observed (M_{obs} , abscissa) frequency of firing patterns for ensembles of neurons separated by $<300 \mu m$ (red dots), but not at larger separations. **d**, Box plots of log-likelihood ratios of the

pairwise model (M_{pair}), referenced to a perfect model (M_{obs}). At distances $<300 \mu\text{m}$ (red), M_{pair} frequently fails to account for the observed distribution of firing patterns; at larger separations, M_{pair} is nearly perfect.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

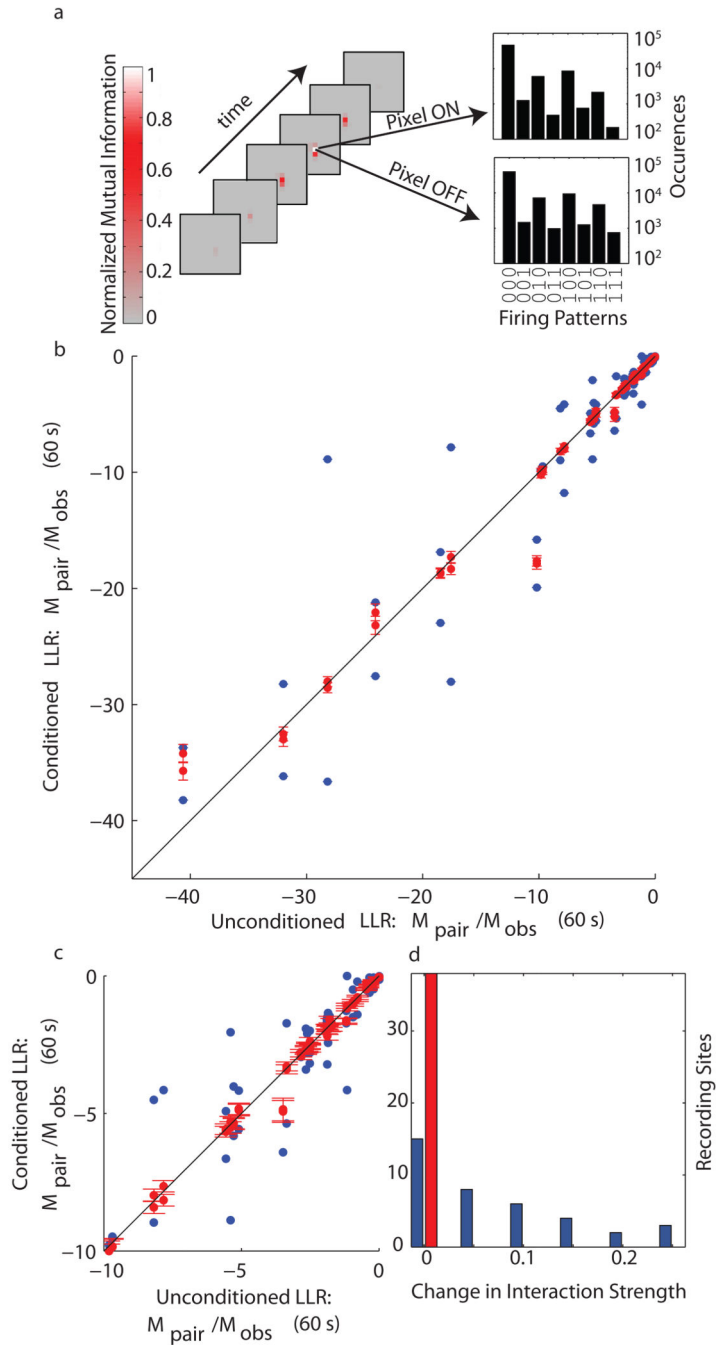


Figure 3. High-order stimulus dependent correlations in multi-neuron firing patterns
a, The Jensen-Shannon divergence between firing pattern distributions conditional on each pixel (insets on right) yields a spatiotemporal map of mutual information between the stimulus and the population response. For a maximally informative pixel, M_{pair} is fit to these conditional distributions. **b**, Conditioning on maximally informative pixels (blue dots) has a large effect on the goodness of fit of M_{pair} ; conditioning on random pixels (red dots) has a small effect. For maximally-informative pixels, confidence intervals (95%, see Methods) are smaller than the plotted symbols. For random pixels, the error bars show 95% confidence

intervals for the mean. **c**, Expanded view of **b**, near the origin. **d**, Histogram of the change in average interaction strength (J_{ij}) following conditioning. Conditioning on informative pixels (blue bars), but not random pixels (red bars) lead to significantly different interaction strengths.

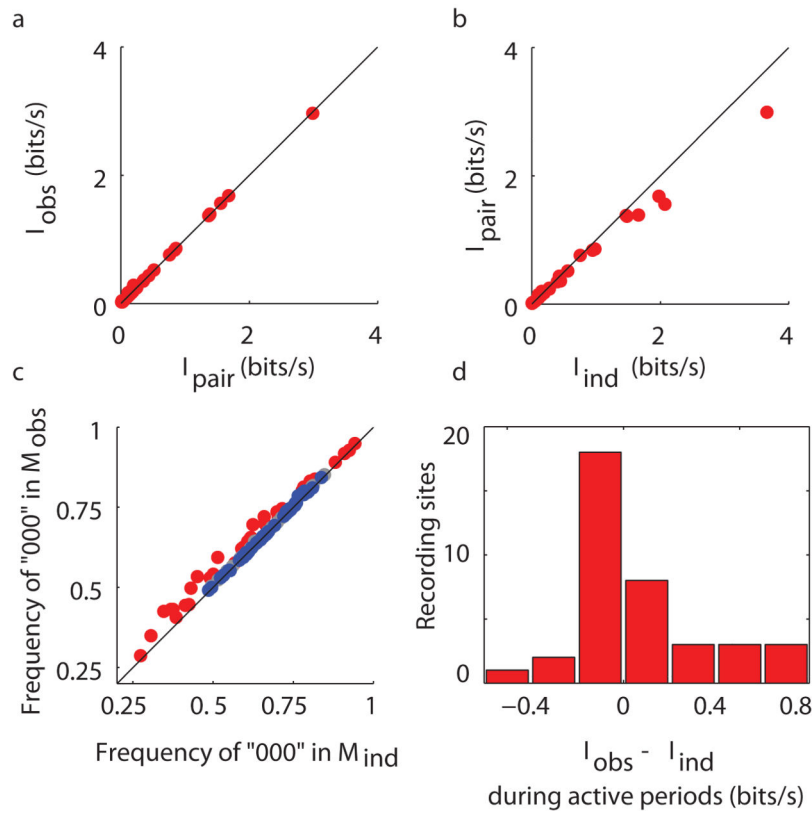


Figure 4. Effect of correlations on stimulus encoding

a, Mutual information between the stimulus (maximally informative pixel) and neural response generated by M_{obs} (ordinate) and M_{pair} (abscissa) for neurons recorded at $<300 \mu\text{m}$. Higher-than-second order correlations make no contribution to the mutual information transmitted. **b**, Pairwise correlations slightly reduce the information encoded by local ensembles. **c**, At distances $<300 \mu\text{m}$ (red), neuronal activity (M_{obs}) has a higher frequency of the all-silent firing pattern than predicted from independent firing (M_{ind}), indicating that correlations in neuronal firing increase the sparseness of local networks. No difference was seen at larger distances ($600 \mu\text{m}$ (blue) or $>1000 \mu\text{m}$ (gray), confluent on the diagonal). **d**, The effect of correlations on the information encoded when the network is active. On average, correlations increase the information transmitted when the network is active.