

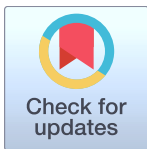
RESEARCH ARTICLE

# Locking of correlated neural activity to ongoing oscillations

Tobias Kühn<sup>1\*</sup>, Moritz Helias<sup>1,2</sup>

**1** Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA BRAIN Institute I, Jülich Research Centre, Jülich, Germany, **2** Department of Physics, Faculty 1, RWTH Aachen University, Aachen, Germany

\* [t.kuehn@fz-juelich.de](mailto:t.kuehn@fz-juelich.de)



## OPEN ACCESS

**Citation:** Kühn T, Helias M (2017) Locking of correlated neural activity to ongoing oscillations. PLoS Comput Biol 13(6): e1005534. <https://doi.org/10.1371/journal.pcbi.1005534>

**Editor:** Bard Ermentrout, University of Pittsburgh, UNITED STATES

**Received:** September 30, 2016

**Accepted:** April 26, 2017

**Published:** June 12, 2017

**Copyright:** © 2017 Kühn, Helias. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All python files are available from the dryad database (doi:[10.5061/dryad.45sg0](https://doi.org/10.5061/dryad.45sg0)).

**Funding:** This work was partly supported by the Helmholtz young investigator group VH-NG-1028 ([https://www.helmholtz.de/en/jobs\\_talent/funding\\_programs/helmholtz\\_young\\_investigators\\_groups/](https://www.helmholtz.de/en/jobs_talent/funding_programs/helmholtz_young_investigators_groups/)), the Helmholtz Portfolio Supercomputing and Modeling for the Human Brain (SMHB, <http://www.fz-juelich.de/inm/inm-1/EN/Forschung/JuBrain/Helmholtz%20Portfolio.html>) and EU Grant 604102 (Human Brain Project, HBP, <https://www.humanbrainproject.eu/>). The funders had no role in

## Abstract

Population-wide oscillations are ubiquitously observed in mesoscopic signals of cortical activity. In these network states a global oscillatory cycle modulates the propensity of neurons to fire. Synchronous activation of neurons has been hypothesized to be a separate channel of signal processing information in the brain. A salient question is therefore if and how oscillations interact with spike synchrony and in how far these channels can be considered separate. Experiments indeed showed that correlated spiking co-modulates with the static firing rate and is also tightly locked to the phase of beta-oscillations. While the dependence of correlations on the mean rate is well understood in feed-forward networks, it remains unclear why and by which mechanisms correlations tightly lock to an oscillatory cycle. We here demonstrate that such correlated activation of pairs of neurons is qualitatively explained by periodically-driven random networks. We identify the mechanisms by which covariances depend on a driving periodic stimulus. Mean-field theory combined with linear response theory yields closed-form expressions for the cyclostationary mean activities and pairwise zero-time-lag covariances of binary recurrent random networks. Two distinct mechanisms cause time-dependent covariances: the modulation of the susceptibility of single neurons (via the external input and network feedback) and the time-varying variances of single unit activities. For some parameters, the effectively inhibitory recurrent feedback leads to resonant covariances even if mean activities show non-resonant behavior. Our analytical results open the question of time-modulated synchronous activity to a quantitative analysis.

## Author summary

In network theory, statistics are often considered to be stationary. While this assumption can be justified by experimental insights to some extent, it is often also made for reasons of simplicity. However, the time-dependence of statistical measures do matter in many cases. For example, time-dependent processes are examined for gene regulatory networks or networks of traders at stock markets. Periodically changing activity of remote brain areas is visible in the local field potential (LFP) and its influence on the spiking activity is

study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

currently debated in neuroscience. In experimental studies, however, it is often difficult to determine time-dependent statistics due to a lack of sufficient data representing the system at a certain time point. Theoretical studies, in contrast, allow the assessment of the time dependent statistics with arbitrary precision. We here extend the analysis of the correlation structure of a homogeneously connected EI-network consisting of binary model neurons to the case including a global sinusoidal input to the network. We show that the time-dependence of the covariances—to first order—can be explained analytically. We expose the mechanisms that modulate covariances in time and show how they are shaped by inhibitory recurrent network feedback and the low-pass characteristics of neurons. These generic properties carry over to more realistic neuron models.

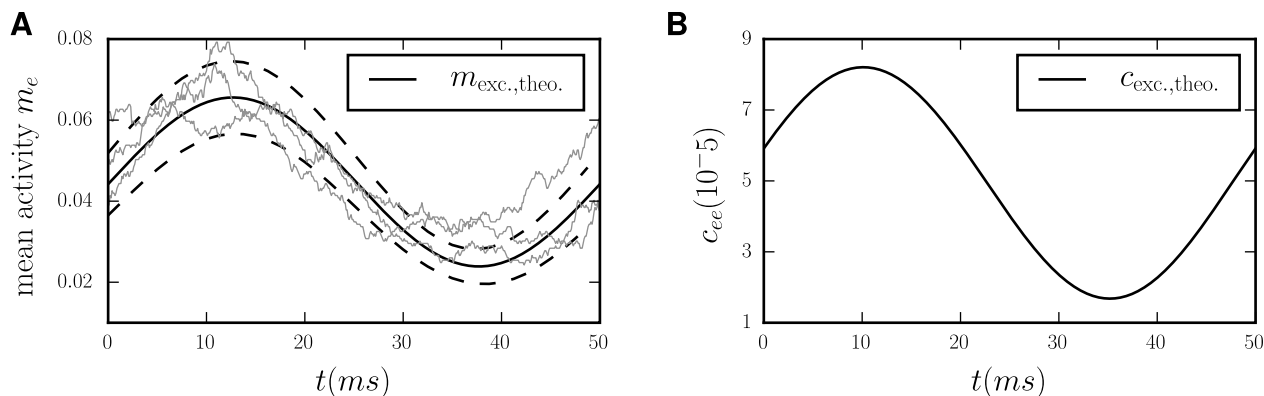
## Introduction

To date it is unclear which channels the brain uses to represent and process information. A rate-based view is argued for by the apparent stochasticity of firing [1] and by the high sensitivity of the network dynamics to single spikes [2]. In an extreme view, correlated firing is a mere epiphenomenon of neurons being connected. Indeed, a large body of literature has elucidated how correlations relate to the connectivity structure [3–14]. But the matter is further complicated by the observation that firing rates and correlations tend to be co-modulated, as demonstrated experimentally and explained theoretically [4, 5]. If the brain employs correlated firing as a means to process or represent information, this requires in particular that the appearance of correlated events is modulated in a time-dependent manner. Indeed, such modulations have been experimentally observed in relation to the expectation of the animal to receive task-relevant information [15, 16] or in relation to attention [17].

Oscillations are an extreme case of a time-dependent modulation of the firing rate of cells. They are ubiquitously observed in diverse brain areas and typically involve the concerted activation of populations of neurons [18]. They can therefore conveniently be studied in the local field potential (LFP) that represents a complementary window to the spiking activity of individual neurons or small groups thereof: It is composed of the superposition of the activity of hundreds of thousands to millions of neurons [19, 20] and forward modeling studies have confirmed [21] that it is primarily driven by the synaptic inputs to the local network [22–24]. As the LFP is a quantity that can be measured relatively easily, this mesoscopic signal is experimentally well documented. Its interpretation is, however, still debated. For example, changes in the amplitude of one of the components of the spectrum of the LFP have been attributed to changes in behavior (cf. e.g. [25]).

A particular entanglement between rates and correlations is the correlated firing of spikes in pairs of neurons in relation to the phase of an ongoing oscillation. With the above interpretation of the LFP primarily reflecting the input to the cells, it is not surprising that the mean firing rate of neurons may modulate in relation to this cycle. The recurrent network model indeed confirms this expectation, as shown in Fig 1A. It is, however, unclear if and by which mechanisms the covariance of firing follows the oscillatory cycle. The simulation shown in Fig 1B indeed exhibits a modulation of the covariance between the activities of pairs of cells. Such modulations have also been observed in experiments:

Denker et al. [26] have shown that the synchronous activation of pairs of neurons within milliseconds preferentially appears at a certain phase of the oscillatory component of the LFP in the beta-range—in their words the spike-synchrony is “phase-locked” to the beta-range of the LFP. They explain their data by a conceptual model, in which an increase in the local



**Fig 1. A** Time-varying mean activity of the excitatory population  $m_E(t) = N_E^{-1} \sum_{i \in E} n_i(t)$  in a balanced EI-network (parameters given in Table 1). Thin gray lines are the outcomes of three independent simulations, the solid black line indicates the mean activity predicted by the theory (Eqs (7) and (8)). Dashed black lines indicate the range of expected fluctuations of the population activity ( $\pm$  one standard deviation): The square of the fluctuation magnitude is given by the variance of the population activity  $\frac{\sigma_E^2}{N_E} + c_{EE}$  (Eqs (3) and (4)). **B** Population-averaged cross covariance  $c_{EE} = \frac{1}{N_E(N_E-1)} \sum_{i \neq j \in E} c_{ij}$ .

<https://doi.org/10.1371/journal.pcbi.1005534.g001>

input, assumed to dominate the LFP, leads to the activation of cell assemblies. The current work investigates an alternative hypothesis: We ask if a periodically-driven random network is sufficient to explain the time-dependent modulation of covariances between the activities of pairs of cells or whether additional structural features of the network are required to explain this experimental observation.

To investigate the mechanisms causing time-dependent covariances in an analytically tractable case, we here present the simplest model that we could come up with that captures the most important features: A local network receiving periodically changing external input. The randomly connected neurons receive sinusoidally modulated input, interpreted as originating from other brain areas and mimicking the major source of the experimentally observed LFP. While it is obvious that the mean activity in a network follows an imposed periodic stimulation, it is less so for covariances. In the following we will address the question why they are modulated in time as well. Extending the analysis of mean activities and covariances in the stationary state [13, 27, 28], we here expose the fundamental mechanisms that shape covariances in periodically driven networks.

Our network model includes five fundamental properties of neuronal dynamics: First, we assume that the state of low and irregular activity in the network [1] is a consequence of its operation in the balanced state [29, 30], where negative feedback dynamically stabilizes the activity. Second, we assume that each neuron receives a large number of synaptic inputs [31], each of which only has a minor effect on the activation of the receiving cell, so that total synaptic input currents are close to Gaussian. Third, we assume the neurons are activated in a threshold-like manner depending on their input. Fourth, we assume a characteristic time scale  $\tau$  that measures the duration of the influence a presynaptic neuron has on its postsynaptic targets. Fifth, the output of the neuron is dichotomous or binary, spike or no spike, rather than continuous. As a consequence, the variance of the single unit activity is a direct function of its mean.

We here show how each of the five above-mentioned fundamental properties of neuronal networks shape and give rise to the mechanisms that cause time-dependent covariances. The presented analytical expressions for the linear response of covariances expose two different paths by which a time-dependence arises: By the modulation of single-unit variances and by the modulation of the linear gain resulting from the non-linearity of the neurons. The

interplay of negative recurrent feedback and direct external drive can cause resonant behavior of covariances even if mean activities are non-resonant. Qualitatively, these results explain the modulation of synchrony in relation to oscillatory cycles that are observed in experiments, but a tight locking of synchronous events to a particular phase of the cycle is beyond the mechanisms found in the here-studied models.

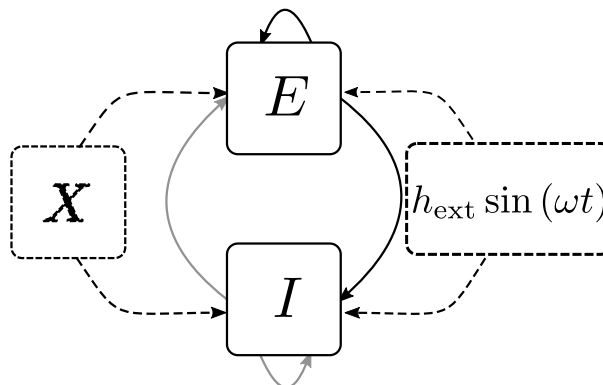
## Results

To understand the locking of synchronous activity to an oscillatory cycle, as observed experimentally, we here need to consider time-dependent network states. We are in particular interested in the covariance between two stochastic variables  $x_1$  and  $x_2$ , which is defined as  $c(t) = \langle \delta x_1(t) \delta x_2(t) \rangle = \langle (x_1(t) - \langle x_1(t) \rangle) (x_2(t) - \langle x_2(t) \rangle) \rangle$ , where  $\langle \dots \rangle$  denotes the average over realizations. In words, the covariance in a time-dependent setting measures the co-variability of a pair of signals with respect to their respective mean. The mean value itself may depend on time. Only if this quantity can be determined with sufficient accuracy, time-dependent covariances can be calculated correctly. This is the source of the technical problems occurring in the context of a time-dependent covariance: It may be hard to assess the covariance, much more its time-dependence, because it is overshadowed by the time-varying mean. If a stochastic model is given, however, disentangling the time dependence of different cumulants, like mean and covariance, is possible. A theoretical study to understand the prevalent mechanisms that cause time-dependent covariances in a network model is therefore a necessary first step. In the following we identify these mechanisms by which time-dependent covariances of activities arise in oscillatory-driven recurrent networks. In Fig 1A we show the population-averaged activity of the excitatory population activity in a balanced EI-network together with the theoretical prediction to be developed in the sequel: The fluctuations around the mean show a wider spread close to the peak of the oscillation than at the trough. Correspondingly, the covariance between pairs of neurons in panel B has its peaks and troughs at points of high and low variability of the population activity in A, respectively.

## Binary network model and its mean field equations

To address our central question, whether a periodically-driven random network explains the experimental observations of time-modulated pairwise covariances, we consider a minimal model here. It consists of one inhibitory (*I*) population and, in the latter part of the paper, additionally one excitatory population (*E*) of binary model neurons [6, 27, 29, 32]. Neurons within these populations are recurrently and randomly connected. All neurons are driven by a global sinusoidal input mimicking the incoming oscillatory activity that is visible in the LFP, illustrated in Fig 2. The local network may in addition receive input from an external excitatory population (*X*), representing the surrounding of the local network. The fluctuations imprinted by the external population, providing shared inputs to pairs of cells, in addition drive the pairwise covariances within the network [13, c.f. especially the discussion]. Therefore we need the external population *X* to arrive at a realistic setting that includes all sources of covariances. In the following, we extend the analysis of cumulants in networks of binary neurons presented in [6, 13, 27, 28, 33] to the time-dependent setting. This formal analysis allows us to obtain analytical approximations for the experimentally observable quantities, such as pairwise covariances, that expose the mechanisms shaping correlated network activity.

Binary model neurons at each point in time are either inactive  $n_i = 0$  or active  $n_i = 1$ . The time evolution of the network follows the Glauber dynamics [34]; the neurons are updated asynchronously. At every infinitesimal time step  $dt$ , any neuron is chosen with probability  $\frac{dt}{\tau}$ . After an update, neuron  $i$  is in the state 1 with the probability  $F_i(\mathbf{n})$  and in the 0-state with



**Fig 2. Recurrent balanced network driven by oscillatory input.** Recurrently connected excitatory ( $E$ ) and inhibitory ( $I$ ) populations (Erdős-Rényi random network with connection probability  $p$ ) receiving input from an external ( $X$ ) excitatory population. Additionally, all neurons in the microcircuit receive a sinusoidal signal of amplitude  $h_{\text{ext}}$  and frequency  $\omega$ , representing the oscillatory activity received from external brain areas.

<https://doi.org/10.1371/journal.pcbi.1005534.g002>

probability  $1 - F_i(\mathbf{n})$ , where the activation function  $F$  is chosen to be

$$\begin{aligned} F_i(\mathbf{n}) &= H(h_i - \theta_i) \\ h_i &= \sum_{k=1}^N J_{ik} n_k + h_{\text{ext}} \sin(\omega t) + \xi_i \\ H(x) &= \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}. \end{aligned} \quad (1)$$

We here introduced the connectivity matrix  $J$  with the synaptic weights  $J_{ij} \in \mathbb{R}$  describing the influence of neuron  $j$  on neuron  $i$ . The weight  $J_{ij}$  is negative for an inhibitory neuron  $j$  and positive for an excitatory neuron. Due to the synaptic coupling the outcome of the update of neuron  $i$  potentially depends on the state  $\mathbf{n} = (n_1, \dots, n_N)$  of all other neurons in the network. Compared to the equations in [13, page 4], we added an external sinusoidal input to the neurons representing the influence of other cortical or subcortical areas and Gaussian uncorrelated noise with vanishing mean  $\langle \xi_i \rangle = 0$  and covariance  $\langle \xi_i \xi_j \rangle = \delta_{ij} \sigma_{\text{noise}}^2$ . The threshold  $\theta_i$  depends on the neuron type and will be chosen according to the desired mean activity.

We employ the neural simulation package NEST [35, 36] for simulations. Analytical results are obtained by mean-field theory [6, 13, 27, 28, 37, 38] and are described for completeness and consistency of notation in the section “Methods”. In the main text we only mention the main steps and assumptions entering the approximations. The basic idea is to describe the time evolution of the Markov system in terms of its probability distribution  $p(\mathbf{n}, t)$ . Using the master Eq 14 we obtain ordinary differential equations (ODEs) for the moments of  $p(\mathbf{n}, t)$ . In particular we are interested in the population averaged mean activities  $m_\alpha$ , variances  $a_\alpha$ , and covariances  $c_{\alpha\beta}$

$$m_\alpha(t) := \frac{1}{N_\alpha} \sum_{i \in \alpha} \langle n_i(t) \rangle \quad (2)$$

$$a_{\alpha}(t) := \frac{1}{N_{\alpha}} \sum_{i \in \alpha} \langle n_i(t) \rangle - \langle n_i(t) \rangle^2 \quad (3)$$

$$c_{\alpha\beta}(t) := \frac{1}{N_{\alpha}N_{\beta}} \sum_{i \in \alpha, j \in \beta, i \neq j} \langle n_i(t)n_j(t) \rangle - \langle n_i(t) \rangle \langle n_j(t) \rangle, \quad (4)$$

which are defined as expectation values  $\langle \rangle$  over realizations of the network activity, where the stochastic update of the neurons and the external noisy input presents the source of randomness in the network. The dynamics couples moments of arbitrarily high order [33]. To close this set of equations, we neglect cumulants of order higher than two, which also approximates the input by a Gaussian stochastic variable with cumulants that vanish for orders higher than two [39]. This simplification can be justified by noticing that the number of neurons contributing to the input is large and their activity is weakly correlated, which makes the central limit theorem applicable. In a homogeneous random network, on expectation there are  $K_{\alpha\beta} = p_{\alpha\beta} N_{\beta}$  synapses from population  $\beta$  to a neuron in population  $\alpha$ . Here  $p_{\alpha\beta}$  is the connection probability; the probability that there is a synapse from any neuron in population  $\beta$  to a particular neuron in population  $\alpha$  and  $N_{\alpha}$  is the size of the population. Mean Eq (2) and covariance Eq (4) then follow the coupled set of ordinary differential equations (ODEs, see section II A in S1 Text for derivation)

$$\tau \frac{d}{dt} m_{\alpha}(t) = -m_{\alpha}(t) + \varphi(\mu_{\alpha}(\mathbf{m}(t), h_{\text{ext}} \sin(\omega t)), \sigma_{\alpha}(\mathbf{m}(t), c(t))) \quad (5)$$

$$\begin{aligned} \tau \frac{d}{dt} c_{\alpha\beta}(t) = & \left\{ -c_{\alpha\beta}(t) + \sum_{\gamma} \left[ S(\mu_{\alpha}(\mathbf{m}(t), h_{\text{ext}} \sin(\omega t)), \sigma_{\alpha}(\mathbf{m}(t), c(t))) \right. \right. \\ & \left. \left. \times K_{\alpha\gamma} J_{\alpha\gamma} \left( c_{\gamma\beta}(t) + \delta_{\gamma\beta} \frac{a_{\beta}(t)}{N_{\beta}} \right) \right] \right\} + \{\alpha \leftrightarrow \beta\}, \end{aligned} \quad (6)$$

where  $\alpha \leftrightarrow \beta$  indicates the transposed term. The Gaussian truncation employed here is parameterized by the mean  $\mu_{\alpha}$  and the variance  $\sigma_{\alpha}^2$  of the summed input to a neuron in population  $\alpha$ . These, in turn, are functions of the mean activity and the covariance, given by Eqs (18) and (19), respectively.

Here  $\varphi$  is the expectation value of the activation function, which is smooth, even though the activation function itself is a step function, therefore not even continuous. The function  $\varphi$  fulfills  $\lim_{m \rightarrow 0} \varphi = 0$  and  $\lim_{m \rightarrow 1} \varphi = 1$  and monotonically increases. Its derivative  $S$  with respect to  $\mu$  has a single maximum and is largest for the mean input  $\mu$  within a region with size  $\sigma$  around the threshold  $\theta$ .  $S$  measures the strength of the response to a slow input and is therefore termed susceptibility. The definitions are given in “Methods” in Eqs (17) and (20).

The stationary solution (indicated by a bar) of the ODEs Eqs (5) and (6) can be found by solving the equations

$$\bar{\mathbf{m}} = \varphi(\bar{\mathbf{m}}) \quad (7)$$

$$2\bar{c} = SKJ \left( \bar{c} + \frac{\bar{a}}{N} \right) + \text{transposed} \quad (8)$$

numerically and self-consistently, as it was done in [13, 27, 33].



The full time-dependent solution of Eqs (5) and (6) can, of course, be determined numerically without any further assumptions. Besides the comparison with simulation results, this will give us a check for the subsequently applied linear perturbation theory. The resulting analytical results allow the identification of the major mechanisms shaping the time-dependence of the first two cumulants. To this end, we linearize the ODEs Eqs (5) and (6) around their stationary solutions. We only keep the linear term of order  $h_{\text{ext}}$  of the deviation, justifying a Fourier ansatz for the solutions. For the mean activities this results in

$$m_{\alpha}(t) = \bar{m}_{\alpha} + \delta m_{\alpha}(t) = \bar{m}_{\alpha} + M_{\alpha}^1 e^{i\omega t} \text{ with}$$

$$M_{\alpha}^1 = \sum_{\beta} U_{\alpha\beta} M_{\beta}^1 = \sum_{\beta} U_{\alpha\beta} \frac{h_{\text{ext}} (U^{-1} S(\bar{\mu}, \bar{\sigma}))_{\beta} (-i\tau\omega + 1 - \lambda_{\beta})}{(\tau\omega)^2 + (1 - \lambda_{\beta})^2}. \quad (9)$$

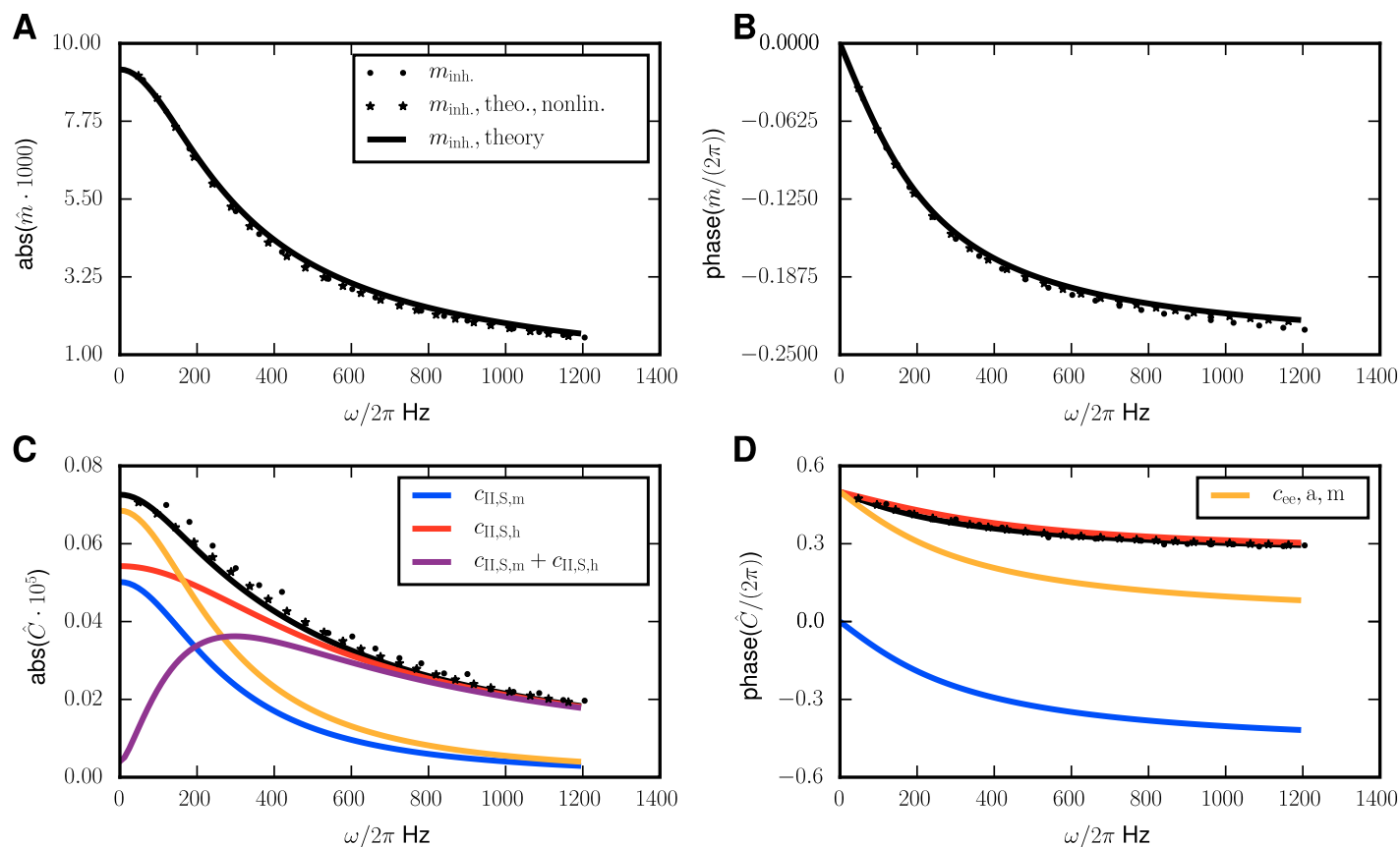
The time-dependence of  $\sigma$  was neglected here, which can be justified for large networks (“Methods”, Eqs (22) and (30)). The matrix  $U$  represents the basis change that transforms  $\bar{W}_{\alpha\beta} := S(\bar{\mu}_{\alpha}, \bar{\sigma}_{\alpha}) K_{\alpha\beta} J_{\alpha\beta}$  into a diagonal matrix with  $\lambda_{\alpha}$  the corresponding eigenvalues. We see that, independent of the number of populations or the detailed form of the connectivity matrix, the amplitude of the time-dependent part of the mean activities has the shape of a low-pass-filtered signal to first order in  $h_{\text{ext}}$ . Therefore the phase of  $\delta m$  lags behind the external drive and its amplitude decreases asymptotically like  $\frac{1}{\omega}$ , as can be seen in Fig 3A and 3B.

If we also separate the covariances into their stationary part and a small deviation that is linear in the external drive,  $c_{\alpha\beta}(t) = \bar{c}_{\alpha\beta} + \delta c_{\alpha\beta}(t)$ , expand  $S(\mu_{\alpha}(t), \sigma_{\alpha}(t))$  and  $a(t)$  around their stationary values, keeping only the terms of order  $h_{\text{ext}}$  and neglect contributions from the time-dependent variation of the variance of the input  $\sigma^2$  (see “Methods”, especially Eq (30) for a discussion of this point), we get the ODE

$$\begin{aligned} & \tau \frac{d}{dt} \delta c(t) + 2\delta c(t) - \bar{W} \delta c(t) - (\bar{W} \delta c(t))^T \\ &= \underbrace{\left\{ \bar{W} \text{diag}\left(\frac{1 - 2\bar{m}}{N}\right) \text{diag}(\delta m(t)) \right\}}_{\text{modulated-autocorrelations-drive}} \\ &+ \left[ \underbrace{\text{diag}(K \odot J \delta m(t))}_{\text{recurrent drive}} + \underbrace{h_{\text{ext}} \sin(\omega t)}_{\text{direct drive}} \right] \text{diag}\left(\frac{\partial S}{\partial \mu}(t)\right) K \odot J \bar{c}^{\text{total}} \\ &+ \{ \dots \}^T, \end{aligned} \quad (10)$$

where we introduced the point-wise (Hadamard) product  $\odot$  of two matrices  $A$  and  $B$  [see 40, for a consistent notation of matrix operations] as  $(A \odot B)_{ij} := A_{ij} B_{ij}$ , defined the matrix with the entries  $\text{diag}(x)_{ij} := \delta_{ij} x_i$  for the vector  $x = (x_1, \dots, x_n)$  and set  $\bar{c}^{\text{total}} := \bar{c} + \text{diag}(\frac{\bar{a}}{N})$  to bring our main equation into a compact form.

We can now answer the question posed in the beginning: Why does a global periodic drive influence the cross covariances in the network at all and does not just make the mean activities oscillate? First, the variances are modulated with time, simply because they are determined via Eq (3) by the modulated mean activities. A neuron  $i$  with modulated autocorrelation  $a_i(t)$  projects via  $J_{ji}$  to another neuron  $j$  and therefore shapes the pairwise correlation  $c_{ji}(t)$  in a time-dependent way. We call this effect the “modulated-autocovariances-drive”, indicated by the curly brace in the second line of Eq (10). Its form in index notation is  $[\bar{W} \text{diag}((1 - 2\bar{m})/N) \text{diag}(\delta m(t))]_{\alpha\beta} = \bar{W}_{\alpha\beta} (1 - 2\bar{m}_{\beta})/N_{\beta} \delta m_{\beta}(t)$ . This is the low-pass-filtered input.



**Fig 3. Periodically driven single population network.** Dependence of the modulations of the mean activity and covariances on the driving frequency  $\omega$ . **A** Amplitude of modulation of mean activity. **B** Phase of modulation of mean activity relative to the external drive. **C** Amplitude of modulation of covariances. **D** Phase of modulation of covariance relative to the external drive. In all panels, the analytical predictions (Eqs (9) and (39)) are shown as solid black curves. The black curve is the complete solution. The different contributions to the time-dependent covariances, identified in Eq (12), are shown separately: The  $S_n$ -term in red, the  $S_m$ -term in blue, their sum in purple, and the  $a$ -term in orange. Numerical solutions of the full mean-field equations (Eqs (5) and (6)) are shown as stars and simulation results by dots (only indicated in the legend of A). The numerical results are obtained by using the integrate.ode-method from the python-package scipy [41] with the option "lsoda", meaning that either implicit Adams- or backward differentiation-algorithms (depending on the given problem) are used. Network parameters: Number of neurons  $N_I = 5000$ , connection probability  $p_{II} = 0.1$ , coupling strength  $J_{II} = -1$ , mean activity  $m_I \approx 0.3$ , and  $\sigma_{\text{noise}} = \sigma_{\text{system}} := \sqrt{J_{II}^2 p_{II} N_I m_I (1 - m_I)} \approx 10.2$ .

<https://doi.org/10.1371/journal.pcbi.1005534.g003>

The other contributions are a bit more subtle and less obvious, as they are absent in networks with a linear activation function. The derivative of the expectation value of the activation function, the susceptibility, contributes linearly to the ODE of the covariances. As the threshold-like activation function gives rise to a nonlinear dependence of  $\varphi$  on the mean input  $\mu$ , the susceptibility  $S = \varphi'$  is not constant, but depends on the instantaneous mean input. The latter changes as a function of time by the direct external drive and by the recurrent feedback of the oscillating mean activity, indicated by the terms denoted by the curly braces in the third line of Eq (10). Together, we call these two term the "susceptibility terms". Both terms are of the same form

$$\left[ \text{diag}(\delta\mu(t)) \text{diag}\left(\frac{\partial S}{\partial \mu(t)}\right) K \otimes J \bar{\tau}^{\text{total}} \right]_{\alpha\beta} = \delta\mu_{\alpha}(t) \frac{\partial S_{\alpha}}{\partial \mu_{\alpha}} \sum_{\gamma} K_{\alpha\gamma} J_{\alpha\gamma} (\bar{c}_{\gamma\beta} + \delta_{\gamma\beta} \frac{\bar{a}_{\beta}}{N_{\beta}}), \quad (11)$$

but with different  $\delta\mu_{\alpha}$ . This form shows how the time-dependent modulation of the mean input  $\delta\mu_{\alpha}$  by the second derivative of the gain function  $\frac{\partial S_{\alpha}}{\partial \mu_{\alpha}} = \varphi''$ , influences the transmission of



covariances. The sum following  $\frac{\partial S_c}{\partial \mu_s}$  is identical to the one in the static case Eq (8). For the “recurrent drive”, the time-dependent input is given by  $\delta\mu_\alpha(t) = \sum_\beta K_{\alpha\beta} J_{\alpha\beta} \delta m_\beta(t)$ , which is a superposition of the time-dependent activities that project to population  $\alpha$  and is therefore low-pass-filtered, too. The term due to “direct drive” is  $\delta\mu_\alpha(t) = h_{\text{ext}} \sin(\omega t)$ .

We solve Eq (10) by transforming into the eigensystem of  $\bar{W}$  and inserting a Fourier ansatz,  $\delta c_{\alpha\beta}(t) = C_{\alpha\beta}^1 e^{i\omega t}$ . The solution consists of a low-pass filtered part coming from the direct drive and two parts that are low-pass filtered twice, coming from the recurrent drive and the modulated-autocovariances-drive. For a detailed derivation, consult the section “Covariances: Stationary part and response to a perturbation in linear order”.

We have calculated higher Fourier modes of the simulated network activity and of the numerical solution of the mean-field equations to check if they are small enough to be neglected, so that the response is dominated by the linear part. Of course, it would be possible to derive analytical expressions for those as well. However, we will see that the linear order and the corresponding first harmonic qualitatively and for remarkably large perturbations even quantitatively gives the right predictions. The limits of this approximation are analyzed in Fig D in S1 Text. We will therefore constrain our analysis to controlling the higher harmonics through the numerical solution.

In the following we will study three different models of balanced neuronal networks to expose the different mechanisms in their respective simplest setting.

**Single population.** As a first example, we quantitatively study the particular case of a single population, which has to be inhibitory to ensure stable stationary activity. Let us look at the behavior of the different contributions in Eq (10) to the modulated covariance and their mutual relation. Written explicitly, the terms driving the time variation of the covariance are

$$\overbrace{\left( \underbrace{K J \delta m(t)}_{S_m\text{-term} \propto \frac{1}{\omega} \text{ for big } \omega} + \underbrace{h_{\text{ext}} \sin(\omega t)}_{S_h\text{-term does not scale with } \omega} \right)}^{\text{susceptibility terms; partly cancel}} \underbrace{\frac{\partial S}{\partial \mu} K J \left( \bar{c} + \frac{\bar{a}}{N} \right)}_{\text{partly cancel}} + \underbrace{\bar{W}(1 - 2\bar{m}) \frac{\delta m(t)}{N}}_{a\text{-term} \propto \frac{1}{\omega} \text{ for big } \omega}. \quad (12)$$

With respect to their dependence on the number of synaptic connections  $|K|$ , the sum of the two susceptibility terms is of the same order of magnitude as the modulated-autocovariances-drive (cf. 37 in the section “Methods”), therefore their interplay determines the shape of the solution of Eq (10) and we cannot neglect either term in favor of the other.

To analyze the contributions to  $\delta c$ , it is reasonable to first focus on the quasi-static case  $\omega \rightarrow 0$ , because its analysis is simplest and, due to the continuity of the observed quantities, it carries over to the case of biologically relevant small frequencies up to the  $\beta$ -range. For  $\omega \rightarrow 0$ , the solution  $\delta c$  in Eq (10) has the same sign as the sum of the inhomogeneities, because it is given by a multiplication with  $0.5(1 - W)^{-1}$ , where  $W < 0$ . The main information—especially about the sign—is therefore already included in these inhomogeneities, that we termed “recurrent drive” and “direct drive” (the susceptibility terms) and “modulated-autocovariances-drive” in the previous section. The modulation of the covariance  $\delta c(t)$  then results by low pass filtering their sum. Individually they yield the  $S_m$ -term and  $S_h$ -term (together the  $S$ -terms) and the  $a$ -term, respectively.

In a general balanced network, the deviation of the mean activity from the stationary solution  $\delta m(t)$  is in phase with the perturbation for  $\omega \approx 0$  and lags behind it for larger  $\omega$  due to the “forgetfulness” of the network caused by the leak term in the ODE. At low frequencies, the recurrent drive  $\propto K \circledast J \delta m(t)$  therefore partly cancels the direct drive  $\propto h_{\text{ext}} \sin(\omega t)$ . This is because the rate response  $\delta m$  is in phase, and the feedback  $KJ < 0$  in the network is negative. The cancellation becomes less efficient at larger frequencies, because the recurrent drive

asymptotically decays like  $\omega^{-1}$  and is phase-shifted; the mean activity is low-pass-filtered Eq (9). The direct drive, in contrast, does not depend on the driving frequency  $\omega$ . Therefore, the  $S_m$ -term is low-pass-filtered twice and the  $S_h$ -term term only once, therefore their sum has a peak at an intermediate frequency, as visible in Fig 3C, purple curve. Note that this cancellation generally appears in the balanced state, because the network feedback is always effectively inhibitory. Furthermore, the modulated-autocovariances-drive only vanishes for  $\overline{m}_x = \frac{1}{2}$ ; for realistic activity  $\overline{m}_x \ll \frac{1}{2}$  it is in anti-phase with  $\delta m(t)$ , because it is defined including  $W < 0$ , which flips the phase by  $\pi$ .

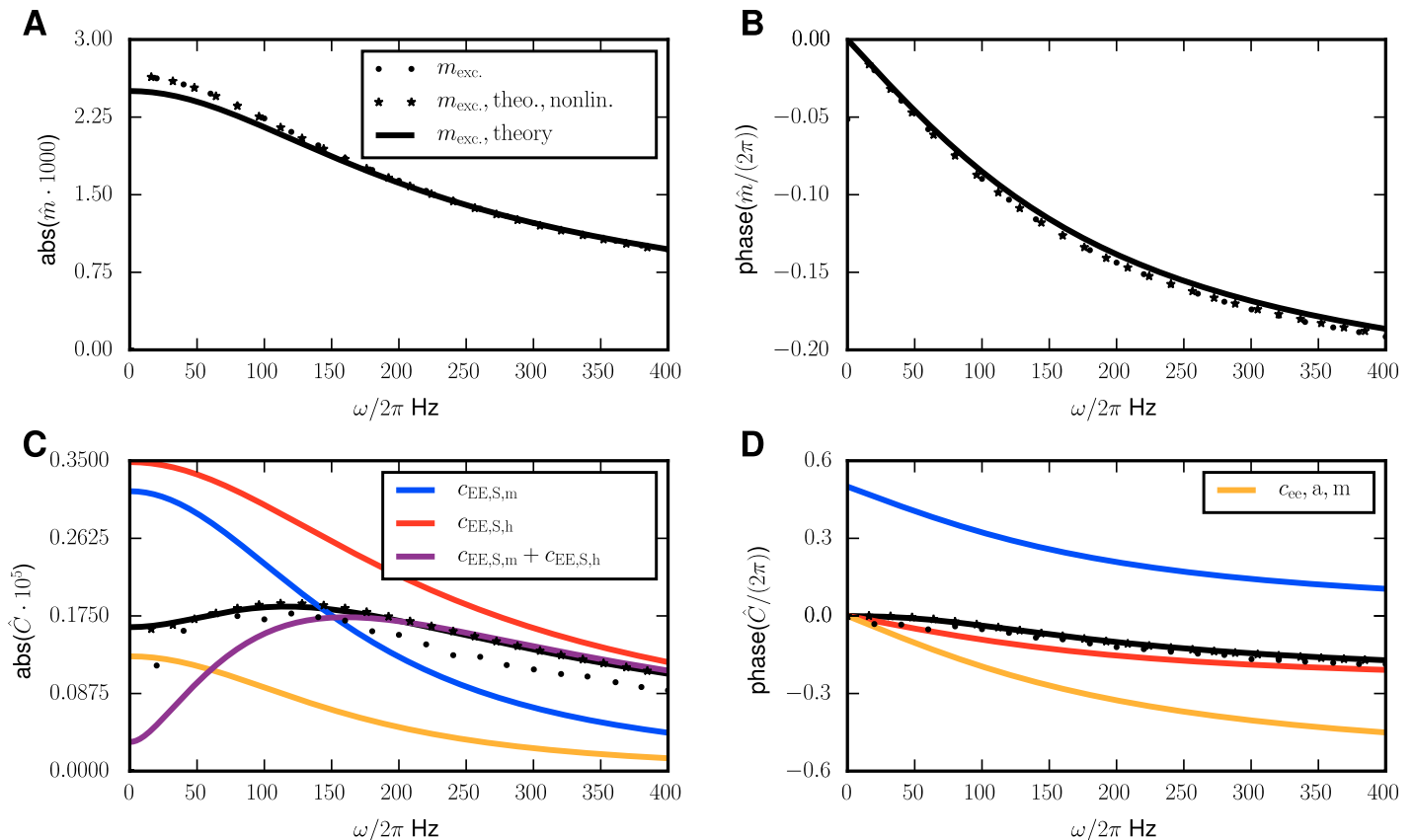
Average covariances in inhibitory networks are negative [13]. As a consequence, in the setting of a single inhibitory population there is a second kind of cancellation: The two terms  $\bar{c}$  and  $N^{-1}\bar{a}$  in the prefactor  $\bar{c} + N^{-1}\bar{a}$  of the susceptibility terms in Eq (12) partly cancel; their sum in fact vanishes in the large  $N$  limit [cf. 13, eq. (32) and their Fig 5]. This leads to the dominance of the  $a$ -term, shown in Fig 3D (orange curve). The maximum in the  $S$ -terms is therefore overshadowed by the  $a$ -term, which asymptotically also shows a second order low pass characteristics with  $\propto \omega^{-2}$ . So in the purely inhibitory network the peak is not visible in the sum of all contributions to  $\hat{C}(\omega)$ .

In summary, the model of a single population in the balanced state exposes several generic features of time-dependent mean activities and covariances: Mean activities and the direct drive contribution to covariances follow the external modulation with first order low pass characteristics. The  $S_m$ -term and the  $a$ -term of the covariances, being mediated by the mean activity, consequently expose a second order low pass filtering. The direct drive and the recurrent drive (the susceptibility terms) to large extent cancel at low frequencies, but not at high ones. Due to their overall decay in amplitude with increasing frequency, an intermediate maximum arises in their sum. In the single population model this peak is typically overshadowed by the  $a$ -term. This is because of the suppression of population fluctuations by negative feedback in the stationary state [10], which causes a small population variance  $N^{-1}\bar{a} + \bar{c}$  and the latter term controls the amplitude of the susceptibility terms.

**Two homogeneously connected populations.** A slightly more realistic, but still simple setup is an EI-network with the same input for the inhibitory and the excitatory neurons, as studied before, in [13, parameters, except  $m_x$  as in fig. 6 there]. This network is also inhibition-dominated, therefore we observe qualitatively the same competition of the two  $S$ -terms leading to the existence of a maximum in the  $\omega$ -dependence of  $|C_i|$ . In contrast to the single population case, in the E-I network the peak may be visible. This is because—in contrast to the single population case—covariances in this setup may also be positive, preventing the cancellation with the variances in the term  $\bar{c} + \frac{\bar{a}}{N}$  that drives the  $S$ -terms. The latter contribution may therefore dominate over the  $a$ -term at small  $\omega$ . Its dominance increases the larger the covariances are, which for example arises when increasing the external drive or by lowering the noise level at the input to the neurons. The “resonance” effect itself increases for weaker the excitatory synapses.

Fig 4C, indeed shows a peak of the response of the covariances at a frequency of about 120 Hz. We here focus on the covariances between excitatory neurons, because their activities are recorded most often and also cell assemblies are normally assumed to consist of excitatory neurons.

**Two populations with inhomogeneous connections.** The example of homogeneous connectivity helps to explain the fundamental mechanisms that shape the covariances; it is, however, certainly not very realistic. Furthermore, in the case of synaptic weights being different for individual receiving populations, the linearized connectivity  $W$  can have a pair of complex eigenvalues, which is qualitatively different to the setup described before. To check if the theory also works for parameters satisfying biological constraints, we choose the



**Fig 4. Periodically driven E-I network.** **A** Amplitude of modulation of the mean activity deviating from the stationary value for the excitatory population. **B** Phase of the modulation of the mean activity. **C** Different contributions to the amplitude of the covariance between pairs of excitatory cells in dependence of the frequency  $\omega$  of the external drive. **D** Phase of covariances relative to the driving signal. Analytical theory (Eqs (9) and (39)) shown by solid black curves, numerical solutions of the full mean field equations (Eqs (5) and (6)) (stars) and simulation results (dots, only indicated in the legend of A). Same color code as in C. In C and D, the contributions to the variation of covariances are shown separately: The  $S_h$ -term in red, the  $S_m$ -term in blue, their sum in purple and the  $a$ -term in yellow. The legend for C and D is split over both panels. Numerical solutions obtained by the same methods as in Fig 3. Parameters:  $N_E = N_I = N_X = 8192$ ,  $p_E = p_I = p_X = p = 0.2$ ,  $m_E = m_I \approx 0.11$ ,  $m_X = 0.25$ , identical to [13, e.g. fig. 6].

<https://doi.org/10.1371/journal.pcbi.1005534.g004>

connectivity and activity levels in accordance to experimental studies. Apart from the results from [26], the parameters were measured in the layer 2/3 in the barrel cortex of mice. We select this layer, because it is the assumed location of cell assemblies [42], allowing us to relate our results to the original hypothesis of excess synchrony by activation of assemblies [26], a feature that could be considered in future studies. The connection probabilities are taken from [43], the fractions of excitatory and inhibitory neurons from [44] and the membrane time constant is extracted from [45, supplementary material]. We adjust the neurons' thresholds such that the stationarity condition  $\varphi(\bar{m}) = \bar{m}$  is fulfilled for  $m_\alpha = \tau v_\alpha$ , where  $\alpha \in \{\text{exc., inh.}\}$ ,  $v_\alpha$  is the firing rate of the respective population and  $\tau$  is the neuronal time constant. Note that the mapping  $m = \tau v$  implies a slightly different notion of a "spike" of a binary neuron than previously used [28]. The two conventions agree in the limit of vanishing firing rates (cf. section II B in S1 Text). The firing rate of 18 Hz given in [26] presumably reflects the activity of excitatory neurons (private communication). To obtain the firing rate of the inhibitory neurons  $v_{\text{inh.}}$ , we scale the measurement from [26] by the ratio  $v_{\text{inh.}}/v_{\text{exc.}}$  from [46]. All parameters are summarized in Table 1. The effective connectivity  $W$  of this

**Table 1. Parameters for the biologically inspired network model used in Figs 5 and 6 and Figs A, B, C and D in S1 Text.**

		exc.	inh.	$\nu$ (Hz)	mean act.	#neurons	
exc.	connection prob.	0.168	0.5	18	0.045	1691	$\tau = 2.5\text{ms}$
	synaptic weight	0.37	-0.52				
inh.	connection prob.	0.327	0.36	108	0.27	230	$m_{\text{ex}} = 0.1$
	synaptic weight	0.82	-0.54				

<https://doi.org/10.1371/journal.pcbi.1005534.t001>

system has two conjugate complex eigenvalues. Therefore, there exists a resonance frequency also for the mean activity, shown in Fig 5C.

In the two upper panels of Figs 5 and 6, we compare the stationary values for the mean activity Eq (7) and the covariances Eq (8) with the respective time averaged results of the simulation and with the numerical solution of the full mean-field equations. The stationary statistics have been investigated before for other parameters in finite networks [13] and in the limit  $N \rightarrow \infty$  [6]. The second harmonics extracted from the simulations and the numerical solution of the full mean-field equations show good agreement and are overall small compared to the zeroth and first harmonics, justifying the truncation of the Fourier series in the analytical theory after the first term.

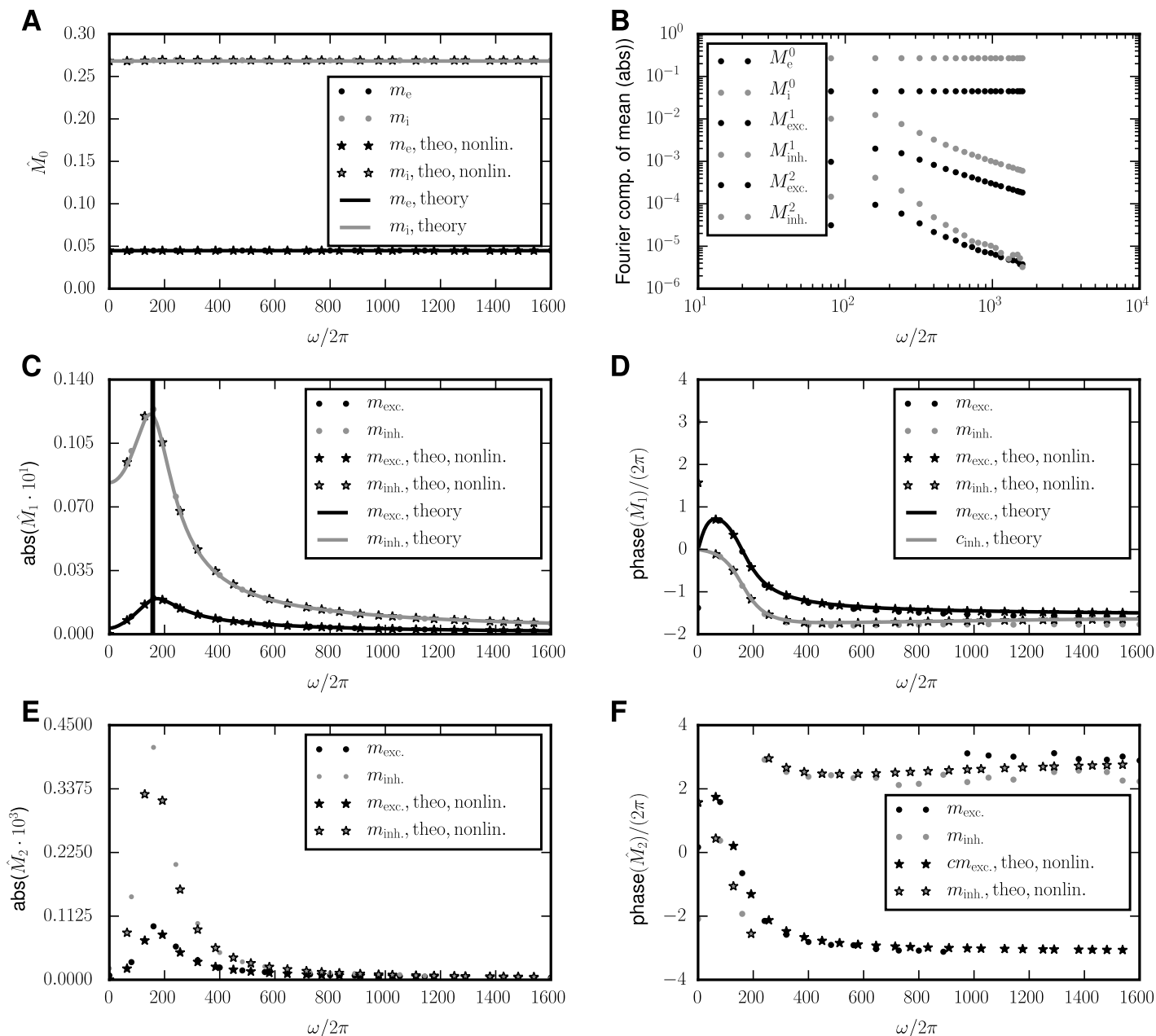
The first harmonics of the mean activity (see Fig 5) and covariances (see Fig 6) predicted by the linear response theory agree well with simulations and the numerical solution. This is not necessarily clear a priori because the perturbation in the input to every neuron is of the order  $\mathcal{O}(\frac{\sigma}{10})$ , where  $\sigma$  is the input noise level of the unperturbed system. However, linear response theory works surprisingly well, even for the covariances caused by a perturbation leading to a response of the same order of magnitude as the stationary value. Increasing the perturbation strength  $h_{\text{ext}}$  further ultimately leads to a breakdown of the linear perturbation theory, visible in the growing absolute values of the second Fourier modes of mean activities and covariances (Fig D in S1 Text). The maximal modulation in the firing rates amounts to  $\approx 0.8$  for the excitatory and 4.9 Hz for the inhibitory neurons.

In this biologically inspired setting, it is also interesting to apply the Unitary Event (UE) analysis to our data, as it was done for experimental data in [26]. Because this is a bit aside the scope of this paper, we present this part in the appendix, Sec. I in S1 Text.

The connectivity matrix has complex eigenvalues  $\lambda_1 = \lambda_2^*$ , so we observe a resonance of the mean activities at the frequencies

$$f_{\text{res,mean}} = \frac{\Im(\lambda_1)}{\tau 2\pi},$$

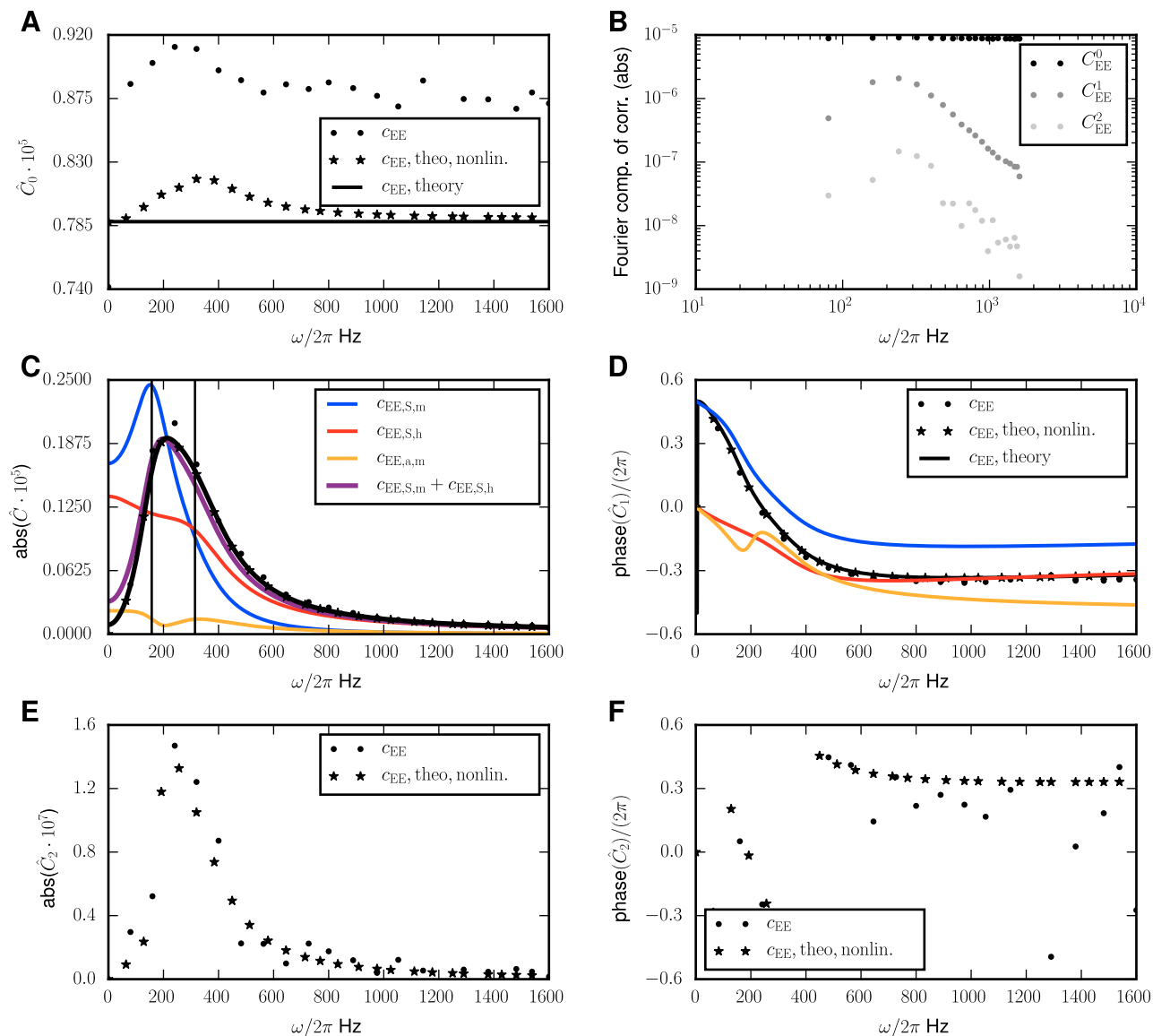
indicated by a vertical line in Fig 5C. The components of  $\delta m$  are composed of different modes, therefore their maximum does not appear exactly at  $f_{\text{res,mean}}$ . The covariances are shaped by more modes: In general, the covariance matrix for a three-dimensional quantity has 6 independent components. In our case,  $c_{XX}$  is always 0, which is a consequence of the missing feedback to  $X$ . Now, the evolution of every mode of  $\tilde{\delta c}$  is given by the sum of two eigenvalues of  $1 - W$ , i.e.  $2 - \lambda$ ,  $2 - \lambda^*$ ,  $2 - 2\lambda$ ,  $2 - 2\lambda^*$  and  $2 - (\lambda^* + \lambda)$ . The missing mode is the “trivial” one owing to the vanishing eigenvalue of  $W$ . So the behavior of the “kernel” of the ODE for  $\delta c$  is given by the resonances at  $\frac{\Im(\lambda)}{\tau 2\pi}$  and  $2 \cdot \frac{\Im(\lambda)}{\tau 2\pi}$ . In addition, the inhomogeneity of the ODE (10) (its right hand side) is already resonant at  $\frac{\Im(\lambda)}{\tau 2\pi}$ . All these modes are mixed with different strength in the different modes of  $\delta c$ , giving rise to a maximum of  $|C_1|$  somewhere in the vicinity of  $f_{\text{res,mean}}$  and  $2f_{\text{res,mean}}$ . In all cases the “resonances” are damped, therefore, a resonance



**Fig 5. Driven E-I network with biologically inspired parameters: Mean activity.** From the first to the third row, the zeroth to second Fourier mode of the mean activity is shown. **A** Constant part of mean activity (zeroth order). **B** First three Fourier-modes of the mean activities on a loglog-scale. **C** Amplitude of first mode of the mean activity. **D** Phase of first mode relative to driving signal. **E** and **F** are structured analogous to **C** and **D** for the second Fourier modes. Solid curves indicate the linear theory (Eq (9)), stars numerical integration of the full mean field equations (Eqs (5) and (6)) and dots the simulation results of the full network. Black symbols indicate the activity of excitatory, gray symbols of inhibitory neurons. Numerical results obtained by the same methods as in Fig 3. Noise amplitudes  $\sigma_{\text{noise},E} = \sigma_{\text{noise},I} = 10$ ,  $\sigma_{\text{network},E} = 2.8$ ,  $\sigma_{\text{network},I} = 4.6$ , other parameters of the network model given in Table 1.

<https://doi.org/10.1371/journal.pcbi.1005534.g005>

catastrophe, induced by  $\delta m$  oscillating with the resonance frequency of  $\delta c$ , cannot occur. We also notice that all resonances are the stronger, the closer  $\Re(\lambda)$  is to 1, the critical point, which makes sense intuitively: The damping comes from the overall inhibitory feedback; at the critical point the leak term is exactly compensated by positive feedback of identical magnitude. It is

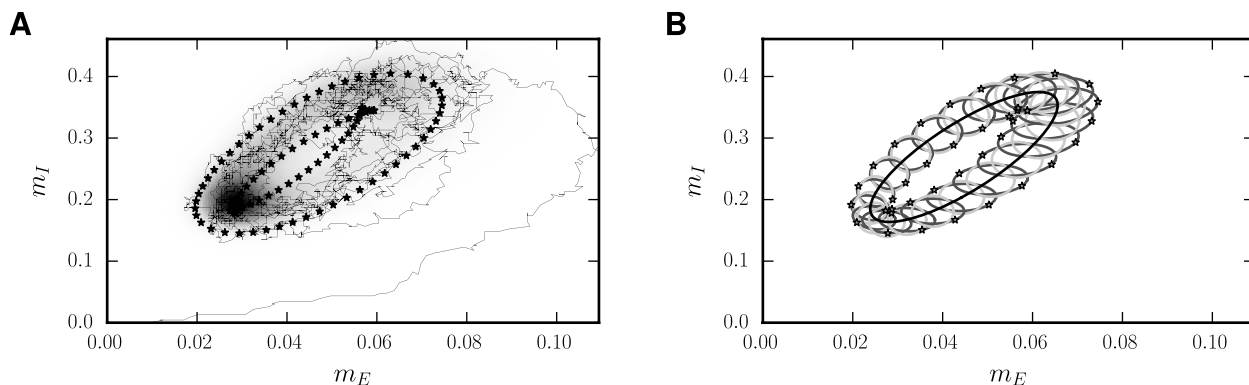


**Fig 6. Driven E-I network with biologically inspired parameters: EE-Covariance.** Response of the covariance to a perturbation with frequency  $\omega$  in the Fourier space. **A** Zeroth Fourier mode (time independent part) of the covariance. **B** Absolute value of the first three Fourier components of the  $c_{EE}$ -covariances on a loglog-scale. **C** Absolute value of the first order of the time-dependent part of the covariance. **D** Phase angle in relation to the driving signal. **E** and **F** are analogous to **C** and **D** for the second Fourier modes. Solid lines indicate the linear theory Eq (39), stars the results of the numerical solution of the full mean-field theory Eqs (5) and (6) and dots the direct simulation of the full network. Numerical results obtained by the same methods as in Fig 3. Parameters of the network model as in Fig 5.

<https://doi.org/10.1371/journal.pcbi.1005534.g006>

worth noticing that the effect of the partial cancellation of the S-terms, which can be read off from Eq (10) and is described in the previous subsections for small  $\omega$  is still valid. The functional form of  $|C_1(\omega)|$ , however, is now mainly determined by the resonances due to the complex eigenvalues of  $W$ .

The  $\omega$ -dependencies of the  $c_{II}$ - and the  $c_{EI}$ -covariances shown in the appendix are qualitatively similar (Fig. B and Fig. C in S1 Text). The stationary covariance is well predicted by the theory [13], which is confirmed here.



**Fig 7. Distribution of population-averaged activity of periodically driven E-I network.** **A** Empirical density of population activity of the E-I network. Gray shading indicates time-averaged occupation of states. The thin mid gray curve is a sample of the binary dynamics of 10 periods after the start of the simulation. The black dots indicate the  $\sigma$ -region predicted by the linear theory as described by Eq (13) in the main text. **B** Limit cycle of the linear theory (black ellipse), together with error ellipses stemming from the sum of covariances and variances (dark gray, slightly tilted) and representing solely variances (light gray). The stars are at the same places as in A. Parameters are given in Table 1, only the perturbation strength is increased to  $h_{\text{ext}} = 6$  (noise level around  $\sigma_E \simeq 14$ ,  $\sigma_I \simeq 23$ ) for reasons of readability (for this value the simulated results already show deviations from the linear approximation as shown in Fig. D in S1 Text). The perturbing frequency is chosen to be  $f = 80$  Hz.

<https://doi.org/10.1371/journal.pcbi.1005534.g007>

Fig 7 illustratively summarizes the results of this section. In panel A, the probability of the binary system to be in a certain activity state  $(m_{\text{inh}}, m_{\text{exc}})^T$  is indicated by different gray shades, the darker, the higher the probability to find it in the respective area. On top, the area including the most probable network states, as predicted by the linear theory, is indicated by black dots. Its construction is depicted in panel B: We draw the limit cycle (black) formed by the points  $(\langle m_{\text{inh}}(t) \rangle, \langle m_{\text{exc}}(t) \rangle)^T$  as a parametric plot with time as parameter. Then, we define the points on the error ellipse  $(m_{\text{inh}}, m_{\text{exc}})^T$  as follows

$$\delta \mathbf{m}(t)^T (c^{\text{pop}}(t))^{-1} \delta \mathbf{m}(t) = 1, \quad (13)$$

where  $\delta \mathbf{m}^T := (m_{\text{inh}}, m_{\text{exc}}, 0)^T - (\langle m_{\text{inh}} \rangle, \langle m_{\text{exc}} \rangle, 0)^T$  and

$$c^{\text{pop}}(t) = \begin{pmatrix} c_{\text{EE}}^{\text{pop}}(t) & c_{\text{EI}}^{\text{pop}}(t) & c_{\text{EX}}^{\text{pop}}(t) \\ c_{\text{EI}}^{\text{pop}}(t) & c_{\text{II}}^{\text{pop}}(t) & c_{\text{IX}}^{\text{pop}}(t) \\ c_{\text{EX}}^{\text{pop}}(t) & c_{\text{IX}}^{\text{pop}}(t) & 0 \end{pmatrix}.$$

In this way, the solutions  $\delta \mathbf{m}(t)$  of Eq (13) are composed of all points that are one standard deviation away from the expected activity. The covariances enter the total population averaged variability, given by

$$\begin{aligned} c_{\alpha\beta}^{\text{pop}}(t) &:= \langle \delta m_{\alpha}(t) \delta m_{\beta}(t) \rangle = \left\langle \frac{1}{N_{\alpha}} \sum_{i \in \alpha} \delta n_i(t) \frac{1}{N_{\beta}} \sum_{i \in \beta} \delta n_i(t) \right\rangle \\ &= \frac{\delta_{\alpha\beta}}{N_{\alpha}^2} \sum_{i \in \alpha} \langle \delta n_i^2(t) \rangle + \frac{1}{N_{\alpha} N_{\beta}} \sum_{i \in \alpha, j \in \beta, i \neq j} \langle \delta n_i(t) \delta n_j(t) \rangle \\ &\approx \delta_{\alpha\beta} \frac{a_{\alpha}(t)}{N_{\alpha}} + c_{\alpha\beta}(t) \end{aligned}$$

with the definitions Eqs (3) and (4).



The two points on the border of the dark gray error-ellipses of the full covariances with the largest distance to the tangent of the limit cycle at  $(\langle m_{\text{inh}} \rangle, \langle m_{\text{exc}} \rangle)$  are marked by a star, which, taken together, form the border of a tube-shaped  $\sigma$ -area. This tube indicates the region in which we most likely expect to find the system. To visualize the contributions of auto- and pairwise covariances, we plot in light gray the error ellipses based solely on the variances ( $c^{\text{pop}}(t)$  is diagonal in this case). The dark error ellipses are bigger than the light ones, indicating that the covariances are positive and their axes are tilted; the  $c_{EI} = c_{IE}$ -component is non-zero. Furthermore, the error ellipses significantly change their size in time, indicative of the modulation of the fluctuations with time. The variances grow monotonically with the respective mean activities, explaining that the light gray ellipses are largest (smallest) where the mean activities are largest (smallest). One can read off the phase shift of  $c_{EE}$  to  $m_E$  to be roughly  $\frac{\pi}{2}$ : the deviation of the dark gray error ellipses from the light gray ones is largest at the points where  $m_E(t) \approx \bar{m}_E$  and  $\delta m_I(t)$  is minimal.

## Methods

### Glauber dynamics in mean-field theory

We have left out so far several steps in the derivation of the results that were not necessary for the presentation of the main ideas. In this section, we will therefore give a self-contained derivation of our results also necessitating paraphrases of some results known from earlier works. The starting point is the master equation for the probability density of the possible network states emerging from the Glauber dynamics [34] described in “Binary network model and its mean field equations” (see for the following also [13, 37])

$$\frac{\partial p}{\partial t}(\mathbf{n}, t) = \underbrace{\frac{1}{\tau}}_{\text{update rate}} \sum_{i=1}^N \sum_{\substack{\in \{-1, 1\}, \text{direction of flux}}} \underbrace{(2n_i - 1)}_{\text{net flux due to neuron } i} \phi_i(\mathbf{n} \setminus n_i, t) \quad \forall \quad \mathbf{n} \in \{0, 1\}^N, \quad (14)$$

where

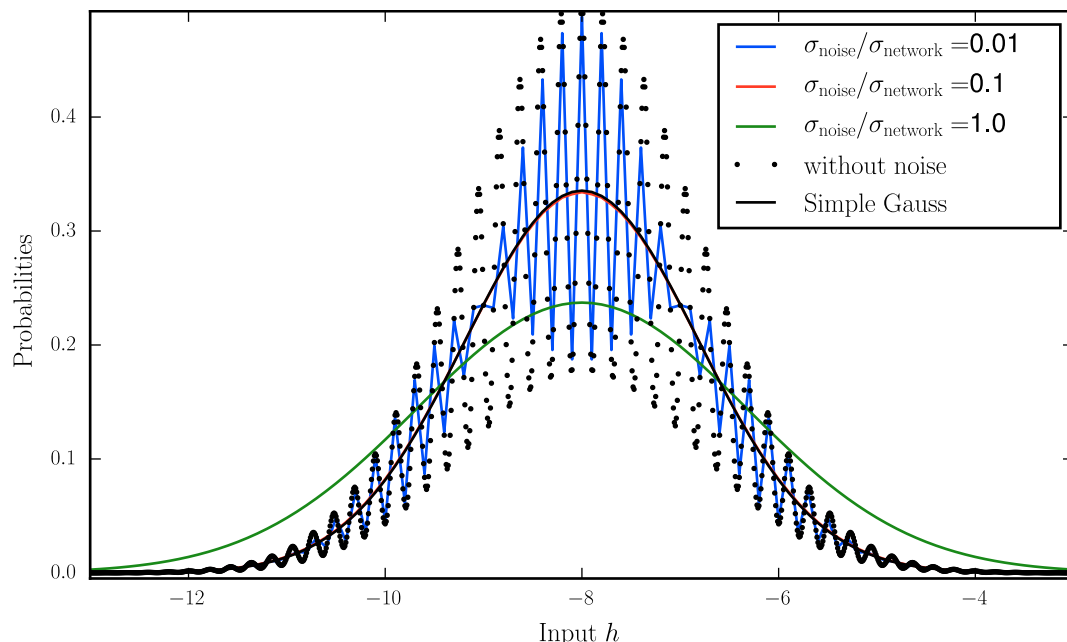
$$\begin{aligned} \phi_i(\mathbf{n} \setminus n_i, t) &= \underbrace{p(\mathbf{n}_{i-}, t) F_i(\mathbf{n}_{i-})}_{\text{neuron } i \text{ transition up}} - \underbrace{p(\mathbf{n}_{i+}, t) (1 - F_i(\mathbf{n}_{i+}))}_{\text{neuron } i \text{ transition down}} \\ &= -p(\mathbf{n}_{i+}) + p(\mathbf{n}_{i-}, t) F_i(\mathbf{n}_{i-}) + p(\mathbf{n}_{i+}, t) F_i(\mathbf{n}_{i+}). \end{aligned}$$

The activation function  $F_i(\mathbf{n})$  is given by Eq (1).

Using the master equation (for details cf. section II A in S1 Text), one can derive a differential equation for the mean activity of the neuron  $i$ ,  $\langle n_i \rangle(t) = \sum_{\mathbf{n}} p(\mathbf{n}, t) n_i$  and the raw covariance of the neurons  $i$  and  $j$ ,  $\langle n_i(t) n_j(t) \rangle = \sum_{\mathbf{n}} p(\mathbf{n}, t) n_i n_j$  [6, 13, 27, 34, 37]. This yields

$$\begin{aligned} \tau \frac{d}{dt} \langle n_k \rangle(t) &= -\langle n_k \rangle(t) + \langle F_k(t) \rangle \\ \frac{d}{dt} \langle n_k(t) n_l(t) \rangle &= \{-\langle n_k(t) n_l(t) \rangle + \langle n_l(t) F_k(t) \rangle\} + \{k \leftrightarrow l\}. \end{aligned} \quad (15)$$

As mentioned in “Binary network model and its mean field equations”, we assume that the input  $h_i$  coming from the local and the external population is normally distributed, say with



**Fig 8. Distribution of inputs from binary neurons for different noise levels.** Probability distribution of synaptic input  $h_i = \sum_j J_{ij} n_j + \xi_i$  of a neuron in a network of independently active cells  $n_j$  with  $\langle n_E \rangle = \langle n_I \rangle = 0.2$  and synaptic weights  $j_i = -0.21$ ,  $j_E = 0.01$ .  $|j_E|$  was deliberately chosen to be large because only then the convolution of a binomial distribution “squeezed” to the step size  $j_E$  with the binomial distribution squeezed to the step size  $|j_i|$  results in a probability distribution with many local maxima leading to the impression of an oscillation. The noiseless case  $\xi_i = 0$  is shown as black dots. The solid black curve indicates the Gaussian approximation (cf. e.g. Eq (16), here without perturbation) of this distribution from the main text. This distribution appears in the expectation values of the activation function  $F$  (cf. e.g. Eq (1)): It is a Gaussian distribution with the mean  $\mu = K_{Ej_E} m_E + K_{ij_i} m_I$  and the variance  $\sigma_{\text{network}}^2 = K_{Ej_E}^2 m_E (1 - m_E) + K_{ij_i}^2 m_I (1 - m_I)$  of the original binomial distributions  $\text{Binom}(m_E, K_E)$ ,  $\text{Binom}(m_I, K_I)$ . The other curves indicate convolutions with the Gaussian noise  $\xi \sim \mathcal{N}(0, \sigma_{\text{noise}})$  of different magnitudes  $\sigma_{\text{noise}}$ , given in units of the noise level  $\sigma_{\text{network}}$  intrinsically produced by the network.

<https://doi.org/10.1371/journal.pcbi.1005534.g008>

mean  $\mu_i$  and standard deviation  $\sigma_i$  given by

$$\begin{aligned} \mu_i(t) &:= \langle h_i \rangle = (J \langle \mathbf{n} \rangle)_i + h_{\text{ext}} \sin(\omega t) \\ \sigma_i^2(t) &:= \langle h_i^2 \rangle - \langle h_i \rangle^2 = \sum_{k,k'=1}^N J_{i,k} J_{i,k'} (\langle n_k n_{k'} \rangle - \langle n_k \rangle \langle n_{k'} \rangle) + (\sigma_i^{\text{noise}})^2 \\ &= (J^T c J)_{ii} + J \otimes J \langle \mathbf{n} \rangle \otimes (1 - \langle \mathbf{n} \rangle) + (\sigma_i^{\text{noise}})^2, \end{aligned} \quad (16)$$

where the average  $\langle \rangle$  is taken over realizations of the stochastic dynamics and we used the element-wise (Hadamard) product (see main text).

The additional noise introduced in Eq (1) effectively leads to a smoothing of the neurons’ activation threshold and broadens the width of the input distribution. It can be interpreted as additional variability coming from other brain areas. Furthermore, it is computationally convenient, because the theory assumes the input to be a (continuous) Gaussian distribution, while in the simulation, the input  $\sum_{i=k}^N J_{ik} n_k$ , being a sum of discrete binary variables, can only assume discrete values. The smoothing by the additive noise therefore improves the agreement of the continuous theory with the discrete simulation. Already weak external noise compared to the intrinsic noise is sufficient to obtain a quite smooth probability distribution of the input (Fig 8).

The description in terms of a coupled set of moment equations instead of the ODE for the full probability distribution here serves to reduce the dimensionality: It is sufficient to describe the time evolution of the moments on the population level, rather than on the level of individual units. To this end we need to assume that the synaptic weights  $J_{ij}$  only depend on the population  $\alpha, \beta \in \{\text{exc.}, \text{inh.}, \text{ext.}\}$  that  $i$  and  $j$  belong to, respectively, and thus (re)name them  $J_{\alpha\beta}$  (homogeneity). Furthermore, we assume that not all neurons are connected to each other, but that  $K_{\alpha\beta}$  is the number of incoming connections a neuron in population  $\alpha$  receives from a neuron in population  $\beta$  (fixed in-degree). The incoming connections to each neuron are chosen randomly, uniformly distributed over all possible sending neurons. This leads to expressions for the population averaged input  $h_\alpha$ , mean activity  $m_\alpha$  and covariance  $c_{\alpha\beta}$ , formally nearly identical to those on the single cell level and analogous to those in [13, sec. Mean-field solution].

**Mean activity: Stationary part and response to perturbation in linear order.** We are now able to calculate the quantity  $\langle F_\alpha(\mathbf{n}(t), t) \rangle = \langle H(h_\alpha(t) - \theta) \rangle$  (recall that  $h_\alpha(t)$  is a Gaussian random variable with mean  $\mu_\alpha(t)$  and standard deviation  $\sigma_\alpha(t)$ ), the nonlinearity of the ODEs 15 on the population level. Multiplying  $H(h_\alpha(t) - \theta_\alpha)$  by the Gaussian probability density for  $h_\alpha(t)$ , we get, after substitution of the integration variable,

$$\begin{aligned} \langle F_\alpha(\mathbf{n}(t), t) \rangle &= \langle H(h_\alpha(t) - \theta_\alpha) \rangle \\ &= \frac{1}{\sqrt{\pi}} \int_{\frac{\theta - \mu_\alpha(t)}{\sqrt{2}\sigma_\alpha(t)}}^{\infty} e^{-x^2} dx = \frac{1}{2} \operatorname{erfc} \left( \frac{\theta_\alpha - \mu_\alpha(t)}{\sqrt{2}\sigma_\alpha(t)} \right) \\ &=: \varphi(\mu_\alpha(\mathbf{m}(t), h_{\text{ext}} \sin(\omega t)), \sigma_\alpha(\mathbf{m}(t), c(t))), \end{aligned} \quad (17)$$

where we defined the average input  $\mu_\alpha$  and the width of the input distribution  $\sigma_\alpha$

$$\mu_\alpha(t) := [(K \circledast J) \mathbf{m}(t)]_\alpha + h_{\text{ext}} \sin(\omega t) \quad (18)$$

$$\begin{aligned} \sigma_\alpha^2(t) &:= [(K \circledast J)^T c(t) (K \circledast J)]_{\alpha\alpha} \\ &\quad + [K \circledast J \circledast J \mathbf{m}(t) \circledast (1 - \mathbf{m}(t))]_\alpha + \sigma_{\alpha, \text{noise}}^2. \end{aligned} \quad (19)$$

Recall that we defined  $\bar{x}$  to be the quantity  $x$  in the stationary case (without external input). For the linear approximation around  $\mu_\alpha = \bar{\mu}_\alpha$ ,  $\sigma_\alpha = \bar{\sigma}_\alpha$  and  $h_{\text{ext}} = 0$ , we have to take into account all dependencies via inner derivatives. We set  $\delta\mu_\alpha = \mu_\alpha - \bar{\mu}_\alpha$  and  $\delta\sigma_\alpha = \sigma_\alpha - \bar{\sigma}_\alpha$ . Note that  $\delta\mu$  includes the variation of  $\mu$  both because of fluctuations in the network and because of the external drive. The Taylor expansion up to linear order is

$$\varphi(\mu_\alpha(\mathbf{m}, c, h_{\text{ext}}), \sigma_\alpha(\mathbf{m}, c)) \approx \varphi(\bar{\mu}_\alpha, \bar{\sigma}_\alpha) + \underbrace{S(\bar{\mu}_\alpha, \bar{\sigma}_\alpha)}_{=S_\alpha} \left[ \delta\mu_\alpha + \frac{\theta - \bar{\mu}_\alpha}{\bar{\sigma}_\alpha} \delta\sigma_\alpha \right],$$

where we introduced the susceptibility on the population level

$$S(\mu_\alpha(t), \sigma_\alpha(t)) := \frac{d}{d\mu_\alpha(t)} \varphi(\mu_\alpha(t), \sigma_\alpha(t)) = \frac{1}{\sqrt{2\pi}\sigma_\alpha(t)} e^{-\frac{(\mu_\alpha(t) - \theta_\alpha)^2}{2\sigma_\alpha^2(t)}}. \quad (20)$$

Now, we express  $\delta\sigma_\alpha$  and  $\delta\mu_\alpha$  via  $\delta\mathbf{m} := \mathbf{m} - \bar{\mathbf{m}}$  and  $\delta c := c - \bar{c}$  (cf. [13, eq. (29)] for the time-

independent case):

$$\delta\mu_\alpha(t) = \sum_{\beta} K_{\alpha\beta} J_{\alpha\beta} \delta m_{\beta}(t) + h_{\text{ext}} \sin(\omega t) \quad (21)$$

$$\delta\sigma_\alpha(t) = \frac{1}{2\sigma_\alpha} \left( \sum_{\beta} K_{\alpha\beta} J_{\alpha\beta}^2 (1 - 2m_{\beta}) \delta m_{\beta}(t) + \sum_{\beta, \gamma} K_{\alpha\beta} K_{\alpha\gamma} J_{\alpha\beta} J_{\alpha\gamma} \delta c_{\beta\gamma}(t) \right). \quad (22)$$

Note that in  $\delta\mu_\alpha$  (but not  $\delta\sigma_\alpha$ ), the perturbation occurs again explicitly. In Eq (33), we demonstrate that  $\delta c$  scales like  $\frac{\delta m}{N}$ , like in the stationary case. Furthermore, we certainly have  $\frac{|K|}{N} = \mathcal{O}(1)$  and  $\sigma = \mathcal{O}(\sqrt{|K|})$ , thus

$$|\delta\mu(t)| = \mathcal{O}(|K||\delta m(t)|) = \mathcal{O}(h_{\text{ext}}), \text{ but } |\delta\sigma(t)| = \mathcal{O}(\sqrt{|K|}|\delta m(t)|) = \mathcal{O}\left(\frac{h_{\text{ext}}}{\sqrt{|K|}}\right).$$

We therefore neglect  $\delta\sigma$  in our calculations for  $\delta m$  from Eq (24) on. This yields for the linearization of the ODE Eq (5)

$$\tau \frac{\partial}{\partial t} \delta m_\alpha(t) + \delta m_\alpha(t) = \bar{S}_\alpha [\delta\mu_\alpha(t) + \underbrace{\frac{\theta_\alpha - \bar{\mu}_\alpha}{\bar{\sigma}_\alpha}}_{=\text{erfc}^{-1}(\bar{m})} \delta\sigma_\alpha(t)] \quad (23)$$

$$\tau \frac{\partial}{\partial t} \delta m_\alpha(t) + \delta m_\alpha(t) = \sum_{\beta} W_{\alpha\beta} \delta m_{\beta}(t) + \bar{S}_\alpha h_{\text{ext}} \sin(\omega t) + \mathcal{O}\left(h_{\text{ext}}^2, \frac{1}{\sqrt{K}}\right), \quad (24)$$

where we used the relation  $\frac{\theta_\alpha - \bar{\mu}_\alpha}{\bar{\sigma}_\alpha} = \sqrt{2} \text{erfc}^{-1}(2\bar{m}_\alpha)$ , derived from Eq (7) in connection with Eq (17), which implies that this expression does not depend on  $K$ , but solely on  $\bar{m}_\alpha$  and we defined

$$W_{\alpha\beta} := \bar{S}_\alpha K_{\alpha\beta} J_{\alpha\beta}.$$

The only change compared to the setup in [13] is again the occurrence of a periodic term, here  $\bar{S}_\alpha h_{\text{ext}} \sin(\omega t)$ .

We solve Eq (24) by transforming it into the eigenbasis of the matrix  $W_{\alpha\beta}$

$$U^{-1} W U = \text{diag}(\lambda_1, \dots, \lambda_N) := \Lambda. \quad (25)$$

We multiply Eq (24) by  $U^{-1}$ , define  $\delta m^\alpha := (U^{-1})^{\alpha\beta} \delta m_{\beta}$  and get

$$\tau \frac{d}{dt} \delta m^\alpha = -\delta m^\alpha + \Lambda_\alpha^\alpha \delta m^\alpha + (U^{-1})^{\alpha\beta} \bar{S}_\beta h_{\text{ext}} \sin(\omega t). \quad (26)$$

Note that the input is projected onto the respective eigenmodes. Eq (26) can be solved including the transient phase by the method of variation of constants.

But as we are only interested in the cyclostationary part of the solution, we can neglect the solution of the homogeneous part and solely compute the particular solution. Observe that  $\frac{d}{dt} \text{Im}(\delta m^\alpha(t)) = \text{Im}\left(\frac{d}{dt} \delta m^\alpha(t)\right)$  for a differentiable function  $\delta m^\alpha$  because  $t \in \mathbb{R}$ . We insert the ansatz  $\delta m^\alpha = M_1^\alpha e^{i\omega t}$  and solve for  $M_1^\alpha$ , which gives Eq (9) of the main text. For further calculations, keep in mind that  $M_1^\alpha$  and therefore  $\delta m_\alpha$  are of order  $\mathcal{O}\left(h_{\text{ext}} \frac{s}{SKJ}\right) = \mathcal{O}\left(h_{\text{ext}} \frac{1}{KJ}\right)$ . In section II C in S1 Text, we describe how to extract the right phase of the real solution from the complex ansatz.

**Covariances: Stationary part and response to a perturbation in linear order.** Using Eq (15) in the population-averaged version, we calculate the derivative of the zero time-lag covariance

$$c_{\alpha\beta}(t) := \frac{1}{N_\alpha N_\beta} \sum_{i \in \alpha, j \in \beta, i \neq j} \langle n_i(t) n_j(t) \rangle - \langle n_i(t) \rangle \langle n_j(t) \rangle$$

getting

$$\tau \frac{dc_{\alpha\beta}(t)}{dt} = -2c_{\alpha\beta}(t) + \frac{1}{N_\alpha N_\beta} \sum_{i \in \alpha, j \in \beta, i \neq j} \langle F_j(\mathbf{n}(t)) \delta n_i(t) \rangle + \langle F_i(\mathbf{n}(t)) \delta n_j(t) \rangle.$$

Neglecting cumulants of order higher than two, we can expand the expectation value  $\langle F_i(\mathbf{n}(t)) \delta n_j(t) \rangle$  (cf. [13, 33, section “Linearized equation for correlations and susceptibility”]) and get

$$\langle F_i(\mathbf{n}(t)) \delta n_j(t) \rangle \approx S(\mu_i(t), \sigma_i(t)) \sum_{k \neq j} J_{ik} c_{kj}(t) + S(\mu_i(t), \sigma_i(t)) J_{ij} a_j(t). \quad (27)$$

After carrying out the population averaging, we get the ordinary differential equation

$$\begin{aligned} & \tau \frac{dc_{\alpha\beta}(t)}{dt} \\ &= \left\{ -c_{\alpha\beta}(t) + \sum_{\gamma} S(\mu_\alpha(t), \sigma_\alpha(t)) K_{\alpha\gamma} J_{\alpha\gamma} \left( c_{\gamma\beta}(t) + \delta_{\gamma\beta} \frac{a_\beta(t)}{N_\beta} \right) \right\} + \{\alpha \leftrightarrow \beta\}. \end{aligned} \quad (28)$$

Therefore, the stationary part  $\bar{c}$  of the covariances fulfills the relation (cf. [13, 33])

$$2\bar{c}_{\alpha\beta} = \sum_{\gamma} S(\bar{\mu}_\alpha, \bar{\sigma}_\alpha) (K \circ J)_{\alpha\gamma} \left( \bar{c}_{\gamma\beta} + \delta_{\gamma\beta} \frac{\bar{a}_\beta}{N_\beta} \right) + \alpha \leftrightarrow \beta. \quad (29)$$

As for the mean activities, we want to make a little step (of order  $h_{\text{ext}}$ , to be precise) away from the stationary state determining the deviation  $\delta c(t) := c(t) - \bar{c}$ . For that, we have to calculate the Taylor expansion of  $S(\mu_\alpha(t), \sigma_\alpha(t))$  in  $\delta \mathbf{m}$ , i.e.

$$\begin{aligned} & S(\mu_\alpha(t), \sigma_\alpha(t)) \\ &:= \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_\alpha(t)} \exp\left(-\frac{(\mu_\alpha(t) - \theta_\alpha)^2}{2(\sigma_\alpha(t))^2}\right) \\ &\approx S(\bar{\mu}_\alpha, \bar{\sigma}_\alpha) + \left( \frac{\partial S}{\partial \mu_\alpha(t)} \delta \mu_\alpha + \frac{\partial S}{\partial \sigma_\alpha(t)} \delta \sigma_\alpha \right) \Big|_{\mu_\alpha = \bar{\mu}_\alpha, \sigma_\alpha = \bar{\sigma}_\alpha}, \end{aligned}$$

where  $\delta\mu_\alpha$  and  $\delta\sigma_\alpha$  are given by Eq (21) and

$$\begin{aligned}\frac{\partial S}{\partial \mu_\alpha(t)}(\bar{\mu}_\alpha, \bar{\sigma}_\alpha) &= \frac{\theta_\alpha - \bar{\mu}_\alpha}{\bar{\sigma}_\alpha^2} S(\bar{\mu}_\alpha, \bar{\sigma}_\alpha) \\ \frac{\partial S}{\partial \sigma_\alpha(t)}(\bar{\mu}_\alpha, \bar{\sigma}_\alpha) &= -\frac{1}{\bar{\sigma}_\alpha} \left( 1 - \left( \frac{\theta_\alpha - \bar{\mu}_\alpha}{\bar{\sigma}_\alpha} \right)^2 \right) S(\bar{\mu}_\alpha, \bar{\sigma}_\alpha) \\ &= \frac{\theta_\alpha - \bar{\mu}_\alpha}{\bar{\sigma}_\alpha^2} \left( \underbrace{\frac{\theta_\alpha - \bar{\mu}_\alpha}{\bar{\sigma}_\alpha}}_{=\mathcal{O}(1)} - \underbrace{\frac{\bar{\sigma}_\alpha}{\theta_\alpha - \bar{\mu}_\alpha}}_{=\mathcal{O}(1)} \right) S(\bar{\mu}_\alpha, \bar{\sigma}_\alpha)\end{aligned}$$

Here again, the relation  $\frac{\theta_\alpha - \bar{\mu}_\alpha}{\bar{\sigma}_\alpha} = \sqrt{2} \operatorname{erfc}^{-1}(2\bar{m}_\alpha)$  was used to estimate the dependence on  $K$ . We insert the linearization of  $S$  and the expressions for  $\delta\mu$  and  $\delta\sigma$ , Eq (21), into the ODE for  $c_{\alpha\beta}(t) = \bar{c}_{\alpha\beta} + \delta c_{\alpha\beta}(t)$  to get, after neglecting the contributions of order  $\mathcal{O}(h_{\text{ext}}^2)$  and sorting the rest into terms proportional to  $\delta c$ ,  $h_{\text{ext}}$  and  $\delta m$  respectively:

$$\begin{aligned}& \tau \frac{d}{dt} \delta c_{\alpha\beta}(t) + \left\{ \sum_{\gamma} (\delta_{\alpha\gamma} - S(\bar{\mu}_\alpha, \bar{\sigma}_\alpha) K_{\alpha\gamma} J_{\alpha\gamma}) \delta c_{\gamma\beta}(t) \right\} + \{\alpha \leftrightarrow \beta\} \\ &= \left\{ \frac{\partial S}{\partial \mu_\alpha(t)} \sum_{\gamma} K_{\alpha\gamma} J_{\alpha\gamma} \left( \bar{c}_{\gamma\beta} + \frac{\bar{a}_\beta}{N_\beta} \delta_{\gamma\beta} \right) \left( h_{\text{ext}} \sin(\omega t) + \underbrace{\sum_{\delta} K_{\alpha\delta} J_{\alpha\delta} \delta m_\delta(t)}_{=\mathcal{O}(h_{\text{ext}})} \right) \right. \\ &+ \frac{\partial S}{\partial \sigma_\alpha(t)} \sum_{\gamma} K_{\alpha\gamma} J_{\alpha\gamma} \left( \bar{c}_{\gamma\beta} + \frac{\bar{a}_\beta}{N_\beta} \delta_{\gamma\beta} \right) \underbrace{\delta \sigma_\alpha(t)}_{=\mathcal{O}\left(\frac{h_{\text{ext}}}{\sqrt{K}}\right)} \\ &\left. + S(\bar{\mu}_\alpha, \bar{\sigma}_\alpha) K_{\alpha\beta} J_{\alpha\beta} \frac{(1 - 2\bar{m}_\beta)}{N_\beta} \delta m_\beta(t) \right\} + \{\alpha \leftrightarrow \beta\} \quad (30)\end{aligned}$$

Before finally solving for  $\delta c(t)$ , we want to justify the assumption  $\delta c = \mathcal{O}\left(\frac{\delta m}{N}\right)$ , which we needed already in the beginning to determine  $\delta m(t)$ , by a short calculation. We insert Eq (22) into Eq (30) and switch to matrix notation for brevity, which yields

$$\begin{aligned}& \tau \frac{d}{dt} \delta c(t) + \{(\mathbb{1} - \bar{S}KJ) \delta c(t)\} + \{\dots\}^T \\ &= \left\{ \frac{\partial S}{\partial \mu} KJ \left( \bar{c} + \frac{\bar{a}}{N} \right) (h_{\text{ext}} \sin(\omega t) + KJ \delta m(t)) + \underbrace{\frac{\partial S}{\partial \sigma} KJ \left( \bar{c} + \frac{\bar{a}}{N} \right)}_{=\frac{\partial S}{\partial \mu} \left( \frac{\theta - \mu}{\sigma - \mu} \right)} \right. \\ &\quad \left. \times \left( \frac{KJ^2(1 - 2\bar{m})}{2\bar{\sigma}} \delta m(t) + \frac{KJ}{2\bar{\sigma}} \delta c(t) (KJ)^T \right) + \bar{S}KJ \frac{1 - 2\bar{m}}{N} \delta m(t) \right\} + \{\dots\}^T. \quad (31)\end{aligned}$$

We can rewrite the left hand side in order to recognize the parts, which are identical to the

right hand side of Eq (23), i.e. the ODE for  $\delta m(t)$  without the neglect of  $\delta\sigma$ , which gives

$$\left\{ \frac{\partial S}{\partial \mu} KJ \left( \bar{c} + \frac{\bar{a}}{N} \right) \underbrace{\left( (h_{\text{ext}} \sin(\omega t) + KJ \delta m(t)) + \frac{\Theta - \mu}{\sigma} \left( \frac{KJ^2(1 - 2\bar{m})}{2\bar{\sigma}} \delta m(t) + \frac{KJ}{2\sigma} \delta c(t) (KJ)^T \right) \right)}_{=\left(\tau \frac{\partial}{\partial t} \delta m(t) + \delta m(t)\right)/S} \right\} \\ - \underbrace{\frac{\partial S}{\partial \mu}}_{=\frac{1}{\sigma} \frac{\Theta - \mu}{\sigma}} \frac{\sigma}{\Theta - \mu} KJ \left( \bar{c} + \frac{\bar{a}}{N} \right) \left( \frac{KJ^2(1 - 2\bar{m})}{2\bar{\sigma}} \delta m(t) + \frac{KJ}{2\sigma} \delta c(t) (KJ)^T \right) + \bar{S} KJ \frac{1 - 2\bar{m}}{N} \delta m(t) \Big\} + \{\dots\}^T. \quad (32)$$

Bringing the  $\delta c$ -terms on the left hand side finally yields

$$\tau \frac{d}{dt} \delta c(t) + \left\{ (1 - \bar{S} KJ) \delta c(t) + \underbrace{\frac{1}{\bar{\sigma}} \bar{S} KJ \left( \bar{c} + \frac{\bar{a}}{N} \right) \frac{KJ}{2\sigma} \delta c(t) (KJ)^T}_{=\mathcal{O}\left(\bar{S} KJ \frac{K}{N} \frac{KJ^2}{\sigma^2} \delta c(t)\right) = \mathcal{O}\left(\bar{S} KJ \frac{K}{N} \delta c(t)\right)} \right\} + \{\dots\}^T \\ = \left\{ \underbrace{\frac{\partial S}{\partial \mu} KJ \left( \bar{c} + \frac{\bar{a}}{N} \right) \left( \tau \frac{\partial}{\partial t} \delta m_x(t) + \delta m_x(t) \right)}_{=\frac{1}{\sigma} \frac{\Theta - \mu}{\sigma}}}_{=\mathcal{O}\left(\bar{S} KJ \frac{K}{N} \delta m(t)\right)} \right\} \\ - \underbrace{\frac{1}{\bar{\sigma}} \bar{S} KJ \left( \bar{c} + \frac{\bar{a}}{N} \right) \frac{KJ^2(1 - 2\bar{m})}{2\bar{\sigma}} \delta m(t) + \bar{S} KJ \frac{1 - 2\bar{m}}{N} \delta m(t)}_{=\mathcal{O}\left(\bar{S} KJ \frac{K}{N} \delta m(t)\right)} \Big\} + \{\dots\}^T. \quad (33)$$

We have therefore shown that—independent of the scaling of the synaptic weights  $J$ —the relation  $\delta c = \mathcal{O}\left(\frac{\delta m}{N}\right)$  holds not only for the zero-mode, i.e. for the stationary case, but also for the time-dependent part. Note that for our actual calculation of  $\delta m$ , we have neglected its dependence on  $\delta\sigma$ , as it is one order  $\sqrt{K}$  smaller than the  $\delta\mu$ -contribution. However, this is not true for  $\delta c$  because of the cancellation of the two contributions to  $\delta\mu$ . Inserting the rhs of the ODE Eq (24) actually used to determine  $\delta m$  and shifting the  $\delta c$ -contribution of  $\delta\sigma$  back to the other



side, we arrive at

$$\begin{aligned} & \tau \frac{d}{dt} \delta c(t) + \{(\mathbb{1} - \bar{S}KJ)\delta c(t)\} + \{\dots\}^T \\ &= \left\{ \frac{\partial S}{\partial \mu} KJ \left( \bar{c} + \frac{\bar{a}}{N} \right) \left( \tau \frac{\partial}{\partial t} \delta m_x(t) + \delta m_x(t) \right) \right. \\ & \quad - \frac{1}{\bar{\sigma}} \left( 1 - \left( \frac{\mu - \Theta}{\sigma} \right)^2 \right) \bar{S}KJ \left( \bar{c} + \frac{\bar{a}}{N} \right) \left( \frac{KJ^2(1 - 2\bar{m})}{2\bar{\sigma}} \delta m(t) + \frac{KJ}{2\bar{\sigma}} \delta c(t)(KJ)^T \right) \\ & \quad \left. + \bar{S}KJ \frac{1 - 2\bar{m}}{N} \delta m(t) \right\} + \{\dots\}^T. \end{aligned} \quad (34)$$

We want to compare the contribution from  $\delta\mu$  in the second line of Eq (34) with the contribution from  $\delta\sigma$  in the third line. As pointed out above, they scale in the same way with the system size  $N$ , given that we do not rescale the driving frequency with  $N$ . Therefore, its contribution stays equally important if we enlarge the network. We neglect it anyway, which can be justified by comparing the decisive part of the prefactors of the  $\delta\sigma$  and the  $\delta\mu$ -parts (the remaining parts are of the same order of magnitude):

$$\frac{\partial S}{\partial \sigma} \frac{\partial \mu}{\partial S} = \frac{\Theta - \bar{\mu}}{\bar{\sigma}} = \sqrt{2} \operatorname{erfc}^{-1}(2\bar{m}) \stackrel{\text{for input fluct. not too small}}{\gg} \frac{1}{\bar{\sigma}} \exp(-\operatorname{erfc}^{-1}(2\bar{m})^2) = \bar{S}.$$

This inequality is fulfilled for the three settings used in this work, whereas the first term is one or two orders of magnitude larger than the second. Especially, this inequality can always be fulfilled if the externally generated noise level is high. Therefore, even if the neglect of the  $\delta\sigma$ -contribution to  $\delta c$  cannot be justified by the standard mean-field argument that it decays faster with the system size than other terms, it is applicable because the input fluctuations are large enough—for all system sizes. This largely simplifies the calculations because the ODE for  $\delta c$  can be solved by transforming into the eigensystem of  $W$ , which would not be possible after including the more involved term emerging from  $\delta\sigma$ . Taking into account the neglected term would require to reformulate the problem as an equation for the vector  $(\delta c_{EE}, \delta c_{EI}, \dots)$ , which would be much less intuitive. Furthermore, there is an indirect argument for high frequencies that does rely on the system size: The  $\omega$ -dependence of the absolute value of the maxima of  $\delta m$  and  $\delta c$  scales with the eigenvalues of  $W$ , which scale with  $\sqrt{K}$ . Thus, changing the system size  $N$  in first order just stretches the  $\omega$ -axis. Therefore, the “interesting” frequencies do scale with  $N$ , which leads to the dominance of the derivative term in the second line of Eq (34) over the  $\delta\sigma$ -term. Note that the observation from Eq (32) that

$$\underbrace{\left( KJ \left( 1 + \mathcal{O}\left( \frac{1}{\sqrt{K}} \right) \right) \delta m \right)^{\text{diag}}}_{=\mathcal{O}(h_{\text{ext}})} + \underbrace{h_{\text{ext}} \sin(\omega t)}_{=\mathcal{O}(h_{\text{ext}})} \quad (35)$$

$$= (\bar{S}^{\text{diag}})^{-1} \left( \tau \frac{\partial}{\partial t} \delta m + \delta m \right) + \mathcal{O}(\sqrt{|K|} |\delta m|) = \mathcal{O}(\sqrt{|K|} |\delta m|) = \mathcal{O}\left( \frac{h_{\text{ext}}}{\sqrt{|K|}} \right) \quad (36)$$

is a direct consequence of the recurrent drive being effectively inhibitory (for other networks, the expansion around the stationary point would not make sense): Any of the two terms in the

susceptibility terms are of order  $\sqrt{K}$  bigger than their sum. Furthermore, we see from Eq (33) that the sum of the susceptibility terms is of the same order of magnitude with respect to its dependence on  $K$  (or, equivalently, the connection probabilities and the system size) as the term coming from the time modulation of the variances (modulated-autocovariances-drive).

We define

$$\begin{aligned} T_{\alpha\beta} &:= K_{\alpha\beta} J_{\alpha\beta} \\ V_{\alpha\beta} &:= \frac{\Theta - \bar{\mu}_\alpha}{(\bar{\sigma}_\alpha)^2} S(\bar{\mu}_\alpha, \bar{\sigma}_\alpha) K_{\alpha\beta} J_{\alpha\beta}, \end{aligned} \quad (37)$$

and

$$\begin{aligned} N_{\alpha\beta}^{\text{diag}} &= \delta_{\alpha\beta} N_\alpha \\ \bar{m}_{\alpha\beta}^{\text{diag}} &= \delta_{\alpha\beta} \bar{m}_\alpha \\ \bar{a}_{\alpha\beta}^{\text{diag}} &= \delta_{\alpha\beta} \bar{a}_\alpha \\ \delta m_{\alpha\beta}^{\text{diag}}(t) &= \delta_{\alpha\beta} \delta m_\alpha(t) \\ (T\delta m(t))_{\alpha\beta}^{\text{diag}} &= \delta_{\alpha\beta} \sum_\gamma T_{\alpha\gamma} \delta m_\gamma(t), \end{aligned} \quad (38)$$

in order to end up with the index-free version Eq (10). The first two inhomogeneities, the susceptibility terms introduced in the main part (“Results”) reflect the nonlinearity of the gain-function.

With  $U$  given in Eq (25), we multiply Eq (30) from the left by  $U^{-1}$  and from the right by  $(U^{-1})^T$  to get (cf. [13, 33])

$$\begin{aligned} & \tau \frac{d}{dt} \underbrace{U^{-1} \delta c(t) (U^{-1})^T}_{:= \widetilde{\delta c}} \\ &= \{(-\mathbb{1} + \underbrace{U^{-1} W U}_{=: \Lambda}) \underbrace{U^{-1} \delta c(t) (U^{-1})^T}_{:= \widetilde{\delta c(t)}}\} \\ &+ U^{-1} \left( (T\delta m(t))^{\text{diag}} + h_{\text{ext}} \sin(\omega t) \right) V \left( \bar{c} + \frac{1}{N^{\text{diag}}} \bar{a}^{\text{diag}} \right) (U^{-1})^T \\ &+ U^{-1} W \left( \mathbb{1} - 2\bar{m}^{\text{diag}} \right) \frac{1}{N^{\text{diag}}} \delta m(t)^{\text{diag}} (U^{-1})^T \\ &+ \{\dots\}^T. \end{aligned}$$

We are only interested in the cyclostationary statistics, so we can ignore again the transient state making the ansatz  $\widetilde{\delta c_{\alpha\beta}^{\text{inhom}}} = \widetilde{C_{\alpha\beta}^{\text{I}}} e^{i\omega t}$ . Inserting this ansatz and transforming back into the

original system, we get

$$\begin{aligned} \widetilde{C}_{\alpha\beta}^1 &= h_{\text{ext}} \frac{-i\tau\omega + 2 - (\lambda_\alpha + \lambda_\beta)}{(\tau\omega)^2 + (2 - (\lambda_\alpha + \lambda_\beta))^2} \\ &\quad \left[ \sum_{\gamma, \delta, \theta, \phi, \eta} U_{\alpha\eta}^{-1} U_{\beta\delta}^{-1} T_{\eta\theta} U_{\theta\phi} (U^{-1}S)_\phi \frac{-i\tau\omega + 1 - \lambda_\phi}{(\tau\omega)^2 + (1 - \lambda_\phi)^2} V_{\eta, \gamma} \left( \bar{c} + \frac{1}{N^{\text{diag}}} \bar{a}^{\text{diag}} \right)_{\gamma\delta} \right. \\ &\quad \left. + \sum_{\gamma, \delta, \epsilon} U_{\alpha\epsilon}^{-1} U_{\beta\delta}^{-1} V_{\epsilon\gamma} \left( \bar{c} + \frac{1}{N^{\text{diag}}} \bar{a}^{\text{diag}} \right)_{\gamma\delta} \right. \\ &\quad \left. + \sum_{\theta, \phi, \gamma} U_{\alpha\gamma}^{-1} U_{\beta\theta}^{-1} W_{\gamma\theta} (1 - 2\bar{m}_\theta^{\text{diag}}) \frac{1}{N_\theta} U_{\theta\phi} (U^{-1}S)_\phi \frac{-i\tau\omega + 1 - \lambda_\phi}{(\tau\omega)^2 + (1 - \lambda_\phi)^2} \right] \end{aligned} \quad (39)$$

Together with Eq (9), this is the main result of this section.

## Discussion

The present work offers an extension of the well-known binary neuronal network model beyond the stationary case [6, 13, 27, 28, 33]. We here describe the influence of a sinusoidally modulated input on the mean activities and the covariances to study the statistics of recurrently generated network activity in an oscillatory regime, ubiquitously observed in cortical activity [18].

Comparing with the results of the simulation of the binary network with NEST [35, 36] and the numerical solution of the full mean-field ODE, we are able to show that linear perturbation theory is sufficient to explain the most important effects occurring due to sinusoidal drive. This enables us to understand the mechanisms by the help of analytical expressions and furthermore we can predict the network response to any time-dependent perturbation with existing Fourier representation by decomposing the perturbing input into its Fourier components.

We find that the amplitude of the modulation of the mean activity is of the order

$h_{\text{ext}} / ((1 - \lambda_\alpha)^2 + (\tau\omega)^2)^{\frac{1}{2}}$ , where  $\lambda_\alpha$ ,  $\alpha \in \{E, I\}$  are the eigenvalues of the effective connectivity matrix  $W$ , i.e. the input is filtered by a first order low-pass filter and the amplitude of the modulation decays like  $\propto \omega^{-1}$  for large frequencies. This finding is in line with earlier work on the network susceptibility [27, esp. section V].

The qualitatively new result here is the identification of two distinct mechanisms by which the covariances  $\delta c$  are modulated in time. First, covariances are driven by the direct modulation of the susceptibility  $S$  due to the time-dependent external input and by the recurrent input from the local network. Second, time-modulated variances, analogous to their role in the stationary setting [13], drive the pairwise covariances.

Our setup is the minimal network model, in which these effects can be observed—minimal in the sense that we would lose these properties if we further simplified the model: The presence of a nonlinearity in the neuronal dynamics, here assumed to be a threshold-like activation function, is required for the modulation of covariances by the time-dependent change of the effective gain. In a linear rate model [10, 46] this effect would be absent, because mean activities and covariances then become independent.

The second mechanism relies on the binary nature of neuronal signal transmission: the variance  $a(t)$  of the binary neuronal signal is, at each point in time, completely determined by its mean  $m(t)$ . This very dependence provides the second mechanism by which the temporally

modulated mean activity causes time-dependent covariances, because all fluctuations and therefore all covariances are driven by the variance  $a(t)$ .

Rate models have successfully been used to explain the smallness of pairwise covariances [6] by negative feedback [10]. A crucial difference is that their state is continuous, rather than binary. As a consequence, the above-mentioned fluctuations present due to the discrete nature of the neuronal signal transmission need to be added artificially: The pairwise statistics of spiking or binary networks are equivalent to the statistics of rate models with additive white noise [46]. To obtain qualitative or even quantitative agreement of time-dependent covariances between spiking or binary networks and rate models, the variance of this additive noise needs to be chosen such that its variance is a function of the mean activity and its time derivative.

The direct modulation of the susceptibility  $S$  due to the time-dependent external input leads to a contribution to the covariances with first order low-pass filter characteristics that dominates the modulated covariances at large frequencies. For small—and probably biologically realistic—frequencies (typically the LFP shows oscillations in the  $\beta$ -range around 20 Hz), however, the modulation of the susceptibility by the local input from the network leads to an equally important additional modulation of the susceptibility. The intrinsic fluctuations of the network activity are moreover driven by the time-dependent modulation of the variance, which is a function of the mean activity as well. Because the mean activity follows the external drive in a low-pass filtered manner, the latter two contributions hence exhibit a second order low-pass-filter characteristics. These contributions are therefore important at the small frequencies we are interested in here.

The two terms modulating the susceptibility, by the direct input and by the feedback of the mean activity through the network, have opposite signs in balanced networks. In addition they have different frequency dependencies. In networks in which the linearized connectivity has only real eigenvalues, these two properties together lead to their summed absolute value having a maximum. Whether or not the total modulation of the covariance shows resonant behavior, however, depends also on the third term that stems from the modulated variances. We find that in purely inhibitory networks, the resonance peak is typically overshadowed by the latter term. This is because inhibitory feedback leads to negative average covariances [13], which we show here reduce the driving force for the two resonant contributions. In balanced E-I networks, the driving force is not reduced, so the resonant contribution can become dominant.

For the biologically motivated parameters used in the last setting studied here, the effective coupling matrix  $W$  has complex eigenvalues which cause resonant mean activities. If the inhomogeneity was independent of the driving frequency,  $\delta c$  would have resonant modes with frequency  $f_{\text{res}}$  and  $2f_{\text{res}}$ . Due to the mixing of the different modes and by the frequency dependence of the inhomogeneity driving the modulation of covariances, these modes determine only the ballpark for the location of the resonance in the covariance. Especially the resonances are not sharp enough so that each of them is visible in any combination of the modes. Different behavior is expected near the critical point where  $\Re(\lambda) \lesssim 1$ .

For predictions of experimental results, however, a more careful choice of reasonable biological parameters would be necessary. In particular, the external drive should be gauged such that the modulations of the mean activities are in the experimentally observed range. Still, our setup shows that the theory presented here works in the biologically plausible parameter range.

The goal of extracting fundamental mechanisms of time-dependent covariances guides the here presented choice of the level of detail of our model. Earlier works [6, 28, 29] showed that our setup without sinusoidal drive is sufficient to qualitatively reproduce and explain phenomena observed in vivo, like high variability of neuronal activity and small covariances. The latter point can be explained in binary networks by the suppression of fluctuations by inhibitory

feedback, which is a general mechanism also applicable to other neuron models [10] and even finds application outside neuroscience, for example in electrical engineering [47]. The high variability observed in binary networks can be explained by the network being in the balanced state, that robustly emerges in the presence of negative feedback [29, 30]. In this state, the mean excitatory and inhibitory synaptic inputs cancel so far that the summed input to a neuron fluctuates around its threshold. This explanation holds also for other types of model networks and also for biological neural networks [48]. We have seen here that the operation in the balanced state, at low frequencies, gives rise to a partial cancellation of the modulation of covariances.

Our assumption of a network of homogeneously connected binary neurons implements the general feature of neuronal networks that every neuron receives input from a macroscopic number of other neurons, letting the impact of a single synaptic afferent on the activation of a cell be small and the summed input be distributed close to Gaussian: For uncorrelated incoming activity, the ratio between the fluctuations caused by a single input and the fluctuations of the total input is  $N^{-\frac{1}{2}}$ , independent of how synapses scale with  $N$ . However, the input to a neuron is actually not independent, but weakly correlated, with covariances decaying at least as fast as  $N^{-1}$  [6, 29]. Therefore this additional contribution to the fluctuations also decays like  $N^{-\frac{1}{2}}$ . The Gaussian approximation of the synaptic input relies crucially on these properties. Dahmen et al. [39] investigated third order cumulants, the next order of non-Gaussian corrections to this approximation. They found that the approximation has a small error even down to small networks of about 500 neurons and 50 synaptic inputs per neuron. These estimates hold as long as all synaptic weights are of equal size. For distributed synaptic amplitudes, in particular those following a wide or heavy-tailed distributions (e.g. [49, 50], reviewed in [51]), we expect the simple mean-field approximation applied here to require corrections due to the strong effect of single synapses.

The generic feature of neuronal dynamics, the threshold-like nonlinearity that determines the activation of a neuron, is shared by the binary, the leaky integrate-and-fire and, approximately, also the Hodgkin-Huxley model neuron. An important approximation entering our theory is the linearity of the dynamic response with respect to the perturbation. We estimate the validity of our theory by comparison to direct simulations. To estimate the breakdown of this approximation we compare the linear response to the first non-linear correction. We observe that the second order harmonics in the considered range of parameters remains as small as about 10 percent of the first harmonics. The quadratic contribution to the transfer properties of the neurons stems from the curvature of the effective gain function  $\varphi$  (Eq (17)). The linear portion of this gain function, in turn, is controlled by the amplitude  $\sigma$  of the synaptic noise. One therefore expects a breakdown of the linear approximation as soon as the temporal modulation of the mean input is of the order of this amplitude. Fig D in S1 Text shows that with the parameters  $h_{\text{ext}} = 1$  and  $\sigma_{\text{exc,inh}} \approx 10$ , used in the plots Figs 5 and 6 and Fig. B and Fig. C in S1 Text, the linear approximation is good, whereas in Fig 7, we used  $h_{\text{ext}} = 6$ , for which the linear perturbation theory already begins to break down. The latter figure is mainly supposed to give an intuitive impression.

A generic property that is shared by nearly all neuron models is the characteristic duration  $\tau$  during which the activity of a sending cell affects the downstream neuron. For the binary neuron model, this time scale is identical to the mean interval  $\tau$  between updates, because, once active, a neuron will stay active until the next update. It most certainly deactivates at that point, because we here consider low activity states prevalent in cortex [1]. In the leaky integrate-and-fire model the exponentially decaying membrane voltage with time constant  $\tau$  is qualitatively similar: it sustains the effect that an input has on the output for this time scale. As

a consequence, neurons transmit their input in a low-pass filtered manner to their output. This feature persists for more realistic spiking models, as shown for the leaky integrate-and-fire model [52, 53], the exponential integrate-and-fire model [52, 53], and the quadratic integrate-and-fire model [54]. We therefore expect that the qualitative properties reported here will carry over to these models.

A possible application of the framework developed in this paper is a quantitative comparison of the neuronal activity in the model network to the analysis of data measured in cortex [26]. Detecting the occurrence of so called Unitary Events (UE, [55–57], see also Sec. I in [S1 Text](#)), the authors observed that the simultaneous activation of neurons above the level expected for independence is locked to certain phases of the LFP. They hypothesized that the reason for this observation is the activation of cell assemblies. The results presented here show that the correlated activation of pairs of neurons is modulated by a sinusoidal drive even in a completely unstructured random network. In consequence, the locking of pairwise events to the cycle of the LFP is more pronounced for correlated events than for single spikes. Future work needs to quantitatively compare experimental data to the results from the model presented here. The closed form expressions for the modulations of the mean activities and covariances enable such an approach and the effective study of the dependence on the model parameters. A quantitative comparison needs to convert mean activities and pairwise covariances for binary neurons into the probability to measure a unitary event, interpreting the binary neuron states as binned spike trains. Preliminary results indicate that already the homogeneous network presented in this work can show some features described in [26]. In Sec. I in [S1 Text](#), we apply the Unitary Event analysis to our setting. The presented methods will be helpful to analyze the modulation of synchrony in the presence of cell assemblies [58] in the model. This can be done by enhancing the connection probability among groups of excitatory neurons, similar as in [59] and will yield a more realistic model, which captures also nonlinear effects in the perturbation. Technically this extension amounts to the introduction of additional populations and the change of the connectivity matrix to reflect that these populations represent cell assemblies.

The relation of spiking activity to mesoscopic measures, such as the LFP, is still an open question. These population measures of neuronal activity naturally depend on the statistics of the microscopic activity they are composed of. Pairwise covariances, the focus of the current work, in particular tend to dominate the variance of any mesoscopic signal of summed activity: The contribution of covariances grows quadratically in the number of components, the contribution of variances only linearly [60, Box 2][10, eq. (1)][21, eq. (1),(2)]. Under the assumption that the LFP mainly reflects the input to a local recurrent network [21, 24], we have shown here that these two signals—spikes and LFPs—are intimately related; not only does the afferent oscillatory drive trivially modulate the propensity to produce spikes, their firing rate, but also the joint statistics of pairs of neurons by the three distinct pathways exposed in the present analysis. Forward modeling studies have shown that the spatial reach of the LFP critically depends on covariances, with elevated covariances leading to larger reach [21]. In this light our work shows that a local piece of neuronal tissue driven by a source of coherent oscillations will more effectively contribute to the local field potential itself: not only the spiking rate is modulated accordingly, but also the covariances are increased and decreased in a periodic manner, further amplifying the modulation of the generated local field potential and temporally modulating the spatial reach of the signal.

Functional consequences of the findings presented here deduce from the hypothesis that communication channels in cortex may effectively be multiplexed by the selective excitation of different areas with coherent oscillations [61, 62]. The presented analysis exposes that oscillatory drive to a local piece of cortex alone already effectively enhances coherent firing beyond

the level expected based on the assumption of independence. If synchronous activity is employed as a dimension to represent information, it is hence tightly entangled with time-dependent changes of the mean activity. A similar conclusion was drawn from the observation that covariance transmission in feed-forward networks is monotonously increasing with firing rate [4, 5]. Any information-carrying modulation of synchronous activity must hence go beyond the here investigated effects, which can be regarded the baseline given by the non-stationary activity in networks without function. Since the mechanisms we have exposed only depend on generic features of cortical tissue—networks of non-linear neurons, connectivity with strong convergence and divergence, and dynamic stabilization by inhibition—the time-dependent entanglement of mean activity and covariances qualitatively exists in any network with these properties. In this view, our analysis can help to distinguish the level of time-modulated covariances in neural tissues that are surprising, and are therefore candidates to be attributed to function, from those that need to be expected in networks due to their generic properties.

## Supporting information

**S1 Text. Appendix.** In this text, we show how to apply the UE-analysis to a model network of binary neurons choosing the parameters from Table 1. We also present the derivation of the ODEs for the first two moments, we discuss the different possibilities to define a spike in a binary network and show how to handle the complex phase jump induced by the usage of the sine-function as a perturbation. Furthermore, we include the plots of the II- and EI-component of the covariances for the parameters of Table 1 and a plot of the mean activities and the EE-covariance for varying perturbation strength  $h_{\text{ext}}$  for the same parameters. (PDF)

## Acknowledgments

The authors would like to thank PierGianLuca Porta Mana for his great support and Michael Denker, Sonja Grün and the whole INM-6 for fruitful discussions. All network simulations were carried out with NEST (<http://www.nest-simulator.org>).

## Author Contributions

**Conceptualization:** MH TK.

**Data curation:** TK MH.

**Formal analysis:** TK MH.

**Funding acquisition:** MH.

**Investigation:** TK MH.

**Methodology:** MH TK.

**Project administration:** MH.

**Resources:** TK MH.

**Software:** TK MH.

**Supervision:** MH.

**Validation:** TK MH.



**Visualization:** TK MH.

**Writing – original draft:** TK MH.

**Writing – review & editing:** TK MH.

## References

1. Softky WR, Koch C. The Highly Irregular Firing of Cortical Cells Is Inconsistent with Temporal Integration of Random EPSPs. *J Neurosci*. 1993; 13(1):334–350. PMID: [8423479](#)
2. London M, Roth A, Beeren L, Häusser M, Latham PE. Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. *Nature*. 2010 July; 466(1):123–128. <https://doi.org/10.1038/nature09086> PMID: [20596024](#)
3. Lindner B, Doiron B, Longtin A. Theory of oscillatory firing induced by spatially correlated noise and delayed inhibitory feedback. *Phys Rev E*. 2005; 72:061919. <https://doi.org/10.1103/PhysRevE.72.061919>
4. De la Rocha J, Doiron B, Shea-Brown E, Kresimir J, Reyes A. Correlation between neural spike trains increases with firing rate. *Nature*. 2007 August; 448(16):802–807. <https://doi.org/10.1038/nature06028> PMID: [17700699](#)
5. Shea-Brown E, Josic K, de la Rocha J, Doiron B. Correlation and synchrony transfer in integrate-and-fire neurons: basic properties and consequences for coding. *Phys Rev Lett*. 2008 March; 100:108102. <https://doi.org/10.1103/PhysRevLett.100.108102> PMID: [18352234](#)
6. Renart A, De La Rocha J, Bartho P, Hollender L, Parga N, Reyes A, et al. The asynchronous State in Cortical Circuits. *Science*. 2010 January; 327:587–590. <https://doi.org/10.1126/science.1179850> PMID: [20110507](#)
7. Pernice V, Staude B, Cardanobile S, Rotter S. How Structure Determines Correlations in Neuronal Networks. *PLoS Comput Biol*. 2011 May; 7(5):e1002059. <https://doi.org/10.1371/journal.pcbi.1002059> PMID: [21625580](#)
8. Trousdale J, Hu Y, Shea-Brown E, Josic K. Impact of network structure and cellular response on spike time correlations. *PLoS Comput Biol*. 2012; 8(3):e1002408. <https://doi.org/10.1371/journal.pcbi.1002408> PMID: [22457608](#)
9. Pernice V, Staude B, Cardanobile S, Rotter S. Recurrent interactions in spiking networks with arbitrary topology. *Phys Rev E*. 2012; 85(3):031916. <https://doi.org/10.1103/PhysRevE.85.031916>
10. Tetzlaff T, Helias M, Einevoll G, Diesmann M. Decorrelation of neural-network activity by inhibitory feedback. *PLoS Comput Biol*. 2012; 8(8):e1002596. <https://doi.org/10.1371/journal.pcbi.1002596> PMID: [23133368](#)
11. Hu Y, Trousdale J, Josic K, Shea-Brown E. Motif statistics and spike correlations in neuronal networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2013; 2013(03):P03012. <https://doi.org/10.1088/1742-5468/2013/03/P03012>
12. Helias M, Tetzlaff T, Diesmann M. Echoes in correlated neural systems. *New J Phys*. 2013; 15:023002. <https://doi.org/10.1088/1367-2630/15/2/023002>
13. Helias M, Tetzlaff T, Diesmann M. The correlation structure of local cortical networks intrinsically results from recurrent dynamics. *PLoS Comput Biol*. 2014; 10(1):e1003428. <https://doi.org/10.1371/journal.pcbi.1003428> PMID: [24453955](#)
14. Rosenbaum R, Smith MA, Kohn A, Rubin JE, Doiron B. The spatial structure of correlated neuronal variability. *Nat Neurosci*. 2017; 20:107–114. <https://doi.org/10.1038/nn.4433> PMID: [27798630](#)
15. Riehle A, Grün S, Diesmann M, Aertsen A. Spike Synchronization and Rate Modulation Differentially Involved in Motor Cortical Function. *Science*. 1997 December; 278(5345):1950–1953. <https://doi.org/10.1126/science.278.5345.1950> PMID: [9395398](#)
16. Kilavik BE, Roux S, Ponce-Alvarez A, Confais J, Gruen S, Riehle A. Long-term Modifications in Motor Cortical Dynamics induced by intensive practice. *J Neurosci*. 2009; 29:12653–12663. <https://doi.org/10.1523/JNEUROSCI.1554-09.2009> PMID: [19812340](#)
17. Cohen MR, Maunsell JHR. Attention improves performance primarily by reducing interneuronal correlations. *Nat Neurosci*. 2009; 12:1594–1600. <https://doi.org/10.1038/nn.2439> PMID: [19915566](#)
18. Buzsáki G, Draguhn A. Neuronal Oscillations in Cortical Networks. *Science*. 2004; 304:1926–1929. <https://doi.org/10.1126/science.1099745> PMID: [15218136](#)
19. Nunez PL, Ramesh S. Electric fields of the brain: the neurophysics of EEG. Oxford University Press; 2006.

20. Ray S, Crone NE, Niebur E, Franaszczuk PJ, Hsiao SS. Neural correlates of high-gamma oscillations (60–200 Hz) in macaque local field potentials and their potential implications in electrocorticography. *J Neurosci*. 2008 November; 28(45):11526–11536. <https://doi.org/10.1523/JNEUROSCI.2848-08.2008> PMID: [18987189](#)
21. Lindén H, Tetzlaff T, Potjans TC, Pettersen KH, Grün S, Diesmann M, et al. Modeling the spatial reach of the LFP. *Neuron*. 2011; 72(5):859–872. <https://doi.org/10.1016/j.neuron.2011.11.006> PMID: [22153380](#)
22. Mitzdorf U. Current Source-Density Method and Application in Cat Cerebral Cortex: Investigation of Evoked Potentials and EEG Phenomena. *Physiol Rev*. 1985; 65(1):37–100. PMID: [3880898](#)
23. Viswanathan A, Freeman RD. Neurometabolic coupling in cerebral cortex reflects synaptic more than spiking activity. *Nature Neuroscience*. 2007; 10:1308–1312. <https://doi.org/10.1038/nn1977> PMID: [17828254](#)
24. Mazzoni A, Lindén H, Cuntz H, Lansner A, Panzeri S, Einevoll G. Computing the Local Field Potential (LFP) from Integrate-and-Fire Network Models. *PLoS Comput Biol*. 2015; 11:1–38. <https://doi.org/10.1371/journal.pcbi.1004584>
25. Scherberger H, Jarvis MR, Andersen RA. Cortical Local Field Potential Encodes Movement Intentions in the Posterior Parietal Cortex. *Neuron*. 2005; 46(2):347–354. Available from: <http://www.sciencedirect.com/science/article/pii/S0896627305001996>. <https://doi.org/10.1016/j.neuron.2005.03.004> PMID: [15848811](#)
26. Denker M, Roux S, Lindén H, Diesmann M, Riehle A, Grün S. The Local Field Potential Reflects Surplus Spike Synchrony. *Cereb Cortex*. 2011 December; 21:2681–2695. <https://doi.org/10.1093/cercor/bhr040> PMID: [21508303](#)
27. Ginzburg I, Sompolinsky H. Theory of correlations in stochastic neural networks. *Phys Rev E*. 1994; 50(4):3171–3191. <https://doi.org/10.1103/PhysRevE.50.3171>
28. Van Vreeswijk C, Sompolinsky H. Chaotic balanced state in a model of cortical circuits. *Neural Comput*. 1998; 10(6):1321–1371. <https://doi.org/10.1162/089976698300017214> PMID: [9698348](#)
29. van Vreeswijk C, Sompolinsky H. Chaos in Neuronal Networks with Balanced Excitatory and Inhibitory Activity. *Science*. 1996 December; 6; 274:1724–1726. <https://doi.org/10.1126/science.274.5293.1724> PMID: [8939866](#)
30. Amit DJ, Brunel N. Dynamics of a recurrent network of spiking neurons before and following learning. *Network: Comput Neural Systems*. 1997; 8:373–404. [https://doi.org/10.1088/0954-898X\\_8\\_4\\_003](https://doi.org/10.1088/0954-898X_8_4_003)
31. Braitenberg V, Schüz A. *Anatomy of the Cortex: Statistics and Geometry*. Berlin, Heidelberg, New York: Springer-Verlag; 1991.
32. van Vreeswijk C, Sompolinsky H. Chaotic Balanced State in a Model of Cortical Circuits. *Neural Comput*. 1998; 10:1321–1371. <https://doi.org/10.1162/089976698300017214> PMID: [9698348](#)
33. Buice MA, Cowan JD, Chow CC. Systematic fluctuation expansion for neural network activity equations. *Neural Comput*. 2010 Feb; 22(2):377–426. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2805768/>. <https://doi.org/10.1162/neco.2009.02-09-960> PMID: [19852585](#)
34. Glauber R. Time-dependent statistics of the Ising model. *J Math Phys*. 1963; 4(2):294–307. <https://doi.org/10.1063/1.1703954>
35. Gewaltig MO, Diesmann M. NEST (NEural Simulation Tool). *Scholarpedia*. 2007; 2(4):1430. <https://doi.org/10.4249/scholarpedia.1430>
36. Eppler JM, Pauli R, Peyser A, Ippen T, Morrison A, Senk J, et al. NEST 2.8.0; 2015. Available from: <http://dx.doi.org/10.5281/zenodo.32969>.
37. Buice MA, Cowan JD, Chow CC. Systematic Fluctuation Expansion for Neural Network Activity Equations. *Neural Comput*. 2010; 22:377–426. <https://doi.org/10.1162/neco.2009.02-09-960> PMID: [19852585](#)
38. van Vreeswijk C. What Is the Neural Code? In: van Hemmen L, Sejnowski T, editors. *23 Problems in Systems Neuroscience*. Oxford University Press; 2006. p. 143–159.
39. Dahmen D, Bos H, Helias M. Correlated Fluctuations in Strongly Coupled Binary Networks Beyond Equilibrium. *Phys Rev X*. 2016 Aug; 6:031024.
40. Cichocki A, Zdunek R, Phan AH, Amari Si. *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons; 2009.
41. Jones E, Oliphant T, Peterson P, et al. *SciPy: Open source scientific tools for Python*; 2001. <http://www.scipy.org/>.
42. Yoshimura Y, Callaway EM. Fine-scale specificity of cortical networks depends on inhibitory cell type and connectivity. *Nat Neurosci*. 2005; 8(11):1552–1559. <https://doi.org/10.1038/nn1565> PMID: [16222228](#)

43. Avermann M, Tomm C, Mateo C, Gerstner W, Petersen C. Microcircuits of excitatory and inhibitory neurons in layer 2/3 of mouse barrel cortex. *J Neurophysiol*. 2012; 107(11):3116–3134. <https://doi.org/10.1152/jn.00917.2011> PMID: 22402650
44. Lefort S, Tomm C, Sarria JCF, Petersen CCH. The Excitatory Neuronal Network of the C2 Barrel Column in Mouse Primary Somatosensory Cortex. *Neuron*. 2009; 61(2):301–316. Available from: <http://www.sciencedirect.com/science/article/pii/S0896627308010921>. <https://doi.org/10.1016/j.neuron.2008.12.020> PMID: 19186171
45. Gentet L, Avermann M, Matyas F, Staiger JF, Petersen CCH. Membrane Potential Dynamics of GABAergic Neurons in the Barrel Cortex of Behaving Mice. *Neuron*. 2010; 65:422–435. <https://doi.org/10.1016/j.neuron.2010.01.006> PMID: 20159454
46. Grytskyy D, Tetzlaff T, Diesmann M, Helias M. A unified view on weakly correlated recurrent networks. *Front Comput Neurosci*. 2013; 7:131. <https://doi.org/10.3389/fncom.2013.00131> PMID: 24151463
47. Oppenheim A, Willsky A. Systems and signals. Prentice Hall; 1996.
48. Shadlen MN, Newsome WT. Noise, neural codes and cortical organization. *Curr Opin Neurobiol*. 1994; 4(4):569–579. [https://doi.org/10.1016/0959-4388\(94\)90059-0](https://doi.org/10.1016/0959-4388(94)90059-0) PMID: 7812147
49. Song S, Sjöström P, Reigl M, Nelson S, Chklovskii D. Highly nonrandom features of synaptic connectivity in local cortical circuits. 2005; 3(3):e68.
50. Ikegaya Y, Sasaki T, Ishikawa D, Honma N, Tao K, Takahashi N, et al. Interpyramid Spike Transmission Stabilizes the Sparseness of Recurrent Network Activity. *Cereb Cortex*. 2013; 23(2):293–304. <https://doi.org/10.1093/cercor/bhs006> PMID: 22314044
51. Buzsaki G, Mizuseki K. The log-dynamic brain: how skewed distributions affect network operations. *Nat Rev Neurosci*. 2014 Feb; 15:264–278. <https://doi.org/10.1038/nrn3687> PMID: 24569488
52. Brunel N, Hakim V. Fast Global Oscillations in Networks of Integrate-and-Fire Neurons with Low Firing Rates. *Neural Comput*. 1999; 11(7):1621–1671. <https://doi.org/10.1162/089976699300016179> PMID: 10490941
53. Lindner B, Schimansky-Geier L. Transmission of noise coded versus additive signals through a neuronal ensemble. *Phys Rev Lett*. 2001; 86:2934–2937. <https://doi.org/10.1103/PhysRevLett.86.2934> PMID: 11290076
54. Brunel N, Latham P. Firing rate of the noisy quadratic integrate-and-fire neuron. *Neural Comput*. 2003; 15(10):2281–2306. <https://doi.org/10.1162/089976603322362365> PMID: 14511522
55. Grün S, Aertsen A, Abeles M, Gerstein G. Unitary Events in Cortical Multi-Neuron Activity. In: van Pelt J, Corner MA, Uylings H, van Veen MP, van Ooyen A, editors. *The Self-Organizing Brain. From Growth Cones to Functional Networks*. Amsterdam, Netherlands Inst Brain Res, Royal Netherl Acad Sci, Grad School Neurosci, Amsterdam; 1993. p. 94–95.
56. Grün S, Diesmann M, Aertsen A. 'Unitary Events' in Multiple Single-Neuron Spiking Activity. I. Detection and Significance. *Neural Comput*. 2002; 14(1):43–80. <https://doi.org/10.1162/089976602753284455> PMID: 11747534
57. Grün S, Diesmann M, Aertsen A. 'Unitary Events' in Multiple Single-Neuron Spiking Activity. II. Non-Stationary Data. *Neural Comput*. 2002; 14(1):81–119. <https://doi.org/10.1162/089976602753284464> PMID: 11747535
58. Yoshimura Y, Dantzker JLM, Callaway EM. Excitatory cortical neurons form fine-scale functional networks. *Nature*. 2005; 433(24):868–873. <https://doi.org/10.1038/nature03252> PMID: 15729343
59. Litwin-Kumar A, Chacron MJ, Doiron B. The Spatial Structure of Stimuli Shapes the Timescale of Correlations in Population Spiking Activity. *PLoS Comput Biol*. 2012; 8(9):e1002667. <https://doi.org/10.1371/journal.pcbi.1002667> PMID: 23028274
60. Harris KD, Thiele A. Cortical state and attention. *Nat Rev Neurosci*. 2011 September; 12:509–523. <https://doi.org/10.1038/nrn3084> PMID: 21829219
61. Singer W. Neuronal synchrony: a versatile code for the definition of relations? *Neuron*. 1999 Sep; 24(1):49–65. [https://doi.org/10.1016/S0896-6273\(00\)80821-1](https://doi.org/10.1016/S0896-6273(00)80821-1) PMID: 10677026
62. Womelsdorf T, Schoffelen JM, Oostenveld R, Singer W, Desimone R, Engel AK, et al. Modulation of Neuronal Interactions Through Neuronal Synchronization. *Science*. 2007 June; 316(15):1609–1612. <https://doi.org/10.1126/science.1139597> PMID: 17569862