ARTICLE **Communicated by Nicolas Brunel**

Correlations and Population Dynamics in Cortical Networks

Birgit Kriener*
kriener@biologie.uni-freiburg.de
Bernstein Center for Computational Neuroscience, and Neurobiology and Biophysics, Faculty of Biology, Albert-Ludwigs-University, D-79104 Freiburg, Germany

Tom Tetzlaff*
tom.tetzlaff@umb.no
Bernstein Center for Computational Neuroscience, Albert-Ludwigs-University, D-79104 Freiburg, Germany, and Institute of Mathematical Sciences and Technology, Norwegian University of Life Sciences, N-1432 Ås, Norway

Ad Aertsen
aertsen@biologie.uni-freiburg.de
Bernstein Center for Computational Neuroscience, and Neurobiology and Biophysics, Faculty of Biology, Albert-Ludwigs-University, D-79104 Freiburg, Germany

Markus Diesmann
diesmann@brain.riken.jp
Bernstein Center for Computational Neuroscience, Albert-Ludwigs-University, D-79104 Freiburg, Germany, and Brain Science Institute, RIKEN, Wako City, Saitama 351-0198, Japan

Stefan Rotter
stefan.rotter@biologie.uni-freiburg.de
Bernstein Center for Computational Neuroscience, Albert-Ludwigs-University, D-79104 Freiburg, Germany, and Theory and Data Analysis, Institute for Frontier Areas of Psychology and Mental Health, D-79098 Freiburg, Germany

The function of cortical networks depends on the collective interplay between neurons and neuronal populations, which is reflected in the correlation of signals that can be recorded at different levels. To correctly interpret these observations it is important to understand the origin of neuronal correlations. Here we study how cells in large recurrent networks of excitatory and inhibitory neurons interact and how the associated correlations affect stationary states of idle network activity. We

*Birgit Kriener and Tom Tetzlaff contributed equally to this work. Tom Tetzlaff is presently affiliated with the Norwegian University of Life Sciences.

demonstrate that the structure of the connectivity matrix of such networks induces considerable correlations between synaptic currents as well as between subthreshold membrane potentials, provided Dale's principle is respected. If, in contrast, synaptic weights are randomly distributed, input correlations can vanish, even for densely connected networks. Although correlations are strongly attenuated when proceeding from membrane potentials to action potentials (spikes), the resulting weak correlations in the spike output can cause substantial fluctuations in the population activity, even in highly diluted networks. We show that simple mean-field models that take the structure of the coupling matrix into account can adequately describe the power spectra of the population activity. The consequences of Dale's principle on correlations and rate fluctuations are discussed in the light of recent experimental findings.

1 Introduction

The collective dynamics of balanced random networks was extensively studied in the past decade assuming different neuron models as constituting dynamical units (van Vreeswijk & Sompolinsky, 1996, 1998; Brunel & Hakim, 1999; Brunel, 2000; Mattia & Del Guidice, 2004). These models have in common that they all assume random network topologies with sparse connectivity for local, but large, neuronal networks that are embedded into an unspecific external population, supplying excitatory drive to the local network. These networks are basic models for cortical networks of about 1 mm^3 volume and display activity states as observed in vivo. In these models, it is commonly assumed that all synapses of inhibitory neurons cause hyperpolarization, while excitatory synapses have a depolarizing effect on their postsynaptic targets. This constraint is an interpretation of Dale's principle (Li & Dayan, 1999; Hoppensteadt & Izhikevich, 1997), and we will refer to it that way throughout this article. Recent experimental evidence (Ren, Yoshimura, Takada, Horibe, & Komatsu, 2007), however, hints at the existence of an inhibitory neocortical pathway that involves interpyramidal inhibitory postsynaptic currents (IPSCs). Ren et al. (2007) observed that in about 28% of tested cell pairs, action potentials generated in a single layer II/III pyramidal neuron (mouse visual cortex) are sufficient to reliably evoke large, constant-latency IPSCs in the other pyramidal cell. Though the exact nature of this pathway is still under debate (Connors & Cruikshank, 2007), Ren et al. hypothesize axo-axonic glutamate receptor–mediated excitation of the nerve endings of inhibitory interneurons, thus bypassing dendrites, soma, and axonal trunk of the involved interneuron. Effectively, this yields an inhibitory projection between two excitatory cells and, hence, a violation of Dale's principle. In this article, we systematically study the effects of a general regard or neglect of Dale's principle on the dynamics and most of all pairwise correlations in balanced random networks.

The perhaps most realistic while still analytically tractable model of balanced random networks was presented by Brunel (2000), describing the dynamical behavior of local networks consisting of leaky integrate-and-fire neurons with current-based synapses, of which 80% are excitatory cells and the residual fraction is inhibitory. Each neuron is supposed to receive a large number of weak synaptic inputs; hence, the input current can be decomposed into a mean and a fluctuating part. The activity dynamics of the network can then be described by a Fokker-Planck equation for the probability distribution $P(V, t)$ of the membrane depolarization $V(t)$, and the stationary rate is derived self-consistently. Further stability analysis of the stationary states revealed four qualitative states the network can assume, depending on the relative strength g of inhibition over excitation and the relative external input rate ν_{ext}. In particular, a state of asynchronous population activity combined with irregular individual neuron spiking (AI state) was described for net recurrent inhibition and sufficiently strong external drive. Numerical simulations of that state, however, displayed residual globally synchronous activity that arises due to the finite size of the network, as was demonstrated in Brunel (2000) and Mattia and Del Guidice (2004). Even for highly diluted networks of size 5×10^5 neurons with synapse numbers of about 10^3 to 10^4 (i.e., connectivities as low as $\epsilon = 0.002$–0.02), the predicted AI state is not asynchronous, as reflected in a high variance of the population activity (Tetzlaff, Morrison, Timme, & Diesmann, 2005).

Hence, to obtain local cortical networks with asynchronous-irregular activity, dilution of the connection density beyond biologically plausible values is required. One way to reduce fluctuations of the AI state is to introduce heterogeneities, for example, a distribution of transmission delays or a distribution of the number of incoming synapses per neuron (cf. Brunel, 2000; Mattia & Del Guidice, 2004; Tetzlaff et al., 2005). As we demonstrate in this article, such modifications have an impact on only limited frequency bands and do not attenuate fluctuations completely. Residual fluctuations are dramatically attenuated, however, if we admit that Dale's principle is violated and allow neurons to form both excitatory and inhibitory synapses on their axons. This hybrid wiring scheme is sketched in Figure 1B, while a coupling matrix respecting Dale's principle is shown in Figure 1A. As demonstrated by numerical simulation (see Figures 1C and 1D) of otherwise identical systems, this connectivity randomization procedure leads to strongly attenuated global fluctuations. The distribution of population spike counts is broad and highly skewed in networks respecting Dale's principle, whereas it is comparably narrow and symmetric for the randomized (hybrid) scenario (see Figure 2C). The variance of the population activity is one order of magnitude larger in the Dale case (see Figure 2B), while the mean firing rate is the same (see Figure 2A). The power spectra of the population activity show that especially the frequency band up to 1 kHz is modulated by this Dale effect (see Figure 2D). Further deviations from the analytical predictions in Brunel (2000) concerning the fast synchronous irregular mode

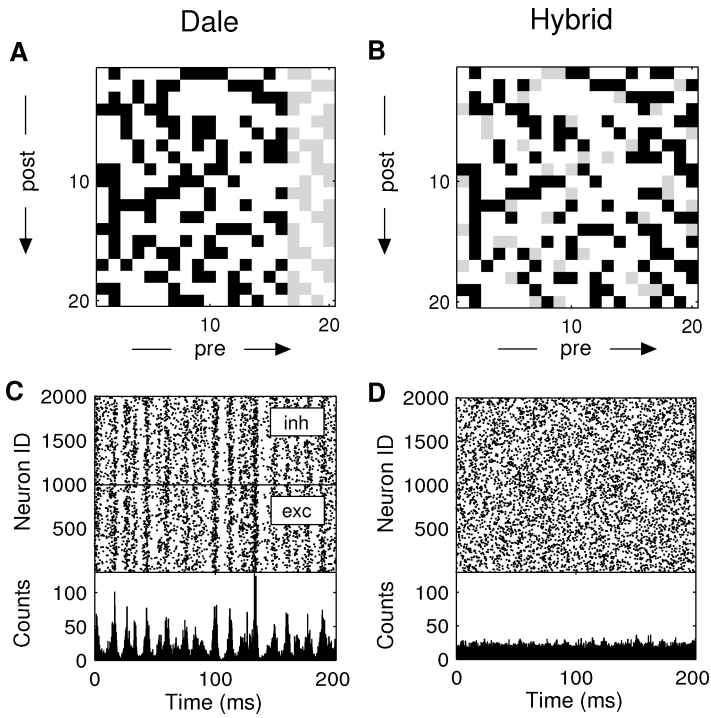

Figure 1: Population spiking activity in Dale-conform (left column) and random (hybrid; right column) networks. (A, B) Connectivity matrix (sketch) in the Dale (A) and hybrid (B) case ($N = 20$ neurons, connection probability $\epsilon = 0.4$, 80% excitatory, 20% inhibitory synapses). Black squares depict positive, gray squares negative, and white squares zero synaptic weights. Each column represents the output (axon) of a particular (presynaptic) neuron. In *A*, a neuron either excites all its postsynaptic targets or inhibits them. The hybrid architecture shown in *B*, allows both types of output synapses for each neuron. The topology and mean input (sum over all entries in a row) are the same in both cases. (C, D) Spiking activity (dot displays) for 2000 randomly selected neurons (1000 inhibitory and 1000 excitatory neurons in C) and population histograms (bottom panels; bin size 0.1 ms) in Dale (C) and hybrid (D) networks of size $N = 12{,}500$ (simulation results; see section 2 for network and simulation parameters).

(SI_{fast} state) are also much weaker if we apply the hybrid coupling scheme (not shown here). Our findings indicate that the functional segregation of excitation and inhibition has a very strong impact on network dynamics. Moreover, the violation of Dale's principle by mechanisms as suggested by Ren et al. (2007) can be a means to stabilize both the asynchronous-irregular and the fast synchronous-irregular states of balanced random networks.

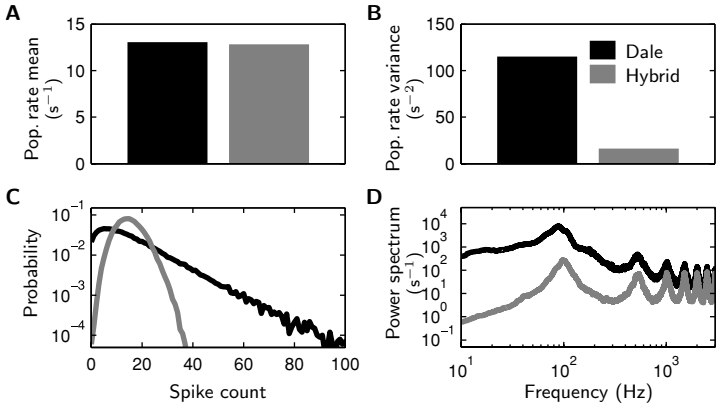

Figure 2: Spike count statistics (bin size 0.1 ms) for a network respecting Dale's principle (black curves and bars) and for a network ignoring it (gray curves and bars). Time- and population-averaged population rates (A) and population rate variances (B), spike count distributions (C), and power spectra of population rates (D, smoothed by a boxcar of width 10 Hz). Graphs show simulation results for networks composed of $N = 12{,}500$ integrate-and-fire neurons ($N_E = 10{,}000$ excitatory, $N_I = 2500$ inhibitory neurons, simulation time 10 s, temporal resolution 0.1 ms).

The article is structured as follows. In section 2, we specify the neuron and network model. In section 3, we show that the fluctuations in the global spiking activity can be understood from an analysis of the joint input statistics of pairs of neurons. The input currents to neurons in Dale-type networks are considerably correlated. These correlations are partially transmitted to the spike outputs. These spike correlations induced by common input cause the strong fluctuations of the population signals seen in the simulations. We point out that this effect has to be taken seriously unless networks are diluted beyond the biological limit. In section 4 we demonstrate that the second-order statistics of the population activity can be described in terms of mean-field models if the coupling strengths are replaced by ensemble expectations such that input correlations are preserved. A simple linear Wilson-Cowan-type rate model (Wilson & Cowan, 1972; Dayan & Abbott, 2001) turns out to be sufficient to capture the most important qualitative features of the population rate power spectra and reproduce the observed amplification of fluctuations by Dale's principle. In section 5, we show that the strong fluctuations in the global activity caused by Dale's principle are also present in heterogeneous networks with distributed neuron and connectivity parameters. Heterogeneities can suppress input correlations only in limited frequency bands. By contrast, the correlations induced by Dale's principle are not restricted to a particular timescale.

2 Model

Throughout the article, we assume the model and parameters as introduced in Brunel (2000), if not stated otherwise. Here, we briefly recapitulate the key features of this model. All notations are summarized in appendix D.

2.1 Neuronal Dynamics and Synaptic Input. The neurons in the network of size N are modeled as leaky integrate-and-fire point neurons with current-based synapses. Their membrane potential dynamics $V_i(t)$, $i \in \{1, \ldots, N\}$, are governed by

$$\tau_{\mathrm{m}} \dot{V}_i(t) = -V_i(t) + RI_i(t), \tag{2.1}$$

with membrane resistance R and membrane time constant τ_{m}. Whenever $V_i(t)$ reaches the threshold θ, a spike is emitted, $V_i(t)$ is reset to V_{res}, and the neuron stays refractory for a period τ_{ref}. Synaptic inputs

$$RI_{\mathrm{loc},i} = \tau_{\mathrm{m}} \sum_{j=1}^{N} J_{ij} \sum_{k} \delta(t - t_{jk} - d) \tag{2.2}$$

from the local network are modeled as delta currents. Whenever a presynaptic neuron j fires an action potential at time t_{jk}, it will evoke an exponential postsynaptic potential (PSP) of amplitude

$$J_{ij} = \begin{cases} J & \text{if the synapse } j \to i \text{ is excitatory,} \\ -gJ & \text{if the synapse } j \to i \text{ is inhibitory,} \\ 0 & \text{if the synapse } j \to i \text{ does not exist} \end{cases} \tag{2.3}$$

after a fixed transmission delay d. Note that multiple connections between two neurons and self-connections are excluded in this framework. In addition to the local input, each neuron receives a constant external current I_{ext} mimicking inputs from other cortical areas or subcortical regions. The total input is thus given by

$$I_i(t) = I_{\mathrm{loc},i}(t) + I_{\mathrm{ext}}. \tag{2.4}$$

2.2 Network Structure. The network consists of N neurons, of which $N_{\mathrm{E}} = \beta N$ are excitatory and $N_{\mathrm{I}} = (1 - \beta)N$ ($\beta \in [0, 1]$) inhibitory. The network topology is random; all neurons are connected with equal probability $\epsilon \in [0, 1]$, regardless of their identity. The only constraint is that all neurons receive the same number of excitatory and inhibitory synapses: $K_{\mathrm{E}} = \epsilon N_{\mathrm{E}}$ and $K_{\mathrm{I}} = \epsilon N_{\mathrm{I}}$, respectively.

Networks composed of neurons that exclusively hyperpolarize (inhibitory neurons) or exclusively depolarize (excitatory neurons) all their postsynaptic targets will be called Dale networks. Whenever this constraint is violated (i.e., any neuron has both hyperpolarizing and depolarizing synaptic projections), the network is no longer regarded as Dale-conform. Specifically, a network where the weights J and $-gJ$ are randomly distributed will be referred to as a hybrid network. A sketch of the respective coupling matrices is given in Figures 1A and 1B.

2.3 Parameters. The neuron parameters are set to $\tau_m = 20$ ms, $R = 80$ MΩ, $J = 0.1$ mV, and $d = 2$ ms. The firing threshold θ is 20 mV and the reset potential $V_{res} = 0$ mV. After a spike event, the neurons are refractory for $\tau_{ref} = 2$ ms. If not stated otherwise, all simulations were performed for networks of size $N = 12{,}500$, with $\beta = 0.8$: $N_E = 10{,}000$ and $N_I = 2{,}500$. The connection probability is set to $\epsilon = 0.1$, such that each neuron receives exactly $K_E = 1000$ excitatory and $K_I = 250$ inhibitory inputs. For $g = 4$, inhibition hence balances excitation in the local network, while for $g > 4$, the local network is net inhibition dominated. Here, we chose $g = 6$.

In Brunel (2000), external inputs were modeled as K_{ext} independent Poissonian sources with frequency ν_{ext} measured in units of $\nu_{thr} = \theta/(J\,K_{ext}\tau_m)$, that is, the input rate a neuron needs to reach the threshold in the absence of recurrent feedback. It has been shown (van Vreeswijk & Sompolinsky, 1996) that the randomness of external inputs is not a crucial premise for the asynchronous irregular dynamics we are primarily interested in. For our simulations, we therefore used an external direct current input with the amplitude $J\,K_{ext}\nu_{ext}\tau_m/R$ (the DC amplitude was set to 375 pA, if not stated elsewise).

All network simulations were performed using the NEST simulation tool (see NEST Initiative, 2006) with a temporal resolution of 0.1 ms. For details of the simulation technique see Morrison, Mehring, Geisel, Aertsen, and Diesmann, 2005.

3 Correlations in Dale and Hybrid Networks

The aim of this section is to explain how the structure of the network determines the spike correlations in the network and, hence, the variance of the population signal. We will quantify the spiking activity of neuron k in terms of a count variable $z_k(t; h)$ representing the number of emitted spikes in the time interval $[t, t + h)$. For convenience, the parameter h will be omitted: $z_k(t) = z_k(t; h)$. The statistics of the compound signal

$$Z(t) = \sum_{k=1}^{M} z_k(t) \tag{3.1}$$

of a population of size M will be described in terms of temporal averages:

$$\mathrm{E}_t[\cdot] := \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathrm{d}t \ \ldots. \tag{3.2}$$

If all involved processes are stationary in time, this average can be replaced by an ensemble expectation (average over trials with different initial conditions) without changing the results of this section.

The variance of the population signal $Z(t)$ is determined by the variances and covariances of the individual spike counts:

$$\mathrm{Var}[Z(t)] = \sum_k \mathrm{Var}[z_k(t)] + \sum_k \sum_{l\neq k} \mathrm{Cov}[z_k(t), z_l(t)]. \tag{3.3}$$

In terms of population-averaged measures, equation 3.3 reads

$$\mathrm{Var}[Z] = M\lambda(1 + c_s[M-1]), \tag{3.4}$$

with

$$\lambda := \frac{1}{M} \sum_k^M \mathrm{Var}[z_k] \tag{3.5}$$

being the population-averaged variance and

$$c_s := \frac{1}{\lambda} \frac{1}{M(M-1)} \sum_k^M \sum_{l\neq k}^M \mathrm{Cov}[z_k, z_l] \tag{3.6}$$

the population-averaged spike train correlation coefficient.

Generally activity in Dale and hybrid networks can differ in both λ and c. In the following, we will treat individual spike trains as Poisson processes. In this case, $\lambda = \nu h$ is related to the average network firing rate ν via the spike count window h. As we will discuss in section 3.1.1, the average firing rates are approximately identical in Dale and hybrid networks (see Figure 2A). Thus, under the Poisson assumption, only the spike train correlation c_s can be responsible for the observed difference in the variance of the population signal in the two network types. It remains to be clarified why spike correlations in networks respecting Dale's principle should differ from those in hybrid networks. For this purpose, we need to take a closer look at the input statistics of two neurons k and l.

3.1 Input Statistics. In the following, we describe the total input $I_k(t)$ of neuron k as a shot-noise process, that is, a superposition of linearly filtered

presynaptic spike trains $S_i(t)$ (see appendix A):

$$I_k(t) = I_{\text{ext}} + \sum_{i=1}^{N} (S_i * f_{ki})(t). \tag{3.7}$$

For our purpose, the choice of the filter kernel $f_{ki}(t)$ is not relevant since we assume that individual spike trains can be modeled as stationary Poisson processes with delta-shaped auto- and cross-correlation functions (see appendix A). In the context of our model, $I_k(t)$ can represent, for example, the weighted input spike count, the synaptic input current ($f_{ik}(t) =$ unit postsynaptic current, PSC), or the subthreshold membrane potential ($f_{ik}(t) =$ unit postsynaptic potential, PSP). As we assume that excitatory and inhibitory synapses differ only in strength J_{ij} but not in their kinetics, individual synapses are fully described by

$$f_{ki}(t) = A_{ki} f(t), \tag{3.8}$$

with $A_{ki} := J_{ki}/J \in \{1, -g, 0\}$ (cf. equation 2.3) and a fixed linear kernel $f(t)$. By defining the filtered activity of neuron i as

$$s_i(t) := (S_i * f)(t), \tag{3.9}$$

we may rewrite equation 3.7 as

$$I_k(t) = I_{\text{ext}} + \sum_{i=1}^{N} A_{ki} s_i(t). \tag{3.10}$$

As the model networks are homogeneous (identical neurons with fixed number of input synapses K_{E} and K_{I} and delays d), we will assume that the time-averaged mean values and auto- and cross-covariance functions of the filtered spike signals are constant:

$$\begin{aligned} \mathrm{E}_t\left[s_i(t)\right] &= \alpha_1 \nu \qquad (\forall i \in [1, N]), \\ \mathrm{Cov}[s_i(t) s_j(t+\tau)] &= \mathrm{E}_t\left[s_i(t) s_j(t+\tau)\right] - \mathrm{E}_t\left[s_i(t)\right] \mathrm{E}\left[s_j(t+\tau)\right] \\ &=: \begin{cases} a_s(\tau) & \forall i = j \\ c_s(\tau) & \forall i \neq j. \end{cases} \end{aligned} \tag{3.11}$$

Assuming that all spike trains can be treated as Poisson processes with delta-shaped correlation functions, the variance and covariance of the filtered processes become

$$a_s(0) = \alpha_2 \nu \quad \text{and} \quad c_s(0) = c_s \alpha_2 \nu. \tag{3.12}$$

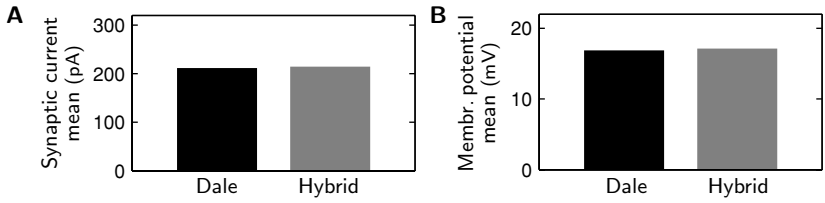

Figure 3: Time- and population-averaged synaptic input currents (A) and free membrane potentials (B, membrane potentials without spiking dynamics) for Dale (black bars) and hybrid networks (gray bars). Graphs show simulation results for networks composed of $N = 12{,}500$ I&F neurons ($\epsilon = 0.1$, $\beta = 0.8$, $g = 6$, simulation time 10 s, temporal resolution 0.1 ms). Synaptic currents were recorded from 2000 randomly selected neurons. Free membrane potentials were obtained by low-pass filtering (membrane time constant $\tau_{\mathrm{m}} = 20$ ms) of synaptic currents.

Here, ν denotes the population-averaged firing rate and c_s the spike correlation coefficient defined in equation 3.6. The constants $\alpha_1 = \int \mathrm{d}t\ f(t)$ and $\alpha_2 = \int \mathrm{d}t\ f(t)^2$ follow from shot-noise theory (see appendix A). As demonstrated in the following discussion, the homogeneity assumption, equation 3.11, enables us to separate the statistics of the spike trains from those of the network structure.

3.1.1 Mean Input. According to equation 3.11, the average synaptic input of a neuron k is given by

$$\mathrm{E}_t\left[I_k(t)\right] = I_{\mathrm{ext}} + \alpha_1 \nu \sum_{i=1}^{N} A_{ki}. \tag{3.13}$$

As we assumed that each postsynaptic neuron k receives the same numbers $K_{\mathrm{E}} = \epsilon\beta N$, $K_{\mathrm{I}} = \epsilon(1-\beta)N$ of excitatory and inhibitory inputs, we obtain

$$\mathrm{E}_t\left[I_k(t)\right] = I_{\mathrm{ext}} + \alpha_1 \nu \epsilon N(\beta - g[1-\beta]). \tag{3.14}$$

Note that equation 3.14 holds for both Dale and hybrid networks.

Network simulations with integrate-and-fire neurons showed that the average firing rates ν are identical in both types of networks (see Figure 2A). For the mean input $\mathrm{E}_t\left[I_k(t)\right]$, it is therefore irrelevant whether the network is wired according to Dale's principle. Indeed, recordings of synaptic input currents and membrane potentials in network simulations revealed the same mean values in both scenarios (see Figure 3).

3.1.2 Input Correlations. Given the homogeneity assumption 3.11, the input covariance function reads

$$\mathrm{Cov}[I_k(t)I_l(t+\tau)] = a_s(\tau)\sum_{i=1}^{N} A_{ki}A_{li} + c_s(\tau)\sum_{i=1}^{N}\sum_{j\neq i}^{N} A_{ki}A_{lj}. \tag{3.15}$$

With the input auto- and cross-covariance functions defined as

$$\mathrm{Cov}[I_k(t)I_l(t+\tau)] =: \begin{cases} a_{\mathrm{in}}(\tau) & \forall k = l \\ c_{\mathrm{in}}(\tau) & \forall k \neq l \end{cases} \tag{3.16}$$

and the constants

$$H := \sum_{i=1}^{N} A_{ki}^2, \quad L := \sum_{i=1}^{N}\sum_{j\neq i}^{N} A_{ki}A_{lj} \quad \text{and} \quad G := \sum_{i=1}^{N} A_{ki}A_{li}, \tag{3.17}$$

we may rewrite equation 3.15 in matrix notation:

$$\begin{pmatrix} a_{\mathrm{in}}(\tau) \\ c_{\mathrm{in}}(\tau) \end{pmatrix} = \begin{pmatrix} H & L \\ G & L \end{pmatrix} \begin{pmatrix} a_s(\tau) \\ c_s(\tau) \end{pmatrix}. \tag{3.18}$$

For a given network realization, the coefficients in equation 3.17, and therefore also the input auto- and cross-covariances, may generally depend on the postsynaptic neuron indices k and l. Since all neurons in the network receive K_{E} excitatory and K_{I} inhibitory inputs, the coefficients H and L are, however, independent of the target cells:

$$\begin{aligned} H &= K_{\mathrm{E}} + g^2 K_{\mathrm{I}} = K(\beta + g^2[1-\beta]) \\ L &= (K_{\mathrm{E}} - gK_{\mathrm{I}})^2 = K^2(\beta - g[1-\beta])^2. \end{aligned} \tag{3.19}$$

Note that for equation 3.19, we assumed that $K_{\mathrm{E}} \gg 1$ and $K_{\mathrm{I}} \gg 1$. The third coefficient G reflects the presynaptic overlap between the two postsynaptic neurons k and l. In networks respecting Dale's principle, the common sources are either excitatory or inhibitory for both neurons k and l (see Figure 4A). In hybrid networks, however, their impact can be excitatory for one and inhibitory for the other target (see Figure 4B). As shown in appendix B, this leads to quite distinct results for the coefficient G in the two scenarios:

$$G_{\mathrm{D}} = \epsilon K(\beta + g^2[1-\beta]) = \epsilon H \quad \text{(Dale)}, \tag{3.20}$$

$$G_{\mathrm{H}} = \epsilon K(\beta - g[1-\beta])^2 = \epsilon\, L/K \quad \text{(hybrid)}. \tag{3.21}$$

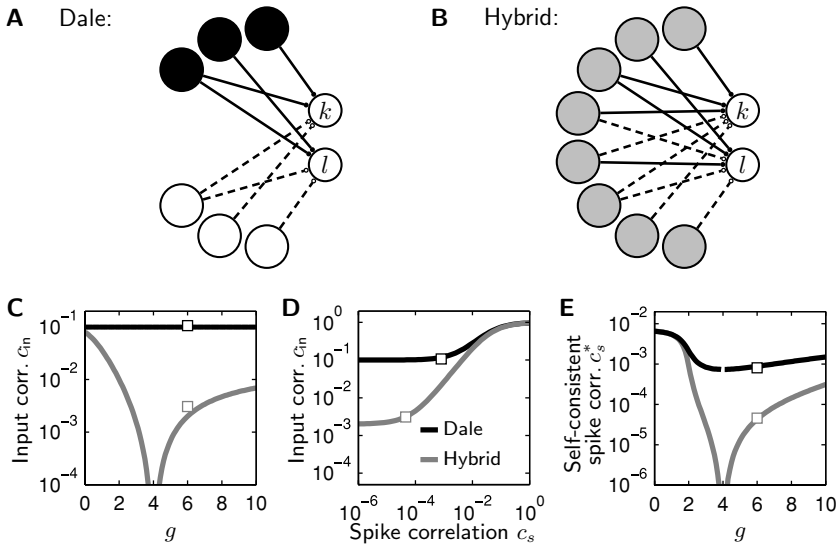

Figure 4: Input scenarios in Dale (A) and hybrid networks (B). Two neurons k and l receive excitatory (solid lines) and inhibitory inputs (dashed lines) from different types of presynaptic sources. While in the Dale scenario presynaptic neurons are either excitatory (black circles) or inhibitory (white circles), this distinction cannot be made in the hybrid case (gray circles). (C) Input correlation c_{in} for uncorrelated spiking ($c_s = 0$) as a function of the relative strength g of inhibition in networks respecting Dale's principle (black curves) and in hybrid networks (gray curves). (D) Dependence of input correlations c_{in} on spike train correlations c_s for fixed $g = 6$. (E) Self-consistent spike correlation c_s^* as a function of the relative strength g of inhibition. Lines show analytical and symbols network simulation results (cf. Figure 7, $\epsilon = 0.1$, $\beta = 0.8$, $K = 1250$).

For our standard parameters $g = 6$, $\beta = 0.8$ they differ by a factor of $G_{\text{D}}/G_{\text{H}} = KH/L = 50$. Intuitively, the difference between the Dale and the hybrid case can be understood as follows. In the Dale scenario, the contributions to the input correlation due to shared presynaptic targets are always positive. In the hybrid case, however, they are negative for those common sources that excite one and inhibit the other target. The total input correlation is therefore reduced.

If the covariance functions $a_s(\tau)$ and $c_s(\tau)$ of the (filtered) spike trains are known, those of the input currents can be easily computed with the help of equation 3.18 for both Dale and hybrid networks. Note that by Fourier-transforming equation 3.18, the input power and cross-spectra $\hat{a}_{\text{in}}(\omega)$ and $\hat{c}_{\text{in}}(\omega)$ are fully determined too. Further, consider that the prefactor $L = K^2(\beta - g[1-\beta])^2$, which scales the impact of the spike train cross-covariance functions $c_s(\tau)$ (or cross-spectra), vanishes if the network

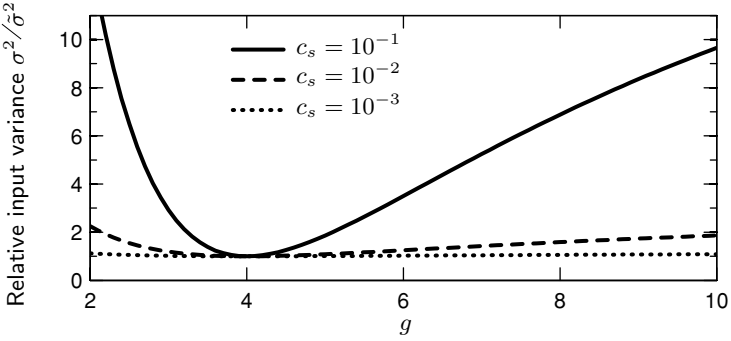

Figure 5: Dependence of relative input variance $\sigma^2/\tilde{\sigma}^2$ on the relative strength g of inhibition and on spike correlations c_s (analytical results for $\epsilon = 0.1$, $\beta = 0.8$, $K = 1250$).

is balanced, that is, if $\beta = g(1-\beta)$. Thus, in the balanced case, spike correlations $c_s(\tau)$ do not contribute to the second-order input statistics in either the Dale or hybrid scenario (cf. Salinas & Sejnowski, 2000).

According to equations 3.18 and 3.12, the variance of the input currents reads

$$\sigma^2 := a_{\text{in}}(0) = \alpha_2 \nu K \left(\beta + g^2[1-\beta] + c_s K[\beta - g(1-\beta)]^2\right). \tag{3.22}$$

In Figure 5 the input variances σ^2 in the presence of spike correlations c_s are compared with those obtained for $c_s = 0$, that is, with $\tilde{\sigma}^2 := \alpha_2 \nu K(\beta + g^2[1-\beta])$. The three curves show the ratio $\sigma^2/\tilde{\sigma}^2$ as a function of the relative strength g of inhibition for three different spike correlations: $c_s = 10^{-1}$ (solid), 10^{-2} (dashed), and 10^{-3} (dotted). In the fully balanced case, here for $g = 4$ ($\beta = 0.8$), the input variance is insensitive to correlations in the presynaptic spike train ensemble: $\sigma^2/\tilde{\sigma}^2 = 1$. For our reference network with dominant inhibition, $g = 6$, the deviation of σ from $\tilde{\sigma}$ is negligible if the spike count correlations are small ($c_s < 10^{-3}$). As we will show in section 3.2, spike count correlations are indeed tiny in both Dale and hybrid networks. We therefore conclude that the variance of the input signal is not substantially affected by Dale's principle. Indeed, in network simulations, we find that the population-averaged variances of the synaptic input current and the free membrane potential are only marginally smaller in the hybrid case (12% for synaptic currents and 8% for membrane potentials; see Figures 6A and 6B, respectively). Although significant, these differences are small compared to the large difference (86%) observed for the population spike count variances (see Figure 2B).

Figures 6C and 6D compare the distributions of synaptic currents and membrane potentials in Dale-type (black) and hybrid (gray) networks. Both

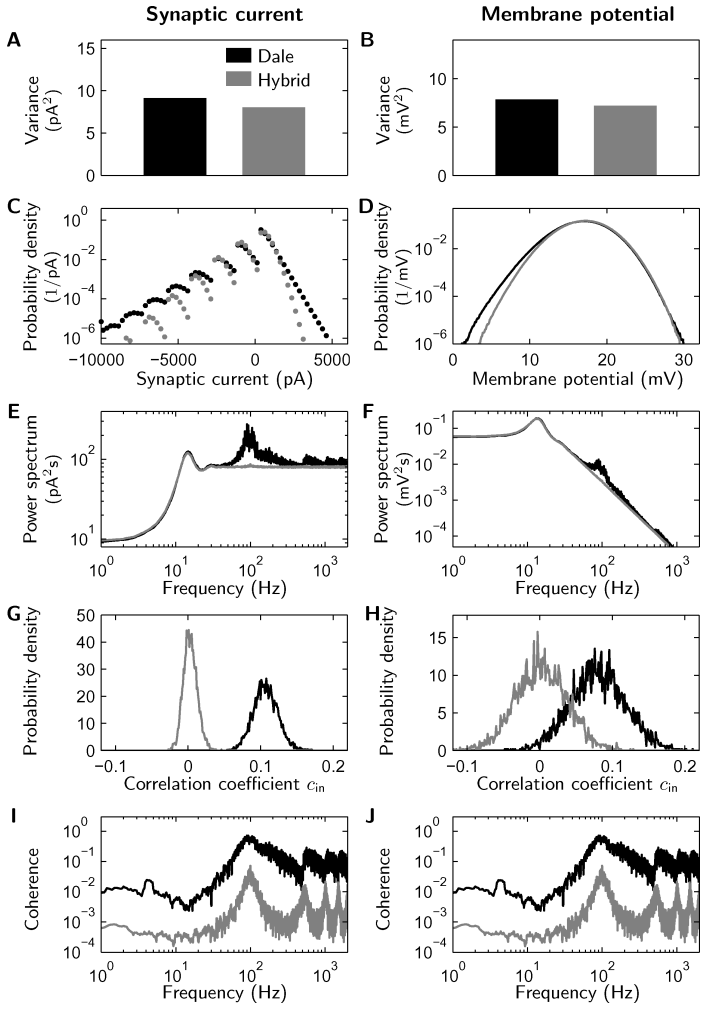

Figure 6: Second-order statistics of synaptic input currents (left column) and free membrane potentials (right column) for Dale (black curves and bars) and hybrid networks (gray curves and bars). (A, B) Time- and population-averaged variances. (C, D) Distributions of synaptic currents and membrane potentials. (E, F) Population-averaged power spectra (smoothed by moving average [frame size 1 Hz]). (G, H) Distributions of input correlation coefficients c_{in}. (I, J) Population-averaged coherences (smoothed by moving average [frame size 1 Hz]). Graphs show simulation results for networks composed of $N = 12{,}500$ I&F neurons ($\epsilon = 0.1$, $\beta = 0.8$, $g = 6$, simulation time 10 s, temporal resolution 0.1 ms). Synaptic currents were recorded from 2000 randomly selected neurons. Free membrane potentials were obtained by low-pass filtering (membrane time constant $\tau_{\mathrm{m}} = 20$ ms) of synaptic currents.

the current and the voltage distribution are slightly narrower in the hybrid case.[1] However, compared to the results obtained for the spike count signals (see Figure 2C), these differences are subtle. Also the power spectra $\hat{a}_{\text{in}}(\omega)$ (see Figures 6E and 6F) are similar in Dale and hybrid networks for frequencies below 50 Hz. At about 100 Hz, a peak is observed in Dale networks that is not visible in the hybrid case. Due to the low-pass filter characteristics of the cell membrane, this peak is flattened and becomes less significant for the membrane voltages (see Figure 6F).

Due to the differences in the coefficient G (see equations 3.20 and 3.21), the input covariances

$$c_{\text{in,D/H}}(0) = \alpha_2 \nu (G_{\text{D/H}} + c_s L) \tag{3.23}$$

in Dale and hybrid networks differ considerably, even if the spike correlations c_s are neglected. With $c_s := c_s(0)/a_s(0)$, the input correlation coefficients in general read

$$c_{\text{in}} := \frac{c_{\text{in}}(0)}{a_{\text{in}}(0)} = \frac{G + c_s L}{H + c_s L}. \tag{3.24}$$

With equation 3.20, we obtain,

$$c_{\text{in,D}} = \frac{\epsilon H + c_s L}{H + c_s L} \tag{3.25}$$

for Dale and with equation 3.21,

$$c_{\text{in,H}} = \frac{\epsilon L/K + c_s L}{H + c_s L} \tag{3.26}$$

for hybrid networks. If we assume that presynaptic neurons fire in an uncorrelated fashion, that is, $c_s = 0$, equations 3.25 and 3.26 simplify to

$$c_{\text{in,D}} = \epsilon \tag{3.27}$$

and

$$c_{\text{in,H}} = \epsilon \frac{L}{KH} = \epsilon \frac{(\beta - g[1-\beta])^2}{\beta + g^2[1-\beta]}. \tag{3.28}$$

[1] The discreteness of the current distributions (see Figure 6C) results from the delta-type currents we used in the simulations.

Thus, the correlation between the synaptic inputs in the Dale network is basically determined by the network connectivity ϵ, and barely affected by other parameters, if spike correlations are small ($c_s < 10^{-3}$). This is different in hybrid networks; here, the input correlations vanish if the network is balanced—i.e., $\beta = g(1-\beta)$. Figure 4C shows the dependence of $c_{\mathrm{in,D}}$ (black curve) and $c_{\mathrm{in,H}}$ (gray curve) on the relative inhibition strength g for $c_s = 0$ ($\epsilon = 0.1$, $\beta = 0.8$). Close to the balanced regime, input correlations in Dale and hybrid networks differ by at least one order of magnitude. This is confirmed by simulation results for our standard network with $g = 6$ (squares in Figure 4C). Figures 6G and 6H illustrate that this effect can be observed at the level of both input currents and membrane potentials, respectively.[2] The measured coherences $\kappa_{\mathrm{in}}(\omega) := |\hat{c}_{\mathrm{in}}(\omega)|/\hat{a}_{\mathrm{in}}(\omega)$ between input currents (see Figure 6I) and membrane potentials (see Figure 6J) show that Dale's principle amplifies common input correlations on all timescales (frequency bands).

The dependence of the input correlations c_{in} on correlations c_s between spike trains is illustrated in Figure 4D for fixed $g = 6$. Spike correlations c_s measured in network simulations (square symbols) are of order 10^{-3} in the Dale and 10^{-4} in the hybrid case. In this range, input correlations are well approximated by assuming $c_s = 0$. Only for large spike correlations ($c_s > 10^{-2}$) do input correlations approach the limiting value 1.

According to equations 3.27 and 3.28, input correlations depend linearly on the network connectivity $\epsilon = K/N$. We tested this by simulating networks of different sizes $N \in \{10{,}000, \ldots, 325{,}000\}$ while keeping the number of synapses per neuron constant at $K = 1{,}250$ ($\beta = 0.8$, $g = 6$). Network connectivities thus range between 0.004 and 0.125. Figure 7A compares the analytical results for $c_s = 0$ (dashed lines) with simulation results (squares) for Dale (black) and hybrid networks (gray). Not only the $1/N$ scaling but also the absolute values are correctly predicted over a wide range of network sizes.

3.2 Output Correlations. In the previous section we showed that input correlations c_{in} induced by common input are much stronger in Dale than in hybrid networks. It has been demonstrated in earlier studies (Shadlen & Newsome, 1998; Stroeve & Gielen, 2001; Tetzlaff, Buschermöhle, Geisel, & Diesmann, 2003; Moreno-Bote & Parga, 2006) that pairs of integrate-and-fire neurons transmit common input correlations to some extent to their output spike signals. Here, this result is confirmed and extended by comparing correlations c_{in} between input currents with spike count correlations c_{out} (bin size 0.1 ms) in Dale and hybrid networks for different network sizes N (see Figures 7A and 7B, squares). Although much smaller, the measured spike count correlations c_{out} reflect similar features as the input correlations

[2] The width of the distributions of correlation coefficients shown in Figures 6G and 6H is determined by the distributions of C_{E}, C_{I}, and C_{EI} (see appendix B).

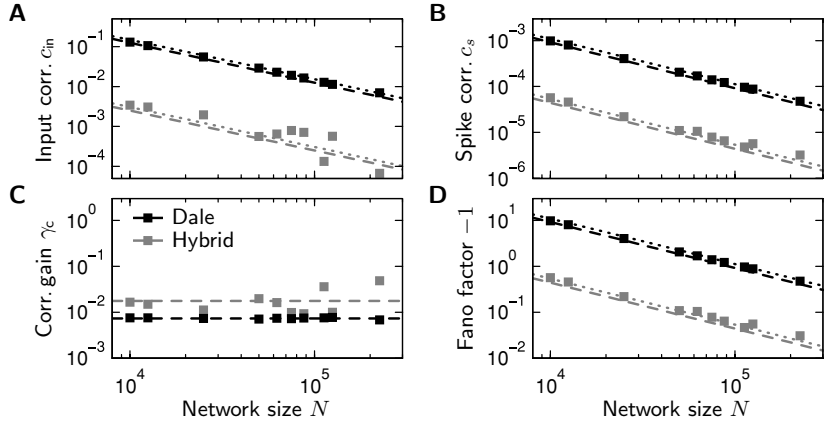

Figure 7: Population-averaged correlation coefficients for synaptic input currents c_{in} recorded from 20 neurons (A) and for spike trains c_s recorded from 2000 neurons (B, bin size 0.1 ms) as functions of the network size N (number of synapses is kept constant at $K_E = 1{,}000$, $K_I = 250$) for Dale-conform (black lines and symbols) and random (hybrid) networks (gray lines and symbols, $\beta = 0.8, \epsilon = 0.1, g = 6$). (C) Correlation gain $\gamma_c = c_s/c_{in}$ obtained from the ratios of correlation coefficients shown in *B* and *A*. (D) Relative amplitude of the population rate fluctuations measured by the Fano factor *F*. Symbols in *A*–*D* show simulation results (simulation time 10 s, temporal resolution 0.1 ms). Lines in *A*, *B*, and *D* show theoretical results for $c_s = 0$ (dashed) and for $c_s = c_s^*$ (dotted). Correlation gains $\gamma_c = 0.007$ (Dale) and 0.018 (hybrid) were obtained from simulation results (dashed lines in *C*).

c_{in} do. Again, output correlations in Dale networks are considerably larger than in hybrid networks. Moreover, we find the same $1/N$ scaling as we did for input correlations.

In a previous study (Tetzlaff et al., 2003) we showed that small input correlations are transmitted to the output in approximately linear fashion: $c_{out} \approx \gamma_c c_{in}$. Furthermore, it is known that the correlation gain γ_c depends on the marginal input statistics (Stroeve & Gielen, 2001; Tetzlaff et al., 2003; Moreno-Bote & Parga, 2006). According to section 3.1, the mean and variance of the input do not depend on the network size as long as the number of synapses $K_{E/I}$ is constant and the network is close to balance. Figure 7C (squares) shows that the correlation gain $\gamma_c = c_{out}/c_{in}$ is indeed independent of the network size N (see Figure 7C, squares). Moreover, we pointed out that (balanced) Dale and hybrid networks do not basically differ in the marginal statistics of the input currents. This is also consistent with Figure 7C: the correlation gains, while being different, are at least of the same order of magnitude in both scenarios. We will use the mean values (0.007 and 0.018 in the Dale and the hybrid case, respectively; dashed lines

in Figure 7C) to predict output correlations c_{out} from the theoretical results for c_{in} (dashed lines in Figure 7B).

Thus far, we have shown that input correlations c_{in} depend on spike correlations c_s (see Figure 4D), and output spike correlations c_{out} result from input correlations c_{in}. In a stationary situation, we must therefore expect a self-consistent spike correlation $c_s = c_{\text{out}} =: c_s^*$. Using equations 3.25 and 3.26 and assuming a linear relationship $c_{\text{out}} = \gamma_c c_{\text{in}}$ between input and output correlations, we obtain the self-consistent spike correlation c_s^* as the (positive) solution(s) of

$$c_s^* = \gamma_c \cdot c_{\text{in}}^* = \gamma_c \cdot \frac{G + c_s^* L}{H + c_s^* L}. \tag{3.29}$$

Figure 4E shows the results for Dale (black curve) and hybrid (gray curve) networks as a function of the relative strength g of inhibition with $\gamma_{c,D} = 0.007$ and $\gamma_{c,H} = 0.018$, respectively. The symbols at $g = 6$ represent spike count correlations measured in simulations. The dotted lines in Figures 7A and 7B show that incorporating the results for $c_{s,D/H}^*$ in the calculation of c_{in} and c_{out} has only a minor effect. Again, since spike correlations are small and the considered networks are close to balance, the marginal and joint input statistics are well described by assuming $c_s = 0$. However, to understand why Dale and hybrid networks differ that much in the statistics of the population signal $Z(t)$ (see Figure 2), the tiny spike correlations c_s are essential and cannot be neglected.

3.3 Network Fluctuations. At the beginning of section 3, we pointed out that the variance of the compound spike count signal $Z(t)$ of a population of size M ($M \leq N$) is given by

$$\text{Var}[Z(t)] = M\lambda\,(1 + c_s[M - 1]) \tag{3.30}$$

with $\lambda = h\nu$. In section 3.2, we argued that the spike correlation $c_s = c_{\text{out}}$ is well approximated by the linear relationship $c_{\text{out}} = \gamma_c c_{\text{in}}$. In section 3.1.2, we showed that the input correlation c_{in} in turn depends only weakly on the spike correlation c_s, provided the network is close to balance. Assuming $c_s = 0$, it shows a linear dependence on the network connectivity $\epsilon = K/N$—$c_{\text{in}} = Q\epsilon$. Dale and hybrid networks basically differ with respect to the factor Q:

$$\begin{aligned} Q_{\text{H}} &= \frac{(\beta - g[1 - \beta])^2}{\beta + g^2(1 - \beta)} \\ Q_{\text{D}} &= 1. \end{aligned} \tag{3.31}$$

In summary, the spike correlation can be written as

$$c_s = \gamma_c Q \epsilon. \tag{3.32}$$

To measure the strength of fluctuations of the population spike count, we consider the Fano factor:

$$F = \frac{\mathrm{Var}[Z(t)]}{\mathrm{E}[Z(t)]}. \tag{3.33}$$

With equations 3.30 and 3.32 and the mean $\mathrm{E}[Z(t)] = M\lambda$, we obtain

$$F = 1 + \frac{M-1}{N}\gamma_c Q K. \tag{3.34}$$

The dependence of F on the network size N (K and M fixed) is shown in Figure 7D for both the Dale and the hybrid scenarios. Results of network simulations (symbols) are in good agreement with the analytical predictions (see equation 3.34, dashed lines). As before, the correction obtained by taking the finite self-consistent spike correlation c_s^* (see equation 3.29) into account is negligible (dotted lines). In Dale networks, the deviations of F from the asynchronous Poissonian case $F = 1$ remain even for large network sizes (i.e., unrealistically small connectivities). To achieve the same Fano factor as that of a hybrid network of size $N = 12{,}500$ ($\epsilon = 0.1$), a network respecting Dale's principle would have to be enlarged to $N \approx 2 \cdot 10^5$, corresponding to a connectivity of less than 1%.

According to equation 3.34, the Fano factor F approaches 1 in the limit $N \to \infty$ if the number of observed neurons M and the number of synapses K per neuron are fixed. Thus, the strong fluctuations in the population signal (see Figure 1C) can be considered as a finite size effect (see, e.g., Brunel, 2000; Mattia & Del Guidice, 2004). They disappear in the above limit. For a broad range of reasonable connection probabilities, however, correlations caused by common input cannot be neglected.

Mass activity signals, such as the electro encephalogram (EEG), measure the activity of very large neuron populations. In our description, this might correspond to $M \sim N$. According to equation 3.34, the limit $N \to \infty$ in this scenario leads to finite deviations from the Poisson case:

$$\lim_{N\to\infty} F = 1 + \gamma_c Q K. \tag{3.35}$$

For our standard parameters $\beta = 0.8$, $g = 6$, $K = 1250$ and the measured correlation gains $\gamma_{c,D} = 0.007$ and $\gamma_{c,H} = 0.018$, the Fano factor of the whole population at large network size becomes $F_D = 9.75$ in Dale and $F_H = 1.45$

in hybrid networks. Thus, in the case $M \sim N$, the enlarged network fluctuations cannot be attributed to a finite size effect.

4 Population Dynamics

4.1 Ensemble Averages. In the previous section, we showed that a correct description of the second-order statistics of the population activity needs to consider correlations between input currents. The classical mean-field approach takes only the mean synaptic input into account. Thus, it cannot predict the amplitude of the population rate fluctuations for networks with significant input correlations. In this section, we derive an extended mean-field description for random networks respecting Dale's principle. For this purpose, we consider the input covariance 3.15 for the case $c_s = 0$ in an ensemble average (averaged across network realizations):

$$\mathrm{E}\left[\mathrm{Cov}[I_k(t), I_l(t)]\right] = \alpha_2 \nu \sum_{i=1}^{N} \mathrm{E}\left[A_{ki} A_{li}\right]. \tag{4.1}$$

In a purely random (i.e., hybrid) network, the elements of the coupling matrix are independent for $k \neq l$; the ensemble average factorizes, that is, $\mathrm{E}\left[A_{ki} A_{li}\right] = \mathrm{E}\left[A_{ki}\right] \mathrm{E}\left[A_{li}\right]$. Since the average coupling strength $\mathrm{E}\left[A_{ki}\right] = \epsilon(\beta - g[1-\beta])$ is independent of the pre- and postsynaptic neuron identities i and k, the average input covariance reads

$$\mathrm{E}\left[\mathrm{Cov}[I_k(t), I_l(t)]\right] = \alpha_2 \nu \epsilon K (\beta - g[1-\beta])^2, \tag{4.2}$$

which is identical to the result we obtained for $c_s = 0$ in the previous section. In the classical mean-field description, the weights A_{ki} in the input current 3.10 are replaced by the ensemble average $\mathrm{E}\left[A_{ki}\right]$. The resulting input covariance is identical to equation 4.2. Thus, in an ensemble average, the mean-field approach correctly predicts the input statistics in random networks, regardless of the network size N.

This is different in networks respecting Dale's principle. Here, the coupling strengths A_{ki} and A_{li} are independent only inside the excitatory or inhibitory subpopulations, that is, for $i \in \mathcal{E}$ or $i \in \mathcal{I}$. By splitting the sum in equation 4.1, we may therefore write

$$\mathrm{E}\left[\mathrm{Cov}[I_k(t), I_l(t)]\right] = \alpha_2 \nu \left(\sum_{i \in \mathcal{E}} \mathrm{E}\left[A_{ki}\right] \mathrm{E}\left[A_{li}\right] + \sum_{j \in \mathcal{I}} \mathrm{E}\left[A_{kj}\right] \mathrm{E}\left[A_{lj}\right] \right). \tag{4.3}$$

With

$$\mathrm{E}\,[A_{ki}] = \begin{cases} \epsilon & \text{if } i \in \mathcal{E} \\ -g\epsilon & \text{if } i \in \mathcal{I} \end{cases}, \tag{4.4}$$

$N_\mathrm{E} = |\mathcal{E}| = \beta N$ and $N_\mathrm{I} = |\mathcal{I}| = (1-\beta)N$, we regain the result of section 3.1.2,

$$\mathrm{E}\,[\mathrm{Cov}[I_k(t), I_l(t)]] = \alpha_2 \nu \epsilon K \left(\beta + g^2[1-\beta]\right), \tag{4.5}$$

for $c_s = 0$. The same input covariance is obtained if the coupling strengths in the input current, equation 3.10, are replaced by their expectations inside each subpopulation:

$$I_k(t) = I_\mathrm{ext} + \epsilon \sum_{i\in\mathcal{E}} s_i(t) - g\epsilon \sum_{j\in\mathcal{I}} s_j(t). \tag{4.6}$$

In such an extended mean-field description, the ensemble averaged input correlations are preserved. In the remainder of this section, we study mean-field population models for Dale and hybrid networks and show that the resulting power spectra of the population rates indeed resemble those obtained by network simulations (see Figure 2D).

4.2 From Single Neuron Dynamics to Population Rates. In the following, we derive equations describing the dynamics of the population activity from the integrate-and-fire dynamics (see section 2) of individual neurons. The subthreshold membrane potential dynamics of neuron i is fully determined by equations 2.1, 2.2, and 2.4. As defined in section 2, spikes are elicited whenever the membrane potential $V_i(t)$ crosses a threshold θ. Immediately after spike emission, the membrane potential is reset to the resting potential $V_\mathrm{res} = 0$. The reset after each spike of neuron i can be treated as a pulse-like negative current, which triggers a depolarizing membrane potential jump of amplitude θ. The $2N$ variables $\{V_i(t), S_i(t) \mid i \in \{1, \ldots, N\}\}$ in the network are then coupled by the N equations

$$\dot{V}_i(t) + \frac{1}{\tau_\mathrm{m}} V_i(t) = -\theta\, S_i(t) + \sum_{j=1}^{N} J_{ij}\, S_j(t-d) + C^{-1} I_\mathrm{ext} \qquad (i \in \{1, \ldots, N\}). \tag{4.7}$$

Here,

$$S_i(t) = \sum_k \delta(t - t_{ik}) \tag{4.8}$$

denotes the spike train generated by neuron i with spikes at times t_{ik} and C the membrane capacitance. Macroscopic mean-field models of the population dynamics can be derived by summating the single neuron signals given by equation 4.7.

In the hybrid scenario, all neurons are statistically equal (cf. Figure 4B). All entries A_{ij} of the coupling matrix are independent and identically distributed according to a common distribution with mean $\mu = J\,\epsilon(\beta - g[1 - \beta])$. In the line of reasoning of section 4.1, we sum equation 4.7 over all neurons $i \in \{1, \ldots, N\}$ and obtain a one-dimensional equation,

$$\dot{V} + \frac{1}{\tau_{\mathrm{m}}} V(t) = -\theta\, S(t) + W'\, S(t - d) + X', \tag{4.9}$$

describing the interplay between the population membrane potential,

$$V(t) := \sum_{i=1}^{N} V_i(t), \tag{4.10}$$

and the compound population spike train,

$$S(t) := \sum_{i=1}^{N} S_i(t), \tag{4.11}$$

with a coupling coefficient $W' = N\mu$ and a constant external input $X' = NC^{-1} I_{\mathrm{ext}}$.

In the Dale scenario, by contrast, all synapses on the axon of one particular neuron have the same sign: nodes are either excitatory or inhibitory. To properly take care of the structural correlations in the mean-field reduction, we sum over the individual subpopulations separately and explicitly account for the four possible interactions among different subpopulations. Let $\mathcal{E}$ collect the indices of all N_{E} excitatory neurons and $\mathcal{I}$ those of the N_{I} inhibitory neurons. Further, let

$$\mathrm{E}\left[J_{ij}\right] = \begin{cases} \mu_{\mathrm{EE}} := \epsilon J & i \in \mathcal{E},\ j \in \mathcal{E} \\ \mu_{\mathrm{EI}} := -g\epsilon J & i \in \mathcal{E},\ j \in \mathcal{I} \\ \mu_{\mathrm{IE}} := \epsilon J & i \in \mathcal{I},\ j \in \mathcal{E} \\ \mu_{\mathrm{II}} := -g\epsilon J & i \in \mathcal{I},\ j \in \mathcal{I} \end{cases} \tag{4.12}$$

denote the ensemble averages of the entries J_{ij} of the coupling matrix. With

$$\begin{aligned} V_{\mathrm{E}}(t) &:= \sum_{i\in\mathcal{E}} V_i(t), \quad V_{\mathrm{I}}(t) := \sum_{i\in\mathcal{I}} V_i(t), \\ S_{\mathrm{E}}(t) &:= \sum_{i\in\mathcal{E}} S_i(t), \quad S_{\mathrm{I}}(t) := \sum_{i\in\mathcal{I}} S_i(t), \end{aligned} \tag{4.13}$$

and $X'_{\mathrm{E}} := N_{\mathrm{E}} C^{-1} I_{\mathrm{ext}}$ and $X'_{\mathrm{I}} := N_{\mathrm{I}} C^{-1} I_{\mathrm{ext}}$, we obtain

$$\begin{aligned} &\dot{V}_{\mathrm{E}}(t) + \frac{1}{\tau_{\mathrm{m}}} V_{\mathrm{E}}(t) = \\ &\quad = -\theta\, S_{\mathrm{E}}(t) + \sum_{i\in\mathcal{E}} \sum_{j\in\mathcal{E}} J_{ij} S_j(t-d) + \sum_{i\in\mathcal{E}} \sum_{j\in\mathcal{I}} J_{ij} S_j(t-d) + X'_{\mathrm{E}} \\ &\quad = -\theta\, S_{\mathrm{E}}(t) + N_{\mathrm{E}}\, \mu_{\mathrm{EE}}\, S_{\mathrm{E}}(t-d) + N_{\mathrm{E}}\, \mu_{\mathrm{EI}}\, S_{\mathrm{I}}(t-d) + X'_{\mathrm{E}} \end{aligned} \tag{4.14}$$

for the excitatory and

$$\dot{V}_{\mathrm{I}}(t) + \frac{1}{\tau_{\mathrm{m}}} V_{\mathrm{I}}(t) = -\theta\, S_{\mathrm{I}}(t) + N_{\mathrm{I}}\, \mu_{\mathrm{IE}}\, S_{\mathrm{E}}(t-d) + N_{\mathrm{I}}\, \mu_{\mathrm{II}}\, S_{\mathrm{I}}(t-d) + X'_{\mathrm{I}} \tag{4.15}$$

for the inhibitory subpopulation. With

$$V(t) = \begin{bmatrix} V_{\mathrm{E}}(t) \\ V_{\mathrm{I}}(t) \end{bmatrix}, \quad S(t) = \begin{bmatrix} S_{\mathrm{E}}(t) \\ S_{\mathrm{I}}(t) \end{bmatrix}, \quad X'(t) = \begin{bmatrix} X'_{\mathrm{E}} \\ X'_{\mathrm{I}} \end{bmatrix}, \tag{4.16}$$

and

$$W' = \begin{bmatrix} N_{\mathrm{E}}\, \mu_{\mathrm{EE}} & N_{\mathrm{E}}\, \mu_{\mathrm{EI}} \\ N_{\mathrm{I}}\, \mu_{\mathrm{IE}} & N_{\mathrm{I}}\, \mu_{\mathrm{II}} \end{bmatrix}, \tag{4.17}$$

equations 4.14 and 4.15 read in matrix notation

$$\dot{V}(t) + \frac{1}{\tau_{\mathrm{m}}} V(t) = -\theta\, S(t) + W'\, S(t-d) + X'. \tag{4.18}$$

Note that these equations are generally not symmetric with respect to E and I. Although we use the same notation in equations 4.18 and 4.9, we refer to a one-dimensional system in the hybrid case and a two-dimensional one in the Dale scenario.

Clearly, equations 4.9 and 4.18 cannot be solved for both the population membrane potential $V(t)$ and the population spike train $S(t)$

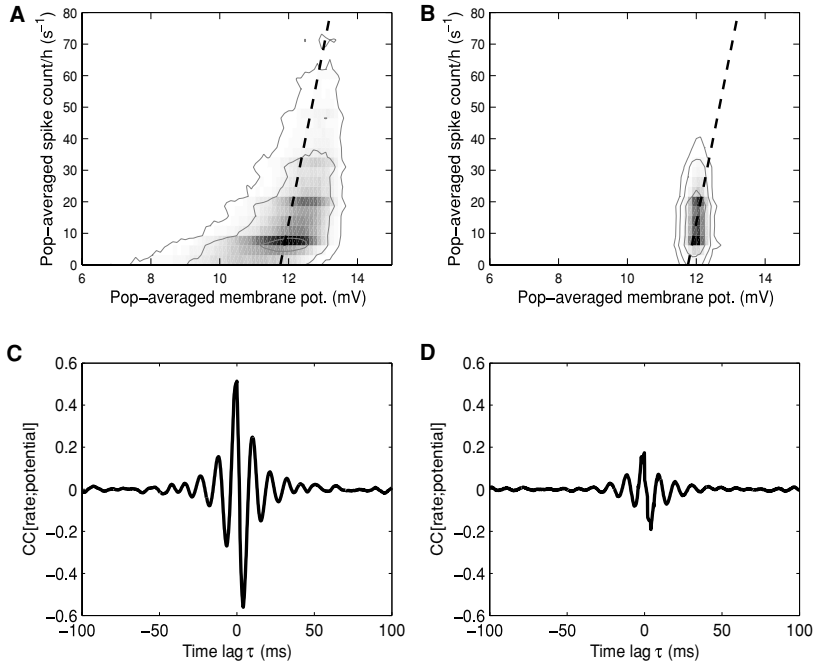

Figure 8: Scatter plots (A, B) and cross-correlation coefficient (C, D) of the population-averaged spike count $S(t)/N$ per time bin $h = 0.1$ ms and the population-averaged membrane potential (including resets) $V(t)/N$ for the Dale and the hybrid scenarios taken from a network simulation of 12,500 neurons. The population size of observed neurons was $M = 4000$ in both cases, with all other parameters as before. In A and B, the relative frequencies are shown in linear (density plots) and logarithmic scaling (contour lines). The black dashed lines represent a linear approximation of the underlying rate intensity function $Y(t)/N$.

simultaneously. By reducing the system to a one- or two-dimensional system, we lose the explicit knowledge about the microstates of the observables $V(t)$ and $S(t)$ during the summation process. Even if we knew the input variable $S(t)$ at all times t, we do not know exactly how it is distributed across neurons, because the detailed connectivity has been averaged out. At the same time, different combinations of membrane potentials $V_i(t)$ in the summation may result in the same sum $V(t)$, while containing $V_i(t) = \theta$ for a number of neurons in one case, or $V_i(t) < \theta$ for all neurons in another case. This will eventually result in different possible values $V(t)$ for the same $S(t)$ (and vice versa). Hence, strictly speaking, we have to deal with the dynamics of probability densities for both $V(t)$ and $S(t)$. This is reflected in the extension of the scatter cloud of $S(t)$ versus $V(t)$ derived from network simulations in Figures 8A and 8B. We observe, however, that

$V(t)$ and $S(t)$ exhibit a close instantaneous relationship in both the Dale and the hybrid case (see Figure 8). We find significant correlation coefficients between $V(t)$ and $S(t)$ of 0.51 in the Dale and 0.13 in the hybrid case (cross-correlation peaks in Figures 8C and 8D, respectively). Moreover, the peaks of the respective cross-correlation functions,

$$\mathcal{R}(\tau) = \mathrm{E}[S(t)V(t+\tau)], \tag{4.19}$$

are centered at time lag zero, indicating an instantaneous relationship (see Figures 8C and 8D). Note that correlations between spike responses and population membrane potentials have also been reported experimentally (e.g., Arieli, Shoham, Hildesheim, & Grinvald, 1995). In the following, we replace the population spike train $S(t)$ by a deterministic signal $Y(t)$—the population firing rate. For simplicity, we model the relation between $Y(t)$ and $V(t)$ (see Figure 8) as

$$Y(t) = \frac{1}{b}[V(t) - a]. \tag{4.20}$$

The parameters $a = 11.8\,\mathrm{mV} \cdot N$ and $b = 18.2\,\mathrm{mVms}$ are roughly estimated from network simulations (see the dashed lines in Figures 8A and 8B). In the Dale case, we find similar estimates for both the excitatory and the inhibitory subpopulation (data not shown). Although a and b slightly differ in the Dale and hybrid scenarios, we use identical values here (the qualitative results of our analysis do not critically depend on the exact choice of these numbers). With equation 4.20 and the abbreviations

$$\tau_\mathrm{r} = \frac{b}{\theta + b\tau_\mathrm{m}^{-1}}, \quad W = \frac{W'}{\theta + b\tau_\mathrm{m}^{-1}}, \quad \text{and} \quad X = \frac{X' - a\,\tau_\mathrm{m}^{-1}}{\theta + b\tau_\mathrm{m}^{-1}}, \tag{4.21}$$

equations 4.9 and 4.18 simplify to

$$\tau_\mathrm{r}\dot{Y} = -Y(t) + WY(t-d) + X. \tag{4.22}$$

Note that the variable $Y(t)$ and the parameters a and X are two-dimensional vectors in the Dale scenario. The relaxation time constant $\tau_\mathrm{r} \approx 0.87\,\mathrm{ms}$ is fully determined by the neuron parameters $\tau_\mathrm{m} = 20\,\mathrm{ms}$ and $\theta = 20\,\mathrm{mV}$ and the slope $b = 18.2\,\mathrm{mVms}$ of the linear relation, equation 4.20.

The time-averaged compound firing rates $\bar{Y} \equiv \nu N$ (hybrid) and $\bar{Y} \equiv \nu(N_\mathrm{E}, N_\mathrm{I})^T$ (Dale) are hence determined by

$$\bar{Y} = (\mathbb{I} - W)^{-1} X. \tag{4.23}$$

For our standard parameter setting and the estimated coefficients a and b, the average single neuron firing rate becomes $\nu \approx 12.83\,\mathrm{Hz}$ in both the Dale and the hybrid scenario, which is in very good agreement with the average firing rates observed in the simulations (cf. Figure 2A, $\nu_{\mathrm{D}} = 12.89\,\mathrm{Hz}$, $\nu_{\mathrm{H}} = 12.83\,\mathrm{Hz}$).

4.3 Power Spectra of the Population Responses. Equation 4.22 describes a deterministic system driven by a constant input X. The recurrent input to the network, however, is mediated by spike trains. Therefore, the system excites itself in all frequency bands. We take this into account by adding a white noise component $R(t)$ with

$$\begin{aligned} \mathrm{E}\,[R(t)] &= 0 \quad \text{and} \\ \mathrm{E}\,[R(t)R(t')] &= \Lambda\delta(t-t'), \end{aligned} \tag{4.24}$$

and obtain a stochastic differential equation (Langevin equation) for the population rate $Y(t)$:

$$\tau_{\mathrm{r}}\dot{Y} = -Y(t) + WY(t-d) + X + R(t). \tag{4.25}$$

In the Fourier domain, equation 4.25 reads

$$-\imath\omega\tau_{\mathrm{r}}\hat{Y}(\omega) = -\hat{Y}(\omega) + \mathrm{e}^{\imath\omega d}W\hat{Y}(\omega) + \hat{X}(\omega) + \hat{R}(\omega). \tag{4.26}$$

If we set $P = [(1-\imath\omega\tau_{\mathrm{r}})\mathbb{I} - \mathrm{e}^{\imath\omega d}W]^{-1}$, we obtain the Fourier-transformed population rates:

$$\hat{Y}(\omega) = P\big[\hat{X}(\omega) + \hat{R}(\omega)\big]. \tag{4.27}$$

For the calculation of the power spectra $\mathrm{E}[\hat{Y}(\omega)\hat{Y}^*(\omega)]$, we ignore the constant external drive X as it contributes only to the zero-frequency component ($\hat{X}(\omega) \sim \delta(\omega)$). From equation 4.24, it follows that $\mathrm{E}[\hat{R}(\omega)\hat{R}^*(\omega)] = \Lambda$. Thus, we obtain

$$\mathrm{E}[\hat{Y}\hat{Y}^*] = \mathrm{E}[(P\hat{R})(P\hat{R})^*] = P\mathrm{E}[\hat{R}\,\hat{R}^*]P^* = P\Lambda P^*. \tag{4.28}$$

To specify the spectral matrix Λ of the noise component $R(t)$, we assume that the compound spiking activity results from the superposition of N uncorrelated Poisson processes with time-averaged firing rates ν. Thus, the noise (co)variances become

$$\Lambda = N\nu \tag{4.29}$$

in the hybrid and

$$\Lambda = \begin{pmatrix} N_{\mathrm{E}}\nu & 0 \\ 0 & N_{\mathrm{I}}\nu \end{pmatrix} \tag{4.30}$$

in the Dale scenario. In the hybrid case, the power spectrum 4.28 of the population rate is therefore given by

$$\mathcal{P}(\omega) := \mathrm{E}\left[\hat{Y}(\omega)\hat{Y}^*(\omega)\right] = \frac{N\nu}{|(1 - \imath\omega\tau_{\mathrm{r}}) - \mathrm{e}^{\imath\omega d}W|^2}, \tag{4.31}$$

whereas we obtain in the Dale scenario

$$\mathcal{P}(\omega) = \begin{pmatrix} \mathcal{P}_{\mathrm{EE}}(\omega) & \mathcal{P}_{\mathrm{EI}}(\omega) \\ \mathcal{P}_{\mathrm{IE}}(\omega) & \mathcal{P}_{\mathrm{II}}(\omega) \end{pmatrix}, \tag{4.32}$$

with the explicit expressions given in appendix C. Finally, to obtain the full spike train spectra, a constant offset $\Lambda = \mathrm{E}[\hat{R}(\omega)\hat{R}^*(\omega)]$ has to be added.

Figures 9A and 9B show the power spectra obtained by numerical simulations of the full system in comparison to the corresponding power spectra obtained from the reduced linear rate model, Figures 9C and 9D, for the same parameters as before. Although we chose identical parameters in the Dale and the hybrid scenario, the power of the population signal $Y(t)$ in the Dale case exceeds that in the hybrid scenario by up to two orders of magnitude at almost all frequencies (see Figure 9C). This is in good agreement with our results obtained from network simulations (see Figure 9A). Figure 9E demonstrates that the overall increase in power in the Dale case does not result from the asymmetry induced by the different sizes of the excitatory and inhibitory neuron population ($\beta = 0.8$) and the dominance of inhibition ($g = 6$). Even for a system with equally sized populations ($\beta = 0.5$), the same absolute strength of inhibition and excitation ($g = 1$), and zero delay $d = 0$, the linear model predicts a massive difference in total power between the Dale (see Figure 9F, gray line) and the hybrid case (black line).

In the Dale case, the total power of the excitatory subpopulation clearly exceeds the power of the inhibitory subpopulation (see Figures 9B and 9D). Even for the case $\beta = 0.5$, $g = 1$, $d = 0$ (see Figure 9F), there is a remaining difference between excitatory and inhibitory subpopulation that is due solely to the asymmetry in the signs of inhibition and excitation.

In addition to the overall differences in power, the presented mean-field models can also qualitatively reproduce the delay resonance peaks observed in the simulations.

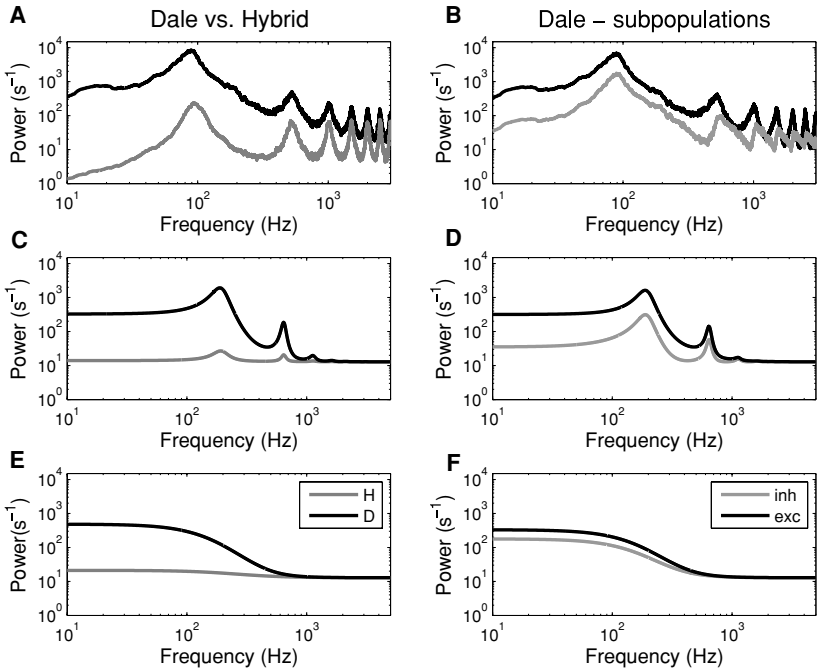

Figure 9: Population size normalized power spectra obtained by simulation of the full nonlinear system (A, B; smoothed by moving average [frame size 1 Hz]) in comparison with the according analytical results for the reduced linear model (C, D: standard parameters; E, F: $d = 0$, $g = 1$, $\beta = 0.5$). The left column (A, C, E) compares the total spectra in Dale (black curves) and hybrid (gray) networks. The right column (B, D, F) shows the power spectra of the excitatory (black) and inhibitory (gray) subpopulations in the Dale scenario.

5 Heterogeneous Networks

It has been shown in previous studies (e.g., Brunel, 2000; Denker, Timme, Diesmann, Wolf, & Geisel, 2004; Tetzlaff et al., 2005) that global network fluctuations are diminished by introducing a reasonable degree of heterogeneity in different parameters, such as spike transmission delay, synaptic strength, and in-degree. One may therefore argue that in realistic heterogeneous networks, the effects of Dale's principle on input correlations and the fluctuations of the population activity play only a minor role and hence can be neglected. In this section, we point out, however, that heterogeneities typically affect correlations only in restricted frequency bands. Dale's principle, in contrast, amplifies correlations on all timescales and in all frequency bands.

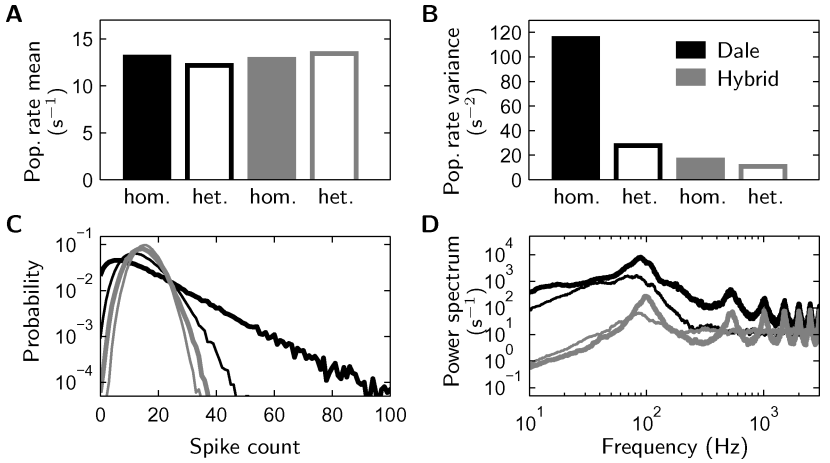

Figure 10: Spike count statistics (bin size 0.1 ms) in Dale (black) and hybrid networks (gray) with homogeneous (solid bars, thick lines) and distributed synaptic delays and in-degrees (open bars, thin lines). (A) Time- and population-averaged ($n = 12{,}500$) population rate. (B) Variance of population rate. (C) Spike count distributions. (D) Power spectra of population activity (smoothed by moving average [frame size 10 Hz]). Graphs show simulation results for networks composed of $N = 12,500$ I&F neurons ($N_E = 10{,}000$ excitatory, $N_I = 2500$ inhibitory neurons, simulation time 10 s, temporal resolution 0.1 ms). In heterogeneous networks, synaptic delays d and in-degrees K_E, K_I are uniformly distributed with $d \in \{0.1, \ldots, 3.9\}$ ms, $K_E \in \{800, \ldots, 1200\}$, and $K_I \in \{200, \ldots, 300\}$.

To illustrate this, we performed network simulations in which synaptic transmission delays $d \in \{0.1, \ldots, 3.9\}$ ms and in-degrees $K_E \in \{800, \ldots, 1200\}$ and $K_I \in \{200, \ldots, 300\}$ were drawn from uniform distributions. According to Figures 10B and 10C, the global network fluctuations are indeed reduced in heterogeneous networks—at least on a submillisecond timescale. The power spectra in Figure 10D reveal that the reduction in the population rate variance (see Figure 10B) is mainly due to a suppression of power at high frequencies (>100 Hz). At lower frequencies, however, networks respecting Dale's principle still exhibit large population activity fluctuations, even if synaptic delays and in-degrees K_E and K_I are randomly distributed. This finding is confirmed by a visual inspection of the spiking activity in Figure 11.

The correlation coefficients between input currents in heterogeneous networks are centered around zero in both the hybrid and the Dale scenario (see Figure 12B). It is not surprising that on short timescales (recall that we used delta-type synaptic currents), correlations induced by

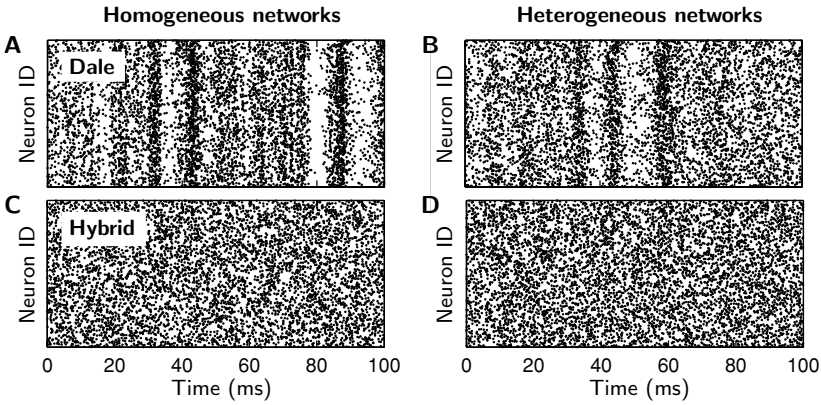

Figure 11: Spiking activity in Dale (top row) and hybrid (bottom row) networks with homogeneous (left column) and heterogeneous synaptic delays and in-degrees (right column; see Figure 10). Dot displays show the activity of 5000 randomly selected neurons.

common input are destroyed if the synaptic transmission delays are distributed. On larger timescales, however, common input correlations remain. In our example simulations, this becomes visible at the level of the membrane potentials (see Figure 12D). Here, the time constants of the cell membranes ($\tau_m = 20\,\text{ms}$) determine the timescale at which correlations are measured. Although the distribution of correlation coefficients in the heterogeneous Dale network is shifted toward smaller values, its mean value is still clearly larger than zero (see Figure 12D). Figures 12E and 12F show the coherences between input currents in the four different networks. In the homogeneous case, correlations are enhanced by Dale's principle in all frequency bands. The heterogeneities in in-degrees and synaptic delays weaken these correlations mostly at high frequencies (above 200 Hz) and also at low frequencies (below 30 Hz). In the intermediate range (around 100 Hz), input correlations are basically unaffected. The attenuation of input correlations at low and high frequencies is reflected in the power spectra of the population activity (see Figure 10D, thin black curve).

We demonstrated that even in heterogeneous networks, the global activity does not exhibit the level of stationarity (i.e., small fluctuations), which is achieved in networks violating Dale's principle. The heterogeneities we studied here (delay and number of inputs per neuron) are effective only in specific frequency bands. By contrast, common input correlations are not restricted to particular timescales or frequency bands. We conclude that the correlations induced by Dale's principle have to be taken seriously and cannot be overcome in a trivial way.

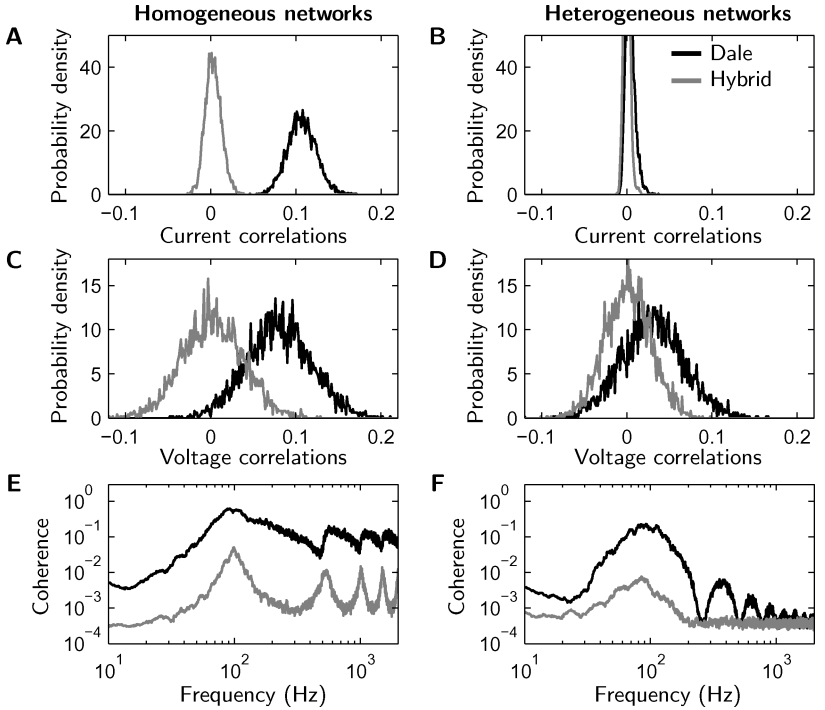

Figure 12: Input correlations at the level of synaptic currents and membrane potentials in Dale (black) and hybrid networks (gray) with homogeneous (left column) and heterogeneous synaptic delays and in-degrees (right column). (A, B) Distributions of correlation coefficients c_{in} between input currents. (C, D) Distributions of correlation coefficients c_{in} between free membrane potentials. (E, F) Coherences between synaptic input currents (smoothed by moving average [frame size 1 Hz]; identical results for membrane potentials). Graphs show simulation results for networks composed of $N = 12{,}500$ I&F neurons ($\epsilon = 0.1$, $\beta = 0.8$, $g = 6$, simulation time 10 s, temporal resolution 0.1 ms). In heterogeneous networks, synaptic delays d and in-degrees K_{E}, K_{I} are uniformly distributed with $d \in \{0.1, \ldots, 3.9\}$ ms, $K_{\text{E}} \in \{800, \ldots, 1200\}$, $K_{\text{I}} \in \{200, \ldots, 300\}$. Synaptic currents and free membrane potentials were recorded from 100 neurons for the measurement of correlation coefficients and from 2000 neurons for the estimation of coherences.

6 Discussion

We showed that the segregation of excitation and inhibition strongly increases the amplitude of fluctuations in the population activity in random recurrent networks of leaky integrate-and-fire neurons. We analyzed the marginal and joint statistics of input currents and found that spike train

correlations in balanced networks are negligible for the marginal and joint input statistics but crucial for the statistics of the population activity. Each pair of neurons in a random network of N neurons with connectivity ϵ shares on average a pool of $N\epsilon^2$ presynaptic neurons. In the Dale case, the resulting common input induces strong correlations, since each spike in one particular neuron simultaneously either hyperpolarizes or depolarizes all postsynaptic partners. In the hybrid case, in contrast, these common inputs partially cancel out because a significant fraction of the shared presynaptic neurons hyperpolarizes one target and at the same time depolarizes another one. Hence, Dale's principle amplifies common input correlations. In consequence, correlations between input currents exceed those in purely random networks (hybrid scenario) by more than one order of magnitude, regardless of the network connectivity ϵ.

We were able to predict output correlations from input correlations in a self-consistent way by assuming a simple linear correlation transmission. The fixed linear gain was measured from simulated data and turned out to be very small ($\sim 10^{-2}$). Still, the small pairwise spike train correlations fully explain the observed strong network fluctuations in our simulations. Both input and output correlations scale linearly with the network connectivity $\epsilon = K/N$. In the biologically relevant parameter range, however, they cannot be ignored for the dynamics of the population activity. Recently, in a study of retinal ganglion cell networks (Schneidman, Berry, Segev, & Bialek, 2006), it was also reported that weak pairwise correlations may lead to strong fluctuations of network activity. We conclude that fluctuations in the population rate cannot be reduced to the Poissonian case even if the network is constantly driven by stationary inputs.

As the classical mean-field approach neglects input correlations, it cannot correctly describe the fluctuations of the population rate. We derived an extended mean-field model from the integrate-and-fire dynamics, which preserves the input correlations caused by Dale's principle. By appropriate summation of the single neuron responses, we obtained simple linear delay-differential equations describing the population rates for both the Dale and the hybrid cases. Solving these equations in the frequency domain leads to power spectra that reproduce the overall features of those of the full spiking dynamics. The simple model is able to account for the observed higher overall power in the Dale case as compared to the hybrid case, as well as for the differences between the inhibitory and excitatory population in the Dale case. Even for equally sized subpopulations, zero delay, and the same absolute strength of inhibitory and excitatory synapses, the segregation of inhibition and excitation causes an increased overall power for the Dale scenario. The difference between subpopulations in the Dale case is primarily due to the asymmetry of the reduced two-dimensional model, that is, the genuine asymmetry between excitatory and inhibitory synaptic interactions (depolarizing versus hyperpolarizing), and is only secondarily caused by the different sizes of the neuronal subpopulations.

Finally, we checked the effect of heterogeneity on network fluctuations. It has been argued (Brunel, 2000; Mattia & Del Guidice, 2004; Denker et al., 2004; Tetzlaff et al., 2005) that introducing heterogeneity in the number of synapses and a broader delay distribution stabilizes the asynchronous-irregular state. Here, we point out that this type of desynchronization occurs only in limited frequency bands. By contrast, the amplification of correlations by Dale's principle is not restricted to any specific timescale. Therefore, the randomization of network and neuron parameters can only partially compensate for the Dale effect. In simulations of spiking networks, we found that distributions of in-degrees and synaptic delays suppress network fluctuations mainly at high, and to some extent at low, frequencies. Intermediate frequencies were unaffected.

A recent experimental study (Ren et al., 2007) demonstrated that couplings between nearby pyramidal cells in the mouse visual cortex can be inhibitory (see also Connors & Cruikshank, 2007). This observation can be interpreted as a violation of Dale's principle and may therefore lead to network models that exhibit much weaker population activity fluctuations. As pointed out in appendix B, common-input correlations depend smoothly on the fraction of neurons or the fraction of synapses that obey Dale's principle (see equation B.5). Therefore, unless Dale's principle turns out to be considerably violated, its amplifying effect on correlated activity and population rate fluctuations has to be taken seriously.

In summary, we conclude that segregated populations of excitatory and inhibitory neurons cause network fluctuations, which were not fully appreciated in previous models of cortical networks. Here, we presented a rather simple shot-noise model and a linear rate model that capture most of the observed effects without obscuring effects by complicated mathematics, underlining the fundamental nature of the residual fluctuations in the AI state. It was outlined before (Hoppensteadt & Izhikevich, 1997; Li & Dayan, 1999) that asymmetric networks of excitatory and inhibitory neurons have different synchronization and computational properties. But so far however, nobody has emphasized or analyzed its most crucial effect on balanced recurrent random networks.

We showed that input correlations in a random network (respecting Dale's principle) are essentially determined by the network connectivity ϵ. According to anatomical data (Abeles, 1991; Braitenberg & Schüz, 1991; Hellwig, 2000), we would therefore expect input correlations in the cortex of about 0.1 (see Figures 6G and 6H, black curves). For the visual cortex of anesthetized cats, Lampl, Reichova, and Ferster (1999) reported a broad range of membrane potential correlations with a mean of about 0.4. Membrane potential correlations of similar quality were reported during slow-wave sleep (Volgushev, Chauvette, Mukovski, & Timofeev, 2006). There are several possible reasons for this deviation from our model. First, there is growing experimental evidence that the topology of local cortical networks differs from a random connectivity (e.g., Song, Per, Reigl, Nelson,

& Chklovskii, 2005). The distributions of in-degrees (number of inputs per neuron) are most likely broader than assumed in our study, and the probability of finding a common presynaptic neuron shared by two (observed) target cells may significantly differ from chance level ϵ^2 in certain neuronal subpopulations (Yoshimura, Dantzker, & Callaway, 2005). Further, spatially correlated or time-dependent external inputs from other cortical or subcortical areas are neglected in our study. Finally, input correlation coefficients are modulated by the filter properties of the synapses and the cell membranes (time constants; see Tetzlaff et al., 2005, and the companion article in this issue: Tetzlaff et al., "Dependence of Neuronal Correlations on Filter Characteristics and Marginal Spike Train Statistics"). In consequence, the distributions of input correlation coefficients can be expected to be broader and shifted toward higher values, as compared to our results.

According to our simulation results, the transmission of input correlations to output spike correlations is very weak, in agreement with several other theoretical studies (Shadlen & Newsome, 1998; Stroeve & Gielen, 2001; Tetzlaff et al., 2003; Moreno-Bote & Parga, 2006; Tetzlaff et al. in this issue). In recurrent random networks with plausible connection probabilities we found spike train correlation coefficients of the order of 10^{-3}. This observation seems to contradict the results of many experimental studies reporting stimulus-unspecific spike train correlation coefficients of 0.1 or higher (e.g., Aertsen & Gerstein, 1985; Vaadia & Aertsen, 1992; Zohary, Shadlen, & Newsome, 1994; Vaadia et al., 1995; Gawne & Richmond, 1993; Shadlen & Newsome, 1998; Bair, Zohary, & Newsome, 2001). In these studies, however, spikes from different neurons were recorded with a single electrode. The reported correlations therefore refer to neurons within a radius of not more than 50 μm (e.g., Henze et al., 2000; Sakurai & Takahashi, 2006). Measured correlations between spike trains recorded on different electrodes are typically weaker (e.g., Ts'o, Gilbert, & Wiesel, 1986; Gochin, Miller, Gross, & Gerstein, 1991; Vaadia & Aertsen, 1992). One must therefore expect that spike correlations averaged over a cortical volume that corresponds to the spatial scale described by our model ($\approx 1\,\mathrm{mm}^3$) are smaller than 0.1. But apart from that, it is likely that the model simplifications mentioned above (random network structure, uncorrelated stationary external inputs) lead to an underestimation of correlations. So far, it is not clear whether the low correlation transmission gain results from an oversimplification of the I&F neuron model or is a realistic feature of cortical neurons. From our study (and those mentioned above), we must conclude that the information about the network structure contained in the correlation between intracellular signals is significantly decreased by the nonlinear spike generation dynamics. To explore the underlying network structure, the study of correlations at the intracellular level (e.g., membrane potentials) appears much more promising than at the level of spikes (see also Tetzlaff et al., this issue).

Apart from the structure imposed by Dale's principle, the networks studied here are purely random. We observed in simulations that depending

on the topology (e.g., locally coupled random networks, small world networks), networks respecting Dale's principle exhibit rich spatiotemporal activity dynamics (e.g., pattern formation). If, however, Dale's principle is discarded, the population activity becomes nearly indistinguishable from that observed in random networks (data not shown here). Thus, studies focusing on the effect of the network topology on dynamics must account for Dale's principle.

Appendix A: Shot-Noise Correlations

Consider the spike train

$$S_i(t) = \sum_k \delta(t - t_{i,k}) \tag{A.1}$$

of neuron i defined as a series of delta pulses at times $t_{i,k}$. Throughout the article, we model spike trains $S_i(t)$ as Poisson processes with average rates $\nu_i = \mathrm{E}_t[S_i(t)]$ and delta-shaped autocovariance functions:

$$\psi_{ii}(\tau) = \mathrm{E}_t[S_i(t)S_i(t+\tau)] - \mathrm{E}_t[S_i(t)]^2 = \nu_i\delta(\tau). \tag{A.2}$$

Correlations between spike trains play a crucial role in this article. Here, we restrict ourselves to the simple case of delta-shaped cross-covariance functions:

$$\psi_{ij}(\tau) = \mathrm{E}_t\left[S_i(t)S_j(t+\tau)\right] - \mathrm{E}_t[S_i(t)]\,\mathrm{E}_t\left[S_j(t)\right] = c_{ij}\sqrt{\nu_i\nu_j}\delta(\tau). \tag{A.3}$$

A linearly filtered version

$$s_i(t) = (S_i * f)(t) \tag{A.4}$$

of $S_i(t)$ with a filter kernel $f(t)$ (and convolution operator $*$) is generally called *shot noise*. Its mean and covariance function are given by (Papoulis & Pillai, 2002)

$$\mathrm{E}_t[s_i(t)] = \nu_i \int_{-\infty}^{\infty} \mathrm{d}t\ f(t) \tag{A.5}$$

and

$$\mathrm{Cov}[s_i(t), s_j(t+\tau)] = (\psi_{ij} * \phi)(\tau). \tag{A.6}$$

Here, $\phi(\tau)$ represents the autocorrelation of the filter kernel $f(t)$:

$$\phi(\tau) = \int_{-\infty}^{\infty} \mathrm{d}t \; f(t) f(t+\tau). \tag{A.7}$$

Given the assumptions above we obtain for the variance and covariance of the shot noise,

$$\begin{aligned} \mathrm{Var}[s_i(t)] &= \nu_i \int_{-\infty}^{\infty} \mathrm{d}t \; f^2(t) \\ \mathrm{Cov}[s_i(t), s_j(t)] &= c_{ij} \sqrt{\nu_i \nu_j} \int_{-\infty}^{\infty} \mathrm{d}t \; f^2(t) \quad (i \neq j). \end{aligned} \tag{A.8}$$

The correlation coefficient between two (linearly filtered) Poissonian spike trains is therefore given by

$$\frac{\mathrm{Cov}[s_i, s_j]}{\sqrt{\mathrm{Var}[s_i]\mathrm{Var}[s_j]}} = c_{ij}. \tag{A.9}$$

Note that it does not depend on the filter kernel $f(t)$.

Appendix B: Common-Input Coefficients

The amplitude of the input covariance function $c_{\mathrm{in}}(\tau)$ of two neurons is determined in a major way by the number of shared presynaptic cells and the potential synapse-pair configurations (see Figures 4A and 4B). In general, we have to distinguish among three different types of common-input populations. Assume the number of purely excitatory sources, that is, sources that are excitatory to both neurons k and l, to be $C_{\mathrm{E}} := |\{i | A_{ki} A_{li} = 1\}|$ and that of purely inhibitory ones (sources inhibiting both k and l) to be $C_{\mathrm{I}} := |\{i | A_{ki} A_{li} = g^2\}|$. The number of presynaptic neurons that are excitatory to k and inhibitory to l (or vice versa) is denoted by $C_{\mathrm{EI}} := |\{i | A_{ki} A_{li} = -g\}|$. The sizes C_{E}, C_{I}, and C_{EI} of the common input pools actually depend on the postsynaptic targets k and l. In the following, we will replace these numbers by their expectation values for an ensemble of network realizations. With a fixed total amount of inputs $K = \epsilon N$, the expected number of shared connections in a random network is given by $C = \epsilon K$. In the Dale scenario, these common sources are either excitatory or inhibitory for both targets. With $\beta = K_{\mathrm{E}}/K$ being the fraction of excitatory synapses, we thus have

$$C_{\mathrm{E}} = \beta C, \quad C_{\mathrm{I}} = (1-\beta) C, \quad C_{\mathrm{EI}} = 0 \quad \text{(Dale)}. \tag{B.1}$$

In hybrid networks the common sources cannot be separated into excitatory and inhibitory pools. A synapse drawn from the common pool of size C becomes excitatory with probability β and inhibitory with probability $1-\beta$. In the hybrid scenario, the expected numbers of purely excitatory, purely inhibitory, and mixed common sources are therefore given by

$$C_{\mathrm{E}} = \beta^2 C, \quad C_{\mathrm{I}} = (1-\beta)^2 C, \quad C_{\mathrm{EI}} = 2\beta(1-\beta)C \quad \text{(Hybrid)}. \tag{B.2}$$

The coefficient $G = \sum_{i=1}^{N} A_{ki} A_{li} = C_{\mathrm{E}} + g^2 C_{\mathrm{I}} - g C_{\mathrm{EI}}$ in equation 3.18 thus becomes

$$G_{\mathrm{D}} = \epsilon H \tag{B.3}$$

in the Dale case and

$$G_{\mathrm{H}} = \epsilon\, L/K \tag{B.4}$$

in the hybrid scenario.

One can interpolate between the two network types by introducing a parameter p_{N} representing the fraction of neurons in the network that follow Dale's principle. Alternatively, one could vary the fraction p_{S} of synapses per neuron that are consistent with Dale's principle. In either case, the Dale and the hybrid scenarios are retrieved for $p_{\mathrm{N/S}} = 1$ or $p_{\mathrm{N/S}} = 0$, respectively. Depending on the interpolation scheme, the common-input coefficient G (and therefore also the input correlation c_{in}) grows either linearly or quadratically with the interpolation parameter:

$$\begin{aligned} G(p_{\mathrm{N}}) &= (G_{\mathrm{D}} - G_{\mathrm{H}})p_{\mathrm{N}} + G_{\mathrm{H}}, \\ G(p_{\mathrm{S}}) &= (G_{\mathrm{D}} - G_{\mathrm{H}})p_{\mathrm{S}}^2 + G_{\mathrm{H}}. \end{aligned} \tag{B.5}$$

In any case, we observe that the transition from the Dale to the hybrid scenario is gradual. Marginal deviations from Dale's principle cause only a small reduction of common-input correlations.

Appendix C: Power Spectra of the Two-Dimensional Reduced Model

For the two-dimensional case and our special parameter selection $\mu_{\mathrm{EE}} = \mu_{\mathrm{IE}} = \mu_{\mathrm{E}}$ and $\mu_{\mathrm{II}} = \mu_{\mathrm{EI}} = \mu_{\mathrm{I}}$, the elements of the matrix P as given by equation 4.28 read:

$$P_{11}(\omega) = \frac{N\tau_{\mathrm{r}}(1-\beta)\mu_{\mathrm{I}} - b\mathrm{e}^{-\imath\omega d}(1-\imath\omega\tau_{\mathrm{r}})}{(1-\imath\omega\tau_{\mathrm{r}})\left[N\tau_{\mathrm{r}}\left(\beta\mu_{\mathrm{E}} + (1-\beta)\mu_{\mathrm{I}}\right) - b\mathrm{e}^{-\imath\omega d}(1-\imath\omega\tau_{\mathrm{r}})\right]}$$

$$P_{12}(\omega) = \frac{-N\,\tau_r\,\beta\,\mu_I}{(1-\imath\omega\tau_r)\,[N\tau_r\left(\beta\,\mu_E+(1-\beta)\mu_I\right)-b\mathrm{e}^{-\imath\omega d}(1-\imath\omega\tau_r)]}$$

$$P_{21}(\omega) = \frac{-N\,\tau_r\,(1-\beta)\,\mu_E}{(1-\imath\omega\tau_r)\,[N\tau_r\left(\beta\,\mu_E+(1-\beta)\mu_I\right)-b\mathrm{e}^{-\imath\omega d}(1-\imath\omega\tau_r)]}$$

$$P_{22}(\omega) = \frac{N\,\tau_r\,\beta\,\mu_E - b\mathrm{e}^{-\imath\omega d}\,(1-\imath\omega\tau_r)}{(1-\imath\omega\tau_r)\,[N\tau_r\left(\beta\,\mu_E+(1-\beta)\mu_I\right)-b\mathrm{e}^{-\imath\omega d}(1-\imath\omega\tau_r)]}.$$

This leads to the corresponding power spectra in the Dale case,

$$\mathcal{P}(\omega) = \hat{Y}(\omega)\hat{Y}^*(\omega) = P\,\Lambda\,P^*,$$

via simple matrix algebra (see equation 4.28).

Appendix D: Notation

A_{ij}	relative strength of synapse $j \to i$
$a_{\mathrm{in}}(\tau)$	input autocovariance function
$\hat{a}_{\mathrm{in}}(\omega)$	input power spectrum
$a_s(\tau)$	spike train autocovariance function
$\hat{a}_s(\omega)$	spike train power spectrum
α_1	area of filter kernel $f(t)$
α_2	area of $f^2(t)$
β	fraction of excitatory neurons/synapses (e.g., $N_E = \beta N$)
C	total number of common (shared) inputs of two neurons
C_E, C_I, C_{EI}	total number of common excitatory or inhibitory synapses of two neurons
$c_{\mathrm{in}}(\tau)$	input (current, voltage) covariance function
$\hat{c}_{\mathrm{in}}(\omega)$	input cross-spectrum
c_{in}	input (current, voltage) correlation coefficient
c_{in}^*	self-consistent input (current, voltage)correlation coefficient
$c_s(\tau)$	spike train covariance function
$\hat{c}_s(\omega)$	spike train cross-spectrum
c_s	spike train correlation coefficient
c_s^*	self-consistent spike train correlation coefficient
c_{out}	output (spike) correlation
D	label referring to "Dale" (e.g., $c_{\mathrm{in,D}}$)
d	synaptic delay
E	label referring to "excitatory" (e.g., N_E)
$\mathcal{E}$	population of excitatory neurons
ϵ	network connectivity
η	$= \beta/(1-\beta)$, ratio between excitatory and inhibitory neurons or synapses

F	Fano factor
$f(t)$	filter kernel
G	coefficient determining the coupling between spike autocorrelation and input cross-correlation
g	relative strength of inhibition
γ_c	input-output correlation gain
H	coefficient determining the coupling between spike autocorrelation and input autocorrelation
H	label referring to hybrid (e.g., $c_{in,H}$)
$I_k(t)$	input of neuron k
I	label referring to inhibitory (e.g., N_I)
$\mathcal{I}$	population of inhibitory neurons
J_{ij}	absolute strength of synapse $j \to i$
K	total number of synapses or neuron (in-degree)
K_E, K_I	number of excitatory or inhibitory synapses per neuron (in-degree)
L	coefficient determining the coupling between spike cross-correlation and input auto- and cross-correlation
M	size of observed neuron population
μ	ensemble averaged coupling strengths (weights)
N	network size
N_E, N_I	number of excitatory or inhibitory neurons
ν	firing rate
ω	angular frequency
p_N	fraction of neurons obeying Dale's principle
p_S	fraction of synapses obeying Dale's principle
$\mathcal{P}(\omega)$	power spectra
Q	ratio between input correlation c_{in} and connectivity ϵ
$R(t)$	white-noise input in mean-field model
$S_i(t)$	spiking activity of neuron i
$s_i(t)$	filtered spiking activity of neuron i
σ^2	variance of currents or voltages
τ_m	membrane time constants
τ_{ref}	absolute refractory period
τ_r	time constant of the rate model
θ	spike threshold
$\mathrm{Var}(x(t))$	variance of x (time average)
$V_i(t)$	membrane potential of neuron i
W	coupling matrix in mean-field model
X	constant input in mean-field model
$Y(t)$	rates (output) in linear model
$z_i(t;h)$	count variable of spike train of neuron i (bin size h)
$Z(t)$	population spike count
$\hat{\cdot}$	Fourier transforms
$.^*$	complex conjugates

Acknowledgments

We thank M. O. Gewaltig for helpful comments on the manuscript. We acknowledge partial support by the German Federal Ministry of Education and Research (BMBF grant 01GQ0420 and BMBF-DIP F1.2), the German-Israeli Foundation for Scientific Research and Development (GIF), and the European Union (EU Grant 15879, FACETS). All network simulations were carried out with the NEST simulation tool (available online at http://www.nest-initiative.org).

References

Abeles, M. (1991). *Corticonics: Neural circuits of the cerebral cortex*. Cambridge: Cambridge University Press.

Aertsen, A., & Gerstein, G. (1985). Evaluation of neuronal connectivity: Sensitivity of crosscorrelation. *Brain Res.*, *340*, 341–354.

Arieli, A., Shoham, D., Hildesheim, R., & Grinvald, A. (1995). Coherent spatiotemporal patterns of ongoing activity revealed by real-time optical imaging coupled with single-unit recording in the cat visual cortex. *J. Neurophysiol.*, *73*(5), 2072–2093.

Bair, W., Zohary, E., & Newsome, W. (2001). Correlated firing in macaque visual area MT: Time scales and relationship to behavior. *J. Neurosci.*, *21*(5), 1676–1697.

Braitenberg, V., & Schüz, A. (1991). *Anatomy of the cortex: Statistics and geometry*. Berlin: Springer-Verlag.

Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.*, *8*(3), 183–208.

Brunel, N., & Hakim, V. (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comput.*, *11*(7), 1621–1671.

Connors, B. W., & Cruikshank, S. J. (2007). Bypassing interneurons: Inhibition in neocortex. *Nat. Neurosci.*, *10*, 808–810.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience*. Cambridge, MA: MIT Press.

Denker, M., Timme, M., Diesmann, M., Wolf, F., & Geisel, T. (2004). Breaking synchrony by heterogeneity in complex networks. *Phys. Rev. Lett.*, *92*(7), 074103–1–074103–4.

Gawne, T. J., & Richmond, B. J. (1993). How independent are the messages carried by adjacent inferior temporal cortical neurons? *J. Neurosci.*, *13*(7), 2758–2771.

Gochin, P. M., Miller, E. K., Gross, C. G., & Gerstein, G. L. (1991). Functional interactions among neurons in inferior temporal cortex of the awake macaque. *Exp. Brain Res.*, *84*(3), 505–516.

Hellwig, B. (2000). A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex. *Biol. Cybern.*, *2*(82), 111–121.

Henze, D., Borhegyi, Z., Csicsvari, J., Mamiya, A., Harris, K., & Buzsaki, G. (2000). Intracellular features predicted by extracellular recordings in the hippocampus in vivo. *J. Neurophysiol.*, *1*(84), 390–400.

Hoppensteadt, F. C., & Izhikevich, E. M. (1997). *Weakly connected neural networks*. Berlin: Springer-Verlag.

Lampl, I., Reichova, I., & Ferster, D. (1999). Synchronous membrane potential fluctuations in neurons of the cat visual cortex. *Neuron*, *22*, 361–374.

Li, Z., & Dayan, P. (1999). Computational differences between asymmetrical and symmetrical networks. *Network: Comput. Neural Systems*, *10*, 59–77.

Mattia, M., & Del Guidice, P. (2004). Finite-size dynamics of inhibitory and excitatory interacting spiking neurons. *Phys. Rev. E*, *70*, 052903.

Moreno-Bote, R., & Parga, N. (2006). Auto- and crosscorrelograms for the spike response of leaky integrate-and-fire neurons with slow synapses. *Phys. Rev. Lett.*, *96*, 028101.

Morrison, A., Mehring, C., Geisel, T., Aertsen, A., & Diesmann, M. (2005). Advancing the boundaries of high connectivity network simulation with distributed computing. *Neural Comput.*, *17*(8), 1776–1801.

NEST Initiative. (2006). *The neural simulation technology*. Available online at http://www.nest-initiative.org.

Papoulis, A., & Pillai, S. U. (2002). *Probability, random variables, and stochastic processes* (4th ed.). New York: McGraw-Hill.

Ren, M., Yoshimura, Y., Takada, N., Horibe, S., & Komatsu, Y. (2007). Specialized inhibitory synaptic actions between nearby neocortical pyramidal neurons. *Science*, *316*, 758–761.

Sakurai, Y., & Takahashi, S. (2006). Dynamic synchrony of firing in the monkey prefrontal cortex during working-memory tasks. *J. Neurosci.*, *6*(40), 10141–10153.

Salinas, E., & Sejnowski, T. J. (2000). Impact of correlated synaptic input on output firing rate and variability in simple neuronal models. *J. Neurosci.*, *20*(16), 6193–6209.

Schneidman, E., Berry, M. J., Segev, R., & Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, *440*, 1007–1012.

Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *J. Neurosci.*, *18*(10), 3870–3896.

Song, S., Per, S., Reigl, M., Nelson, S., & Chklovskii, D. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *Public Library of Science, Biology*, *3*(3), 0507–0519.

Stroeve, S., & Gielen, S. (2001). Correlation between uncoupled conductance-based integrate-and-fire neurons due to common and synchronous presynaptic firing. *Neural Comput.*, *13*(9), 2005–2029.

Tetzlaff, T., Aertsen, A., & Diesmann, M. (2005). Time-scale dependence of inter-neuronal spike correlations. In *Proceedings of the 30th Göttingen Neurobiology Conference*. Available online at http://www.neuroanatomie.uni-goettingen.de/neurobio_archiv/2005/pdf/Proceedings-Goettingen2005.pdf.

Tetzlaff, T., Buschermöhle, M., Geisel, T., & Diesmann, M. (2003). The spread of rate and correlation in stationary cortical networks. *Neurocomputing*, *52–54*, 949–954.

Tetzlaff, T., Morrison, A., Timme, M., & Diesmann, M. (2005). Heterogeneity breaks global synchrony in large networks. In *Proceedings of the 30th Göttingen*

Neurobiology Conference. Available online at http://www.neuroanatomie.uni-goettingen.de/neurobio_archiv/2005/pdf/Proceedings-Goettingen2005.pdf.

Tetzlaff, T., Rotter, S., Stark, E., Abeles, M., Aertsen, A., & Diesmann, M. (2008). Dependence of neuronal correlations on filter characteristics and marginal spike train statistics. *Neural Comput.*, *20*(9), 2133–2184.

Ts'o, D. Y., Gilbert, C. D., & Wiesel, T. N. (1986). Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. *J. Neurosci.*, *6*(4), 1160–1170.

Vaadia, E., & Aertsen, A. (1992). Coding and computation in the cortex: Single-neuron activity and cooperative phenomena. In A. Aertsen & V. Braitenberg (Eds.), *Information processing in the cortex: Experiments and theory* (pp. 81–121). Berlin: Springer-Verlag.

Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H., et al. (1995). Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature*, *373*(6514), 515–518.

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

van Vreeswijk, C., & Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural Comput.*, *10*, 1321–1371.

Volgushev, M., Chauvette, S., Mukovski, M., & Timofeev, I. (2006). Precise long-range synchronization of activity and silence in neocortical neurons during slow-wave sleep. *J. Neurosci.*, *26*(21), 5665–5672.

Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.*, *12*(1), 1–24.

Yoshimura, Y., Dantzker, J., & Callaway, E. (2005). Excitatory cortical neurons form fine-scale functional networks. *Nature*, *433*(24), 868–873.

Zohary, E., Shadlen, M. N., & Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, *370*, 140–143.

Received February 16, 2007; accepted December 19, 2007.