Supplementary Methods.

High resolution temporal analysis. Classifying near simultaneous (<1.2 ms) spiking events from multiple neurons recorded at the same tetrode is difficult because multiple action potential (AP) waveforms overlap. Here, we show that findings are not a result of detection failures due to this overlap, and that its quantitative effects are minor. We do this both by analyzing the size of the correction due to detection failures, and the fine temporal structure of the recorded spike trains. In the analyses below, we consider only the responses to pseudorandom stimuli. First, we consider the analysis used in the main text. We corrected for underestimates in near-synchronous firing events by a simple lockout correction, based on the assumption that firing probabilities are constant on a timescale of 10 to 15 ms (see main Methods). Supp. Fig. 2 shows that this correction has only a minimal effect on the fit of $M_{\text{pair}}$, and, indeed, typically worsens it.

Next, because synchronous firing on a millisecond timescale may differ from that on a timescale of 10-15 ms, we apply a more sophisticated correction (Supp. Fig. 3). Specifically, we add in an excess of near simultaneous spikes on a timescale of 1.5 ms. To do this, we bin each spike train into intervals of 0.5 ms and construct pairwise cross-correlograms for all pairs of neurons (Supp. Fig. 3a). For each pair of neurons ($i$ and $j$), we find the peak number of coincident spikes ($C_{ij}$) in a window +/- 1.5 ms around 0 lag. Then, for each bin ($t$) on the interval [-1.5 ms 1.5 ms], we add-in $\lambda C_{ij}/2$ spikes to each train, with each train’s added spikes placed $t$ ms after a randomly-chosen spike in the other train, and avoiding overlaps. For $\lambda =1$, this generates spike trains for which the zero-lag cross-correlation matches the peak cross-correlation (Supp. Fig. 3b), and for $\lambda =2$, it generates spike trains in which zero-lag cross-correlation is twice the peak (Supp. Fig. 3c). We then evaluate the performance of $M_{\text{pair}}$ and $M_{\text{obs}}$ on these datasets as described in Methods. As shown in Supp. Fig. 3d, adding fine-scale correlations worsens the fit of the pairwise model.
Finally, we remove near simultaneous multi-neuron firing events from networks separated by 600 and >1000μm and demonstrate that these networks still display only second-order correlations (Supp. Fig. 4). To do this, spike trains (now recorded from separate tetrodes) are binned at 0.5 ms intervals. For each pair of neurons (i and j) we find all instances of a spike in i within a +/- 1.5 ms window around a spike in j and randomly delete a spike from one train or the other. This leads to spike trains which have no simultaneous events within this window. We apply $M_{\text{pair}}$ and $M_{\text{obs}}$ to these datasets. As shown in Supp. Fig. 4, this has little effect on model performance. In particular, it does not lead to the behavior seen for local neuronal populations – the pairwise model fits data from networks at 600 and >1000 microns, even when simultaneous events are removed (simulating a recording from a single tetrode).

**Cross-validation.** To verify that our findings do not result from over-fitting, we performed a cross-validation. For each recording site, we assigned interleaved presentations to a training set and a testing set. We calculated the log likelihood ratio for two models applied to the test set: the training set’s pairwise model ($M_{\text{pair}}$), and the training set’s full distribution of firing probabilities ($M_{\text{obs train}}$). As seen in Supp. Fig. 5, $M_{\text{obs train}}$ typically provided a significant improvement over $M_{\text{pair}}$ for the 300-micron datasets (most points above the diagonal), but not for the others. Note that out-of-sample fits of $M_{\text{pair}}$ (abscissa) are not nearly as good as in-sample fits (Main Fig. 2d), owing both to data limitations and to overall fluctuations in firing rates. The crucial observation is that for the 300-micron datasets, the out-of-sample fit was improved by adding the triplet firing probability ($M_{\text{obs train}}$), while for the more widely separated neurons, it was not.
Supplementary Figure Legends

**Figure 1 |** $M_{\text{pair}}$ frequently fails for local networks of neurons driven by natural images. **Left,** Under naturalistic stimulation, the pairwise model (ordinate) fails to predict the observed (abscissa) frequency of firing patterns for ensembles of neurons separated by <300 μm (red dots). **Right,** Box plots of log-likelihood ratios of the pairwise model ($M_{\text{pair}}$), referenced to a perfect model ($M_{\text{obs}}$) under naturalistic stimulation for 60 s of data. At distances <300 μm (red), failures of $M_{\text{pair}}$ are more pronounced for natural stimuli than for pseudorandom stimuli (Main Fig. 2b).

**Figure 2 |** The lock-out correction has only a minor effect on the fit of the pairwise model $M_{\text{pair}}$. The goodness of fit of $M_{\text{pair}}$ when applied to raw data (no lock-out correction, abscissa), was compared to its fit when applied to data corrected for the lockout bias, as in the main text (ordinate). In no case did this result in a substantial change of goodness-of-fit.

**Figure 3 |** Failure of $M_{\text{pair}}$ in fine-scale networks does not reflect “missed” correlations on timescales of ~1 ms. **a,** Cross-correlograms for three simultaneously recorded neurons. Overlapping multi-neuron AP waveforms preclude accurate spike sorting at this resolution and leads to an absence of correlations at lags of 1.5 ms. **b,** Cross-correlograms illustrating the effect of adding an estimate of near-simultaneous events obscured by lockout ($\lambda=1$ in Supplementary Methods.) **c,** Cross-correlograms illustrating the effect of adding double the estimated near-simultaneous events obscured by lockout ($\lambda=2$ in Supplementary Methods). **d,** Adding a correction for missed near-simultaneous events worsens $M_{\text{pair}}$ fits. On the abscissa are log-likelihood ratios of $M_{\text{pair}}/M_{\text{obs}}$ for the uncorrected data (e.g., panel **a**), as in Supp. Fig. 2. The ordinate shows log-likelihood ratios of $M_{\text{pair}}/M_{\text{obs}}$ based on adding an estimate of
simultaneous events ($\lambda=1$, as in panel b), plotted as points, and adding double that estimate ($\lambda=2$, as in panel c), plotted as arrowheads. Both corrections worsen the fit of $M_{\text{pair}}$.

**Figure 4 | Artificial removal of spiking events within 1.5 ms does not lead to failure of $M_{\text{pair}}$ in networks separated by 600 μm or more.** The abscissa shows log-likelihood ratios of $M_{\text{pair}}/M_{\text{obs}}$ for datasets recorded at distances of >300 μm, analyzed as in the main text. The ordinate shows how lockout of near-simultaneous events due to single-tetrode recording would have altered the analysis. To do this, we removed simultaneous spiking events (recorded separately on different tetrodes) that occurred within <1.5 ms, prior to fitting the models. This has a negligible effect on model fit, thus demonstrating that the differences between correlation patterns in local (<300 μm) and more distant (600 μm or more) clusters is not merely due to an inability to record simultaneous events on a 1 ms timescale.

**Figure 5 | Cross-validation of maximum entropy models.** To separate intrinsic model failure from the effects of gradual changes in neural activity or errors in parameter estimation, maximum-entropy models were fit to one half of the data, and then applied to the second half (see Supplementary Methods). We then compared the out-of-sample fit of $M_{\text{pair}}$ (abscissa) with the out-of-sample fit of $M_{\text{obs \ train}}$ (ordinate), which also included the triplet interaction; both were referenced to $M_{\text{obs \ test}}$, the perfect model for the out-of-sample data. For distant (600 μm and >1000 μm) clusters, $M_{\text{obs \ train}}$ only rarely improved the fit over $M_{\text{pair}}$ (most points near the diagonal). For local (<300 μm) clusters, the out-of-sample fits were often substantially improved by including the triplet interaction (most points above the diagonal), confirming its statistical significance. Note that the out-of-sample fits of $M_{\text{pair}}$ are worse than the within-
sample fits (Main Fig. 2d), reflecting both random variation in neural activity and a reduction in the amount of data used to build the model.

**Figure 6 | The pairwise maximum entropy model fails for ensembles of 3-6 neurons.** A histogram of log-likelihood ratios per minute of the pairwise model to a perfect model. Ensembles of 3 neurons (black bars), 4 neurons (blue), 5 neurons (red) and 6 neurons (green bars) often produce firing patterns unlikely to result from interactions between pairs of neurons, indicating the presence of higher-order correlations.

**Figure 7 | Conditioning on the maximally informative pixel often leads to significant changes in the fit of pairwise maximum entropy model.** As described in the main text, pairwise maximum entropy models were fit to subsets of the data, defined by the state of the maximally informative pixel. The difference in goodness of fit of these two models is shown here, as quantified by the difference in their log-likelihood ratios per minute.
Supplemental Figure

Supp. 1

Graphs showing observed firing pattern rates for different conditions.

Supp. 2

Graph showing lockout-corrected data with LLR against raw data.

Supp. 3

Graph and plots showing spike rates and firing patterns for different conditions.

Neuron 1 vs. 2
Neuron 2 vs. 3
Neuron 1 vs. 3

Data with simultaneous events added

Legend:
- < 300 µm (n=46)
- 300-600 µm (n=481)
- > 1000 µm (n=265)

Supplemental Figure
