Sparse balance: excitatory-inhibitory networks with small bias currents and broadly distributed synaptic weights

Ramin Khajeh[1*], Francesco Fumarola[2,1], LF Abbott[1]

1 Mortimer B. Zuckerman Mind Brain Behavior Institute, Department of Neuroscience, Columbia University, New York, NY 10027, USA
2 Laboratory for Neural Computation and Adaptation, RIKEN Center for Brain Science, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

* rk2899@columbia.edu

Abstract

Cortical circuits generate excitatory currents that must be cancelled by strong inhibition to assure stability. The resulting excitatory-inhibitory (E-I) balance can generate spontaneous irregular activity but, in standard balanced E-I models, this requires that an extremely strong feedforward bias current be included along with the recurrent excitation and inhibition. The absence of experimental evidence for such large bias currents inspired us to examine an alternative regime that exhibits asynchronous activity without requiring unrealistically large feedforward input. In these networks, irregular spontaneous activity is supported by a continually changing sparse set of neurons. To support this activity, synaptic strengths must be drawn from high-variance distributions. Unlike standard balanced networks, these sparse balance networks exhibit robust nonlinear responses to uniform inputs and non-Gaussian statistics. In addition to simulations, we present a mean-field analysis to illustrate the properties of these networks.

Introduction

A typical cortical pyramidal cell receives thousands of excitatory inputs [1] that, without the influence of inhibition, would drive extremely high firing rates. It has been suggested that the inhibition that moderates these rates sets up a balanced condition that causes neurons to operate in a regime where fluctuations, not the mean, of their inputs drive spiking, resulting in irregular sequences of action potentials [2–5]. A number of theoretical models have been developed to address E-I balance and the irregular firing of cortical neurons (see [6] for a review). In one class of balanced models [7, 8], the input to each neuron has three strong components – recurrent excitation, recurrent inhibition and feedforward excitation. These balance automatically as part of the network dynamics, leaving residual fluctuations that drive neuronal firing at reasonable rates. Although the presence of strong excitation and inhibition is compatible with the data, there is no evidence for the strong feedforward inputs required in these models [9], and some evidence against them [10–13]. For this reason, we examine the consequences of removing strong feedforward input in balanced models.

In balanced models, synaptic strengths are drawn independently from two probability distributions, one for excitation and another for inhibition. For standard recurrent models to generate spontaneous irregular (chaotic) activity, the synaptic weight distributions must have a variance of order $1/K$, where K is the in-degree, i.e., the number of synapses per neuron [8, 14–16]. The excitatory (inhibitory) distributions are only non-zero for non-negative (non-positive) values, and typically their mean is of the same order as the square-root of their variance, with both being of order $1/\sqrt{K}$. Summing this mean over the K synapses to each neuron gives a total input, which for reasons of stability is inhibitory, of order $\sqrt{K}$. This large mean input must be cancelled and, in conventional models, this is done by adding constant feedforward excitation that is also of order $\sqrt{K}$. This is the large feedforward input that we aim to avoid.

A first question to ask is what happens if the order $\sqrt{K}$ input is simply left out of the standard models, and replaced by an input of order 1. This results in a constraint on the firing rates; specifically, the average firing rate in the network must be of order $1/\sqrt{K}$. This implies that, although there can be irregular spontaneous activity without strong feedforward input, it involves neurons firing at very low rates. One way around this problem is to note that a small average firing rate is not incompatible with having individual neurons with significant firing rates if the activity is sparse. In other words, the average rate can be of order $1/\sqrt{K}$ if, as in the standard models, activity is dense and individual rates are of order $1/\sqrt{K}$, or if individual rates are of order 1 and the fraction of active neurons is of order $1/\sqrt{K}$. Here, we explore this latter possibility.

We mentioned above that standard balanced models require synaptic distributions with variance of order $1/K$ to generate irregular spontaneous activity. More precisely, the requirement is that, for each neuron, the sum of the variances of the strengths of its active inputs must be of order 1. In the standard model, this is satisfied because the product of K, the order of magnitude of the number of active inputs, and $1/K$, the variance per synapse, is 1. In the sparse models proposed in the previous paragraph, the number of active inputs is only of order $\sqrt{K}$, so the total variance computed in this way would be $\sqrt{K}/K = 1/\sqrt{K}$, which is not sufficient to generate irregular activity. To solve this problem, we consider distributions of synaptic strength with means of order $1/\sqrt{K}$, as in the standard model, but with much larger variances of order $1/\sqrt{K}$. In this case, the total variance is of order $\sqrt{K}/\sqrt{K} = 1$, and irregular activity is restored.

Another feature of standard balanced models that seems at odds with the data is that they have attenuated linear responses to input that is uniform across neurons [9, 17]. This linearity is not present in the networks we study. In summary, the combination of small feedforward inputs and broadly distributed synaptic strengths gives rise to a novel E-I regime that exhibits asynchronous irregular sparse firing. In the following, we illustrate prominent features of this regime, such as nonlinear response to feedforward input and non-Gaussian current distributions, and we highlight the mechanisms that maintain sparsity and distributed firing across network neurons.

Results

The model

A common simplification for analyzing E-I networks is to consider a single population of inhibitory units driven by excitatory input from an external source [15, 16]. After analyzing such purely inhibitory networks, we will show that our results apply to networks with both excitatory and inhibitory units. We consider standard 'rate' models. The inhibitory networks we study have currents x_i for $i = 1, 2, \ldots, N$ and firing rates $\phi(x_i)$ that obey

$$\tau_x \frac{dx_i}{dt} = -x_i - \sum_{j=1}^{N} J_{ij}\phi(x_j) + I_0 \,, \tag{1}$$

where ϕ is a nonlinear function and $J_{ij} \geq 0$. We call the variable x a current because it represents the total current generated by the recurrent synaptic and feedforward inputs in the second and third terms on the right side of the above equation, and because it determines the firing rate through the 'F-I' function $\phi(x)$. We also use the terms 'response' or 'firing rate' for $\phi(x)$ and 'activity' for non-zero rates. In our plots, we measure time in units of τ_x, making it a dimensionless variable. Connectivity can be all to all $(K = N)$ or we can restrict the connectivity so that only $K < N$ of the elements in each row of J are non-zero. I_0 is a positive bias input that is identical for all units; it is the feedforward input discussed in the Introduction. Standard balanced models assume the unrealistically large scaling $I_0 \sim \sqrt{K}$; we consider, instead, models with I_0 of order 1.

The non-zero elements of J are drawn independently from a distribution with mean $J_0/\sqrt{K}$, with J_0 an order 1 parameter. We express the variance of this distribution as g^2/K^ν, where g is another parameter of order 1, and ν allows us to vary the scaling with K. The standard scaling is $\nu = 1$, which we call low variance. As we will show, the novel sparse balance regime we explore comes about from setting $\nu = 1/2$, which we call high variance.

It is awkward to use clipped Gaussians for sign-constrained synapses, especially in the large-variance case we consider. We use, instead, distributions with positive support, such as lognormal and gamma, focusing particularly on gamma-distributed synapses for reasons given below. This specific choice is not essential; network behavior remains qualitatively the same across a range of weight distributions, including a binary distribution (Fig S1).

We begin (Fig 1) by setting the response nonlinearity ϕ to a rectified hyperbolic tangent,

$$\phi(x) = \begin{cases} \tanh(x) & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{2}$$

but later we also consider

$$\phi(x) = \begin{cases} x^\lambda & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{3}$$

focusing, in particular, on the case $\lambda = 0$ (Heaviside function) for the analysis, but we also consider $\lambda = 1$ (rectified linear), and $\lambda = 2$ (rectified quadratic).

Throughout, $[\cdot]$ denotes averages over units, $\langle\cdot\rangle$ denotes averages over time, and an overline represents averages across both units and time. For fixed K, the results we present are independent of network size N, provided that the networks are large enough. For this reason and because we are interested in large K, we restrict our studies to the case $K = N$, but the results reported extend to partially-connected networks as well ($K < N$; Fig S2).

Simulation results

With the usual $I_0 \sim \sqrt{K}$ bias reduced to an input of order 1, the network behaves very differently in the low- ($\nu = 1$) and high- ($\nu = 1/2$) variance cases (Fig 1). For low variance, many units are active, but their responses are small (Fig 1A). In contrast, for high variance, activity in the network is sparse but individual units exhibit robust responses (Fig 1B). Scaling of the firing rate as a function of connectivity K can be quantified by computing

$$\overline{\phi} = [\langle\phi\rangle] = \frac{1}{N}\sum_{j=1}^{N}\langle\phi_j\rangle\,. \tag{4}$$

We can break down this average by writing it as the product of f, the fraction of units that are active $(\phi > 0)$, and μ, the average firing rate of the active units, $\overline{\phi} = f\mu$. In both the low- and high-variance cases, the average firing rate $\overline{\phi}$ scales as $1/\sqrt{K}$ (Fig 1C) but, for low variance, f is fairly independent of K (Fig 1D) and μ scales as $1/\sqrt{K}$ (Fig 1E). The scaling is different for high variance where f scales closer to $1/\sqrt{K}$ (Fig 1D) and μ is relatively independent of K (Fig 1E). Thus, the high-variance case, which we call sparse balance, results in networks in which activity is sparse but individual units have appreciable responses.

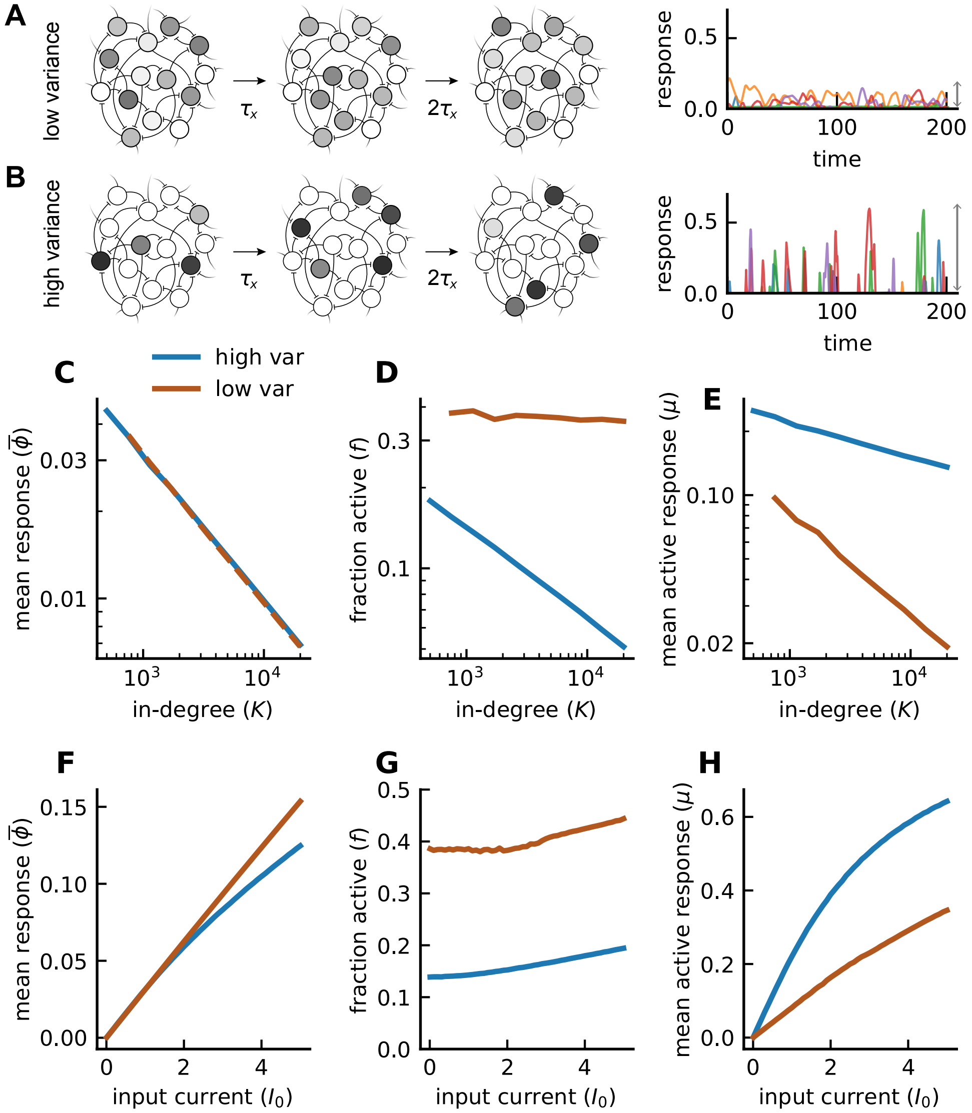

Fig 1. Comparison of low- and high-variance networks. A) Cartoon of network dynamics in time. Light (dark) gray corresponds to low (high) firing rates. With low synaptic variance, fluctuations in firing rates are small, and a relatively fixed and dense subset of units contribute to firing. Right: firing rate traces of five example units (each with a distinct color). Gray arrow indicates the extent of fluctuations in the network. **B)** Same as **A** except for the high-variance model. The network exhibits a small and shifting ensemble of cells that respond robustly at any given time. The magnitude of fluctuations is increased substantially (right). **C)** Mean response in both networks follows a $1/\sqrt{K}$ scaling (fits to the data yield $\overline{\phi} \sim 1/K^{0.503}$ for high variance and $\sim 1/K^{0.513}$ for low variance; J_0 is adjusted so that $\overline{\phi}$ values in the two networks overlaps). **D)** Fraction of active units (inverse sparsity). High-variance model exhibits a rapid sparsening in K while, in the low-variance network, this fraction remains roughly constant. **E)** Mean response of the active subset. The trend in **D** is flipped: the low-variance network demonstrates a rapidly vanishing μ, which is not the case in the high-variance model. Input current I_0 is set to one. **F)** Network's response to the external input current I_0 with $K = 1000$. ***(continued on next page)***

Fig 1. G) Despite similar $\overline{\phi}$ values, the high-variance network is more sparsely active by more than twofold. **H)** Active neurons respond more robustly in the high-variance network than in its low-variance counterpart. (Model parameters: $J_0 = 2$ for high variance and 1.05 for low variance, $g = 2$, $J_{ij} \sim$ gamma, $\phi = [\tanh]_+$).

A well-known distinctive feature of standard balanced networks ($I_0 \sim \sqrt{K}$ and J-variance $\sim 1/K$) is that the average firing rate $\overline{\phi}$ is a linear function of the bias input I_0 despite the presence of a nonlinear response function in the model. This feature extends to the low bias model ($I_0 \sim 1$) in the case of low variance but, for high J-variance ($\sim 1/\sqrt{K}$), the average response $\overline{\phi}$ has a nonlinear dependence on I_0 (Fig 1F). In both the low- and high-variance cases, f is insensitive to I_0 (Fig 1G), meaning that the mean firing rate of the active units μ is also linear for low variance and nonlinear for high variance (Fig 1H). Thus, the restriction to linear responses for uniform input does not apply to the sparse balance networks.

We also examined the distribution of x values in these networks (Fig 2). In the low-variance case, these distributions are Gaussian, and both their mean and variance decrease with K (Fig 2A). The result of these two effects is that the fraction of the x distribution above threshold ($x = 0$) remains fairly constant as a function of K, corresponding to the roughly constant fraction of active units (Fig 1D), but the range of the distribution above threshold drops with K, matching the drop in activity (Fig 1E). For high variance (Fig 2B), the distribution is non-Gaussian, the fraction above threshold drops with K, and the range remains constant, again corresponding to the dependence of the average firing-rate response on K (Fig 1D & E). The mean of the x distribution for the sparse balance network is insensitive to K and lies below threshold. The mean of the distribution for low variance is also negative, but it moves toward zero as K increases.

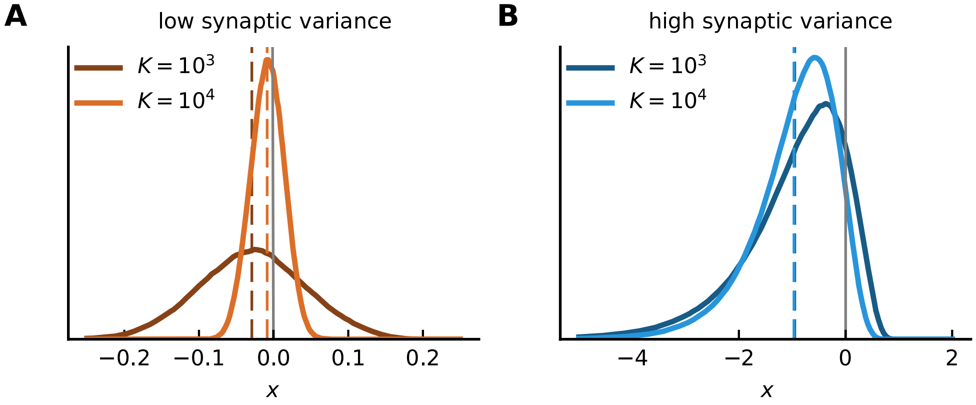

Fig 2. Sparse balance yields non-Gaussian dynamics and a subthreshold mean. Distribution of currents x (over time and units) for gamma-distributed synapses. Dashed lines denote the mean of each distribution, i.e., $\overline{x}$. Area above threshold (set to zero; solid line) corresponds to the fraction of active units f. **A)** With low synaptic variance ($\nu = 1$), the distribution of x is a Gaussian centered around a mean that tends to zero for larger K. **B)** Same as in A except for high synaptic variance ($\nu = 1/2$). Note the larger range of the horizontal axis compared to B. The distribution is no longer Gaussian. $\overline{x}$ is relatively insensitive to K and lies below threshold. (Model parameters are $g = J_0 = 2$, $I_0 = 1$, $J_{ij} \sim \text{gamma}$, $\phi = [\tanh]_+$)

The results for networks with small input biases and large synaptic-weight variances, shown in Fig 1 for a rectified hyperbolic tangent nonlinearity, extend to other nonlinear response functions as well (Fig 3A-B). The response in these networks is distributed across almost all of the units, but at any given time only a sparse distinct subset of units is active (Fig 3C). This active population constantly changes, and firing rates appear chaotic. The fraction of time that units are active is skewed toward small values (Fig 3D), indicating that the majority of units respond infrequently. For all choices of ϕ, the distribution of x is non-Gaussian with only a small fraction of units above threshold (Fig 3E), consistent with the sparsity of the firing. Finally, the dynamics in these networks can be characterized by the population-averaged autocorrelation function of the currents (Fig 3F), which we consider in more detail in a later section.

In summary, these simulations illustrate an alternative regime for E-I networks in which the activity of individual units remains robust, despite the absence of order $\sqrt{K}$ feedforward bias inputs. Furthermore, in these networks, mean firing rate exhibits a nonlinear dependence on bias input. We now analyze in detail the features illustrated in these simulations.

Analysis of sparse balance networks

How does high-variance connectivity support sparse but robust activity with low bias input, and what is the nature of this activity? Addressing these questions is simplified by considering a Heaviside response function (Eq (3) with $\lambda = 0$; we comment on extensions to other nonlinearities in the Materials & Methods). For a Heaviside nonlinearity, the firing rate of an active unit is always one, so $\mu = 1$ and the average response is equal to the sparsity, $\overline{\phi} = f$. We consider a general J-variance scaling, $1/K^{\nu}$, so that we can compare results to the low-variance $\nu = 1$ case, but we are primarily interested in the high-variance case $\nu = 1/2$.

We begin the analysis by defining the recurrent synaptic input as

$$\eta_i(t) = \sum_{j=1}^{N} J_{ij}\ \phi(x_j)\,, \tag{5}$$

so that Eq (1) can be written as

$$\tau_x \frac{dx_i}{dt} = -x_i(t) - \eta_i(t) + I_0\,. \tag{6}$$

We consider non-zero weights drawn from a gamma distribution, gamma(κ, θ), where κ and θ are the 'shape' and 'scale' parameters of the gamma distribution in terms of which its mean is $\kappa\theta$, and its variance is $\kappa\theta^2$. To achieve a mean $J_0/\sqrt{K}$ and variance g^2/K^{ν} we set

$$\kappa = \frac{J_0^2}{g^2} K^{\nu-1}, \qquad \theta = \frac{g^2}{J_0} K^{1/2-\nu}\,. \tag{7}$$

For a Heaviside nonlinearity, the sum in Eq (5) is only over active units with $\phi = 1$, and the probability of a unit being active is equal to the sparsity f. This means that of the K non-zero elements of J for each unit, fK will be active. As a result, η is given by the sum of fK random variables drawn independently from the distribution gamma(κ, θ). The sum of random variables that are gamma-distributed with the same scale parameter is itself gamma-distributed with that scale parameter and a shape parameter equal to the sum of the shape parameters of the variables being summed [18]. Thus,

$$\eta \sim \text{gamma}(\alpha, \theta) \tag{8}$$

with $\alpha = fK\kappa$, which has mean

$$[\eta] = \alpha\theta = fK\kappa\theta = J_0 f\sqrt{K} \tag{9}$$

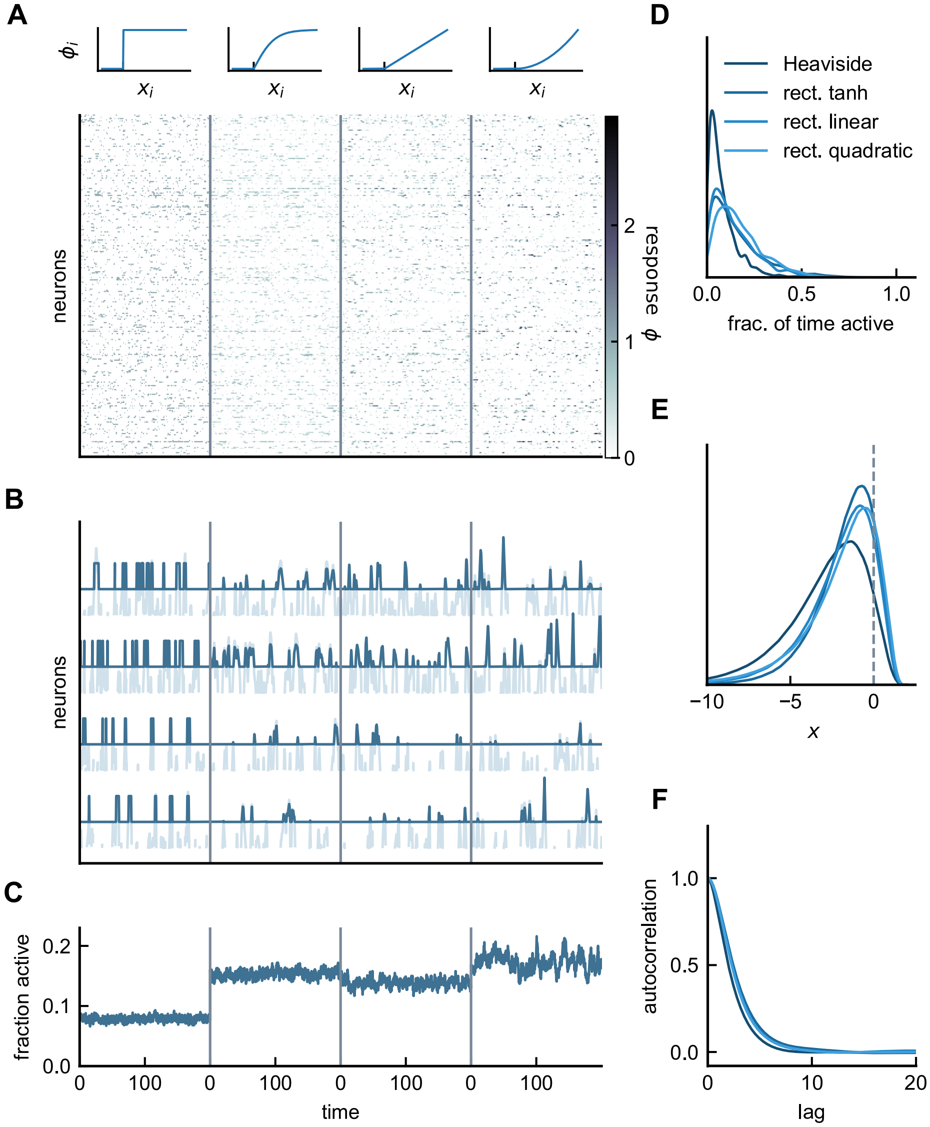

Fig 3. Asynchronous irregular activity in the sparse balance model. A) Responses of network neurons in time for four different nonlinear response functions: Heaviside step function, rectified $\tanh$, rectified linear, and rectified quadratic. **B)** Rates $\phi(x)$ (dark) superimposed on the currents x (light) for four example units. Cells respond robustly and infrequently across choices of the response functions. **C)** Fractions of active neurons, or the inverse sparsity. **D)** Normalized distributions for the fraction of ON-time, defined as the fraction of (simulation) time a unit spends above threshold. For better visualization, histograms are smoothened using kernel density estimation. **E)** Normalized distributions of x, showing non-Gaussian dynamics. **F)** Population-averaged autocorrelation functions of x. At this fixed value of in-degree ($K = 1000$), all response functions produce qualitatively similar results. (Model parameters: $g = J_0 = I_0 = 2$, $J_{ij} \sim$ gamma).

and variance

$$\text{var}(\eta) = \alpha\theta^2 = fK\kappa\theta^2 = g^2 fK^{1-\nu}\,. \tag{10}$$

To maintain a finite mean input as K grows, Eq (9) implies that we must have $f \sim 1/\sqrt{K}$, which implies, from Eq (10), that the fluctuations in the synaptic input scale as $K^{1/2-\nu}$. Thus, the only solution with finite fluctuations as K grows is the high-variance case, $\nu = 1/2$. For $\nu = 1/2$ and with $f \sim 1/\sqrt{K}$, Eqs (9) and (10) show that the distribution of synaptic inputs is independent of K with var$(\eta) = g^2 f\sqrt{K}$. This feature may appear surprising given that the sparseness of network activity is proportional to $1/\sqrt{K}$. We resolve this paradox in a later section.

A naive application of the central limit theorem would suggest that for sufficiently large K, the synaptic input would be normally distributed. Independent of θ, the larger the shape parameter, the closer a gamma distribution approximates a Gaussian (in particular, the approximation is good for shape parameters $\sim$20 or larger). The shape parameter for the distribution of η, from Eq (8), is $fK\kappa = fJ_0^2 g^{-2} K^{-\nu} \sim K^0$. Thus, unless J_0 is large or g is small, even in the limit of large K, the η distribution remains non-Gaussian (Fig S4).

If we average Eq (6) over both units and time and use $\overline{\eta} = J_0 f\sqrt{K}$, we obtain $\overline{x} = J_0 f\sqrt{K} + I_0$ or, equivalently,

$$f = \overline{\phi} = \frac{I_0 - \overline{x}}{\sqrt{K} J_0} \, . \tag{11}$$

In the standard balanced model and in the low-variance case considered above, $I_0 \gg \overline{x}$, so the mean response is linear in I_0. This is no longer true for high synaptic variance ($\nu = 1/2$) for which I_0 and $\overline{x}$ are both of order 1. The nonlinear mean response seen in Fig 1F arises because the dependence of $\overline{x}$ on I_0 is nonlinear. Depending on the choice of ϕ, the sparse balance model exhibits sublinear or subralinear mean population response (Fig S5).

These analyses show that when feedforward bias input is of order one, large synaptic variance is required to generate robust fluctuations, with a synaptic variance of order $1/\sqrt{K}$ producing order-one fluctuations.

Sparse activity arises from network dynamics

We noted in the previous section that the distribution of synaptic inputs is independent of K, and yet the mean network firing rate $\overline{\phi}$ varies as $1/\sqrt{K}$. Network currents x are generated through Eq (6), which involves low-pass filtering of the synaptic input. This suggests that the response sparseness is related not to the distribution of synaptic inputs but rather to their dynamics.

To explore these dynamics, we consider the population-averaged autocorrelation function of η,

$$R_\eta(\tau) = [\langle (\eta_i(t) - \langle \eta_i \rangle)(\eta_i(t+\tau) - \langle \eta_i \rangle) \rangle] \tag{12}$$

which captures the extent to which η at time $t+\tau$ is affected by η at time t. R_η is a decaying function of the lag τ (Fig 4A), and its decay defines a correlation time-scale denoted by τ_η. One way to define this correlation time-constant is by considering the normalized area underneath the autocorrelation function,

$$\tau_\eta = \frac{1}{R_\eta(0)} \int_0^\infty d\tau R_\eta(\tau) \tag{13}$$

We characterize the dynamics of η using the dimensionless constant $\beta = \tau_x/\tau_\eta$, and find that β increases logarithmically with K, an increase that does not occur in the low-variance (Fig 4B) or for conventional balanced networks. As K increases, the time-scale of the fluctuations in η becomes more rapid, although their variance remains constant. This makes it increasingly harder for x to keep up with the fluctuations, due to the low-pass filtering in Eq (6). As a result, the fraction of x above threshold decreases and the overall activity decreases with K. Thus, interestingly, it is the dynamics of the recurrent synaptic inputs, not their size, that leads to sparse activity at large K.

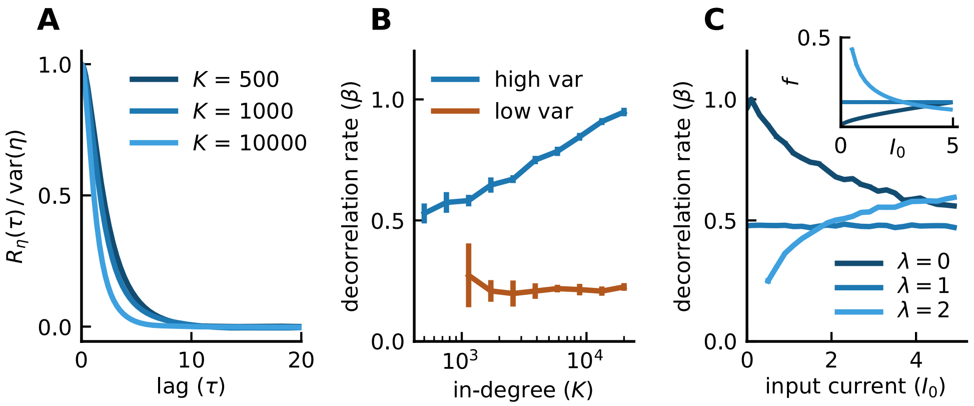

Fig 4. Time-scale of fluctuations adjusts to maintain sparse activity. A) Population-averaged autocorrelation function of the synaptic input normalized by its zero-lag value. Note the faster decay of the autocorrelation for increasing K. **B)** The decorrelation rate β is constant in the low-variance network but increases logarithmically with K in the sparse balance model, resembling the (inverted) trends of sparsity (Fig 1D). **C)** β also exhibits a nonlinear dependence on I_0 similar, but opposite to, that of the sparsity (inset). Error bars indicate the 95% confidence interval around the mean, averaged over 10 random realizations of the connectivity. (Model parameters: $g = J_0 = 2$, $I_0 = 1$, $K = 1000$, $J_{ij} \sim$ gamma, $\phi = [x]_+^\lambda$ for λ sweeps in **C**, $[\tanh]_+$ otherwise.)

Consistent with the argument above that links the time-scale of dynamics to the level of currents above threshold, we find that changes in β account for the degree of sparsity (Fig 4C). In particular, trends in β are opposite to those in f, with sparser activity (smaller f) corresponding to faster time-scale (larger β) and vice versa. These results highlights how the time-scale of synaptic fluctuations dynamically adjust to maintain sparse activity.

Mean-field analysis

In a previous section we noted that, for a Heaviside response nonlinearity and gamma-distributed synaptic weights, the recurrent synaptic input η is gamma distributed with shape parameter (for the high-variance $\nu = 1/2$ case) $\alpha = f J_0^2 \sqrt{K}/g^2$ and scale parameter $\theta = g^2/J_0$. Although θ is completely determined by parameters characterizing J (g and J_0), α is not determined because it depends on the fraction of active units f or, equivalently, on the mean firing rate $\overline{\phi}$. We begin our mean-field analysis by deriving a self-consistent equation for the mean response that determines α and, thereby, the full distribution of recurrent synaptic inputs. For this purpose, we introduce the variable m, which is the mean-field approximation for $\overline{\phi}$. The above equations imply that α is given in terms of m by

$$\alpha = \frac{m J_0^2 \sqrt{K}}{g^2} \,. \tag{14}$$

Our goal is therefore to compute m as a function of α so that the above equation becomes a closed self-consistent condition for determining α.

Conventionally, in a dynamic mean-field approach, the full autocorrelation of η is computed self-consistently [15, 16, 19, 20]. This computation is difficult in the high-variance case because of the non-Gaussian statistics of η. Instead, we consider a 'static' mean-field approximation motivated by the logarithmic scaling of the decorrelation rate β in the sparse balance model. When β is small, x roughly tracks the slow fluctuations in η. This tracking holds particularly well for $\beta \ll 1$. On the contrary, for $\beta \gg 1$, x cannot keep up with the fluctuations in η and averages them. As Fig 4B suggests, for in-degrees of interest, $K \sim 10^3$, the network operates closer to the tracking regime, allowing us to make the approximation $x \approx -\eta + I_0$. Indeed, for $K \sim 500$–1000, the distribution of x resembles that of η, except for a rightward shift by I_0 (Fig 5A). Note that this approach differs from mean-field approaches in which the limit $K \to \infty$ is assumed. Here K is considered large but finite. In both cases, however, the limit $N \to \infty$ is assumed.

When x tracks η, we can use the distribution of η values, $\eta \sim \text{gamma}(\alpha, \theta)$, to perform averages over x. For example, for the mean-field calculation of $[\langle\phi(x)\rangle]$, we can make the substitution $\phi(x) \to \phi(I_0 - \eta)$ and write

$$m = \int_0^\infty d\eta \, p_\gamma(\eta; \alpha, \theta)\phi(I_0 - \eta) \,, \tag{15}$$

where p_γ is the probability density function of the gamma distribution with shape and scale parameters α and θ. Eqs (14) and (15) together form a closed self-consistent condition that determines α, with results that are in decent agreement with numerical simulations (Fig 5B). In particular, the nonlinear relationship between α and I_0 is captured by the theory. Larger values of K exhibit slightly larger deviations from the theory, which hints at the violation of the static assumption. However, even with $K \sim 10^3$, which is in the range of interest, the theoretical α is close to the empirical results (Fig 5B).

Although we have computed α and thereby determined the distribution of η values, this does not completely characterize the nature of the fluctuations in the recurrent synaptic input. The total variance of the recurrent synaptic input can be divided into temporal and quenched parts by writing

$$\text{var}(\eta) = \left[\left\langle (\eta - [\langle\eta\rangle])^2 \right\rangle\right] = \sigma_T^2 + \sigma_Q^2 \,, \tag{16}$$

where

$$\sigma_T^2 = \left[\left\langle (\eta - \langle\eta\rangle)^2 \right\rangle\right] \tag{17}$$

and

$$\sigma_Q^2 = \left[(\langle\eta\rangle - [\langle\eta\rangle])^2\right] \,. \tag{18}$$

The quenched variance arises because the different units of the network fluctuate around different time-averaged values. To analyze this quenched variance, we need to specify how the total variance is divided into temporal and quenched components. For this purpose, we decompose η as

$$\underbrace{\eta}_{\sim\text{gamma}(\alpha,\,\theta)} = \underbrace{\eta_T}_{\sim\text{gamma}(\alpha-b,\,\theta)} + \underbrace{\eta_Q}_{\sim\text{gamma}(b,\,\theta)} \tag{19}$$

where η_T is a time-dependent variable with no quenched variance, and η_Q is a static variable that embodies the influence of quenched disorder. The shape parameters of the temporal and quenched components must add up to α, and the variance of the quenched component determines b through

$$\sigma_Q^2 = b\theta^2 \,, \tag{20}$$

where b is the scaled quenched variance. This decomposition assumes that the time-averaged synaptic input in the full model is gamma-distributed (Fig S6).

Our mean-field analysis of the quenched variance, b, is based on computing a mean-field approximation, s, of $[\langle\phi\rangle^2]$. Using the decomposition into temporal and quenched components of the gamma distribution, we can write the mean-field approximation of s as

$$s = \int_0^\infty d\eta_Q \, p_\gamma(\eta_Q; b, \theta)\left(\int_0^\infty d\eta_T \, p_\gamma(\eta_T; \alpha - b, \theta)\, \phi\big(I_0 - \eta_T - \eta_Q\big)\right)^2 . \tag{21}$$

This expression depends on the parameter α computed above and, in addition, on b, so we need a self-consistency condition to determine the value of this second parameter. Using standard mean-field approximations, we can write

$$[\langle\eta\rangle^2] = \left[\sum_j \sum_k J_{ij} J_{ij} \langle\phi_j\rangle\langle\phi_k\rangle\right] \approx K\left[J^2\right][\langle\phi\rangle\langle\phi\rangle] + K^2 [J]^2 [\langle\phi\rangle]^2 \qquad (22)$$

Inserting now $[J^2] \approx g^2/\sqrt{K}$ and $[J]^2 = J_0^2/K$ gives

$$\left[\langle\eta\rangle^2\right] \approx g^2\sqrt{K}\,[\langle\phi\rangle\langle\phi\rangle] + J_0^2 K\,[\langle\phi\rangle]^2 = b\theta^2 + [\langle\eta\rangle]^2\ . \qquad (23)$$

Finally, using $[\langle\eta\rangle]^2 = J_0^2 K m^2$ and $[\langle\phi\rangle] = m$, this yields

$$s = \frac{b\theta^2}{g^2\sqrt{K}}\ . \qquad (24)$$

Eqs (21) and (24) form a closed system that can be used to determine b (Materials & Methods). Fig 5C depicts this self-consistent solution. Notably, the theory captures the nonlinear relationship between b and I_0 as well as the trend with K. Since the mean-field solution to α is used in computing b, any error in the estimate for α is carried over to the solution for b. The primary source of error is the violation of the assumption that the distribution of time-averages are gamma-distributed, which is used in the decomposition of η in Eq (19). The deviation is pronounced for larger bias inputs but for $I_0 \sim 1$, which is closer to the range of input currents of interest, the theory is in decent agreement.

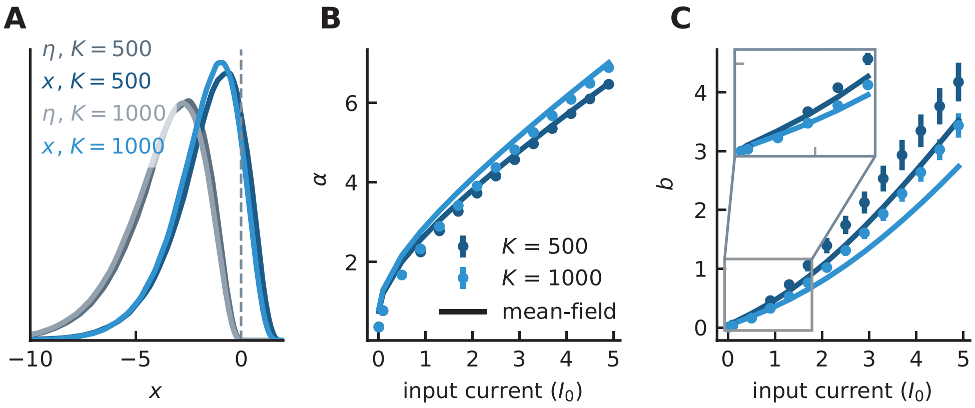

Fig 5. Mean-field approximation captures the mean and quenched variance of the synaptic input. A) Distribution of the recurrent synaptic input η (gray) and current x (blue) for two K values. **C-D)** α and b as functions of input current I_0. The nonlinear features of α and b are captured by the theory. The approximation for b works better with small currents (zoomed-in gray box). The trend (larger K produces smaller quenched variance and larger α) is also accounted for by the theory. Error bars indicate the 95% confidence interval around the mean, averaged over 10 random realizations of the connectivity. (Model parameters: $I_0 = 2$ in **A**, $g = J_0 = 1$, $J_{ij} \sim$ gamma, $\phi = \mathrm{Heaviside}$)

These results suggest that, despite non-Gaussian statistics, the sparse balance model is amenable to a mean-field treatment that is in similar spirit to what has been applied previously to recurrent networks.

Sparse balance in an E-I network

Finally, we illustrate that all of the features we have discussed for a purely inhibitory network are present in mixed excitatory-inhibitory networks for two choices (gamma and lognormal) of the connectivity distribution (Fig 6). These networks exhibit asynchronous irregular activity with chaotic responses of individual units (Fig 6A-B) and constant population activity (Fig 6C). Responses are sparse across the population: roughly 10% of excitatory and 20-30% of inhibitory units are active at any given time (Fig 6C). Individual units show sporadic response for both E and I cells in time (Fig 6B) with spatiotemporal variability that is purely internally generated.

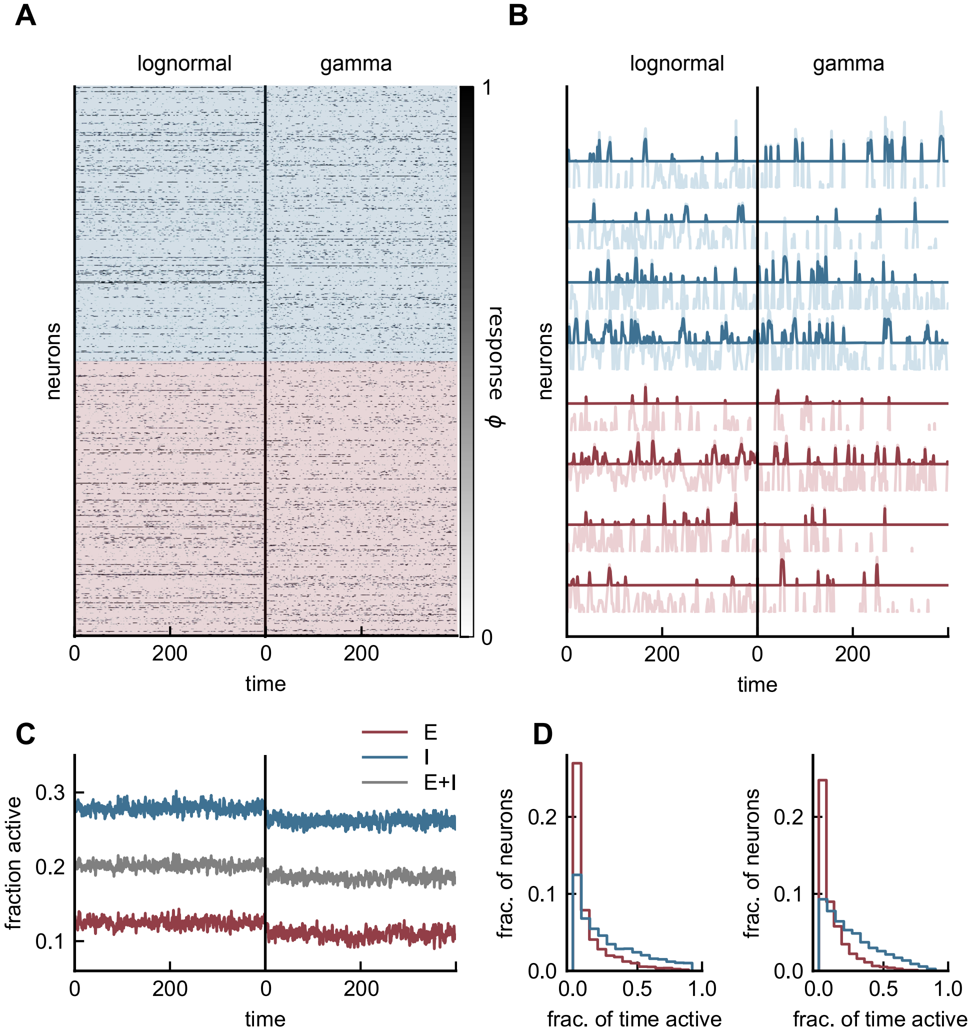

Fig 6. Asynchronous irregular activity in an E-I network with small input current **A)** Responses of 500 excitatory (red) and 500 inhibitory (blue) units in two networks, one with lognormal (left) and the other with gamma (right) weight distributions. Responses are sparse and distributed across the population. **B)** Rates (dark) superimposed on the currents (light) for four example cells from each population. Response is infrequent as fluctuations occasionally push the current above threshold. **C)** Fraction of active units for individual populations (red and blue) and across the entire network (gray). The inhibitory population is more active than its excitatory counterpart. ***(continued on next page)***

Fig 6. D) Fraction of (simulation) time units spend above threshold for each population and connectivity distribution. This distribution is wide and skewed. Both choices of the connectivity distribution produce qualitatively similar results. (Model parameters: $g = 1$, $J_{EE} = J_{IE} = 1$, $J_{EI} = 2$, $J_{II} = 1.2$, $I_E = 2$, $I_I = 1$, $N_E = N_I = 3000$, $K = 600$, $\phi = [\tanh]_+$).

Responses are robust and shared across the entire population as opposed to a fixed subset of units. We characterize this feature by considering the distribution of ON-time fraction, i.e., the fraction of time individual cells spend above threshold (Fig 6D). This quantity shows a wide and skewed distribution across both E and I populations. The majority of units spend very little time above the threshold, with only a few (5% of E, 20% of I cells with lognormal; 2% of E, 15% of I cells with gamma) spending more than half the time above threshold, and none responding at all times. We note that gamma and lognormal synaptic distributions produce similar activity patterns across the population.

Discussion

We have uncovered a novel regime of E-I networks that exhibits asynchronous irregular activity without the need for unrealistically large external input currents. We have done so by taking advantage of widely distributed synapses that generate fluctuations that would otherwise be minuscule in the absence of large feedforward currents. We highlighted a number of properties including sparse activity, non-Gaussian dynamics and a nonlinear population response. We also revealed the mechanism by which the time-scale of the dynamics generates sparse network activity. Using mean-field theory, we computed the statistical features of the recurrent input. This model demonstrates the important role of synaptic variance in the dynamics of recurrent networks.

Robust network responses with small input currents are especially interesting in light of the fact that experiments suggest the feedforward component of the input in cortical circuits is comparable in magnitude to the total synaptic input (see [9] for a review). For example, in olfactory cortex, the feedforward excitation from the olfactory bulb accounts for only a quarter of the net excitation into a pyramidal cell [10]; in the visual cortex, thalamic input accounts for roughly 30-40% of the net excitation to a cell [11, 12]. Similar results have been reported in the auditory cortex [21]. Our model provides a novel theoretical insight into these observations and highlights the importance of the scaling of the input current on network dynamics.

To provide a more quantitative link to these experimental findings, it is appropriate to define $\chi = \overline{x}/I_0$, referred to as the 'balanced index' [9]. The ratio χ captures the relative contribution of the feedforward input I_0 to the mean of the total current $\overline{x}$. The aforementioned experiments suggest a χ of order 1. In both the standard balanced and the low-variance networks result in $\chi \sim 1/\sqrt{K}$. In the sparse balance model, widely distributed synapses together with small input currents yield a χ of order 1 in agreement with experimental findings in cortex.

Despite the absence of cancellation of large excitatory and inhibitory currents in the sparse balance model, the mean of the net synaptic input, $\overline{x}$, lies well below threshold. One consequence of this, as mentioned above, is that the net input and the feedforward input have comparable contributions to the mean response. This results in a nonlinear population response to uniform input that is absent in the standard balanced regime where the strength of the feedforward input overwhelms the influence of the net input. Nonlinear mean responses are known to be necessary for a variety of cortical computations such as response normalization and surround suppression in visual cortex [9, 17, 22–24] and concentration invariance in olfactory cortex [25–27]. In the sparse balance model, the shape of this nonlinear population response depends on the choice of neuronal response function.

Small input currents impose a constraint on the mean response. To ensure this constraint is carried over to the sparsity, but not the mean response of the active neurons, we considered widely distributed synapses through an unconventional scaling of synaptic variance. Models that address the role of heavy-tailed connectivity distributions are timely because it has been shown that the distribution of synaptic efficacies in cortex are compatible with a lognormal fit [1, 28, 29]. Experiments and modeling studies have also suggested that strong synapses in the tail of such distributions, although less frequent, can have a strong influence on postsynaptic firing and network dynamics [30–33].

Our choice of the gamma distribution, as opposed to the lognormal, was motivated by its analytical tractability. The scaling of variance we consider results in an effectively sparse connectivity distribution where the majority of synapses are weak and thus neuronal activity is heavily influenced by the minority of strong synapses. Similar to this idea, recent modeling work has demonstrated that networks with power-law synaptic weights exhibit self-sustained activity [33]. In these 'Cauchy networks', the variance of the connectivity is infinite for finite K, and network behavior is dominated by large tails in the weight distribution. In another modeling study, a lognormal distribution of synaptic weights in a network of spiking neurons gave rise to self-sustained asynchronous firing in the absence of any bias input current [31]. Together with these results, our model highlights the degree to which heterogeneity in connectivity can compensate for the absence of large input currents and help sustain rich network dynamics.

The emergence of sparse activity in our model is interesting given that only a small fraction of cortical cells, particularly in the superficial layers, are active in response to many stimulus or spontaneously (see [34] for a review). The level of sparsity in the model is related to the degree of network connectivity as opposed to single-neuron properties, such as the threshold. Standard balanced models can also exhibit a high degree of population sparseness, as in the spiking models of [35]. In networks with intrinsic chaotic activity, whether and how the degree of sparsity can be used to perform computations that require high-dimensional representation [36, 37] of the stimulus remains to be investigated.

In our model, we revealed that the time-scale of the synaptic input, not its distribution, adjusts to maintain the degree of sparsity in the network. When the constraint on the mean response demands a small level of current above threshold, fluctuations at the synaptic input speed up; since the dynamic equation of each cell is a low-pass filter of its synaptic input, the output distribution narrows ever so slightly and the tail above threshold retreats in a manner that satisfies the constraint on the mean rate. Changes in sparsity are made possible by changes in the autocorrelation time-scale, and this phenomenon appears to apply generally. This feature highlights the flexibility of recurrent dynamics in adjusting not only the mean and variance of the distribution of its firing rates, but also their correlation time-scales. The role of this flexibility in building functional networks of rate neurons is a potential subject of future work.

Mean-field theory is an important tool for understanding network behavior. We examined a mean-field theory based on the approximation that synaptic fluctuations are instantaneously tracked by neuronal responses. This adequately predicted the network behavior and was extended to include both time-dependent and quenched fluctuations. This theory assumes a finite, but large in-degree K. Biological values of K are roughly of order 10^3, the range we considered. Specifically, pyramidal cells receive ~ 7000 excitatory synapses [1], but when we account for the average number of synapses per connection, ~ 4 [38, 39] and the number of non-intracortical synapses, the in-degree comes out to be $\sim 10^3$.

One prominent theory that addresses the issue of bias input is the stabilized supralinear network (SSN) [9, 24, 40]. Important features of SNNs include supralinear neuronal response function, small bias current, and weak synaptic coupling. These models have been extensively and successfully used to describe steady-state responses in sensory cortex. Our model differs from the SSN in that it generates chaotic activity, has strong synaptic coupling [41] and has widely distributed synaptic weights. With a supralinear response function ($\lambda > 1$), the large synaptic variance endows our model with a higher degree of chaos than other strongly-coupled networks; without this large degree of heterogenity in the weights, responses are susceptible to resting at fixed points [15]. This degree of chaos may aid in the learning of functional trajectories [42–46].

We examined a model with random connectivity, but it would be interesting to investigate stimulus selectivity in sparse balance networks with structured connections. The large degree of variability in the synapses could route stimulus information along particular paths across network neurons. Structured connectivity is of particular interest given compelling evidence that the recurrent contribution of the synaptic input, not just the feedforward component, exhibits selectivity [10–12]. We believe that the variance of connectivity, in addition to its mean structure, is important to consider for addressing the way feedforward and recurrent components shape selective responses.

Materials & methods

Numerical simulations

Numerical simulations were performed using Euler integration with time-steps less than 0.05, $\tau_x = 1$, and simulation time $T = 1000$. Other network parameters are included in the figure captions. The code (written in Julia v1.3.0) is available with the online version of the manuscript.

Non-Heaviside nonlinearities

Our analysis, leading to the high variance scaling ($\nu = 1/2$) is based on networks with a Heaviside ($\lambda = 0$) response function, which simplifies the calculations. Generally, the variance of the recurrent synaptic input is var$(\eta) = K$var$(J)\overline{\phi^2} \sim K^{1-\nu}\overline{\phi^2}$. With a Heaviside nonlinearity, $\overline{\phi^2} = \overline{\phi}$ and since $\overline{\phi} \sim 1/\sqrt{K}$, we conclude that $\nu = 1/2$ is the only solution with finite fluctuations as K grows. Non-binary responses do not guarantee the equivalence of $\overline{\phi^2}$ and $\overline{\phi}$, so we introduce $\psi(\lambda) \equiv \overline{\phi^2}/\overline{\phi}$. In the case of a Heaviside ($\lambda = 0$), $\psi = 1$. For $\lambda > 0$, the distribution of x dictates this ratio. We find numerically that, while not exactly one, ψ is slowly varying in K (Fig S3).

The variance scaling result obtained with a Heaviside nonlinearity extends to rectified $\tanh$ (and similarly rectified linear) in that the $1/\sqrt{K}$ scaling of $\overline{\phi}$ is predominantly inherited from f, and not μ. For $\lambda \geq 2$, this is not necessarily the case and μ can exhibit a non-negligible scaling with K. In spite of this, with a fixed value of K in the range of interest, $\sim 10^3$, the responses in the high-variance case, as opposed to its low-variance counterpart, are much more appreciable and irregular, and the low-variance model is prone to fixed point states for $\lambda > 1$.

Existence of the mean-field solution

From Eq 15,

$$m = \frac{1}{\Gamma(\alpha)} \int_0^{I_0/\theta} du\; e^{-u} u^{\alpha-1} = \tilde{\gamma}(\alpha, I_0/\theta)\,, \tag{25}$$

where $\tilde{\gamma}$ is the regularized lower incomplete gamma function. From Eq 14,

$$\alpha = \frac{J_0^2}{g^2}\sqrt{K}\,\tilde{\gamma}\left(\alpha, \frac{I_0 J_0}{g^2}\right). \tag{26}$$

Solving this equation for α produces the result shown in Fig 5B.

Combining Eqs (21) and (24), we obtain

$$\frac{b\theta^2}{g^2\sqrt{K}} = \frac{1}{\Gamma(b)} \int_0^{I_0/\theta} du\ e^{-u} u^{b-1} \Big(\tilde{\gamma}\Big(\alpha - b, \frac{I_0}{\theta} - u\Big)\Big)^2 \tag{27}$$

Here, we comment on the existence of a solution to Eq (27). On one hand, the maximal possible value of b is $b = \alpha$. At this extremum value, it follows from Eq (14) that the left-hand-side (LHS) of Eq (27) is equal to m. In this limit, the right-hand-side (RHS) of Eq (27) is also equal to m. This is because $\lim_{\epsilon \to 0} \tilde{\gamma}(\epsilon, u) = 1$. In other words,

$$\text{LHS}(b = \alpha) = \text{RHS}(b = \alpha) = m \tag{28}$$

This, however, is not a viable solution because $b = \alpha$ describes a fixed-point state with no temporal variability, rather than describing a chaotic state.

On the other hand, the minimal value of b is $b = 0$, and a similar manipulation of Eq (27) shows that RHS$(b = 0) = m^2$

$$\text{LHS}(b = 0) = 0 < \text{RHS}(b = 0) = m^2 \tag{29}$$

Therefore, a solution is guaranteed as long as the slope of the RHS at $b = \alpha$, $\partial(\text{RHS})/\partial b$, is larger than that of the LHS. These slopes depend on K, through the explicit appearance of K in the LHS of Eq (27) and through the dependence of the mean-field value of α on K. Numerically, we find that a solution exists for the K values of interest. Solving Eq (27) for b produces the result shown in Fig 5C.

Acknowledgments

We thank Ken Miller, Gregory Handy, Alessandro Ingrosso, and Rainer Engelken for helpful discussions. This research was supported by the NSF NeuroNex Award DBI-1707398 and The Gatsby Charitable Foundation.

References

1. Iascone DM, Li Y, Sümbül U, Doron M, Chen H, Andreu V, et al. Whole-neuron synaptic mapping reveals spatially precise excitatory/inhibitory balance limiting dendritic and somatic spiking. Neuron. 2020;.

2. Holt GR, Softky WR, Koch C, Douglas RJ. Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. Journal of neurophysiology. 1996;75(5):1806–1814.

3. Shadlen MN, Newsome WT. Noise, neural codes and cortical organization. Current opinion in neurobiology. 1994;4(4):569–579.

4. Tsodyks MV, Sejnowski T. Rapid state switching in balanced cortical network models. Network: Computation in Neural Systems. 1995;6(2):111–124.

5. Troyer TW, Miller KD. Integrate-and-fire neurons matched to physiological fI curves yield high input sensitivity and wide dynamic range. In: Computational Neuroscience. Springer; 1997. p. 197–201.

6. Vogels TP, Abbott LF. Signal propagation and logic gating in networks of integrate-and-fire neurons. Journal of neuroscience. 2005;25(46):10786–10795.

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

8. Vreeswijk Cv, Sompolinsky H. Chaotic balanced state in a model of cortical circuits. Neural computation. 1998;10(6):1321–1371.

9. Ahmadian Y, Miller KD. What is the dynamical regime of cerebral cortex? arXiv preprint arXiv:190810101. 2019;.

10. Poo C, Isaacson JS. A major role for intracortical circuits in the strength and tuning of odor-evoked excitation in olfactory cortex. Neuron. 2011;72(1):41–48.

11. Lien AD, Scanziani M. Tuned thalamic excitation is amplified by visual cortical circuits. Nature neuroscience. 2013;16(9):1315–1323.

12. Li Yt, Ibrahim LA, Liu Bh, Zhang LI, Tao HW. Linear transformation of thalamocortical input by intracortical excitation. Nature neuroscience. 2013;16(9):1324–1330.

13. Marshel JH, Kim YS, Machado TA, Quirin S, Benson B, Kadmon J, et al. Cortical layer–specific critical dynamics triggering perception. Science. 2019;365(6453):eaaw5202.

14. van Vreeswijk C, Sompolinsky H. Les Houches Lectures LXXX on Methods and models in neurophysics. Elsevier; 2005.

15. Kadmon J, Sompolinsky H. Transition to chaos in random neuronal networks. Physical Review X. 2015;5(4):041030.

16. Harish O, Hansel D. Asynchronous rate chaos in spiking neuronal circuits. PLoS computational biology. 2015;11(7):e1004266.

17. Sanzeni A, Histed MH, Brunel N. Response nonlinearities in networks of spiking neurons. PLoS computational biology. 2020;16(9):e1008165.

18. Kenney J, Keeping E. The distribution of the standard deviation. Mathematics of Statistics, Pt. 1951;2:170–173.

19. Sompolinsky H, Crisanti A, Sommers HJ. Chaos in random neural networks. Physical review letters. 1988;61(3):259.

20. Rajan K, Abbott L, Sompolinsky H. Stimulus-dependent suppression of chaos in recurrent neural networks. Physical Review E. 2010;82(1):011903.

21. Li Ly, Li Yt, Zhou M, Tao HW, Zhang LI. Intracortical multiplication of thalamocortical signals in mouse auditory cortex. Nature neuroscience. 2013;16(9):1179–1181.

22. Carandini M, Heeger DJ. Normalization as a canonical neural computation. Nature Reviews Neuroscience. 2012;13(1):51–62.

23. Angelucci A, Bijanzadeh M, Nurminen L, Federer F, Merlin S, Bressloff PC. Circuits and mechanisms for surround modulation in visual cortex. Annual review of neuroscience. 2017;40:425–451.

24. Rubin DB, Van Hooser SD, Miller KD. The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. Neuron. 2015;85(2):402–417.

25. Stettler DD, Axel R. Representations of odor in the piriform cortex. Neuron. 2009;63(6):854–864.

26. Bolding KA, Franks KM. Complementary codes for odor identity and intensity in olfactory cortex. Elife. 2017;6:e22630.

27. Roland B, Deneux T, Franks KM, Bathellier B, Fleischmann A. Odor identity coding by distributed ensembles of neurons in the mouse olfactory cortex. Elife. 2017;6:e26337.

28. Song S, Sjöström PJ, Reigl M, Nelson S, Chklovskii DB. Highly nonrandom features of synaptic connectivity in local cortical circuits. PLoS Biol. 2005;3(3):e68.

29. Loewenstein Y, Kuras A, Rumpel S. Multiplicative dynamics underlie the emergence of the log-normal distribution of spine sizes in the neocortex in vivo. Journal of Neuroscience. 2011;31(26):9481–9488.

30. Lefort S, Tomm C, Sarria JCF, Petersen CC. The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex. Neuron. 2009;61(2):301–316.

31. Teramae Jn, Tsubo Y, Fukai T. Optimal spike-based communication in excitable networks with strong-sparse and weak-dense links. Scientific reports. 2012;2(1):1–6.

32. Ikegaya Y, Sasaki T, Ishikawa D, Honma N, Tao K, Takahashi N, et al. Interpyramid spike transmission stabilizes the sparseness of recurrent network activity. Cerebral Cortex. 2013;23(2):293–304.

33. Kuśmierz Ł, Ogawa S, Toyoizumi T. Edge of Chaos and Avalanches in Neural Networks with Heavy-Tailed Synaptic Weight Distribution. Physical Review Letters. 2020;125(2):028101.

34. Barth AL, Poulet JF. Experimental evidence for sparse firing in the neocortex. Trends in neurosciences. 2012;35(6):345–355.

35. Pehlevan C, Sompolinsky H. Selectivity and sparseness in randomly connected balanced networks. PLoS One. 2014;9(2):e89992.

36. Babadi B, Sompolinsky H. Sparseness and expansion in sensory representations. Neuron. 2014;83(5):1213–1226.

37. Barak O, Rigotti M, Fusi S. The sparseness of mixed selectivity neurons controls the generalization–discrimination trade-off. Journal of Neuroscience. 2013;33(9):3844–3856.

38. Feldmeyer D, Egger V, Lübke J, Sakmann B. Reliable synaptic connections between pairs of excitatory layer 4 neurones within a single 'barrel'of developing rat somatosensory cortex. The Journal of physiology. 1999;521(1):169–190.

39. Gal E, London M, Globerson A, Ramaswamy S, Reimann MW, Muller E, et al. Rich cell-type-specific network topology in neocortical microcircuitry. Nature neuroscience. 2017;20(7):1004.

40. Ahmadian Y, Rubin DB, Miller KD. Analysis of the stabilized supralinear network. Neural computation. 2013;25(8):1994–2037.

41. Barral J, Reyes AD. Synaptic scaling rule preserves excitatory–inhibitory balance and salient neuronal network dynamics. Nature neuroscience. 2016;19(12):1690–1696.

42. Sussillo D, Abbott LF. Generating coherent patterns of activity from chaotic neural networks. Neuron. 2009;63(4):544–557.

43. DePasquale B, Cueva CJ, Rajan K, Escola GS, Abbott L. full-FORCE: A target-based method for training recurrent networks. PloS one. 2018;13(2):e0191527.

44. Schuecker J, Goedeke S, Helias M. Optimal sequence memory in driven random networks. Physical Review X. 2018;8(4):041029.

45. Toyoizumi T, Abbott L. Beyond the edge of chaos: Amplification and temporal integration by recurrent networks in the chaotic regime. Physical Review E. 2011;84(5):051908.

46. Legenstein R, Maass W. Edge of chaos and prediction of computational performance for neural circuit models. Neural networks. 2007;20(3):323–334.

47. Chung S, Ferster D. Strength and orientation tuning of the thalamic input to simple cells revealed by electrically evoked cortical suppression. Neuron. 1998;20(6):1177–1189.

48. Abramowitz M, Stegun IA, Romer RH. Handbook of mathematical functions with formulas, graphs, and mathematical tables; 1988.

49. Renart A, De La Rocha J, Bartho P, Hollender L, Parga N, Reyes A, et al. The asynchronous state in cortical circuits. science. 2010;327(5965):587–590.

50. Kuśmierz Ł, Ogawa S, Toyoizumi T. Edge of Chaos and Avalanches in Neural Networks with Heavy-Tailed Synaptic Weight Distribution. Physical Review Letters. 2020;125(2):028101.

51. Bertschinger N, Natschläger T. Real-time computation at the edge of chaos in recurrent neural networks. Neural computation. 2004;16(7):1413–1436.

Supplementary figures

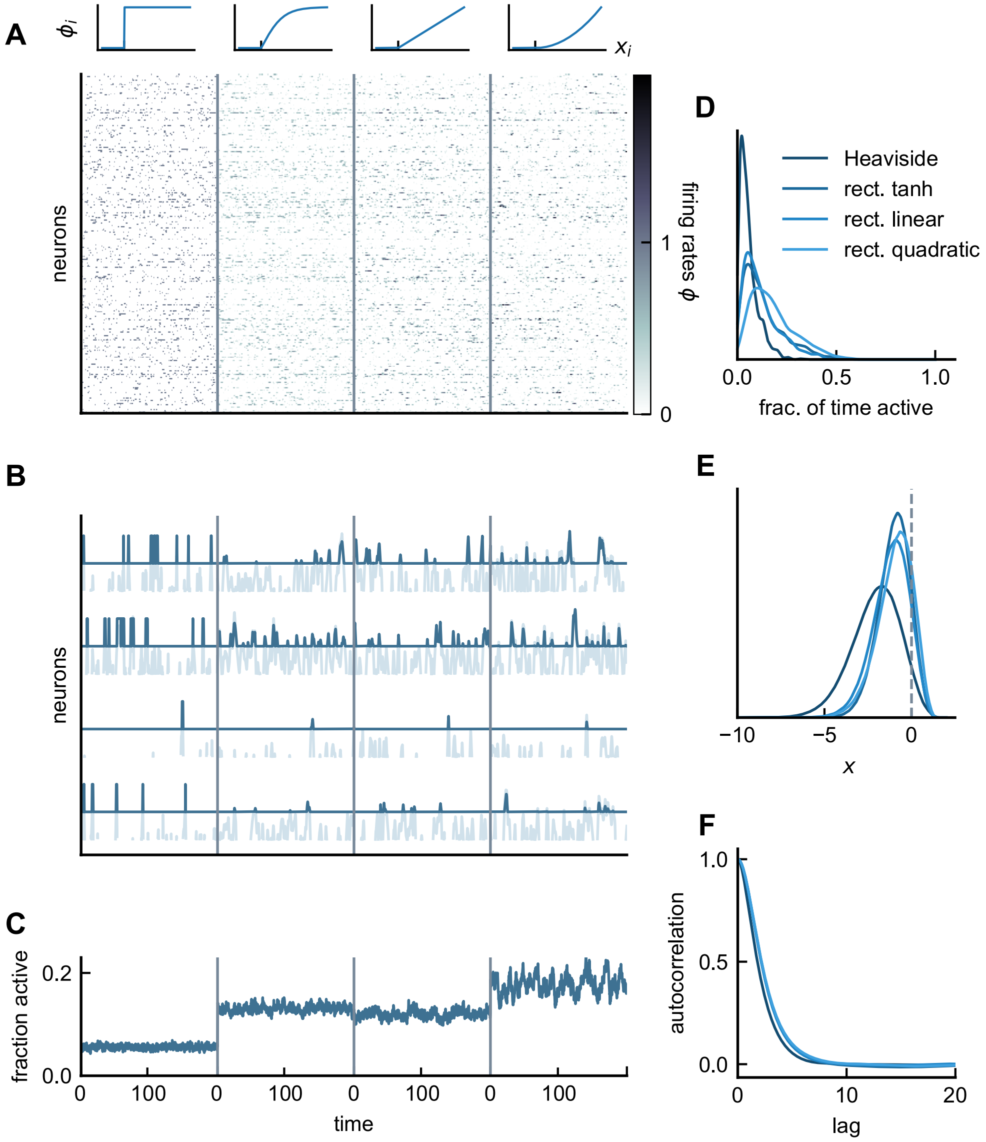

Fig S1. Asynchronous irregular activity in the sparse balance model with binary weights. Same as Fig 3, except with a Bernoulli connectivity distribution of mean $J_0/\sqrt{N}$. (Model parameters: $J_0 = 2$, $g = \sqrt{J_0(1 - J_0/\sqrt{N})}$, $I_0 = 1.5$, $N = 1000$).

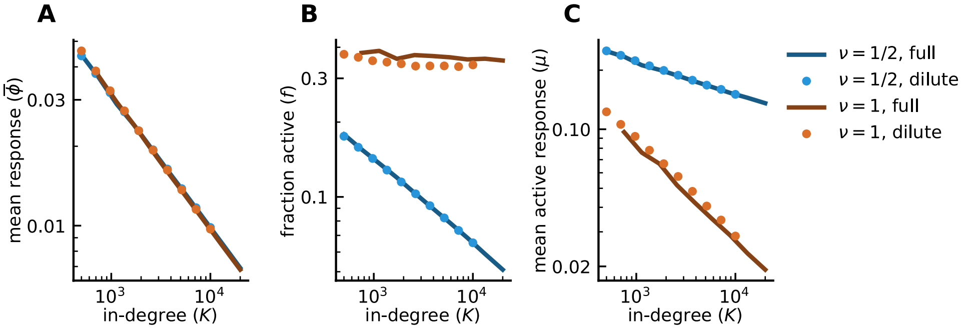

Fig S2. Equivalence between dilute and full connectivity. A-C) Same as Fig 1C-E, but with the addition of results from a dilutely-connected network (dots). With dilute connectivity, the source of variability in the connections are twofold: each neuron receives inputs from, on average, K other neurons out of the total N network neurons; additionally, each existing connection is drawn from a distribution of mean $J_0/\sqrt{K}$ and variance $g^2/\sqrt{K}$. In the fully-connected case, each neuron receives input from all other neurons with connections drawn from a distribution of mean $J_0/\sqrt{N}$ and variance $g^2/\sqrt{N}$. Note that in the main text, we considered fully-connected networks and denoted the mean and variance by $J_0/\sqrt{K}$ and $g^2/\sqrt{K}$ with $K = N$. (Model parameters: $g = J_0 = 2$, $I_0 = 1$, $J_{ij} \sim \text{gamma}, \phi = [\tanh]_+$; $N = K$ in full (solid), $N = 20000$ in dilute (dotted)).

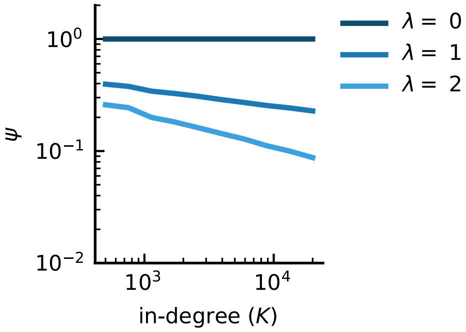

Fig S3. Slow variations in ψ for various nonlinear response functions. $\psi(\lambda)$, defined as $\overline{\phi^2}/\overline{\phi}$, exhibits sub-power law scalings with K. (Model parameters: $g = J_0 = 2$, $I_0 = 1$, $J_{ij} \sim$ gamma, $\phi = [x]_+^\lambda$)

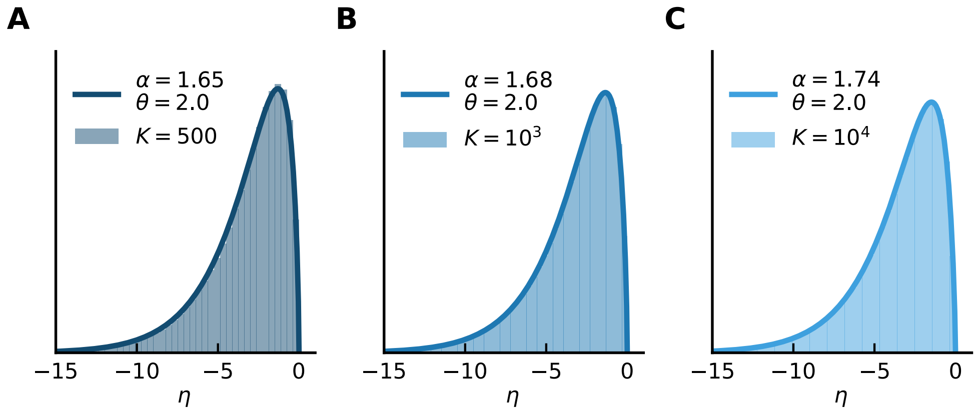

Fig S4. Recurrent synaptic input is described by a gamma distribution. A-C) The distribution of the synaptic input η (across population and time) with a Heaviside response nonlinearity for three different values of K (shaded histograms). The solid line is the gamma distribution in Eq (8) with scale parameter $\theta = g^2/J_0$ and shape parameter $\alpha = f\sqrt{K}J_0^2/g^2$, where f is the measured sparsity in the simulations, and J_0, g, K are network parameters. The distribution accurately describes the histograms. Also note that, due to the high J-variance in the sparse balance model, the distributions of synaptic input hardly change as K increases. (Model parameters: $g = J_0 = 2$, $I_0 = 1$, $J_{ij} \sim$ gamma, $\phi = \text{Heaviside}$)

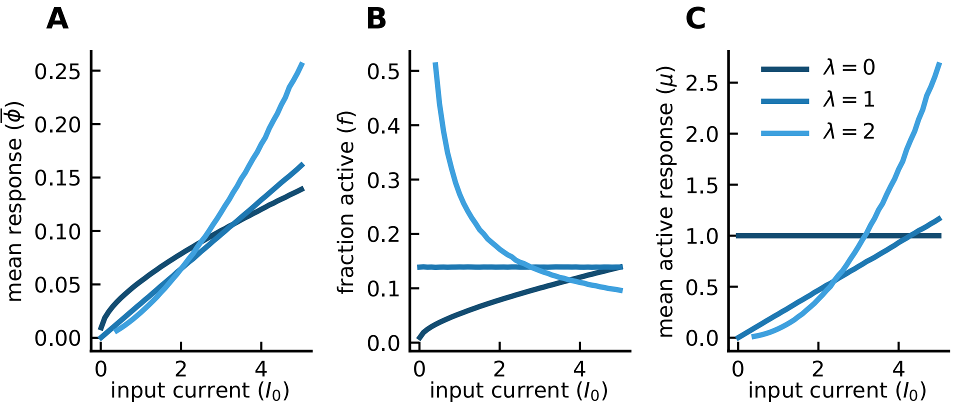

Fig S5. Sparse balance responds nonlinearly to input current. A) Mean response, $\bar{\phi} = f\mu$, increases nonlinearly with input current. This relationship depends on the shape of the neuronal response function: $\lambda < 1$ and $\lambda > 1$ give rise to sublinear and supralinear mean responses, respectively. **B-C)** Fraction, f, and mean response, μ, of the active neurons versus input current. With rectified linear ($\lambda = 1$), this fraction remains constant since x can be rescaled by I_0^{-1} without changing the shape of the x distribution; this feature also makes the mean response linear. The nonlinear trend in sparsity switches at $\lambda = 1$. The fraction active f increases with I_0 at an ever-decreasing rate with $\lambda > 1$, while the opposite is true for $\lambda < 1$. For $\lambda > 1$, with stronger feedforward excitation, threshold crossings that result in sufficiently large responses become amplified. This amplification produces a large, supralinear μ, which in turn comes at the cost of sparsening the population activity with I_0. For the Heaviside ($\lambda = 0$), responses are binary, so $\mu = 1$ independent of I_0 and $\bar{\phi} = f$. (Model parameters: $g = J_0 = 2$, $J_{ij} \sim$ gamma, $\phi = [x]_+^\lambda$, $K = 1000$)

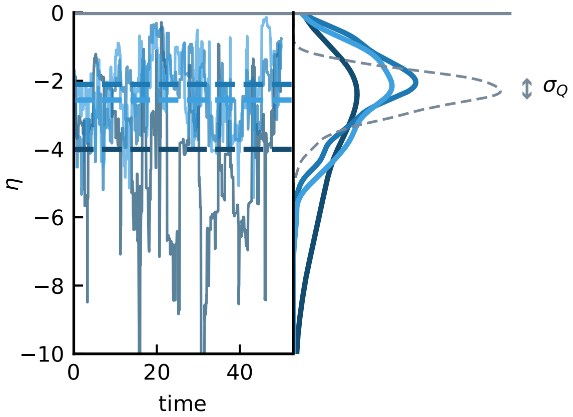

Fig S6. Distribution of time-averaged activities. Recurrent synaptic inputs of three example neurons with their corresponding distributions on the right (solid curves). Due to quenched disorder in the network, each neuron fluctuates around a different time-average (horizontal dashed lines). These time-averages form a distribution shown on the right (gray dashed curve) whose width is captured by $\sigma_Q = \theta\sqrt{b}$ and is approximated by a gamma distribution. This approximation is the main source of error in Fig 5C. (Model parameters: $g = J_0 = I_0 = 1$, $K = 1000$, $J_{ij} \sim$ gamma, $\phi = \mathrm{Heaviside}$)