Interpretation of Correlated Neural Variability from Models of Feed-Forward and Recurrent Circuits

Volker Pernice\textsuperscript{1,2} \\
Rava Azeredo da Silveira\textsuperscript{1,2,3,*}

\textsuperscript{1} Department of Physics, École Normale Supérieure, PSL Research University, 75005 Paris, France \\
\textsuperscript{2} Laboratoire de Physique Statistique, Centre National de la Recherche Scientifique, Université Pierre et Marie Curie, Université Denis Diderot, 75005 Paris, France \\
\textsuperscript{3} Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA \\
* corresponding author \\
February 7, 2017

Abstract

The correlated variability in the responses of a neural population to the repeated presentation of a sensory stimulus is a universally observed phenomenon. Such correlations have been studied in much detail, both with respect to their mechanistic origin and to their influence on stimulus discrimination and on the performance of population codes. In particular, recurrent neural network models have been used to understand the origin (or lack) of correlations in neural activity. Here, we apply a model of recurrently connected stochastic neurons to interpret correlations found in a population of neurons recorded from mouse auditory cortex. We study the consequences of recurrent connections on the stimulus dependence of correlations, and we compare them to those from alternative sources of correlated variability, like correlated gain fluctuations and common input in feed-forward architectures.

We find that a recurrent network model with random effective connections reproduces observed statistics, like the relation between noise and signal correlations in the data, in a natural way. In the model, we can analyze directly the relation between network parameters, correlations, and how well pairs of stimuli can be discriminated based on population activity. In this way, we can relate circuit parameters to information processing.

1 Author Summary

The response of neurons to a stimulus is variable across trials. A natural solution for reliable coding in the face of noise is the averaging across a neural population. The nature of this averaging depends on the structure of noise correlations in the neural population. In turn, the correlation structure depends on the way noise and correlations are generated in neural
circuits. It is in general difficult, however, to tease apart the origin of correlations from the observed population activity alone. In this article, we explore different theoretical scenarios of the way in which correlations can be generated, and we relate these to the architecture of feed-forward and recurrent neural circuits. Analyzing population recordings of the activity in mouse auditory cortex in response to sound stimuli, we find that population statistics are consistent with those generated in a recurrent network model. Using this model, we can then quantify the effects of network properties on average population responses, noise correlations, and the representation of sensory information.

2 Introduction

In the search for clues about the function of neural circuits, it has become customary to rely upon recordings of the responses of large populations of neurons. These measurements exhibit the concerted activity of neural populations in different conditions, such as presentations of different stimuli, summarized by stimulus-dependent, high-dimensional statistics. The lowest moment of these statistics are also the best characterized, namely the mean response of neurons, the variability of response of single neurons, and their pairwise correlations. With these statistics in hand, one can ask two questions: How are they generated in the neural population? What purpose, if any, do they serve? While the first question is mechanistic and the second is functional, the two are intimately linked. We focus here on these questions and on their connection.

Along the mechanistic line of research, a number of network models have been proposed to account for the statistics in measurements as well as to determine the relationship between anatomical and physiological parameters, on the one hand, and observed dynamics, on the other hand. In this spirit, for example models of balanced networks have been proposed to explain asynchronous and irregular spike trains [1, 2, 3]. Mechanistic explanations for the origin of pairwise correlation include the influence of recurrent connections [4, 5, 6] and global fluctuations (e.g., from top-down afferents) [7, 8, 9, 10, 11, 12]. Reading the origin of correlations from the recorded activity in a population of neurons is, however, a difficult task [13, 14, 7, 15].

Along the functional line of research, the objective is to relate the statistics of neural responses to the function of neural circuits; for example, to elucidate the role of the statistics in the representation of sensory information. The relation between correlated variability in the population response and the accuracy of stimulus representation, in particular, has been the object of much study in recent years [16, 17, 18, 19, 20, 21, 22, 23, 24, 25]. One reason for the focus on correlations is their possible effect in suppressing noise along relevant dimensions [26, 2, 27, 28]. Recent work has also illuminated the importance of the origin of correlations for their structure as it relates to the effect on stimulus representation [29, 30, 31].

In the present paper, we follow a three-pronged approach to relate the possible mechanistic origins of correlation to recorded statistics in cortical populations [32], on the one hand, and to relate network mechanisms to the representation of information, on the other hand. First, we study models of neural populations using the framework of Poisson processes [33] and we identify signatures in the population statistics of different mechanistic motifs. Second, we examine data on cortical populations in the light of our model results, and we find that the measured statistics are consistent with those generated by a recurrent network of Poisson neurons. Third, we use an idealized model of a recurrent cortical population, in which we can manipulate mechanistic parameters, to evaluate how the latter affect the accuracy of the representation of information. We find that, if correlations are not too weak, correlations generated within recurrent networks can be distinguished from those that are due to external
signals. The estimated parameters from the data suggest that the interpretation of neural spike trains as Poisson processes implies a strong amplification together with noise generation by the network dynamics.

3 Materials and Methods

3.1 Experimental data set

The data set was first published and analyzed in [32]. It consists of the activity of neural populations (46-99 neurons) recorded using calcium imaging in the auditory cortex of mice. Animals were isoflurane anesthetized (1 %). Signals were obtained from neurons labeled with the synthetic calcium indicator OGB1. Fluorescence was measured at 30 Hz sampling rate, and firing rates inferred from temporal deconvolution of the fluorescence signal. Up to 6 neural populations in each of 14 animals were recorded. The data points we use are the average firing rates over a window of 250 ms after presentation of each of 65 different sound stimuli, each for 15 trials. Responses were measured relative to spontaneous activity. Hence, negative responses occurred, if the stimulus-evoked firing rates were smaller than the spontaneous ones. The stimuli consisted of a range of different pure tones recorded for different sound intensities, as well as a number of natural sounds. We do not distinguish individual stimulus identities, but instead regard the set of stimuli as a diverse and, to some degree, generic ensemble.

3.2 Description of response variability

If the vector of population activity in trial $T$ of stimulus $s$ is $\tilde{r}(s, T)$, the average response across trials for this stimulus is denoted by $r(s) = \langle \tilde{r}(s, T) \rangle_T$. To measure covariability across trials in a given stimulus condition between neurons $i$ and $j$, we use the noise covariances, defined as

$$C_{ij}(s) = \text{cov}(\tilde{r}_i(s, T), \tilde{r}_j(s, T))_T = \langle \tilde{r}_i(s, T)\tilde{r}_j(s, T) \rangle_T - \langle \tilde{r}_i(s, T) \rangle_T \langle \tilde{r}_j(s, T) \rangle_T.$$

(1)

A measure of the strength of pairwise noise correlations across stimuli is the average correlation coefficient,

$$c_{ij}^N = \left\langle \frac{C_{ij}(s)}{\sqrt{C_{ii}(s)C_{jj}(s)}} \right\rangle_s.$$

(2)

Here, $\langle \cdot \rangle_s$ denotes the average over all stimuli presented. This quantity is to be contrasted with the signal correlation,

$$c_{ij}^S = \frac{\text{cov}(r_i(s), r_j(s))_s}{\sqrt{\text{var}(r_i(s))_s\text{var}(r_j(s))_s}},$$

(3)

which measures the similarity of average responses across stimuli. As a measure of how the orientation of the high-dimensional distributions changes across stimuli, we use the variance of the population activity projected along the direction of the normalized mean response, $\bar{r}(s) = r(s)/|r(s)|$, namely,

$$\sigma^2_{\mu}(s) = \sum_{ij} C_{ij}(s)\bar{r}_j(s)\bar{r}_i(s).$$

(4)
This quantity can be compared to the variance projected along the diagonal direction, \( \bar{d} = (1, \ldots, 1)^T / \sqrt{N} \),

\[
\sigma_d^2(s) = \sum_{ij} C_{ij}(s) \bar{d}_i \bar{d}_j, \tag{5}
\]

which corresponds to a uniform averaging of the covariances. To compare these quantities across stimuli, \( \sigma_\mu^2 \) and \( \sigma_d^2 \) are normalized by the sum of the variances, \( \sigma_{\text{all}}^2(s) = \sum_i C_{ii}(s) \).

If, for a given stimulus, all neurons are equally active on average, then \( \bar{r} = \bar{d} \) and \( \sigma_d^2 = \sigma_\mu^2 \).

As we show, different circuit models predict different stimulus dependencies of \( \sigma_\mu^2 \) and \( \sigma_d^2 \).

These differences are most apparent for the stimuli for which the average population response differs strongly from a uniform population response. A deviation from a uniform response can be measured by the angle between mean response and diagonal, or \( \cos(d, r) = \bar{r}^T \cdot \bar{d} \).

For a graphical illustration of these measures, see Fig. 1.

Figure 1: Properties of response distributions and network scenarios. A: Examples for distributions where the variability is to a large degree in the diagonal direction (response distribution indicated by blue ellipse), resulting in a large value of \( \sigma_d \), and where the variability is mainly in the direction of the average response (red ellipse), resulting in a larger value of \( \sigma_\mu \). B: Response distributions for two stimulus ensembles. Either \( \sigma_\mu \) (red set) or \( \sigma_d \) (blue set) remains large across stimuli. A dependence as in the red set appears in a model of shared gain fluctuations, while a dependence as in the blue set may arise in densely connected recurrent or feed-forward networks. C: Different network layouts that induce correlated activity. Connections (arrows) to and between neurons (dots) vary in strength. Dashed arrows indicate multiplicative effects on firing rates.
3.3 Recurrent model of Poisson neurons with noisy input

We model an intrinsically noisy process of spike generation in neurons, in which the effect of presynaptic spikes on neural activity is captured through linear modulations of an underlying firing rate. Specifically, the spike train of neuron $i$ is a realization of an inhomogeneous Poisson process with a time-dependent firing rate, $\tilde{r}_i(t)$, calculated as

$$\tilde{r}_i(t) = \tilde{r}_{\text{ext},i}(t) + \sum_j \int_0^\infty \tilde{g}_{ij}(\tau)\tilde{s}_j(t-\tau)d\tau,$$

where $\tilde{r}_{\text{ext},i}(t)$ is the external input, $\tilde{s}_j(t) = \sum_{t'} \delta(t-t')$ is the recurrent input from presynaptic spike trains, and the causal coupling kernel, $\tilde{g}_{ij}(\tau)$, defines the interactions between neurons in the network. The external input is an analog quantity which can be viewed as resulting from a convolution between a linear filter and an incoming spike train from “non-local” presynaptic neurons. We treat it as a noisy, but stationary, signal, i.e., its mean and higher moments do not depend on time. While this model of interacting point processes [33] defines the full temporal dynamics of the system, we are only interested in time-averaged rates or, equivalently, spike counts. Their expected values and (co-)variances across realizations can be obtained if the coupling matrix is known.

Assuming stationarity of input spiking and firing rates, we can solve for the average rate vector, $r = (r_1, \ldots, r_N)$, across trial realizations, as

$$r = (I - G)^{-1}r_{\text{ext}}.$$  \hfill (7)

Here, $I$ denotes the identity matrix, $G$ is the steady-state coupling matrix with elements

$$G_{ij} = \int_0^\infty \tilde{g}_{ij}(\tau)d\tau,$$

and $r_{\text{ext}}$ is the average input vector. To describe the correlation in the activity, we consider the spike count in a long time bin, $\Delta$, defined as

$$n_i(\Delta) = \int_t^{t+\Delta} \tilde{s}_i(t)dt.$$  \hfill (9)

The total input in the time bin is given by

$$r_{\text{ext}}(\Delta) = \int_t^{t+\Delta} \tilde{r}_E(t)dt,$$

and is a random variable with a normalized variance, $V_{\text{ext}} \equiv \text{var}(r_{\text{ext}}(\Delta))/\Delta$ (across choices of different time bins). From spectral properties of the covariance [33] and the relation between covariance and count correlations [34, 5], it can be shown that the count covariance normalized by the time bin duration, $C$, with elements

$$C_{ij} \equiv \lim_{\Delta \to \infty} \frac{\text{cov}(n_i(\Delta), n_j(\Delta))}{\Delta},$$

is given by the matrix

$$C = (I - G)^{-1}(D[r] + D[V_{\text{ext}}])(I - G^T)^{-1}.$$  \hfill (12)

The term $D[r]$ denotes a diagonal matrix with elements $D[r]_{ij} = r_i\delta_{ij}$, and similarly $D[V_{\text{ext}}]_{ij} = \delta_{ij}V_{\text{ext},i}$. The effect of the recurrent network interactions is reflected in the transfer matrix,

$$B \equiv (I - G)^{-1}.$$  \hfill (13)
In Eq. (12), one component of the covariance arises from the variance of the external input, $V_{\text{ext}}$. Although the intrinsic noise from the spike-generating Poisson process is dynamical, its effect is that of an additive contribution to the external variance (see also [35]). The dependence on the rate vector, in Eq. (12), results from the Poisson character of the noise, in which the variance equates the mean.

In the experimental data set, firing rates are measured relative to spontaneous activity. To take this into account, an offset, $a$, can be added to the rates in Eq. (12). Henceforth, we examine the model through Eqs. (7) and (12) for the mean neural activities and their covariances, respectively.

### 3.3.1 Random effective network and stimulus ensemble

The rates and covariances given in Eqs. (7) and (12) depend on the recurrent coupling matrix within the network, $G$, only via effective couplings defined by the transfer matrix, $B = (\mathbb{I} - G)^{-1}$. In numerical calculations, we choose the entries of $G$ independently from a normal distribution with positive mean; this affords a specific structure to the matrix $B$. The empirical mean and variance of the elements of this matrix will be called $\langle B \rangle$ and var$(B)$. In a number of analytical calculations, for the sake of simplicity, we neglect the specific structure of $B$ and model it as a random matrix with elements independently drawn from a normal distribution with corresponding mean and variance. The elements of $B$ are not direct, but effective connections, because they reflect also the effects of indirect connections in the network, hence the expression “random effective network.” The variability of effective connections can be quantified by the ratio $\rho = \text{var}(B)/\langle B \rangle^2$.

The stimulus ensemble is modeled as a set of random vectors, $r_{\text{ext}}(s)$. We assume that the elements of these vectors are independent across neurons and stimuli, and normal, and we call the mean and variance of the corresponding normal distribution $\langle r_{\text{ext}} \rangle$ and var$(r_{\text{ext}})$. The relative variability of the inputs is defined as $\rho_E = \text{var}(r_{\text{ext}})/\langle r_{\text{ext}} \rangle$. This variability is high if both excitatory and inhibitory inputs are present, and, together with the network parameters, it determines the variability of the average responses across stimuli. In numerical calculations, we set $V_{\text{ext},i}(s) = \langle |r_{\text{ext},i}(s_j)| \rangle$, as dictated by a Poisson process. We allow for input signal correlations, i.e., the average inputs to a pair of neurons can come with a non-vanishing correlation coefficient, $c_{\text{in}} = \langle r_{\text{ext},i}(s_j), r_{\text{ext},j}(s) \rangle$.

### 3.4 Models of correlated activity from shared inputs or gain modulations

Correlations in the activity of neurons can have other origins besides recurrent connections. In parallel with the recurrent network model, we consider two alternative prototypical models in which correlations originate from shared inputs or from shared gain fluctuations, respectively. The three different scenarios are illustrated in Fig. 1C. In Appendix 8.2, we show how, formally, these models can also be cast as special cases of the recurrent model.

#### 3.4.1 Feed-forward model with shared input

We consider a simple, two-layer, feed-forward network, in which $N$ output neurons receive inputs from $N$ independent presynaptic neurons with spike trains $\tilde{s}_j(t)$. The contributions of the input neurons to the firing rates of the output neurons are determined by feed-forward connection kernels $f_{ij}(\tau)$, so that the observed firing rates are
\[
\tilde{r}_i(t) = \sum_j \int_0^\infty \tilde{f}_{ij}(\tau) \tilde{s}_j(t-\tau) d\tau.
\]

(14)

As we show in Appendix 8.2, this scenario represents a special case of the framework described in Section 3.3, so that firing rates and count covariances can be calculated correspondingly. We define the feed-forward coupling matrix \(F\) by its elements, \(F_{ij} = \int_0^\infty \tilde{f}_{ij}(\tau) d\tau\). With the time-averaged firing rates of the external input for stimulus \(s\), \(r_{\text{ext}}(s)\), the average spike counts of the output neurons are given by

\[
r(s) = Fr_{\text{ext}}(s).
\]

(15)

If the spike trains of the input neurons are Poisson processes, the input variance is equal to the rate, \(V_{\text{ext}} = r_{\text{ext}}\), but we will allow for more general inputs. The count covariances are then given by

\[
C = FD[V_{\text{ext}}]F^T + D[r].
\]

(16)

We denote again by \(D[r], D[V_{\text{ext}}]\) diagonal matrices with diagonal elements given by the vectors \(r, V_{\text{ext}}\). The first term describes covariances resulting from the combined shared inputs. The second term results from the contribution to the count variances of the Poisson spike generation in output neurons, which is independent across neurons.

If rates are observed only up to a stimulus-dependent offset, \(a\), Eq. (16) has to be replaced by

\[
C = FD[V_{\text{ext}}]F^T + D[r + a].
\]

(17)

This scenario can be compared directly to the random effective recurrent network model, if one uses \(F = B\) for an identical ensemble of stimuli (see Section 3.3.1). Comparing Eq. (17) to Eq. (12), we note that the difference between the two models is that the internally generated noise in the feed-forward model, \(D[r + a]\), is uncorrelated and contributes only to the variances, while in the recurrent network it is filtered by the network and therefore correlated.

### 3.4.2 Model with shared gain fluctuations

In addition to recurrence and shared inputs, shared gain fluctuations have also been proposed as a source of correlation in neural populations by various authors [71, 36, 24, 10]. In this model, the firing rate of the neural population is given by the product of a constant, stimulus dependent vector, \(r(s)\), and a scalar fluctuating signal, \(\epsilon(t)\). We set the time-averaged value of the fluctuations to unity, \(\langle \epsilon(t) \rangle_t = 1\), so that the time-averaged firing rate is \(r\). The resulting count covariance matrix is given by

\[
C = D[r] + rr^TV_{\text{ext}},
\]

(18)

where the scalar variance, \(V_{\text{ext}}\), reflects the strength of the variation of the fluctuating signal. In other words, the firing rate of a Poisson neuron is modulated multiplicatively by a fluctuating “gain” signal. One consequence is that the covariance of a neuron pair is proportional to the product of their firing rates. The general model described in Section 3.3 reduces to this model if for any given stimulus the activity of the population results from a single input neuron projecting to the output neurons with corresponding weights (see Appendix 8.2).

Here, the average responses can be chosen freely. When comparing this scenario to the recurrent and feed-forward scenarios, we use the ensemble of average responses resulting from the corresponding network scenario. If the stimulus dependence of the firing rates is measured relative to an offset \(a\), the covariances are given by

\[
C = D[r + a] + (r + a)(r + a)^TV_{\text{ext}}.
\]

(19)
3.5 Stimulus discriminability

In order to evaluate the influence of correlation on neural coding, we examine the discriminability of a set of stimuli from the population activity. In the models, a stimulus, \( s \), evokes activity characterized by an average response vector, \( r(s) \), and a covariance matrix, \( C(s) \). We seek a simple measure for evaluating the possibility of attributing a given population response to one of two discrete stimuli, \( s_1 \) and \( s_2 \), unambiguously. For this, we assume that the high-dimensional distributions of responses are Gaussian, and project the two distributions on a single dimension where we calculate the signal-to-noise ratio as our measure.

The best projection follows from Fisher’s linear discriminant analysis: the most informative linear combination of neuron responses, denoted by \( w \cdot r \), is achieved if the vector \( w \) points in the most discriminant direction, 

\[
 w = \left( C(s_1) + C(s_2) \right)^{-1} \left( r(s_1) - r(s_2) \right). 
\]

(20)

The mean and variance of the projected distributions onto the normalized direction, \( \bar{w} \), are 

\[
 \bar{w}^T r(s_i) \quad \text{and} \quad \sigma^2_{s_i} = \bar{w}^T C(s_i) \bar{w}, \quad \text{for} \quad i \in \{1, 2\}. 
\]

We can then define the “signal-to-noise ratio,” as

\[
 S = \frac{|\bar{w}^T \cdot (r(s_1) - r(s_2))|}{\sigma_{s_1} + \sigma_{s_2}}. 
\]

(21)

Larger values of \( S \) correspond to better discriminability.

To quantify the effect of correlation on discriminability, we compare the quantity \( S \), defined in Eq. (21), with the quantity \( S_{\text{shuffled}} \) obtained from a “shuffled data set”, in which responses are shuffled across trials (in experimental data) or off-diagonal covariance elements, \( C_{i \neq j}(s) \), are set to 0 (in models). A ratio \( S_{\text{shuffled}} / S \) smaller than unity indicates that correlation is beneficial to discriminability.

We note that \( S \) is an approximation for a measure based on the optimal linear classifier, for which the threshold separating the two one-dimensional projected distributions has to be calculated numerically. A measure akin to \( S \) used in similar contexts is the linear Fisher information [16], valid for continuous stimuli. An advantageous property of \( S \) is its invariance under linear transformations: if all responses, \( r \), are fed into another network whose output, \( Br \), results from a product with an invertible matrix \( B \), \( S \) does not change. This obtains because the mean responses are transformed into \( Br(s) \), while the covariances are transformed into \( BC(s)B^T \). Intuitively, a simple matrix multiplication is accommodated for by a corresponding change in the most discriminative direction \( w \) on which we project.

3.6 Two-population model

In order to develop an intuitive understanding of the way in which correlations generated in a recurrent network influence coding, we examine a highly simplified model. The recurrent network consists of two excitatory sub-populations, labeled \( E \) and \( E' \), each made up of \( N \) neurons. We ask to what extent their activity discriminates two stimuli when these elicit preferential responses in the two sub-populations, respectively.

The activity in the network is determined by Eqs. (7) and (12). For the sake of simplicity, we can set the external variance to zero, such that the input to the network is fully defined by its mean input, \( r_{\text{ext}} \). We assume some further simplifications, for the sake of calculational ease: each neuron in sub-population \( L \) projects to a fixed number, \( n_{KL} \), of neurons in sub-population \( K \); all non-zero coupling weights are identical, \( G_{ij} = g_E \). Additionally, all neurons in the same population receive identical external input. Then, the two-component vectors \( R_{\text{ext}}(s_1) = (1 + \Delta, 1)^T \) and \( R_{\text{ext}}(s_2) = (1, 1 + \Delta)^T \) define a pair of stimuli. Each component denotes the input to each neuron in the corresponding population.
We reduce the dimensionality of the system by considering the population activity on a macroscopic level. The average response of population $K$ across trials is defined as

$$R_K = \sum_{k \in K} r_k,$$

and its trial-to-trial variability is described by the population covariance matrix, with elements defined as

$$\Sigma_{KL} = \sum_{k \in K, l \in L} C_{kl}.$$

It turns out that these macroscopic quantities depend only on the overall number of connections between populations, not on a specific network realization. For example, the sum of the rates of neurons in one population depends only on the sum of the inputs to all neurons, but not on how these inputs are distributed among the neurons. This is a consequence of the linearity of the dynamics and the assumption that each neuron has a fixed number of output connections, as we show in detail in Appendix 8.4. This idea was already applied in Refs. [5, 6].

More precisely, we define the coupling within each population, $\Gamma_s = n_{EE}g_E = n_{E'E}g_E$, and across populations, $\Gamma_c = n_{EE}g_E = n_{E'E}g_E$, which make up a population coupling matrix

$$\Gamma = \begin{pmatrix} \Gamma_s & \Gamma_c \\ \Gamma_c & \Gamma_s \end{pmatrix}.$$  

Applying the definition in Eqs. (22) and (23), we can rewrite the microscopic Eqs. (7) and (12) as population equations,

$$R(s) = (I - \Gamma)^{-1} NR_{\text{ext}}(s)$$  

and

$$\Sigma(s) = (I - \Gamma)^{-1} D[R(s)](I - \Gamma)^{-1}.$$

The population transfer matrix is defined by $P = (I - \Gamma)^{-1}$. Up to a factor $N$, these equations are equivalent to the microscopic equations in the restricted context of the two-population model.

4 Results

We discuss the characteristics of the noise statistics that emerge from neural dynamics in models of recurrent networks, feed-forward networks, and networks with global gain fluctuations. We also look for coarse statistical signatures of each of these three model structures, which can then be compared with data. In this perspective, we analyze cortical recordings in mouse. These recordings were made in the auditory cortex, while the animal was stimulated with musical tones [32]. We then discuss the implications of the neural noise statistics for stimulus encoding. Finally, we present a highly simplified model of network dynamics, paired down to include only a few parameters, in order to develop an intuition of the link between network structure and stimulus encoding.

4.1 Population signatures of noise statistics in a recurrent network model

One population feature which distinguishes the possible mechanisms of generation of noise correlation is the relation between the population-averaged response, $\langle r \rangle(s) = \langle r_i(s) \rangle_i = \langle r_{ss}(s) \rangle_s$,
\[ \frac{1}{N} \sum_i r_i(s) \text{, on the one hand, and the population-averaged variance, } \langle C_{ii} \rangle_i(s) = \frac{1}{N} \sum_i C_{ii}(s), \text{ or the noise covariance averaged across pairs } \langle C_{ij} \rangle_{i \neq j} = \frac{1}{N(N-1)} \sum_{i \neq j} C_{ij}, \text{ on the other hand. In this section, we analyze these relations for Poisson neurons in a recurrent network with random effective connections, as described in Section 3.3.1. Using Eq. (12), which reads } C_{ij} = \sum_k B_{ik} B_{jk} (r_k(s) + a + V_{ext,k}) \text{ for the pairwise covariances, we derive the expression } \]

\[ \langle C_{ii} \rangle_i(s) \approx N \langle B^2 \rangle \left( \langle r \rangle(s) + a + \langle V_{ext} \rangle \right) \tag{27} \]

for the average variance, and the expression

\[ \langle C_{ij} \rangle_{i \neq j}(s) \approx N \langle B \rangle^2 \left( \langle r \rangle(s) + a + \langle V_{ext} \rangle \right) \tag{28} \]

for the average covariance. (See Appendix 3.3.4 for mathematical details and Fig. 2 for numerical results.) The quantity \( \langle B \rangle = \frac{1}{N^2} \sum_{ij} B_{ij} \) denotes the average strength of the effective connections in the network across neuron pairs and, correspondingly, \( \langle B^2 \rangle = \frac{1}{N^2} \sum_{ij} B^2_{ij} \) the average of their square. The variance of the input, averaged across neurons, is \( \langle V_{ext} \rangle \), and \( a \) denotes a potential constant offset in the observed firing rates.

The variance of a Poisson variable is proportional to its mean. This contribution from the Poisson spike generation is reflected by the term \( \langle r \rangle \) in the two equations and is the reason for the linear relation between average response and average variance/covariance - for non-Poisson spike generation, these relations would be non-linear. As we see, not only the external noise, \( \langle V_{ext} \rangle \), but also the internally generated noise are amplified by the recurrent connections. The amplification factor, or the slope of the linear relation, \( N \langle B \rangle^2 \) in Eq. (28), is smaller for the covariances, where it depends only on the mean strength \( \langle B \rangle \) of the effective connections. For the variances, expressed in Eq. (27), the averaging in the slope \( N \langle B^2 \rangle \) is carried out on the square of the effective connections. The slope is therefore larger, and depends both on the average strength of connections and their variability. In the relations for both average population variance and covariance, the ratio between the intercept and the slope is \( \langle V_{ext} \rangle + a \). Up to a possible offset, \( a \), that can result from a measurement of firing rates relative to a constant baseline activity, it represents the strength of the external noise.

In summary, in the recurrent network model, both internally generated noise and external noise are amplified (or attenuated) by the recurrent connections, and the parameters in the relationship between (co-)variances and population response can be related to network and input parameters.
Figure 2: **Variability in the recurrent network model.** A: Relation between response and variance averaged across population for randomly chosen stimulus vectors (each dot one stimulus). Four networks are generated with normally distributed connection strengths. Larger $\rho$, calculated from effective connections in the transfer matrix $B$, indicate larger variability of effective weights. Lines indicate analytic predictions, Eqs. (27) and (28). B: Same for covariances, averaged across all neuron pairs. C: The ratio of slope and intercept of linear fits corresponds to the strength of the input variance, which is identical for all networks, and is consistent for variances and covariances. See Appendix 8.1 for further details and the numerical parameters.

4.2 Population signatures of noise statistics in feed-forward network models

In a feed-forward network, both shared input units and global gain fluctuations can yield noise correlations [7, 10, 24]. Here, we compare the population signatures of the noise statistics in the case of these two alternatives with those from a recurrent model. This serves to point out differences between the outputs of the various models as well as to motivate the data analysis to which we turn below.

The three scenarios differ in the scaling of the population-averaged covariance, $\langle C_{ij} \rangle_{i \neq j}(s)$, with respect to the population response, $\langle r \rangle(s)$ (Fig. 3). We already demonstrated in the previous section that both variances and covariances scale linearly with the population response in the recurrent network. In a feed-forward network, pairwise covariances are given by Eq. (16), which can be rewritten as

$$C_{ij} = \delta_{ij}(r_i + a) + \sum_k F_{ik}F_{jk}V_{\text{ext},k}.$$ (29)

Note that, in contrast to the recurrent network, the neural firing rates, $r_i$, only affect diagonal entries, and therefore the variances. The average covariance can be expressed as

$$\langle C_{ij} \rangle_{i \neq j}(s) \approx N\langle F \rangle^2\langle V_{\text{ext}} \rangle(s),$$ (30)

and does not directly depend on the average population response, $\langle r \rangle(s)$ (see Appendix 8.3.3). If there are excitatory as well as inhibitory inputs, also the average input variance, $\langle V_{\text{ext}} \rangle$, can be uncorrelated to the population response: for Poisson external input, in each input channel $i$ the variance is as large as the mean input, $V_{\text{ext},i} = |r_{\text{ext},i}|$. Nonetheless, because the variance is always positive, including for negative input, the average variance $\langle V_{\text{ext}} \rangle$ can be
decorrelated from the average input, $\langle r_{\text{ext}} \rangle$, and thus from the average output, $\langle r \rangle$. Intuitively, because inhibitory channels add variance, but decrease the average input, the combined input does not have to be proportional to the combined variance. The extent to which this is true depends on the balance between positive and negative elements in $r_{\text{ext}}$, which we measure by the input variability $p_E = \text{var}(r_{\text{ext}})/\langle r_{\text{ext}} \rangle$.

In a model network with global gain fluctuations, from Eq. (18), the covariances are given by

$$C_{ij} = \delta_{ij} r_i + r_i r_j V_{\text{ext}}.$$  

(31)

Again averaging only across the off-diagonal elements, one finds that the average covariances scale quadratically with the population response, as

$$\langle C_{ij} \rangle_{i \neq j} = V_{\text{ext}} \langle r \rangle^2.$$  

(32)

This is a direct consequence of the fact that pairwise covariances are proportional to the product of the firing rates of the pair of neurons, which holds for large $N$ independent of the distribution of average responses. By contrast, as is apparent from the expression for $C_{ij}$, the individual variances and the population averaged variance each have both a quadratic and a linear contribution in $r$ (and $\langle r \rangle$, respectively) (see also e.g. [7]). (Further details are provided in Appendix 8.3.)

Until now, we examined how the average covariance $\langle C_{ij} \rangle_{i \neq j}$ changes across stimuli. Further differences between the different scenarios can be related to the detailed structure of $C$, which determines the shape of the response distributions. By shape, we refer to the geometric orientation and extent of the multi-dimensional ellipsoid cloud that corresponds to the response distribution in the space spanned by the responses of the individual neurons. A two-dimensional sketch of such an ellipse, and the geometric interpretation of the quantities we use is depicted in Fig. 1A (see also Methods 3.2). In words, the way the orientation of the response distribution changes with the average response varies between the three scenarios: the long axis of the ellipsoid indicating the direction of the largest variance is aligned along the diagonal for the feed-forward and the recurrent network models, but in the direction of the mean response for the gain fluctuation model (Fig. 1B). To capture this relation quantitatively, we consider the projected variances, $\sigma^2_\mu$ and $\sigma^2_d$ (see Fig. 3 and methods for details). There, the “diagonal direction” corresponds to a response where all neurons are equally active.

Quantitatively, we find that in the recurrent and feed-forward network models with relatively strong correlations, there is always a large fraction of variance in the diagonal direction, independent of the stimulus. The variance projected along the direction of the average response is only large if the latter happens to be similar to the diagonal direction. By contrast, in the gain fluctuation model, the variance projected along the direction of the average response is approximately constant across stimuli. This behavior of the gain fluctuation model is due to the fact that the variability across trials is multiplicative on the mean response, and hence variability is large in this direction for all stimuli.

In the network models, strong correlations result from effective interactions between neurons or from shared input. Each neuron in the recurrent network, or each input channel in the feed-forward network, is effectively connected to a large part of the remaining neurons and contributes to their correlations. Because our networks are not balanced, these contributions are mostly positive, even for inhibitory input channels. Thus, across trials, responses of all neurons are strongly correlated and the trial-to-trial fluctuations tend to be of similar size for each neuron for a given stimulus. By contrast, the variability for any given neuron varies more strongly across stimuli.
Figure 3: **Response-covariance relations depend on origin of correlations.** A-C: Average population response $\langle r \rangle$ versus average (co-)variances in recurrent network (top), feed-forward network (middle) and gain-fluctuation model (bottom). Each dot corresponds to a random stimulus, blue dashed lines indicate analytic results, Eqs. (28)-(32). D-F: Dependence of normalized variability projected on mean direction on response direction for the three scenarios. In D,E colors indicate results for different networks, in F size of markers indicates strength of gain fluctuations. Square markers on the left indicate numerical value of $\sqrt{c_N}$. G-I: Same for the normalized variability projected on diagonal direction. See Appendix for further details and the numerical parameters.

### 4.3 Relation between signal correlation and noise correlation in recurrent network models

In the simple scenario we analyze here, average responses and covariances are related because they both depend on the network architecture. By expressing signal correlations, $c_{ij}^S$, and the noise correlations, $c_{ij}^N$, in terms of network and input parameters, we can derive a relation between these two sets of quantities (Fig. 4). For a random transfer matrix, $B$, all neuron pairs are statistically equivalent, and the strength of correlations can be characterized by the average of the pairwise signal and noise correlations. In our recurrent network model, the network architecture affects noise and signal correlations through an effective quantity,
namely the signal-to-noise ratio of the elements of the transfer matrix, \( \rho = \text{var}(B)/\langle B \rangle^2 \):

\[
c_N \equiv \langle c_{ij}^N \rangle_{i \neq j} \approx \frac{1}{1+\rho},
\]

\[
c_S \equiv \langle c_{ij}^S \rangle_{i \neq j} \approx \frac{1 + (N-1)c_{\text{in}}}{1 + \rho + (N-1)c_{\text{in}}},
\]

(see Appendix 8.3.4 Eqs. (70) and (77)). Here, \( c_{\text{in}} \) denotes the strength of signal correlations in the input across stimuli. Both signal and noise correlations are larger in networks with more homogeneous entries (smaller \( \rho \)). We already showed in Section 4.1 that noise covariances, \( \langle C_{ij} \rangle_{i \neq j} \), depend on the average strength of the effective connections, while the variances, \( \langle C_{ii} \rangle_i \), depend on \( \text{var}(B) \) as well. If the average input for pairs of neurons across stimuli is uncorrelated (no input signal correlations, \( c_{\text{in}} = 0 \)), noise and signal correlations are identical on average. However, due to a prefactor which grows with system size (\( N - 1 \) in Eq. (34)), even weak input signal correlations are strongly amplified and can yield a large effect. Because noise in the population response to a given stimulus is produced internally in the process of spike generation, noise covariances are unaffected by this mechanism, so that strong signal correlations can coexist with weaker noise correlations.

These relations depend on network parameters and hold for the average correlations (across pairs in a network). The pairwise noise and signal correlation coefficients, \( c_{ij}^N \) and \( c_{ij}^2 \), can vary widely, but numerical calculations indicate that they are correlated in a given network (across pairs) as well.

Figure 4: Relation between signal and noise correlations in the recurrent network model. A: Dependence of average signal correlations, \( c_S \), (continuous lines) and average noise correlations, \( c_N \), (dashed lines) on input correlation and and network properties, Eqs. (33), (34). Low variability of the transfer matrix elements, \( \rho \), increases \( c_N \) and \( c_S \). Input signal correlations affect only \( c_S \). B: Same data, average signal versus noise correlation. C: Scatter plot of all pairwise signal versus pairwise noise correlations (dots) in five network realizations for \( c_{\text{in}} = 0.05 \). Circles indicate network average across pairs, orange dotted line corresponds to analytic expressions shown in B for \( c_N \) and \( c_S \).

4.4 Stimulus discrimination in a recurrent network model

Since the network architecture affects both signal and noise correlation, it is natural to ask how the discriminability of stimuli is affected in turn, as it depends on both quantities. Indeed, noise correlations can affect the coding properties of a population appreciably [19]. Noise
correlations are referred to as favorable for the discrimination of a pair of stimuli when the shapes of the two correlated response distributions (corresponding to the two stimuli) is such that variability occurs predominantly in a direction orthogonal to the informative direction (Fig. 5). The relevance of correlations can be quantified by comparing the discriminability of stimulus pairs in the case of the full population output and in the case in which noise correlations are removed by shuffling trials.

In Fig. 5, we visualize, for a recurrent network model, the interplay of the two effects which affect the influence of correlations: on the one hand, how likely it is that the influence of correlations be favorable for a pair of stimuli, and, on the other hand, how strong the effect of correlations is for a given pair. For the first question, we utilize the fact that the noise is predominantly along the diagonal direction. Consequently, the effect of correlations is favorable, if average responses differ in a direction orthogonal to the diagonal. The distribution of the angle between these two directions will therefore reflect the likelihood of beneficial correlations. To confirm this picture, and to examine the second question, we plot how the effect of shuffling on the signal-to-noise ratio, $S$, depends on this angle (see Eq. 21 for the definition of $S$). The network parameters, in this case the variability $\rho$ of the elements of the effective connections, affect both the distribution of angles and the effects on $S$ - the distribution of angles via the distribution of average responses, or the signal correlations, and $S$ by the strength of the noise correlations. However, it turns out that the effect of shuffling, averaged across all pairs, is the same in each network. This is due to the fact that signal correlations are on average as strong as noise correlations in all of these networks. In Section 4.8 we analyze a simplified network model to better understand some of these effects.
Figure 5: Covariances affect stimulus discriminability

A: Sketch of response distributions for two neurons and two stimuli (blue ellipses). Neural responses are correlated for both stimuli, so that main axis of variability is parallel to linear separator (orange line). B: Larger ensemble of stimuli for correlated neurons (large variance along diagonal directions). The effect of correlations is favorable for stimulus pairs where difference in means is orthogonal to diagonal (red arrow), or unfavorable when difference in means tends to be parallel to diagonal (green arrow). C: Distribution of cosines of angle between diagonal and difference in means across stimulus pairs in recurrent networks with different parameter $\rho$. In networks with low $\rho$, unfavorable angles are more frequent. D: Ratio between discriminability for correlated and shuffled distributions, dots correspond to stimulus pairs. Lower values indicate that correlations are more beneficial. Dashed horizontal lines indicate average across all stimulus pairs.

4.5 Variability and correlations in populations of mouse cortical auditory neurons

We use the theoretical results presented in the previous sections to analyze the responses of a population of neurons to different stimuli. The data set contains the firing rates, collected during a certain time interval, in response to the presentation of different sound stimuli in a number of trials (see Section 3.1 and [32]). We will compare the properties of covariances (across trials) and average responses with the predictions of our different models. Based on the models, we can then evaluate the effect of network generated variability on stimulus discrimination.

The trial-to-trial variability in single-neuron output is large (supra-Poisson, Fig. 6). Neurons with larger average response exhibit also a larger variability in their responses. This tendency is observable across different neurons responding to a given stimulus, for individual neurons across stimuli and for the average population response across stimuli. The pairwise correlations in trial-to-trial variability are generally high and pairs with high noise correlations tend to have strong signal correlations. In other words, because signal correlations measure the similarity of average responses across different stimuli, neurons with similar tuning properties also have similar trial-to-trial variability. Additionally, in populations with strong average signal correlations, noise correlations are strong as well. These relations be-
tween noise and signal correlations can be reproduced in the recurrent network model (Fig. 4). However, similar results may be possible also in alternative scenarios for the generation of correlated variability. We compare the different models in the next section.

Figure 6: **Variability and correlation in mouse auditory cortex.** A: Average response versus response variance in a single population. Colors correspond to 10 randomly chosen stimuli. Dots correspond to single neurons, circles to averaged values across population. Dashed line indicates identity. Inset: All neurons for all stimuli. B: Scatter plot of signal correlation coefficient (across stimuli) vs. noise correlation coefficient (averaged across stimuli) for different neural populations measured in the same animal, marginal histograms at top and bottom. Signal and noise correlations are correlated across pairs. Correlations are high in general, but the amount of signal and noise correlations varies strongly across populations. Circles denote average across pairs in each populations. Inset: Black circles: Average signal and noise correlations in all measured populations. Grey dots: individual pairs, 5% randomly chosen from all experiments.

### 4.6 Analysis of population response variability in mouse auditory cortex

Motivated by our investigation of network models, we first examine the relation between the shape of the response distributions for each stimulus and the pattern of average responses for the entire set of stimuli. In Fig. 7, we display the dependence of the normalized standard deviation projected on the direction of the mean response, $\sigma_\mu/\sigma_{all}$, and the normalized standard deviation projected on the diagonal direction, $\sigma_d/\sigma_{all}$, on the angle between average response and diagonal, through $\cos(d,r)$ (see Section 3.2 for details and definitions). We observe a strong dependence on $\cos(d,r)$ of the variance projected on the direction of the mean response, but not of the variance projected on the diagonal direction. The dependence of $\sigma_\mu/\sigma_{all}$ on the stimulus is much stronger in nearly all the measured populations. This behavior is consistent with a network model, either feed-forward or recurrent, but not with a model with shared gain fluctuations, where a large part of the variance is consistently in the direction of the mean response (see Section 4.2). Note, however, that in the data firing rates are measured relative to the spontaneous activity; this is reflected in the presence of negative values in the stimulus evoked activity. Such an offset could affect our comparison because we do not know the true value of the mean response. To account for this possibility, we searched for the best possible offset, assuming the model of gain fluctuations for the
stimulus-dependent covariances. We then corrected the firing rates by this offset, and evaluated the stimulus dependence of $\sigma_\mu$ and $\sigma_d$, as before (see Appendix 8.3.2 for details), but found no qualitative change in the results. However, we did not test for mixed models, and it is possible that part of the correlated variability can be explained by shared fluctuations.

Based on the results presented above we conclude that, of our three scenarios, feed-forward and recurrent network models are more consistent with the data. In the following, we fit the parameters of these models to the data to find which of these two provides a better fit. For this comparison, we analyze the dependence of the average variances, $\langle C_{ij} \rangle_{i\neq j}$, and covariances, $\langle C_{ii} \rangle_i$, on the population-averaged response, $\langle r \rangle(s)$. In the recurrent network (Eqs. (27) and (28)), both variances and covariances increase linearly with $\langle r \rangle$ across stimuli, while in the feed-forward network (Eq. (32)), mean covariances are not expected to depend strongly on $\langle r \rangle$.

To examine whether the experimental data is consistent with the feed-forward model, we fitted parameters to the activity of each experimentally observed population (see Appendix 8.3.5), and generated a surrogate data set of responses and covariance matrices, with a matching number of neurons and stimuli. In particular, we obtained values for the variability of the input ensemble, $\rho_E = \text{var}(r_{\text{ext}})/\langle r_{\text{ext}} \rangle^2$, and of the network elements, $\rho = \text{var}(F)/\langle F \rangle^2$. It turns out that $\rho_E, \rho > 1$: for both input and network elements the variance is much larger than the mean (see Fig. 8A-B). The reason is the high variability of average responses across stimuli. With these parameters, the statistics of the distribution of average responses are well reproduced (Fig. 8C). As we discussed in Section 4.2, for a high value of $\rho_E$ the feed-forward model does not predict a strong relation between $\langle C_{ij} \rangle_{i\neq j}$ and $\langle r \rangle(s)$. We quantify this relation by the ratio of slope and intercept from a linear fit to the data, which indicates how strongly the average covariances increase for stimuli that evoke strong population responses. In the feed-forward scenario, this behavior does not obtain: the slopes of the linear fits in the model relative to the intercepts are too low in comparison to what is observed in the experiment (Fig. 8D, E). The feed-forward model thus cannot reproduce both the large variability of neural responses across stimuli and the increase of the average covariance with the average response.

In a recurrent model, this increase is expected, because noise generated by the neurons is propagated through the network: the covariances are proportional to the average response (see Eq. (28)). Moreover, it predicts that the ratios of intercept and slope in the linear behaviors of variances and covariances are the same (it indicates the variance of the external input). Indeed, linear fits to these relationships reveal approximately consistent ratios of intercept and slope. The estimated parameters are summarized in Fig. 8G and H. The external noise, estimated from the ratio intercept/slope, turns out to be of the same magnitude as the average rate. Interpreted in terms of the model, the noise resulting from Poisson spike generation thus contributes as much to neuron variance as the external input. This combined variability is propagated through the network, and, on average, multiplied by a factor $N\langle B \rangle^2 > 1$, corresponding to the slope in Eq. (28). Because this factor is larger than one, both average response and noise are amplified by the recurrent connections.

In summary, we find that a model of recurrently connected neurons with random effective connections captures the observed activity statistics, in particular the relations between average response and average covariances as well as the consistent direction of population fluctuations across stimuli. Both purely feed-forward networks and gain fluctuation models are not consistent with all of these observations. We note, however, that our conclusions are based on a relatively small data set. A larger number of trials and measurements of absolute values of firing rates would be desirable for a more stringent test.
Figure 7: **Dependence of noise distribution orientation on average response in data.** A, B: Typical example, 6 neural populations recorded in the same animal. Relative variances projected on mean and diagonal direction, versus cosine of angle between mean response and diagonal. Each marker corresponds to a different stimulus. Different markers/colors denote different populations. Squares to the left side indicate $\sqrt{\langle C_{ij} \rangle_{i\neq j}/\langle C_{ii} \rangle_i}$. Solid lines: linear fit. C: Slopes from linear fits as in A/B from $\sigma_\mu/\sigma_{all}$ vs. slope from $\sigma_d/\sigma_{all}$ for all measured populations. Circles correspond to slopes for positive, squares to negative $\cos(r, d)$. Colors in all panels indicate value of $c_N$ in populations.
4.7 Influence of noise correlations on stimulus discrimination

We quantify the influence of noise correlation on stimulus discrimination by the ratio $S_{\text{shuffled}} / S_{\text{original}}$. $S$ is calculated for the data set before and after shuffling trials. It is defined in Eq. (21) and denotes, for a pair of stimuli, the difference in average response divided by the standard deviation of the responses, both projected on the most discriminative direction. Larger values indicate that stimuli are easier to discriminate. If $S_{\text{shuffled}} / S_{\text{original}}$ is larger than one, then...
removing correlations by shuffling trials improves stimulus discrimination. Across pairs of stimuli, this signal-to-noise ratio varies strongly (Fig. 9). On average, stimuli are slightly easier to discriminate in the shuffled data, so that noise correlations are weakly unfavorable to the encoding of the stimulus set used in the experiments.

A closer analysis of the response distributions reveals that, to a large degree, the effect of shuffling can be explained by the relative locations of the two mean responses and the diagonal, as measured by \( \cos(d, r(s_1) - r(s_2)) \) (Fig. 5). This is because, due to the noise correlations, the main direction of variability is along the diagonal. The overall effect of shuffling on stimulus discrimination depends on the relation between noise and signal correlations: stronger signal correlations lead to a stronger dominance of angles that are unfavorable if noise is aligned along the diagonal direction, and in this case shuffling correlations away will benefit stimulus discrimination.

Figure 9: **Effect of correlations on stimulus discrimination.** A: Signal-to-noise ratio \( S \), dots correspond to pairs of stimuli. Value from observed covariance matrix versus the one based on shuffled trials, for different populations in the same animal. B: Same data, distribution of \( S(\text{shuffled})/S(\text{original}) \). Broad distribution, but mean value greater one for all populations (short solid lines) indicates that shuffling increases information on average. C: Distribution of cosine of angle between diagonal and average response difference across stimulus pairs for neural populations. Large cosines/small angles are most frequent. D: Scatter plot of effect of correlations on discrimination. Small dots correspond to stimulus pairs. Connected large dots indicate average value in a bin centered at the corresponding location on the x-axes. E: The distribution of mean values across all measured populations shows that correlations are on average unfavorable for almost all populations. F: Average effect of correlations is stronger for increasing signal correlations.

### 4.8 Intuition from a two-population model

The relation between network properties and the statistics of variability, on the one hand, and the relation between the statistics of variability and the encoding of stimuli, on the other hand, have been studied widely. Above, we describe the structures of correlations that emerge from different network architecture. We also analyze the structure of correlations in data from mouse cortex, and we relate them to stimulus encoding. Here, we study a highly idealized recurrent cortical network in order to develop an understanding of the relation between network characteristics, response statistics, and stimulus encoding.
To understand the effect of recurrent connections on the discriminability of a pair of stimuli, we consider two populations of excitatory neurons. The neurons in each of these populations prefer one of the stimuli, and we want to understand the effect of connections between these differently tuned neurons. The connectivity is regular, in the sense that each neuron projects to a fixed number of neurons in its own population and to a fixed number of neurons in the other population. Neurons in each population also receive the same external input. We start by considering two stimuli, $s_1, s_2$, where for each stimulus one of the populations receives a stronger input. To discriminate between these stimuli, it is sufficient to consider the network response at the population level, that is, the sum of the response of neurons in each of the two populations. This follows from the regularity of the connections. The distribution of responses across trials can thus be described by a two-dimensional vector, $R$, of the average population response and a covariance matrix, $\Sigma$, of the variability. The sum of the couplings within a population, $\Gamma_s$, and across populations, $\Gamma_c$, determine the response via a symmetric $2 \times 2$ coupling matrix, $\Gamma$. Here, we will manipulate the entries of $\Gamma$, rather than the effective population coupling matrix, $P = (I - \Gamma)^{-1}$ (see Fig. 10 and Methods for further details of the model).

We start by examining how cross-coupling between the populations can induce beneficial correlations for stimulus discrimination, by considering the effect of connections between the populations on the covariances. If we introduce more connections across populations, the cross-coupling $\Gamma_c$ increases and the response distributions evoked by the two stimuli change (Fig. 10). The responses of the two populations are amplified, but also more similar - the difference in population response, $|R(s_1) - R(s_2)|$, decreases. Simultaneously, the correlation between the population responses increases, rendering the distributions more elongated. To evaluate the effect of shuffling trials, we need to consider the covariances between individual neuron pairs. As we show in detail in Appendix 8.4, covariances reduce the overlap of the distributions if $\Gamma_c > \Gamma_s$. Then, covariances between neurons belonging to different populations are larger, on average, than for neurons of the same sub-population, and the signal-to-noise ratio, $S$, becomes smaller if covariances are shuffled away (Fig. 10C, D). While it thus appears that strong coupling between differently tuned neurons induces beneficial pairwise correlations this is not always the case.

To see this, we compare networks with different parameters, and disentangle the effects of cross-connections on average responses, variances, and covariances. Increasing the coupling between populations decreases $S$ both in the shuffled and in the original cases (Fig. 10E), but this suppression is stronger when covariances are shuffled. The suppression can be traced back to the amount of noise produced through spike generation. Recurrent amplification increases firing rates, and the variance of Poisson neurons increases with rate. If the amount of noise remained constant, $S$ would not depend on $\Gamma_c$: larger coupling induces stronger correlations which reshape noise along a direction irrelevant for discrimination, but at the same time average responses approach each other. These two effects precisely cancel each other, ultimately because a linear transformation of responses by the network cannot reduce noise.

Above, we only considered a pair of stimuli for which beneficial correlations are realized in our network. We now evaluate the effect of cross-connections on a more general stimulus ensemble. We define it via the input vectors for the two populations, $(1+\Delta \cos(\phi), 1+\Delta \sin(\phi))^T$ for $\phi \in (0, 2\pi)$ and a constant $\Delta$ (see Fig. 10G and Appendix 8.4.1). For this ensemble we see that inevitably correlations will be unfavorable for certain stimulus pairs, depending on the stimulus dimension in which their inputs differ. The averaged effect of shuffling across this two-dimensional stimulus ensemble depends not only on the noise correlations, but also on the distribution of the input, which affects the distribution of the average responses, i.e. the signal correlations.
If excitatory recurrent networks amplify internally generated noise, are feed-forward architectures generally more efficient? In the following, we compare the performances in the two scenarios. As we noted before, the difference between the two scenarios is that the noise produced by spike generation in the recurrent network is fed back into the network, and therefore correlated, while it is independent in the feed-forward network. The measure of performance we used so far is the discriminability of pairs of stimuli. We want to compare the effect on information of a recurrent network, with a transfer matrix $P$ defined by the connectivity parameters, to a feed-forward network with identical transfer matrix, for changes of a high-dimensional input in arbitrary directions. A quantity that measures the information of a network response statistics in the case of a continuous stimulus is the linear Fisher information matrix. In the two-population model, it is given by

$$I_R = (D[R] + \Sigma_{ext})^{-1}$$  \hspace{1cm} (35)

in the recurrent case and by

$$I_F = P^T(P\Sigma_{ext}P^T + D[R])^{-1}P$$ \hspace{1cm} (36)

in the feed-forward case. Here, $\Sigma_{ext}$ denotes the variance of the external input to the two populations (see Appendix 8.4.3 for details). As figures of merit for the coding of high-dimensional stimuli, we use the traces of $IR$ and $IF$, for the recurrent and feed-forward cases respectively. In the recurrent network, this quantity depends only on the firing rates. These, in turn, depend on the average strength of the effective connections, $\langle P \rangle$ (Fig. 10H): the larger $\langle P \rangle$, and therefore the response, the more noise is generated, and the more information is lost. In a corresponding feed-forward network, information depends also on the correlations. However, because the network-generated noise is uncorrelated, increasing firing rates effectively improves the signal-to-noise ratio, and therefore increases the information. Whether a feed-forward or a recurrent architecture is preferable thus depends on the output firing rate. From this model, it is not clear how firing rates should be chosen, but additional constraints, for example imposed by energy consumption are conceivable.

A simple model with two coupled, uniform populations illustrates how variations in the network parameters can affect stimulus discriminability, through modulations in the average responses and variability. Broadly speaking, stimulus discriminability depends on the interplay between noise creation, average response, and noise correlation, and the relative behavior of these quantities depends on the network architecture.
Figure 10: Model of two coupled excitatory populations. A: Circuit model. Neurons are coupled with excitatory connections within and across two populations. For each stimulus, one population receives stronger input. Population responses are described by macroscopic variables. B: Responses to two stimuli (colored dots on dotted red line) are amplified for stronger cross-coupling (brighter colors). Simultaneously, response fluctuations are more correlated. If internally generated noise does not increase with rate, variability is smaller (dashed ellipses). C and D: Distribution of pairwise covariances within (dark bars) and across populations (light bars), vertical lines indicate average. If cross-coupling is weak (C), average covariance is larger within populations than across, if it is strong (D), covariances between neurons of different populations are larger. E: Dependence of stimulus discriminability $S$ on $\Gamma_c$. Strong couplings implies smaller discriminability and even more so for shuffled trials. F: Increasing cross-coupling reduces difference in mean response (black line). Effect on variance in the relevant direction differs for correlated populations (solid orange line) and model with independent neurons/shuffled correlations (dash-dotted orange line). G: Red dots: average response for stimulus ensemble. Ellipses: Response distributions for four stimuli for the recurrent network (blue) and a feed-forward network with identical average responses (red). Dotted/dashed lines: stimulus pairs where correlations have positive/negative influence. H: Linear Fisher information, dependence on average of population transfer matrix elements (horizontal axes) for recurrent and feed-forward scenario. Darker lines indicate larger variance of transfer matrix elements. Lines overlap in the recurrent scenario.
5 Discussion

Despite large amounts of work on the topic, the interpretation of the observed response fluctuations in population activity, in the context of computation by neural circuits and information coding, is not elucidated [11]. In the present study, we seek to understand the influence of correlated variability on stimulus representation, based on the exploration of model networks of stochastic neurons together with data analysis. We modeled neural activity using a set of coupled Poisson processes in order to study the relations between covariances and average responses in a neural population, across different stimuli. We found that these relations differ depending on how covariances are generated, and that correlations observed in mouse auditory cortex are consistent with those generated in a recurrent network model. This model also helps us interpret the effect of the observed correlations on the discriminability of stimuli and the nature of the propagation of information through the network.

5.1 Origins of the structure of correlations in neural circuits

In the examination of the impact of correlated response variability on stimulus representation, it is useful to be aware of the possible range of its mechanistic origin. Spike train variability has been attributed to the dynamics in recurrent networks [1, 37]. Along with this source of variability, in stable states of asynchronous activity, recurrent network models can also produce transitions between different states of global activity [3, 38, 39]. Alternatively, external signals fed to sub-populations of neurons, such as gain fluctuations induced by upstream areas, were pointed to as causing high variability and correlations [7, 15, 9, 8, 24]. Population-wide gain fluctuations can also reproduce dependencies between noise and signal correlations [10].

In general, it is difficult to differentiate shared input or gain fluctuations, and recurrent connections, as the origin of correlated variability, based on measured activity. Here, instead of focusing on temporal dynamics [40, 13, 41], we approach the problem by considering the relation between the network architecture and the structure of population activity statistics, when a neural population is presented with an ensemble of stimuli. Our analysis makes the assumption that noise resulting from spike generation in neurons can be modeled as a Poisson process, and, thus, depends on the firing rate. If the noise generated within the network is purely additive and independent of firing rates, it is indistinguishable from external noise that is simply filtered by the network [35]. In the Poisson model, however, the interplay of internally generated noise and variability due to external signals imprints its signature on the dependence of covariances on firing rates.

5.2 Influence of correlations on stimulus discrimination in neural populations

We use a recurrent network with random effective connections to estimate the influence of the variability generated by the network dynamics on stimulus representation. Based on the paradigm that response variability amounts to noise that downstream neurons have to cope with, a number of authors have argued that noise correlations can benefit information coding [19, 21, 23, 22, 24, 25]. Furthermore, recent studies have examined specifically the impact of correlations generated by recurrent networks. Reference [30] stresses that it is the structure of the noise, more than its magnitude, which determines the role it plays for information coding. Harmful structures of noise correlation, causing information to saturate in large neural populations, are the consequence of the amount of irreducible external noise and potentially sub-optimal processing [31], but these structures are difficult to pinpoint.
in detail in measured activity [30]. By contrast, it was demonstrated that noise internally generated in a network via spike generation can be averaged out in feed-forward or recurrent networks, depending on the connectivity in the network and the redundancy of the population code [27].

In this work, the goal was to start from the observed properties of the activity, and to interpret mean activity and noise correlations based on outcomes of model network dynamics. Variations of network properties, such as the architecture of connections, change both the mean activity and noise correlations, both of which influence the accuracy of the neural code. Consequently, it is useful to consider both simultaneously to understand network effects. The effect of recurrent dynamics on input information depends mostly on the amount of internally generated noise. Our toy model, as well as Ref. [27], argue that recurrent amplification suppresses information. The reason is essentially that, for Poisson neurons, larger firing rates imply a higher variance. The stronger the recurrent connections, the more the average response is amplified, and the more noise is produced internally. Because this noise is fed back into the network and thus amplified furthermore, the variance is increased to the point that the signal-to-noise ratio decreases. We also show that under simplifying assumptions, including a random effective network architecture and a large, random stimulus ensemble, the amplification factor of the network and the amount of noise can be estimated from measured responses.

Limitations of our analysis result, in particular, from the assumption of linearly interacting Poisson processes. Nonlinear transfer functions can affect the relation between rates and correlation in individual pairs [42, 11]. Using Poisson neurons also implies that the generation of noise in the network is attributed to individual neurons. The tractability of the model, however, allows us to obtain relations between observable statistics, like noise and signal correlations, which can be used as a starting point to interpret correlated variability in terms of its mechanistic origins.

5.3 Variability in mouse auditory cortex

We tested the consistency of different scenarios involving interacting Poisson neurons with measured neural population responses, and we interpreted the observed variability in terms of a recurrent network model. The measured population activity is given as spike counts in 200 ms bins, so that correlations are defined over a relatively slow timescale and, therefore, do not take into account fast temporal structure (as noted in, e.g., Ref. [43]). Both experimental variability and correlations are high [44], so that the effects of correlations are potentially strong and amenable to an analysis such as ours. Although variability and correlations are often high in anesthetized animals [9], noise and signal correlations were somewhat weaker in comparable experiments with anesthetized animals [45]. It is conceivable that part of the correlations are due to experimental artifacts from the calcium imaging technique used in the experiments (like scattering of fluorescent light by the neuropil), but intracellular recordings of rates were consistent with calcium recordings [32].

We model the set of stimuli as a generic, high-dimensional ensemble, and the network connections at the level of a transfer matrix. When analyzing the discriminability of pairs of stimuli, this enables a broader analysis than one relying on a single stimulus dimension or average correlation. The recurrent network model explains not only the scaling of variances and covariances with firing rates, but also the relation between signal and noise correlations, and the effect of shuffling away covariances on stimulus discrimination, in a natural way. Differences among populations of neurons can be attributed to changes in the effective connectivity within populations.

Recent studies have explained correlated variability in terms of “external” gain fluctu-
ations acting on a (sub-)population \[7, 15, 9, 8, 24\]. We showed that recurrent networks of strongly connected neurons can exhibit population-wide fluctuations as well. Estimated model parameters indicate an amplification of input signals by excitatory recurrent connections. This suggests that the noise that can be traced back to spike generation in the population — noise harmful to coding — is comparable in magnitude to the external input noise.

Our data analysis suffers from an unknown offset in the firing rates, a relatively small number of trials, and the low temporal resolution. We therefore only compared different prototypical scenarios without trying to evaluate the outcome of a mixed model. In experimental populations of neurons, the statistics of activity can obviously result from a combination of both feed-forward and recurrent processes.

6 Acknowledgments

We thank Brice Bathellier for providing the experimental data set as well as commenting on it, and for discussions.

This work was supported by the CNRS through UMR 8559 and the SNSF Sinergia Project CRSII3_141801.

7 Author Contributions

Conceived and designed the experiments: VP RS. Performed the experiments: VP. Analyzed the data: VP. Contributed reagents/materials/analysis tools: VP RS. Wrote the paper: VP RS.
8 Appendix

8.1 Numerical simulations: details and parameters

In Fig. 2, covariances were calculated for recurrent networks from Eq. (12), with the addition of a rate offset \(a = 4\), so that \(C = (I - G)^{-1} (D[r] + a + D[V_{\text{ext}}]) (I - G')^{-1}\). The stimulus ensemble consisted of 200 input vectors \(r_{\text{ext}}\) of size \(N = 60\), with entries independently chosen from a normal distribution with mean 0.2 and standard deviation 1. Rates \(r\) were calculated as in Eq. (7), and we assumed Poisson input with \(V_{\text{ext}} = |r_{\text{ext}}|\). For the networks, four connectivity matrices \(G\) of size \(N = 60\) were generated. Their entries were chosen from a normal distribution with mean 0.4 and variance 2. Connectivity in the recurrent network was as in Fig. 2. For rates and covariances of the feed-forward scenario, we use Eqs. (15) and (16). We set \(F = 10 \cdot B\), and scaled the input vectors by a factor 1/10 to obtain rates and correlations of comparable size as in the recurrent network. Negative output rates were set to zero. For the gain fluctuation model, responses of the network with \(\rho = 1.4\) were used as the ensemble of stimulus responses. From these, covariances were calculated, see Eq. (18), for which the variance of external fluctuations \(V_{\text{ext}}\) was varied. For this figure, no rate offset was used, \(a = 0\).

The networks in Fig. 4C were identical to the ones used above. 50 stimuli \(r_{\text{ext}}\) of length \(N = 60\) were chosen from a normal distribution with mean 0.2 and standard deviation 1.5. Inputs for different neurons were correlated with correlation coefficient \(c_{\text{in}} = 0.05\) across stimuli. Rate offset was 4.

In Fig. 5, networks and stimulus parameters were the same as in Fig. 4 but there were no input signal correlations, \(c_{\text{in}} = 0\).

In Fig. 10, the network connects two populations of 100 neurons each. For the pair of stimuli, panel B, neurons in the population that was more strongly excited received an input of \(1 + \Delta\), with \(\Delta = 0.2\), and the others input 1. More general inputs (panel G) were defined by setting the inputs \(R_{\text{ext}} = (1 + \cos(\phi)\Delta, 1 + \sin(\phi)\Delta)\) for \(\phi \in (0, 2\pi)\). Each neuron had a fixed number of \(n_{EE}\) postsynaptic partners within the same populations. Coupling strength between connected neurons was set to \(g = 0.01\). We chose a population coupling within populations of \(\Gamma_s = 0.2\), and the number of connections between neurons within populations such that \(\Gamma_s = n_{EE}g\). Connections across populations were chosen accordingly to realize across coupling values \(\Gamma_c = 0, \ldots, 0.4\). To visualize the effect of the increase in internally generated noise due to increased rates (panel B), we compared the covariances \(\Sigma = B^\rho D[R](B^\rho)^T\) for increasing \(\Gamma_c\) to the ones where in \(D[R]\) the rates of a network with \(\Gamma_c = 0\) were used.

In panels E and F, we calculated \(S_{\text{original}}\) by inserting the population covariances and responses in Eq. (21) for the two stimuli, for the correlated case. For the shuffled case, the full matrix covariance matrix \(C\) was generated and the off-diagonal elements set to 0, such that

\[
\Sigma_{\text{shuffled}} = \begin{pmatrix}
\sum_{i \in E} C_{ii} & 0 \\
0 & \sum_{i \in E^c} C_{ii}
\end{pmatrix}.
\] (37)

For panels E-H, population level parameters in the matrices \(\Gamma\) or \(P\) were varied directly. We set the external noise \(\Sigma_{\text{ext}} = ND[R_{\text{ext}}]\). In panel G, \(\Gamma_c = 0.4\) and \(\Gamma_s = 0.2\). In panel H, we fixed for the population transfer matrix \(P\) the ratio \(\rho = \langle P \rangle^2 / \text{var}(P)\) to the values \(\rho = 0.5, 1, 2\). For each of these values, \(\langle P \rangle\) was varied between 0.1...5, so that \(\text{var}(P)\) was fixed. From \(\langle P \rangle\) and \(\text{var}(P)\), we calculated \(P_s\) and \(P_c\), see Appendix 8.4.1.
8.2 Common framework for the generation of correlations

Here we formulate the three model scenarios described in Sections 3.3 and 3.4 as special cases of the interacting point processes framework defined in Eq. (6). We consider an extended network of two populations which receive only constant input \( r_{\text{full}}(t) \equiv r_{\text{full}} \). Neurons are divided into an observed population \( O \) and an unobserved external population \( U \). The external population projects to the observed population, but does not receive feedback. The coupling matrix of integrated kernels of the full system is a block matrix of the shape

\[
G_{\text{full}} = \begin{pmatrix} E & 0 \\ F & G \end{pmatrix}.
\]  

(38)

The coupling between nodes within the external network is described by the matrix \( E \). If \( E = 0 \), external inputs are independent Poisson processes. The feed-forward weights to the observed network are defined by \( F \), and by \( G \) the recurrent connections within the observed population. The input vector to the system is \( r_{\text{full}} = (r_0, 0) \). The components of the vector \( r_0 \) together with \( E \) determine the firing rates of the input nodes. The 0 represents a vector of zeros, so that there is no constant input directly to the observed population.

The time dependent firing rates of neurons \( k \in U \) in the external input population are determined by

\[
\tilde{r}_k(t) = \sum_{t \in U} \int_0^\infty \tilde{e}_{kl}(\tau) \tilde{s}_l(t - \tau) d\tau + r_{0,k},
\]  

(39)

and the rates of neurons \( i \in O \) in the observed population by

\[
\tilde{r}_i(t) = \sum_{j \in O} \int_0^\infty \tilde{g}_{ij}(\tau) \tilde{s}_j(t - \tau) d\tau + \sum_{k \in U} \int_0^\infty \tilde{f}_{ik}(\tau) \tilde{s}_k(t - \tau) d\tau.
\]  

(40)

The transfer matrix of the full system is

\[
B_{\text{full}} = (I - G_{\text{full}})^{-1} = \begin{pmatrix} (I - E)^{-1} \\ (I - G)^{-1}F(I - E)^{-1} \end{pmatrix} = \begin{pmatrix} B_E & 0 \\ BFB_E & B \end{pmatrix}.
\]  

(41)

The average rates of the system are, from applying Eq. (7) to the full system,

\[
(I - G_{\text{full}})^{-1}r_{\text{full}} = B_{\text{full}}r_{\text{full}} = (B_{E}r_{0}, BFB_{E}r_{0})^T \equiv (r_{\text{ext}}, r)^T,
\]  

(42)

where \( r_{\text{ext}} = B_{E}r_{0} \) are the rates of the input neurons and \( r = BFB_{E}r_{0} = BFR_{\text{ext}} \) correspondingly, the rates of the observed neurons. From Eq. (12), the covariance matrix is given by

\[
C_{\text{full}} = \begin{pmatrix} B_{E}D[r_{\text{ext}}]B_{E}^T \\ BFB_{E}D[r_{\text{ext}}]B_{E}^T \end{pmatrix} B_{E}D[r_{\text{ext}}]B_{E}^TFTB_{E}^T \begin{pmatrix} B_{E}D[r_{\text{ext}}]B_{E}^TFTB_{E}^T & B_{E}D[r_{\text{ext}}]B_{E}^TFTB_{E}^T \end{pmatrix} \equiv B(D[r]) + FCE_{\text{ext}}F^T)
\]  

(43)

The block \( C_{E} \equiv B_{E}D[r_{\text{ext}}]B_{E}^T \) at the upper left describes the covariances of the external population. The block at the lower right,

\[
C \equiv B(D[r]) + FC_{E}F^T)
\]  

(44)

describes how covariances in the observed network depend on the properties of the unobserved input neurons.

The recurrent network scenario is obtained for an input population and an output population of equal size \( N \), with \( F = I \) and any diagonal \( E \). In this case, \( C_{E} \) is diagonal, and we can identify the elements on the diagonal of \( C_{E} \) with \( V_{\text{ext}} \).
For the feed-forward scenario set $G = 0$ and again identify the diagonal elements of $C_E$ with $V_{\text{ext}}$.

The gain fluctuation model is obtained for $G = 0$, and if the input population consists of a single neuron. The matrix $F$ then corresponds to a single column vector, and $C = D[F r_{\text{ext}}] + F C_E F^T$. With $r = F r_{\text{ext}}$ and setting $V_{\text{ext}} = C_E / r_{\text{ext}}^2$ ($r_{\text{ext}}$ is a number in this case), $C = D|r| + r r^T V_{\text{ext}}$.

### 8.3 Further details on statistics in different model scenarios

#### 8.3.1 Gain fluctuation model

We derive relations between covariances and average response in the gain fluctuation model, in particular the scaling of average covariances with average rates and the orientation of the response distribution measured by the projections of the variances in different directions. From Eq. (18), pairwise covariances are directly related to average responses,

$$C_{ij} = \delta_{ij} r_i + r_i r_j V_{\text{ext}}.$$  

(45)

The relation between average covariance and the population averaged response, Eq. (32), follows from

$$\langle C_{ij} \rangle_{i \neq j} = \frac{1}{N(N - 1)} \sum_{i \neq j} r_i r_j V_{\text{ext}} \approx \frac{V_{\text{ext}}}{N^2} \left( \sum_i r_i \right) \left( \sum_j r_j \right) = V_{\text{ext}} \langle r \rangle^2,$$

(46)

where the terms $r_i^2$ can be neglected if $N$ is large.

To measure changes of the response distribution across stimuli, we use the variance projected in the direction of mean response and diagonal direction, $\sigma_d^2$ and $\sigma_a^2$, respectively. In the following, we give an argument that in this model $\sigma_a^2$ does not strongly depend on the stimulus, while $\sigma_d^2$ does.

Both quantities are normalized by the sum of the variances

$$\sigma_{all}^2 = \sum_i C_{ii} = \sum_i (r_i^2 V_{\text{ext}} + r_i) = |r|^2 V_{\text{ext}} + N \langle r \rangle.$$

(47)

The variance projected in the diagonal direction $d = (1, \ldots, 1)^T / \sqrt{N}$ is

$$\sigma_d^2 = \sum_{ij} C_{ij}(s) d_i d_j = \sqrt{N} \langle r_i r_j \rangle / N = N V_{\text{ext}} \langle r \rangle^2 + \langle r \rangle.$$ (48)

To see that there is a strong dependence of $\sigma_d^2 / \sigma_{all}^2$ on stimulus direction, note that

$$\cos(d, r) = d r^T = \sum_{i=1}^{N} \frac{1}{\sqrt{N}} \frac{r_i}{|r|} = \sqrt{N} \langle r \rangle / |r|.$$ (49)

For non-vanishing $r$, it follows from $\cos(d, r) = 0$ that $\langle r \rangle = 0$ and in this case $\sigma_d^2 / \sigma_{all}^2 = 0$.

If $\cos(d, r) = 1$, $r \propto d$ and $|r|^2 = N \langle r \rangle^2$, and so $\sigma_d^2 / \sigma_{all}^2 = \frac{V_{\text{ext}} \langle r \rangle^2 + \langle r \rangle}{N V_{\text{ext}} \langle r \rangle^2 + N \langle r \rangle}$. This value is of the order of one, if the ratio of $V_{\text{ext}} \langle r \rangle^2$ and $\langle r \rangle$, which is of the order of the average noise correlation coefficient, is not very small (in comparison to one). Consequently, if noise correlations are not too small, the normalized variance projected on the diagonal direction strongly depends on the direction of the response vector.

The variance projected on the mean direction is

$$\sigma_d^2 = \sum_{ij} \bar{r}_i \bar{r}_j C_{ij} = \sum_{ij} \bar{r}_i \bar{r}_j (r_i r_j V_{\text{ext}} + \delta_{ij} r_i) > V_{\text{ext}} \sum_{ij} r_i^2 r_j^2 / |r|^2 = V_{\text{ext}} |r|^2$$ (50)
Hence, $\sigma_{\mu}^2/\sigma_{\text{all}}^2 > \frac{V_{\text{ext}}|r|^2}{|r|^2V_{\text{ext}} + N\langle r \rangle^2}$, which depends only weakly on $\langle r \rangle$, if $N\langle r \rangle$ is not much bigger than $V_{\text{ext}}|r|^2 \geq V_{\text{ext}}N\langle r \rangle^2$, that is once again if noise correlations are not too small.

### 8.3.2 Additional test for a common origin of shared variability

If rates are only measured up to an unknown offset $a$, the variance $\sigma_{\mu}^2/\sigma_{\text{all}}^2$, which is obtained from a projection on the apparent average response $r - a$ may not be constant across stimuli. In this case, the variation of this quantity across stimuli is not a reliable indicator to exclude a gain fluctuation model as the source of observed correlations. However, if a large part of the covariances can be explained by a single component, it should be possible to reconstruct the common origin from the stimulus dependent covariance matrices. From

$$C(s) = (r(s) + a)^T (r(s) + a)V_{\text{ext}} + D(r(s) + a)$$

the vectors $v(s) = r(s) + a$ can be obtained approximately by finding the eigenvector with the largest eigenvalue of $C$, by neglecting the contribution of $D(r(s) + a)$. This approximation can also be avoided by applying a factor analysis with a single latent component. Because $a$ is constant across stimuli, up to measurement errors of the estimated covariances, the straight lines through the mean responses $r + x\bar{v}$, for $x \in \mathbb{R}$ and normalized directions $\bar{v}$, will intersect in the point $-a$. The best intersection point of multiple lines in the least square sense is given by

$$-\hat{a} = \left( \sum_s (\mathbb{I} - \bar{v}(s)\bar{v}(s)^T) \right)^{-1} \left( \sum_s (\mathbb{I} - \bar{v}(s)\bar{v}(s)^T)r(s) \right)$$

and can be used to correct the average responses to $r'(s) = r(s) + \hat{a}$. We found that the analysis of $\sigma_{\mu}^2$ and $\sigma_{\text{all}}^2$ based on the corrected responses lead to no qualitative change in the stimulus dependence, indicating that a potential shared component is too weak to be identified based on the available data.

### 8.3.3 Feed-forward model

Here we derive characteristic relations for the response statistics resulting in a feed-forward network. They illustrate qualitative differences between the predictions of different models and will be used to extract model parameters from the data. In the feed-forward model the mean responses and covariances are, from Eqs. (15) and (16) and allowing an offset in observed rates,

$$r = Fr_{\text{ext}},$$

$$C_{ij} = \delta_{ij}(r_i + a) + \sum_k F_{ik}F_{jk}V_{\text{ext},k}.$$  

In contrast to the recurrent network model (see below), the average covariance is not correlated to the population averaged rate $\langle r \rangle$ across stimuli, Eq. (30). This follows after taking the average across neurons:

$$\langle C_{ij} \rangle_{i \neq j} = \frac{1}{N(N-1)} \sum_{i \neq j,k} F_{ik}F_{jk}V_{\text{ext},k} \approx N\langle F_{ik} \rangle_{ik}\langle F_{jk} \rangle_{jk}\langle V_{\text{ext},k} \rangle_k = N\langle F \rangle^2\langle V_{\text{ext}} \rangle.$$  

Input variances $V_{\text{ext}}$ are assumed to be independent of the network structure and $F_{ik}$ independent of $F_{jk}$, which means that the strengths of connection of an external neuron to different internal neurons are independent.

If $\langle C_{ij} \rangle_{i \neq j}$ is to be uncorrelated to $\langle r \rangle$ across stimuli, the average input variance $\langle V_{\text{ext}} \rangle$ needs to be approximately uncorrelated to $\langle r \rangle$. This holds in our model even though $r_k =$
$\sum_{i} F_{ik} r_{ext,k}$ and in each input channel the variance equals the strength of the input, $V_{ext,k} = |r_{ext,k}|$, if either the distribution of inputs across neurons is approximately symmetric around 0, or the distribution of column sums $\sum_{i} F_{ik}$ of the feed-forward matrix (across columns) is symmetric around 0. The first case can be realized in a feed-forward network with a similar number of excitatory and inhibitory input channels. Intuitively, inhibitory inputs decrease the average output, but contribute positively to the average variance, and thus decorrelate the two quantities.

Formally, one compares the random variable $\sum_{k} r_{k} = \sum_{k} F_{ik} r_{ext,k} = \sum_{k} r_{ext,k} \sum_{i} F_{ik}$ to the variable $\sum_{k} V_{ext,k} = \sum_{k} |r_{ext,k}|$. A random variable $x$ is uncorrelated to its absolute value $|x|$ if its distribution is symmetric around 0. Here, the relevant variables are the elements of the input vector $r_{ext}$, and their distribution is approximately symmetric around 0, if the variance $\text{var}(r_{ext})$ is much larger than their mean $\langle r_{ext} \rangle$, that is if $\rho_{E} = \text{var}(r_{ext})/\langle r_{ext} \rangle \gg 1$. Consequently, if the variance of external inputs across stimuli is high, the population averaged covariances in a feed-forward network are uncorrelated to the population response.

In the following, we motivate that the stimulus dependence of the variances projected along different directions in the feed-forward model is different from the one in the gain fluctuation model. The sum of the variances is

$$\sigma_{all}^2 = N(\langle r \rangle + a) + \sum_{i,k} F_{ik}^2 V_{ext,k} \approx N(\langle r \rangle + a) + N^2 \langle V_{ext} \rangle \langle F^2 \rangle. \quad (56)$$

Due to the Poisson spike generation, there is a linear contribution in $\langle r \rangle$ to the average variance, $\langle C_{ii} \rangle_i = \sigma_{all}^2/N$. This variability does not contribute to covariances, because spikes are generated independently across neurons, in contrast to the recurrent model, where the spiking or not spiking of a neuron directly influences post-synaptic firing rates.

The variance projected onto the diagonal direction is

$$\sigma_{d}^2 = \langle r \rangle + a + \frac{1}{N} \sum_{ijk} F_{ik} F_{jk} V_{ext,k} \approx \langle r \rangle + a + N^2 \langle V_{ext} \rangle \langle F^2 \rangle. \quad (57)$$

Consequently, the ratio $\sigma_{d}^2/\sigma_{all}^2$ depends only weakly on $\langle r \rangle$ if the term $N \langle V_{ext} \rangle \langle F^2 \rangle$ is large against $\langle r \rangle + a$, that is if the sum of covariances is larger than the sum of variances. This is the case, if correlation coefficients are larger than of the order $O(1/N)$. The variance projected onto the direction of the mean response is

$$\sigma_{\mu}^2 = \sum_{i} \langle r_i + a \rangle r_i^2 + \sum_{ijk} F_{ik} F_{jk} V_{ext,k} \bar{r}_i \bar{r}_j \approx a + \sum_{i} \frac{r_i^2}{\langle r \rangle^2} + \frac{\langle r \rangle^2}{\langle r \rangle^2} N^3 \langle F^2 \rangle \langle V_{ext} \rangle. \quad (58)$$

To see that the ratio $\sigma_{\mu}^2/\sigma_{all}^2$ strongly depends on the population response, in contrast to $\sigma_{d}^2/\sigma_{all}^2$, assume again that the sum across variances in the first term $a + \frac{\sum_i r_i^2}{\langle r \rangle^2}$ is not too large against the sum across covariances in the second term. The second term depends strongly on $\langle r \rangle$: it is 0 for $\cos(r, d) = 0$. For $\cos(r, d) = 1$ it becomes $N^2 \langle F^2 \rangle \langle V_{ext} \rangle$. In this case, $\sigma_{\mu}^2/\sigma_{all}^2$ is not too small, if $(\langle r \rangle + a)/(N V_{ext})$ is not much larger than 1 (assuming that $\langle F^2 \rangle/\langle F \rangle^2$ is of order one).

Finally, we calculate the average noise correlation coefficient (across stimuli and neuron pairs)

$$c_N = \left( \frac{C_{ij}(s)}{\sqrt{C_{ii}(s) C_{jj}(s)}} \right)_{s,i \neq j}. \quad (59)$$

For a given stimulus, we assume that the $C_{ij}$ are approximately independent, so that we can write

$$\left( \frac{C_{ij}(s)}{\sqrt{C_{ii}(s) C_{jj}(s)}} \right)_{i \neq j} \approx \langle C_{ij}(s) \rangle_{i \neq j} = \langle C_{ij}(s) \rangle_{i \neq j}. \quad (60)$$
Then, from Eqs. (55) and (56),

\[ c_N = \frac{N \langle F \rangle^2 \langle V_{\text{ext}} \rangle}{\langle r \rangle + a + N \langle V_{\text{ext}} \rangle \langle F^2 \rangle}. \tag{61} \]

Signal correlations can be calculated analogously as for the recurrent model, see below.

### 8.3.4 Recurrent network model

The response distributions in the recurrent network are characterized by the average and covariances of responses given in Eqs. (7) and (12),

\[ r = B r_{\text{ext}} \tag{62} \]

and

\[ C = BD[r_{\text{eff}}]B, \tag{63} \]

with \( r_{\text{eff}} = r + a + V_{\text{ext}} \), including an offset \( a \). To compare this model to the alternative scenarios described in the previous sections, we calculate the relation between average (co)variances and the projected variances and \( r \) as well as the signal and noise correlations.

We first derive Eqs. (27) and (28). The average covariance is

\[ \langle C_{ij} \rangle_{i \neq j} = \frac{1}{N(N - 1)} \sum_{i \neq j, k} B_{ik} B_{jk} r_{\text{eff}, k} \approx N \langle B_{ik} B_{jk} \rangle_{ijk} \langle r_{\text{eff}, k} \rangle_k. \tag{64} \]

We assume that \( N \) is sufficiently large and that the external input characterized by \( V_{\text{ext}}, r_{\text{ext}} \) does not depend on the recurrent network. In that case \( r_{\text{eff}, k} \) is approximately uncorrelated to \( B_{ik} \) (including the contribution of \( r_k \), because \( \langle B_{ij} r_k \rangle = \langle B_{ij} \sum_l B_{il} r_{\text{ext}, l} \rangle \approx \langle B_{ij} \rangle \langle \sum_l B_{il} r_{\text{ext}, l} \rangle \)). The term \( \langle r_{\text{eff}} \rangle = \langle r \rangle + a + \langle V_{\text{ext}} \rangle \) is linear in \( \langle r \rangle \), with an uncorrelated contribution \( \langle V_{\text{ext}} \rangle \), see Appendix 8.3.3. If the elements of \( B \) are pairwise independent,

\[ \langle C_{ij} \rangle_{i \neq j} = N \langle B \rangle^2 (\langle r \rangle + a + \langle V_{\text{ext}} \rangle). \tag{65} \]

Under the same assumptions, the sum of variances is

\[
\sigma_{\text{all}}^2 = \sum_{i, k} B_{ik}^2 r_{\text{eff}, k} = \sum_{i, k} B_{ik}^2 (V_{\text{ext}, k} + r_k + a) \\
\approx N \langle B \rangle^2 \sum_k (V_{\text{ext}, k} + r_k + a) = N^2 \langle B \rangle^2 (\langle r \rangle + a + \langle V_{\text{ext}} \rangle). \tag{66}\]

Using the same argument as in the feed-forward scenario following Eq. (55), the term \( N^2 \langle B \rangle^2 \langle V_{\text{ext}} \rangle \) only contributes with an offset to the relation between \( \langle r \rangle \) and \( \langle C_{ii} \rangle_i = \sigma_{\text{all}}/N \).

Next, we show that the projected variances in this scenario have a similar dependence on \( \langle r \rangle \) as the ones in the feed-forward model. The projected variance on the diagonal is

\[ \sigma_d^2 = \frac{1}{N} \sum_{i, j, k} B_{ik} B_{jk} r_{\text{eff}, k} \approx N^2 \langle B \rangle^2 (\langle r \rangle + a + \langle V_{\text{ext}} \rangle). \tag{67} \]

Consequently, in the ratio \( \sigma_d^2 / \sigma_{\text{all}}^2 \) the dependence on \( \langle r \rangle \) is canceled out.

The projection along the mean response direction,

\[ \sigma_{\mu}^2 = \sum_{i, j, k} B_{ik} B_{jk} (r_k + a + V_{\text{ext}, k}) \bar{r}_i \bar{r}_j \approx N^3 \frac{\langle r \rangle^2}{|\langle r \rangle|^2} \langle B \rangle^2 (\langle r \rangle + a + \langle V_{\text{ext}} \rangle), \tag{68} \]

33
depends strongly on \( r \). Because \( \cos(d, r) = 0 \) implies \( \langle r \rangle = 0 \), \( \sigma^2 / \sigma^2_{\text{all}} = 0 \) in this case, and for \( \cos(d, r) = 1 \), with \( |r|^2 = N\langle r \rangle^2 \) one gets \( \sigma^2 / \sigma^2_{\text{all}} = \langle B \rangle^2 / \langle B^2 \rangle \).

Apart from the raw covariances, we are interested the average noise correlation coefficient (across stimuli and neuron pairs). As for the feed-forward model, we approximate \( c_N \approx \frac{\langle C_{ij} \rangle_{s, i \neq j}}{\langle C_{ii} \rangle_{s, i}} \). From Eqs. (65) and (66)

\[
c_N = \frac{N\langle B \rangle^2(\langle r \rangle + a + (V_{\text{ext}}))}{N\langle B^2 \rangle(\langle r \rangle + a + (V_{\text{ext}}))} = \frac{\langle B \rangle^2}{\langle B^2 \rangle},
\]

and because \( \langle B^2 \rangle = \text{var}(B) + \langle B \rangle^2 \)

\[
c_N = \frac{1}{1 + \text{var}(B)/\langle B \rangle^2} = \frac{1}{1 + \rho}.
\]

We see that the correlation coefficient depends on the relative variability of the network elements. The variances, resulting from the variances of the input channels, are determined by the mean of the square elements of the network. By contrast, the covariances depend on the effective weights of inputs to the neuron pairs, and hence on the square of the mean weight. Correspondingly, we consider the signal covariances \( C^S_{ij}(s) = \text{cov}(r_i(s), r_j(s))_s \),

\[
C^S_{ij} = \text{cov}\left( \sum_{kl} B_{ik} r_{\text{ext},k}(s), B_{jl} r_{\text{ext,l}}(s) \right)_s = \sum_{kl} \text{cov}(B_{ik} r_{\text{ext},k}(s), B_{jl} r_{\text{ext,l}}(s))_s.
\]

For \( i \neq j \), one gets

\[
C^S_{ij} = \sum_{kl} B_{ik} B_{jk} \text{cov}(r_{\text{ext},k}(s), r_{\text{ext,l}}(s))_s = \sum_k B_{ik} B_{jk} \text{var}(r_{\text{ext},k}(s)) + \sum_{k \neq l} B_{ik} B_{jk} \text{cov}(r_{\text{ext},k}(s), r_{\text{ext,l}}(s))_s.
\]

Averaged across neurons

\[
\langle C^S_{ij} \rangle_{i \neq j} = N\langle B \rangle^2 \text{var}(r_{\text{ext}}) + N(N - 1)\langle B \rangle^2 c_{\text{in}} \text{var}(r_{\text{ext}}).
\]

For the signal variances, \( i = j \), which correspond to the variance of the rates across stimuli,

\[
\langle C^S_{ii}(s) \rangle_s = \sum_k \text{var}(B_{ik} r_{\text{ext},k})(s) + \sum_{k \neq l} \text{cov}(B_{ik} r_{\text{ext},k}, B_{il} r_{\text{ext,l}}) = \sum_k B^2_{ik} \text{var}(r_{\text{ext},k}) + \sum_{k \neq l} \text{cov}(r_{\text{ext},k}, r_{\text{ext},l}) B_{ik} B_{il}.
\]

Averaged across neurons, with \( \langle B^2 \rangle = \langle B \rangle^2 + \text{var}(B) \):

\[
\langle C^S_{ii}(s) \rangle_{s,i} = N \text{var}(r_{\text{ext}})(\langle B \rangle^2 + \text{var}(B)) + N(N - 1)\langle B \rangle^2 \text{var}(r_{\text{ext}}) c_{\text{in}}.
\]

Their ratio is

\[
\frac{\langle C_{ij} \rangle}{\langle C_{ii} \rangle} = \frac{N\langle B \rangle^2 \text{var}(r_{\text{ext}})(1 + (N - 1)c_{\text{in}})}{N\text{var}(r_{\text{ext}})(\langle B \rangle^2 + \text{var}(B)) + N(N - 1)\langle B \rangle^2 \text{var}(r_{\text{ext}}) c_{\text{in}}} = \frac{1 + (N - 1)c_{\text{in}}}{1 + \text{var}(B)/\langle B \rangle^2 + (N - 1)c_{\text{in}}}.
\]

This results in an approximate expression for the average signal correlation coefficient

\[
c_S = \langle \frac{C^S_{ij}}{\sqrt{C^S_{ii}C^S_{jj}}} \rangle_{i,j} \approx \frac{\langle C^S_{ij} \rangle}{\langle C^S_{ii} \rangle} = \frac{1 + (N - 1)c_{\text{in}}}{1 + \text{var}(B)/\langle B \rangle^2 + (N - 1)c_{\text{in}}}.
\]

34
8.3.5 Estimation of network model parameters

We extract parameters of the recurrent and feed-forward network models from the data, both to test the consistency of the models with the data and to interpret the observed variability.

Because rates and covariance matrices were measured for many different stimuli and thus provide a large number of constraints, one approach would be to infer as much information as possible about the full connectivity matrices $B$ or $F$ from the data. However, due to the relatively small number of trials for each stimulus, we use a model with few parameters. The set of parameters consists of the network parameters $⟨B⟩$ and $\text{var}(B)$ (and $⟨F⟩$ and $\text{var}(F)$, respectively) as well as the parameters of the input ensemble, $⟨r_{\text{ext}}⟩$, $\text{var}(r_{\text{ext}})$ and $c_{\text{in}}$.

In particular, we want to infer the ratios $\rho_E = \frac{\text{var}(r_{\text{ext}})}{⟨r_{\text{ext}}⟩^2}$ and $\rho = \frac{\text{var}(B)}{⟨B⟩^2}$ (and correspondingly for $F$). Based on these, we can generate surrogate data to test if the observed scaling of average covariances with average rates is more consistent with the recurrent or the feed-forward model, Eq. (28) or (30). The experimental data provides constraints in the form of the population and stimulus averaged rates, their variances and the noise and signal correlation coefficients. In the models, rates are given by Eqs. (7) and (15), respectively. The population averaged mean response thus is

$$⟨r⟩ = N⟨B⟩⟨r_{\text{ext}}⟩ \text{ or } ⟨r⟩ = N⟨F⟩⟨r_{\text{ext}}⟩.$$  \hspace{1cm} (78)

The variance of rates across stimuli, see Eq. (75), is

$$\text{var}(r) = N\text{var}(r_{\text{ext}})(⟨B⟩^2 + \text{var}(B) + (N-1)(B)_{\text{in}}^2)$$  \hspace{1cm} (79)

such that the relative variance of rates

$$\frac{\text{var}(r)}{⟨r⟩^2} = \frac{\text{var}(r_{\text{ext}})1 + \rho + (N-1)c_{\text{in}}}{N}$$  \hspace{1cm} (80)

depends on the input signal to noise ratio $\rho_E = \frac{\text{var}(r_{\text{ext}})}{⟨r_{\text{ext}}⟩^2}$ and the variability of the network elements, $\rho = \frac{\text{var}(B)}{⟨B⟩^2}$, or $\rho = \frac{\text{var}(F)}{⟨F⟩^2}$ respectively.

The estimates for the remaining parameters $\rho$ and $c_{\text{in}}$ are obtained from the measured values of the ratio of covariances to variances, which correspond approximately to the average coefficients of noise and signal correlations,

$$\frac{⟨C_{ij}⟩_{i\neq j}}{⟨C_{ii}⟩_i} \approx c_N,$$  \hspace{1cm} (81)

and

$$\frac{⟨\text{cov}(r_i, r_j)⟩_{i\neq j}}{⟨\text{var}(r_i)⟩_i} \approx c_S,$$  \hspace{1cm} (82)

using Eqs. (70) and (77). Together with Eq. (80), these relations provide the necessary constraints for the network models. Strictly speaking, these equations are valid only for the recurrent network, while for the feed-forward model, Eq. (61) is relevant. In this case $\rho$ is overestimated, which results in a lower bound for $\rho_E$, and thus a conservative estimate of the input variability.

Under additional assumptions, we can also choose the absolute values of the parameters, for example $⟨r_{\text{ext}}⟩$ and $⟨F⟩$, such that mean rates and mean covariances correspond to experimental ones. The mean rates (78) constrain the product of the two parameters

$$⟨r⟩ = N⟨F⟩⟨r_{\text{ext}}⟩.$$  \hspace{1cm} (83)
From Eq. (55) it follows that

\[ \langle C_{ij} \rangle_{i \neq j} = N \langle F \rangle^2 \langle V_{\text{ext}} \rangle, \]  

and we assume that \( V_{\text{ext}} = |r_{\text{ext}}| \) to relate mean input and input variance. The distribution of \( V_{\text{ext}} \) thus is a folded normal distribution, with

\[ \langle |r_{\text{ext}}| \rangle = \sqrt{\text{var}(r_{\text{ext}})} \sqrt{2/\pi} e^{-1/2\rho_x} + \langle r_{\text{ext}} \rangle \left( 1 - 2\Phi(-\sqrt{1/\rho_x}) \right) \]  

(85)

(with the cumulative normal \( \Phi \)), and from this one finds

\[ \langle C_{ij} \rangle = \langle F \rangle \langle r \rangle \left[ \sqrt{\rho_x} e^{-1/2\rho_x} + 1 - 2\Phi(-\sqrt{1/\rho_x}) \right]. \]  

(86)

To set the absolute values of input versus network strength we made assumptions regarding the relation between input strength and input variance across trials. If we measure the strength of the dependence between mean response and mean covariance, Eq. (55) by the ratio intercept/slope of a linear fit, however, the result is independent of the absolute values of \( \langle F \rangle \) and \( \langle r_{\text{ext}} \rangle \), as well as a potential linear factor relating \( V_{\text{ext}} \) and \( |r_{\text{ext}}| \).

### 8.4 Covariances and rates on a population level in regular networks

Here we derive the macroscopic Eqs. (25) and (26) for the responses of the two populations: the average response of population \( I \), \( R_I \equiv \sum_{i \in I} r_i \), and the covariances between responses of populations \( L, K \), \( \Sigma_{KL} \equiv \sum_{k \in K, l \in L} C_{kl} \), in the regular networks defined in Section 3.6 depend only on the coupling between populations, see below.

The two populations are of size \( N \), and weights between connected neurons are of equal strength. By definition, the regularity of the network connectivity implies that for any neuron \( l' \in L \), the sum \( \sum_{k \in K} G_{kl'} \) is identical, and we can define the population coupling matrix, \( \Gamma \), by

\[ \Gamma_{KL} \equiv \sum_{k \in K} G_{kl'} = \frac{1}{N} \sum_{k \in K, l \in L} G_{kl} = n_{KL} g, \]  

(87)

where \( g \) is the weight of the connections. We want to express \( R \) and \( \Sigma \) in terms of \( \Gamma \). On the neuron level, rates, \( r \), and covariances, \( C \), depend on the transfer matrix, \( B = (I - G)^{-1} \). We will define an analogous population transfer matrix \( P \) that depends only on \( \Gamma \) and show that \( R \) and \( \Sigma \) can be written in terms of \( P \). We define

\[ P \equiv (I - \Gamma)^{-1} = I + \Gamma + \Gamma^2 + \cdots = \sum_{m=0}^{\infty} \Gamma^m. \]  

(88)

To relate \( P \) to the microscopic quantities \( r \) and \( C \), we will use that

\[ P_{KL} = \frac{1}{N} \sum_{k \in K, l \in L} B_{kl} = \sum_{k \in K} B_{kl'} \]  

(89)

for any \( l' \in L \), in complete analogy to \( \Gamma \). To see this, we note the corresponding relation for all of the individual terms \( m \)

\[ [\Gamma^m]_{KL} = \frac{1}{N} \sum_{k \in K, l \in L} [G^m]_{kl}. \]  

(90)
As an example consider \( m = 2 \). Using the regularity, we see that

\[
\sum_{k \in K, l \in L} [G^2]_{kl} = \sum_{l \in L, k, i \in I, I} G_{ki} G_{il} = \sum_{l \in L, i \in I, I} \Gamma_{KI} G_{il} = N \sum_{I} \Gamma_{KI} \Gamma_{IL} = [\Gamma^2]_{KL} N. \tag{91}
\]

For the rate of population \( I, R_I \), we then find with (89) that

\[
R_I \equiv \sum_{i \in I} r_i = \sum_{i \in I} \sum_{k \in K} \sum_{k, k'} B_{ikr_{ext, k}} = \sum_{k} \sum_{i \in I, k \in K} B_{ikr_{ext, k}} = \sum_{k} P_{IK} NR_{ext, K}, \tag{92}
\]

because \( r_{ext, k} = R_{ext, K} \) for any \( k \in K \), and thus

\[
R = PNR_{ext} = (I - \Gamma)^{-1} NR_{ext}, \tag{93}
\]

which is Eq. (25). Similarly, for the covariances of the population responses, \( \Sigma \):

\[
\Sigma_{KL} \equiv \sum_{k \in K, l \in L} C_{kl} = \sum_{I} \sum_{k \in K, l \in L, i \in I} B_{il} B_{ij} r_i = \sum_{l \in L} \sum_{i \in I} \sum_{i} P_{KI} B_{il} r_i
\]

\[
= \sum_{l \in L} \sum_{i \in I} P_{KI} P_{li} r_i = \sum_{l} P_{KI} P_{LI} R_I, \tag{94}
\]

which corresponds to Eq. (26),

\[
\Sigma = (I - \Gamma)^{-1} D[R](I - \Gamma)^{-1}, \tag{95}
\]

(with the diagonal matrix \( D[R] \) with \( D[R]_{I,I} = \delta_{I,I} R_I \)).

### 8.4.1 Effect of cross-coupling on correlations on a population level

In this section, we evaluate how noise and signal correlations in the population activity arise from direct and effective coupling. The strength of the cross-coupling between populations, \( \Gamma_c \), determines via the effective cross-coupling, \( P_c \), of the population transfer matrix,

\[
P = (I - \Gamma)^{-1} = \frac{1}{(1 - \Gamma_s)^2 - \Gamma_c^2} \begin{pmatrix} 1 - \Gamma_s & \Gamma_c \\ \Gamma_c & 1 - \Gamma_s \end{pmatrix} \equiv \begin{pmatrix} P_s & P_c \\ P_c & P_s \end{pmatrix}, \tag{97}
\]

the population rates and covariances. The network is unstable if the eigenvalues of \( \Gamma \) are larger than one, which corresponds to the constraint \( \Gamma_s + \Gamma_c < 1 \), and therefore \( P_c < P_s \).

On the population level, the effects of cross-coupling between differently tuned neurons on correlations can be expressed in terms of the variance and the mean of the elements of \( P \), similar as for the random effective network model. To see this, we consider an ensemble of (two-component) input vectors \( R_{ext}(s) = (r_{ext}(s), r_{ext}'(s))^T \) with external input components \( r_{ext}, r_{ext}' \) chosen with mean and variance \( \mu_{r_{ext}}, \sigma_{r_{ext}}^2 \), independently across stimuli \( s \).

The signal covariance matrix of mean responses \( R = PR_{ext} \) across the stimulus ensemble is \( \Sigma^S_{ij} = \text{cov}(R_i, R_j) \) which is

\[
\Sigma^S = \sigma_{r_{ext}}^2 \begin{pmatrix} P_s^2 + P_c^2 & 2P_s P_c \\ 2P_s P_c & P_s^2 + P_c^2 \end{pmatrix} \tag{98}
\]

The average signal correlation for the population activity is

\[
\langle c_s \rangle = \frac{1}{4} \sum_{ij} \frac{\Sigma^S_{ij}}{\sqrt{\Sigma^S_{ii} \Sigma^S_{jj}}} = 1 + \frac{2P_s P_c}{P_s^2 + P_c^2} = \frac{(P_s + P_c)^2}{(P_s + P_c)^2 + (P_s - P_c)^2}, \tag{99}
\]

\[
= \frac{\langle P \rangle^2}{\langle P \rangle^2 + \text{var}(P)} = \frac{\langle P \rangle^2}{\langle P \rangle^2} = \frac{1}{1 + \text{var}(P)/\langle P \rangle^2}
\]
using $\langle P \rangle = (P_s + P_c)/2$ and $2\text{var}(P) = (P_s - (P_s + P_c)/2)^2 + (P_c - (P_s + P_c)/2)^2 = (P_s - P_c)^2/2$. The normalized variance of the elements of $P$ thus determines the strength of signal correlations, and one can interpret $\rho = \text{var}(P)/\langle P \rangle^2$ as a measure for coupling strength.

A similar relation holds for the average noise correlations, if we average first across stimuli, then across neurons: the average of the noise covariance matrix across stimuli is $\langle \Sigma(s) \rangle_s = (PR(s)PT) = P(R)PT$. Because $\langle R_{\text{ext}} \rangle = \langle R_{\text{ext}}' \rangle$, the average rates of both populations are identical, and the average across population pairs can be evaluated just as for the signal correlations.

### 8.4.2 Condition for favorable correlations

Here, we show that strong connections between populations can induce correlations that are beneficial for stimulus discrimination, when compared to shuffled trials. Shuffling correlations is favorable for stimulus discrimination if the covariances within each population are stronger than the ones across populations. To see this, assume that the responses of the excitatory populations for two stimuli are $R(s_1) = (R_0 + \Delta R, R_0)$ and $R(s_2) = (R_0, R_0 + \Delta R)$. We also assume that the population covariance matrix is stimulus independent, $\Sigma(s_1) = \Sigma(s_2) = (\Sigma_{EE}, \Sigma_{E'E'})$, and that $\Sigma_{EE} = \Sigma_{E'E'}$. Calculating the most discriminative direction $w$ (using Eq. $(20)$ on the macroscopic variables) yields an eigenvector of $\Sigma$, $(1, -1)$, and the variance in that direction corresponds to the smaller eigenvalue of $\Sigma$,

$$
\sigma^2_w = \Sigma_{EE} - \Sigma_{E'E'} = \sum_{i\in E} C_{ii} + \sum_{i\neq j\in E} C_{ij} - \sum_{i\in E, k\in E'} C_{ik}.
$$

Covariances can be considered helpful, if this eigenvalue is smaller when compared to a shuffled version where all covariances are set to 0, such that $\sigma^2_w' = \sum_{i\in E} C_{ii}$. This is exactly the case if the sum of covariances across populations, $\sum_{i\in E, k\in E'} C_{ik}$, is larger than the one within a population, $\sum_{i\neq j\in E} C_{ij}$. Although $\Sigma_{EE} \geq \Sigma_{E'E'}$, it is possible that $\langle C_{ij} \rangle_{i\neq j\in E} < \langle C_{ij} \rangle_{i\in E, j\in E'}$, because of the contribution of the variances $C_{ii}$ to $\Sigma_{EE}$.

We want to find a condition for which such a removal of covariances by shuffling decreases noise in the relevant direction. For simplicity, let us assume that $R_{\text{ext}} = R_{\text{ext}}' \equiv R$. Then

$$
\Sigma = R \begin{pmatrix} P_s^2 + P_c^2 & 2P_s P_c \ \\
2P_s P_c & P_s^2 + P_c^2 \end{pmatrix}
$$

(100)

The average pairwise across-covariance is $C_{\text{across}} = \Sigma_{E'E'}/N^2$. Approximately, the average neuron variance differs from the average within-covariance only by the contribution of the rate: $\langle C_{ii} \rangle \approx R/N + C_{\text{within}}$. From this we get for the population variance $\Sigma_{EE} = R(P_s^2 + P_c^2) = N\langle C_{ii} \rangle + N(N-1)C_{\text{within}} = R + N^2C_{\text{within}}$. For favorable correlations, we need $C_{\text{across}} > C_{\text{within}}$,

$$
\frac{2P_s P_c R}{N^2} > \frac{R(P_s^2 + P_c^2) - R}{N^2}
$$

(101)

and since $P_s > P_c$

$$
1 > (P_s - P_c)^2 \iff 1 > P_s - P_c
$$

(102)

Using Eq. (97) for the entries of $P$, this means that

$$
1 > \frac{1 - \Gamma_s - \Gamma_c}{(1 - \Gamma_s)^2 - \Gamma_c^2} \iff \Gamma_s (1 - \Gamma_s) > \Gamma_c (1 - \Gamma_c).
$$

(103)

The function $\Gamma(1 - \Gamma)$ is a parabola through the points $(0,0)$ and $(0,1)$, with its minimum at $(1/2, 1/4)$. We have $|\Gamma_s|, |\Gamma_c| < 1$, so both sides are $< 0$ and $\Gamma_c < 1 - \Gamma_s$. Thus, the condition can be fulfilled if $\Gamma_s < 1/2$ and within-coupling is smaller than cross-coupling, $\Gamma_s < \Gamma_c < 1 - \Gamma_s$. 38
8.4.3 Linear Fisher information

We use the signal-to-noise ratio, $S$, to measure how well discrete pairs of stimuli in a given ensemble can be discriminated, Section 3.5. In a model network, we have access to all possible input dimensions, and we can use an alternative measure that combines these possible stimulus dimensions. We will use this measure to compare the effects of different network scenarios on the representation of general stimuli. If a high-dimensional stimulus varies continuously, the information in the response distribution about the stimulus can be measured by the Fisher information matrix \[16\]. An approximate value, the ‘linear Fisher information’, is obtained by neglecting the information in the stimulus dependence of the covariance matrix, and the entries of the linear Fisher Information matrix are defined by

$$I_{mn} = \frac{\partial}{\partial m} R^T \Sigma^{-1} \frac{\partial}{\partial n} R,$$  \hspace{1cm} (104)

for the population covariance matrix $\Sigma$ and population response $R$. Here, $\partial_m R$ is the derivative of the response vector with respect to stimulus coordinate $m$. Our stimulus dimensions are the coordinates of the input vector $R_{\text{ext}}$ and output in a network model is $R = PNR_{\text{ext}}$. It is easy to see that $\partial_m R$ is given by the $m$th column of $P$, and hence

$$I = P^T \Sigma^{-1} P.$$ \hspace{1cm} (105)

Generally, if the input depends linearly on a one-dimensional stimulus $s$, $R_{\text{ext}}(s) = R_{\text{ext,0}} + s \Delta R_{\text{ext}}$, the information about changes in this direction is $\Delta R_{\text{ext}}^T I \Delta R_{\text{ext}}$. As a measure for the information about stimulus changes in all possible directions, we use the trace of $I$, $\text{Tr}(I) = \sum_i I_{ii}$.

Covariances in a recurrent network with external noise are $\Sigma = P(D[R] + \Sigma_{\text{ext}})P^T$, where $\Sigma_{\text{ext}}$ is the diagonal $2 \times 2$ matrix describing the variance of external input to the two populations. This results in a recurrent linear Fisher Information matrix

$$I^r = (D[R] + \Sigma_{\text{ext}})^{-1}.$$ \hspace{1cm} (106)

This can be compared to a feed-forward scenario with identical transfer matrix, $F = P$, where $\Sigma = P\Sigma_{\text{ext}}P^T + D[R]$, so that

$$I^f = P^T(P\Sigma_{\text{ext}}P^T + D[R])^{-1}P.$$ \hspace{1cm} (107)

References

[1] van Vreeswijk C, Sompolinsky H. Chaotic balanced state in a model of cortical circuits. Neural Comput. 1998;10(6):1321–71.
[2] Shadlen MN, Newsome WT. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. J Neurosci. 1998;18(10):3870–96.
[3] Litwin-Kumar A, Doiron B. Slow dynamics and high variability in balanced cortical networks with clustered connections. Nat Neurosci. 2012;15(11):1498–505. doi:10.1038/nn.3220.
[4] Renart A, la Rocha JD, Bartho P. The asynchronous state in cortical circuits. Science. 2010;327(5965):587–590. doi:10.1126/science.1179850.
[5] Pernice V, Staude B, Cardanobile S, Rotter S. How Structure Determines Correlations in Neuronal Networks. PLoS Comput Biol. 2011;7(5):e1002059. doi:10.1371/journal.pcbi.1002059.
[6] Trousdale J, Hu Y, Shea-Brown E, Josić K. Impact of network structure and cellular response on spike time correlations. PLoS Comput Biol. 2012;8(3):e1002408. doi:10.1371/journal.pcbi.1002408.
[7] Goris RLT, Movshon JA, Simoncelli EP. Partitioning neuronal variability. Nat Neurosci. 2014;17(6):858–65. doi:10.1038/nn.3711.

[8] Scholvinck ML, Saleem AB, Benucci A, Harris KD, Carandini M. Cortical State Determines Global Variability and Correlations in Visual Cortex. J Neurosci. 2015;35(1):170–178. doi:10.1523/JNEUROSCI.4994-13.2015.

[9] Lin IC, Okun M, Carandini M, Harris KD. The Nature of Shared Cortical Variability. Neuron. 2015;87(3):644–656. doi:10.1016/j.neuron.2015.06.035.

[10] Ecker AS, Denfield GH, Bethge M, Tolias AS. On the Structure of Neuronal Population Activity under Fluctuations in Attentional State. J Neurosci. 2016;36(5):1775–1789. doi:10.1523/JNEUROSCI.2044-15.2016.

[11] Doiron B, Litwin-Kumar A, Rosenbaum R, Ocker GK, Josić K. The mechanics of state-dependent neural correlations. Nat Neurosci. 2016;19(3):383–393. doi:10.1038/nn.4242.

[12] Kohn A, Coen-cagli R, Kanitscheider I, Pouget A. Correlations and Neuronal Population Information. Annu Rev Neurosci. 2016;39:237–256. doi:10.1146/annurev-neuro-070815-013851.

[13] Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, et al. Spatiotemporal correlations and visual signalling in a complete neuronal population. Nature. 2008;454(7207):995–9. doi:10.1038/nature07140.

[14] Vidne M, Ahmadian Y, Shlens J, Pillow JW, Kulkarni J, Litke AM, et al. Modeling the impact of common noise inputs on the network activity of retinal ganglion cells. J Comput Neurosci. 2012;33(1):97–121. doi:10.1007/s10827-011-0376-2.

[15] Ecker AS, Berens P, Cotton RJ, Subramaniyan M, Denfield GH, Cadwell CR, et al. State dependence of noise correlations in macaque primary visual cortex. Neuron. 2014;82(1):235–48. doi:10.1016/j.neuron.2014.02.006.

[16] Abbott LF, Dayan P. The effect of correlated variability on the accuracy of a population code. Neural Comput. 1999;11(1):91–101.

[17] Wilke SD, Erich CW. Representational accuracy of stochastic neural populations. Neural Comput. 2002;14(1):155–189. doi:10.1162/089976602753284482.

[18] Shamir M, Sompolinsky H. Nonlinear population codes. Neural Comput. 2004;16(6):1105–1136. doi:10.1162/089976604773371755.

[19] Averbeck BB. Neural correlations, population coding and computation. Nat Rev Neurosci. 2006;7(5):358–366. doi:10.1038/nrn1888.

[20] Tkacik G, Prentice JS, Balasubramanian V, Schneiderman E. Optimal population coding by noisy spiking neurons. P Natl Acad Sci USA. 2010;107(32):14419–24. doi:10.1073/pnas.100496107.

[21] Ecker AS, Berens P, Tolias AS, Bethge M. The effect of noise correlations in populations of diversely tuned neurons. J Neurosci. 2011;31(40):14272–83. doi:10.1523/JNEUROSCI.2539-11.2011.

[22] da Silveira RA, Berry MJ. High-Fidelity Coding with Correlated Neurons. PLoS Comput Biol. 2014;10(11):e1003970. doi:10.1371/journal.pcbi.1003970.

[23] Hu Y, Zylberberg J, Shea-Brown E. The sign rule and beyond: boundary effects, flexibility, and noise correlations in neural population codes. PLoS Comput Biol. 2014;10(2):e1003469. doi:10.1371/journal.pcbi.1003469.

[24] Franke F, Fiscella M, Sevelev M, Roska B, Hierlemann A, da Silveira RA. Structures of Neural Correlation and How They Favor Coding. Neuron. 2016;89(2):409–422. doi:10.1016/j.neuron.2015.12.037.

[25] Zylberberg J, Cafaro J, Turner MH, Shea-Brown E, Rieke F. Direction-Selective Circuits Shape Noise to Ensure a Precise Population Code. Neuron. 2016;89(2):369–383. doi:10.1016/j.neuron.2015.11.019.
[26] Zohary E, Shadlen MN, Newsome WT. Correlated neuronal discharge rate and its implications for psychophysical performance. Nature. 1994;370(6485):140–3. doi:10.1038/370140a0.

[27] Beck J, Bejjanki VR, Pouget A. Insights from a simple expression for linear fisher information in a recurrently connected population of spiking neurons. Neural Comput. 2011;23(6):1484–1502. doi:10.1162/NECO_a_00125.

[28] Renart A, van Rossum MCW. Transmission of population-coded information. Neural Comput. 2012;24(2):391–407. doi:10.1162/NECO_a_00227.

[29] Sériès P, Latham PE, Pouget A. Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations. Nat Neurosci. 2004;7(10):1129–35. doi:10.1038/nn1321.

[30] Moreno-Bote R, Beck J, Kanitscheider I, Pitkow X, Latham P, Pouget A. Information-limiting correlations. Nat Neurosci. 2014;17(10):1410–1417. doi:10.1038/nn.3807.

[31] Kanitscheider I, Coen-Cagli R, Pouget A. Origin of information-limiting noise correlations. P Natl Acad Sci USA. 2015;112(50):E6973–E6982. doi:10.1073/pnas.1508738112.

[32] Bathellier B, Ushakova L, Rumpel S. Discrete neocortical dynamics predict behavioral categorization of sounds. Neuron. 2012;76(2):435–49. doi:10.1016/j.neuron.2012.07.008.

[33] Hawkes AG. Point spectra of some mutually exciting point processes. J R Stat Soc B. 1971;33(3):438–443.

[34] Brody CD. Correlations without synchrony. Neural Comput. 1999;11(7):1537–1551.

[35] Grytskyy D, Tetzlaff T, Diesmann M, Helias M. A unified view on weakly correlated recurrent networks. Front Comput Neurosci. 2013;7(October):131. doi:10.3389/fncom.2013.00131.

[36] Okun M, Steinmetz NA, Cossell L, Iacaruso MF, Ko H, Barthó P, et al. Diverse coupling of neurons to populations in sensory cortex. Nature. 2015;521(7553):511–515. doi:10.1038/nature14273.

[37] Brunel N. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. J Comput Neurosci. 2000;8(3):183–208.

[38] Ponce-Alvarez A, Thiele A, Albright TD, Stoner GR, Deco G. Stimulus-dependent variability and noise correlations in cortical MT neurons. P Natl Acad Sci USA. 2013;110(32):13162–7. doi:10.1073/pnas.1300098110.

[39] Ostojic S. Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. Nat Neurosci. 2014;17(4):594–600. doi:10.1038/nn.3658.

[40] Paninski L. Maximum likelihood estimation of cascade point-process neural encoding models. Network - Comp Neural. 2004;15(4):243–262. doi:10.1088/0954-898X/15/4/002.

[41] Macke JH, Buesing L, Cunningham JP, Yu BM, Shenoy KV, Sahani M. Empirical models of spiking in neuronal populations. Adv Neur In. 2011;24:1350—1358. doi:10.1.1.230.7630.

[42] Rocha J, Doiron B, Reyes A, Shea-brown E, Josić K. Correlation between neural spike trains increases with firing rate. Nature. 2007;448(7155):802–6. doi:10.1038/nature06028.

[43] Lučzak A, Barthó P, Harris KD. Gating of sensory input by spontaneous cortical activity. J Neurosci. 2013;33(4):1684–95. doi:10.1523/JNEUROSCI.2928-12.2013.

[44] Cohen MR, Kohn A. Measuring and interpreting neuronal correlations. Nat Neurosci. 2011;14(7):811–9. doi:10.1038/nn.2842.

[45] Rothschild G, Nelken I, Mizrahi A. Functional organization and population dynamics in the mouse primary auditory cortex. Nat Neurosci. 2010;13(3):353–60. doi:10.1038/nn.2484.