Decorrelation of Neural-Network Activity by Inhibitory Feedback

Tom Tetzlaff[1,2,9]*, Moritz Helias[1,3,9], Gaute T. Einevoll[2], Markus Diesmann[1,3,4]

1 Institute of Neuroscience and Medicine (INM-6), Computational and Systems Neuroscience, Research Center Jülich, Jülich, Germany, 2 CIGENE, Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, Ås, Norway, 3 RIKEN Brain Science Institute and Brain and Neural Systems Team, RIKEN Computational Science Research Program, Wako, Japan, 4 Medical Faculty, RWTH Aachen University, Aachen, Germany

Abstract

Correlations in spike-train ensembles can seriously impair the encoding of information by their spatio-temporal structure. An inevitable source of correlation in finite neural networks is common presynaptic input to pairs of neurons. Recent studies demonstrate that spike correlations in recurrent neural networks are considerably smaller than expected based on the amount of shared presynaptic input. Here, we explain this observation by means of a linear network model and simulations of networks of leaky integrate-and-fire neurons. We show that inhibitory feedback efficiently suppresses pairwise correlations and, hence, population-rate fluctuations, thereby assigning inhibitory neurons the new role of active decorrelation. We quantify this decorrelation by comparing the responses of the intact recurrent network (feedback system) and systems where the statistics of the feedback channel is perturbed (feedforward system). Manipulations of the feedback statistics can lead to a significant increase in the power and coherence of the population response. In particular, neglecting correlations within the ensemble of feedback channels or between the external stimulus and the feedback amplifies population-rate fluctuations by orders of magnitude. The fluctuation suppression in homogeneous inhibitory networks is explained by a negative feedback loop in the one-dimensional dynamics of the compound activity. Similarly, a change of coordinates exposes an effective negative feedback loop in the compound dynamics of stable excitatory-inhibitory networks. The suppression of input correlations in finite networks is explained by the population averaged correlations in the linear network model: In purely inhibitory networks, shared-input correlations are canceled by negative spike-train correlations. In excitatory-inhibitory networks, spike-train correlations are typically positive. Here, the suppression of input correlations is not a result of the mere existence of correlations between excitatory (E) and inhibitory (I) neurons, but a consequence of a particular structure of correlations among the three possible pairings (EE, EI, II).

Introduction

Neurons generate signals by weighting and combining input spike trains from presynaptic neuron populations. The number of possible signals which can be read out this way from a given spike-train ensemble is maximal if these spike trains span an orthogonal basis, i.e. if they are uncorrelated [1]. If they are correlated, the amount of information which can be encoded in the spatio-temporal structure of these spike trains is limited. In addition, correlations impair the ability of readout neurons to decode information reliably in the presence of noise. This is often discussed in the context of *rate coding*: for N uncorrelated spike trains, the signal-to-noise ratio of the compound spike-count signal can be enhanced by increasing the population size N. In the presence of correlations, however, the signal-to-noise ratio is bounded [2,3]. The same reasoning holds for any other linear combination of spike trains, also for those where exact spike timing matters (for example for the coding scheme presented in [4]). Thus, the robustness of neuronal responses against noise critically depends on the level of correlated activity within the presynaptic neuron population.

Several studies suggested that correlated neural activity could be beneficial for information processing: Spike-train correlations can modulate the gain of postsynaptic neurons and thereby constitute a gating mechanism (for a review, see [4]). Coherent spiking activity might serve as a means to bind elementary representations into more complex objects [5,6]. Information represented by correlated firing can be reliably sustained and propagated through feedforward subnetworks ('synfire chains'; [7,8]). Whether correlated firing has to be considered favorable or not largely depends on the underlying hypothesis, the type of correlation (e.g. the time scale or the affected frequency band) or which subpopulations of neurons are involved. Most ideas suggesting a functional benefit of correlated activity rely on the existence of an asynchronous 'ground state'. Spontaneously emerging correlations, i.e. correlations which are not triggered by internal or external events, would impose a serious challenge to many of these hypotheses. Functionally relevant synfire activity, for example, cannot be

Author Summary

The spatio-temporal activity pattern generated by a recurrent neuronal network can provide a rich dynamical basis which allows readout neurons to generate a variety of responses by tuning the synaptic weights of their inputs. The repertoire of possible responses and the response reliability become maximal if the spike trains of individual neurons are uncorrelated. Spike-train correlations in cortical networks can indeed be very small, even for neighboring neurons. This seems to be at odds with the finding that neighboring neurons receive a considerable fraction of inputs from identical presynaptic sources constituting an inevitable source of correlation. In this article, we show that inhibitory feedback, abundant in biological neuronal networks, actively suppresses correlations. The mechanism is generic: It does not depend on the details of the network nodes and decorrelates networks composed of excitatory and inhibitory neurons as well as purely inhibitory networks. For the case of the leaky integrate-and-fire model, we derive the correlation structure analytically. The new toolbox of formal linearization and a basis transformation exposing the feedback component is applicable to a range of biological systems. We confirm our analytical results by direct simulations.

guaranteed in the presence of correlated background input from the embedding network [9]. It is therefore–from several perspectives–important to understand the origin of uncorrelated activity in neural networks.

It has recently been shown that spike trains of neighboring cortical neurons can indeed be uncorrelated [10]. Similar results have been obtained in several theoretical studies [11–17]. From an anatomical point of view, this observation is puzzling: in general, neurons in finite networks share a certain fraction of their presynaptic sources. In particular for neighboring neurons, the overlap between presynaptic neuron populations is expected to be substantial. This feedforward picture suggests that such presynaptic overlap gives rise to correlated synaptic input and, in turn, to correlated response spike trains.

A number of theoretical studies showed that shared-input correlations are only weakly transferred to the output side as a consequence of the nonlinearity of the spike-generation dynamics [15,18–21]. Unreliable spike transmission due to synaptic failure can further suppress the correlation gain [22]. In [9], we demonstrated that spike-train correlations in finite-size recurrent networks are even smaller than predicted by the low correlation gain of pairs of neurons with nonlinear spike-generation dynamics. We concluded that this suppression of correlations must be a result of the recurrent network dynamics. In this article, we compare correlations observed in feedforward networks to correlations measured in systems with an intact feedback loop. We refer to the reduction of correlations in the presence of feedback as "decorrelation". Different mechanisms underlying such a dynamical decorrelation have been suggested in the recent past. Asynchronous states in recurrent neural networks are often attributed to chaotic dynamics [23,24]. In fact, networks of nonlinear units with random connectivity and balanced excitation and inhibition typically exhibit chaos [11,25]. The high sensitivity to noise may however question the functional relevance of such systems ([26,27]; cf., however, [28]). [29] and [27] demonstrated that asynchronous irregular firing can also emerge in networks with stable dynamics. Employing an analytical framework of correlations in recurrent networks of binary neurons [30], the

balance of excitation and inhibition has recently been proposed as another decorrelation mechanism [17]: In large networks, fluctuations of excitation and inhibition are in phase. Positive correlations between excitatory and inhibitory input spike trains lead to a negative component in the net input correlation which can compensate positive correlations caused by shared input.

In the present study, we demonstrate that dynamical decorrelation is a fundamental phenomenon in recurrent systems with negative feedback. We show that negative feedback alone is sufficient to efficiently suppress correlations. Even in purely inhibitory networks, shared-input correlations are compensated by feedback. A balance of excitation and inhibition is thus not required. The underlying mechanism can be understood by means of a simple linear model. This simplifies the theory and helps to gain intuition, but it also confirms that low correlations can emerge in recurrent networks with stable, non-chaotic dynamics.

The suppression of pairwise spike-train correlations by inhibitory feedback is reflected in a reduction of population-rate fluctuations. The main effect described in this article can therefore be understood by studying the dynamics of the macroscopic population activity. This approach leads to a simple mathematical description and emphasizes that the described decorrelation mechanism is a general phenomenon which may occur not only in neural networks but also in other (biological) systems with inhibitory feedback. In **"Results: Suppression of population-rate fluctuations in LIF networks"**, we first illustrate the decorrelation effect for random networks of N leaky integrate-and-fire (LIF) neurons with inhibitory or excitatory-inhibitory coupling. By means of simulations, we show that low-frequency spike-train correlations, and, hence, population-rate fluctuations are substantially smaller than expected given the amount of shared input. As shown in the subsequent section, the **"Suppression of population-activity fluctuations by negative feedback"** can readily be understood in the framework of a simple one-dimensional linear model with negative feedback. In **"Results: Population-activity fluctuations in excitatory-inhibitory networks"**, we extend this to a two-population system with excitatory-inhibitory coupling. Here, a simple coordinate transform exposes the inherent negative feedback loop as the underlying cause of the fluctuation suppression in inhibition-dominated networks. The population-rate models of the inhibitory and the excitatory-inhibitory network are sufficient to understand the basic mechanism underlying the decorrelation. They do, however, not describe how feedback in cortical networks affects the detailed structure of pairwise correlations. In **"Results: Population averaged correlations in cortical networks"**, we therefore compute self-consistent population averaged correlations for a random network of N linear excitatory and inhibitory neurons. By determining the parameters of the linear network analytically from the LIF model, we show that the predictions of the linear model are—for a wide and realistic range of parameters—in excellent agreement with the results of the LIF network model. In **"Results: Effect of feedback manipulations"**, we demonstrate that the active decorrelation in random LIF networks relies on the feedback of the (sub)population averaged activity but not on the precise microscopic structure of the feedback signal. In the **"Discussion"**, we put the consequences of this work into a broader context and point out limitations and possible extensions of the presented theory. The **"Methods"** contain details on the LIF network model, the derivation of the linear model from the LIF dynamics and the derivation of population-rate spectra and population averaged correlations in the framework of the linear model. This section is meant as a supplement; the basic ideas and the main results can be extracted from the **"Results"**.

Results

In a recurrent neural network of size N, each neuron $i \in [1,N]$ receives in general inputs from two different types of sources: External inputs $\xi_i(t)$ representing the sum of afferents from other brain areas, and local inputs resulting from the recurrent connectivity within the network. Depending on their origin, external inputs ξ_i and ξ_j to different neurons i and j can be correlated or not. Throughout this manuscript, we ignore correlations between these external sources, thereby ensuring that correlations within the network activity arise from the local connectivity alone and are not imposed by external inputs [17]. The local inputs feed the network's spiking activity $s(t) = (s_1(t), \ldots, s_N(t))^{\mathrm{T}}$ back to the network (we refer to spike train $s_i(t)$, the i th component of the column vector $s(t)$ [the superscript "T" denotes the transpose], as a sum over delta-functions centered at the spike times t_i^k: $s_i(t) = \sum_k \delta(t - t_i^k)$; the abstract quantity 'spike train' can be considered as being derived from the observable quantity 'spike count' $n_i^{\Delta t}(t)$, the number of spikes occurring in the time interval $[t, t + \Delta t]$, by taking the limit $\Delta t \to 0$: $s_i(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} n_i^{\Delta t}(t)$). The structure and weighting of this feedback can be described by the network's connectivity matrix J (see Fig. 1 A). In a finite network, the local connectivity typically gives rise to overlapping presynaptic populations: in a random (Erdös-Rényi) network with connection probability ϵ, for example, each pair of postsynaptic neurons shares, on average, $\epsilon^2 N$ presynaptic sources. For a network size of, say, $N = 10^4$ and a connection probability $\epsilon = 0.1$, this corresponds to a fairly large number of 100 identical inputs. For other network structures, the amount of shared input may be smaller or larger. Due to this presynaptic overlap, each pair of neurons receives, to some extent, correlated input (even if the external inputs are uncorrelated). One might therefore expect that the network responses $s_1(t), \ldots, s_N(t)$ are correlated as well. In this article, we show that, in the presence of negative feedback, the effect of shared input caused by the structure of the network is compensated by its recurrent dynamics.

Suppression of population-rate fluctuations in LIF networks

To illustrate the effect of shared input and its suppression by the recurrent dynamics, we compare the spike response $s(t) = (s_1(t), \ldots, s_N(t))^{\mathrm{T}}$ of a recurrent random network (*feedback scenario*; Fig. 1 A,C,E) of N LIF neurons to the case where the feedback is cut and replaced by a spike-train ensemble $q(t) = (q_1(t), \ldots, q_N(t))^{\mathrm{T}}$, modeled by N independent realizations of a stationary Poisson point process (*feedforward scenario*; Fig. 1 B,D,F). The rate of this Poisson process is identical to the time and population averaged firing rate in the intact recurrent system. In both the feedback and the feedforward case, the (local) presynaptic spike trains are fed to the postsynaptic population according to the same connectivity matrix J. Therefore, not only the in-degrees and the synaptic weights but also the shared-input statistics are exactly identical.

For realistic size N and connectivity ϵ, asynchronous states of random neural networks [12,31] exhibit spike-train correlations which are small but not zero (compare raster displays in Fig. 1 C and D; see also [15]). Although the presynaptic spike trains are, by construction, independent in the feedforward case (Fig. 1 D), the resulting response correlations, and, hence, the population-rate fluctuations, are substantially stronger than those observed in the feedback scenario (compare Fig. 1 F and E). In other words: A theory which is exclusively based on the amount of shared input but neglects the details of the presynaptic spike-train statistics can significantly overestimate correlations and population-rate fluctuations in recurrent neural networks.

The same effect can be observed in LIF networks with both purely inhibitory and mixed excitatory-inhibitory coupling (Fig. 2). To demonstrate this quantitatively, we focus on the fluctuations of the population averaged activity $s(t) = N^{-1} \sum_{i=1}^{N} s_i(t)$. Its power-spectrum (or auto-correlation, in the time domain)

$$
\begin{aligned}
C_{SS}(\omega) \quad &= |S(\omega)|^2 = |\mathfrak{F}[s(t)](\omega)|^2 \\
&= \frac{1}{N^2} \left[\sum_{i=1}^{N} A_i(\omega) + \sum_{i=1, j \neq i}^{N} C_{ij}(\omega) \right]
\end{aligned}
\tag{1}
$$

is determined both by the power-spectra (auto-correlations) $A_i(\omega) = |S_i(\omega)|^2$ of the individual spike trains and the cross-spectra (cross-correlations) $C_{ij}(\omega) = S_i(\omega) S_j(\omega)^*$ $(i \neq j)$ of pairs of spike trains (throughout the article, we use capital letters to represent quantities in frequency [Fourier] space; $S_k(\omega) = \mathfrak{F}[s_k(t)](\omega) = \int \mathrm{d}t\, s_k(t) e^{-i\omega t}$ represents the Fourier transform of the spike train $s_k(t)$). We observe that the spike-train power-spectra $A_i(\omega)$ (and auto-correlations) are barely distinguishable in the feedback and in the feedforward case (not shown here; the main features of the spike-train auto-correlation are determined by the average single-neuron firing rate and the refractory mechanism; both are identical in the feedback and the feedforward scenario). The differences in the population-rate spectra $C_{SS}(\omega)$ are therefore essentially due to differences in the spike-train cross-spectra $C_{ij}(\omega)$. In other words, the fluctuations in the population activity serve as a measure of pairwise spike-train correlations [32]: small (large) population averaged spike-train correlations are accompanied by small (large) fluctuations in the population rate (see lower panels in Fig. 1 C–F). The power-spectra $C_{SS}(\omega)$ of the population averaged activity reveal a feedback-induced suppression of the population-rate variance at low frequencies up to several tens of Hertz. For the examples shown in Fig. 2, this suppression spans more than three orders of magnitude for the inhibitory and more than one order of magnitude for the excitatory-inhibitory network.

The suppression of low-frequency fluctuations does not critically depend on the details of the network model. As shown in Fig. 2, it can, for example, be observed for both networks with zero rise-time synapses (δ-shaped synaptic currents) and short delays and for networks with delayed low-pass filtering synapses (α-shaped synaptic currents). In the latter case, the suppression of fluctuations is slightly more restricted to lower frequencies ($< 10\,\mathrm{Hz}$). Here, the fluctuation suppression is however similarly pronounced as in networks with instantaneous synapses.

In Fig. 2 C,D, the power-spectra of the population activity converge to the mean firing rate at high frequencies. This indicates that the spike trains are uncorrelated on short time scales. For instantaneous δ-synapses, neurons exhibit an immediate response to excitatory input spikes [33,34]. This fast response causes spike-train correlations on short time scales. Hence, the compound power at high frequencies is increased. In a recurrent system, this effect is amplified by reverberating simultaneous excitatory spikes. Therefore, the high-frequency power of the compound activity is larger in the feedback case (Fig. 2 B). Note that this high-frequency effect is absent in networks with more realistic low-pass filtering synapses (Fig. 2 C,D) and in purely inhibitory networks (Fig. 2 A).

Synaptic delays and slow synapses can promote oscillatory modes in certain frequency bands [12,31], thereby leading to peaks in the population-rate spectra in the feedback scenario which exceed the power in the feedforward case (see peaks at $\sim 25\,\mathrm{Hz}$ in Fig. 2 C,D). Note that, in the feedforward case, the
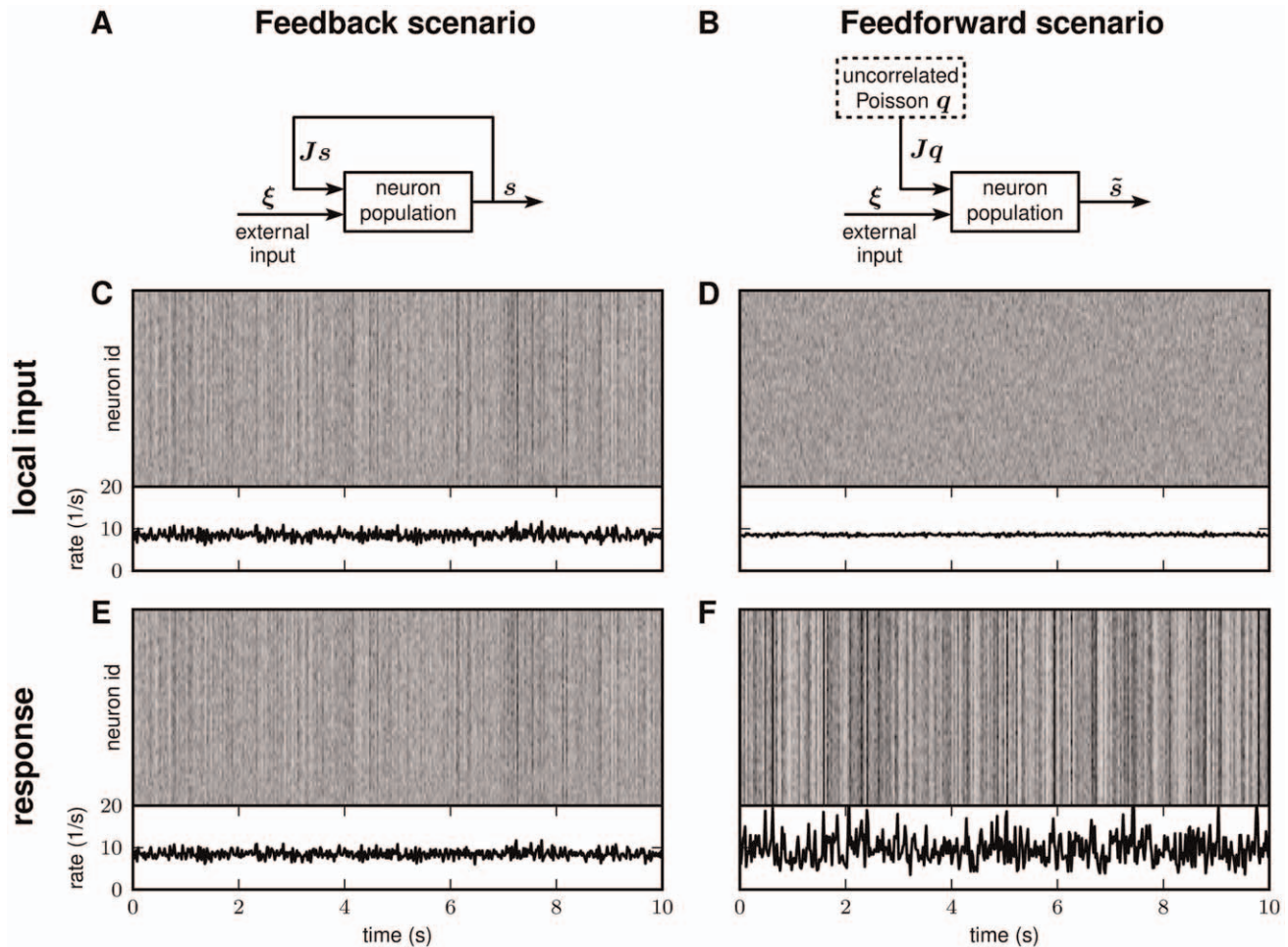
Figure 1. Spiking activity in excitatory-inhibitory LIF networks with intact (left column; *feedback scenario*) **and opened feedback loop (right column;** *feedforward scenario*). **A,B**: Network sketches for the feedback (A) and feedforward scenario (B). **C,D**: Spiking activity (top panels) and population averaged firing rate (bottom panels) of the local presynaptic populations. **E,F**: Response spiking activity (top panels) and population averaged response rate (bottom panels). In the top panels of C–F, each pixel depicts the number of spikes (gray coded) of a subpopulation of 250 neurons in a 10 ms time interval. In both the feedback and the feedforward scenario, the neuron population $\{1,\ldots,N\}$ is driven by the same realization $x(t)=(\xi_1(t),\ldots,\xi_N(t))^{\mathrm{T}}$ of an uncorrelated white-noise ensemble; local input is fed to the population through the same connectivity matrix J. The in-degrees, the synaptic weights and the shared-input statistics are thus exactly identical in the two scenarios. In the feedback case (A), local presynaptic spike-trains are provided by the network's response $s(t)=(s_1(t),\ldots,s_N(t))^{\mathrm{T}}$, i.e. the pre- (C) and postsynaptic spike-train ensembles (E) are identical. In the feedforward scenario (B), the local presynaptic spike-train population is replaced by an ensemble of N independent realizations $q(t)=(q_1(t),\ldots,q_N(t))$ of a Poisson point process (D). Its rate is identical to the time- and population-averaged firing rate in the feedback case. See Table 1 and Table 2 for details on network models and parameters.
doi:10.1371/journal.pcbi.1002596.g001

local input was replaced by a stationary Poisson process, whereas in the recurrent network (feedback case) the presynaptic spike trains exhibit oscillatory modes. By replacing the feedback by an inhomogeneous Poisson process with a time dependent intensity which is identical to the population rate in the recurrent network, we found that these oscillatory modes are neither suppressed nor amplified by the recurrent dynamics, i.e. the peaks in the resulting power-spectra have the same amplitude in the feedback and in the feedforward case (data not shown here). At low frequencies, however, the results are identical to those obtained by replacing the feedback by a homogeneous Poisson process (i.e. to those shown in Fig. 2; see **"Results: Effect of feedback manipulations"**). In the present study, we mainly focus on these low-frequency effects.

The observation that the suppression of low-frequency fluctuations is particularly pronounced in networks with purely

inhibitory coupling indicates that inhibitory feedback may play a key role for the underlying mechanism. In the following subsection, we demonstrate by means of a one-dimensional linear population model that, indeed, negative feedback alone leads to an efficient fluctuation suppression.

Suppression of population-activity fluctuations by negative feedback

Average pairwise correlations can be extracted from the spectrum (1) of the compound activity, provided the single spike-train statistics (auto-correlations) is known (see previous section). As the single spike-train statistics is identical in the feedback and in the feedforward scenario, the mechanism underlying the decorrelation in recurrent networks can be understood by studying the dynamics of the population averaged activity. In this and in the next subsection, we consider the linearized dynamics of random
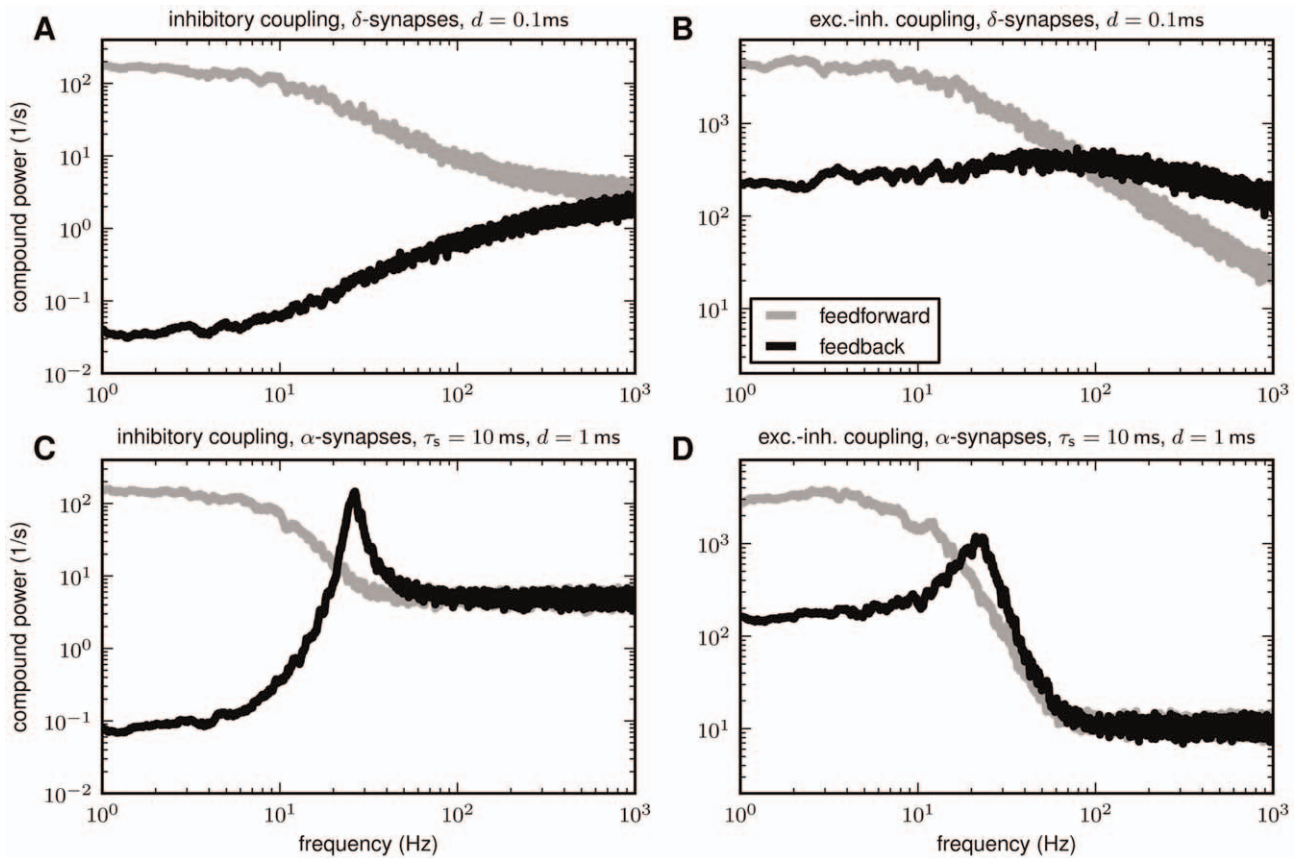
Figure 2. Suppression of low-frequency fluctuations in recurrent LIF networks with purely inhibitory (A, C) and mixed excitatory-inhibitory coupling (B, D) for instantaneous synapses with delay $d = 0.1$ ms (A, B) and low-pass synapses with $d = 1$ ms (C, D). Power-spectra NC_{SS} of population rates $s(t)$ for the feedback (black) and the feedforward case (gray; cf. Fig. 1). See Table 1 and Table 2 for details on network models and parameters. In C and D, local synaptic inputs are modeled as currents $I_i(t) = \sum_j J_{ij} \sum_l \mathrm{psc}(t - t_l^j - d)$ with α-function shaped kernel $\mathrm{psc}(t) = et\tau_s^{-1} \exp(-t/\tau_s)\Theta(t)$ with time constant $\tau_s = 10$ ms ($\Theta(\cdot)$ denotes Heaviside function). (Excitatory) Synaptic weights are set to $J = 1$ pA (see Table 1 for details). Simulation time $T = 100$ s. Single-trial spectra smoothed by moving average (frame size 1 Hz).
doi:10.1371/journal.pcbi.1002596.g002

networks composed of homogeneous subpopulations of LIF neurons. The high-dimensional dynamics of such systems can be reduced to low-dimensional models describing the dynamics of the compound activity (for details, see **"Methods: Linearized network model"**). Note that this reduction is exact for networks with homogeneous out-degree (number of outgoing connections). For the networks studied here (random networks with homogeneous in-degree), it serves as a sufficient approximation (in a network of size N where each connection is randomly and independently realized with probability ϵ [Erdös-Rényi graph], the [binomial] in- and out-degree distributions become very narrow for large N [relative to the mean in/out-degree]; both in- and out-degree are therefore approximately constant across the population of neurons). In this subsection, we first study networks with purely inhibitory coupling. In **"Results: Population-activity fluctu-ations in excitatory-inhibitory networks"**, we investigate the effect of mixed excitatory-inhibitory connectivity.

Consider a random network of N identical neurons with connection probability ϵ. Each neuron $i = 1, \ldots, N$ receives $K = \epsilon N$ randomly chosen inputs from the local network with synaptic weights $-J$. In addition, the neurons are driven by external uncorrelated Gaussian white noise $\xi_i(t)$ with amplitude η, i.e. $E_t[\xi_i(t)] = 0$ and $E_t[\xi_i(t)]\xi_j(t+\tau) = \delta_{ij}\eta^2\delta(\tau)$. For small input fluctuations, the network dynamics can be linearized. This

linearization is based on the averaged response of a single neuron to an incoming spike and describes the activity of an individual neuron i by an abstract fluctuating quantity $r_i(t)$ which is defined such that within the linear approximation its auto- and cross-correlations fulfill the same linearized equation as the spiking model in the low-frequency limit. Consequently, also the low-frequency fluctuations of the population spike rate are captured correctly by the reduced model up to linear order. This approach is equivalent to the treatment of finite-size fluctuations in spiking networks (see, e.g., [31]). For details, see **"Methods: Linearized network model"**. For large N, the population averaged activity $r(t) = E_i[r_i(t)] = N^{-1} \sum_{i=1}^{N} r_i(t)$ can hence be described by a one-dimensional linear system

$$r(t) = ([-\bar{w}r + x] * h)(t) \qquad (2)$$

with linear kernel $h(t)$, effective coupling strength $\bar{w} = Kw$ and the population averaged noise $x(t) = E_i[x_i](t)$ (see **"Methods: Linearized network model"** and Fig. 3 B). The coupling strength $\bar{w}$ represents the integrated linear response of the neuron population to a small perturbation in the input rate of a single presynaptic neuron. For a population of LIF neurons, its relation to the synaptic weight J (PSP amplitude) is derived in **"Methods: Linearized network model"** and **"Methods: Response**

kernel of the LIF model''. The normalized kernel $h(t)$ (with $\int_0^\infty dt\, h(t) = 1$) captures the time course of the linear response. It is determined by the single-neuron properties (e.g. the spike-initiation dynamics [35,36]), the properties of the synapses (e.g. synaptic weights and time constants [37,38]) and the properties of the input (e.g. excitatory vs. inhibitory input [39]). For many real and model neurons, the linear population-rate response exhibits low-pass characteristics [13,34–46]. For illustration (Fig. 3), we consider a 1st-order low-pass filter, i.e. an exponential impulse response $h(t) = \tau^{-1} \exp(-t/\tau)\Theta(t)$ with time constant τ (cutoff frequency $f_c = (2\pi\tau)^{-1}$; see Fig. 3 A, light gray curve in E). The results of our analysis are however independent of the choice of the kernel $h(t)$. The auto-correlation $E_t[x(t)x(t+\tau)] = \bar{\rho}^2 \delta(\tau)$ of the external noise is parametrized by the effective noise amplitude $\bar{\rho} = \rho/\sqrt{N}$.

Given the simplified description (2), the suppression of response fluctuations by negative feedback can be understood intuitively: Consider first the case where the neurons in the local network are unconnected (Fig. 3 A; no feedback, $\bar{w} = 0$). Here, the response $r(t)$ (Fig. 3 A_3) is simply a low-pass filtered version of the external input $x(t)$ (Fig. 3 A_1), resulting in an exponentially decaying response auto-correlation (Fig. 3 D; light gray curve) and a drop in the response power-spectrum at the cutoff frequency f_c (Fig. 3 E). At low frequencies, $r(t)$ and $x(t)$ are in phase; they are correlated. In the presence of negative feedback (Fig. 3 B), the local input $-\bar{w}r(t)$ (Fig. 3 B_2) and the low-frequency components of the external input $x(t)$ (Fig. 3 B_1) are anticorrelated. They partly cancel out, thereby reducing the response fluctuations $r(t)$ (Fig. 3 B_3). The auto-correlation function and the power-spectrum are suppressed (Fig. 3 D,E; black curves). Due to the low-pass characteristics of the system, mainly the low-frequency components of the external drive
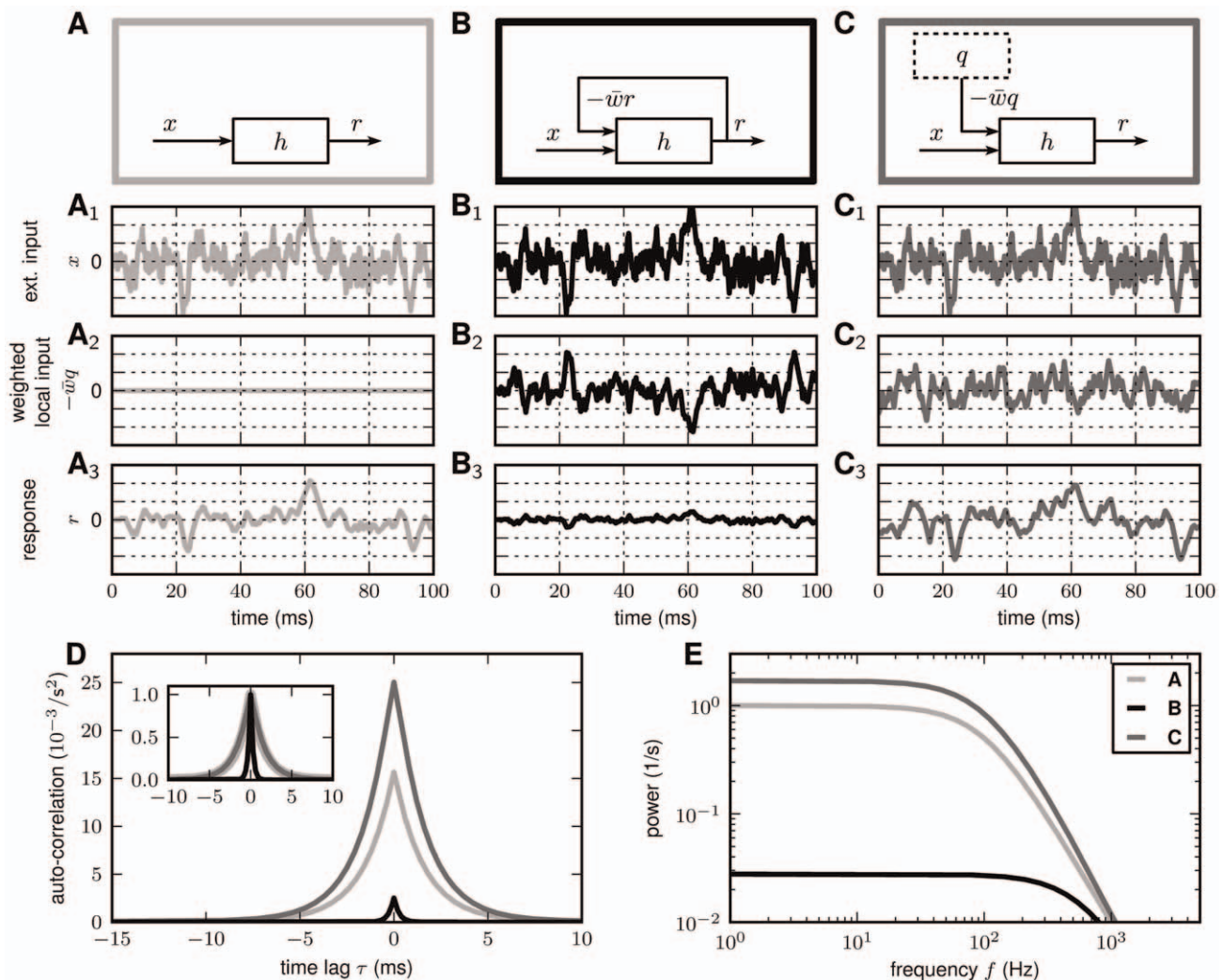

Figure 3. Partial canceling of fluctuations in a linear system by inhibitory feedback. Response $r(t)$ of a linear system with impulse response $h(t)$ (1st-order low-pass, cutoff frequency $100\,\text{Hz}$) to Gaussian white noise input $x(t)$ with amplitude $\bar{\rho}=1$ for three local-input scenarios. **A** (light gray): No feedback (local input $q(t)=0$). **B** (black): Negative feedback ($q(t)=r(t)$) with strength $\bar{w}=5$. The fluctuations of the weighted local input $-\bar{w}q(t)$ (B_2) are anticorrelated to the external drive $x(t)$ (B_1). **C** (dark gray): Feedback in B is replaced by uncorrelated feedforward input $q(t)$ with the same auto-statistics as the response $r(t)$ in B_3. The local input $q(t)=F^{-1}\left[|R(\omega)|e^{i\xi(\omega)}\right](t)$ is constructed by assigning a random phase $\xi(\omega)$ to each Fourier component $R(\omega)=F[r(t)](\omega)$ of the response in B_3. Fluctuations in C_2 and C_1 are uncorrelated. **A**, **B**, **C**: Network sketches. A_1, B_1, C_1: External input $x(t)$. A_2, B_2, C_2: Weighted local input $-\bar{w}q(t)$. A_3, B_3, C_3: Responses $r(t)$. **D**, **E**: Response auto-correlation functions (D) and power-spectra (E) for the three cases shown in A,B,C (same gray coding as in A,B,C; inset in D: normalized auto-correlations).
doi:10.1371/journal.pcbi.1002596.g003

$x(t)$ are transferred to the output side and, in turn, become available for the feedback signal. Therefore, the canceling of input fluctuations and the resulting suppression of response fluctuations are most efficient at low frequencies. Consequently, the auto-correlation function is sharpened (see inset in Fig. 3 D). The cutoff frequency of the system is increased (Fig. 3 E; black curve). This effect of negative feedback is very general and well known in the engineering literature. It is employed in the design of technical devices, like, e.g., amplifiers [47]. As the zero-frequency power is identical to the integrated auto-correlation function, the suppression of low-frequency fluctuations is accompanied by a reduction in the auto-correlation area (Fig. 3 D; black curve). Note that the suppression of fluctuations in the feedback case is not merely a result of the additional inhibitory noise source provided by the local input, but follows from the precise temporal alignment of the local and the external input. To illustrate this, let's consider the case where the feedback channel is replaced by a feedforward input $q(t)$ (Fig. 3 C) which has the same auto-statistics as the response $r(t)$ in the feedback case (Fig. 3 B_3) but is uncorrelated to the external drive $x(t)$. In this case, external input fluctuations (Fig. 3 C_1) are not canceled by the local input $-\bar{w}q(t)$ (Fig. 3 C_2). Instead, the local feedforward input acts as an additional noise source which leads to an increase in the response fluctuations (Fig. 3 C_3). The response auto-correlation and power-spectrum (Fig. 3 D,E; dark gray curves) are increased. Compared to the unconnected case (Fig. 3 E; light gray curve), the cutoff frequency remains unchanged.

The feedback induced suppression of response fluctuations can be quantified by comparing the response power-spectra

$$C_{RR}(\omega) = E_x\left[|R(\omega)|^2\right] = \frac{\bar{\rho}^2|H(\omega)|^2}{|1+\bar{w}H(\omega)|^2} \qquad (3)$$

and

$$C_{\tilde{R}\tilde{R}}(\omega) = E_x\left[|\tilde{R}(\omega)|^2\right] = |H(\omega)|^2(\bar{w}^2 C_{RR}(\omega) + \bar{\rho}^2) \qquad (4)$$

in the feedback (Fig. 3 B) and the feedforward case (Fig. 3 C), respectively (see **"Methods: Population-activity spectrum of the linear inhibitory network"**). Here, $R(\omega)$ and $\tilde{R}(\omega)$ denote the Fourier transforms of the response fluctuations in the feedback and the feedforward scenario, respectively, $H(\omega)$ the transfer function (Fourier transform of the filter kernel $h(t)$) of the neuron population, and $E_x[]$ the average across noise realizations. We use the power ratio

$$\alpha(\omega) = \frac{C_{RR}(\omega)}{C_{\tilde{R}\tilde{R}}(\omega)} = \frac{1}{\bar{w}^2|H(\omega)|^2 + |1+\bar{w}H(\omega)|^2} \qquad (5)$$

as a measure of the relative fluctuation suppression caused by feedback. For low frequencies ($\omega \to 0$) and strong effective coupling $|\bar{w}| = |Kw| \gg 1$, the power ratio (5) decays as $\bar{w}^{-2}$ (see Fig. 4 A): the suppression of population-rate fluctuations is promoted by strong negative feedback. In line with the observations in **"Results: Suppression of population-rate fluctuations in LIF networks"**, this suppression is restricted to low frequencies; for high frequencies ($\omega \to \infty$, i.e. $H(\omega) \to 0$), the power ratio $\alpha(\omega)$ approaches 1. Note that the power ratio (5) is independent of the amplitude $\bar{\rho}$ of the population averaged external input $x(t)$. Therefore, even if we dropped the assumption of the external inputs $x_i(t)$ being uncorrelated, i.e. if $E_t[x_i(t)x_j(t+\tau)] \neq 0$ for $i \neq j$, the power ratio (5) remained the same. For correlated

external input, the power $\bar{\rho}$ of the population average $x(t)$ is different from $\rho/\sqrt{N}$. The suppression factor $\alpha(\omega)$, however, is not affected by this. Moreover, it is straightforward to show that the power ratio (5) is, in fact, independent of the shape of the external-noise spectrum $C_{XX}(\omega) = E_x[|X(\omega)|^2]$. The same result (5) is obtained for any type of external input (e.g. colored noise or oscillating inputs).

For low frequencies, the transfer function $H(\omega)$ approaches unity ($\lim_{\omega \to 0} H(\omega) = 1$); the exact shape of the kernel $h(t)$ becomes irrelevant. In particular, the cutoff frequency (or time constant) of a low-pass kernel has no effect on the zero-frequency power (integral correlation) and the zero-frequency power ratio $\alpha(0)$ (Fig. 4). Therefore, the suppression of low-frequency fluctuations does not critically depend on the exact choice of the neuron, synapse or input model. The same reasoning applies to synaptic delays: Replacing the kernel $h(t)$ by a delayed kernel $h(t-d)$ leads to an additional phase factor $e^{-i\omega d}$ in the transfer function $H(\omega)$. For sufficiently small frequencies (long time scales), this factor can be neglected ($\lim_{\omega \to 0} e^{-i\omega d} = 1$).

For networks with purely inhibitory feedback, the absolute power (3) of the population rate decreases monotonously with increasing coupling strength $\bar{w}$. As we will demonstrate in **"Results: Population-activity fluctuations in excitatory-inhibitory networks"** and **"Results: Population averaged correlations in cortical networks"**, this is qualitatively different in networks with mixed excitatory and inhibitory coupling $\bar{w}_E = \bar{w} > 0$ and $\bar{w}_I = -\bar{g}\bar{w} < 0$, respectively: here, the fluctuations of the compound activity increase with $\bar{w}$. The power ratio $\alpha(\omega)$, however, still decreases with $\bar{w}$.

Population-activity fluctuations in excitatory-inhibitory networks

In the foregoing subsection, we have shown that negative feedback alone can efficiently suppress population-rate fluctuations and, hence, spike-train correlations. So far, it is unclear whether the same reasoning applies to networks with mixed excitatory and inhibitory coupling. To clarify this, we now consider a random network composed of a homogeneous excitatory and inhibitory subpopulation $\mathcal{E}$ and $\mathcal{I}$ of size $N_E = |\mathcal{E}|$ and $N_I = |\mathcal{I}| = \gamma N_E$, respectively. Each neuron receives $K = \epsilon N_E$ excitatory and $\gamma K = \epsilon N_I$ inhibitory inputs from $\mathcal{E}$ and $\mathcal{I}$ with synaptic weights $w > 0$ and $-gw < 0$, respectively. In addition, the neurons are driven by external Gaussian white noise. As demonstrated in **"Methods: Linearized network model"**, linearization and averaging across subpopulations leads to a two-dimensional system

$$r(t) = ([\mathbf{W}r + x] * h)(t) \qquad (6)$$

describing the linearized dynamics of the subpopulation averaged activity $r(t) = (r_E(t), r_I(t))^T$. Here, $x(t) = (x_E(t), x_I(t))^T$ denotes the subpopulation averaged external uncorrelated white-noise input with correlation functions $E_{x,t}[x_p(t), x_q(t+\tau)] = \bar{\rho}_p^2 \delta_{pq} \delta(\tau)$ ($\bar{\rho}_p = \rho/\sqrt{N_p}$, $p,q \in \{E,I\}$), and $h(t)$ a normalized linear kernel with $\int_0^\infty dt\, h(t) = 1$. The excitatory and inhibitory subpopulations are coupled through an effective connectivity matrix

$$W = \bar{w}\begin{pmatrix} 1 & -\bar{g} \\ 1 & -\bar{g} \end{pmatrix} \qquad (7)$$

with effective weight $\bar{w} = Kw > 0$ and balance parameter $\bar{g} = \gamma g > 0$.
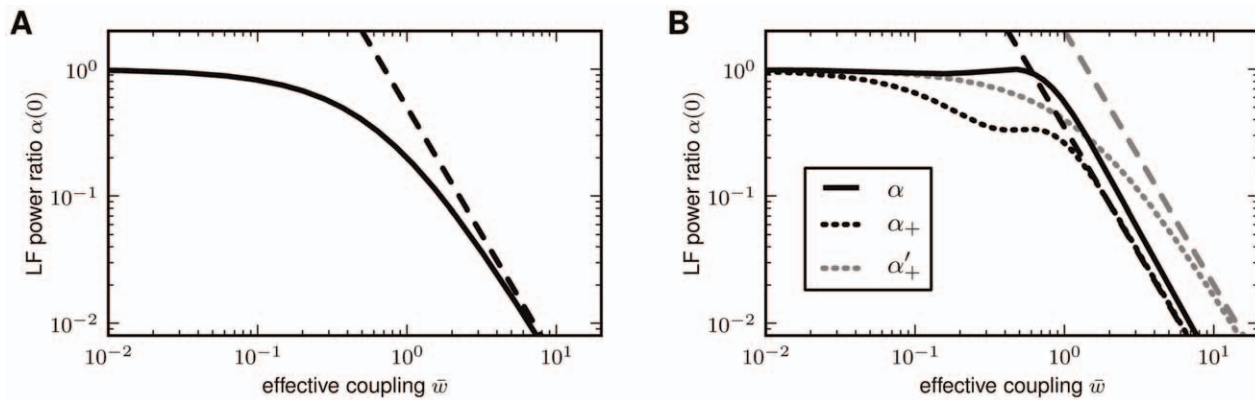
Figure 4. Suppression of low-frequency (LF) population-rate fluctuations in linearized homogeneous random networks with purely inhibitory (A) and mixed excitatory-inhibitory coupling (B). Dependence of the zero-frequency power ratio $\alpha(0)$ on the effective coupling strength $\bar{w}$ (solid curves: full solutions; dashed lines: strong-coupling approximations). The power ratio $\alpha(0)$ represents the ratio between the low-frequency population-rate power in the recurrent networks (**A**: Fig. 3 B; **B**: Fig. 5 A,B) and in networks where the feedback channels are replaced by uncorrelated feedforward input (**A**: Fig. 3 C; **B**, black: Fig. 5 C,D; **B**, gray: Fig. 5 D′). Dotted curves in **B** depict power ratio of the sum modes r_+ and $\tilde{r}_+$ (see text). B: Balance factor $\bar{g} = 1.5$.
doi:10.1371/journal.pcbi.1002596.g004

The two-dimensional system (6)/(7) represents a recurrent system with both positive and negative feedback connections (Fig. 5 A). By introducing new coordinates

$$r_+(t) = (r_E(t) + r_I(t))/\sqrt{2}, \quad r_-(t) = (r_E(t) - r_I(t))/\sqrt{2} \quad (8)$$

and $x_+(t) = (x_E(t) + x_I(t))/\sqrt{2}$, $x_-(t) = (x_E(t) - x_I(t))/\sqrt{2}$, we obtain an equivalent representation of (6)/(7),

$$\begin{pmatrix} r_+(t) \\ r_-(t) \end{pmatrix} = \left(\left[S \begin{pmatrix} r_+ \\ r_- \end{pmatrix} + \begin{pmatrix} x_+ \\ x_- \end{pmatrix} \right] * h \right)(t), \quad (9)$$

describing the dynamics of the sum and difference activity $r_+(t)$ and $r_-(t)$, respectively, i.e. the in- and anti-phase components of the excitatory and inhibitory subpopulations (see [48–50]). The new coupling matrix

$$S = \begin{pmatrix} -w_+ & w_{FF} \\ 0 & 0 \end{pmatrix} \quad (10)$$

reveals that the sum mode $r_+(t)$ is subject to self-feedback ($S_{11} = -w_+ = \bar{w}(1-\bar{g})$) and receives feedforward input from the difference mode $r_-(t)$ ($S_{12} = w_{FF} = \bar{w}(1+\bar{g})$). All remaining connections are absent ($S_{21} = S_{22} = 0$) in the new representation (8) (see Fig. 5 B). The correlation functions of the external noise in the new coordinates are given by $E_{x,t}[x_p(t), x_q(t+\tau)] = \bar{\rho}_{pq}^2 \delta(\tau)$ with $\bar{\rho}_{pq} = \rho^2/N_E (\gamma^{-1}\delta_{pq} + (1-\gamma^{-1})/2)$ ($p,q \in \{+, -\}$).

The feedforward coupling is positive ($w_{FF} > 0$): an excitation surplus ($r_-(t) > 0$) will excite all neurons in the network, an excitation deficit ($r_-(t) < 0$) will lead to global inhibition. In inhibition dominated regimes with $\bar{g} = \gamma g > 1$, the self-feedback of the sum activity $r_+(t)$ is effectively negative ($-w_+ < 0$). The dynamics of the sum rate in inhibition-dominated excitatory-inhibitory networks is therefore qualitatively similar to the dynamics in purely inhibitory networks (**"Results: Suppression of population-activity fluctuations by negative feedback"**). As shown below, the negative feedback loop exposed by the transform (8) leads to an efficient relative suppression of population-rate fluctuations (if compared to the feedforward case).

Mathematically, the coordinate transform (8) corresponds to a *Schur decomposition* of the dynamics: Any recurrent system of type (6) (with arbitrary coupling matrix W) can be transformed to a system with a triangular coupling matrix (see, e.g., [50]). The resulting coupling between the different Schur modes can be ordered so that there are only connections from modes with lower index to modes with the same or larger index. In this sense, the resulting system has been termed 'feedforward' [50]. The original coupling matrix W is typically not normal, i.e. $W^T W \neq W W^T$. Its eigenvectors do not form an orthogonal basis. By performing a Gram-Schmidt orthonormalization of the eigenvectors, however, one can obtain a (normalized) orthogonal basis, a Schur basis. Our new coordinates (8) correspond to the amplitudes (the time evolution) of two orthogonal Schur modes.

The spectra $C_{R_E R_E}(\omega)$, $C_{R_I R_I}(\omega)$, $C_{R_E R_I}(\omega)$ and $C_{R_+ R_+}(\omega)$ of the subpopulation averaged rates r_E, r_I and the sum mode r_+, respectively, are derived in **"Methods: Population-activity spectra of the linear excitatory-inhibitory network"**. In contrast to the purely inhibitory network (see **"Results: Suppression of population-activity fluctuations by negative feedback"**), the population-rate fluctuations of the excitatory-inhibitory network increase monotonously with increasing coupling strength $\bar{w}$. For strong coupling, $C_{R_+ R_+}(\omega)$ approaches

$$\lim_{\bar{w} \to \infty} C_{R_+ R_+}(\omega) = |H(\omega)|^2 \rho^2 \frac{1 + \gamma^{-1}}{2 N_E} \frac{w_{FF}^2}{w_+^2} \quad (11)$$

from below with $w_{FF}/w_+ = (\bar{g}+1)/(\bar{g}-1)$. Close to the critical point ($\bar{g} \simeq 1$), the rate fluctuations become very large; (11) diverges. Increasing the amount of inhibition by increasing $\bar{g}$, however, leads to a suppression of these fluctuations. In the limit $\bar{g} \to \infty$, $C_{R_+ R_+}(\omega)$ and (11) approach the spectrum $\lim_{\bar{w} \to 0} C_{R_+ R_+} = |H|^2 \rho^2 (1 + \gamma^{-1})/(2 N_E)$ of the unconnected network. For strong coupling ($\bar{w} \gg 1$), the ratio $C_{R_E R_E}(\omega)/C_{R_I R_I}(\omega)$ approaches $\bar{g}^2$: the fluctuations of the population averaged excitatory firing rate exceed those of the inhibitory population by a factor $\bar{g}^2$ (independently of $H(\omega)$ and ω).

Similarly to the strategy we followed in the previous subsections, we will now compare the population-rate fluctuations of the
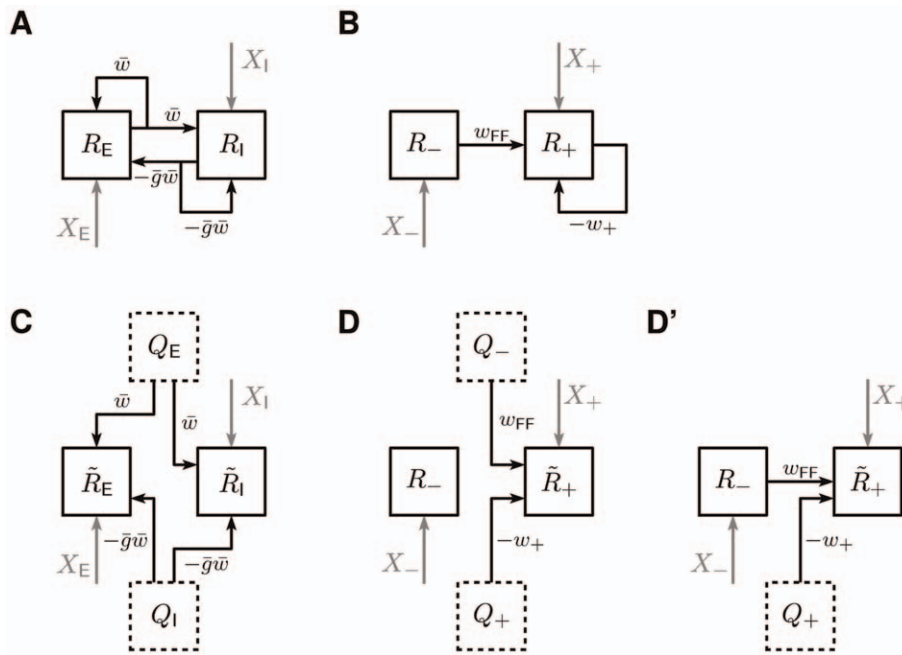
Figure 5. Sketch of the 2D (excitatory-inhibitory) model for the feedback (A,B) and the feedforward scenario (C,D) in normal (A,C) and Schur-basis representation (B,D). A: Original 2D recurrent system. **B**: Schur-basis representation of the system shown in A. **C**: Feedforward scenario: Excitatory and inhibitory feedback connections of the original network (A) are replaced by feedforward input from populations with rates Q_E, Q_I, respectively. **D**: Schur-basis representation of the system shown in C. **D'**: Alternative feedforward scenario: Here, the feedforward channel (weight w_{FF}) of the original system in Schur basis (B) remains intact. Only the inhibitory feedback (weight $-w_+$) is replaced by feedforward input Q_+.
doi:10.1371/journal.pcbi.1002596.g005

feedback system (6), or equivalently (9), to the case where the feedback channels are replaced by feedforward input with identical auto-statistics. A straight-forward implementation of this is illustrated in Fig. 5 C: Here, the excitatory and inhibitory feedback channels R_E and R_I are replaced by uncorrelated feedforward inputs Q_E and Q_I, respectively. The Schur representation of this scenario is depicted in Fig. 5 D. According to (6), the Fourier transforms of the response fluctuations of this system read

$$\begin{pmatrix} \tilde{R}_E(\omega) \\ \tilde{R}_I(\omega) \end{pmatrix} = H(\omega)\left[W\begin{pmatrix} Q_E(\omega) \\ Q_I(\omega) \end{pmatrix} + \begin{pmatrix} X_E(\omega) \\ X_I(\omega) \end{pmatrix} \right]. \quad (12)$$

With $\tilde{R}_+ = (\tilde{R}_E + \tilde{R}_I)/\sqrt{2}$, and using $C_{Q_E Q_E} = C_{R_E R_E}$, $C_{Q_I Q_I} = C_{R_I R_I}$, $C_{Q_E Q_I} = C_{Q_E X_E} = C_{Q_E X_I} = C_{Q_I X_E} = C_{Q_I X_I} = 0$, we can express the spectrum $C_{\tilde{R}_+ \tilde{R}_+}(\omega)$ of the sum activity in the feedforward case in terms of the spectra $C_{R_E R_E}(\omega)$ and $C_{R_I R_I}(\omega)$ of the feedback system (see eq. (55)). For strong coupling ($\bar{w} \gg 1$), the zero-frequency component ($H(0) = 1$) becomes

$$C_{\tilde{R}_+ \tilde{R}_+}(0) \simeq \bar{w}^2 \rho^2 \frac{1 + \gamma^{-1}}{N_E} \frac{4\bar{g}^2}{(\bar{g} - 1)^2}. \quad (13)$$

Thus, for strong coupling, the zero-frequency power ratio

$$\alpha_+(0) = \frac{C_{R_+ R_+}(0)}{C_{\tilde{R}_+ \tilde{R}_+}(0)} \simeq \frac{(\bar{g} + 1)^2}{8\bar{w}^2 \bar{g}^2} \quad (14)$$

reveals a relative suppression of the population-rate fluctuations in the feedback system which is proportional to $1/\bar{w}^2$ (see Fig. 4 B; black dashed line). The power ratio $\alpha_+(0)$ for arbitrary weights $\bar{w}$

is depicted in Fig. 4 B (black dotted curve). For a network at the transition point $\bar{g} = 1$, (14) equals $1/(2\bar{w}^2)$. Increasing the level of inhibition by increasing $\bar{g}$ leads to a decrease in the power ratio: in the limit $\bar{g} \to \infty$, (14) approaches $1/(8\bar{w}^2)$ monotonously.

Above, we suggested that the negative self-feedback of the sum mode R_+, weighted by $-w_+$ (Fig. 5 B), is responsible for the fluctuation suppression in the recurrent excitatory-inhibitory system. Here, we test this by considering the case where this feedback loop is opened and replaced by uncorrelated feedforward input Q_+, weighted by $-w_+$, while the feedforward input from the difference mode R_-, weighted by w_{FF}, is left intact (see Fig. 5 D'). As before, we assume that the auto-statistics of Q_+ is identical to the auto-statistics of R_+ as obtained in the feedback case, i.e. $C_{Q_+ Q_+}(\omega) = C_{R_+ R_+}(\omega)$. According to the Schur representation of the population dynamics (9)/(10), the Fourier transform of the sum mode of this modified system is given by

$$\tilde{R}_+(\omega) = H(\omega)\left(-w_+ Q_+(\omega) + w_{FF} \tilde{R}_-(\omega) + X_+(\omega) \right). \quad (15)$$

With $C_{\tilde{R}_+ \tilde{R}_+}(\omega)$ given in (54) and $C_{R_+ R_+}(\omega)$, we obtain the power ratio

$$\alpha'_+(\omega) = \frac{C_{R_+ R_+}(\omega)}{C_{\tilde{R}_+ \tilde{R}_+}(\omega)} = \frac{1}{w_+^2 |H(\omega)|^2 + |1 + w_+ H(\omega)|^2}. \quad (16)$$

Its zero-frequency component $\alpha'_+(0)$ is shown in Fig. 4 B (gray dotted curve). For strong coupling, the power ratio decays as $1/(2w_+^2)$ (gray dashed line in Fig. 4 B). Thus, the (relative) power in the recurrent system is reduced by strengthening the negative self-feedback loop, i.e. by increasing w_+.

So far, we have presented results for the subpopulation averaged firing rates $r_E(t)$ and $r_I(t)$ and the sum mode $r_+(t)$. The spectrum of the compound rate $r(t) = N^{-1} \sum_{i=1}^{N} r_i(t) = N^{-1}[N_E r_E(t) + N_I r_I(t)]$, i.e. the activity averaged across the entire population, reads

$$C_{RR}(\omega) = N^{-2}\Big(N_E^2 C_{R_E R_E}(\omega) + N_I^2 C_{R_I R_I}(\omega) + \\ N_E N_I [C_{R_E R_I}(\omega) + C_{R_E R_I}(\omega)^*]\Big). \qquad (17)$$

In the feedforward scenario depicted in Fig. 5 C, the spectrum of the compound rate $\tilde{R} = H(\bar{w}Q_E + \bar{w}\bar{g}Q_I + X)$ (with $X = N^{-1}\sum_{i=1}^{N} X_i$) is given by

$$C_{\tilde{R}\tilde{R}}(\omega) = |H(\omega)|^2\Big(\bar{w}^2 C_{R_E R_E} + \bar{w}^2\bar{g}^2 C_{R_I R_I} + \rho^2/N\Big). \qquad (18)$$

For strong coupling, the corresponding low-frequency power ratio $\alpha(0) = C_{RR}(0)/C_{\tilde{R}\tilde{R}}(0)$ (black solid curve in Fig. 4 B) exhibits qualitatively the same decrease $\propto \bar{w}^{-2}$ as the sum mode.

To summarize the results of this subsection: the population dynamics of a recurrent network with mixed excitatory and inhibitory coupling can be mapped to a two-dimensional system describing the dynamics of the sum and the difference of the excitatory and inhibitory subpopulation activities. This equivalent representation uncovers that, in inhibition dominated networks ($\bar{g} > 1$), the sum activity is subject to negative self-feedback. Thus, the dynamics of the sum activity in excitatory-inhibitory networks is qualitatively similar to the population dynamics of purely inhibitory networks (see **"Results: Suppression of population-activity fluctuations by negative feedback"**). Indeed, the comparison of the compound power-spectra of the intact recurrent network and networks where the feedback channels are replaced by feedforward input reveals that the (effective) negative feedback in excitatory-inhibitory networks leads to an efficient suppression of population-rate fluctuations.

Population averaged correlations in cortical networks

The results presented in the previous subsections describe the fluctuations of the compound activity. Pairwise correlations $c_{ij}(t) = E_{t'}[\bar{s}_i(t+t')\bar{s}_j(t')]$ between the (centralized) spike trains $\bar{s}_i(t) = s_i(t) - E_{t'}[s_i(t')]$ are outside the scope of such a description. In this subsection, we consider the same excitatory-inhibitory network as in **"Results: Population-activity fluctuations in excitatory-inhibitory networks"** and present a theory for the population averaged spike-train cross-correlations. In general, this is a hard problem. To understand the structure of cross-correlations, it is however sufficient to derive a relationship between the cross- and auto-covariances in the network, because the latter can, to good approximation, be understood in mean-field theory. The integral of the auto-covariance function of spiking LIF neurons can be calculated by Fokker-Planck formalism [12,31,51]. To determine the relation between the cross-covariance and the auto-covariance, we replace the spiking dynamics by a reduced linear model with covariances obeying, to linear order, the same relation. We present the full derivation in **"Methods: Linearized network model"**. There, we first derive an approximate linear relation between the auto- and cross-covariance functions $a(\tau)$ and $c(\tau)$, respectively, of the LIF network. A direct solution of this equation is difficult. In the second step, we therefore show that there exists a linear stochastic system with activity u and correlations $a^u(\tau)$ and $c^u(\tau)$ fulfilling the same equation as the original LIF model. This reduced model can be solved in the frequency domain by standard Fourier methods. Its

solution allows us, by construction, to determine the relation between the integral cross-covariances $C(0)$ and the integral auto-covariances $A(0)$ up to linear order.

As we are interested in the covariances averaged over many pairs of neurons, we average the resulting set of N linear self-consistency equations (56) for the covariance matrix in the frequency domain $C(\omega)$ over statistically identical pairs of neurons and many realizations of the random connectivity (see **"Methods: Population averaged correlations in the linear EI network"**). This yields a four-dimensional linear system (76) describing the population averaged variances A_E and A_I of the excitatory and inhibitory subpopulations, and the covariances $C_{E \not\rightleftharpoons E}$ and $C_{I \not\rightleftharpoons I}$ for unconnected excitatory-excitatory and inhibitory-inhibitory neuron pairs, respectively (note that we use the terms "variance" and "covariance" to describe the *integral* of the auto- and cross-correlation function, respectively; in many other studies, they refer to the zero-lag correlation functions instead). The dependence of the variances and covariances on the coupling strength $\bar{w}$, obtained by numerically solving (76), is shown in Fig. 6. We observe that the variances A_E and A_I of excitatory and inhibitory neurons are barely distinguishable (Fig. 6 A). With the approximation $A := A_E = A_I$, explicit expressions can be obtained for the covariances (thick dashed curves Fig. 6 E):

$$C_{EE/II} = \underbrace{\frac{1}{(1-\bar{w}(1-\bar{g}))^2}}_{I} C_{shared}^{in}$$
$$+ \underbrace{\frac{1}{1-\bar{w}(1-\bar{g})}}_{II} \underbrace{2\bar{w}A \begin{cases} \dfrac{1}{N_E} & \text{for EE} \\[2mm] \dfrac{-\bar{g}}{N_I} & \text{for II} \end{cases}}_{III}, \qquad (19)$$
$$C_{EI} = \frac{1}{2}(C_{EE} + C_{II})$$
$$\text{with} \quad C_{shared}^{in} = \bar{w}^2\left(\frac{1}{N_E} + \frac{\bar{g}^2}{N_I}\right)A.$$

The deviations from the full solutions (thin solid curves in Fig. 6 E), i.e. for $A_E \neq A_I$, are small. In the reduced model, both the external input and the spiking of individual neurons contribute to an effective noise. As the fluctuations in the reduced model depend linearly on the amplitude ρ of this noise, the variances A and covariances C_{pq} ($p,q \in \{E,I\}$) can be expressed in units of the noise variance ρ^2. Consequently, the correlation coefficients C_{pq}/A are independent of ρ^2 (see Fig. 6).

The analytical form (19) of the result shows that the correlations are smaller than expected given the amount of shared input a pair of neurons receives: The quantity C_{shared}^{in} in the first line is the contribution of shared input to the covariance. For strong coupling $\bar{w} \gg 1$, the prefactor **I** causes a suppression of this contribution. Its structure is typical for a feedback system, similar to the solution (3) of the one-population or the solution (52) of the two-population model. The term $\bar{w}(1-\bar{g})$ in the denominator represents the negative feedback of the compound rate. The prefactor **II** in the second line of (19) is again due to the feedback and suppresses the contribution of the factor **III**, which represents the effect of direct connections between neurons.

Our results are consistent with a previous study of the decorrelation mechanism: In [17], the authors considered how correlations scale with the size N of the network where the synaptic weights are chosen as $J \propto 1/\sqrt{N}$. As a result, the covariance C_{shared}^{in} in (19) caused by
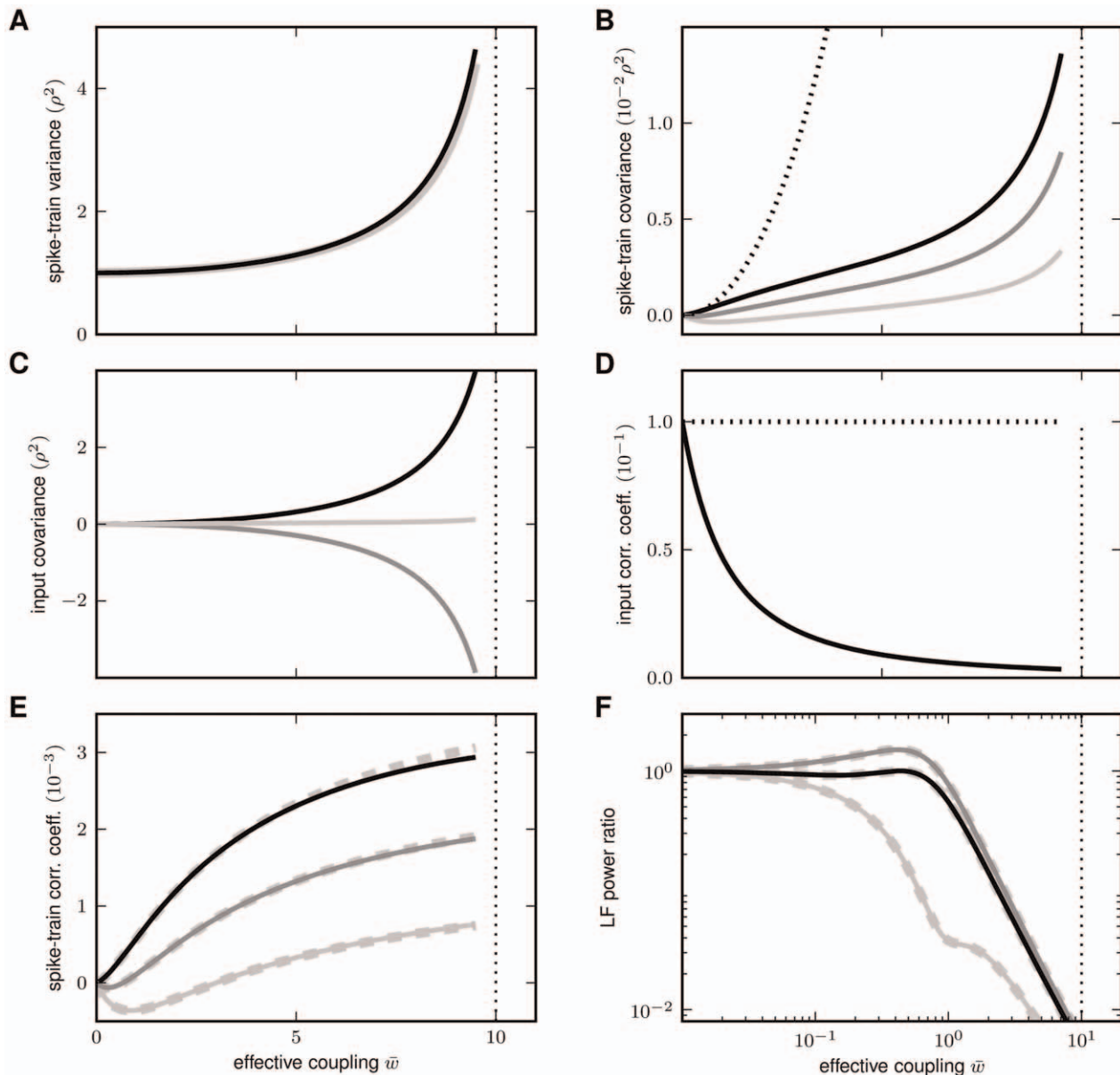
Figure 6. Dependence of population averaged correlations and population-rate fluctuations on the effective coupling $\bar{w} = Kw$ in a linearized homogeneous network with excitatory-inhibitory coupling. A: Spike-train variances A_E (black) and A_I (gray) of excitatory and inhibitory neurons. **B**: Spike-train covariances C_{EE} (black solid), C_{EI} (dark gray solid) and C_{II} (light gray solid) for excitatory-excitatory, excitatory-inhibitory and inhibitory-inhibitory neuron pairs in the recurrent network, respectively, and shared-input contribution C_{shared}^{in} (black dotted curve; 'feedforward case'). **C**: Decomposition of the total input covariance C^{in} (light gray) into shared-input covariance C_{shared}^{in} (black) and weighted spike-train covariance C_{corr}^{in} (dark gray). Covariances in A, B and C are given in units of the noise variance ρ^2. **D**: Input-correlation coefficient C^{in}/A^{in} in the recurrent network (black solid curve). In the feedforward case, the input-correlation coefficient is identical to the network connectivity ϵ (horizontal dotted line). **E**: Spike-train correlation coefficients C_{EE}/A_E (black), $C_{EI}/\sqrt{A_E A_I}$ (dark gray) and C_{II}/A_I (solid light gray curve) for excitatory-excitatory, excitatory-inhibitory and inhibitory-inhibitory neuron pairs, respectively. Thick dashed curves represent approximate solutions assuming $A_E = A_I$. **F**: Low-frequency (LF) power ratios α (black), α_E (dark gray), α_I (solid light gray) for the population rate $r(t)$ and the excitatory and inhibitory subpopulation rates $r_E(t)$ and $r_I(t)$, respectively. The LF power ratio represents the ratio between the LF spectra in the recurrent network and for the case where the feedback channels are replaced by feedforward input with $C_{EI} = 0$ (cf. Fig. 5 C). Thick dashed curves in F show power ratios obtained by assuming that the auto-correlations are identical in the feedback and the feedforward scenario (see main text). Vertical dotted lines mark the stability limit of the linear model (see "**Methods: Linearized network model**"). A–F: $K = 1000$, $\epsilon = 0.1$, $\gamma = 1/4$, $g = 6$, $\bar{g} = \gamma g = 3/2$, $N = K(1+\gamma)/\epsilon = 12500$.
doi:10.1371/journal.pcbi.1002596.g006

shared input is independent of the network size, while the feedback $\bar{w}(1-\bar{g}) \propto \epsilon N(1-\bar{g})(J + O(J^2))$ scales—to leading order—as $\sqrt{N}$ (see (45)). Consequently, the first line in (19) scales as $1/N$. The same scaling holds for the second line in (19), explaining the decay of correlations as $1/N$ found in [17].

The first line in (19) is identical for any pair of neurons. The second line is positive for a pair of excitatory neurons and negative for a pair of inhibitory neurons. In other words, excitatory neurons are more correlated than inhibitory ones. Together with the third line in (19), this reveals a characteristic correlation structure:

$C_{\text{EE}} > C_{\text{EI}} > C_{\text{II}}$ (Fig. 6 B,E). For strong coupling $\bar{w} \gg 1$, the difference between the excitatory and inhibitory covariance is $C_{\text{EE}} - C_{\text{II}} \simeq \dfrac{2}{\bar{g}-1}(\dfrac{1}{N_{\text{E}}} + \dfrac{\bar{g}}{N_{\text{I}}})A$. The difference decreases as the level $\bar{g}$ of inhibition is increased, i.e. the further the network is in the inhibition dominated regime, away from the critical point $\bar{g} = 1$.

To understand the suppression of shared-input correlations in recurrent excitatory-inhibitory networks, consider the correlation between the local inputs $I_{k/l} = [\boldsymbol{W}\boldsymbol{r}]_{k/l}$ of a pair of neurons k, l. The input-correlation coefficient $C^{\text{in}}/A^{\text{in}} = \text{Cov}[I_k, I_l]\sqrt{\text{Var}[I_k]\text{Var}[I_l]}$ can be expressed in terms of the averaged spike-train covariances:

$$
\begin{aligned}
C^{\text{in}} &= \text{Cov}[I_k, I_l]C^{\text{in}}_{\text{shared}} + C^{\text{in}}_{\text{corr}} \\
A^{\text{in}} &= \text{Var}[I_k]\epsilon^{-1}\bar{w}^2\left(\frac{1}{N_{\text{E}}} + \frac{\bar{g}^2}{N_{\text{I}}}\right)A + C^{\text{in}}_{\text{corr}} \qquad (20) \\
\text{with} \quad & C^{\text{in}}_{\text{corr}} = \bar{w}^2(C_{\text{EE}} - 2\bar{g}C_{\text{EI}} + \bar{g}^2 C_{\text{II}})
\end{aligned}
$$

(see **"Methods: Population averaged correlations in the linear EI network"**: The input covariance C^{in} equals the average quantity $C_{xy,\text{B}}$ given in (67), the input variance A^{in} is given by (63) as $A_{x,\text{B}}$. The term $C^{\text{in}}_{\text{shared}}$ represents the contribution due to the spike-train variances of the shared presynaptic neurons (see (19)). This contribution is always positive (provided the network architecture is consistent with Dale's law; see [15]). In a purely feedforward scenario with uncorrelated presynaptic sources, $C^{\text{in}}_{\text{shared}}$ is the only contribution to the input covariance of postsynaptic neurons. The resulting response correlation for this feedforward case is much larger than in the feedback system (Fig. 6 B, black dotted curve). The correlation coefficient between inputs to a pair of neurons in the feedforward case is identical to the network connectivity ϵ (horizontal dotted curve in Fig. 6 D; see [15]). In an inhibition dominated recurrent network, spike-train correlations between pairs of different source neurons contribute the additional term $C^{\text{in}}_{\text{corr}}$, which is negative and of similar absolute value as the shared-input contribution $C^{\text{in}}_{\text{shared}}$. Thus, the two terms $C^{\text{in}}_{\text{shared}}$ and $C^{\text{in}}_{\text{corr}}$ partly cancel each other (see Fig. 6 C). In consequence, the resulting input correlation coefficient $C^{\text{in}}/A^{\text{in}}$ is smaller than ϵ (see Fig. 6 D; here: $\epsilon = 0.1$).

The correlations in a purely inhibitory network can be obtained from (19) by replacing $N_{\text{E}} \to N$, taking into account the negative sign of w in $\bar{w} = -Kw$ and setting $g = 0$ and $\gamma = 0$:

$$
C = \left(-1 + \frac{1}{(1+\bar{w})^2}\right)\frac{A}{N}. \qquad (21)
$$

For finite coupling strength $\bar{w} > 0$, this expression is negative. The contributions of shared input and spike-train correlations to the input correlation are given by $C^{\text{in}}_{\text{shared}} = \bar{w}^2 \dfrac{A}{N} > 0$ and $C^{\text{in}}_{\text{corr}} = \bar{w}^2 C$, respectively (see (19) and (20)). Using (21), we can directly verify that $C^{\text{in}}_{\text{corr}} < 0$, because pairwise correlations C are negative, leading to a partial cancellation $C^{\text{in}}_{\text{shared}} + C^{\text{in}}_{\text{corr}} = \bar{w}^2 \dfrac{1}{(1+\bar{w})^2}\dfrac{A}{N}$: the right hand side is smaller in magnitude by a factor of $\simeq \dfrac{1}{\bar{w}^2}$ compared to each individual contribution. Hence, as in the network with excitation and inhibition, shared-input correlations are partly canceled by the contribution due to presynaptic pairwise spike-train correlations. In the feedforward scenario with zero presynaptic spike-train correlations, in contrast, the response correlations are determined by shared input alone and are therefore increased. The suppression of shared-input correlations

in the feedback case is what we call 'decorrelation' in the current work. In purely inhibitory networks, this decorrelation is caused by weakly negative pairwise correlations (21). For sufficiently strong negative feedback, correlations are smaller in absolute value as compared to the feedforward case. The absolute value of these anti-correlations is bounded by A/N.

The similarity in the results obtained for purely inhibitory networks and excitatory-inhibitory networks demonstrates that the suppression of pairwise correlations and population-activity fluctuations is a generic phenomenon in systems with negative feedback. It does not rely on an internal balance between excitation and inhibition.

As discussed in **"Results: Suppression of population-rate fluctuations in LIF networks"**, the suppression of correlations in the recurrent network is accompanied by a reduction of population-activity fluctuations. With the population averaged correlations (19), the power (1) of the population activity $r(t)$ reads

$$
\begin{aligned}
C_{RR} = \frac{1}{N^2}[&N_{\text{E}}A_{\text{E}} + N_{\text{I}}A_{\text{I}} + N_{\text{E}}(N_{\text{E}}-1)C_{\text{EE}} + \\
&N_{\text{I}}(N_{\text{I}}-1)C_{\text{II}} + 2N_{\text{E}}N_{\text{I}}C_{\text{EI}}]. \qquad (22)
\end{aligned}
$$

In **"Results: Population-activity fluctuations in excitatory-inhibitory networks"**, we showed that the population-activity fluctuations are amplified if the local input in the recurrent system is replaced by feedforward input from independent excitatory and inhibitory populations (see Fig. 5 C). This manipulation corresponds to a neglect of correlations C_{EI} between excitatory and inhibitory neurons. All remaining correlations (A_{E}, A_{I}, C_{EE}, C_{II}) are preserved. With the resulting response auto- and cross-correlations $\tilde{A}$ and $\tilde{C}$ given by (84), the power (1) of the population activity becomes

$$
C_{\tilde{R}\tilde{R}} = \frac{1}{N}\tilde{A} + \left(1 - \frac{1}{N}\right)\tilde{C}. \qquad (23)
$$

For large effective coupling $\bar{w}$, the power ratio $\alpha = C_{RR}/C_{\tilde{R}\tilde{R}}$ decays as $1/\bar{w}^2$ (black curve in Fig. 6 F). Note that the power ratio α derived here is indistinguishable from the one we obtained in the framework of the population model in **"Results: Population-activity fluctuations in excitatory-inhibitory networks"** (black solid curve in Fig. 4 B). Although the derivation of the macroscopic model in **"Results: Population-activity fluctuations in excitatory-inhibitory networks"** is different from the one leading to the population averaged correlations described here, the two models are consistent: They describe one and the same system and lead to identical power ratios.

The fluctuation suppression is not only observed at the level of the entire network, i.e. for the population activity $r(t)$, but also for each individual subpopulation $\mathcal{E}$ and $\mathcal{I}$, i.e. for the subpopulation averaged activities $r_{\text{E}}(t)$ and $r_{\text{I}}(t)$. The derivation of the corresponding power ratios α_{E} and α_{I} is analog to the one described above. As a result of the correlation structure $C_{\text{EE}} > C_{\text{II}}$ in the feedback system (see Fig. 6 B), the power of the inhibitory population activity is smaller than the power of the excitatory population activity. In consequence, $\alpha_{\text{E}} > \alpha_{\text{I}}$ (gray curves in Fig. 6 F).

In (22) and (23), the auto-correlations are scaled by $1/N$, while the cross-correlations enter with a prefactor of order unity. For large N, one may therefore expect that the suppression of population-activity fluctuations is essentially mediated by pairwise correlations. In the recurrent system, however, the cross-correlations C_{xy} ($x,y \in \{\text{E},\text{I}\}$) are of order A/N (see Fig. 6 and (19)). It is therefore a priori not clear whether the fluctuation suppression is

indeed dominated by pairwise correlations. In our framework, one can explicitly show that the auto-correlation is irrelevant: Replacing the auto-correlation $\tilde{A}$ in (23) by the average auto-correlation $(N_E A_E + N_I A_I)/N$ of the intact feedback system has no visible effect on the resulting power ratio (dashed curves in Fig. 6 F). The difference in the spectra of the population activities C_{RR} and $C_{\tilde{R}\tilde{R}}$ is therefore essentially caused by the cross-correlations.

The absolute population-activity fluctuations in purely inhibitory and in excitatory-inhibitory networks show a qualitatively different dependence on the synaptic coupling $\bar{w}$, in agreement with the previous sections. In networks with excitation and inhibition, the correlation coefficient increases with increasing synaptic coupling (see Fig. 6 E). Hence, the population-activity fluctuations grow with increasing coupling strength. In purely inhibitory networks, in contrast, the pairwise spike-train correlation decreases monotonously with increasing magnitude of the coupling strength $\bar{w}$, see (21). In consequence, the population-activity fluctuations decrease. The underlying reason is that, in the inhibitory network, the power of the population activity is directly proportional to the covariance of the input currents, which is actively suppressed, as shown above. For excitatory-inhibitory networks, these two quantities are not proportional (compare (20) and (1)) due to the different synaptic weights appearing in the input covariance.

To compare our theory to simulations of spiking LIF networks, we need to determine the effect of a synaptic input on the response activity of the neuron model. To this end, we employ the Fokker-Planck theory of the LIF model (see **"Methods: Response kernel of the LIF model"**). In this context, the steady state of the recurrent network is characterized by the mean μ and the standard deviation σ of the total synaptic input. Both μ and σ depend on the steady-state firing rate in the network. The steady-state firing rate can be determined in a self-consistent manner [12] as the fixed point of the firing rate approximation (42). The approximation predicts the firing rate to sufficient accuracy of about $\pm 1\,\text{s}^{-1}$ (see Fig. 7 A). We then obtain an analytical expression of the low-frequency transfer which relates the fluctuation $v_j(t) = \bar{v} + \epsilon\delta(t)$ of a synaptic input to neuron i to the fluctuation of neuron i's response firing rate to linear order, so that $\int_0^\infty \delta v_i(t)\,dt = \epsilon w(J_{ij})$. This relates the postsynaptic potential J_{ij} in the LIF model to the effective linear coupling $w_{ij} = w(J_{ij})$ in our linear theory. The functional relation $w(J)$ can be derived in analytical form by linearization of (42) about the steady-state working point. Note that $w(J)$ depends on μ and σ and, hence, on the steady-state firing rate in the network. The derivation outlined in **"Methods: Response kernel of the LIF model"** constitutes an extension of earlier work [21,33] to quadratic order in J. The results agree well with those obtained by direct simulation for a large range of synaptic amplitudes (see Fig. 8).

Fig. 7 B compares the population averaged correlation coefficients C/A obtained from the linear reduced model, see (19), and simulations of LIF networks. Note that the absolute value of the noise amplitude ρ in the reduced model does not influence the correlation coefficient C/A, as both quantities C and A depend linearly on ρ^2. Theory and simulation agree well for synaptic weights up to $J \approx 1\,\text{mV}$. For larger synaptic amplitudes, the approximation of the effective linear transfer for a single neuron obtained from the Fokker-Planck theory deviates from its actual value (see Fig. 8 B). Fig. 7 C shows that the cancellation of the input covariance in the LIF network is well explained by the theory.

Previous work [17] suggested that positive correlations between excitatory and inhibitory inputs lead to a negative component in the input correlation which, in turn, suppresses shared-input correlations. The mere existence of positive correlations between excitatory and inhibitory inputs is however not sufficient. To explain the effect, it is necessary to take the particular correlation structure $C_{EE} > C_{EI} > C_{II}$ into account. To illustrate this, consider the case where the correlation structure is destroyed by replacing all pairwise correlations in the input spike-train ensemble by the overall population average $C = (N_E C_{EE} + N_I C_{II})/(N_E + N_I) > 0$ (homogenization of correlations). The resulting response correlations (upper gray curve in Fig. 7 B) are derived in **"Methods: Population averaged correlations in the linear EI network"**, eq. (86). In simulations of LIF networks, we study the effect of homogenized spike-train correlations by first recording the activity of the intact recurrent network, randomly reassigning the neuron type (E or I) to each recorded spike train, and feeding this activity into a second population of neurons. Compared to the intact recurrent network, the response correlations are significantly larger (Fig. 7 B). The contribution of homogenized spike-train correlations to the input covariance C^{in} (see (20)) is given by $C^{\text{in}}_{\text{corr,hom}} = \bar{w}^2(1 - \bar{g})^2 C \geq 0$. For positive spike-train correlations $C > 0$, this contribution is greater or equal zero (zero for $\bar{g} = 1$). Hence, it cannot compensate the (positive) shared-input contribution $C^{\text{in}}_{\text{shared}}$ (see Fig. 7 C). In consequence, input correlations, output correlations and, in turn, population-rate fluctuations (Fig. 7 D) cannot be suppressed by homogeneous positive correlations in the input spike-train ensemble. Canceling of shared-input correlations requires either negative spike-train correlations (as in purely inhibitory networks) or a heterogeneity in correlations across different pairs of neurons (e.g. $C_{EE} > C_{EI} > C_{II}$).

Effect of feedback manipulations

In the previous subsections, we quantified the suppression of population-rate fluctuations in recurrent networks by comparing the activity in the intact recurrent system (feedback scenario) to the case where the feedback is replaced by feedforward input with some predefined statistics (feedforward scenario). We particularly studied the effect of neglecting the auto-statistics of the compound feedback, (the structure of) correlations within the feedback ensemble and/or correlations between the feedback and the external input. In all cases, we observed a significant amplification of population-activity fluctuations in the feedforward scenario. In this subsection, we further investigate the role of different types of feedback manipulations by means of simulations of LIF networks with excitatory-inhibitory coupling. To this end, we record the spiking activity of the recurrent network (feedback case), apply different types of manipulations to this activity (described in detail below) and feed this modified activity into a second population of identical (unconnected) neurons (feedforward case). As before, the connectivity structure (in-degrees, shared-input structure, synaptic weights) is exactly identical in the feedback and the feedforward case.

In **"Methods: Linearized network model"**, we show that the low-frequency fluctuations of the population rate $s(t)$ of the spiking model are captured by the reduced model $r(t)$ presented in the previous subsections. To verify that the theory based on excitatory and inhibitory population rates is indeed sufficient to explain the decorrelation mechanism, we first consider the case where the sender identities of the presynaptic spike train are randomly shuffled. Fig. 9 A shows the power-spectrum of the population activity recorded in the original network (FB) as well as the spectra obtained after shuffling spike-train identities within the excitatory and inhibitory subpopulations separately (Shuff2D), or across the entire network (Shuff1D). As shuffling of neuron identities does not change the population rates, all three
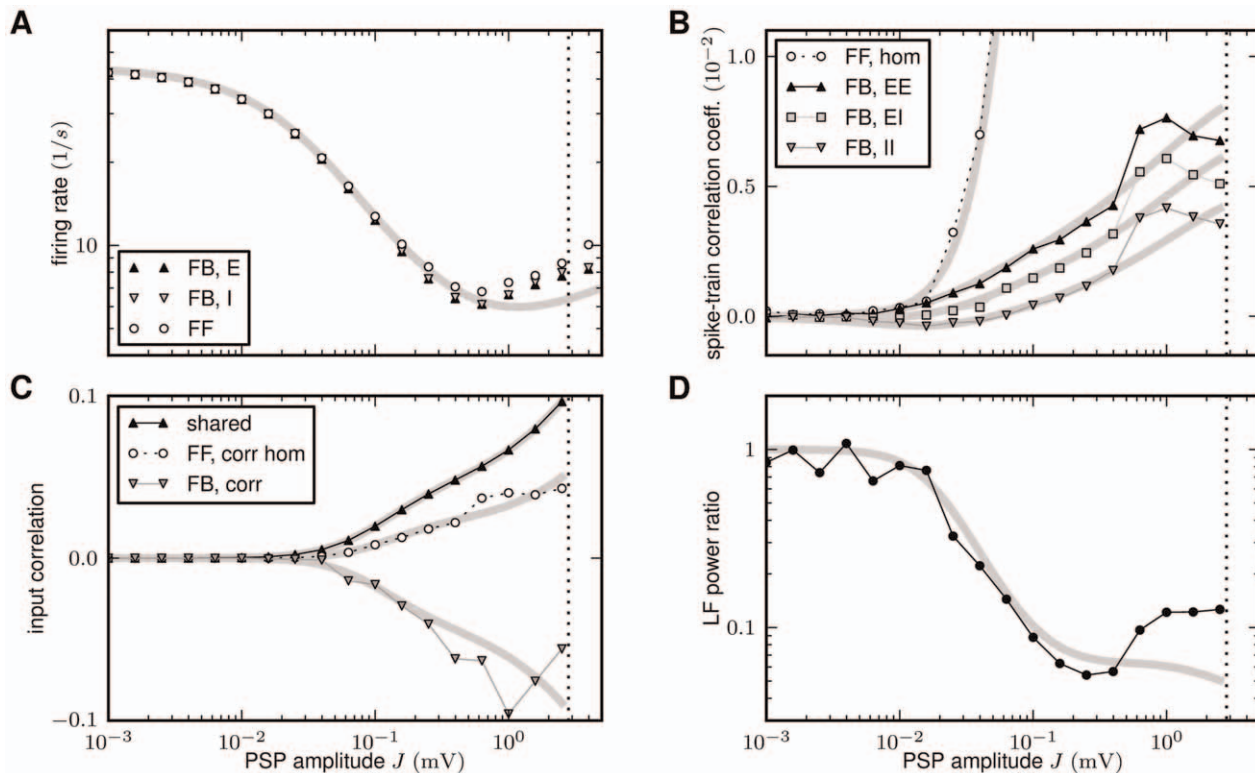
Figure 7. Comparison between predictions of the linear theory (thick gray curves) and direct simulation of the LIF-network model (symbols and thin lines). Dependence of the spike-train and population-rate statistics on the synaptic weight J (PSP amplitude) in a recurrent excitatory-inhibitory network ('feedback system', 'FB') and in a population of unconnected neurons receiving randomized feedforward input ('feedforward system', 'FF') from neurons in the recurrent network. Average presynaptic firing rates and shared-input structure are identical in the two systems. In the FF case, the average correlations between presynaptic spike-trains are homogenized (i.e. $C_{EE} = C_{EI} = C_{II}$) as a result of the random reassignment of presynaptic neuron types. The mapping of the LIF dynamics to the linear reduced dynamics (**"Methods: Response kernel of the LIF model"**) relates the PSP amplitude J to the effective coupling strength $w(J)$ by (45), as shown in Fig. 8 B. **A**: Average firing rates ν_0 in the FB (black up-triangles: excitatory neurons; gray down-triangles: inhibitory neurons) and in the FF system (open circles). Analytical prediction (??) (gray curve). **B**: Spike-train correlation coefficients C_{EE}/A_E (black up-triangles), $C_{EI}/\sqrt{A_E A_I}$ (gray squares) and C_{II}/A_I (gray down-triangles) for excitatory-excitatory, excitatory-inhibitory, and inhibitory-inhibitory neuron pairs, respectively, in the FB system. Analytical prediction (19) (gray curves). Spike-train correlation coefficient $\tilde{C}/\tilde{A}$ (open circles) in the FF system with homogenized presynaptic spike-train correlations. Analytical prediction (86) (underlying gray curve). **C**: Shared-input (C_{shared}^{in}; black up-triangles) and spike-correlation contribution C_{corr}^{in} (FB: gray down-triangles; FF: open circles) to the input correlation C^{in} (normalized by $\sqrt{A_E A_I}$). Analytical predictions (20). **D**: Low-frequency (LF) power ratio of the compound activity. Vertical dotted lines in A–D mark the stability limit of the linear model (see **"Methods: Linearized network model"**). $N = 12500$, $K = 1000$, $\gamma = 1/4$, $g = 6$. Size of postsynaptic population in the FF case: $M = 2000$. Simulation time: $T = 100$ s.
doi:10.1371/journal.pcbi.1002596.g007

compound spectra are identical. Fig. 9 B shows the response power-spectra of the neuron population receiving the shuffled spike trains. Shuffling within the subpopulations (Shuff2D) preserves the population-specific fluctuations and average correlations. The effect on the response fluctuations is negligible (compare black and light gray curves in Fig. 9 B). In particular, the power of low-frequency fluctuations remains unchanged (Fig. 9 C). This result confirms that population models which take excitatory and inhibitory activity separately into account are sufficient to explain the observations. Shuffling of spike-train identities across subpopulations (Shuff1D), in contrast, causes an increase in the population fluctuations by about one order of magnitude (Fig. 9 B,C; dark gray). This outcome is in agreement with the result obtained by homogenizing pairwise correlations (see Fig. 7) and demonstrates that the excitatory and inhibitory subpopulation rates have to be conserved to explain the observed fluctuation suppression.

The shuffling experiments and the results of the linear model in the previous subsections suggest that the precise temporal structure of the *population averaged* activities within homogeneous subpopu-

lations is essential for the suppression of population-rate fluctuations. Preserving the exact structure of individual spike trains is not required. This is confirmed by simulation experiments where new sender identities were randomly reassigned for each individual presynaptic spike (rather than for each spike train; data not shown). This operation destroys the structure of individual spike trains but preserves the compound activities. The results are similar to those reported here.

So far, it is unclear how sensitive the fluctuation-suppression mechanism is to perturbations of the temporal structure of the population rates. To address this question, we replaced the excitatory and inhibitory spike trains in the feedback ensemble by independent realizations of inhomogeneous Poisson processes (PoissI) with intensities given by the measured excitatory and inhibitory population rates $s_E(t)$ and $s_I(t)$ of the recurrent network, respectively. Note that the compound rates of a single realization of this new spike-train ensemble are similar but not identical to the original population rates $s_E(t)$, $s_I(t)$ (in each time window $[t + \Delta t)$, the resulting spike count is a random number drawn from a Poisson distribution with mean and variance
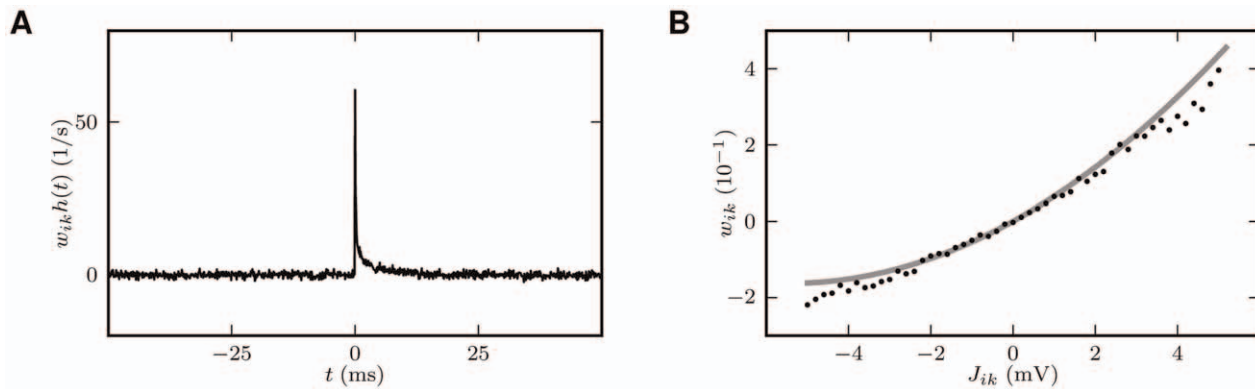
Figure 8. Linear response and relation between synaptic weight J and effective coupling strength w. A: Firing-rate deflection $w_{ik}h(t)$ of a LIF neuron caused by an incoming spike event of postsynaptic amplitude $J_{ik} = 0.6$ mV. **B**: Integral $w_{ik} = w_{ik} \int_0^\infty h(t)\, dt$ of the firing rate deflection shown in A as a function of the postsynaptic amplitude J_{ik} (simulation: black dots; analytical approximation (45) : gray curve). The neuron receives constant synaptic background input with $J = 0.1$ mV, $g = 4$, and rates $\nu_E = 5960\,\frac{1}{s}$, $\nu_I = 1190\,\frac{1}{s}$ resulting in a first and second moment (42) $\bar{\mu}_i = 12$ mV and $\bar{\sigma}_i = 5$ mV. Simulation results are obtained by averaging over 1000 trials of 10 s duration each with 25000 input impulses on average. For further parameters of the neuron model, see Table 1 and Table 2.
doi:10.1371/journal.pcbi.1002596.g008

proportional to $s_E(t)$ and $s_I(t)$, respectively). Although the compound spectrum of the resulting local input is barely distinguishable from the compound spectrum of the intact recurrent system (Fig. 9 D; black and dark gray curves), the response spectra are very different: replacing the feedback ensemble by inhomogeneous Poisson processes leads to a substantial amplification of low-frequency fluctuations (Fig. 9 E; compare black and dark gray

curves). The effect is as strong as if the temporal structure of the population rates was completely ignored, i.e. if the feedback channels were replaced by realizations of homogeneous Poisson processes with constant rates (PoissH; light gray curves in Fig. 9 D,E). This result indicates that the precise temporal structure of the population rates is essential and that even small deviations can significantly weaken the fluctuation-suppression mechanism. The
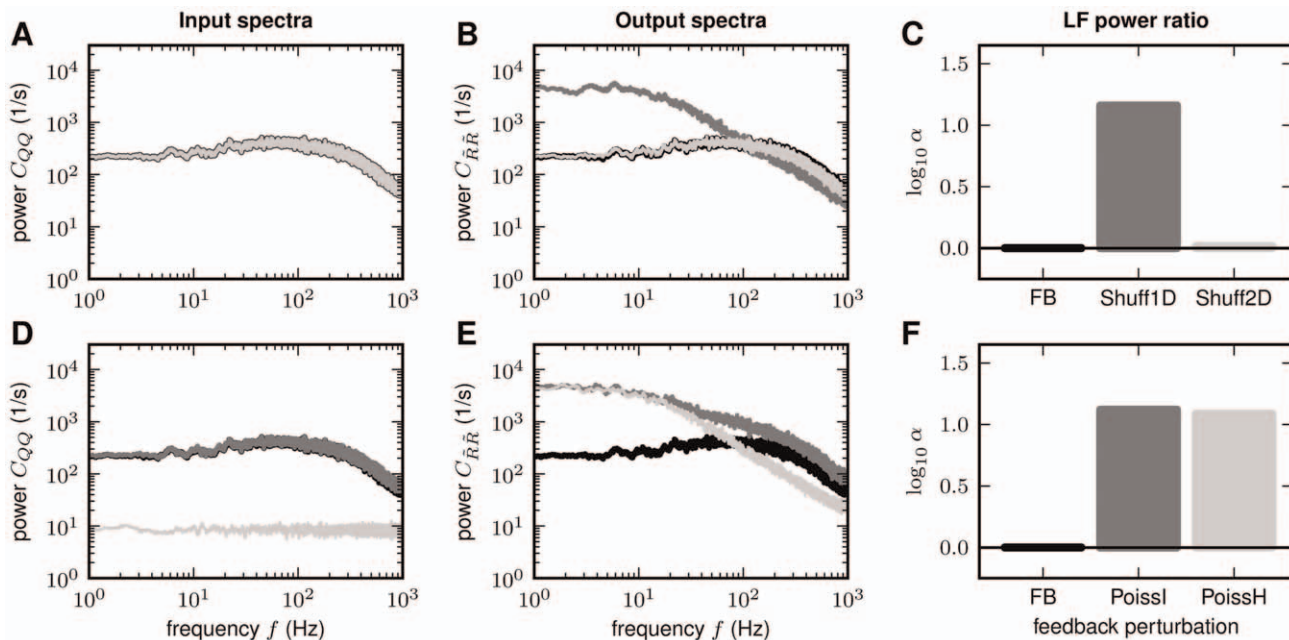
Figure 9. Amplification of population-rate fluctuations by different types of feedback manipulations in a random network of excitatory and inhibitory LIF neurons (simulation results). Top row (**A–C**): Unperturbed feedback (FB; black), shuffling of spike-train senders across entire network (Shuff1D; dark gray) and within each subpopulation (E,I) separately (Shuff2D; light gray). Bottom row (**D–F**): Unperturbed feedback (FB; black), replacement of spike trains by realizations of inhomogeneous (PoissI; dark gray) and homogeneous Poisson processes (PoissH; light gray). In the PoissI (PoissH) case, the (time averaged) subpopulation rates are approximately preserved. **A, D**: Compound power-spectra C_{QQ} of input spike-train ensembles. **B, E**: Power-spectra $C_{\tilde{R}\tilde{R}}$ of population-response rates. **C, F**: Low-frequency (LF; 1–20 Hz) power ratio α (increase in LF power relative to the unperturbed case [FB]; logarithmic scaling). Note that in A, the compound-input spectra (FB, Shuff1D, Shuff2D) are identical. In D, the input spectra for the intact recurrent network (FB) and the inhomogeneous-Poisson case (PoissI) are barely distinguishable. See Table 1 and Table 2 for details on the network model and parameters. Simulation time $T = 100$ s. Single-trial spectra smoothed by moving average (frame size 1 Hz).
doi:10.1371/journal.pcbi.1002596.g009

results of the Poisson experiments can be understood by considering the effect of the additional noise caused by the stochastic realization of individual spikes. Considering the auto-correlation, a Poisson spike-train ensemble with rate profile $v(t)$ is equivalent to a sum of the rate profile and a noise term resulting from the stochastic (Poissonian) realization of spikes, $\boldsymbol{q}(t) = v(t)\boldsymbol{1} + \sqrt{v_0}\boldsymbol{z}(t)$. Here, $\boldsymbol{z}(t)$ denotes a Gaussian white noise with auto-correlation $\mathrm{E}_t[\boldsymbol{z}(t)\boldsymbol{z}(t+s)] = \boldsymbol{1}\delta(s)$ and $v_0 = \mathrm{E}_t[v(t)]$ the mean firing rate. The response fluctuations of the population driven by the rate modulated Poisson activity are, to linear approximation, given by $\tilde{\boldsymbol{R}} = H(\boldsymbol{WQ} + \boldsymbol{X})$. Inserting $\boldsymbol{Q}$, we obtain an additional noise term $v_0|H|^2\boldsymbol{WW}^T$ in the spectrum $\boldsymbol{C}_{\tilde{R}\tilde{R}} = \tilde{\boldsymbol{R}}\tilde{\boldsymbol{R}}^*$ which explains the increase in power compared to the spectrum $\boldsymbol{C}_{RR}$ of the recurrent network. As a generalization of the Poisson model, one may replace the noise amplitude $\sqrt{v_0}$ by some arbitrary prefactor η. In simulation experiments, we observed a gradual amplification of the population-rate fluctuations with increasing noise amplitude η (data not shown).

Discussion

We have shown that negative feedback in recurrent neural networks actively suppresses low-frequency fluctuations of the population activity and pairwise correlations. This mechanism allows neurons to fire more independently than expected given the amount of shared presynaptic input. We demonstrated that manipulations of the feedback statistics, e.g. replacing feedback by uncorrelated feedforward input, can lead to a significant amplification of response correlations and population-rate fluctuations.

The suppression of correlations and population-rate fluctuations by feedback can be observed in networks with both purely inhibitory and mixed excitatory-inhibitory coupling. In purely inhibitory networks, the effect can be understood by studying the role of the effective negative feedback experienced by the compound activity. In networks of excitatory and inhibitory neurons, a change of coordinates, technically a Schur decomposition, exposes the underlying feedback structure: the sum of the excitatory and inhibitory activity couples negatively to itself if the network is in an inhibition dominated regime (which is required for its stability; see, e.g., [12]. This negative feedback suppresses fluctuations in a similar way as in purely inhibitory networks. The fluctuation suppression becomes more efficient the further the network is brought into the inhibition dominated regime, away from the critical point of equal recurrent excitation and inhibition ($\bar{g} = 1$). Having identified negative feedback as the underlying cause of small fluctuations and correlations, we can rule out previous explanations based on a balance between (correlated) excitation and inhibition [17]. We presented a self-consistent theory for the average pairwise spike-train correlations which illuminates that the suppression of population-rate fluctuations and the suppression of pairwise correlations are two expressions of the same effect: as the single spike-train auto-covariance is the same in the feedforward and the feedback case, the suppression of population-rate fluctuations implies smaller correlations. Our theory enables us to identify the cancellation of input correlations as a hallmark of small spike-train correlations.

In previous studies, shared presynaptic input has often been considered a main source of correlation in recurrent networks (e.g. [15,52]). Recently [17], suspected that correlations between excitatory and inhibitory neurons and the fast tracking of external input by the excitatory and the inhibitory population are responsible for an active decorrelation. We have demonstrated here that the mere fact that excitatory and inhibitory neurons are

correlated is not sufficient to suppress shared-input correlations. Rather, we find that the spike-train correlation structure in networks of excitatory and inhibitory networks arranges such that their overall contribution to the covariance between the summed inputs to a pair of neurons becomes negative, canceling partly the effect of shared inputs. This cancellation becomes more precise the stronger the negative compound feedback $Kw(1 - \gamma g)$ is. In homogeneous networks where excitatory and inhibitory neurons receive statistically identical input, the particular structure of correlations is $C_{\mathrm{EE}} > C_{\mathrm{EI}} > C_{\mathrm{II}}$. It can further be shown that this structure of correlations is preserved in the limit of large networks $N \to \infty$ ($K/N = \mathrm{const.}$). For non-homogeneous synaptic connectivity, if the synaptic amplitudes depend on the type of the target neuron (i.e. $J_{\mathrm{EE}} \neq J_{\mathrm{IE}}$ or $J_{\mathrm{EI}} \neq J_{\mathrm{II}}$), the structure of correlations may be different. Still, the correlation structure arranges such that shared input correlation is effectively suppressed. Formally, this can be seen from a self-consistency equation similar to our equation (80).

The study by [17] has shown that correlations are suppressed in the limit of infinitely large networks of binary neurons receiving randomly drawn inputs from a common external population. Its argument rests on the insight that the population-activity fluctuations in a recurrent balanced network follow the fluctuations of the external common population. An elegant scaling consideration for infinitely large networks $N \to \infty$ with vanishing synaptic efficacy $\propto 1/\sqrt{N}$ shows that this fast tracking becomes perfect in the limit. This allows to determine the zero-lag pairwise correlations caused by the external input. The analysis methods and the recurrent networks presented here differ in several respects from these previous results: We study networks of a finite number of spiking model neurons. The neurons receive uncorrelated external input, so that correlations are due to the local recurrent connectivity among neurons, not due to tracking of the common external input [17]. Moreover, we consider homogeneous connectivity where synaptic weights depend only on the type of the presynaptic neuron (as, e.g., in [12]), resulting in a correlation structure $C_{\mathrm{EE}} > C_{\mathrm{EI}} > C_{\mathrm{II}}$. For such connectivity, networks of binary neurons with uncorrelated external input exhibit qualitatively the same correlation structure as reported here (results not shown).

In purely inhibitory networks, the decorrelation occurs in an analog manner as in excitatory-inhibitory networks. As only a single population of neurons is available here, population averaged spike-train correlations C_{II} are negative. This negative contribution compensates the positive contribution of shared input.

The structure of integrated spike-train covariances in networks constitutes an experimentally testable prediction. Note, however, that the prediction (19) obtained in the current work rests on two simplifying assumptions: identical internal dynamics of excitatory and inhibitory neurons and homogeneous connectivity (i.e. $J_{\mathrm{EE}} = J_{\mathrm{IE}}$, $J_{\mathrm{EI}} = J_{\mathrm{II}}$; see **"Results: Population-activity fluctuations in excitatory-inhibitory networks"**). For such networks, the structure of correlations is given by $C_{\mathrm{EE}} > C_{\mathrm{EI}} > C_{\mathrm{II}}$. Further, the relation between subthreshold membrane-potential fluctuations and spike responses is the same for both neuron types. Consequently, the above correlation structure can be observed not only at the level of spike trains but also for membrane potentials, provided the assumptions hold true. A recent experimental study [53] reports neuron-type specific cross-correlation functions in the barrel cortex of behaving mice, both for spike trains and membrane potentials. It is however difficult to assess the integral correlations from the published data. A direct test of our predictions requires either a reanalysis of the data or a theory predicting the entire correlation functions. The raw (unnorma-

lized) II and EI spike-train correlations in [53] are much more pronounced than the EE correlations (Fig. 6 in [53]). This seems to be in contradiction to our results. Note, however, that the firing rates of excitatory and inhibitory neurons are very different in [53]. In our study, in contrast, the average firing rates of excitatory and inhibitory neurons are identical as a consequence of the assumed network homogeneity. Future theoretical work is needed to generalize our model to networks with heterogeneous firing rates and non-homogeneous connectivity. Recent results on the dependence of the correlation structure on the connectivity may prove useful in this endeavor [54–56].

Correlations in spike-train ensembles play a crucial role for the en- and decoding of information. A set of uncorrelated spike trains provides a rich dynamical basis which allows readout neurons to generate a variety of responses by tuning the strength and filter properties of their synapses [1]

In the presence of correlations, the number of possible readout signals is limited. Moreover, spike-train correlations impair the precision of such readout signals in the presence of noise. Consider, for example, a linear combination $y(t) = \sum_{i=1}^{N} (s_i * h_i)(t)$ of N presynaptic spike trains with arbitrary (linear) filter kernels $h_i(t)$ (e.g. synaptic filters). In a realistic scenario, the individual spike trains $s_i(t)$ typically vary across trials [3,57]. To understand how robust the resulting readout signal $y(t)$ is against this spike-train variability, let's consider the variability of its Fourier transform $Y(\omega) = F[y(t)](\omega) = \sum_{i=1}^{N} S_i(\omega) H_i(\omega)$. Assuming homogeneous spike-train statistics,

$$S(\omega) := E_{S_i(\omega)}[\quad](\forall i) \qquad \text{(mean)}$$
$$V(\omega) := E[|S_i(\omega) - S(\omega)|^2](\forall i) \qquad \text{(variance)}, \quad (24)$$
$$C(\omega) := E[(S_i(\omega) - S(\omega))(S_j(\omega) - S(\omega))^*](\forall i, j \neq i) \quad \text{(covariance)}$$

the (squared) signal-to-noise ratio of the readout signal $Y(\omega)$ is given by

$$\begin{aligned} SN^2(\omega) :&= \frac{|E_{Y(\omega)}[Y(\omega)]|^2}{E[|Y(\omega) - E[Y(\omega)]|^2]} \\ &= \frac{|S|^2 |\bar{H}_1|^2}{N^{-1}(1-\kappa)V\bar{H}_2 + \kappa V |\bar{H}_1|^2}. \end{aligned} \qquad (25)$$

Here, $E[\ldots]$ denotes the average across the ensemble of spike-train realizations, $\kappa(\omega) = C(\omega)/V(\omega)$ the spike-train coherence, and the coefficients $\bar{H}_1 := E_i[H_i]$ and $\bar{H}_2 := E_i[|H_i|^2]$ the 1st- and 2nd-order filter statistics. For uncorrelated spike trains, i.e. $\kappa(\omega) = 0$, and $S(\omega) \neq 0$, the signal-to-noise ratio SN^2 grows unbounded with the population size N. Thus, even for noisy spike trains ($V > 0$), the compound signal $y(t)$ can be highly reliable if the population size N is sufficiently large. In the presence of correlations, $\kappa(\omega) \neq 0$, however, SN^2 converges towards a constant value $\kappa^{-1}|S|^2 V^{-1}$ as N grows. Even for large populations, the readout signal remains prone to noise. These findings constitute a generalization of the results reported for population-rate coding, i.e. sums of unweighted spike counts (see, e.g., [2,3]). The above arguments illustrate that the same reasoning applies to coding schemes which are based on the spatio-temporal structure of spike patterns.

In a previous study [9], we demonstrated that active decorrelation in recurrent networks is a necessary prerequisite for a controlled propagation of synchronous volleys of spikes in embedded feedforward subnetworks ('synfire chains'; Fig. 10): A synfire chain receiving background input from a finite population

of independent Poisson sources amplifies the resulting shared-input correlations, thereby leading to spontaneous synchronization within the chain (Fig. 10 B). A distinction between these spurious synchronous events and those triggered by an external stimulus is impossible. The synfire chain loses its asynchronous ground state [58]. A synfire chain receiving background inputs from a recurrent network, in contrast, is much more robust. Here, shared-input correlations are actively suppressed by the recurrent-network dynamics. Synchronous events can be triggered by external stimuli in a controlled manner (Fig. 10 A). Apart from the spontaneous synchronization illustrated in Fig. 10, decorrelation by inhibition might solve another problem arising in embedded synfire structures: In the presence of feedback connections between the synfire chain and the embedding background network, synchronous spike volleys can excite (high-frequency) oscillatory modes in the background network which, in turn, interfere with the synfire dynamics and prevent a robust propagation of synchronous activity within the chain ('synfire explosion'; see [59,60]). The decorrelation mechanism we refer to in our work is efficient only at low frequencies. It cannot prevent the build-up of these oscillations. [61] demonstrated that the 'synfire explosion' can be suppressed by adding inhibitory neurons to each synfire layer ('shadow inhibition') which diffusely project to neurons in the embedding network, thereby weakening the impact of synfire activity on the embedding network.

In the present work we focus on the integral of the correlation function, nurtured by our interest in the low-frequency fluctuations. An analog treatment can however easily be performed for the zero-lag correlations. In contrast to infinite networks with sparse connectivity ($N \to \infty$, $K = \text{const}$), in the case of finite networks, pairs of neurons must be distinguished according to whether they are synaptically connected or not in order to arrive at a self-consistent theory for the averaged correlations. Providing explicit expressions for correlations between connected and unconnected neurons, the current work provides the tools to relate experimentally observed spiking correlations to the underlying synaptic connectivity.

The quantification of pairwise correlations is a necessary prerequisite to understand how correlation sensitive synaptic plasticity rules, like spike-timing dependent plasticity [62], interact with the recurrent network dynamics [63]. Existing theories quantifying correlations employ stochastic neuron models and are limited to purely excitatory networks [63–65]. Here, we provide an analytical equivalence relation between a reduced linear model and spiking integrate-and-fire neurons describing fluctuations correctly up to linear order. A formally similar approach has been employed earlier to study delayed cumulative inhibition in spiking networks [66]. We show that the correlations observed in recurrent networks in the asynchronous irregular regime are quantitatively captured for realistic synaptic coupling with postsynaptic potentials of up to about 1 mV. The success of this approach can be explained by the linearization of the neural threshold units by the afferent noise experienced in the asynchronous regime. For linear neural dynamics, the second-order description of fluctuations is closed [67]. We exploit this finding by applying perturbation theory to the Fokker-Planck description of the integrate-and-fire neuron to obtain the linear input-output transfer at low frequencies [33], thereby determining the effective coupling in our linear model.

The scope of the theory presented in the current work is limited mainly by three assumptions. The first is the use of a linear theory which exhibits an instability as soon as a single eigenvalue of the effective connectivity matrix assumes a positive real part. This ultimately happens when increasing the synaptic coupling
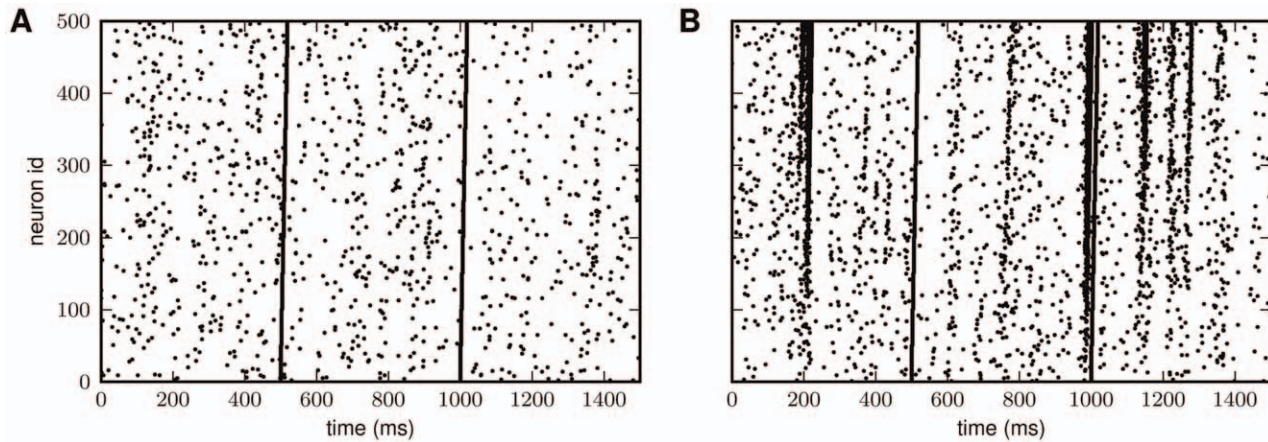
Figure 10. Recurrent network dynamics stabilizes dynamics of embedded synfire chains. Spiking activity in a synfire chain ($L=10$ layers, layer width $b=50$) receiving background input from an excitatory-inhibitory network (**A**, cf. Fig. 1 C) or from a finite pool of excitatory and inhibitory Poisson processes (**B**, cf. Fig. 1 D). Average input firing rates, in-degrees and amount of shared input are identical in both cases. Neurons of the first synfire layer (neuron ids $1,...,b$) are stimulated by current pulses at times $t=500$ and 1000 ms. Each neuron in layer $k \in [2,L]$ receives inputs from all b neurons in the preceding layer $k-1$ (synaptic weights $J_{sfc}=0.8$ mV, spike transmission delays $d_{sfc}=2$ ms), and $K-b$ and γK excitatory and inhibitory background inputs, respectively, randomly drawn from the presynaptic populations. Neurons in the first layer $k=1$ receive K and γK excitatory and inhibitory background inputs, respectively. Note that there is no feedback from the synfire chain to the embedding network. See Table 2 for network parameters.
doi:10.1371/journal.pcbi.1002596.g010

strength, because the eigenvalues of the random connectivity matrix are located in a circle centered in the left half of the complex plain with a radius given by the square root of the variance of the matrix elements [68,69]. Nonlinearities, like those imposed by strictly positive firing rates, prevent such unbounded growth (or decay) by saturation. For nonlinear rate models with sigmoidal transfer functions it has been shown that the activity of recurrent random networks of such units makes a transition to chaos at the point where the linearized dynamics would loose stability [70]. However, this point of transition is sharp only in the limit of infinitely large networks. From the population averaged firing rate and the pairwise correlations averaged over pairs of neurons considered in Fig. 7 we cannot conclude whether or not a transition to chaos occurs in the spiking network. In simulations and in the linearized reduced model, we could however observe that the distribution of pairwise correlations broadens when approaching the point of instability. Future work needs to examine this question in detail, e.g. by considering measures related to the Lyapunov exponent. Recently developed semi-analytical theories accounting for nonlinear neural features [71] may be helpful to answer this question. The second limiting factor of the current theory is the use of a perturbative approach to quantify the response of the integrate-and-fire model. Although the steady-state firing rate of the network is found as the fixed point of the nonlinear self-consistency equation, the response to a synaptic fluctuation is determined up to linear order in the amplitude of the afferent rate fluctuation, which is only valid for sufficiently small fluctuations. For larger input fluctuations, nonlinear contributions to the neural response can become more important [33]. Also for strong synaptic coupling, deviations from our theory are to be expected. Thirdly, the employment of Fokker-Planck theory to determine the steady-state firing rate and the response to incoming fluctuations assumes uncorrelated presynaptic firing with Poisson statistics and synaptic amplitudes which are vanishingly small compared to the distance between reset and threshold. For larger synaptic amplitudes, the Fokker-Planck theory becomes approximate and deviations are expected [33,34,72,73]. This can be observed in Fig. 7 A, showing a deviation between the self-

consistent firing rate and the analytical prediction at about $J \simeq 1$ mV. In this work, we obtained a sufficiently precise self-consistent approximation of the correlation coefficient C/A by relating the random recurrent network of spiking neurons in the asynchronous irregular state to a reduced linear model which obeys the same relation between C and A up to linear order. This reduced linear model, however, does not predict the absolute values of the variance A and covariance C. The variance A of the LIF model, for example, is dominated by nonlinear effects, such as the reset mechanism after each action potential. Previous work [12,31] has shown that the single spike-train statistics can be approximated in the diffusion approximation if the recurrent firing rate in the network is determined by mean-field theory. One may therefore extend our approach and determine the integral auto-correlation function as $A = \nu F$ with the Fano factor F (see [51]). For a renewal process and long observation times, the Fano factor is given by $F = CV^2$ [74,75]. The coefficient of variation CV can be obtained from the diffusion approximation of the membrane-potential dynamics (App. A.1 in [12]). The covariance C can then be determined by (19). Another possibility is the use of a refractory-density approach [76,77].

The spike-train correlation as a function of the time lag is an experimentally accessible measure. Future theoretical work should therefore also focus on the temporal structure of correlations in recurrent networks, going beyond zero-lag correlations [15,17] and the integral measures studied in the current work. This would allow to compare the theoretical predictions to direct experimental observations in a more detailed manner. Moreover, the relative spike timing between pairs of neurons is a decisive property for Hebbian learning [78] in recurrent networks, as implemented by spike timing-dependent plasticity [62], and suspected to play a role for synapse formation and elimination [79].

The simulation experiments performed in this work revealed that the suppression of correlations is vulnerable to certain types of manipulations of the feedback loop. One particular biological source of additional variability in the feedback loop is probabilistic vesicle release at synapses [80]. In feedforward networks, such unreliable synaptic transmission has been shown to decrease the

transmission of correlations by pairs of neurons [22]. Stochastic synaptic release is very similar to the replacement of the population activity in the feedback branch by a rate modulated Poisson processes that conserves the population rate. In these simulations we observed an increase of correlations due to the additional noise caused by the stochastic Poisson realization. Future work should investigate more carefully which of the two opposing effects of probabilistic release on correlations dominates in recurrent networks.

The results of our study do not only shed light on the decorrelation of spiking activity in recurrent neural networks. They also demonstrate that a standard modeling approach in theoretical neuroscience is problematic: When studying the dynamics of a local neural network (e.g. a "cortical column"), it is a common strategy to replace external inputs to this neural population $\mathcal{P}$ by spike-train ensembles with some predefined statistics, e.g. by stationary Poisson processes. Most neural systems, however, exhibit a high degree of recurrence. Nonlocal input to the population $\mathcal{P}$, i.e. input from other brain areas, therefore has to be expected to be shaped by the activity within $\mathcal{P}$. The omission of these feedback loops can lead to qualitatively wrong predictions of the population statistics. The analytical results for the correlation structure of recurrent networks presented in this study provide the means to a more realistic specification of such external activity.

Methods

LIF network model

In the present study, we consider two types of sparsely connected random networks: networks with purely inhibitory coupling ("I networks") and networks with both excitatory and inhibitory interactions ("EI networks"). To illustrate the main findings of this study and to test the predictions of the linear model described in **"Methods: Linearized network model"**, both architectures were implemented as networks of leaky integrate-and-fire (LIF) neurons. The model details and parameters are reported in Table 1 and Table 2, respectively. All network simulations were carried out with NEST (www.nest-initiative.org, [81]).

Linearized network model

In this section we show how the dynamics of the spiking network can be reduced to an effective linear model with fluctuations fulfilling, by construction, the same relationship as the original system up to linear order. We first outline the conceptual steps of this reduction, and then provide the formal derivation.

We make use of the observation that the effect of a single synaptic impulse on the output activity of a neuron is typically small. Writing the response spike train of a neuron as a functional of the history of all incoming impulses therefore allows us to perform a linearization with respect to each of the afferent spike trains. Formally, this corresponds to a Volterra expansion up to linear order, the generalization of a Taylor series to functionals. In **"Methods: Response kernel of the LIF model"**, we perform this linearization explicitly for the example of the LIF model. This determines how the linear response kernel depends on the parameters of the LIF model. The linear dependence on the input leads to an approximate convolution equation (31) linearly connecting the auto- and the cross-correlation functions in the network. As this equation is complicated to solve directly, we introduce a reduced linear model (35) obeying the same convolution equation. The reduced linear model can be solved by standard Fourier methods and yields an explicit form for the covariance matrix in the frequency domain (37). The diagonal and off-diagonal elements of the $N = N_\mathrm{E} + N_\mathrm{I}$ dimensional covariance

matrix $\boldsymbol{C}(\omega)$ in (56) correspond to the power-spectra of individual neurons and the cross-spectra of individual neuron pairs, respectively. As, in this linear approximation, both the auto- and the cross-covariances are proportional to the variance of the driving noise, the resulting correlation coefficients are independent of the noise amplitude (see **"Methods: Population averaged correlations in the linear EI network"**). As shown in **"Results: Suppression of population-activity fluctuations by negative feedback"** and **"Results: Population-activity fluctuations in excitatory-inhibitory networks"**, the suppression of fluctuations in recurrent networks is most pronounced at low frequencies. It is therefore sufficient to restrict the discussion to the zero-frequency limit $\omega \to 0$. Note that the zero-frequency variances and covariances correspond to the integrals of the auto- and cross-correlation functions in the time domain. In this limit, we may combine the two different sources of fluctuations caused by the spiking of the neurons and by external input to the network into a single source of white noise with variance ρ^2 (see (39)).

In general, the spiking activity $s_i(t)$ of neuron i at time t is determined by the entire history $\{s(t')|t' < t\}$ of the activity of all neurons $s = (s_1, \ldots, s_N)$ in the network up to time t. Formally, this dependence can be expressed by a functional

$$s_i(t) = G_t^i[s(t')]. \tag{26}$$

The subscript t in G_t^i indicates that $t' < t$ (causality). In the following, we use the abbreviation $G_t^i[s] \equiv G_t^i[s(t')]$. The effect of a single synaptic input on the state of a neuron is typically small. We therefore approximate the influence of an incoming spike train on the activity of the target neuron up to linear order. The sensitivity of neuron i's activity to the input from neuron k can be expressed by the functional derivative of G_t^i with respect to input spike train s_k:

$$\frac{\delta G_t^i[s(t')]}{\delta s_k(t'')} = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left(G_t^i[s(t') + \epsilon\,\delta(t' - t'')\boldsymbol{e}_k] - G_t^i[s(t')] \right). \tag{27}$$

It represents the response of the functional to a single δ-shaped perturbation in input channel k at time t'', normalized by the perturbation amplitude ϵ. In (27), $\boldsymbol{e}_k = (0, \ldots, 0, 1, 0, \ldots, 0)$ denotes the unity vector with elements $e_{k_k} = 1$ and $e_{k_i} = 0$ for all $i \neq k$. By introducing the vector $s_{\hat{k}}(t) = (s_1(t), \ldots, s_{k-1}(t), 0, s_{k+1}(t), \ldots, s_N(t))$ of spike trains with the k-th component set to zero, $G_t^i[s]$ can be approximated by

$$G_t^i[s] \simeq \sum_{k=1}^N \int_{-\infty}^t \frac{\delta G_t^i[s_{\hat{k}}]}{\delta s_k(t'')}\, s_k(t'')\, dt''. \tag{28}$$

Eq. (28) is a Volterra expansion up to linear order, the formal extension of a Taylor expansion of a function of N variables to a functional, truncated after the linear term. With the linearized dynamics (28), the pairwise spike-train cross-correlation function between two neurons i and $j \neq i$ is given by

$$
\begin{aligned}
c_{ij}(\tau) &= \langle s_i(t+\tau)\tilde{s}_j(t)\rangle_\mathbf{s} \\
&= \langle G_{t+\tau}^i[s]\tilde{s}_j(t)\rangle_\mathbf{s} \\
&= \sum_{k=1}^N \int_{-\infty}^{t+\tau} \langle \frac{\delta G_{t+\tau}^i[s_{\hat{k}}]}{\delta s_k(t'')} \langle s_k(t'')\tilde{s}_j(t)\rangle_{s_k}\rangle_{s\backslash s_k}\, dt'' \qquad (\forall \tau > 0).
\end{aligned}
\tag{29}
$$

Table 1. LIF network: Model overview.

A Model summary	
Populations	one (inhibitory network) or two (excitatory-inhibitory network)
Connectivity	random, fixed in-degrees
Neuron	leaky integrate-and-fire (LIF)
Synapse	current based, delta-shaped postsyn. currents with constant amplitudes
Input	uncorrelated Gaussian white noise currents

B Populations

Inhibitory network

Name	**Elements**	**Size**
I	LIF	$N = K/\epsilon$

Excitatory-inhibitory network

Name	**Elements**	**Size**
E	LIF	$N_\mathrm{E} = K/\epsilon$
I	LIF	$N_\mathrm{I} = \gamma N_\mathrm{E} = \gamma K/\epsilon$

C Connectivity

Inhibitory network

Source	**Target**	**Pattern**
I	I	random convergent $K \rightarrow 1$, delay d, weight $-J$

Excitatory-inhibitory network

Source	**Target**	**Pattern**
E	E	random convergent $K \rightarrow 1$, delay d, weight J
E	I	random convergent $K \rightarrow 1$, delay d, weight J
I	E	random convergent $\gamma K \rightarrow 1$, delay d, weight $-gJ$
I	I	random convergent $\gamma K \rightarrow 1$, delay d, weight $-gJ$

D Neuron

Type	**Leaky integrate-and-fire (LIF, [86])**
Description	**Dynamics of membrane potential $V_i(t)$ ($i \in [1,N]$):**
	- Spike emission at times t_k^i with $V_i(t_k^i) \geq \theta$
	- Subthreshold dynamics:
	- $\tau_\mathrm{m} \dot{V}_i = -V_i + R_\mathrm{m} I(t)$ if $\forall k:\ t \notin (t_k^i, t_k^i + \tau_\mathrm{ref}]$
	- Reset + refractoriness: $V_i(t) = V_\mathrm{reset}$ if $\forall k:\ t \in (t_k^i, t_k^i + \tau_\mathrm{ref}]$
	Exact integration [87] with temporal resolution $\mathrm{d}t$
	Initial membrane-potential distribution at $t = 0$: uniform between 0 and θ

E Synapse

Type	**Current synapse with δ-shaped postsyn. currents (PSCs)**
Description	**Input current of neuron i: $I_i(t) = C_m \sum_j J_{ij} \sum_l \delta(t - t_l^i - d) + I_{i,\mathrm{ext}}(t)$**
	Static synaptic weights J_{ij} (see Connectivity)

F Input

Type	**uncorrelated Gaussian white noise $R_\mathrm{m} I_{i,\mathrm{ext}}(t) = \mu_\mathrm{ext} + \sqrt{\tau_\mathrm{m}}\, \xi_i(t)$ (for $i \in [1,N]$)**
Description	**mean $\mu_\mathrm{ext} = R_\mathrm{m} \mathrm{E}_t[I_{i,\mathrm{ext}}(t)]$, auto-correlation $R_\mathrm{m}^2 \mathrm{E}_t[I_{i,\mathrm{ext}}(t)I_{j,\mathrm{ext}}(t+\tau)] = \mu_\mathrm{ext}^2 + \eta^2 \tau_\mathrm{m} \delta_{ij} \delta(\tau)$**
	in discrete time $t \in \{n \cdot \mathrm{d}t \mid n \in \mathbb{N}\}$, $\xi(n \cdot \mathrm{d}t)$ piecewise constant within time interval $\mathrm{d}t$, value drawn independently for each time point from a normal distribution with zero mean and standard deviation $1/\sqrt{\mathrm{d}t}$

Note that (29) is valid only for positive time lags $\tau > 0$, because for $\tau < 0$ a possible causal influence of s_i on s_j is not expressed by the functional. Here, $\langle \cdot \rangle_s$ denotes the average across the ensemble of realizations of spike trains in the stationary state of the network (e.g. the ensemble resulting from different initial conditions), and $\tilde{s}(t) = s(t) - \langle s \rangle_\mathbf{s}$ the centralized (zero mean) spike train. In the last line in (29), the average $\langle \cdot \rangle_s = \langle \langle \cdot \rangle_{s_k} \rangle_{s \backslash s_k}$ is split into the average $\langle \cdot \rangle_{s \backslash s_k}$ across all realizations of spike trains excluding s_k and the average $\langle \cdot \rangle_{s_k}$ across all realizations of s_k. Note that the latter does not affect the functional derivative because it is, by construction,

Table 2. LIF network: Parameters (default values).

A   Connectivity

Name	Value	Description
K	1250 (inhibitory network)	in-degree
	1000 (E-I network)	excitatory in-degree
ϵ	0.1	network connectivity
$\gamma = N_I/N_E$	1/4 (E-I network)	relative size of inhibitory subpopulation

B   Neuron

Name	Value	Description
R_m	80 MΩ	membrane resistance
τ_m	20 ms	membrane time constant
τ_{ref}	2 ms	refractory period
V_{reset}	0 mV	reset potential
θ	15 mV	spike threshold

C   Synapse

Name	Value	Description
J	0.2 mV	EPSP amplitude
g	6 (E-I network)	relative IPSP amplitude
d	0.1 ms	synaptic delay

D   Input

Name	Value	Description
μ_{ext}	1.5θ	mean external GWN input
η	0.3θ	SD of external GWN input

E   Simulation

Name	Value	Description
T	100 s 10 or	simulation time
dt	0.1 ms	time resolution

independent of the actual realization of s_k. A consistent approximation up to linear order is equivalent to the assumption that for all j the linear dependence of the functional on s_j is completely contained in the respective derivative with respect to s_j (28). Dependencies beyond linear order include higher-order derivatives and are neglected in this approximation. This is equivalent to neglecting the dependence of $\frac{\delta G^i_{t+\tau}[s_k]}{\delta s_k(t'')}$ on s_j for any $j \neq k$. Hence, we can average the inner term over s_k and s_j separately. In the stationary state, this correlation can only depend on $t'' - t$ and equals the auto- or cross-correlation function:

$$\langle s_k(t'')\tilde{s}_j(t)\rangle_{s_k,s_j} = \begin{cases} a_k(t''-t) & \text{for } k=j \\ c_{kj}(t''-t) & \text{for } k \neq j. \end{cases}$$

The pairwise spike-train correlation function is therefore given by

$$c_{ij}(\tau) = \sum_{k=1}^{N} \int_{-\infty}^{t+\tau} dt'' \left\langle \frac{\delta G^i_{t+\tau}[s_k]}{\delta s_k(t'')} \right\rangle_s \begin{cases} a_k(t''-t) & \text{for } k=j \\ c_{kj}(t''-t) & \text{for } k \neq j \end{cases} \quad (\forall \tau > 0),$$

where we used the fact that $\langle f[s_k]\rangle_{s \setminus s_k} = \langle f[s_k]\rangle_s$ for any functional f that does not depend on s_k. The average of the functional

derivative has the intuitive meaning of a response kernel with respect to a δ-shaped perturbation of input s_k at time t''. Averaged over the realizations of the stationary network activity this response can only depend on the relative time $t+\tau-t''$. In a homogeneous random network, the input statistics (number of synaptic inputs and synaptic weights) and the parameters of the internal dynamics are identical for each cell, so that the temporal shape $h(t)$ of the response kernel can be assumed to be the same for all neurons. The synaptic coupling strength from neuron k to neuron i determines the prefactor w_{ik}:

$$w_{ik}h(t+\tau-t'') \equiv \left\langle \frac{\delta G^i_{t+\tau}[s_k]}{\delta s_k(t'')} \right\rangle_s. \quad (30)$$

In this notation, the linear equation connecting the auto-correlations a_k and the cross-correlations c_{ij} takes the form

$$c_{ij}(\tau) = \sum_{k=1}^{N} w_{ik} \int_{-\infty}^{\tau} dt\, h(\tau-t) \begin{cases} a_k(t) & \text{for } k=j \\ c_{kj}(t) & \text{for } k \neq j \end{cases} \quad (\forall \tau > 0). \quad (31)$$

Eq. (31) can be solved numerically or by means of Wiener-Hopf theory taking the symmetry $c_{ij}(\tau) = c_{ji}(-\tau)$ into account [82].

Our aim is to find a simpler model which is equivalent to the LIF dynamics in the sense that it fulfills the same equation (31). Let's $u(t)$ denote the vector of dynamic variables of this reduced model. Analog to the original model, we define the cross-correlation for $i \neq j$ and $\tau > 0$ as

$$c_{ij}^u(\tau) = \langle u_i(t+\tau)\tilde{u}_j(t)\rangle_u$$
$$= \langle L_{t+\tau}^i[u]\tilde{u}_j(t)\rangle_u. \qquad (32)$$

The simplest functional $L_t^i[u]$ consistent with equation (31) is linear in u. Since we require equivalence only with respect to the ensemble averaged quantities, i.e. $c_{ij}^u(\tau) = c_{ij}(\tau)$, the reduced activity and therefore $L_t^i[u]$ can contain a stochastic element which would disappear after averaging. The linear functional

$$u_i(t) = L_t^i[u] = \sum_{k=1}^N w_{ik}\int_{-\infty}^t h(t-t')u_k(t')dt' + z_i(t) \qquad (33)$$

with a pairwise uncorrelated, centralized white noise $z_i(t)$ ($\langle z_i(t+\tau)z_k(t)\rangle_z = \rho_z^2\delta_{ik}\delta(\tau)$) fulfills the requirement, since for $\tau > 0$ and $i \neq j$

$$c_{ij}^u(\tau) = \langle u_i(t+\tau)\tilde{u}_j(t)\rangle_z = \sum_{k=1}^N w_{ik}\int_{-\infty}^{t+\tau} h(t+\tau-t')$$

$$\langle u_k(t')\tilde{u}_j(t)\rangle_z dt' + \underbrace{\langle z_i(t+\tau)\tilde{u}_j(t)\rangle_z}_{=0}$$

$$= \sum_{k=1}^N w_{ik}\int_{-\infty}^\tau dt\, h(\tau-t)\begin{cases} a_k^u(t) & \text{for } k=j \\ c_{kj}^u(t) & \text{for } k\neq j \end{cases}.$$

This equation has the same form as (31), so both models, within the linear approximation, exhibit an identical relationship between the auto- and cross-covariances. The physical meaning of the noise $z(t)$ is the variance caused by the spiking of the neurons. The auto-correlation function of a spike train of rate v has a δ-peak of weight v. The reduced model (33) exhibits such a δ-peak if we set $\rho_z^2 = v$. A related approach has been pursued before (see Sec. 3.5 in [31]) to determine the auto-correlation of the population averaged firing rate. This similarity will be discussed in detail below.

So far, we considered a network without external drive, i.e. all spike trains $s(t)$ originated from within the network. If the network is driven by external input, each neuron receives, in addition, synaptic input y_i from neurons outside the network. We assume uncorrelated external drive $\langle y_i(t+\tau)y_j(t)\rangle = \rho_y^2\delta_{ij}\delta(\tau)$. In the reduced model, this input constitutes a separate source of noise:

$$u_i(t) = L_t^i[u,y] = \sum_{k=1}^N w_{ik}(u_k * h) + (y_i * h_{iy}) + z_i(t). \qquad (34)$$

Here, $(f*g)(t) = \int_{-\infty}^t dt' f(t')g(t-t')$ denotes the convolution and $h_{iy}(t)$ the response kernel with respect to an external input. For simplicity, let's assume that the shape of these kernels is identical for all pairs of pre- and postsynaptic sources, i.e. $h_{ix} = h(t)$. If we further absorb the synaptic amplitude of the external drive in the strength of the noise ρ_y, the linearized dynamics (34) can be written in matrix notation

$$u(t) = ([Wu+y]*h)(t) + z(t) \qquad (35)$$

with $W = \{w_{ij}\}$. The reduced model (35) can be solved directly by means of Fourier transform:

$$U(\omega) = [1 - WH(\omega)]^{-1}(H(\omega)Y(\omega) + Z(\omega)). \qquad (36)$$

The full covariance matrix follows by averaging over the sources of noise Z and Y as

$$C^u(\omega) = \langle U(\omega)U^T(-\omega)\rangle_{Z,Y}$$
$$= (\rho_z^2 + |H(\omega)|^2\rho_y^2)[1 - WH(\omega)]^{-1}[1 - W^TH(-\omega)]^{-1}. \qquad (37)$$

The diagonal elements of C^u represent the auto-covariances, the off-diagonal elements the cross-covariances. Both are proportional to the driving noise $\rho_z^2 + |H(\omega)|^2\rho_y^2$. This is consistent with (31) which is a linear relationship between the cross- and auto-covariances.

For networks which can be decomposed into homogeneous subpopulations, the N dimensional system (35) can be further simplified by population averaging. Consider, for example, a homogeneous random network with purely inhibitory coupling. Assume that the neurons are randomly connected with probability ϵ and coupling strength $-w < 0$. The average number of in/outputs per neuron (in/out-degree) is thus given by $K = \epsilon N$. By introducing the population averaged external input $y(t) = E_i[y_i(t)]$, the averaged spiking noise $z(t) = E_i[z_i(t)]$, and the effective coupling strength $\bar{w} = Kw$, the dynamics of the population averaged activity becomes

$$u(t) = E_i[u_i(t)] = \left(\left[\sum_j E_i[w_{ij}]u_j + E_i[y_i]\right]*h\right)(t) + E_i[z_i(t)] = ([-\bar{w}u+y]*h)(t) + z(t). \qquad (38)$$

Here we assumed that $E_i[w_{ij}]$ is independent of the presynaptic neuron j and can be replaced by $-\epsilon w = -\bar{w}/N$. Note that this replacement is exact for networks with homogeneous out-degree, i.e. if the number of outgoing connections is identical for each neuron j. For large random networks with binomially distributed out-degrees (e.g. Erdös-Rényi networks or random networks with constant in-degree), (38) serves as an approximation.

To relate our approach to the treatment of finite-size fluctuations in [31], consider the population-averaged dynamics (38) of a single population with mean firing rate v. We set $\rho_z^2 = v$ for all single neuron noises z_i in order for the reduced model's auto-covariances to reproduce the δ-peak of the spiking dynamics. In the population averaged dynamics, this leads to the variance of the noise $z(t)$ given by $\langle z(t+\tau)z(t)\rangle = \frac{1}{N}\rho_z^2\delta(\tau) = \frac{v}{N}\delta(\tau)$. This agrees with the variance of the population rate in [31]. Therefore, the dynamics of the population averaged quantity u in (38) agrees with the earlier definition of a population averaged firing rate $s(t) = \frac{1}{N}\sum_i s_i(t)$ for the spiking network [31].

In equation (38), two distinct sources of noise appear: The noise due to external uncorrelated activity y and the noise z which is required to obtain the δ-peak of the auto-correlation functions of the reduced model. The qualitative results of **"Results: Suppression of population-activity fluctuations by negative feedback"** and **"Results: Population-activity fluctuations in excitatory-inhibitory networks"**, however can be understood with an even simpler model. As we are mainly concerned with the low-frequency fluctuations, we only need a

model with the same limit $\omega \rightarrow 0$. As we normalized the kernel so that $H(0) = 1$ we can combine both sources of noise and require $X(0) \equiv Y(0) + Z(0)$ in (36) in the zero frequency limit. Hence, in **"Results: Suppression of population-activity fluctuations by negative feedback"** and **"Results: Population-activity fluctuations in excitatory-inhibitory networks"**, we consider the model

$$r(t) = ([-Wr + x] * h)(t) \tag{39}$$

with a pairwise uncorrelated centralized white noise $E_x[x_i(t + \tau)x_j(t)] = \rho^2 \delta_{ij}\delta(\tau)$ to explain the suppression of fluctuations at low frequencies.

As a second example, consider a random network composed of an excitatory and an inhibitory subpopulation $\mathcal{E}$ and $\mathcal{I}$ with population sizes $N_E = |\mathcal{E}|$ and $N_I = |\mathcal{I}| = \gamma N_E$, respectively. Assume that each neuron receives excitatory and inhibitory inputs from $\mathcal{E}$ and $\mathcal{I}$ with coupling strengths $w > 0$ and $-gw < 0$, respectively, and probability ϵ, such that the average excitatory and inhibitory in/out-degrees are given by $K = \epsilon N$ and γK, respectively. The dynamics of the subpopulation averaged activities $u(t) = (u_E(t), u_I(t))^T$ is given by (35) with subpopulation averaged noise $y(t) = (y_E(t), y_I(t))^T$ and $z(t) = (z_E(t), z_I(t))^T$ and effective coupling

$$W = \bar{w}\begin{pmatrix} 1 & -\bar{g} \\ 1 & -\bar{g} \end{pmatrix}. \tag{40}$$

Here, $\bar{w} = Kw$ denotes the effective coupling strength, $\bar{g} = \gamma g$ the effective balance parameter and $y_{E/I}(t) = E_{i \in \mathcal{E}/\mathcal{I}}[y_i(t)]$ and $z_{E/I}(t) = E_{i \in \mathcal{E}/\mathcal{I}}[z_i(t)]$ the (sub)population averaged external and spiking sources of noise, respectively. Again, the reduction of the N-dimensional linear dynamics to the two-dimensional dynamics (40) is exact if the out-degrees are constant within each subpopulation. As before, both sources of noise can be combined into a single source of noise, if we are only interested in the low-frequency behavior of the model, leading to the dynamics (39) with the effective coupling (40).

The linear theory is only valid in the domain of its stability, which is determined by the eigenvalue spectrum of the effective coupling matrix W. For random coupling matrices, the eigenvalues are located within a circle with a radius equal to the square root of the variance of the matrix entries [69] $\sqrt{\text{Var}[w_{ij}]} = w\sqrt{N\epsilon(1 - \epsilon)(1 + \gamma g^2)}$. Writing the effective dynamics for the exponential kernel as a differential equation $\tau \frac{\partial r}{\partial t} = (W - 1)r + x(t)$, the eigenvalues of the right hand side matrix $W - 1$ are confined to a circle centered at -1 in the complex plain with radius $\sqrt{\text{Var}[w_{ij}]}$. Given $\text{Var}[w_{ij}] > 1$, eigenvalues might exist which have a positive real part, leading to unstable dynamics. This condition is indicated by the vertical dotted lines in Fig. 6 A–F and Fig. 7 B–D near $J = 2.8$ mV. Beyond this line, the linear model predicts an explosive growth of fluctuations. In the LIF-network model, an unbounded growth is avoided by the nonlinearities of the single-neuron dynamics.

Response kernel of the LIF model

We now perform the formal linearization (30) for a network of N LIF neurons $i = 1, \ldots, N$. A similar approach has been employed in previous studies to understand the population dynamics in these networks [12,31]. We consider the input $\sum_j J_{ij}s_j(t)$ received by neuron i from the local network, where s_j denotes the spike train of the neuron j projecting to neuron i with synaptic weight J_{ij}. Given the time dependent firing rate $v_j(t)$ of each afferent, and assuming small correlations and small synaptic weights, the total input to neuron i can be replaced by a Gaussian white noise with mean $\mu_i(t)$ and variance $\sigma_i^2(t)$,

$$\mu_i(t) = \tau_m \sum_j J_{ij}v_j(t)$$
$$\sigma_i^2(t) = \tau_m \sum_j J_{ij}^2 v_j(t), \tag{41}$$

where j sums over all synaptic inputs. $J_{ij} \in \{J, -gJ\}$ denotes the amplitude of the postsynaptic potential evoked by synapse $j \rightarrow i$. τ_m is the membrane time constant of the model. In the stationary state, the firing rate of each afferent is well described by the constant time average $\bar{v}_j = E_t[v_j]$. The working point at which we perform the linearization of the neural response (30) is then given by analog equations as (41), resulting in a constant mean $\bar{\mu}_i = \tau_m \sum_j J_{ij}\bar{v}_j$ and variance $\bar{\sigma}_i^2 = \tau_m \sum_j J_{ij}^2\bar{v}_j$. If the amplitude of each postsynaptic potential is small compared to the distance of the membrane potential to threshold, the dynamics of the LIF model can be approximated by a diffusion process, employing Fokker-Planck theory [83]. The stationary firing rate of the neuron is then given by [12,31,84]

$$\bar{v}_i^{-1}(\bar{\mu}_i, \bar{\sigma}_i) = \tau_{\text{ref}} + \sqrt{\pi}\tau_m(F(y_{\theta,i}) - F(y_{r,i}))$$
with
$$F(y) = \int^y f(y)\, dy \quad f(y) = e^{y^2}(\text{erf}(y) + 1) \tag{42}$$
$$y_{\theta,i} = \frac{\theta - \bar{\mu}_i}{\bar{\sigma}_i} \quad y_{r,i} = \frac{V_{\text{reset}} - \bar{\mu}_i}{\bar{\sigma}_i},$$

with the reset voltage V_{reset}, the threshold voltage θ and the refractory time τ_{ref}. In homogeneous random networks, the stationary rate (Fig. 7 A) is the same for all neurons. It is determined in a self-consistent manner [12] as the fixed point of (42). The stationary mean $\bar{\mu}_i$ and variance $\bar{\sigma}_i^2$ are determined by the stationary rate. To determine the kernel (30) we need to consider how a δ-shaped deflection in the input to this neuron at time point t' affects its output up to linear order in the amplitude of the fluctuation. In the stationary state, we may set $t' = 0$. It is therefore sufficient to focus on the effect of a single fluctuation

$$s_k(t) = a\delta(t). \tag{43}$$

We therefore ask how the density of spikes per time $v_i(t) = \langle G_t^i[s] \rangle_{s \setminus s_k}$ of neuron i, averaged over different realizations of the remaining inputs to neuron i, changes in response to the fluctuation (43) of the presynaptic neuron k in the limit of vanishing amplitude a. This kernel $w_{ik}h$ (30) is identical to the impulse response of the neuron and can directly be measured in simulation by trial averaging over many responses to the given δ-deflection (43) in the input (see Fig. 8 A). For the theory of low-frequency fluctuations, we only need the integral of the kernel, also known as the DC susceptibility,

$$w_{ik} = w_{ik} \int_0^\infty h(t)\, dt$$
$$= \lim_{a \rightarrow 0} \frac{\bar{v}_i(\bar{\mu}_i + \delta\mu_i, \bar{\sigma}_i + \delta\sigma_i) - \bar{v}_i(\bar{\mu}_i, \bar{\sigma}_i)}{a} \tag{44}$$
$$= \frac{\partial \bar{v}_i}{\partial \bar{\mu}_i}\tau_m J_{ik} + \frac{\partial \bar{v}_i}{\partial \bar{\sigma}_i}\frac{\tau_m}{2\bar{\sigma}_i}J_{ik}^2 + O(a^2).$$

The second equality follows from the equivalence of the integral of the impulse response and the step response in linear approximation [21,33]. Following from [41], both mean and variance are perturbed as $\delta\mu_i = a\tau_m J_{ik}$ and $\delta\sigma_i^2 = a\tau_m J_{ik}^2$ in response to a step a in the afferent rate v_j. Moreover, we used the chain rule $\delta\sigma_i = \frac{1}{2\bar{\sigma}_i}\delta\sigma_i^2$. The variation of the afferent firing rate hence co-modulates the mean and the variance and both modulations need to be taken into account to derive the neural response [31]. Although the finite amplitude of postsynaptic potentials has an effect on the response properties [33,34], the integral response is rather insensitive to the granularity of the noise [33]. We therefore employ the diffusion approximation to linearize the dynamics of the LIF neuron around its working point characterized by the mean $\bar{\mu}_i$ and the variance $\bar{\sigma}_i^2$ of the total synaptic input. In (44), we evaluate the partial derivatives of $\bar{v}_i$ with respect to $\bar{\mu}_i$ and $\bar{\sigma}_i^2$ using (42). First, observe that by chain rule $\frac{\partial\bar{v}_i}{\partial\bar{\mu}_i} = -\bar{v}_i^2\frac{\partial\bar{v}_i^{-1}}{\partial\bar{\mu}_i}$. We then again make use of the chain rule $\frac{\partial\bar{v}_i^{-1}}{\partial\bar{\mu}_i} = \frac{\partial\bar{v}_i^{-1}}{\partial y_{\theta,i}}\frac{\partial y_{\theta,i}}{\partial\bar{\mu}_i} + \frac{\partial\bar{v}_i^{-1}}{\partial y_{r,i}}\frac{\partial y_{r,i}}{\partial\bar{\mu}_i}$. Analog expressions hold for the derivative with respect to $\bar{\sigma}_i$. The first derivative yields $\frac{\partial\bar{v}_i^{-1}}{\partial y_{\theta,i}} = \sqrt{\pi}\tau_m f(y_{\theta,i})$, the one with respect to $y_{r,i}$ follows analogously, but with a negative sign. We further observe that $\frac{\partial y_A}{\partial\bar{\mu}_i} = \frac{-1}{\bar{\sigma}_i}$ and $\frac{\partial y_A}{\partial\bar{\sigma}_i} = \frac{-y_A}{\bar{\sigma}_i}$ with $y_A \in \{y_{r,i}, y_{\theta,i}\}$. Taken together, we obtain the explicit result for (44)

$$w_{ik} = (\bar{v}_i\tau_m)^2\sqrt{\pi}\frac{J_{ik}}{\bar{\sigma}_i}\left(f(y_{\theta,i})(1 + \frac{J_{ik}}{2\bar{\sigma}_i}y_{\theta,i}) - f(y_{r,i})(1 + \frac{J_{ik}}{2\bar{\sigma}_i}y_{r,i})\right). \quad (45)$$

Note that the modulation of μ_i results in a contribution to w_{ik} that is linear in J_{ik}, whereas the modulation of σ_i causes a quadratic dependence on J_{ik}. This expression therefore presents an extension to the integral response presented in [21,85]. Fig. 8 B shows the comparison of the analytical expression (45) and direct simulation. The agreement is good over a large range of synaptic amplitudes $J_{ik} \in [-4,4]\,\text{mV}$ in the case of constant background noise caused by small synaptic amplitudes (here $0.1\,\text{mV}$ for excitation and $-0.4\,\text{mV}$ for inhibition). For background noise caused by stronger impulses, the deviations are expected to grow [33].

Population-activity spectra in the linear model: feedback vs. feedforward scenario

The recurrent linear neural dynamics defined in the previous section is conveniently solved in the Fourier domain. The driving external Gaussian white noise X is mapped to the response $R(\omega) = T(\omega)X(\omega)$ by means of the transfer matrix $T(\omega)$. According to (39), it is given by $T(\omega) = H(\omega)(1 - H(\omega)W)^{-1}$. The covariance matrix in the frequency domain, the spectral matrix, thus reads

$$C_{RR}(\omega) = E_x[R(\omega)R(\omega)^*] = T(\omega)\rho^2 1 T(\omega)^*, \quad (46)$$

where we used $E_x[XX^*] = \rho^2 1$ and the expectation operator $E_x[]$ represents an average over noise realizations. To identify the effect of recurrence on the network dynamics, we replace the local feedback input by a feedforward input Q with spectral matrix C_{QQ}. The resulting response firing rate is given by $\tilde{R} = H(WQ + X)$. Assuming that the feedforward input Q is uncorrelated to the external noise source X ($C_{QX} = 0$) yields a response spectrum

$$C_{\bar{R}\bar{R}} = E_x[\tilde{R}\tilde{R}^*] = |H|^2(WC_{QQ}W^* + \rho^2 1). \quad (47)$$

Population-activity spectrum of the linear inhibitory network

In the Fourier domain, the solution of the mean-field dynamics (38) of the inhibitory network is $R(\omega) = H(\omega)X(\omega)/(1 + \bar{w}H(\omega))$. The power-spectrum $C_{RR}(\omega) = E_x[R(\omega)R(\omega)^*]$ hence becomes

$$C_{RR}(\omega) = \frac{|H(\omega)|^2}{|1 + \bar{w}H(\omega)|^2}\rho^2, \quad (48)$$

using the spectrum of the noise $E_x[X(\omega)X(\omega)^*] = \rho^2$.

We compare this power-spectrum to the case where the feedback loop is opened, i.e. where the recurrent input is replaced by feedforward input with unchanged auto-statistics $C_{QQ}(\omega) = C_{RR}(\omega)$, but which is uncorrelated to the external input $C_{QX}(\omega) = 0$. The resulting power-spectrum is given by (47) as $C_{\bar{R}\bar{R}} = |H|^2(\bar{w}^2 C_{RR} + \rho^2)$.

Population-activity spectra of the linear excitatory-inhibitory network

In a homogeneous random network of excitatory and inhibitory neurons, the population averaged activity (40) can be solved in the Schur basis (9) introduced in **"Results: Population-activity fluctuations in excitatory-inhibitory networks"**

$$R_+(\omega) = H(\omega)\frac{H(\omega)w_{\text{FF}}X_-(\omega) + X_+(\omega)}{1 + H(\omega)w_+}$$
$$R_-(\omega) = H(\omega)X_-(\omega), \quad (49)$$

with $w_+ = -\bar{w}(1 - \bar{g})$ and $w_{\text{FF}} = \bar{w}(1 + \bar{g})$. The power of the population rate therefore is

$$C_{R_+R_+}(\omega) = \frac{\rho^2|H(\omega)|^2}{2N_E}\cdot\frac{|H(\omega)w_{\text{FF}} + 1|^2 + \gamma^{-1}|H(\omega)w_{\text{FF}} - 1|^2}{|1 + H(\omega)w_+|^2}$$
$$C_{R_-R_-}(\omega) = \frac{\rho^2|H(\omega)|^2}{2N_E}(1 + \gamma^{-1}). \quad (50)$$

The fluctuations of the excitatory and the inhibitory population follow as

$$R_{E/I}(\omega) = \frac{1}{\sqrt{2}}\frac{H(\omega)}{1 + H(\omega)w_+}X_+(\omega)$$
$$+ \frac{1}{\sqrt{2}}\left(\frac{H(\omega)w_{\text{FF}}}{1 + H(\omega)w_+} \pm 1\right)H(\omega)X_-(\omega). \quad (51)$$

So the power-spectra are

$$C_{R_E R_E}(\omega) = \frac{|H(\omega)|^2\rho^2}{N_E}\cdot\frac{|1 + \bar{w}\bar{g}H(\omega)|^2 + \gamma^{-1}(\bar{w}\bar{g})^2|H(\omega)|^2}{|1 + H(\omega)w_+|^2}$$
$$C_{R_I R_I}(\omega) = \frac{|H(\omega)|^2\rho^2}{N_E}\cdot\frac{\bar{w}^2|H(\omega)|^2 + \gamma^{-1}|1 - \bar{w}H(\omega)|^2}{|1 + H(\omega)w_+|^2}$$
$$C_{R_E R_I}(\omega) = \frac{|H(\omega)|^2\rho^2}{N_E}\cdot\bar{w}\frac{\bar{w}\bar{g}(1 + \gamma^{-1})|H(\omega)|^2 + H^*(\omega) - \bar{g}\gamma^{-1}H(\omega)}{|1 + H(\omega)w_+|^2}. \quad (52)$$

Replacing the recurrent input of the sum activity R_+ by activity Q_+ with the same auto-statistics, but which is uncorrelated to the remaining input into R_+ (Fig. 5 D') results in the fluctuations

$$\tilde{R}_+(\omega) = H(\omega)\left(-w_+ Q_+(\omega) + w_{\mathrm{FF}}\tilde{R}_-(\omega) + X_+(\omega)\right),$$
$$\tilde{R}_-(\omega) = H(\omega) X_-(\omega). \tag{53}$$

The power-spectrum of the sum activity therefore becomes

$$C_{\bar{R}_+\bar{R}_+}(\omega) = |H(\omega)|^2 \left[w_+^2 C_{R_+R_+} + \frac{\rho^2}{2N_{\mathrm{E}}}\left(|H(\omega)w_{\mathrm{FF}}+1|^2 \right. \right.$$
$$\left. \left. + \gamma^{-1}|H(\omega)w_{\mathrm{FF}}-1|^2 \right)\right]. \tag{54}$$

If, alternatively, the excitatory and the inhibitory feedback terms R_{E} and R_{I} are replaced by uncorrelated feedforward input Q_{E} and Q_{I} with power-spectra $C_{R_{\mathrm{E}}R_{\mathrm{E}}}$ and $C_{R_{\mathrm{I}}R_{\mathrm{I}}}$ (Fig. 5 C,D), the spectrum of the sum activity reads

$$C_{\bar{R}_+\bar{R}_+}(\omega)$$
$$= |H(\omega)|^2 \left[2\bar{w}^2 \left(C_{R_{\mathrm{E}}R_{\mathrm{E}}}(\omega) + \bar{g}^2 C_{R_{\mathrm{I}}R_{\mathrm{I}}}(\omega) \right) + \frac{\rho^2}{2N_{\mathrm{E}}}(1+\gamma^{-1}) \right]. \tag{55}$$

The limit (14) for inhibition dominated networks with $\bar{g} > 1$ can be obtained from this and the former expressions by taking $H(0)=1$ and assuming strong coupling $\bar{w} \gg 1$.

Population averaged correlations in the linear EI network

In this subsection, we derive a self-consistency equation for the covariances in a recurrent network. We start from (37) (we drop the superscript u of C^u for brevity) multiply by $1-\boldsymbol{W}H(\omega)$ from left and its transpose from right to obtain

$$\boldsymbol{C}(\omega) = H(\omega)\boldsymbol{W}\boldsymbol{C}(\omega) + \boldsymbol{C}(\omega)H(-\omega)\boldsymbol{W}^{\mathrm{T}}$$
$$- |H(\omega)|^2 \boldsymbol{W}\boldsymbol{C}(\omega)\boldsymbol{W}^{\mathrm{T}} + 1(|H(\omega)|^2\rho_y^2 + \rho_z^2). \tag{56}$$

We assume a recurrent network of N_{E} excitatory and N_{I} inhibitory neurons, in which each neuron receives K excitatory inputs of weight w and γK inhibitory inputs of weight $-gw$ drawn randomly from the presynaptic pool of neurons. To obtain a theory for the variances and covariances at zero frequency (with $H(0)=1$) we may abbreviate $\rho_z^2(0) + |H(0)|^2\rho_y^2(0)$ by $\rho^2(0)$. For a population averaged theory, we need to replace in (56) the variances A_i of an individual neuron by the population average and replace the covariance C_{ij} for a given pair of neurons (i,j) by the average over pairs that are statistically equivalent to (i,j). For a pair (i,j) of neurons we will show that the set of equivalent pairs depends on the current realization of the connectivity since unconnected pairs are not equivalent to connected ones. Therefore it is necessary to first average the covariance matrix over statistically equivalent neuron pairs given a fixed connectivity and to subsequently average over all possible realizations of the connectivity. The latter will be denoted as $\mathrm{E}_W[]$. For compactness of the notation, first we perform the averaging for the general case, where neuron i belongs to population x and neuron j to population y. We denote by $\mathcal{X}$, $\mathcal{Y}$ the sets of neuron indices belonging to populations x and y, respectively. Subsequently replacing x and y by all possible combinations $x,y \in \{\mathrm{E,I}\}$, we obtain the averaged self-consistency equations for the network. We denote the number of incoming connections to a neuron of type x from the population of neurons of type y as K_{xy} and the strength of a synaptic coupling as w_{xy}. Rewriting the self-consistency equation

(56) explicitly with indices yields

$$C_{ij} = \sum_k w_{ik} C_{kj} + \sum_k w_{jk} C_{ik} - \sum_{k,l} w_{ik} C_{kl} w_{jl} + \rho^2 \delta_{ij}. \tag{57}$$

The last equation shows that for a connected pair (i,j) of neurons $(w_{ij} \neq 0$ or $w_{ji} \neq 0)$ either of the first two sums contains a contribution $w_{ij}C_{jj}$ or $w_{ji}C_{ii}$ proportional to the variance of the projecting neuron. We therefore need to perform the averaging separately for connected and for unconnected pairs of neurons. We use the notation

$$C_{x \leftarrow y} = \mathrm{E}_W\left[\frac{1}{N_{\mathrm{pairs},x\leftarrow y}} \sum_{i\in\mathcal{X}, j\in\mathcal{Y}, i\leftarrow j} C_{ij} \right] \tag{58}$$

for the average covariance over pairs of neurons of types $x,y\in\{\mathrm{E,I}\}$ with a connection from neuron $j\in\mathcal{Y}$ to neuron $i\in\mathcal{X}$, where $N_{\mathrm{pairs},x\leftarrow y}$ is the number neuron pairs connected in this way. An arrow to the right, $i\rightarrow j$, denotes a connection from neuron i to neuron j. Note that we use the same letter C for the population averaged covariances and for the covariances of individual pairs. The distinction can be made by the indices: i,j,k,l throughout indexes a single neuron, u,v,x,y,z identifies one of the populations $\{\mathrm{E,I}\}$. We denote the covariance averaged over unconnected pairs as

$$C_{x\not\leftrightarrow y} = \mathrm{E}_W\left[\frac{1}{N_{\mathrm{pairs},x\not\leftrightarrow y}} \sum_{i\in\mathcal{X}, j\in\mathcal{Y}, i\not\leftrightarrow j} C_{ij} \right]. \tag{59}$$

We further use

$$A_x = \mathrm{E}_W\left[\frac{1}{N_x} \sum_{i\in\mathcal{X}} C_{ii} \right] \tag{60}$$

for the integrated variance averaged over all neurons of type x. Connected and the unconnected averaged covariances differ by the term proportional to the variance of the projecting neuron, as mentioned above

$$C_{x\leftarrow y} = C_{x\not\leftrightarrow y} + w_{xy}A_y$$
$$C_{x\rightarrow y} = C_{x\not\leftrightarrow y} + w_{yx}A_x. \tag{61}$$

As a consequence, we can express all quantities in terms of the averaged variance (60) and the covariance averaged over unconnected pairs (59). We now proceed to average the integrated variance over population x. Since there are no self-connections in the network, we do not need to distinguish two cases here. Replacing C_{ii} on the right hand side of (60), the first term of (57) contributes

$$A_{x,\mathrm{A}} = \mathrm{E}_W\left[\frac{1}{N_x} \sum_{i\in\mathcal{X}} \sum_{k\in\mathcal{E}\vee\mathcal{I}} w_{ik} C_{ki} \right]$$
$$= \mathrm{E}_W\left[\frac{1}{N_x} \sum_{i\in\mathcal{X}} \left(\sum_{k\in\mathcal{E}} w_{ik} C_{ki} + \sum_{k\in\mathcal{I}} w_{ik} C_{ki} \right) \right] \tag{62}$$
$$= \sum_{z\in\{\mathrm{E,I}\}} K_{xz} w_{xz} C_{z\rightarrow x} = \sum_{z\in\{\mathrm{E,I}\}} K_{xz} w_{xz}(C_{z\not\leftrightarrow x} + w_{xz}A_z).$$

From the second to the third step we used that the sum over k (l) yields non-zero contributions only if neuron k (l) connects to neuron i. This happens in K_{xE} (K_{xI}) cases with the coupling weight w_{xE} (w_{xI}). Therefore the covariance averaged over connected pairs appears on the right hand side. In the last line we used the relation (61) to express the connected covariance in terms of the variance and the covariance over unconnected pairs. The second term in (60) is identical because of the symmetry $C_{ik} = C_{ki}$. Up to here, the structure of the network only entered in terms of the in-degree of the neurons. The contribution of the third term follows from a similar calculation

$$
\begin{aligned}
&\mathrm{E}_W\left[\frac{1}{N_x}\sum_{i\in\mathcal{X}}\sum_{k,l\in\mathcal{E}\vee\mathcal{I}}w_{ik}C_{kl}w_{il}\right]\\
&=\mathrm{E}_W\left[\frac{1}{N_x}\sum_{i\in\mathcal{X}}\left(\sum_{k\neq l\in\mathcal{E}\vee\mathcal{I}}w_{ik}C_{kl}w_{il}+\sum_{k\in\mathcal{E}\vee\mathcal{I}}w_{ik}^2 C_{kk}\right)\right]\\
&=\sum_{u,v\in\{\mathrm{E,I}\}}K_{xu}K_{xv}w_{xu}w_{xv}\left(\frac{K_{wu}}{N_u}(C_{v\leftarrow u}-C_{v\not\rightleftarrows u})+\right.\\
&\quad\left.\frac{K_{uv}}{N_v}(C_{v\rightarrow u}-C_{v\not\rightleftarrows u})+C_{v\not\rightleftarrows u}\right)+\sum_{z\in\{\mathrm{E,I}\}}K_{xz}w_{xz}^2 A_z\\
&=\sum_{u,v\in\{\mathrm{E,I}\}}K_{xu}K_{xv}w_{xu}w_{xv}C_{vu}+\sum_{z\in\{\mathrm{E,I}\}}K_{xz}w_{xz}^2 A_z.
\end{aligned}
\tag{63}
$$

From the second to the third step we assumed that among the $K_{xz}K_{xw}$ pairs of neurons $k\in\mathcal{Z},l\in\mathcal{W}$ projecting to neuron i, the fraction $\dfrac{K_{wz}}{N_z}$ has a connection $k\to l$. These pairs contribute with the connected covariance. The connections in opposite direction contribute the other term of similar structure. We ignore multiple and reciprocal connections here, assuming the connection probability is low. We introduce the shorthand C_{xy} for the covariance averaged over all neuron pairs including connected and unconnected pairs

$$
C_{xy}=C_{x\not\rightleftarrows y}+w_{xy}\frac{K_{xy}}{N_y}A_y+w_{yx}\frac{K_{yx}}{N_x}A_x.
\tag{64}
$$

This is the covariance which is observed on average when picking a pair of neurons of type x and y randomly. In this step, beyond the in-degree, the structure of the network entered through the expected number of connections between two populations. Taken all three terms together, we arrive at

$$
\begin{aligned}
A_x&=\rho^2+\sum_{z\in\{\mathrm{E,I}\}}K_{xz}w_{xz}(2C_{z\not\rightleftarrows x}+w_{xz}A_z)-C_{xx,\mathrm{corr}}\\
C_{xy,\mathrm{corr}}&\stackrel{\text{def}}{=}\sum_{u,v\in\{\mathrm{E,I}\}}K_{xu}K_{yv}w_{xu}w_{yv}C_{uv}.
\end{aligned}
\tag{65}
$$

The averaged covariances follow by similar calculations. Here we only need to calculate the average over unconnected pairs (i,j) given by (59), because the connected covariance follows from (61). The first sum in (57) contributes

$$
\begin{aligned}
C_{xy,\mathrm{A}}&\stackrel{\text{def}}{=}\mathrm{E}_W\left[\frac{1}{N_{\mathrm{pairs},x\not\rightleftarrows y}}\sum_{i\in\mathcal{X},j\in\mathcal{Y},i\neq j,i\not\rightleftarrows j}\sum_{k\in\mathcal{E}\vee\mathcal{I}}w_{ik}C_{kj}\right]\\
&=\sum_{z\in\{\mathrm{E,I}\}}K_{xz}w_{xz}C_{zy},
\end{aligned}
\tag{66}
$$

where due to the absence of a direct connection between i and j, the term linear in the coupling and proportional to the variance is absent. From the symmetry $C_{kl}=C_{lk}$ it follows that the second term corresponds to an exchange of x and y in the last expression. The third sum in (57) follows from an analog calculation as before

$$
\begin{aligned}
C_{xy,\mathrm{B}}&\stackrel{\text{def}}{=}\mathrm{E}_W\left[\frac{1}{N_{\mathrm{pairs},x\not\rightleftarrows y}}\sum_{i\in\mathcal{X},j\in\mathcal{Y},i\neq j,i\not\rightleftarrows j}\sum_{k,l\in\mathcal{E}\vee\mathcal{I}}w_{ik}C_{kl}w_{jl}\right]\\
&=\sum_{z\in\{\mathrm{E,I}\}}w_{xz}w_{yz}\frac{K_{xz}K_{yz}}{N_z}A_z+C_{xy,\mathrm{corr}}.
\end{aligned}
\tag{67}
$$

In summary, the contributions from (66) and (67) together result in the self-consistency equation for the covariance

$$
C_{x\not\rightleftarrows y}=C_{xy,\mathrm{A}}+C_{yx,\mathrm{A}}-C_{xy,\mathrm{B}}.
\tag{68}
$$

We now simplify the expressions by assuming that the in-degree of a neuron and the incoming synaptic amplitudes do not depend on the type of the neuron, i.e. that excitatory and inhibitory neurons receive statistically the same input. Formally this means that we need to replace K_{xy} by K_y, the number of incoming connections from population y and w_{xy} by w_y, the coupling strength of a projection from a neuron of type y. The covariance $C_{x\not\rightleftarrows y}$ then has two distinct contributions, $C_{xy,\mathrm{sep}}$ that depends on the type of neurons x,y, and C_{com} that does not. In particular $C_{xy,\mathrm{B}}$ and $C_{xy,\mathrm{corr}}$ do not depend on x,y and we omit their subscripts in the following. The variances fulfill

$$
\begin{aligned}
A_x&=A_{x,\mathrm{sep}}+A_{\mathrm{com}}+\rho^2\\
A_{x,\mathrm{sep}}&=\sum_{u\in\{\mathrm{E,I}\}}2K_u w_u C_{u\not\rightleftarrows x}\\
A_{\mathrm{com}}&=\sum_{u\in\{\mathrm{E,I}\}}K_u w_u^2 A_u-C_{\mathrm{corr}},
\end{aligned}
\tag{69}
$$

the covariances satisfy

$$
\begin{aligned}
C_{x\not\rightleftarrows y}&=C_{xy,\mathrm{sep}}+C_{\mathrm{com}}\\
C_{xy,\mathrm{sep}}&=\sum_{u\in\{\mathrm{E,I}\}}K_u w_u\left(\frac{K_x}{N_x}w_x A_x+\frac{K_y}{N_y}w_y A_y+C_{u\not\rightleftarrows y}+C_{u\not\rightleftarrows x}\right)\\
C_{\mathrm{com}}&=\sum_{u\in\{\mathrm{E,I}\}}\frac{K_u^2}{N_u}w_u^2 A_u-C_{\mathrm{corr}}\\
C_{\mathrm{corr}}&=\sum_{u,v\in\{\mathrm{E,I}\}}2K_v w_v\frac{K_u^2}{N_u}w_u^2 A_u+K_u K_v w_u w_v C_{u\not\rightleftarrows v}.
\end{aligned}
\tag{70}
$$

The disjoint part $C_{xy,\mathrm{sep}}$ determines the difference between the covariances for pairs of neurons of different type. Using the parameters $K_{\mathrm{E}}=K$, $K_{\mathrm{I}}=\gamma K$, $w_{\mathrm{E}}=w$, $w_{\mathrm{I}}=-gw$, the explicit form is

$$
\begin{aligned}
C_{\mathrm{EE,sep}}&=2Kw^2(1-\gamma g)\frac{K}{N_{\mathrm{E}}}A_{\mathrm{E}}+2KwC_{\mathrm{E}\not\rightleftarrows\mathrm{E}}-2\gamma gKwC_{\mathrm{E}\not\rightleftarrows\mathrm{I}}\\
C_{\mathrm{II,sep}}&=-2Kgw^2(1-\gamma g)\frac{\gamma K}{N_{\mathrm{I}}}A_{\mathrm{I}}+2\gamma gKwC_{\mathrm{I}\not\rightleftarrows\mathrm{I}}+2KwC_{\mathrm{E}\not\rightleftarrows\mathrm{I}}\\
C_{\mathrm{EI,sep}}&=\frac{1}{2}(C_{\mathrm{EE,sep}}+C_{\mathrm{II,sep}}).
\end{aligned}
\tag{71}
$$

Therefore, also the covariances in the network obey the relation

$$C_{\text{E}\not\rightleftarrows\text{E}} + C_{\text{I}\not\rightleftarrows\text{I}} = 2C_{\text{E}\not\rightleftarrows\text{I}}, \tag{72}$$

i.e. the mixed covariance can be eliminated and is given by the arithmetic mean of the covariances between neurons of same type. In matrix representation with the vector $Q = (A_E, A_I, C_{\text{E}\not\rightleftarrows\text{E}}, C_{\text{I}\not\rightleftarrows\text{I}})$, the self-consistency equation is

$$M_{\text{dis}} =$$

$$Kw \begin{pmatrix} 0 & 0 & 2-\gamma g & -\gamma g \\ 0 & 0 & 1 & 1-2\gamma g \\ 2w(1-\gamma g)\dfrac{K}{N_E} & 0 & 2-\gamma g & -\gamma g \\ 0 & -2gw(1-\gamma g)\dfrac{\gamma K}{N_I} & 1 & 1-2\gamma g \end{pmatrix} \tag{73}$$

$$M_{\text{com}} = (Kw)^2 \begin{pmatrix} \dfrac{1}{K} & \dfrac{g^2\gamma}{K} & 0 & 0 \\ \dfrac{1}{K} & \dfrac{g^2\gamma}{K} & 0 & 0 \\ \dfrac{1}{N_E} & \dfrac{(g\gamma)^2}{N_I} & 0 & 0 \\ \dfrac{1}{N_E} & \dfrac{(g\gamma)^2}{N_I} & 0 & 0 \end{pmatrix} - M_{\text{FF}} \tag{74}$$

$$M_{\text{FF}} = (Kw)^2(1-\gamma g) \begin{pmatrix} 2\dfrac{Kw}{N_E} & 2(\gamma g)^2\dfrac{Kw}{N_I} & 1 & -\gamma g \\ 2\dfrac{Kw}{N_E} & 2(\gamma g)^2\dfrac{Kw}{N_I} & 1 & -\gamma g \\ 2\dfrac{Kw}{N_E} & 2(\gamma g)^2\dfrac{Kw}{N_I} & 1 & -\gamma g \\ 2\dfrac{Kw}{N_E} & 2(\gamma g)^2\dfrac{Kw}{N_I} & 1 & -\gamma g \end{pmatrix}. \tag{75}$$

The self consistent covariance can then be obtained by solving the system of linear equations

$$(I - M_{\text{dis}} - M_{\text{com}})Q = \rho^2(1,1,0,0)^{\text{T}}. \tag{76}$$

The numerical solution shows that the variances for excitatory and inhibitory neurons are approximately the same, as depicted in Fig. 6 A. In the following we therefore assume $A_E = A_I = A$ and then solve (76) for the covariances. With the abbreviation $G = \left(\dfrac{1}{N_E} + \dfrac{(\gamma g)^2}{N_I}\right)$, the third and fourth line yields the equation for the covariances

$$C_{\text{E}\not\rightleftarrows\text{E}/\text{I}\not\rightleftarrows\text{I}} = (Kw)^2 A \left[G(1-2Kw(1-\gamma g)) + 2(1-\gamma g) \begin{cases} \dfrac{1}{N_E} & \text{for EE} \\ \dfrac{-\gamma g}{N_I} & \text{for II} \end{cases} \right]$$

$$+ Kw(C_{\text{E}\not\rightleftarrows\text{E}} - \gamma g C_{\text{I}\not\rightleftarrows\text{I}})(1 - Kw(1-\gamma g))$$

$$+ Kw(1-\gamma g) \begin{cases} C_{\text{E}\not\rightleftarrows\text{E}} & \text{for EE} \\ C_{\text{I}\not\rightleftarrows\text{I}} & \text{for II} \end{cases} \tag{77}$$

The structure of the equation suggests to introduce the linear combination $m = C_{\text{EE}} - \gamma g C_{\text{II}}$ which satisfies

$$m = (Kw)^2(1-\gamma g)G(3-2Kw(1-\gamma g))A$$
$$+ Kw(1-\gamma g)(2-Kw(1-\gamma g))m \tag{78}$$
$$m = (Kw)^2 G(1-\gamma g)\frac{3-2Kw(1-\gamma g)}{(1-Kw(1-\gamma g))^2}A.$$

We solve (77) for $C_{\text{E}\not\rightleftarrows\text{E}}$ and $C_{\text{I}\not\rightleftarrows\text{I}}$ and insert (78) for m to obtain the covariances as

$$C_{\text{E}\not\rightleftarrows\text{E}/\text{I}\not\rightleftarrows\text{I}} =$$

$$(Kw)^2 A \left[G\frac{1-2Kw(1-\gamma g)}{1-Kw(1-\gamma g)} + 2\frac{1-\gamma g}{1-Kw(1-\gamma g)} \begin{cases} \dfrac{1}{N_E} & \text{for EE} \\ \dfrac{-\gamma g}{N_I} & \text{for II} \end{cases} \right]$$

$$+ Kwm = G\frac{(Kw)^2}{(1-Kw(1-\gamma g))^2}A \tag{79}$$

$$+ 2\frac{Kw(1-\gamma g)}{1-Kw(1-\gamma g)}A \begin{cases} \dfrac{Kw}{N_E} & \text{for EE} \\ \dfrac{-K\gamma wg}{N_I} & \text{for II} \end{cases}.$$

The covariance $C_{x\not\rightleftarrows y}$ between unconnected neurons can be related to the covariance between the incoming currents this pair of neurons receives. Expressing the self-consistency (68) in terms of the covariances averaged over connected and unconnected pairs (64) uncovers the connection

$$C_{x\not\rightleftarrows y} = \sum_{z\in\{\text{E,I}\}} K_z w_z (C_{zx} + C_{zy})$$
$$- \sum_{z\in\{\text{E,I}\}} w_z^2 \frac{K_z^2}{N_z} A_z - \sum_{u,v\in\{\text{E,I}\}} K_u K_v w_u w_v C_{uv}$$
$$= Kw\left[C_{\text{E}x} + C_{\text{E}y} - \gamma g(C_{\text{I}x} + C_{\text{I}y})\right]$$
$$- (Kw)^2\left[\frac{1}{N_E}A_E + \frac{(\gamma g)^2}{N_I}A_I + C_{\text{EE}} - 2\gamma g C_{\text{EI}} + (\gamma g)^2 C_{\text{II}}\right]. \tag{80}$$

This self-consistency equation yields the argument, why the shared-input correlation $C_{\text{shared}}^{\text{in}}$ (19) cancels the contribution $C_{\text{corr}}^{\text{in}}$ (20) due to spike-train correlations in the covariance to the input currents (see Fig. 6 C,D). Rewriting (80) in terms of these quantities results in

$$\frac{C_{x\not\rightleftarrows y}}{Kw} - \left[C_{\text{E}x} + C_{\text{E}y} - \gamma g(C_{\text{I}x} + C_{\text{I}y})\right]$$
$$= Kw\left[C_{\text{shared}}^{\text{in}}/(Kw)^2 + C_{\text{corr}}^{\text{in}}/(Kw)^2\right]. \tag{81}$$

If a self-consistent solution with small correlation $|C_{x\not\rightleftarrows y}|, |C_{xy}| < \epsilon$ exists, the right hand side of (81) must be of the same order of magnitude. The right hand side of this equation has a prefactor Kw which typically is $\gg 1$ (for the parameters in Fig. 6, Kw becomes larger than 1 for $w > 10^{-3}$). The first term in the bracket is proportional to the contribution of shared input, the second term is due to correlations among pairs of different neurons. Each of these terms is of order ϵ. Due to the prefactor Kw, however, the sum of the two terms needs to be of order $\epsilon/(Kw)$ to fulfill the equation. Hence, the terms must have different signs to cause the mutual cancellation.

To illustrate how the correlation structure is affected by feedback, let us now consider the case where the feedback activity is perturbed ("feedforward scenario"). We start from (47) and, again, only consider the fluctuations at zero frequency,

$$\boldsymbol{C}_{\tilde{R}\tilde{R}}(0) = \boldsymbol{W}\boldsymbol{C}_{QQ}\boldsymbol{W}^{\mathrm{T}} + 1\rho^2. \tag{82}$$

First, we consider a manipulation that preserves the single-neuron statistics A_{E}, A_{I} and the pairwise correlations C_{EE}, C_{II} within each subpopulation, but neglects correlations C_{EI} between excitatory and inhibitory neurons. Formally, this corresponds to the block diagonal correlation matrix

$$\boldsymbol{C}_{QQ_{ij}} = \begin{cases} \delta_{ij}A_{\mathrm{E}} + (1-\delta_{ij})C_{\mathrm{EE}} & i,j \in \mathcal{E} \\ \delta_{ij}A_{\mathrm{I}} + (1-\delta_{ij})C_{\mathrm{II}} & i,j \in \mathcal{I}. \end{cases} \tag{83}$$

Here, we have replaced the individual entries of the correlation matrix by the corresponding subpopulation averaged correlations. The calculation of the response auto- and cross-correlation $\tilde{A}$ and $\tilde{C}$ is similar as for the expressions (63) and (67), with the difference that terms containing C_{EI} are absent:

$$\begin{aligned} \tilde{A} &= Kw^2\left(A_{\mathrm{E}} + \gamma g^2 A_{\mathrm{I}}\right) + c + \rho^2 \\ \tilde{C} &= (Kw)^2\left(\frac{A_{\mathrm{E}}}{N_{\mathrm{E}}} + (\gamma g)^2\frac{A_{\mathrm{I}}}{N_{\mathrm{I}}}\right) + c \\ \text{with} \quad & c = (Kw)^2\left(C_{\mathrm{EE}} + (\gamma g)^2 C_{\mathrm{II}}\right). \end{aligned} \tag{84}$$

As an alternative type of feedback manipulation, we assume that all correlations are equal, irrespective of the neuron type. To this end, we replace all spike correlations by the population average $C = (N_{\mathrm{E}}^2 C_{\mathrm{EE}} + N_{\mathrm{I}}^2 C_{\mathrm{II}} + 2N_{\mathrm{E}}N_{\mathrm{I}}C_{\mathrm{EI}})/(N_{\mathrm{E}} + N_{\mathrm{I}})^2 = (N_{\mathrm{E}}C_{\mathrm{EE}} + N_{\mathrm{I}}C_{\mathrm{II}})/(N_{\mathrm{E}} + N_{\mathrm{I}})$. Thus, the covariance matrix reads

$$\boldsymbol{C}_{QQ_{ij}} = \delta_{ij}A + (1-\delta_{ij})C. \tag{85}$$

The calculation follows the one leading to the expressions (63) and (67) and results in

$$\begin{aligned} \tilde{A} &= w^2K(1+\gamma g^2)A + (wK)^2(1-\gamma g)^2 C + \rho^2 \\ \tilde{C} &= w^2K^2\left(\frac{1}{N_{\mathrm{E}}} + \frac{(\gamma g)^2}{N_{\mathrm{I}}}\right)A + (wK)^2(1-\gamma g)^2 C. \end{aligned} \tag{86}$$

References

1. Tripp B, Eliasmith C (2007) Neural populations can induce reliable postsynaptic currents without observable spike rate changes or precise spike timing. Cereb Cortex 17: 1830–1840.
2. Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. Nature 370: 140–143.
3. Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. J Neurosci 18: 3870–3896.
4. Salinas E, Sejnowski TJ (2001) Correlated neuronal activity and the ow of neural information. Nat Rev Neurosci 2: 539–550.
5. von der Malsburg C (1981) The correlation theory of brain function. Internal report 81-2, Department of Neurobiology, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany.
6. Bienenstock E (1995) A model of neocortex. Network: Comput Neural Sys 6: 179–224.
7. Abeles M (1991) Corticonics: Neural Circuits of the Cerebral Cortex. Cambridge: Cambridge University Press, 1st edition.
8. Diesmann M, Gewaltig MO, Aertsen A (1999) Stable propagation of synchronous spiking in cortical neural networks. Nature 402: 529–533.
9. Tetzlaff T, Morrison A, Geisel T, Diesmann M (2004) Consequences of realistic network size on the stability of embedded synfire chains. Neurocomputing 58–60: 117–121.
10. Ecker AS, Berens P, Keliris GA, Bethge M, Logothetis NK (2010) Decorrelated neuronal firing in cortical microcircuits. Science 327: 584–587.
11. van Vreeswijk C, Sompolinsky H (1996) Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science 274: 1724–1726.
12. Brunel N (2000) Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. J Comput Neurosci 8: 183–208.
13. Gerstner W (2000) Population dynamics of spiking neurons: fast transients, asynchronous states, and locking. Neural Comput 12: 43–89.
14. Tetzlaff T, Rotter S, Stark E, Abeles M, Aertsen A, et al. (2008) Dependence of neuronal correlations on filter characteristics and marginal spike-train statistics. Neural Comput 20: 2133–2184.
15. Kriener B, Tetzlaff T, Aertsen A, Diesmann M, Rotter S (2008) Correlations and population dynamics in cortical networks. Neural Comput 20: 2185–2226.
16. Hertz J (2010) Cross-correlations in high-conductance states of a model cortical network. Neural Comput 22: 427–447.
17. Renart A, De La Rocha J, Bartho P, Hollender L, Parga N, et al. (2010) The asynchronous state in cortical ciicuits. Science 327: 587–590.
18. Stroeve S, Gielen S (2001) Correlation between uncoupled conductance-based integrate-and-fire neurons due to common and synchronous presynaptic firing. Neural Comput 13: 2005–2029.
19. Tetzlaff T, Buschermöhle M, Geisel T, Diesmann M (2003) The spread of rate and correlation in stationary cortical networks. Neurocomputing 52–54: 949–954.
20. Moreno-Bote R, Parga N (2006) Auto- and crosscorrelograms for the spike response of leaky integrate-and-fire neurons with slow synapses. Phys Rev Lett 96: 028101.
21. De la Rocha J, Doiron B, Shea-Brown E, Kresimir J, Reyes A (2007) Correlation between neural spike trains increases with firing rate. Nature 448: 802–807.
22. Rosenbaum R, Josic K (2011) Mechanisms that modulate the transfer of spiking correlations. Neural Comput 23: 1261–1305.
23. Battaglia D, Brunel N, Hansel D (2007) Temporal decorrelation of collective oscillations in neural networks with local inhibition and long-range excitation. Phys Rev Lett 99: 238106.
24. Monteforte M, Wolf F (2010) Dynamical entropy production in spiking neuron networks in the balanced state. Phys Rev Lett 105: doi: 268104
25. Jahnke S, Memmesheimer RM, Timme M (2009) How chaotic is the balanced state? Front Comput Neurosci 3: 1–15.
26. Legenstein R, Maass W (2007) Edge of chaos and prediction of computational performance for neural circuit models. Neural Netw 20: 323–334.
27. Jahnke S, Memmesheimer R, Timme M (2008) Stable irregular dynamics in complex neural networks. Phys Rev Lett 100: 048102.
28. Toyoizumi T, Abbott LF (2010) Beyond the edge: Amplification and temporal integration by recurrent networks in the chaotic regime. Front Neurosci: doi: 10.3389/conf.fnins.2010.03.00155.
29. Zillmer R, Livi R, Politi A, Torcini A (2006) Desynchronization in diluted neural networks. Phys Rev E 74: 036203.
30. Ginzburg I, Sompolinsky H (1994) Theory of correlations in stochastic neural networks. Phys Rev E 50: 3171–3191.
31. Brunel N, Hakim V (1999) Fast global oscillations in networks of integrate-and-_re neurons with low firing rates. Neural Comput 11: 1621–1671.
32. Harris KD, Thiele A (2011) Cortical state and attention. Nat Rev Neurosci 12: 509–523.
33. Helias M, Deger M, Rotter S, Diesmann M (2010) Instantaneous non-linear processing by pulsecoupled threshold units. PLoS Comput Biol 6: e1000929.
34. Richardson MJE, Swarbrick R (2010) Firing-rate response of a neuron receiving excitatory and inhibitory synaptic shot noise. Phys Rev Lett 105: 178102.

35. Fourcaud-Trocmé N, Hansel D, van Vreeswijk C, Brunel (2003) How spike generation mechanisms determine the neuronal response to ﬂuctuating inputs. J Neurosci 23: 11628–11640.

36. Naundorf B, Geisel T, Wolf F (2005) Action potential onset dynamics and the response speed of neuronal populations. J Comput Neurosci 18: 297–309.

37. Brunel N, Chance FS, Fourcaud N, Abbott LF (2001) Effects of synaptic noise and filtering on the frequency response of spiking neurons. Phys Rev Lett 86: 2186–2189.

38. Nordlie E, Tetzlaff T, Einevoll GT (2010) Rate dynamics of leaky integrate-and-fire neurons with strong synapses. Front Comput Neurosci 4: 149.

39. Pressley J, Troyer TW (2009) Complementary responses to mean and variance modulations in the perfect integrate-and-fire model. Biol Cybern 101: 63–70.

40. Knight BW (1972) The relationship between the firing rate of a single neuron and the level of activity in a population of neurons. J Gen Physiol 59: 767–778.

41. Köndgen H, Geisler C, Fusi S, Wang XJ, Lüscher HR, et al. (2008) The dynamical response properties of neocortical neurons to temporally modulated noisy inputs in vitro. Cereb Cortex 18: 2086–2097.

42. Boucsein C, Tetzlaff T, Meier R, Aertsen A, Naundorf B (2009) Dynamical response properties of neocortical neuron ensembles: multiplicative versus additive noise. J Neurosci 29: 1006–1010.

43. Blomquist P, Devor A, Indahl UG, Ulbert I, Einevoll GT, et al. (2009) Estimation of thalamocortical and intracortical network models from joint thalamic single-electrode and cortical laminar electrode recordings in the rat barrel system. PLoS Comput Biol 5: e1000328.

44. Knight BW (1972) Dynamics of encoding in a population of neurons. J Gen Physiol 59: 734–766.

45. Lindner B, Schimansky-Geier L (2001) Transmission of noise coded versus additive signals through a neuronal ensemble. Phys Rev Lett 86: 2934–2937.

46. Fourcaud N, Brunel N (2002) Dynamics of the firing probability of noisy integrate-and-fire neurons. Neural Comput 14: 2057–2110.

47. Oppenheim A, Wilsky A (1996) Systems and signals. Upper Saddle River, NJ: Prentice Hall.

48. Troyer TW, Krukowski AE, Miller KD (2002) LGN input to simple cells and contrast-invariant orientation tuning: An analysis. J Neurophysiol 87: 2741–2752.

49. Zhaoping L, Lewis A, Scarpetta S (2004) Mathematical analysis and simulations of the neural circuit for locomotion in lampreys. Phys Rev Lett 92: 198106.

50. Murphy BK, Miller KD (2009) Balanced amplification: A new mechanism of selective amplification of neural activity patterns. Neuron 61: 635–648.

51. Moreno-Bote R, Renart A, Parga N (2008) Theory of input spike auto- and cross-correlations and their effect on the response of spiking neurons. Neural Comput 20: 1651–1705.

52. Shadlen MN, Newsome WT (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J Neurophysiol 86: 1916–1936.

53. Gentet L, Avermann M, Matyas F, Staiger JF, Petersen CC (2010) Membrane potential dynamics of GABAergic neurons in the barrel cortex of behaving mice. Neuron 65: 422–435.

54. Pernice V, Staude B, Cardanobile S, Rotter S (2011) How structure determines correlations in neuronal networks. PLoS Comput Biol 7: e1002059.

55. Pernice V, Staude B, Cardanobile S, Rotter S (2012) Recurrent interactions in spiking networks with arbitrary topology. Phys Rev E 85: 031916.

56. Trousdale J, Hu Y, Shea-Brown E, Josic K (2012) Impact of network structure and cellular response on spike time correlations. PLoS Comput Biol 8: e1002408.

57. Softky WR, Koch C (1993) The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. J Neurosci 13: 334–350.

58. Tetzlaff T, Geisel T, Diesmann M (2002) The ground state of cortical feed-forward networks. Neurocomputing 44–46: 673–678.

59. Mehring C, Hehl U, Kubo M, Diesmann M, Aertsen A (2003) Activity dynamics and propagation of synchronous spiking in locally connected random networks. Biol Cybern 88: 395–408.

60. Aviel Y, Mehring C, Abeles M, Horn D (2003) On embedding synfire chains in a balanced network. Neural Comput 15: 1321–1340.

61. Aviel Y, Horn D, Abeles M (2005) Memory capacity of balanced networks. Neural Comput 17: 691–713.

62. Bi G, Poo M (1998) Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. J Neurosci 18: 10464–10472.

63. Gilson M, Burkitt AN, Grayden DB, Thomas DA, van Hemmen JL (2009) Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. I. Input selectivity - strengthening correlated input pathways. Biol Cybern 101: 81–102.

64. Burkitt AN, Gilson M, van Hemmen J (2007) Spike-timing-dependent plasticity for neurons with recurrent connections. Biol Cybern 96: 533–546.

65. Pfister JP, Tass PA (2010) Stdp in oscillatory recurrent networks: theoretical conditions for desynchronization and applications to deep brain stimulation. Front Comput Neurosci 454: doi: 10.3389/fncom.2010.00022.

66. Lindner B, Doiron B, Longtin A (2005) Theory of oscillatory firing induced by spatially correlated noise and delayed inhibitory feedback. Phys Rev E 72: 061919.

67. Buice MA, Cowan JD, Chow CC (2009) Systematic ﬂuctuation expansion for neural network activity equations. Neural Comput 22: 377–426.

68. Sommers H, Crisanti A, Sompolinsky H, Stein Y (1988) Spectrum of large random asymmetric matrices. Phys Rev Lett 60: 1895–1898.

69. Rajan K, Abbott L (2006) Eigenvalue spectra of random matrices for neural networks. Phys Rev Lett 97: 188104.

70. Sompolinsky, Crisanti, Sommers (1988) Chaos in random neural networks. Phys Rev Lett 61: 259–262.

71. Toyoizumi T, Rad KR, Paninski L (2009) Mean-field approximations for coupled populations of generalized linear model spiking neurons with markov refractoriness. Neural Comput 21: 1203–1243.

72. Sirovich L, Omurtag A, Knight BW (2000) Dynamics of neuronal populations: The equilibrium solution. SIAM J Appl Math 60: 2009–2028.

73. Jacobsen M, Jensen AT (2007) Exit times for a class of piecewise exponential markov processes with two-sided jumps. Stoch Proc Appl 117: 1330–1356.

74. Cox DR (1962) Renewal Theory. Science Paperbacks. London: Chapman and Hall.

75. Nawrot MP, Boucsein C, Rodriguez Molina V, Riehle A, Aertsen A, et al. (2008) Measurement of variability dynamics in cortical spike trains. J Neurosci Methods 169: 374–390.

76. Chizhov AV, Graham LJ (2008) Efficient evaluation of neuron populations receiving colored-noise current based on a refractory density method. Phys Rev E 77: 011910.

77. Meyer C, van Vreeswijk C (2002) Temporal correlations in stochastic networks of spiking neurons. Neural Comput 14: 369–404.

78. Hebb DO (1949) The organization of behavior: A neuropsychological theory. New York: John Wiley & Sons.

79. Helias M, Rotter S, Gewaltig M, Diesmann M (2008) Structural plasticity controlled by calcium based correlation detection. Front Comput Neurosci 2: doi:10.3389/neuro.10.007.2008.

80. Loebel A, Silberberg G, Helbig D, Markram H, Tsodyks M, et al. (2009) Multiquantal release underlies the distribution of synaptic efficacies in the neocortex. Front Comput Neurosci 3: 27.

81. Gewaltig MO, Diesmann M (2007) NEST (NEural Simulation Tool). Scholarpedia 2: 1430.

82. Hawkes A (1971) Point spectra of some mutually exciting point process. J R Statist Soc Ser B 33: 438–443.

83. Risken H (1996) The Fokker-Planck Equation. Berlin: Springer Verlag.

84. Siegert AJ (1951) On the first passage time probability problem. Phys Rev 81: 617–623.

85. Helias M, Deger M, Diesmann M, Rotter S (2010) Equilibrium and response properties of the integrate-and-fire neuron in discrete time. Front Comput Neurosci 3: doi:10.3389/neuro.10.029.2009.

86. Tuckwell HC (1988) Introduction to Theoretical Neurobiology, volume 1. Cambridge: Cambridge University Press.

87. Rotter S, Diesmann M (1999) Exact digital simulation of time-invariant linear systems with applications to neuronal modeling. Biol Cybern 81: 381–402.