bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Inhibitory control of correlated variability in cortical networks

Marius Pachitariu[*,1], Carsen Stringer[*,2], Michael Okun[1], Peter Bartho[3], Kenneth Harris[1], Peter Latham[2], Maneesh Sahani[2], Nicholas Lesica[4]

[*] equal contribution first authors

[1] Institute of Neurology, UCL, London, UK

[2] Gatsby Computational Neuroscience Unit, UCL, London, UK

[3] Hungarian Academy of Sciences, Budapest, Hungary

[4] Ear Institute, UCL, London, UK

SUMMARY

Cortical networks exhibit intrinsic dynamics that drive coordinated, large-scale fluctuations across neuronal populations and create noise correlations that impact sensory coding. To investigate the network-level mechanisms that underlie these dynamics, we developed novel computational techniques to fit a deterministic spiking network model directly to multi-neuron recordings from different species, sensory modalities, and behavioral states. The model accurately reproduced the wide variety of activity patterns in our recordings, and analysis of its parameters suggested that differences in noise correlations across recordings were due primarily to differences in the strength of feedback inhibition. Further analysis of our recordings confirmed that putative inhibitory interneurons were indeed more active during desynchronized cortical states with weak noise correlations. Our results demonstrate the power of fitting spiking network models directly to multi-neuron recordings and suggest that inhibition modulates the interactions between intrinsic dynamics and sensory inputs by controlling network stability.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

INTRODUCTION

The patterns of cortical activity evoked by sensory stimuli provide the internal representation of the outside world that underlies perception. However, these patterns are driven not only by sensory inputs, but also by the intrinsic dynamics of the underlying cortical network. These dynamics can create correlations in the activity of neuronal populations with important consequences for coding and computation [Shadlen et al., 1996, Abbott and Dayan, 1999]. The correlations between pairs of neurons have been studied extensively [Cohen and Kohn, 2011, Ecker et al., 2010, Averbeck et al., 2006] and substantial effort has been directed toward understanding their origin [Renart et al., 2010]. Recent studies have demonstrated that correlations are driven by dynamics involving coordinated, large-scale fluctuations in the activity of many neurons [Sakata and Harris, 2009, Pachitariu et al., 2015, Okun et al., 2015] and, importantly, that the nature of these dynamics and the correlations that they create are dependent on the state of the underlying network; it has been shown that various factors modulate the strength of correlations, such as anaesthesia [Harris and Thiele, 2011, Schölvinck et al., 2015, Constantinople and Bruno, 2011], attention [Cohen and Maunsell, 2009, Mitchell et al., 2009, Buran et al., 2014], locomotion [Schneider et al., 2014, Erisken et al., 2014], and alertness [Vinck et al., 2015, McGinley et al., 2015a]. In light of these findings, it is critical that we develop a deeper understanding of the origin and consequences of correlations at the biophysical network level.

In this study, we use a large number of multi-neuron recordings and a model-based analysis to investigate the mechanisms that control noise correlations, a manifestation of intrinsic dynamics during sensory processing in which the variability in responses to identical stimuli is shared between neurons. For our results to provide direct insights into physiological mechanisms, we required a model with several properties: (1) the model must be able to internally generate the complex intrinsic dynamics of cortical networks, (2) it must be possible to fit the model parameters directly to spiking activity from individual multi-neuron recordings, and (3) the model must be biophysically interpretable and enable predictions that can be tested experimentally. Thus far, the only network models that have been fit directly to multi-neuron recordings have relied on either

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

abstract dynamical systems [Curto et al., 2009] or probabilistic frameworks in which variability is modelled as stochastic and correlated variability arises through abstract latent variables whose origin is assumed to lie either in unspecified circuit processes [Ecker et al., 2014, Macke et al., 2011, Pachitariu et al., 2013] or elsewhere in the brain [Goris et al., 2014, de la Rocha et al., 2007]. While these models are able to accurately reproduce many features of cortical activity and provide valuable summaries of the phenomenological and computational properties of cortical networks, their parameters are difficult to interpret at a biophysical level.

One alternative to these abstract stochastic models is a biophysical spiking network, which can generate variable neural activity through chaotic amplification of different initial conditions [van Vreeswijk and Sompolinsky, 1996, Amit and Brunel, 1997, Renart et al., 2010, Litwin-Kumar and Doiron, 2012, Wolf et al., 2014]. These networks can be designed to have interpretable parameters, but have not yet been fit directly to multi-neuron recordings and, thus, their use has been limited to attempts to explain qualitative features of cortical dynamics through manual tuning of network parameters. This approach has revealed a number of different network features that are capable of controlling dynamics, such as clustered connectivity [Litwin-Kumar and Doiron, 2012], synaptic coupling strength [Ostojic, 2014], or adaptation currents [Latham et al., 2000, Destexhe, 2009], but the inability to fit the networks directly to recordings has made it difficult to identify which network features play an important role in vivo. To overcome this limitation, we used a novel computational approach that allowed us to fit a spiking network directly to individual multi-neuron recordings. By taking advantage of the computational power of graphics processing units (GPUs), we were able to sample from the network with millions of different parameter values to find those that best reproduced the activity in a given recording.

We developed a biophysical spiking network with intrinsic variability and a small number of parameters that was able to capture the apparently doubly chaotic structure of cortical activity [Churchland and Abbott, 2012]. Like classical excitatory-inhibitory networks, the model generates deterministic microscopic trial-to-trial variability in the spike times of individual neurons [van Vreeswijk and Sompolinsky, 1996], as well as

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

macroscopic variability in the form of coordinated, large-scale fluctuations that are shared across neurons. Because these fluctuations are of variable duration, arise at random times, and do not necessarily phase-lock to external input, they create noise correlations in evoked responses.

To gain insight into the mechanisms that control noise correlations in vivo, we took the following approach: (1) we assembled multi-neuron recordings from different species, sensory modalities, and behavioral states to obtain a representative sample of cortical dynamics; (2) we generated activity from the network model to understand how each of its parameters controls its dynamics, and we verified that it was able to produce a variety of spike patterns that were qualitatively similar to those observed in vivo; (3) we fit the model network directly to the spontaneous activity in each of our recordings, and we verified that the spike patterns generated by the network quantitatively matched those in each recording; (4) we examined responses to sensory stimuli to determine which of the model parameters could account for the differences in noise correlations across recordings – the results of this analysis identified the strength of feedback inhibition as a key parameter and predicted that the activity of inhibitory interneurons should vary inversely with the strength of noise correlations; (5) we confirmed this prediction through additional analysis of our recordings showing that the activity of fast-spiking (FS) neurons is increased during periods of cortical desynchronization with weak noise correlations in both awake and anesthetized animals. Our results suggest that weak inhibition allows activity to be dominated by coordinated, large-scale fluctuations that cause the state of the network to vary over time and, thus, create variability in the responses to successive stimuli that is correlated across neurons. In contrast, when inhibition is strong, these fluctuations are suppressed and the network state remains constant over time, allowing the network to respond reliably to successive stimuli and eliminating noise correlations.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

RESULTS

Cortical networks exhibit a wide variety of intrinsic dynamics

To obtain a representative sample of cortical activity patterns, we collected multi-neuron recordings from different species (mouse, gerbil, or rat), sensory modalities (A1 or V1), and behavioral states (awake or under one of several anesthetic agents). We compiled recordings from a total of 59 multi-neuron populations across 6 unique recording types (i.e. species/modality/state combinations; Table S1). The spontaneous activity in different recordings exhibited striking differences not only in overall activity level, but also in the spatial and temporal structure of activity patterns; while concerted, large-scale fluctuations were prominent in some recordings, they were nearly absent in others (Figure 1A). In general, large-scale fluctuations were weak in awake animals and strong under anesthesia, but this was not always the case (see further examples in Figure 3 and summary statistics for each recording in Figure S1).

The magnitude and frequency of the large-scale fluctuations in each recording were reflected in the auto-correlation function of the multi-unit activity (MUA, the summed spiking of all neurons in the population in 15 ms time bins). The autocorrelation function of the MUA decayed quickly to zero for recordings with weak large-scale fluctuations, but had oscillations that decayed slowly for recordings with stronger fluctuations (Figure 1B). The activity patterns in recordings with strong large-scale fluctuations were characterized by clear transitions between up states, where most of the population was active, and down states, where the entire population was silent. These up and down state dynamics were reflected in the distribution of the MUA across time bins; recordings with strong large-scale fluctuations had a large percentage of time bins with zero spikes (Figure 1C).

To summarize the statistical structure of the activity patterns in each recording, we measured four quantities. We used mean spike rate to describe the overall level of activity, mean pairwise correlations to describe the spatial structure of the activity patterns, and two different measures to describe the temporal structure of the

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

activity patterns – the decay time of the autocorrelation function of the MUA, and the percentage of MUA time bins with zero spikes. While there were some dependencies in the values of these quantities across different recordings (Figure 1D), there was also considerable scatter both within and across recording types. This scatter suggests that there is no single dimension in the space of cortical dynamics along which the overall level of activity and the spatial and temporal structure of the activity patterns all covary, but rather that cortical dynamics span a multi-dimensional continuum [Harris and Thiele, 2011]. This was confirmed by principal component analysis; even in the already reduced space described by our summary statistics, three principal components were required to account for the differences in spike patterns across recordings (Figure 1E).

A deterministic spiking network model of cortical activity

To investigate the network-level mechanisms that control cortical dynamics, we developed a biophysically-interpretable model that was capable of reproducing the wide range of activity patterns observed in vivo. We constructed a minimal deterministic network of excitatory spiking integrate-and-fire neurons with non-selective feedback inhibition and single-neuron adaptation currents (Figure 2A). Each neuron receives constant tonic input, and the neurons are connected randomly and sparsely with 5% probability. The neurons are also coupled indirectly through global, supralinear inhibitory feedback driven by the spiking of the entire network [Rubin et al., 2015], reflecting the near-complete interconnectivity between pyramidal neurons and interneurons in local populations [Hofer et al., 2011, Fino and Yuste, 2011, Packer and Yuste, 2011]. The supralinearity of the inhibitory feedback is a critical feature of the network, as it shifts the balance of excitation and inhibition in favor of inhibition when the network is strongly driven [Haider et al., 2013].

The model has five free parameters: three controlling the average strength of excitatory connectivity, the strength of inhibitory feedback, and the strength of adaptation, respectively, and two controlling the strength of the tonic input to each neuron, which is chosen from an exponential distribution. The timescales that control the decay of the excitatory, inhibitory and adaptation currents are fixed at 5 ms, 3.75 ms and 375

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ms, respectively. Note that no external noise input is required to generate variable activity; population-wide fluctuations over hundreds of milliseconds are generated when the slow adaptation currents synchronize across neurons to maintain a similar state of adaptation throughout the entire network, which, in turn, results in coordinated spiking [Latham et al., 2000, Destexhe, 2009].

The variability in the model arises through chaotic amplification of small changes in initial conditions or small perturbations to the network that cause independent simulations to diverge. In some parameter regimes, the instability of the network is such that the structure of the spike patterns generated by the model is sensitive to changes in the spike times of individual neurons. In fact, a single spike added randomly to a single neuron during simulated activity is capable of changing the time course of large-scale fluctuations, in some cases triggering immediate population-wide spiking (Figure 2B, top rows). Similar phenomena have been observed in vivo previously [London et al., 2010] and were also evident in our recordings when comparing different extracts of cortical activity; spike patterns that were similar for several seconds often then began to diverge almost immediately (Figure 2B, bottom rows).

Multiple features of the network model can control its dynamics

The dynamical regime of the network model is determined by the interactions between its different features. To determine the degree to which each feature of the network was capable of influencing the structure of its activity patterns, we analyzed the effects of varying the value of each model parameter. We started from a fixed set of parameter values and simulated activity while independently sweeping each parameter across a wide range of values. The results of these parameter sweeps clearly demonstrate that each of the five parameters can exert strong control over the dynamics of the network, as both the overall level of activity and the spatial and temporal structure of the patterns in simulated activity varied widely with changes in each parameter (Figure 2C-D).

With the set of fixed parameter values used for the parameter sweeps, the network is in a regime with slow,

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ongoing fluctuations between up and down states. In this regime, the amplification of a small perturbation results in a sustained, prolonged burst of activity (up state), which, in turn, drives a build-up of adaptation currents that ultimately silences the network for hundreds of milliseconds (down state) until the cycle repeats. These fluctuations can be suppressed by an increase in the strength of feedback inhibition, which eliminates slow fluctuations and shifts the network into a regime with weak, tonic spiking and weak correlations (Figure 2C-D, first column); in this regime, small perturbations are immediately offset by the strong inhibition and activity is returned to baseline [Renart et al., 2010]. The fluctuations between up and down states can also be suppressed by decreasing adaptation (Figure 2C-D, second column); without adaptation currents to create slow, synchronous fluctuations across the network, neurons exhibit strong, tonic spiking.

The dynamics of the network can also be influenced by changes in the strength of the recurrent excitation or tonic input. Increasing the strength of excitation results in increased activity and stronger fluctuations, as inhibition is unable to compensate for the increased amplification of small perturbations (Figure 2C-D, third column). Increasing the spread or baseline level of tonic input also results in increased activity, but with suppression, rather than enhancement, of slow fluctuations (Figure 2C-D, fourth and fifth column). As either the spread or baseline level of tonic input is increased, more neurons begin to receive tonic input that is sufficient to overcome their adaptation current and, thus, begin to quickly reinitiate up states after only brief down states and, eventually, transition to tonic spiking.

The network model reproduces the dynamics observed in vivo

The network simulations demonstrate that each of its features is capable of controlling its dynamics and shaping the structure of its activity patterns. To gain insight into the mechanisms that may be responsible for creating the differences in dynamics observed in vivo, we fit the model to each of our recordings. We optimized the model parameters so that the patterns of activity generated by the network matched those observed in spontaneous activity (Figure 3A). We measured the agreement between the simulated and recorded activity by a cost function which was the sum of discrepancies in the autocorrelation function of the

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

MUA, the distribution of MUA values across time bins, and the mean pairwise correlations. Together, these statistics describe the overall level of activity in each recording, as well as the spatial and temporal structure of its activity patterns.

Fitting the model to the recordings required us to develop new computational techniques. The network parametrization is fundamentally nonlinear, and the statistics used in the cost function are themselves non-linear functions of a dynamical system with discontinuous integrate-and-fire mechanisms. Thus, as no gradient information was available to guide the optimization, we used Monte Carlo simulations to generate activity and measure the relevant statistics with different parameter values. By using GPU computing resources, we were able to design and implement network simulations that ran 10000x faster than real time, making it feasible to sample the cost function with high resolution and locate its global minimum to identify the parameter configuration that resulted in activity patterns that best matched those of each recording.

The model was flexible enough to capture the wide variety of activity patterns observed across our recordings, producing both decorrelated, tonic spiking and coordinated, large-scale fluctuations between up and down states as needed (see examples in Figure 3B, statistics for all recordings and models in Figure S1, and parameter values and goodness-of-fit measures for all recordings in Figure S2). Because we used a cost function that captured many different properties of the recorded activity while fitting only a very small number of model parameters, the risk of network degeneracies was relatively low [Marder et al., 2015]. Nonetheless, we also confirmed that analysis of model parameters corresponding to local minima of the cost function did not lead to a different interpretation of our results (see Figure S3).

Strong inhibition suppresses noise correlations

Our main interest was in understanding how the different network-level mechanisms that are capable of controlling intrinsic dynamics contribute to the correlated variability in responses evoked by sensory stimuli. The wide variety of intrinsic dynamics in our recordings was reflected in the differences in evoked responses

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

across recording types; while some recordings contained strong, reliable responses to the onset of a stimulus, other recordings contained responses that were highly variable across trials (Figure 4A). There were also large differences in the extent to which the variability in evoked responses was correlated across the neurons in each recording; pairwise noise correlations were large in some recordings and extremely weak in others, even when firing rates were similar (Figure 4B).

Because evoked spike patterns can depend strongly on the specifics of the sensory stimulus, we could not make direct comparisons between experimental responses across different species and modalities; our goal was to identify the internal mechanisms that are responsible for the differences in noise correlations across recordings and, thus, any differences in spike patterns due to differences in external input would confound our analysis. To overcome this confound and enable the comparison of noise correlations across recording types, we simulated the response of the network to the same external input for all recordings. We constructed the external input using recordings of spiking activity from the inferior colliculus (IC), a primary relay nucleus in the subcortical auditory pathway (Figure 4C-D). Using the subset of our cortical recordings in which we presented the same sounds that were also presented during the IC recordings, we verified that the noise correlations in the simulated cortical responses were similar to those in the recordings (Figures 4E).

The parameter sweeps described in Figure 2 demonstrated that there are multiple features of the model network that can control its intrinsic dynamics, and a similar analysis of the noise correlations in simulated responses to external input produced similar results (Figure S3).To gain insight into which of these features could account for the differences in noise correlations across our recordings, we examined the dependence of the strength of the noise correlations in each recording on each of the model parameters. While several parameters were able to explain a significant amount of the variance in noise correlations across recordings, the amount of variance explained by the strength of inhibitory feedback was by far the largest (Figure 5A). The predominance of inhibition in the control of noise correlations was confirmed by the measurement of partial correlations (the correlation between the noise correlations and each parameter that remains

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

after factoring out the influence of the other parameters; partial r^2 for inhibition: 0.67, excitation: 0.02, adaptation: 0.08, tonic input spread: 0.17, and tonic input baseline: 0.04). We also performed parameter sweeps to confirm that varying only the strength of inhibition was sufficient to result in large changes in noise correlations in the parameter regime of each recording (Figure 5B).

Strong inhibition sharpens tuning and enables accurate decoding

We also examined how different features of the network controlled other aspects of evoked responses. We began by examining the extent to which differences in the value of each model parameter could explain differences in stimulus selectivity across recordings. To estimate selectivity, we drove the model network that was fit to each cortical recording with external inputs constructed from IC responses to tones, and used the simulated responses to measure the width of the frequency tuning curves of each model neuron. Although each model network received the same external inputs, the selectivity of the neurons in the different networks varied widely. The average tuning width of the neurons in each network varied most strongly with the strength of the inhibitory feedback in the network (Figure 5C; partial r^2 for inhibition: 0.74, excitation: 0.06, adaptation: 0.48, tonic input spread: 0.01, and tonic input baseline: 0.37), and varying the strength of inhibition alone was sufficient to drive large changes in tuning width (Figure 5D). These results are consistent with experiments demonstrating that inhibition can control the selectivity of cortical neurons [Lee et al., 2012], but suggest that this control does not require structured lateral inhibition.

We also investigated the degree to which the activity patterns generated by the model fit to each cortical recording could be used to discriminate different external inputs. We trained a decoder to infer which of seven possible stimuli evoked a given single-trial activity pattern and examined the extent to which differences in the value of each model parameter could account for the differences in decoder performance across recordings. Again, the amount of variance explained by the strength of inhibitory feedback was by far the largest (Figure 5E; partial r^2 for inhibition: 0.5, excitation: 0.16, adaptation 0.27, tonic input spread 0.02, and tonic input baseline 0.03); decoding was most accurate for activity patterns generated by networks with

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

strong inhibition, consistent with the weak noise correlations and high selectivity of these networks. Parameter sweeps confirmed that varying only the strength of inhibition was sufficient to result in large changes in decoder performance (Figure 5F).

Activity of fast-spiking (FS) neurons is increased during periods of cortical desynchronization with weak noise correlations

Our model-based analyses suggest an important role for feedback inhibition in controlling the way in which responses to sensory inputs are shaped by intrinsic dynamics. In particular, our results predict that inhibition should be strong in dynamical regimes with weak noise correlations. To test this prediction, we performed further analysis of our recordings to estimate the strength of inhibition in each recorded population. We classified the neurons in each recording based on the width of their spike waveforms (Figure S5); the waveforms fell into two distinct clusters, allowing us to separate fast-spiking (FS) neurons, which are mostly parvalbumin-positive (PV+) inhibitory interneurons, from regular-spiking (RS) neurons, which are mostly excitatory pyramidal neurons [Kawaguchi and Kubota, 1997, Barthó et al., 2004, Povysheva et al., 2008, Cho et al., 2010, Madisen et al., 2012, Roux and Buzsáki, 2015, Cardin et al., 2009].

Given the results of our model-based analyses, we hypothesized that the overall level of activity of FS neurons should vary inversely with the strength of noise correlations. To identify sets of trials in each recording that were likely to have either strong or weak noise correlations, we measured the level of cortical synchronization. Previous studies have shown that noise correlations are strong when the cortex is in a synchronized state, where activity is dominated by concerted, large-scale fluctuations, and weak when the cortex is in a desynchronized state, where these fluctuations are suppressed [Pachitariu et al., 2015, Schölvinck et al., 2015].

We began by analyzing our recordings from V1 of awake mice. We classified the cortical state during each stimulus presentation based on the ratio of low-frequency LFP power to high-frequency LFP power [Sakata

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and Harris, 2012] and compared evoked responses across the most synchronized and desynchronized subsets of trials (Figure 6A). As expected, noise correlations were generally stronger during synchronized trials than during desynchronized trials, and this variation in noise correlations with cortical synchrony was evident both within individual recordings and across animals (Figure 6B-C). As predicted by our model-based analyses, the change in noise correlations with cortical synchrony was accompanied by a change in FS activity; there was a four-fold increase in the mean spike rate of FS neurons from the most synchronized trials to the most desynchronized trials, while RS activity remained constant (Figure 6D-F).

We next examined our recordings from gerbil A1 under urethane in which the cortex exhibited transitions between distinct, sustained synchronized and desynchronized states (Figure 6G). As in our awake recordings, cortical desynchronization under urethane was accompanied by a decrease in noise correlations and an increase in FS activity (Figures 6H-K). In fact, both FS and RS activity increased with cortical desynchronization under urethane, but the increase in FS activity was much larger.

Finally, we examined our remaining recordings from gerbil A1 under either ketamine/xylazine (KX) or fentanyl/medetomidine/midazolam (FMM) anesthesia. In these recordings, the cortex did not transition between different dynamical regimes, so we could not track changes in noise correlations and FS activity within individual recordings. However, recordings under KX and FMM exhibited stable states with strong and weak noise correlations respectively [Pachitariu et al., 2015] (Figure 7A), so we were able to make comparisons across recordings. Noise correlations under FMM were extremely weak, while those under KX were the largest in any of our recordings, so we expected FS activity under FMM to be much higher than that under KX. Surprisingly, our initial analysis suggested the opposite; the average spike rate of FS neurons under KX was larger than that under FMM (Figure 7B). Further analysis revealed, however, that there were many fewer FS neurons in our KX recordings than in our FMM recordings (Figure 7C), and this effect was most pronounced for FS neurons with low spike rates (Figure 7D). The low number of FS neurons in our KX recordings suggests that many FS neurons become completely silent under KX (all recordings were made in the same region of gerbil A1 with the same multi-tetrode arrays, so a similar number of FS neurons should

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

be expected). When we measured FS activity as the sum of all spiking in each recording rather than the average spike rate of each neuron, the amount of FS activity was indeed much larger under FMM than under KX, consistent with our observations in other recording types and the predictions of our model-based analyses (Figure 7E-F).

DISCUSSION

We have shown that a deterministic spiking network model is capable of reproducing the wide variety of multi-neuron cortical activity patterns observed in vivo. Through chaotic amplification of small perturbations, the model generates activity with both trial-to-trial variability in the spike times of individual neurons and coordinated, large-scale fluctuations of the entire network. Although several features of the model network are capable of controlling its intrinsic dynamics, our analysis suggests that the differences in the correlated variability in evoked responses across our in vivo recordings can be accounted for by differences in feedback inhibition. When we fit the model to each of our individual recordings, we found that noise correlations, as well as stimulus selectivity and decoding accuracy, varied strongly with the strength of inhibition in the network. We also found that the activity of fast-spiking neurons in our recordings was increased during periods of cortical desynchronization with weak noise correlations. Taken together, these results suggest that the control of network stability by inhibition plays a critical role in modulating the impact of intrinsic cortical dynamics on sensory responses.

Inhibition controls the strength of the large-scale fluctuations that drive noise correlations

Our results are consistent with experiments showing that one global dimension of variability largely explains both the pairwise correlations between neurons [Okun et al., 2015] and the time course of population activity [Ecker et al., 2014]. In our network model, the coordinated, large-scale fluctuations that underlie this global dimension of variability are generated primarily by the interaction between recurrent excitation and

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

adaptation. When inhibition is weak, small deviations from the mean spike rate can be amplified by strong, non-specific, recurrent excitation into population-wide events (up states). These events produce strong adaptation currents in each activated neuron, which, in turn, result in periods of reduced spiking (down states) [Latham et al., 2000, Destexhe, 2009, Curto et al., 2009, Mochol et al., 2015]. The alternations between up states and down states have an intrinsic periodicity given by the timescale of the adaptation currents, but the chaotic nature of the network adds an apparent randomness to the timing of individual events, thus creating intrinsic temporal variability.

The intrinsic temporal variability in the network imposes a history dependence on evoked responses; because of the build-up of adaptation currents during each spiking event, external inputs arriving shortly after an up state will generally result in many fewer spikes than those arriving during a down state [Curto et al., 2009]. This history dependence creates a trial-to-trial variability in the total number of stimulus-evoked spikes that is propagated and reinforced across consecutive stimulus presentations to create noise correlations. However, when the strength of the inhibition in the network is increased, the inhibitory feedback is able to suppress some of the amplification by the recurrent excitation, and the transitions between clear up and down states are replaced by weaker fluctuations of spike rate that vary more smoothly over time. If the strength of the inhibition is increased even further, such that it becomes sufficient to counteract the effects of the recurrent excitation entirely, then the large-scale fluctuations in the network disappear, weakening the history dependence of evoked responses and eliminating noise correlations.

Strong inhibition sharpens tuning curves and enables accurate decoding by stabilizing network dynamics

Numerous experiments have demonstrated that inhibition can shape the tuning curves of cortical neurons, with stronger inhibition generally resulting in sharper tuning [Isaacson and Scanziani, 2011]. The mechanisms involved are still a subject of debate, but this sharpening is often thought to result from structured connectivity that produces differences in the tuning of the excitatory and inhibitory synaptic inputs to individ-

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ual neurons; lateral inhibition, for example, can sharpen tuning when neurons with similar, but not identical, tuning properties inhibit each other. Our results, however, demonstrate that strong inhibition can sharpen tuning in a network without any structured connectivity simply by controlling its dynamics.

In our model, broad tuning curves result from the over-excitability of the network. When inhibition is weak, every external input will eventually excite every neuron in the network because those neurons that receive the input directly will relay indirect excitation to the rest of the network. When inhibition is strong, however, the indirect excitation is largely suppressed, allowing each neuron to respond selectively to only those external inputs that it receives either directly or from one of the few other neurons to which it is strongly coupled. Thus, when inhibition is weak and the network is unstable, different external inputs will trigger similar population-wide events [Bathellier et al., 2012], so the selectivity of the network in this regime is weak and its ability to encode differences between sensory stimuli is poor. In contrast, when inhibition is strong and the network is stable, different external inputs will reliably drive different subsets of neurons, and the activity patterns in the network will encode different stimuli with high selectivity and enable accurate decoding.

Two different dynamical regimes with weak noise correlations

A number of studies have observed that the noise correlations in cortical networks can be extremely weak under certain conditions [Ecker et al., 2010, Renart et al., 2010, Hansen et al., 2012, Pachitariu et al., 2015]. It was originally suggested that noise correlations were weak because the network was in an asynchronous state in which neurons are continuously depolarized with a resting potential close to the spiking threshold [Renart et al., 2010, van Vreeswijk and Sompolinsky, 1996]. Experimental support for this classical asynchronous state has been provided by intracellular recordings showing that the membrane potential of cortical neurons is increased during locomotion [McGinley et al., 2015a] and hyper-arousal [Constantinople and Bruno, 2011], resulting in tonic spiking. However, other experiments have shown that the membrane potential of cortical neurons in behaving animals can also be strongly hyperpolarized with clear fluctuations

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

between up and down states [Sachidhanandam et al., 2013, Tan et al., 2014, McGinley et al., 2015a, Polack et al., 2013].

These apparently conflicting results suggest that there may be multiple dynamical regimes in behaving animals that are capable of producing weak noise correlations. There is mounting evidence suggesting that different forms of arousal may have distinct effects on neural activity [McGinley et al., 2015b]. While most forms of arousal tend to reduce the power of low-frequency fluctuations in membrane potential [Bennett et al., 2013, Polack et al., 2013, McGinley et al., 2015a], locomotion tends to cause a persistent depolarization of cortical neurons and drive tonic spiking. In contrast, task-engagement in stationary animals is generally associated with hyperpolarization and suppression of activity [Vinck et al., 2015, Polack et al., 2013, McGinley et al., 2015a, Otazu et al., 2009, Buran et al., 2014]. The existence of two different dynamical regimes with weak noise correlations was also apparent in our recordings; while some recordings with weak noise correlations resembled the classical asynchronous state with spontaneous activity consisting of strong, tonic spiking (e.g. desynchronized urethane recordings and some awake recordings), other recordings with weak noise correlations exhibited a suppressed state with relatively low spontaneous activity that contained clear, albeit weak, up and down states (e.g. FMM recordings and other awake recordings). Our model was able to accurately reproduce spontaneous activity patterns and generate evoked responses with weak noise correlations in both of these distinct regimes.

In addition to strong inhibition, the classical asynchronous state with strong, tonic spiking appears to require a combination of weak adaptation and an increase in the number of neurons receiving strong tonic input (see parameter sweeps in Figures 2C-D and parameter values for awake mouse V1 recordings in Figure S2). Since large-scale fluctuations arise from the synchronization of adaptation currents across the population, reducing the strength of adaptation diminishes the fluctuations [Destexhe, 2009, Curto et al., 2009, Mochol et al., 2015]. Increasing tonic input also diminishes large-scale fluctuations, but in a different way [Latham et al., 2000]; when a subset of neurons receive increased tonic input, their adaptation currents may no longer be sufficient to silence them for prolonged periods, and the activity of these neurons during what

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

would otherwise be a down state prevents the entire population from synchronizing. When the network in the asynchronous state is driven by an external input, it responds reliably and selectively to different inputs. Because the fluctuations in the network are suppressed and its overall level of activity remains relatively constant, every input arrives with the network in the same moderately-adapted state, so there is no history dependence to create noise correlations in evoked responses.

Unlike in the classical asynchronous state, networks in the suppressed state have slow fluctuations in their spontaneous activity, and the lack of noise correlations in their evoked responses is due to different mechanisms (see parameter values for gerbil A1 FMM recordings in Figure S2). The fluctuations in the hyperpolarized network are only suppressed when the network is driven by external input. In our model, this suppression of the correlated variability in evoked responses is caused by the supralinearity of the feedback inhibition [Rubin et al., 2015]. The level of spontaneous activity driven by the tonic input to each neuron results in feedback inhibition with a relatively low gain, which is insufficient to suppress the fluctuations created by the interaction between recurrent excitation and adaptation. However, when the network is strongly driven by external input, the increased activity results in feedback inhibition with a much higher gain, which stabilizes the network and allows it to respond reliably and selectively to different inputs. This increase in the inhibitory gain of the driven network provides a possible mechanistic explanation for the recent observation that the onset of a stimulus quenches variability [Churchland et al., 2010] and switches the cortex from a synchronized to a desynchronized state [Tan et al., 2014], as well as the suppression of responses to high-contrast stimuli in alert animals [Zhuang et al., 2014].

Experimental evidence for inhibitory stabilization of cortical dynamics

The results of several previous experimental studies also support the idea that strong inhibition can stabilize cortical networks and enhance sensory coding. In vitro studies have shown that pharmacologically reducing inhibition increases the strength of the correlations between excitatory neurons in a graded manner [Sippy and Yuste, 2013]. In vivo whole-cell recordings in awake animals have demonstrated that the stimulus-

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

evoked inhibitory conductance is much larger than the corresponding excitatory conductance [Haider et al., 2013]. This strong inhibition in awake animals quickly shunts the excitatory drive and results in sharper tuning and sparser firing than the balanced excitatory and inhibitory conductances observed under anesthesia. While some of the increased inhibition in awake animals may be due to inputs from other brain areas [Yu et al., 2015], the increased activity of local inhibitory interneurons appears to play an important role [Schneider et al., 2014, Kato et al., 2013, Kuchibhotla et al., 2016]. However, not all studies have observed increased inhibition in behaving animals [Zhou et al., 2014], and the effects of behavioral state on different inhibitory interneuron types are still being investigated [Gentet et al., 2010, Gentet et al., 2012, Polack et al., 2013].

The effects of local inhibition on sensory coding have also been tested directly using optogenetics. While the exact roles played by different inhibitory neuron types are still under investigation [Lee et al., 2014, Seybold et al., 2015], the activation of inhibitory interneurons generally results in sharper tuning, weaker correlations, and enhanced behavioral performance [Wilson et al., 2012, Lee et al., 2012, Chen et al., 2015], while suppression of inhibitory interneurons has the opposite effect, decreasing the signal-to-noise ratio and reliability of evoked responses across trials [Zhu et al., 2015, Chen et al., 2015]. These results demonstrate that increased inhibition enhances sensory processing and are consistent with the overall suppression of cortical activity that is often observed during active behaviors [Otazu et al., 2009, Schneider et al., 2014, Kuchibhotla et al., 2016, Buran et al., 2014]. In fact, one recent study found that the best performance in a detection task was observed on trials in which the pre-stimulus membrane voltage was hyperpolarized and low-frequency fluctuations were absent [McGinley et al., 2015a], consistent with a suppressed, inhibition-stabilized network state.

Neuromodulators and inhibitory control of cortical dynamics

Neuromodulators can exert a strong influence on cortical dynamics by regulating the balance of excitation and inhibition in the network. While the exact mechanisms by which neuromodulators control cortical dy-

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

namics are not clear, several lines of evidence suggest that neuromodulator release serves to enhance sensory processing by increasing inhibition. Increases in acetylcholine (ACh) and norepinephrine (NE) have been observed during wakefulness and arousal [Berridge and Waterhouse, 2003, Jones, 2008], and during periods of cortical desynchronization in which slow fluctuations in the LFP are suppressed [Goard and Dan, 2009, Chen et al., 2015, Castro-Alamancos and Gulati, 2014]. Stimulation of the basal forebrain has been shown to produce ACh-mediated increases in the activity of FS neurons and decrease the variability of evoked responses in cortex [Sakata, 2016, Castro-Alamancos and Gulati, 2014, Goard and Dan, 2009]. In addition, optogenetic activation of cholinergic projections to cortex resulted in increased firing of SOM+ inhibitory neurons and reduced slow fluctuations [Chen et al., 2015]. The release of NE in cortex through microdialysis had similar effects, increasing fast-spiking activity and reducing spontaneous spike rates [Castro-Alamancos and Gulati, 2014], while blocking NE receptors strengthened slow fluctuations in membrane potential [Constantinople and Bruno, 2011]. More studies are needed to tease apart the effects of different neurotransmitters on pyramidal neurons and interneurons [Castro-Alamancos and Gulati, 2014, Chen et al., 2015, Sakata, 2016], but most of the existing evidence is consistent with our results in suggesting that neuromodulators can suppress intrinsic fluctuations and enhance sensory processing in cortical networks by increasing inhibition.

References

Abbott, L. F. and Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural Computation*, 11(1):91–101.

Amit, D. J. and Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex*, 7(3):237–252.

Averbeck, B. B., Latham, P. E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, 7(5):358–366.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Barthó, P., Hirase, H., Monconduit, L., Zugaro, M., Harris, K. D., and Buzsáki, G. (2004). Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. *Journal of Neurophysiology*, 92(1):600–608.

Bathellier, B., Ushakova, L., and Rumpel, S. (2012). Discrete neocortical dynamics predict behavioral categorization of sounds. *Neuron*, 76(2):435–449.

Bennett, C., Arroyo, S., and Hestrin, S. (2013). Subthreshold mechanisms underlying state-dependent modulation of visual responses. *Neuron*, 80(2):350–357.

Berridge, C. W. and Waterhouse, B. D. (2003). The locus coeruleus-noradrenergic system: Modulation of behavioral state and state-dependent cognitive processes. *Brain Research Reviews*, 42(1):33–84.

Buran, B. N., von Trapp, G., and Sanes, D. H. (2014). Behaviorally gated reduction of spontaneous discharge can improve detection thresholds in auditory cortex. *The Journal of Neuroscience*, 34(11):4076–4081.

Cardin, J. A., Carlén, M., Meletis, K., Knoblich, U., Zhang, F., Deisseroth, K., Tsai, L.-H., and Moore, C. I. (2009). Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature*, 459(7247):663–667.

Castro-Alamancos, M. A. and Gulati, T. (2014). Neuromodulators produce distinct activated states in neocortex. *The Journal of Neuroscience*, 34(37):12353–12367.

Chen, N., Sugihara, H., and Sur, M. (2015). An acetylcholine-activated microcircuit drives temporal dynamics of cortical activity. *Nature Neuroscience*, 18(6):892–902.

Cho, K.-H., Jang, J. H., Jang, H.-J., Kim, M.-J., Yoon, S. H., Fukuda, T., Tennigkeit, F., Singer, W., and Rhie, D.-J. (2010). Subtype-specific dendritic ca2+ dynamics of inhibitory interneurons in the rat visual cortex. *Journal of Neurophysiology*, 104(2):840–853.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Churchland, M. M. and Abbott, L. F. (2012). Two layers of neural variability. *Nature Neuroscience*, 15(11):1472–1474.

Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., Newsome, W. T., Clark, A. M., Hosseini, P., Scott, B. B., Bradley, D. C., Smith, M. a., Kohn, A., Movshon, J. A., Armstrong, K. M., Moore, T., Chang, S. W., Snyder, L. H., Lisberger, S. G., Priebe, N. J., Finn, I. M., Ferster, D., Ryu, S. I., Santhanam, G., Sahani, M., and Shenoy, K. V. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature Neuroscience*, 13(3):369–378.

Cohen, M. R. and Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature Neuroscience*, 14(7):811–819.

Cohen, M. R. and Maunsell, J. H. R. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience*, 12(12):1594–1600.

Constantinople, C. M. and Bruno, R. M. (2011). Effects and mechanisms of wakefulness on local cortical networks. *Neuron*, 69(6):1061–1068.

Curto, C., Sakata, S., Marguet, S., Itskov, V., and Harris, K. D. (2009). A simple model of cortical dynamics explains variability and state dependence of sensory responses in urethane-anesthetized auditory cortex. *The Journal of Neuroscience*, 29(34):10600–10612.

de la Rocha, J., Doiron, B., Shea-Brown, E., Josić, K., and Reyes, A. (2007). Correlation between neural spike trains increases with firing rate. *Nature*, 448(7155):802–806.

Destexhe, A. (2009). Self-sustained asynchronous irregular states and Up-Down states in thalamic, cortical and thalamocortical networks of nonlinear integrate-and-fire neurons. *Journal of Computational Neuroscience*, 27(3):493–506.

Ecker, A. S., Berens, P., Cotton, R. J., Subramaniyan, M., Denfield, G. H., Cadwell, C. R., Smirnakis, S. M., Bethge, M., and Tolias, A. S. (2014). State dependence of noise correlations in macaque primary visual cortex. *Neuron*, 82(1):235–248.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Ecker, A. S., Berens, P., Keliris, G. a., Bethge, M., Logothetis, N. K., and Tolias, A. S. (2010). Decorrelated neuronal firing in cortical microcircuits. *Science*, 327(5965):584–587.

Erisken, S., Vaiceliunaite, A., Jurjut, O., Fiorini, M., Katzner, S., Busse, L., Reichardt, W., and Neuroscience, I. (2014). Article Effects of Locomotion Extend throughout the Mouse Early Visual System. *Current Biology*, 24(24):1–9.

Fino, E. and Yuste, R. (2011). Dense inhibitory connectivity in neocortex. *Neuron*, 69(6):1188–1203.

Garcia-Lazaro, J. A., Belliveau, L. A., and Lesica, N. A. (2013). Independent population coding of speech with sub-millisecond precision. *The Journal of Neuroscience*, 33(49):19362–19372.

Gentet, L. J., Avermann, M., Matyas, F., Staiger, J. F., and Petersen, C. C. (2010). Membrane potential dynamics of gabaergic neurons in the barrel cortex of behaving mice. *Neuron*, 65(3):422–435.

Gentet, L. J., Kremer, Y., Taniguchi, H., Huang, Z. J., Staiger, J. F., and Petersen, C. C. (2012). Unique functional properties of somatostatin-expressing gabaergic neurons in mouse barrel cortex. *Nature Neuroscience*, 15(4):607–612.

Goard, M. and Dan, Y. (2009). Basal forebrain activation enhances cortical coding of natural scenes. *Nature Neuroscience*, 12(11):1444–1449.

Goris, R. L. T., Movshon, J. A., and Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nature Neuroscience*, 17(6):858–65.

Haider, B., Häusser, M., and Carandini, M. (2013). Inhibition dominates sensory responses in the awake cortex. *Nature*, 493(7430):97–100.

Hansen, B. J., Chelaru, M. I., and Dragoi, V. (2012). Correlated variability in laminar cortical circuits. *Neuron*, 76(3):590–602.

Harris, K. D. and Thiele, A. (2011). Cortical state and attention. *Nature Reviews Neuroscience*, 12(9):509–523.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Hofer, S. B., Ko, H., Pichler, B., Vogelstein, J., Ros, H., Zeng, H., Lein, E., Lesica, N. A., and Mrsic-Flogel, T. D. (2011). Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nature Neuroscience*, 14(8):1045–1052.

Isaacson, J. S. and Scanziani, M. (2011). How inhibition shapes cortical activity. *Neuron*, 72(2):231–243.

Jones, B. E. (2008). Modulation of cortical activation and behavioral arousal by cholinergic and orexinergic systems. *Annals of the New York Academy of Sciences*, 1129:26–34.

Kato, H. K., Gillet, S. N., Peters, A. J., Isaacson, J. S., and Komiyama, T. (2013). Parvalbumin-expressing interneurons linearly control olfactory bulb output. *Neuron*, 80(5):1218–1231.

Kawaguchi, Y. and Kubota, Y. (1997). Gabaergic cell subtypes and their synaptic connections in rat frontal cortex. *Cerebral Cortex*, 7(6):476–486.

Kuchibhotla, K., Gill, J., Papadoyannis, E., Hindmarsh Sten, T., and Froemke, R. (2016). A synaptic and circuit switch for control of flexible behavior. *Cosyne Abstracts 2016*.

Latham, P. E., Richmond, B. J., Nelson, P. G., and Nirenberg, S. (2000). Intrinsic dynamics in neuronal networks. I. Theory. *Journal of Neurophysiology*, 83(2):808–827.

Lee, A. M., Hoy, J. L., Bonci, A., Wilbrecht, L., Stryker, M. P., and Niell, C. M. (2014). Identification of a brainstem circuit regulating visual cortical state in parallel with locomotion. *Neuron*, 83(2):455–466.

Lee, S.-H., Kwan, A. C., Zhang, S., Phoumthipphavong, V., Flannery, J. G., Masmanidis, S. C., Taniguchi, H., Huang, Z. J., Zhang, F., Boyden, E. S., et al. (2012). Activation of specific interneurons improves ĭ feature selectivity and visual perception. *Nature*, 488(7411):379–383.

Litwin-Kumar, A. and Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature Neuroscience*, 15(11):1498–1505.

London, M., Roth, A., Beeren, L., Häusser, M., and Latham, P. E. (2010). Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. *Nature*, 466(7302):123–127.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Luczak, A., Barthó, P., and Harris, K. D. (2009). Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron*, 62(3):413–425.

Macke, J. H., Buesing, L., Cunningham, J. P., Byron, M. Y., Shenoy, K. V., and Sahani, M. (2011). Empirical models of spiking in neural populations. *Advances in Neural Information Processing Systems*, pages 1350–1358.

Madisen, L., Mao, T., Koch, H., Zhuo, J.-m., Berenyi, A., Fujisawa, S., Hsu, Y.-W. A., Garcia III, A. J., Gu, X., Zanella, S., et al. (2012). A toolbox of cre-dependent optogenetic transgenic mice for light-induced activation and silencing. *Nature Neuroscience*, 15(5):793–802.

Marder, E., Goeritz, M. L., and Otopalik, A. G. (2015). Robust circuit rhythms in small circuits arise from variable circuit components and mechanisms. *Current Opinion in Neurobiology*, 31:156–163.

McGinley, M. J., David, S. V., and McCormick, D. A. (2015a). Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron*, 87(1):179–192.

McGinley, M. J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C. R., Tolias, A. S., Cardin, J. A., and McCormick, D. A. (2015b). Waking state: rapid variations modulate neural and behavioral responses. *Neuron*, 87(6):1143–1161.

Mitchell, J. F., Sundberg, K. a., and Reynolds, J. H. (2009). Spatial Attention Decorrelates Intrinsic Activity Fluctuations in Macaque Area V4. *Neuron*, 63(6):879–888.

Mochol, G., Hermoso-Mendizabal, A., Sakata, S., Harris, K. D., and de la Rocha, J. (2015). Stochastic transitions into silence cause noise correlations in cortical circuits. *Proceedings of the National Academy of Sciences*, 112(11):201410509.

Okun, M., Steinmetz, N. A., Cossell, L., Iacaruso, M. F., Ko, H., Barthó, P., Moore, T., Hofer, S. B., Mrsic-Flogel, T. D., Carandini, M., et al. (2015). Diverse coupling of neurons to populations in sensory cortex. *Nature*, 521(7553):511–515.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Ostojic, S. (2014). Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nature Neuroscience*, 17(4):594–600.

Otazu, G. H., Tai, L.-H., Yang, Y., and Zador, A. M. (2009). Engaging in an auditory task suppresses responses in auditory cortex. *Nature Neuroscience*, 12(5):646–654.

Pachitariu, M., Lyamzin, D. R., Sahani, M., and Lesica, N. A. (2015). State-dependent population coding in primary auditory cortex. *The Journal of Neuroscience*, 35(5):2058–2073.

Pachitariu, M., Petreska, B., and Sahani, M. (2013). Recurrent linear models of simultaneously-recorded neural populations. *Advances in Neural Information Processing Systems*, pages 3138–3146.

Packer, A. M. and Yuste, R. (2011). Dense, unspecific connectivity of neocortical parvalbumin-positive interneurons: a canonical microcircuit for inhibition? *The Journal of Neuroscience*, 31(37):13260–13271.

Polack, P.-O., Friedman, J., and Golshani, P. (2013). Cellular mechanisms of brain state-dependent gain modulation in visual cortex. *Nature Neuroscience*, 16(9):1331–9.

Povysheva, N. V., Zaitsev, A. V., Rotaru, D. C., Gonzalez-Burgos, G., Lewis, D. A., and Krimer, L. S. (2008). Parvalbumin-positive basket interneurons in monkey and rat prefrontal cortex. *Journal of Neurophysiology*, 100(4):2348–2360.

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K. D. (2010). The asynchronous state in cortical circuits. *Science*, 327(5965):587–590.

Roux, L. and Buzsáki, G. (2015). Tasks for inhibitory interneurons in intact brain circuits. *Neuropharmacology*, 88:10–23.

Rubin, D. B., Van Hooser, S. D., and Miller, K. D. (2015). The stabilized supralinear network: A unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron*, 85(2):402–417.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Sachidhanandam, S., Sreenivasan, V., Kyriakatos, A., Kremer, Y., and Petersen, C. C. (2013). Membrane potential correlates of sensory perception in mouse barrel cortex. *Nature Neuroscience*, 16(11):1671–1677.

Sakata, S. (2016). State-dependent and cell type-specific temporal processing in auditory thalamocortical circuit. *Scientific Reports*, 6.

Sakata, S. and Harris, K. D. (2009). Laminar structure of spontaneous and sensory-evoked population activity in auditory cortex. *Neuron*, 64(3):404–418.

Sakata, S. and Harris, K. D. (2012). Laminar-dependent effects of cortical state on auditory cortical spontaneous activity. *Frontiers in Neural Circuits*, 6(109):6.

Schneider, D. M., Nelson, A., and Mooney, R. (2014). A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature*, 513(7517):189–194.

Schölvinck, M. L., Saleem, A. B., Benucci, A., Harris, K. D., and Carandini, M. (2015). Cortical state determines global variability and correlations in visual cortex. *The Journal of Neuroscience*, 35(1):170–178.

Seybold, B. A., Phillips, E. A. K., Schreiner, C. E., and Hasenstaub, A. R. (2015). Inhibitory Actions Unified by Network Integration. *Neuron*, 87(6):1181–1192.

Shadlen, M. N., Britten, K. H., Newsome, W. T., and Movshon, J. a. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *Journal of Neuroscience*, 16(4):1486–1510.

Sippy, T. and Yuste, R. (2013). Decorrelating action of inhibition in neocortical networks. *The Journal of Neuroscience*, 33(23):9813–9830.

Tan, A. Y. Y., Chen, Y., Scholl, B., Seidemann, E., and Priebe, N. J. (2014). Sensory stimulation shifts visual cortex from synchronous to asynchronous states. *Nature*, 509(7499):226–9.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

van Vreeswijk, C. and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726.

Vinck, M., Batista-Brito, R., Knoblich, U., and Cardin, J. A. (2015). Arousal and Locomotion Make Distinct Contributions to Cortical Activity Patterns and Visual Encoding. *Neuron*, 86(3):740–754.

Wilson, N. R., Runyan, C. A., Wang, F. L., and Sur, M. (2012). Division and subtraction by distinct cortical inhibitory networks in vivo. *Nature*, 488(7411):343–348.

Wolf, F., Engelken, R., Puelma-Touzel, M., Weidinger, J. D. F., and Neef, A. (2014). Dynamical models of cortical circuits. *Current Opinion in Neurobiology*, 25:228–236.

Yu, X., Ye, Z., Houston, C. M., Zecharia, A. Y., Ma, Y., Zhang, Z., Uygun, D. S., Parker, S., Vyssotski, A. L., Yustos, R., et al. (2015). Wakefulness is governed by gaba and histamine cotransmission. *Neuron*, 87(1):164–178.

Zhou, M., Liang, F., Xiong, X. R., Li, L., Li, H., Xiao, Z., Tao, H. W., and Zhang, L. I. (2014). Scaling down of balanced excitation and inhibition by active behavioral states in auditory cortex. *Nature Neuroscience*, 17(6):841–850.

Zhu, Y., Qiao, W., Liu, K., Zhong, H., and Yao, H. (2015). Control of response reliability by parvalbumin-expressing interneurons in visual cortex. *Nature Communications*, 6.

Zhuang, J., Bereshpolova, Y., Stoelzel, C. R., Huff, J. M., Hei, X., Alonso, J.-M., and Swadlow, H. A. (2014). Brain State Effects on Layer 4 of the Awake Visual Cortex. *The Journal of Neuroscience*, 34(11):3888–3900.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

FIGURE LEGENDS

Figure 1. Cortical networks exhibit a wide variety of intrinsic dynamics

(A) Multi-neuron raster plots showing examples of a short segment of spontaneous activity from each of our recording types. Each row in each plot represents the spiking of one single unit. Note that recordings made under urethane were separated into two different recording types, synchronized (sync) and desynchronized (desync), as described in the Methods.

(B) The autocorrelation function of the multi-unit activity (MUA, the summed spiking of all neurons in the population in 15 ms time bins) for each example recording. The timescale of the autocorrelation function (the autocorr decay) was measured by fitting an exponential function to its envelope as indicated.

(C) The values of the MUA across time bins sorted in ascending order. The percentage of time bins with zero spikes (the % silence) is indicated.

(D) Scatter plots showing all possible pairwise combinations of the summary statistics for each recording. Each point represents the values for one recording. Colors correspond to recording types as in A. The recordings shown in A are denoted by open circles. The best fit line and the fraction of the variance that it explained are indicated on each plot.

(E) The percent of the variance in the summary statistics across recordings that is explained by each principal component of the values.

Figure 2. A deterministic spiking network model of cortical activity

(A) A schematic diagram of our deterministic spiking network model. An example of a short segment of the intracellular voltage of a model neuron is also shown, along with the corresponding excitatory, inhibitory and adaptation currents.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(B) An example of macroscopic variability in cortical recordings and network simulations. The top two multi-neuron raster plots show spontaneous activity generated by the model. By adding a very small perturbation, in this case one spike added to a single neuron, the subsequent activity patterns of the network can change dramatically. The middle traces show the intracellular voltage of the model neuron to which the spike was added. The bottom two raster plots show a similar phenomenon observed in vivo. Two segments of activity extracted from different periods during the same recording were similar for three seconds, but then immediately diverged.

(C) The autocorrelation function of the MUA measured from network simulations with different model parameter values. Each column shows the changes in the autocorrelation function as the value of one model parameter is changed while all others are held fixed. The fixed values used were $w_I = 0.22$, $w_A = 0.80$, $w_E = 4.50$, $b_1 = 0.03$, $b_0 = 0.013$.

(D) The summary statistics measured from network simulations with different model parameter values. Each line shows the changes in the indicated summary statistic as one model parameter is changed while all others are held fixed. Fixed values were as in C.

Figure 3. Deterministic spiking networks reproduce the dynamics observed in vivo

(A) A schematic diagram illustrating how the parameters of the network model were fit to individual multi-neuron recordings.

(B) Examples of spontaneous activity from different recordings, along with spontaneous activity generated by the model fit to each recording.

(C) The left column shows the autocorrelation function of the MUA for each recording, plotted as in Figure 1. The black lines show the autocorrelation function measured from spontaneous activity generated by the model fit to each recording. The middle column shows the sorted MUA for each recording along with the corresponding model fit. The right column shows the mean pairwise correlations between the spiking activity

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of all pairs of neurons in each recording (after binning activity in 15 ms bins). The colored circles show the correlations measured from the recordings and the black open circles show the correlations measured from from spontaneous activity generated by the model fit to each recording.

Figure 4. Deterministic spiking networks reproduce the noise correlations observed in vivo

(A) Multi-neuron raster plots and PSTHs showing examples of evoked responses from each of our recording types. Each row in each raster plot represents the spiking of one single unit. Each raster plot for each recording type shows the response on a single trial. The PSTH shows the MUA averaged across all presentations of the stimulus. Different stimuli were used for different recording types (see Methods).

(B) A scatter plot showing the mean spike rates and mean pairwise noise correlations (after binning the evoked responses in 15 ms bins) for each recording. Each point represents the values for one recording. Colors correspond to recording types as in A. The recordings shown in A are denoted by open circles. Values are only shown for the 38 of 59 recordings that contained both spontaneous activity and evoked responses.

(C) A schematic diagram illustrating the modelling of evoked responses. We constructed the external input using recordings of responses from more than 500 neurons in the inferior colliculus (IC), the primary relay nucleus of the auditory midbrain that provides the main input to the thalamocortical circuit. We have shown previously that the Fano factors of the responses of IC neurons are close to 1 and the noise correlations between neurons are extremely weak [Garcia-Lazaro et al., 2013], suggesting that the spiking activity of a population of IC neurons can be well described by series of independent, inhomogeneous Poisson processes. To generate the responses of each model network to the external input, we averaged the activity of each IC neuron across trials, grouped the IC neurons by their preferred frequency, and selected a randomly chosen subset of 10 neurons from the same frequency group to drive each cortical neuron.

(D) The top left plot shows the sound waveform presented in the IC recordings used as input to the model

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cortical network. The top right plot shows PSTHs formed by averaging IC responses across trials and across all IC neurons in each preferred frequency group. The raster plots show the recorded responses of two cortical populations on successive trials, along with the activity generated by the network model fit to each recording when driven by IC responses to the same sounds.

(E) A scatter plot showing the noise correlations of responses measured from the actual recordings and from simulations of the network model fit to each recording when driven by IC responses to the same sounds.

Figure 5. Strong inhibition suppresses noise correlations, enhances selectivity, and enables accurate decoding

(A) Scatter plots showing the mean pairwise noise correlations measured from simulations of the network model fit to each recording when driven by external input versus the value of the different model parameters. Colors correspond to recording types as in Figure 4. The recordings shown in Figure 4D are denoted by open circles.

(B) The mean pairwise noise correlations measured from network simulations with different values of the inhibition parameter w_I. The values of all other parameters were held fixed at those fit to each recording. Each line corresponds to one recording. Colors correspond to recording types as in Figure 4.

(C,E) Scatter plots showing tuning width and decoding error, plotted as in A.

(D,F) The tuning width and decoding error measured from network simulations with different values of the inhibition parameter w_I, plotted as in B.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 6. Fast-spiking neurons are more active during periods of cortical desynchronization with weak noise correlations

(A) The cortical synchrony at different points during two recordings from V1 of awake mice, measured as the log of the ratio of low-frequency (3 -10 Hz) LFP power to high-frequency (11 - 96 Hz). The distribution of synchrony values across each recording is also shown. The lines indicate the median of each distribution.

(B) A scatter plot showing the noise correlations measured during trials in which the cortex was in either a relatively synchronized (sync) or desynchronized (desync) state for each recording. Each point indicates the mean pairwise correlations between the spiking activity of all pairs of neurons in one recording (after binning the activity in 15 ms bins). Trials with the highest 50% of synchrony values were classified as sync and trials with the lowest 50% of synchrony values were classified as desync. Values for 13 different recordings are shown.

(C) A scatter plot showing noise correlations versus the mean synchrony for trials with the highest and lowest 50% of synchrony values for each recording. Colors indicate different recordings.

(D) Spectrograms showing the average LFP power during trials with the highest (sync) and lowest (desync) 20% of synchrony values across all recordings. The values shown are the deviation from the average spectrogram computed over all trials.

(E) The average PSTHs of FS and RS neurons measured from evoked responses during trials with the highest (sync) and lowest (desync) 20% of synchrony values across all recordings. The lines show the mean across all neurons, and the error bars indicate +/-1 SEM.

(F) The average spike rate of FS and RS neurons during the period from 0 to 500 ms following stimulus onset, averaged across trials in each synchrony quintile. The lines show the mean across all neurons, and the error bars indicate +/-1 SEM.

(G) The cortical synchrony at different points during a urethane recording, plotted as in A. The line indicates

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the value used to classify trials as synchronized (sync) or desynchronized (desync).

(H) A scatter plot showing the noise correlations measured during trials in which the cortex was in either a synchronized (sync) or desynchronized (desync) state. Values for two different recordings are shown. Each point for each recording shows the noise correlations measured from responses to a different sound.

(I) Spectrograms showing the average LFP power during synchronized and desynchronized trials, plotted as in D.

(J) The average PSTHs of FS and RS neurons during synchronized and desynchronized trials, plotted as in E.

(K) The average spike rate of FS and regular-spiking RS neurons during the period from 0 to 500 ms following stimulus onset during synchronized and desynchronized trials. The bars show the mean across all neurons, and the error bars indicate +/-1 SEM.

Figure 7. Many fast-spiking neurons are silent under ketamine/xylazine anesthesia

(A) The noise correlations measured from recordings of responses to speech in gerbil A1 under ketamine/xylazine (KX) and fentanyl/medetomidine/midazolam (FMM). Each point indicates the mean pairwise correlations between the spiking activity of all pairs of neurons in one recording (after binning the activity in 15 ms bins). Recordings were sorted in order of increasing noise correlations.

(B) The average PSTHs of FS and RS neurons under FMM or KX, plotted as in Figure 6.

(C) The average number of FS and RS neurons in recordings under FMM and KX. The bars show the mean across all recordings, and the error bars indicate +/-1 SEM.

(D) Histograms of the average spike rate of FS neurons during the period from 0 to 500 ms following stimulus onset in recordings under FMM or KX. The arrows indicate the median values.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(E) The summed PSTHs of FS and RS neurons under FMM or KX, plotted as in Figure 6.

(F) The ratio of the total number of spikes from FS and RS neurons during the period from 0 to 500 ms following stimulus onset. Each point shows the value for one recording.

METHODS

All of the recordings analyzed in this study have been described previously. Only a brief summary of the relevant experimental details are provided here.

Mouse V1

The experimental details for the mouse V1 recordings have been previously described [Okun et al., 2015]. Briefly, mice were implanted with head plates under anaesthesia, and after a few days of recovery were accustomed to having their head fixed while sitting or standing in a custom built tube. On the day of the recording, the mice were briefly anaesthetised with isoflurane, and a small craniectomy above V1 was made. Recordings were performed at least 1.5h after the animals recovered from the anaesthesia. Buzsaki32 or A4x8 silicon probes were used to record the spiking activity of populations of neurons in the infragranular layers of V1.

Visual stimuli were presented on two of the three available LCD monitors, positioned 25 cm from the animal and covering a field of view of 120 60, extending in front and to the right of the animal. Visual stimuli consisted of multiple presentations of natural movie video clips. For recordings of spontaneous activity, the monitors showed a uniform grey background.

Rat A1

The experimental procedures for the rat A1 recordings have been previously described [Luczak et al., 2009]. Briefly, head posts were implanted on the skull of Sprague Dawley rats (300500 g) under ketaminexylazine anesthesia, and a hole was drilled above the auditory cortex and covered with wax and dental acrylic. After

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

recovery, each animal was trained for 68 d to remain motionless in the restraining apparatus for increasing periods (target, 12 h). On the day of the recording, each animal was briefly anesthetized with isoflurane and the dura resected; after a 1 h recovery period, recording began. The recordings were made from infragranular layers of auditory cortex with 32-channel silicon multi-tetrode arrays.

Sounds were delivered through a free-field speaker. As stimuli we used pure tones (3, 7, 12, 20, or 30 kHz at 60 dB). Each stimulus had duration of 1s followed by 1s of silence.

Gerbil A1

The gerbil A1 recordings have been described in detail previously [Pachitariu et al., 2015]. Briefly, adult male gerbils (70-90 g, P60-120) were anesthetized with one of three different anesthetics: ketamine/xylazine (KX), fentanyl/medetomidine/midazolam (FMM), or urethane. A small metal rod was mounted on the skull and used to secure the head of the animal in a stereotaxic device in a sound-attenuated chamber. A craniotomy was made over the primary auditory cortex, an incision was made in the dura mater, and a 32-channel silicon multi-tetrode array was inserted into the brain. Only recordings from A1 were analyzed. Recordings were made between 1 and 1.5 mm from the cortical surface (most likely in layer V).

Sounds were delivered to speakers coupled to tubes inserted into both ear canals for diotic sound presentation along with microphones for calibration. Repeated presentations of a 2.5 s segment of human speech were presented at a peak intensity of 75 dB SPL. For analyses of responses to different speech tokens, seven 0.25 s segments were extracted from the responses to each 2.5 s segment.

Gerbil IC

The gerbil IC recordings have been described in detail previously [Garcia-Lazaro et al., 2013]. Recordings were made under ketamine/xylazine anesthesia using a multi-tetrode array placed in the low-frequency laminae of the central nucleus of the IC. Experimental details were otherwise identical to those for gerbil A1. In addition to the human speech presented during the A1 recordings, tones with a duration of 75 ms and frequencies between 256 Hz and 8192 Hz were presented at intensities between 55 and 85 dB SPL with a

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

75 ms pause between each presentation.

Spiking network model

We developed a network model using conductance-based quadratic integrate and fire neurons. There are three currents in the model: an excitatory, an inhibitory and an adaptation current. The subthreshold membrane potential for a single neuron i obeys the equation

$$\tau_m \frac{\mathrm{d}V_i}{\mathrm{d}t} = -(V_i - E_L) * (V_i - V_{th}) - g_{E_i}(V_i - E_E) - g_{I_i}(V_i - E_I) - g_{A_i}(V_i - E_A).$$

When $V > V_{th}$, a spike is recorded in the neuron and the neuron's voltage is reset to $V_{reset} = 0.9V_{th}$. For simplicity, we set $V_{th} = 1$ and the leak voltage $E_L = 0$. The excitatory voltage $E_E = 2V_{th}$ and $E_I = E_A = -0.5V_{th}$. Each of the conductances has a representative differential equation which is dependent on the spiking of the neurons in the network at the previous time step, $\mathbf{s}_{t-1}$. The excitatory conductance obeys

$$\tau_E \frac{\mathrm{d}\mathbf{g}_E}{\mathrm{d}t} = -\mathbf{g}_E + J\mathbf{s}_{t-1} + \mathbf{b}.$$

where J is the matrix of excitatory connectivity and $\mathbf{b}$ is the vector of tonic inputs to the neurons. The matrix of connectivity is random with a probability of 5% for the network of 512 neurons and their connectivities are randomly chosen from a uniform distribution between 0 and w_E. The tonic inputs $\mathbf{b}$ have a minimum value b_0, which we call the tonic input baseline added to a random draw from an exponential distribution with mean b_1, which we call the tonic input spread, such that for neuron i, $\mathbf{b}(i) = b_0 + \text{exprnd}(b_1)$. The inhibitory conductance obeys

$$\tau_I \frac{\mathrm{d}\mathbf{g}_I}{\mathrm{d}t} = -\mathbf{g}_I + w_I * (\exp(\sum \mathbf{s}_{t-1} * c) - 1).$$

where c controls the gain of the inhibitory conductance.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The adaptation conductance obeys

$$\tau_A \frac{\mathrm{d}\mathbf{g}_A}{\mathrm{d}t} = -\mathbf{g}_A + w_A \mathbf{s}_{t-1}.$$

The simulations are numerically computed using Eulers method with a time-step of 0.75 ms (this was the lock-out window used for spike-sorting the in vivo recordings and allowed for fast simulations). To avoid numerical instabilities at low voltages, we rectified the voltages at the activation potential of the inhibitory conductance. Each parameter set was simulated for 900 seconds. The timescales are set to $\tau_m = 20$ ms, $\tau_E = 5.10$ ms, $\tau_I = 3.75$ ms, $\tau_A = 375$ ms, and the inhibitory non-linearity controlled by $c = 0.25$. The remaining five parameters (w_I, w_A, w_E, b_1, and b_0) were fit to the spontaneous activity from multi-neuron recordings using the techniques described below. Their ranges were (0.01-0.4), (0.4-1.45), (2.50-5.00), (0.005-0.10), and (0.0001-0.05) respectively.

To illustrate the ability of the network to generate activity patterns with macroscopic variability, we simulated spontaneous activity with a parameter set that produces up and down state dynamics. Figure 2A shows the membrane potential of a single neuron in this simulation and its conductances at each time step. Figure 2B shows the model run twice with the same set of initial conditions and parameters, but with an additional single spike inserted into the network on the second run (circled in green).

Parameter sweep analysis

Figure 2C and D summarize the effects of changing each parameter on the structure of the spontaneous activity patterns generated by the model. We held the values for all but one parameter fixed and swept the other parameter across a wide range of values. The fixed parameter values were set to approximately the median values obtained from fits to all in vivo recordings. Figure 5 B, D, and F and Figure S4 show the results of similar parameter sweep analyses for stimulus-driven activity with the external input to the network derived from IC activity as described below. For these analyses, the values of the parameters that were not swept were fixed at those fit to each individual recording.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

GPU implementation

We accelerated the network simulations by programming them on graphics processing units (GPUs) such that we were able to run them at 650x real time with 15 networks running concurrently on the same GPU. We were thus able to simulate ≈10000 seconds of simulation time in 1 second of real time. To achieve this acceleration, we took advantage of the large memory bandwidth of the GPUs. For networks of 512 neurons, the state of the network (spikes, conductances and membrane potentials) can be stored in the very fast shared memory available on each multiprocessor inside a GPU. A separate network was simulated on each of the 8 or 15 multiprocesssors available (video cards were GTX 690 or Titan Black). Low-level CUDA code was interfaced with Matlab via mex routines.

Summary statistics

Several statistics of spikes were used to summarize the activity patterns observed in the in vivo recordings and in the network simulations. Because there were on the order of 50 neurons in each recording, all of the statistics below were influenced by small sample effects. To replicate this bias in the analysis of network simulations, we subsampled 50 neurons from the network randomly and computed the same statistics we computed from the in vivo recordings.

The noise correlations between each pair of neurons in each recording were measured from responses to speech. The response of each neuron to each trial was represented as a binary vector with 15 ms time bins. The total correlation for each pair of neurons was obtained by computing the correlation coefficient between the actual responses. The signal correlation was computed after shuffling the order of repeated trials for each time bin. The noise correlation was obtained by subtracting the signal correlation from the total correlation.

The multi-unit activity (MUA) was computed as the sum of spikes in all neurons in bins of 15 ms.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The autocorrelation function of the MUA at time-lag τ was computed from the formula

$$\text{ACF}(\tau) = \frac{1}{N_{samples}} \sum \text{MUA}(t) * \text{MUA}(t+\tau)$$

To measure the autocorrelation timescale, we fit one side of the ACF with a parametric function

$$\text{ACF}(\tau) \sim a \exp(-\tau/T) \cdot \cos(\tau/(2\pi t_{\text{period}}))$$

where a is an overall amplitude, T is a decay timescale and t_{period} is the oscillation period of the autocorrelation function. There was not always a significant oscillatory component in the ACF, but the timescale of decay accurately captured the duration over which the MUA was significantly correlated.

Parameter searches

To find the best fit parameters for each individual recording, we tried to find the set of model parameters for which the in vivo activity and the network simulations had the same statistics. We measured goodness of fit for each of the three statistics: pairwise correlations, the MUA distribution, the MUA ACF. Each statistic was normalized appropriately to order 1, and the three numbers obtained were averaged to obtain an overall goodness of fit.

The distance measure D_c between the mean correlations c_θ obtained from a set of parameters θ and the mean correlations c_n in recording n was simply the squared error $D_c(c_n, c_\theta) = (c_n - c_\theta)^2$. This was normalized by the variance of the mean correlations across recordings to obtain the normalized correlation cost Cost_c, where $\langle x_n \rangle_n$ is used to denote the average of a variable x over recordings indexed by n.

$$\text{Cost}_c = \frac{D_c(c_n, c_\theta)}{\langle D_c(c_n, \langle c_n \rangle) \rangle}$$

The distance measure D_m for the MUA distribution was the squared difference summed over the order

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

rank bins k of the distribution $D_m(\text{MUA}_n, \text{MUA}_\theta) = \sum_k (\text{MUA}_n(k) - \text{MUA}_\theta(k))^2$. This was normalized by the distance between the data MUA and the mean data MUA. In other words, the cost measures how much closer the simulation is to the data distribution than the average of all data distributions.

$$\text{Cost}_m = \frac{D_m(\text{MUA}_n, \text{MUA}_\theta)}{D_m(\text{MUA}_n, \langle \text{MUA}_n \rangle)}$$

Finally, the distance measure D_a for the autocorrelation function of the MUA was the squared difference summed over time lag bins t of the distribution $D_a(\text{ACF}_n, \text{ACF}_\theta) = \sum_t (\text{ACF}_n(t) - \text{ACF}_\theta(t))^2$. This was normalized by the distance between the data ACF and the mean data ACF.

$$\text{Cost}_a = \frac{D_a(\text{ACF}_n, \text{ACF}_\theta)}{D_a(\text{ACF}_n, \langle \text{ACF}_n \rangle)}$$

The total cost of parameters θ on recording n is therefore $\text{Cost}(n, \theta) = \text{Cost}_c + \text{Cost}_m + \text{Cost}_a$. Approximately two million networks were simulated on a grid of parameters for 600 seconds each of spontaneous activity, and their summary statistics (c_θ, MUA_θ and ACF_θ) were retained. The Cost was smoothed for each recording by averaging with the nearest 10 other simulations on the grid. This ensured that some of the sampling noise was removed and parameters were estimated more robustly. The best fit set of parameters was chosen as the minimizer of this smoothed cost function, on a recording by recording basis.

Stimulus-driven activity

Once the simulated networks were fit to the spontaneous neuronal activity, we drove them with an external input to study their evoked responses. The stimulus was either human speech (as presented during our gerbil A1 recordings) or pure tones. The external input to the network was constructed using recordings from 563 neurons from the inferior colliculus (IC). For all recordings in the IC the mean pairwise noise correlations were near-zero and the Fano Factors of individual neurons were close to 1 [Garcia-Lazaro et al., 2013], suggesting that responses of IC neurons on a trial-by-trial basis are fully determined by the

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

stimulus alone, up to Poisson-like variability. Thus, we averaged the responses of IC neurons over trials and drove the cortical network with this trial-averaged IC activity. We binned IC neurons by their preferred frequency in response to pure tones, and drove each model cortical neuron with a randomly chosen subset of 10 neurons from the same preferred-frequency bin. We rescaled the IC activity so that the input to the network had a mean value of 0.06 and a maximum value of 0.32, which was three times greater than the average tonic input.

We kept the model parameters fixed at the values fit to spontaneous activity and drove the network with 330 repeated presentations of the stimulus. We then calculated the statistics of the evoked activity. Noise correlations were measured in 15-ms bins as the residual correlations left after subtracting the mean response of each neuron to the stimulus across trials:

$$c_{ij} = \frac{1}{N_{samples}} \sum_t (s_i(t) - < s_i(t) >)(s_j(t) - < s_j(t) >)$$

where $s_i(t)$ is the summed spikes of neuron i in a 15-ms bin and $< s_i(t) >$ is the mean response of neuron i to the stimulus. The noise correlation value given for each recording is the mean of c_{ij}.

Tuning width

To determine tuning width to sound frequency, we used responses of IC neurons to single tones as inputs to the model network. The connections from IC to the network were the same as described in the previous section. Because the connectivity was tonotopic and IC responses are strongly frequency tuned, the neurons in the model network inherited the frequency tuning. We did not model the degree of tonotopic fan-out of connections from IC to cortex and, as a result, the tuning curves of the model neurons were narrow relative to those observed in cortical recordings [Pachitariu et al., 2015]. We chose the full width of the tuning curve at half-max as a standard measure of tuning width.

Decoding tasks

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We computed decoding error for a classification task in which the single-trial activity of all model neurons was used to infer which of seven different speech tokens was presented. The classifier was built on training data using a linear discriminant formulation in which the Gaussian noise term was replaced by Poisson likelihoods. Specifically, the activity of a neuron for each 15-ms bin during the response to each token was fit as a Poisson distribution with the empirically-observed mean. To decode the response to a test trial, the likelihood of each candidate token was computed and the token with the highest likelihood was assigned as the decoded class. This classifier was chosen because it is very fast and can be used to model Poisson-like variables, but we also verified that it produced decoding performance as good as or better than classical high-performance classifiers like support vector machines.

Classifying FS and RS neurons

We classified fast-spiking and regular-spiking neurons based on their spike shape [Okun et al., 2015]. We determined the trough-to-peak time of the mean spike waveform after smoothing with a gaussian kernel of $\sigma = 0.5$ samples. The distribution of the trough-to-peak time τ was clearly bimodal in all types of recordings. Following [Okun et al., 2015] we classified FS neurons in the awake data with $\tau < 0.6ms$ and RS neurons with $\tau > 0.8ms$. The distributions of τ in the anesthetized data, although bimodal, did not have a clear separation point, so we conservatively required $\tau < 0.4$ms to classify an FS cell in these recordings and $\tau > 0.65$ms to classify RS neurons (see Figure S5). The rest of the neurons were not considered for the plots in Figures 7 and 8 and are shown in gray on the histogram in Figure S5.

Local field potential

The low-frequency potential (LFP) was computed by low-pass filtering the raw signal with a cutoff of 300 Hz. Spectrograms with adaptive time-frequency resolution were obtained by filtering the LFP with Hamming-windowed sine and cosine waves and the spectral power was estimated as the sum of their squared amplitudes. The length of the Hamming-window was designed to include two full periods of the sine and cosine function at the respective frequency, except for frequencies of 1 Hz and above 30 Hz, where the window

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

length was clipped to a single period of the sine function at 1 Hz and two periods of the sine function at 30 Hz respectively. The synchrony level was measured as the log of the ratio of the low to high frequency power (respective bands: 3-10 Hz and 11-96 Hz, excluding 45-55 Hz to avoid the line noise). We did not observe significant gamma power peaks except for the line noise, in either the awake or anesthetized recordings.

Dividing trials by synchrony

We computed a synchrony value for each trial in the 500-ms window following stimulus onset. For urethane recordings, the values had a clear bimodal distribution and we separated the top and bottom of these distributions into synchronized and desynchronized trials respectively. For awake recordings, the synchrony index was not clearly bimodal, but varied across a continuum of relatively synchronized and desynchronized states. To examine the effect of synchrony on noise correlations, we sorted all trials by their synchrony value, classified the 50% of trials with the lowest values as desynchronized and the 50% of trials with the highest values as synchronized, and computed the noise correlations for each set of trials for each recording. To examine the effect of synchrony on FS and RS activity, we pooled all trials from all recordings, divided them into quintiles by their synchrony value, and computed the average spike rates of FS and RS neurons for each set of trials. For Figures 6 and 7, noise correlations were computed aligned to the stimulus onsets in windows of 500ms, to match the window used for measuring FS and RS activity as well as LFP power.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

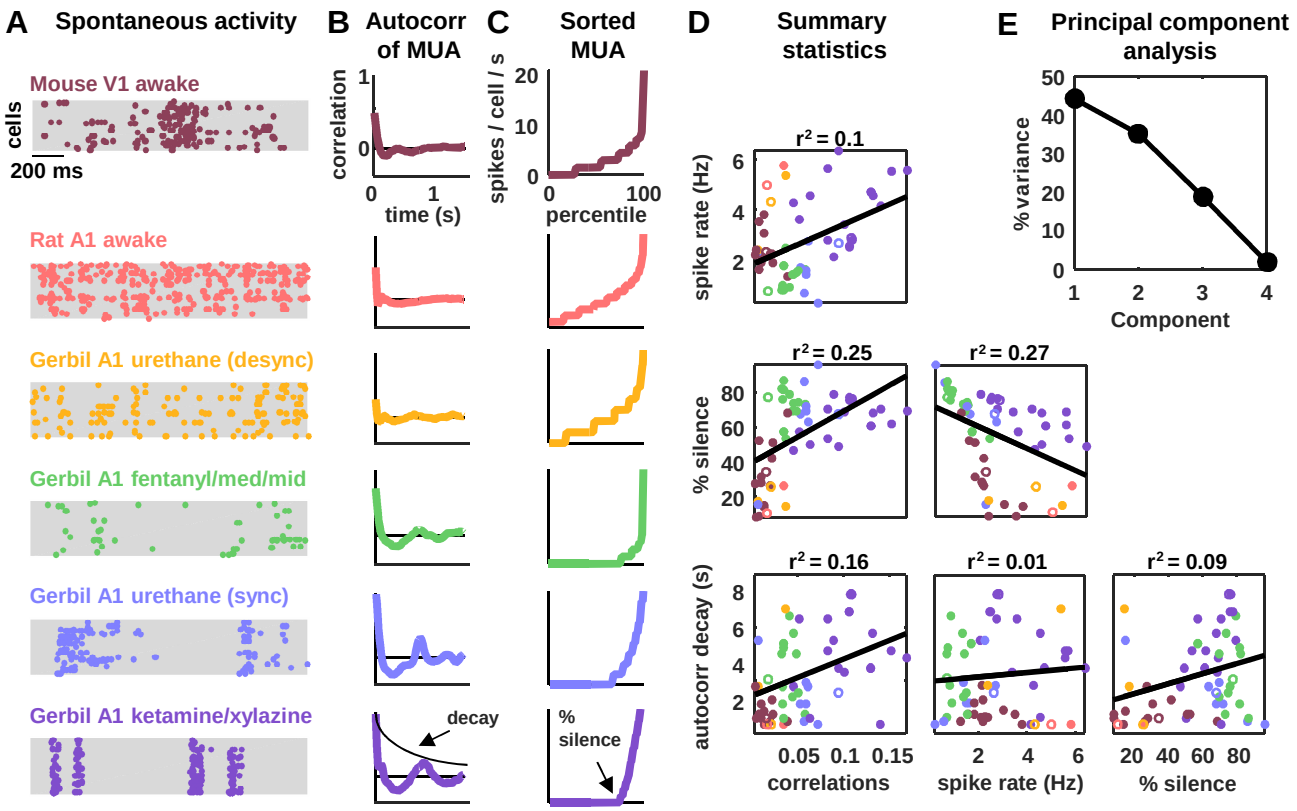

Figure 1

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

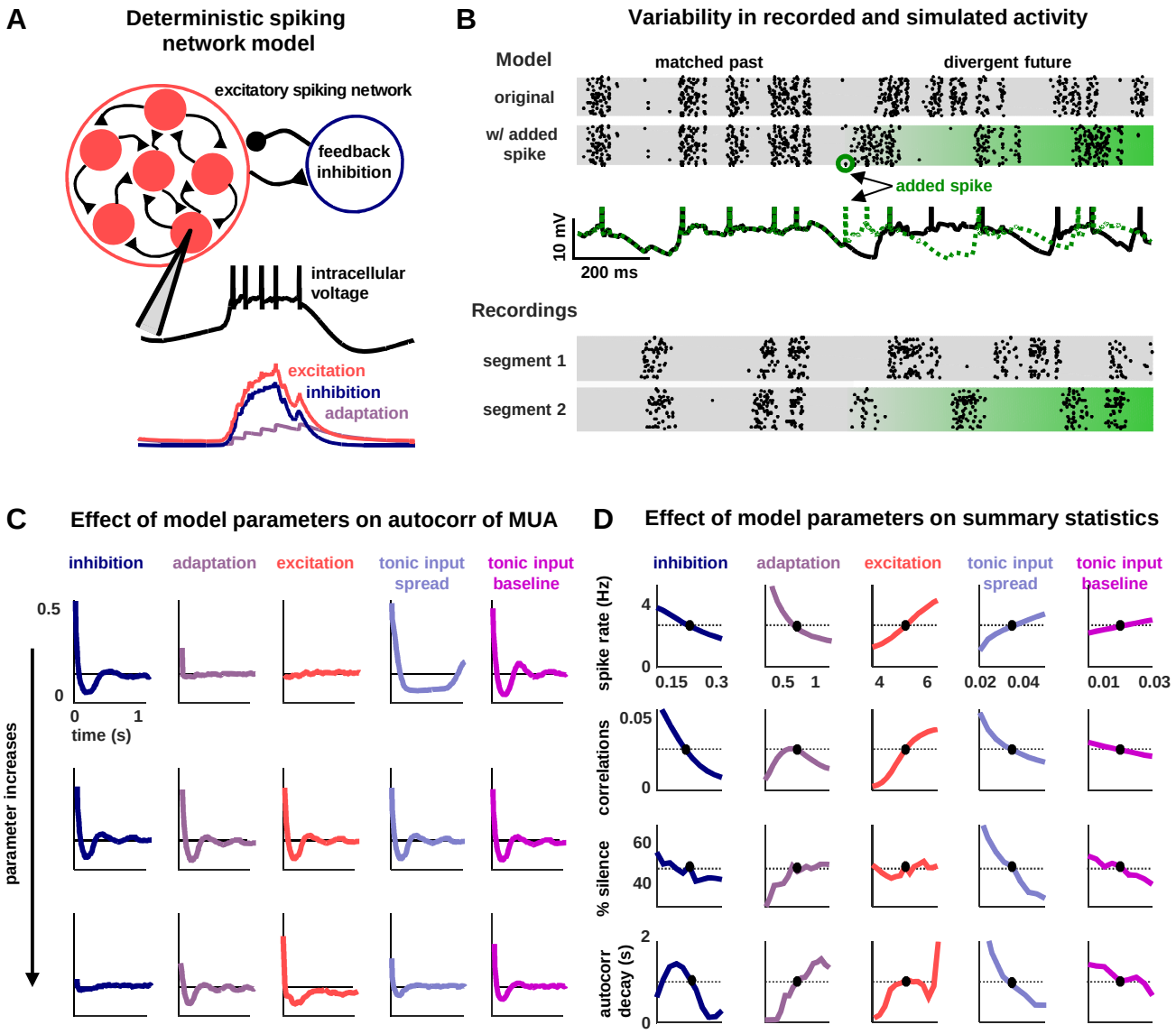

Figure 2

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

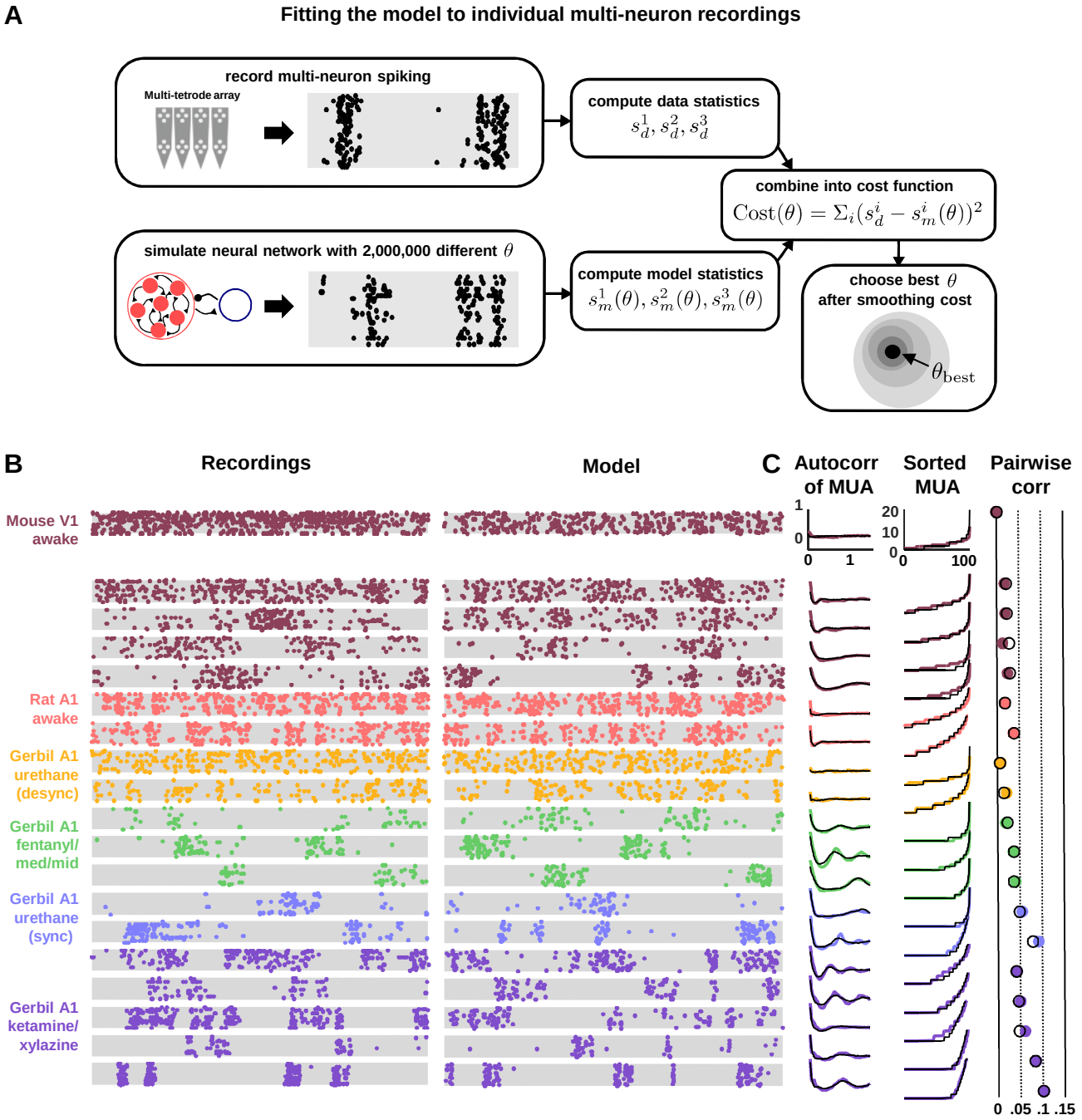

Figure 3

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

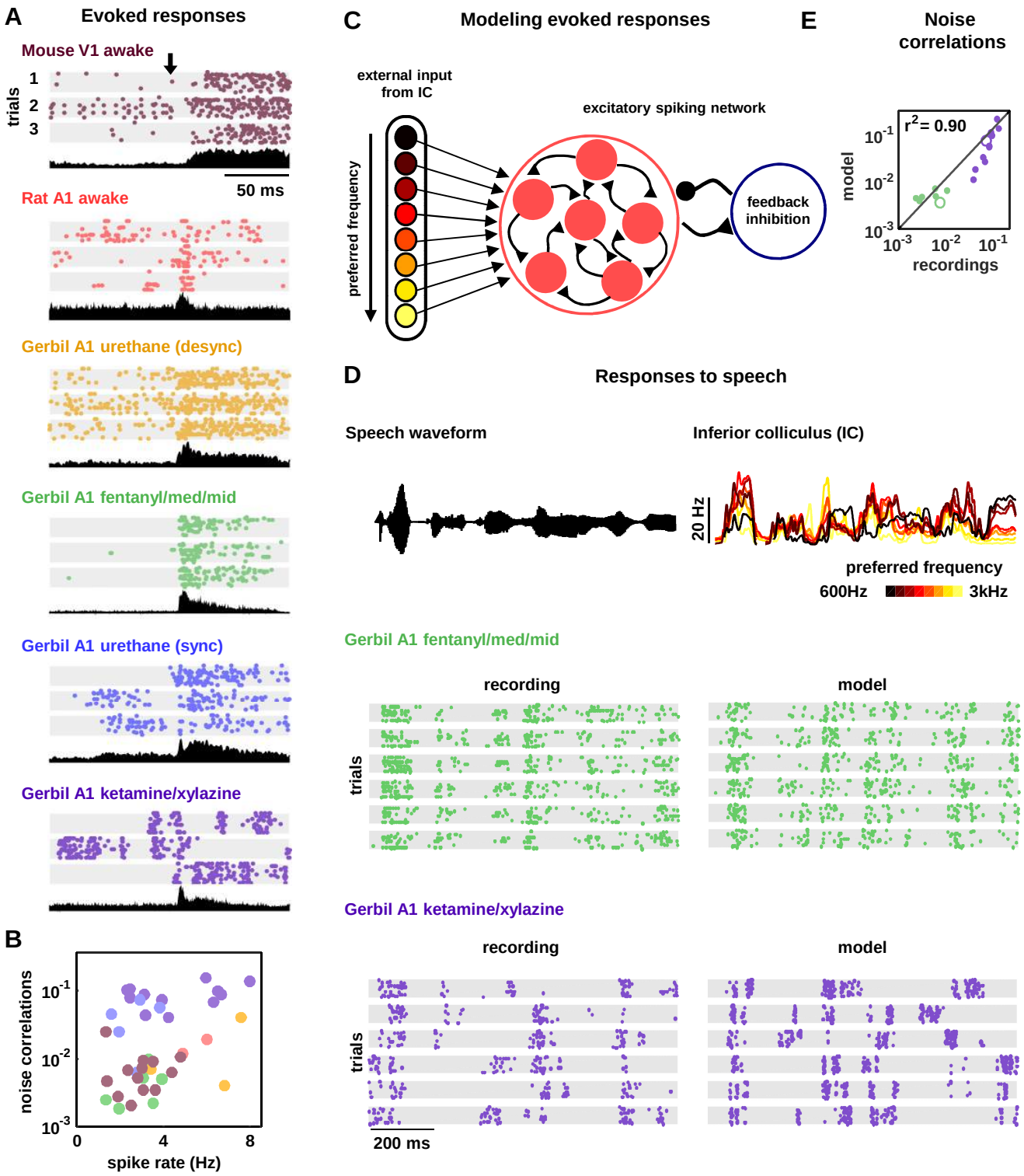

Figure 4

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

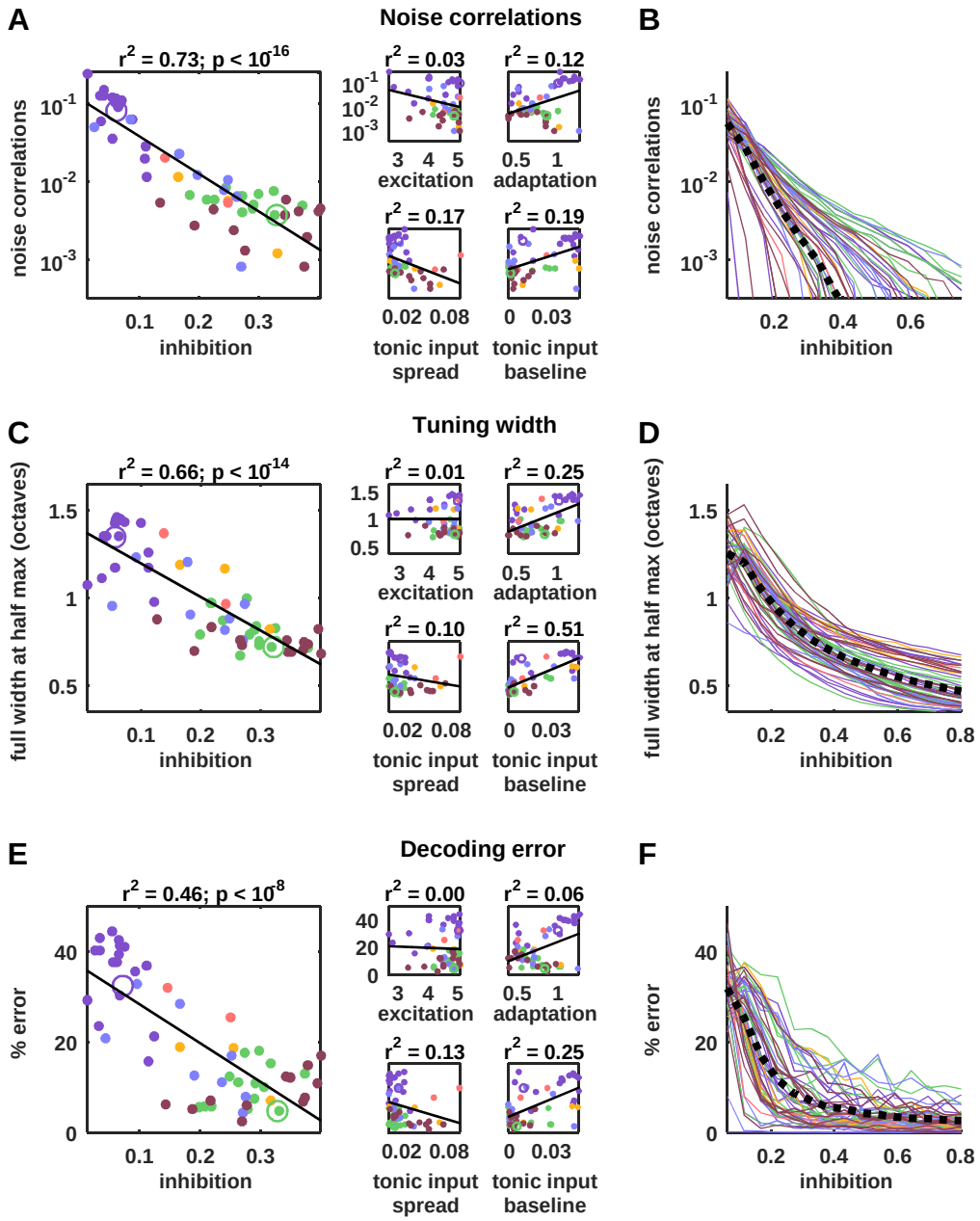

Figure 5

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

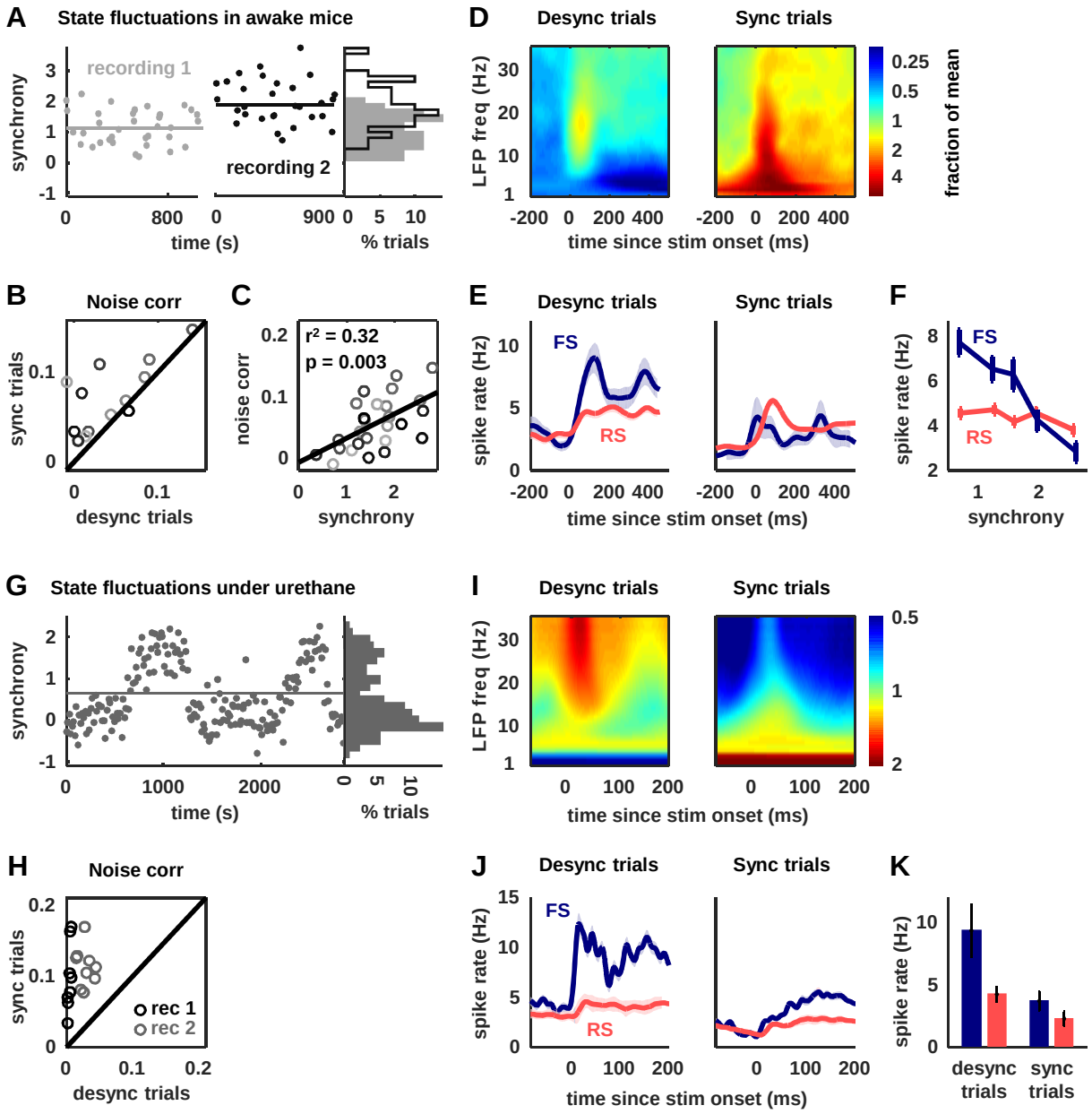

Figure 6

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

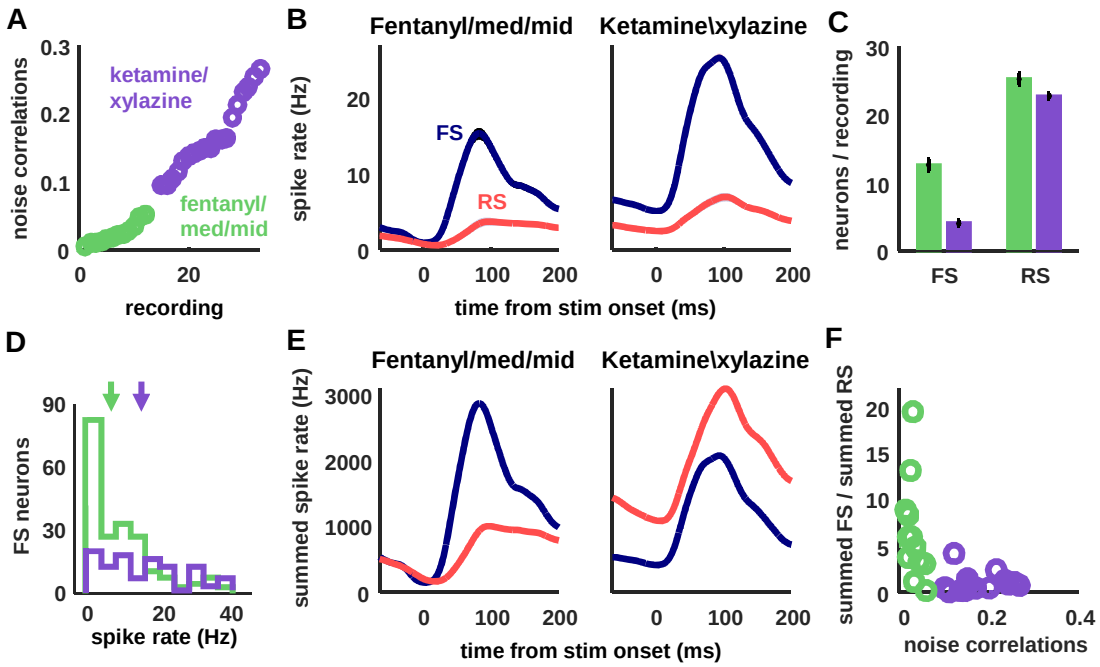

Figure 7

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplemental information for
Inhibitory control of correlated variability in cortical networks

Marius Pachitariu[*,1], Carsen Stringer[*,2], Michael Okun[1], Peter Bartho[3], Kenneth Harris[1], Peter Latham[2], Maneesh Sahani[2], Nicholas Lesica[4]

[*] equal contribution first authors

[1] Institute of Neurology, UCL, London, UK

[2] Gatsby Computational Neuroscience Unit, UCL, London, UK

[3] Hungarian Academy of Sciences, Budapest, Hungary

[4] Ear Institute, UCL, London, UK

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Statistics for all fits. Related to Figures 1 and 3

We fit the model to three statistics: (1) the autocorrelation function of the multi-unit activity (MUA, the summed spiking of all neurons in the population in 15 ms time bins), (2) the values of the MUA across time bins sorted in ascending order, and (3) the mean pairwise correlations across all pairs of neurons (in 15 ms time bins). The statistics for all 59 recordings are shown here in color. The model was fit to each of these recordings and the statistics of the activity generated by the model are shown in black.

Figure S2. Costs and parameter fits. Related to Figures 3 and 5

(A) The three separate terms that combine into our cost function are shown for each recording. Open circles indicate datasets shown in Figure 3.

(B) All parameter fits for each recording. Colors are used to indicate recordings of the same type. A small jitter was added to the horizontal location of each point. Black lines indicate median values for each recording type.

(C) A three-dimensional principal component analysis shows that recordings generally cluster by type, but there is considerable variability both across and within recording types. The inset shows the 5-dimensional PCs.

Figure S3. Analysis of local minima. Related to Figures 3 and 5

Parameter identifiability has recently been raised as a potential problem in interpreting the results of network simulations. To mitigate this problem, we designed our model to have a very small number of parameters and we fit three different functions of the recordings, two of which varied as a function of time or rank. To confirm

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

that the analysis of parameter combinations other than those corresponding to the global minimum of the cost function for each recording would not lead to a different interpretation of our results, we also considered local minima in regions of parameter space that were distant from the global minimum. It is possible that such local minima correspond to parameter regimes that are qualitatively different from the global minimum, yet still capture the statistics of the recordings relatively well. We found the parameters corresponding to local minima did not consistently emphasize the role of any parameter other than inhibition; the strength of inhibitory feedback remained the dominant influence on noise correlations, even for local minima far removed from the global minimum.

(A) A schematic diagram showing an example nonlinear cost function. Several different threshold values are indicated by the colored lines. All costs below threshold are considered and the parameter θ furthest away from the global minimum is chosen to plot in (B).

(B) As the threshold value is increased from the global minimum, the distance of the θ with Cost(θ) $<$ threshold that is furthest from the global minimum is plotted. Discontinuities are visible when the threshold surpasses values at local minima.

(C) Same as B, but for the actual model fits to each recording. The values on the vertical axis are specified in terms of the grid spacing used for the Monte Carlo simulations. While some discontinuities are visible, the functions tend to increase gradually.

(D) For each threshold value, we computed the r^2 between the value of each of the five model parameters and noise correlations, as in Figure 5A. This analysis shows that considering local minima situated far from the global minimum serves only to diminish the relationship between inhibition and noise correlations, without revealing any strong relationships between noise correlations and any other parameter.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure S4. Parameter sweeps for responses to external input. Related to Figures 4 and 5

Parameter sweeps were performed for each recording and each parameter. Each line corresponds to the model fit to one of the 59 recordings. All parameters were fixed at the values fit to each recording, except for the parameter indicated on the horizontal axis, which was swept across a wide range. Activity was generated from the model with these parameters and driven by the IC-derived external input. The spike rate, noise correlations, tuning width and percent decoding error were computed as described for Figure 5.

Figure S5. Classification of neuron types by spike width. Related to Figures 6 and 7

The spike widths of the classified spike waveforms are plotted as histograms for the different recording types. The cells classified as FS based on their width are colored in red and the RS in blue. These were the cells used to compute FS and RS spike rates in Figures 6 and 7. Cells that were not clearly part of either class, which were not included in the analyses in Figures 6 and 7, are shown in gray.

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Animals	Brain region	State / anesthesia	# recordings	# animals	Spont duration (s)	Stimulus type	Trials x duration (s)	# neurons	Figures	Comments
Rats	A1	Awake	2	2	750	Tones	~750 x 1s	82 total	1,3,4,5	only cells >0.5 Hz
Mice	V1	Awake	13	8	300	Natural movie	~30 x 3s	72 FS, 722 RS	1,3,4,5,6	only cells >0.5 Hz
Gerbils	A1	Ketamine/ Xylazine	18	7	600	Speech	~256 x 2.5s	92 FS, 501 RS	1,3,4,5,7	
Gerbils	A1	Fentanyl/ Medetomdine/ Midazolam	15	8	600	Speech	~256 x 2.5s	204 FS, 381 RS	1,3,4,5,7	
Gerbils	A1	Urethane synchronized	8	2	600	Speech	~256 x 2.5s	48 FS, 140 RS	1,3,4,5,6	
Gerbils	A1	Urethane desynchronized	3	2	600	Speech	~256 x 2.5s	66 FS 220 RS	1,3,4,5,6	

Table S1: Metadata for all recordings

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

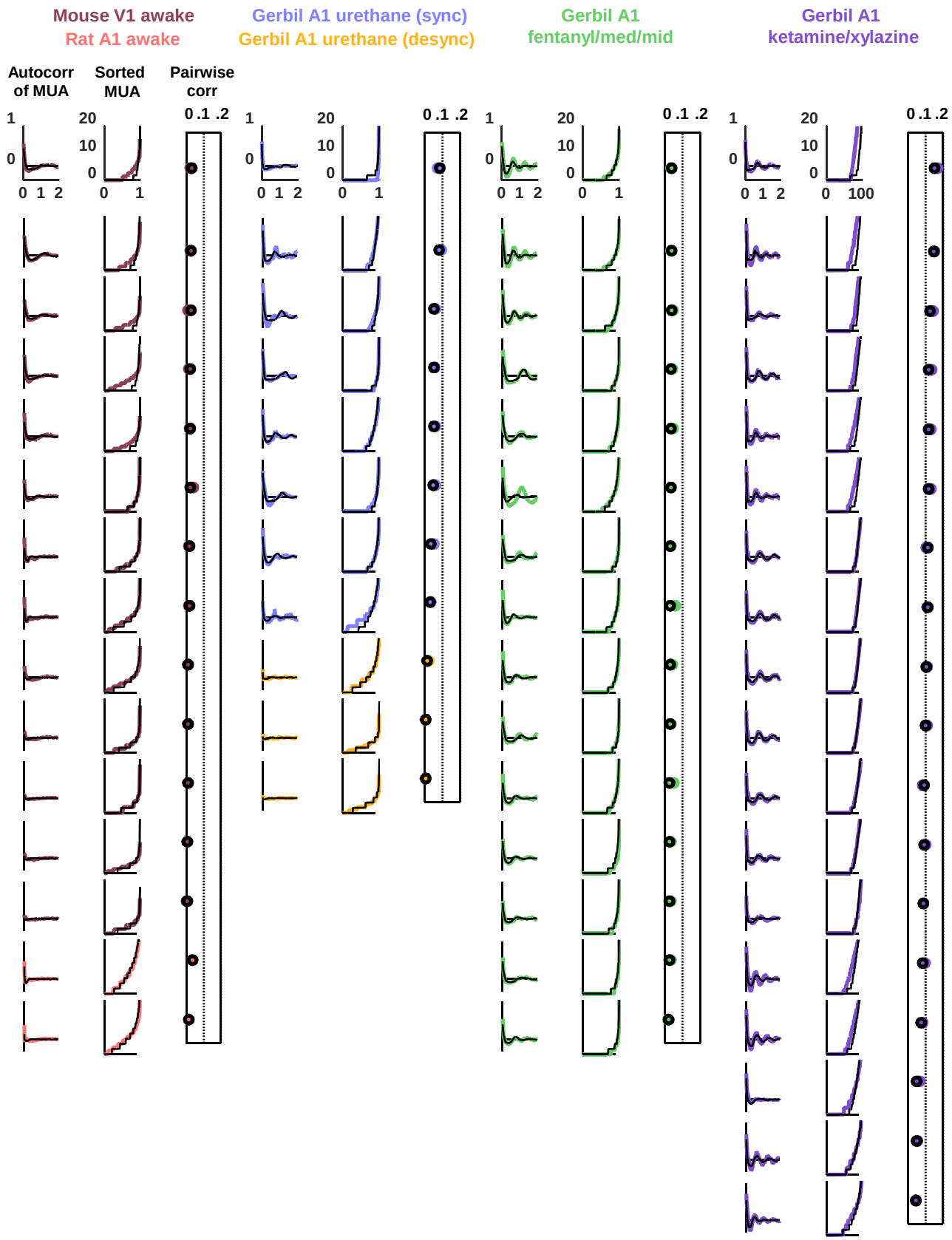

Figure S1. Related to Figures 1 and 3

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

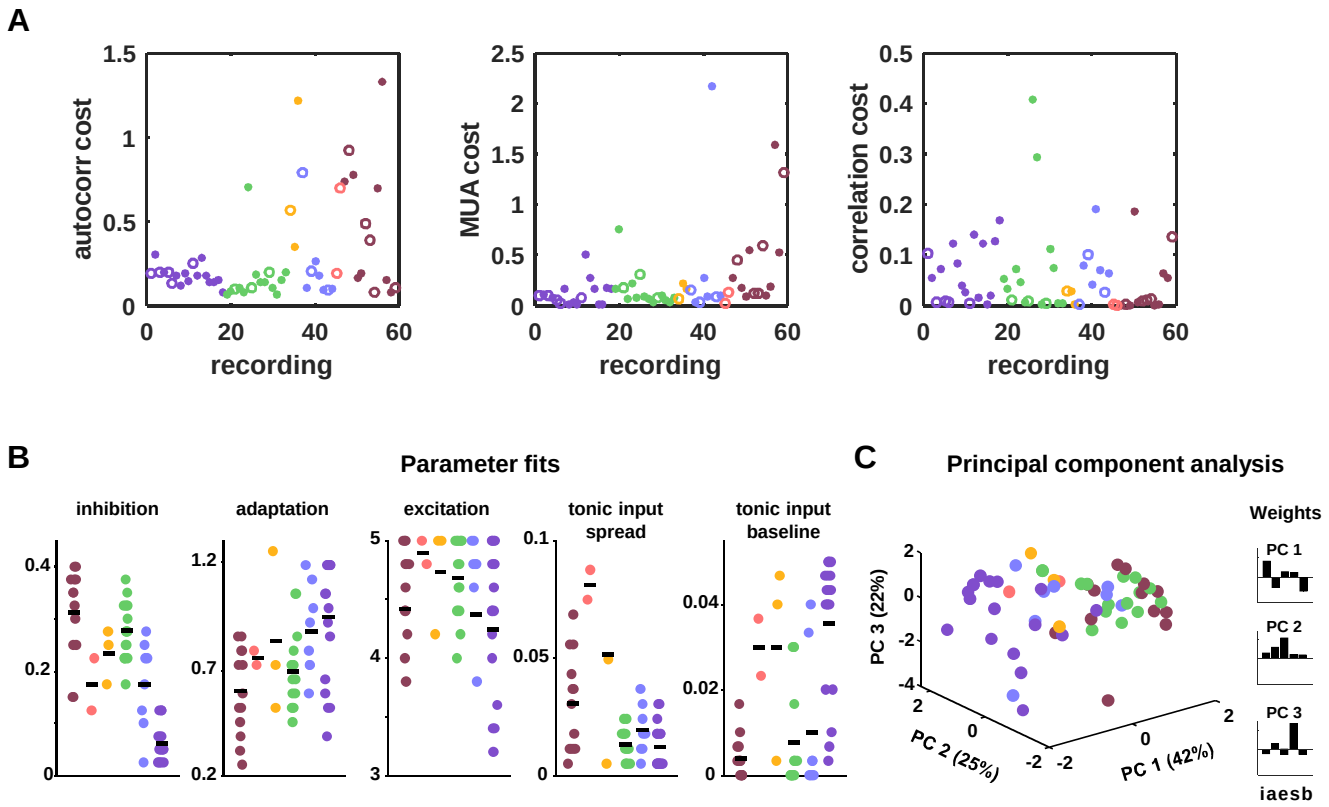

Figure S2: Related to Figures 3 and 5

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

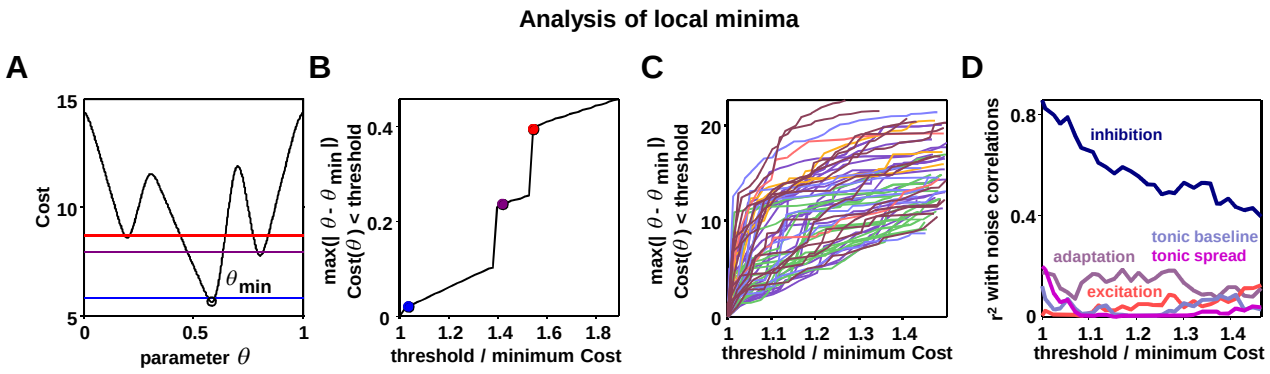

Figure S3: Related to Figures 3 and 5

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

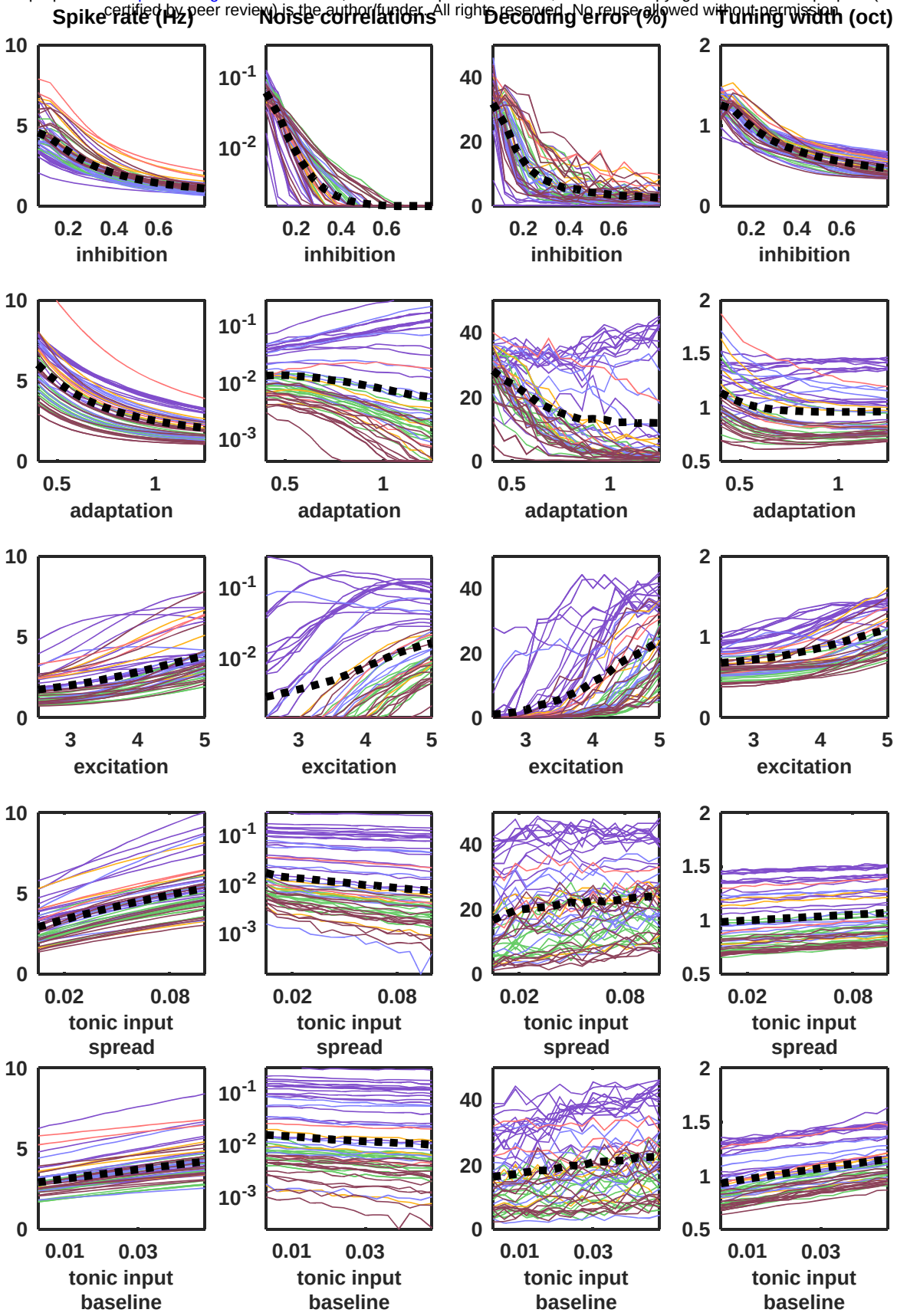

Figure S4: Related to Figures 4 and 5

bioRxiv preprint doi: https://doi.org/10.1101/041103; this version posted March 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

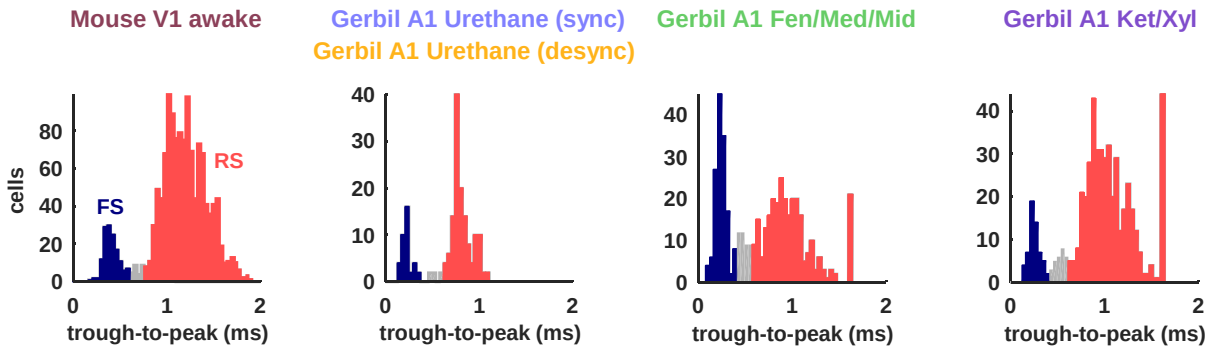

Figure S5: Related to Figures 6 and 7